Systems Biology and Livestock Science

Systems Biology and Livestock Science

Edited by

Marinus F.W. te Pas
Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, PO Box 65, 8200 AB Lelystad, The Netherlands

Henri Woelders
Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, PO Box 65, 8200 AB Lelystad, The Netherlands

André Bannink
Centre of Animal Nutrition, Wageningen UR Livestock Research, PO Box 65, 8200 AB Lelystad, The Netherlands

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

LIVESTOCK RESEARCH
WAGENINGEN UR

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

Registered office: John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial offices: 2121 State Avenue, Ames, Iowa 50014-8300, USA
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK
9600 Garsington Road, Oxford, OX4 2DQ, UK

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data

Systems biology and livestock science / edited by Marinus F.W. te Pas, André Bannink, Henri Woelders.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-8138-1174-1 (hardcover : alk. paper)
ISBN-10: 0-8138-1174-0
1. Veterinary physiology–Computer simulation. 2. Livestock. 3. Livestock–Genetics–Computer simulation. I. Pas, M. F. W. te. II. Bannink, André. III. Woelders, Henri.
SF768.S97 2011
636.089′2–dc23

2011020692

A catalogue record for this book is available from the British Library.

This book is published in the following electronic formats: ePDF [9780470962930]; Wiley Online Library [9780470963012]; ePub [9780470962961]; Mobi [9780470962992]

Set in 10/12 pt Dutch801BT by Aptara® Inc., New Delhi, India

1 2011

Contents

Color plates appear between pages 160 and 161.

All chapter abstracts are available on Wiley Online Library. Please visit Wiley Online Library at http://onlinelibrary.wiley.com/.

List of Contributors

André Bannink
Centre of Animal Nutrition
Wageningen UR Livestock Research
PO Box 65
8200 AB Lelystad, The Netherlands

Stefan Bauersachs
Laboratory for Functional Genome Analysis (LAFUGA) and Chair for Molecular Animal Breeding and Biotechnology
Gene Center of the Ludwig-Maximilians Universität München
Feodor Lynen Street 25
81377 Munich, Germany

Massimo Bionaz
Mammalian NutriPhysioGenomics
Department of Animal Sciences and Division of Nutritional Sciences
University of Illinois
Urbana, IL 61801, USA

Fred C. Boogerd
Molecular Cell Physiology
VU University Amsterdam
de Boelelaan 1085
1081 HV, Amsterdam, The Netherlands

Frank J. Bruggeman
Regulatory Networks Group
Netherlands Institute of Systems Biology
Amsterdam, The Netherlands

Life Sciences
Centre for Mathematics and Computer Science (CWI) Science Park 123
1098 XG, Amsterdam, The Netherlands

Leo K. Cheng
Auckland Bioengineering Institute
The University of Auckland
Private Bag 92019
Auckland 1142
New Zealand

Willem M. de Vos
Laboratory of Microbiology
Human and Animal Physiology
Wageningen University Dreijenplein 10
6710 HB Wageningen, The Netherlands

Department of Veterinary Biosciences
University of Helsinki
PO Box 66
00014 University of Helsinki, Finland

Jan Dijkstra
Animal Nutrition Group
Wageningen University
PO Box 338
6700 AH Wageningen, The Netherlands

Peng Du
Auckland Bioengineering Institute
The University of Auckland
Private Bag 92019
Auckland 1142
New Zealand

James France
Centre for Nutrition Modelling
Department of Animal and Poultry Science
University of Guelph
Guelph, ON, N1G 2W1, Canada

Arend J.W. Hoekman
Animal Breeding and Genomics Centre
Wageningen UR Livestock Research
PO Box 65
8200 AB Lelystad, The Netherlands

Vasant G. Honavar
Center for Integrated Animal Genomics
Iowa State University
Ames, IA 50011, USA

Bioinformatics and Computational Biology Program
Iowa State University
Ames, IA 50011, USA
Department of Computer Science
Iowa State University
Ames, IA 50011, USA

Artificial Intelligence Research Laboratory
Iowa State University
Ames, IA 50011, USA

Center for Computational Intelligence, Learning, and Discovery
Iowa State University
Ames, IA 50011, USA

Guido Hooiveld
Nutrition Metabolism and Genomics Group
University of Wageningen
Wageningen, The Netherlands

Ina Hulsegge
Animal Breeding and Genomics Centre
Wageningen UR Livestock Research
PO Box 65
8200 AB Lelystad, The Netherlands

Guido Invernizzi
Mammalian NutriPhysioGenomics
Department of Animal Sciences and Division of Nutritional Sciences
University of Illinois
Urbana, IL 61801, USA

Jaap Keijer
Department of Human and Animal Physiology
Wageningen University
Wageningen, The Netherlands

RIKILT-Institute of Food Safety
Wageningen, The Netherlands

Gordon M. Kirby
Department of Biomedical Sciences
Ontario Veterinary College
Guelph, Ontario, N1G 2W1, Canada

Alexey N. Kolodkin
Molecular Cell Physiology
VU University Amsterdam
de Boelelaan 1085
1081 HV, Amsterdam, The Netherlands

Juan J. Loor
Mammalian NutriPhysioGenomics
Department of Animal Sciences and Division of Nutritional Sciences
University of Illinois
Urbana, IL 61801, USA

Vitor A.P. Martins dos Santos
Laboratory of Systems and Synthetic Biology
University of Wageningen
Wageningen, The Netherlands

Michael Müller
Nutrition Metabolism and Genomics Group
University of Wageningen
Wageningen, The Netherlands

Gregory O'Grady
Department of Surgery and Auckland Bioengineering Institute
The University of Auckland
Private Bag 92019
Auckland 1142
New Zealand

Dirkjan Schokker
Wageningen UR Livestock Research
Animal Breeding and Genomics Centre
PO Box 65
8200 AB Lelystad, The Netherlands

Mari A. Smits
Animal Breeding and Genomics Centre
Wageningen UR Livestock Research
PO Box 65
8200 AB Lelystad, The Netherlands

Marinus F.W. te Pas
Animal Breeding and Genomics Centre
Wageningen UR Livestock Research
PO Box 65
8200 AB Lelystad, The Netherlands

Fadi Towfic
Bioinformatics and Computational Biology Program
Iowa State University
Ames, IA 50011, USA

Department of Computer Science
Iowa State University
Ames, IA 50011, USA

Artificial Intelligence Research Laboratory
Iowa State University
Ames, IA 50011, USA

Center for Computational Intelligence, Learning, and Discovery
Iowa State University
Ames, IA 50011, USA

Christopher K. Tuggle
Center for Integrated Animal Genomics
Iowa State University
Ames, IA 50011, USA

Department of Animal Science
Iowa State University
Ames, IA 50011, USA

Bioinformatics and Computational Biology Program
Iowa State University
Ames, IA 50011, USA

Center for Computational Intelligence, Learning, and Discovery
Iowa State University
Ames, IA 50011, USA

Peter van Baarlen
Department of Host-Microbe Interactomics
University of Wageningen
6700 AH, Wageningen, The Netherlands

Jerry Wells
Department of Host-Microbe Interactomics
University of Wageningen
6700 AH, Wageningen, The Netherlands

Hans V. Westerhoff
Molecular Cell Physiology
VU University Amsterdam
de Boelelaan 1085
1081 HV, Amsterdam, The Netherlands

Manchester Centre for Integrative Systems Biology Manchester Interdisciplinary Biocentre
University of Manchester
131 Princess Street
Manchester, M1 7DN, UK

Henri Woelders
Animal Breeding and Genomics Centre
Wageningen UR Livestock Research
PO Box 65
8200 AB Lelystad, The Netherlands

Eckhard Wolf
Laboratory for Functional Genome Analysis (LAFUGA) and Chair for Molecular Animal Breeding and Biotechnology
Gene Center of the Ludwig-Maximilians Universität München
Feodor Lynen Street 25
81377 Munich, Germany

Preface

From livestock production to biological science: from systems biology to livestock production

Biological processes underlie livestock production, many of them only broadly investigated or even unknown. Understanding regulation of livestock production capacity and related traits requires an understanding of life itself. It has been a long waited goal of the biological science to understand life in all its complexity. However, it is also recognized for long that life is too complex for the research with the aim to get an understanding of life to be performed *in situ*. For this reason, biology has broken up research into increasingly smaller parts and specializations. Each organ, tissue, and cell-type was investigated separately, from the morphology of the entire organ down to the molecular level in specific cell types. Simply, the summing up of all these scientific efforts still does not explain life in all its facets and organizational complexity. In fact, the complete system is more than the sum of its parts. Insight into the numerous interactions between molecules, within and between cells, within and between tissues, within and between organs, within and between organisms, and perhaps also levels of interactions not yet discovered, has to merge with research on the smaller parts to approximate life. It should be taken into account that the isolated study of the different parts may even have resulted in erroneous conclusions due to omission of signals related to these interactions.

The genomics revolution provided the sequence of whole genomes. The analyses of the sequence led to the discovery of (almost) all genes, although the function of most genes is still not understood, or partly understood at best. Fortunately, the genomics revolution supplied biology with amazing new tools enabling the investigation of the expression levels of all genes, proteins, and metabolites in a cell, tissue, or organ at the same time. Thus, interactions between genes and their environment were taken into account for the expression levels. These techniques allowed defining and describing the cellular components in large detail and completeness, but again life was more complex than the summing up of all "omics" data.

Parallel to the genomics revolution, a computer technology revolution enabled the analysis of increasingly large datasets. This enables integration of the datasets resulting from all different levels of research. Such an analysis can produce a more complete description of life that includes the knowledge obtained at separate levels of biological organization *and* the interactions within and between these diverse levels. This picture is still far too complex to understand at this moment. Therefore, to help understand life and to make testable predictions of complex parts of life, quantitative modeling approaches are necessary. While mathematical equations are based upon the results of investigations, their combination in more complex models can predict new complex patterns of life, and interactions between various elements of life, that were outside the scope of previous research. By doing so, they make these areas accessible to research thereby verifying and extending the possibilities for a mathematical representation of life.

These developments together initiated this whole new phase in biology together called "systems biology." Systems biology provides the opportunity to obtain a wider and deeper understanding of life because of the incorporation of the knowledge at all separate levels of investigation, including the mathematical representation of the complex interactions between these levels. In this way, the accuracy and completeness of that understanding may be advanced further in an iterative process, as the modeling produces testable predictions, which can be tested empirically, followed by refinement of the modeling, leading to closer understanding of the organization of life.

As said, biological processes underlie livestock production, most of them only broadly investigated and mostly from a practical viewpoint. Research has already shown that most quantitative (production) traits depend on both the genotype and on interactions of the genotype with the environment. The genetic background for a number of traits has been partly uncovered. There are interactions between the genetic background and food (components), animal handling (stress), or temperature and housing that (partly) have been quantified. Also, complex interactions between traits have been shown and a main goal in livestock production science is to improve wanted traits without compromising the basic animal needs and requirements, without deterioration of essential physiological traits negatively affecting other important traits (e.g., fertility), and without inducing problems such as leg weakness, stress, and other (acute as well as long-term) health and welfare problems. Such an aim requires full understanding of the essence of life, including the physiological processes of traits and the interactions at various levels of organization of this physiology. Mathematical modeling of the known processes and of the known interactions between them will increase our understanding of the regulation and mutual dependence of both production (efficiency, product quality) and other traits (robustness, fertility, health well-being) of livestock animals. The testable predictions of the mathematical models enable real progress toward better foods for healthier consumers in a healthy animal with a high well-being, and with less impact on the environment. Systems biology has the potential to make a large contribution to the long-range goals of livestock production science.

At the moment, systems biology is still in its infancy in livestock science research in particular. However, it has high potential in stimulating new directions of research and development of new concepts and new ways of thinking about familiar problems. This book describes several aspects of system biology and gives research examples from other biological disciplines, including simple model organisms and human medicine. The authors draw lines from their own research toward livestock science. We, the editors, hope and expect that this book will introduce systems biology to a wide audience involved with livestock science. We thank especially the publisher, Wiley-Blackwell, and the Executive Editor Mr. Justin Jeffryes, who initiated and stimulated the work on this book. We also thank all contributors to this book. The editors acknowledge the financial contribution of the Ministry of Agriculture and Nature Management through the IP/OP grant KB-04-004-049. We also thank the head of our department Dr. Ir. Roel F. Veerkamp for enabling and stimulating the work on this book.

The Editors
Dr. Marinus F.W. te Pas
Dr. Henri Woelders
Dr. André Bannink

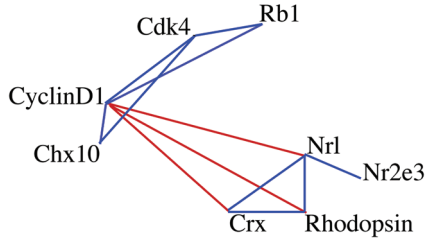


Figure 1.3 Seed network reconstructed from multiple gene expression datasets. Blue lines indicate positive correlations and red lines indicate negative correlations (Hecker et al., 2008).

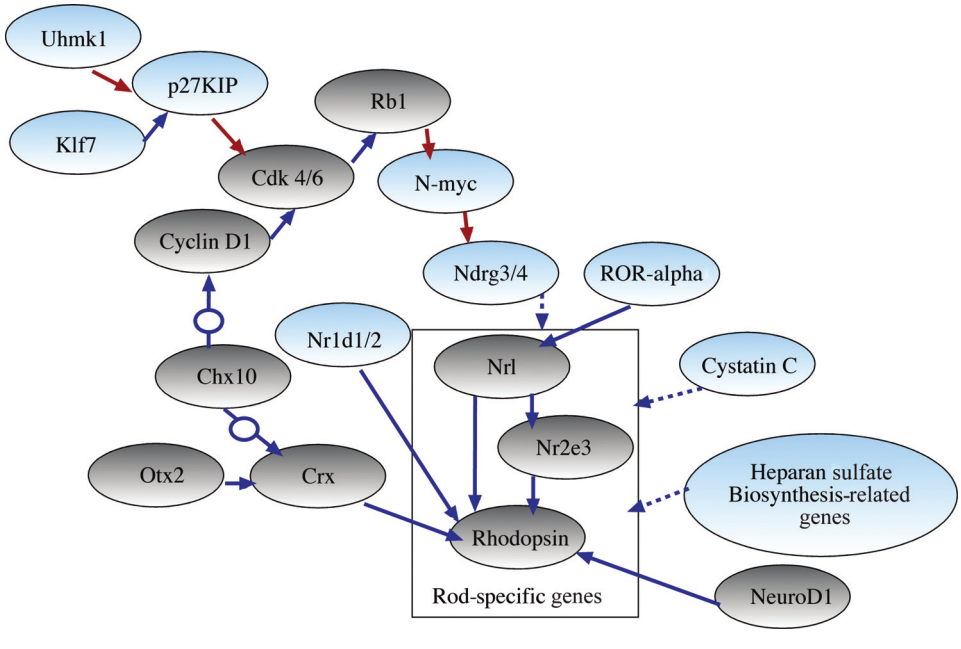


Figure 1.4 Expanded network. Original seed network nodes are shown in gray, and added network nodes are shown in blue. Blue arrows correspond to excitatory interactions, and red arrows correspond to negative interactions. Direct influences are shown using solid lines and indirect ones (through possible intermediaries) using dotted lines (Hecker et al., 2008).

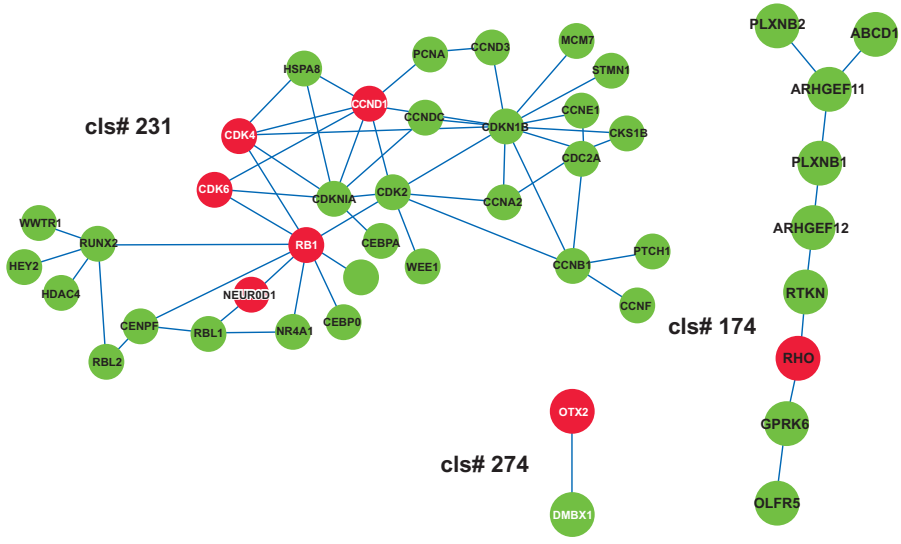


Figure 1.5 Some representative clusters within the mouse protein–protein interaction network obtained by spectral clustering. Seed network genes are shown in red.

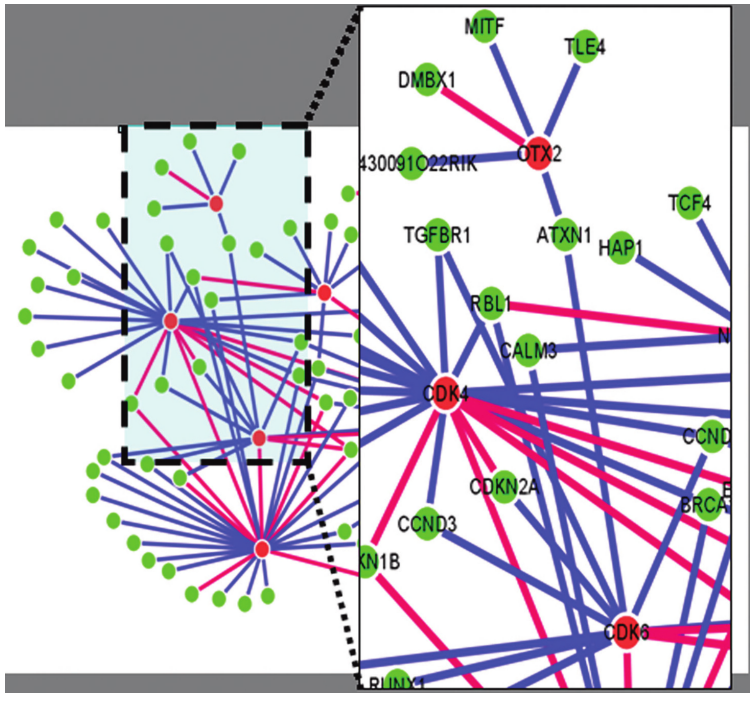


Figure 1.6 A fragment of the mouse protein–protein interaction network expanded on the basis of the human counterpart. Seed network mouse proteins and the links in the mouse protein–protein interaction network are shown in red. Nodes and links added based on orthologs in the human network are shown in green and blue, respectively. Note that two proteins that were not initially linked in the mouse Otx2 and CDK6 are now connected through ATXN1.

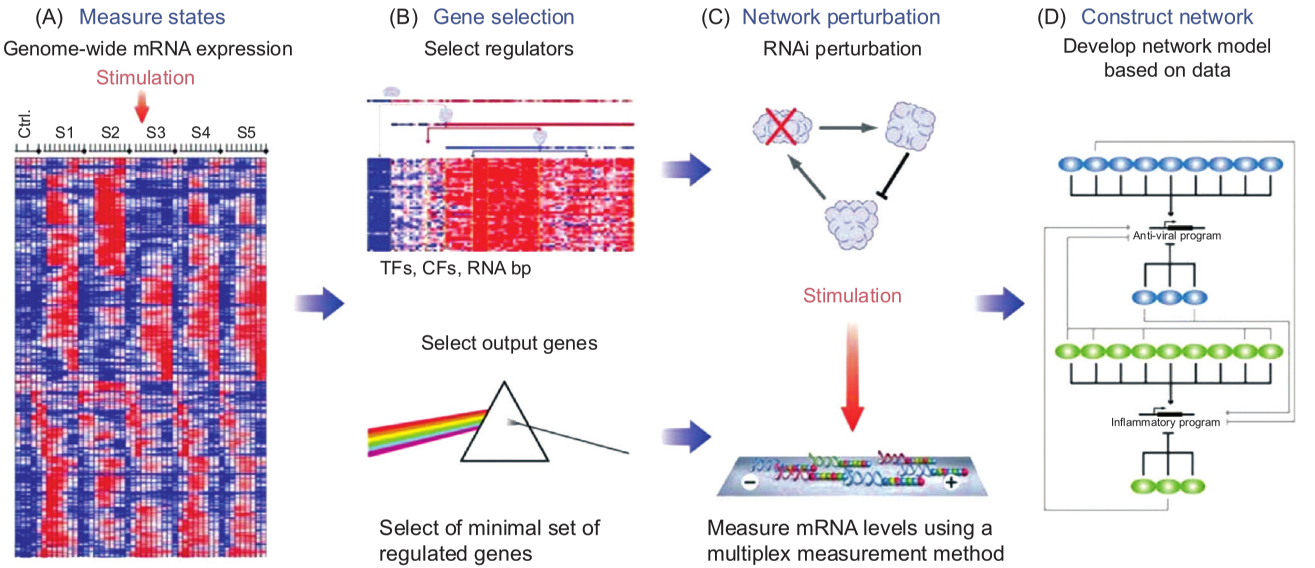


Figure 1.7 An example of an unbiased and systematic strategy for regulatory network reconstruction. Amit et al. (2009) strategy consists of four steps. (A) State measurement. They used genome-wide expression profiles under different stimuli (S1–S5), at different time points (tick marks). Rows—genes; columns—experiments; red—induced; blue—repressed; white—unchanged. (B) Gene selection. Amit et al. (2009) then identify candidate regulators that are transcriptionally regulated and predictive of the expression of gene modules (top) and select a signature of target genes that maximally represents the full expression profile (bottom). (C) Network perturbation. They then generated a functionally validated shRNA library for all potential regulators and used it to knockdown each regulator (top). Following stimulation of genetically perturbed cells (red arrow), Amit et al. (2009) measured the expression of the signature genes using the nCounter multiplex mRNA detection system (bottom). (D) Network reconstruction. Combining genome-wide expression profiles and perturbed multiplex measurements was performed to reconstruct a regulatory network associating regulators with individual targets and overall responses. (Adapted from Amit et al., (2009) *Science* **326**, 257–263. Reprinted with permission from AAAS.)

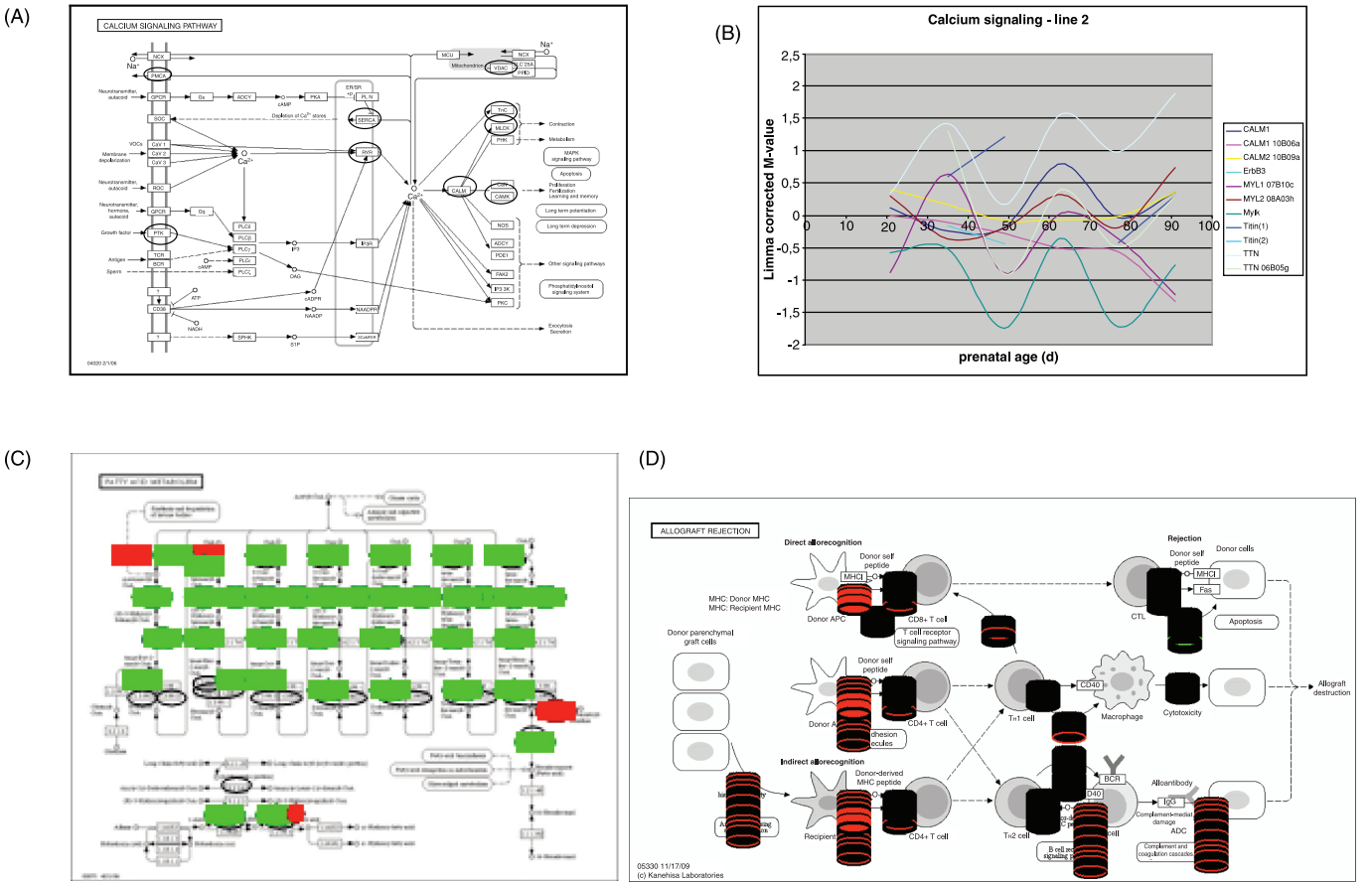


Figure 6.3 Examples visualizing experimental results investigating complex traits on biological pathways. (A, B). Using a pathway downloaded from the KEGG database first the genes under investigation are encircled (A) and the expression pattern of each gene in the time is indicated (B). Most genes of the pathway have similar regulated expression patterns in the time. (C) Alternatively, the genes in the pathway can be color coded: green color denotes up regulation, red color denotes down regulation of the gene expression in the present state of the complex trait as compared to another state. (D) Similar as (C), but each investigated animal is shown as a circle around the gene and each animal underneath the circle of the previous animal.

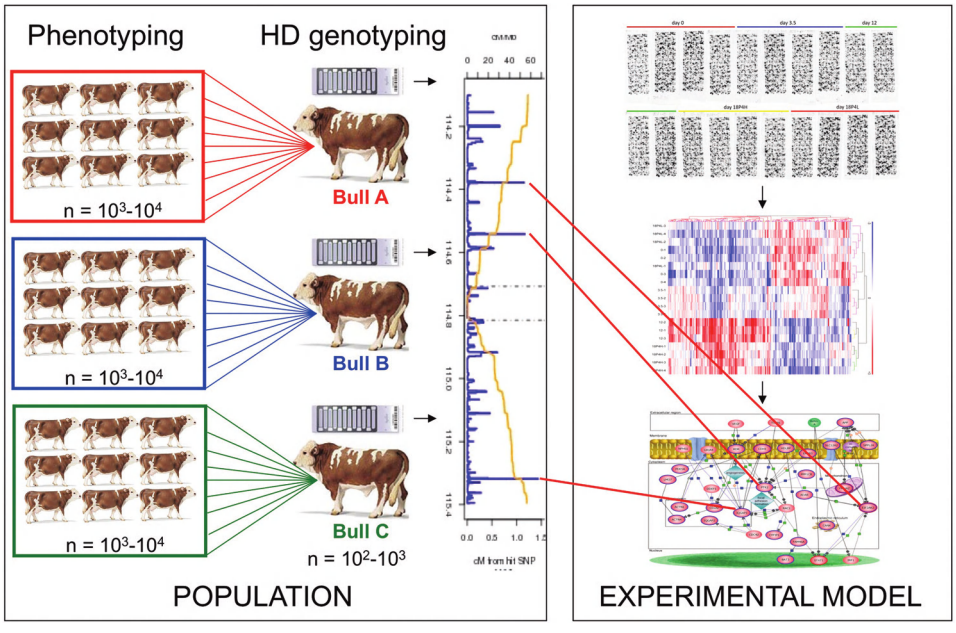


Figure 7.2 Linking data from gene expression and genome-wide association studies.

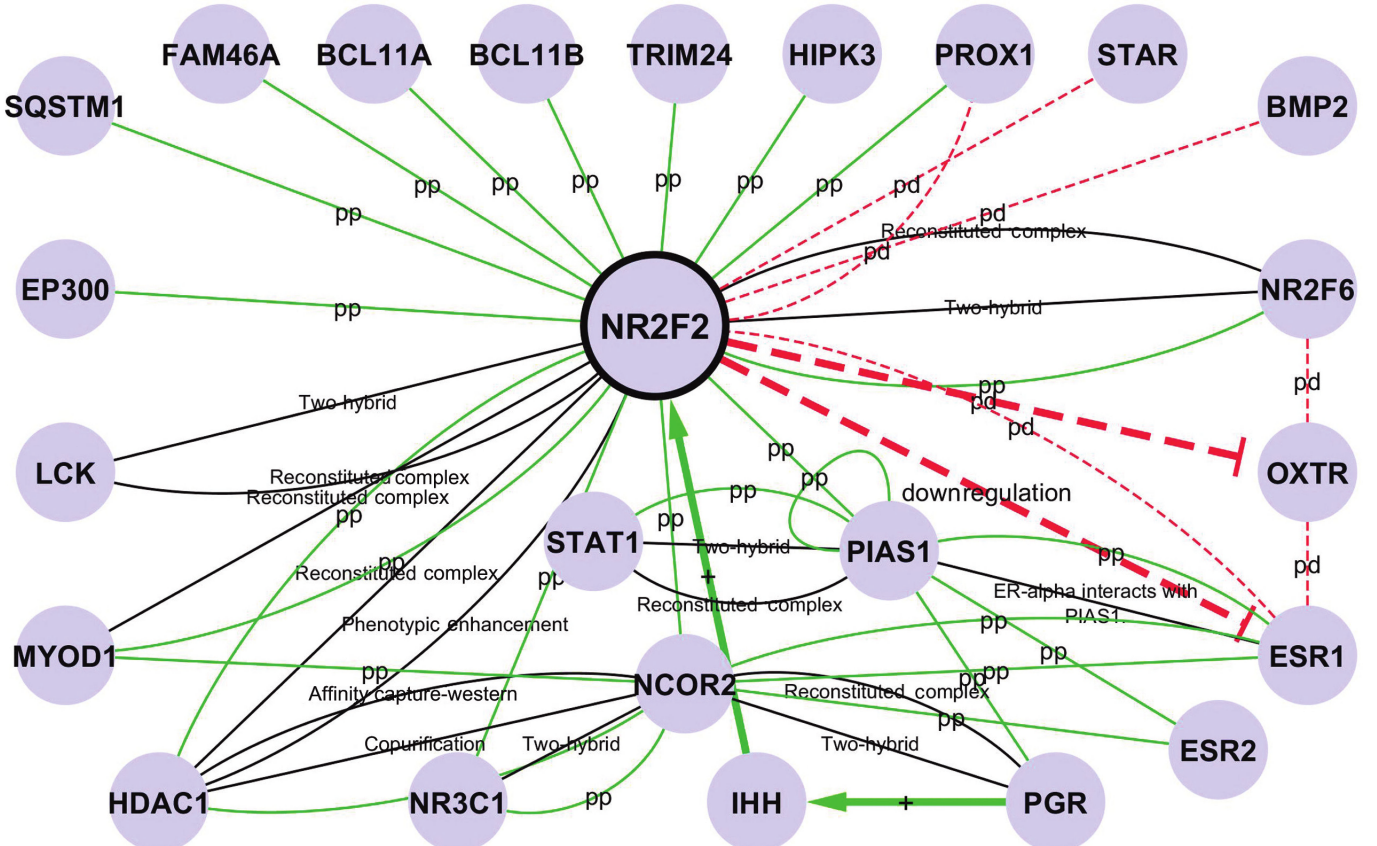


Figure 7.3 Interaction network for the essential transcription factor nuclear receptor subfamily 2, group F, member 2 (NR2F2). Interactors are shown with their official gene symbol. pp, protein–protein interaction; pd, protein–DNA interaction (promoter-binding).

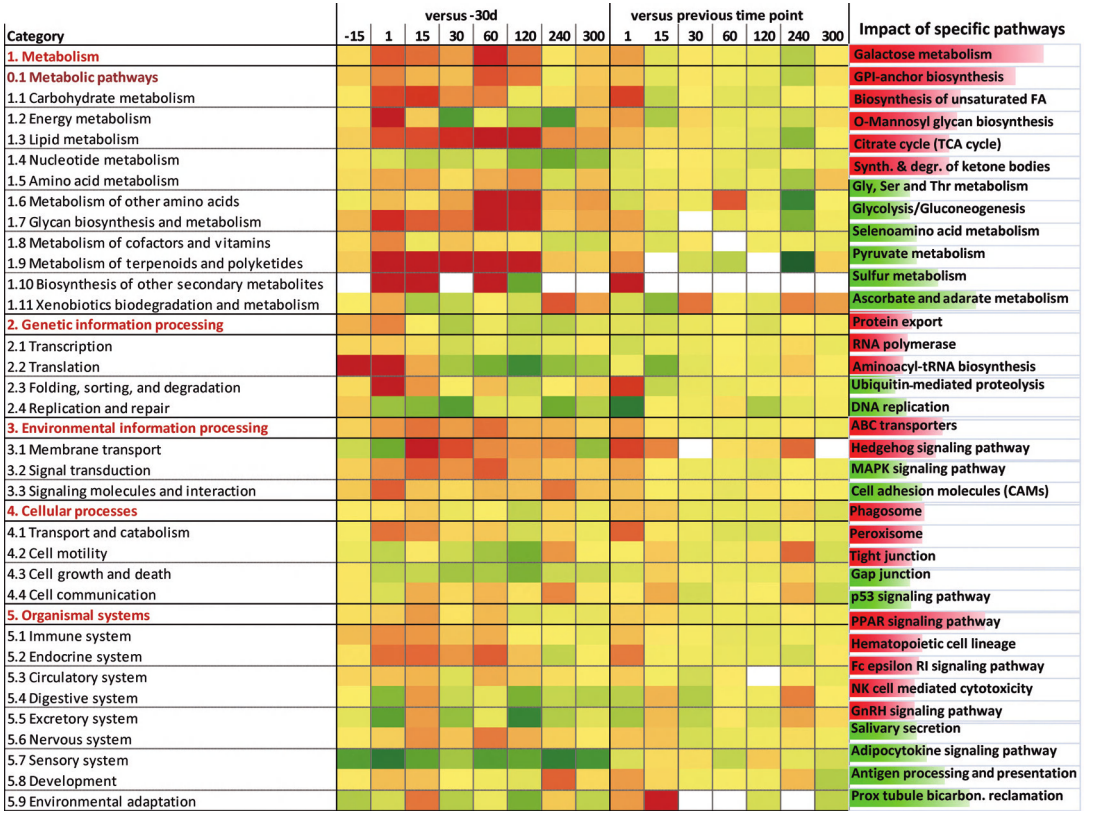


Figure 9.1 Overall calculated dynamic impact (DIA) among KEGG pathways using differentially expressed genes (DEG) in bovine mammary tissue during the lactation cycle. Shown are the main pathway classification category (e.g., metabolism, genetic information processing; left column) and corresponding subcategories (e.g., carbohydrate metabolism in metabolism category, transcription in genetic information processing). The heat map denotes estimated increase (red shade) or decrease (green shade) of metabolic flux or signaling through the pathway at each time point (−15, 1, 15, 30, 60, 120, 240, and 300 days from parturition) relative to −30 days or relative to a consecutive time points. The estimates are calculated as ((% of DEG/genes for the particular pathway in the microarray) × average $\log_2$ fold change of DEG × average $-\log p$-value of DEG). The most impacted (denoted by the length of the bar) pathways with an estimated increase (red bar) or decrease (green bar) in flux in each general category during lactation are shown in the right column.

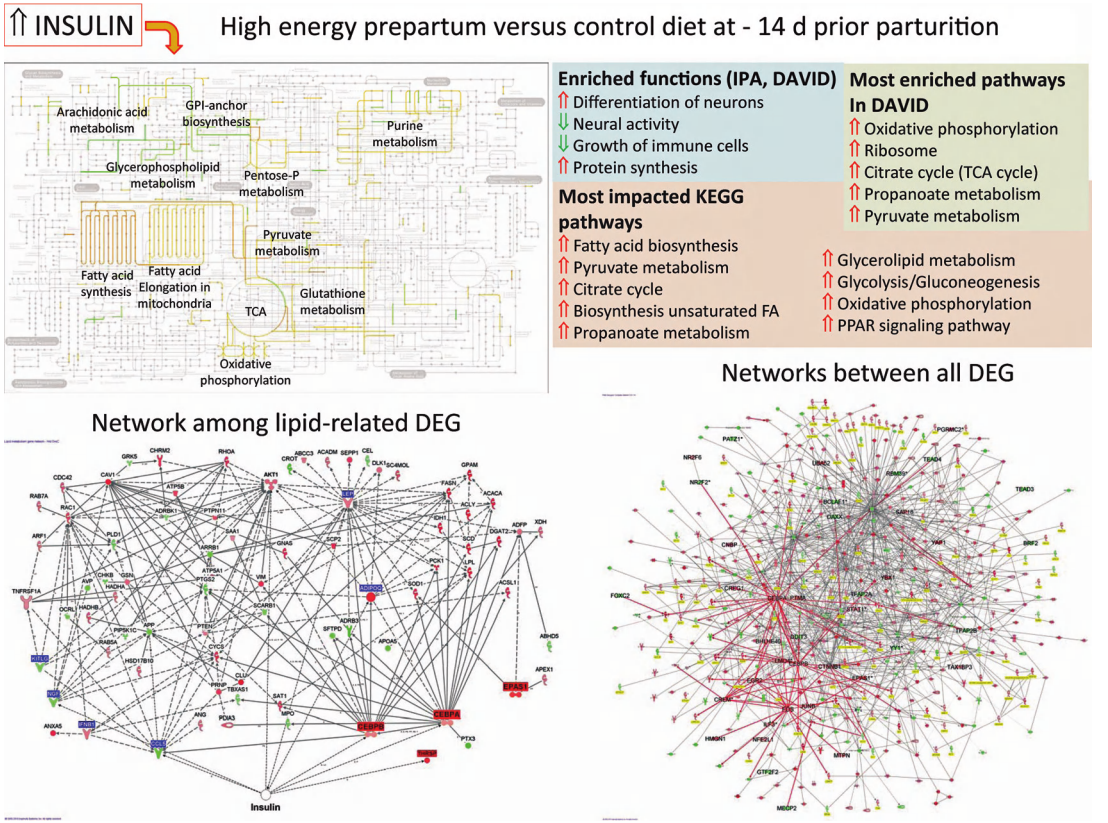


Figure 9.2 Summary of the effects of overfeeding energy during the dry period on bovine subcutaneous adipose tissue metabolic flux (KEGG pathways), main biological functions (DAVID and Ingenuity Pathways Analysis® (IPA)), and network analysis to uncover putative links between blood insulin and DEG in energy-overfed cows versus controls. Arrows denote upregulation (↑) or downregulation (↓) in energy overfed cows versus controls. In the KEGG pathway orange-to-red denotes an increase in flux, while green denotes a decrease in flux through the specific pathway. Among lipid-related DEG highlighted in red in the insulin gene network are transcription factors (TF), and in blue are cytokines and growth factors. Within this network and the network of all DEG, red denotes upregulation and green downregulation of expression of the particular gene in cows overfed energy versus controls. In addition, in the network encompassing all DEG, the TF are in large bold font. Microarray data are from the study of Janovick et al. (2009).

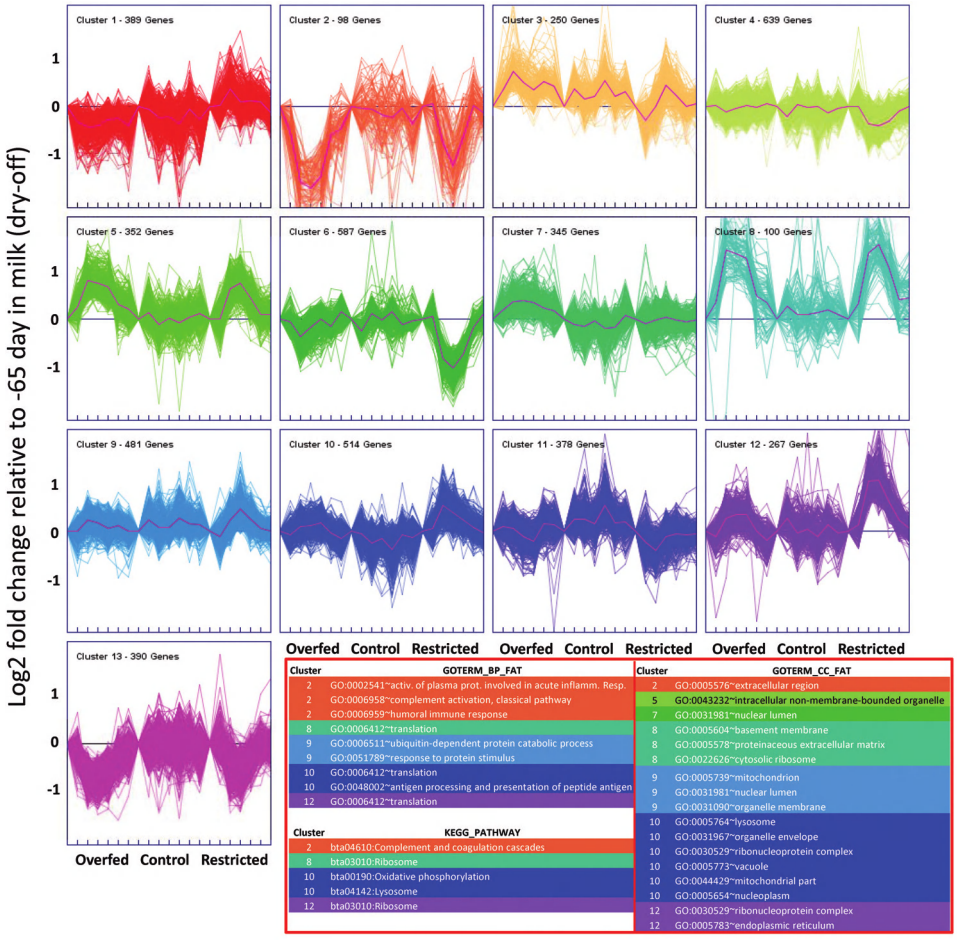


Figure 9.3 *k*-means clustering analysis using Genesis software (Sturn et al., 2002) of ∼4790 DEG (false discovery rate ≤ 0.05) due to treatment × time in liver of cows underfed energy (ca. 80% of energy requirements), overfed energy (ca. 150% of energy requirements), or fed energy to meet requirements (control) prepartum (from dry-off to parturition). The X-axis depicts the dietary treatments imposed by time point (−30, −14, 1, 14, 28, and 49 days relative to parturition). The Y-axis depicts fold change in expression compared with −65 days relative to parturition (i.e., the first sample before cows were assigned to treatments). The average trend in expression pattern for all the genes composing each cluster is shown in pink. Genes composing each cluster have correlated expression profiles and may take part in the same or similar biological processes. Reported also are the GO biological process, GO cellular component, and KEGG pathways in DAVID that were enriched significantly (Benjamini–Hochberg multiple comparison correction <0.05). To help in data interpretation, the color of the table matches the color of the cluster. Microarray data are from a reanalysis of Loor et al. (2005, 2006).

Category	FLUX and IMPACT			Most impacted pathways in each major category	
	EB versus CTR	FSO versus CTR	EB versus FSO	EB versus CTR	FSO versus CTR
1. Metabolism				Phenylalanine metabolism	FA elongation in mitochondria
0.1 Metabolic pathways				FA elongation in mitochondria	Histidine metabolism
1.1 Carbohydrate metabolism				Histidine metabolism	Phenylalanine metabolism
1.2 Energy metabolism				Primary bile acid biosynthesis	Folate biosynthesis
1.3 Lipid metabolism				Cyanoamino acid metabolism	Tyrosine metabolism
1.4 Nucleotide metabolism				Taurine & hypotaurine metabolism	Tryptophan metabolism
1.5 Amino acid metabolism				Linoleic acid metabolism	Primary bile acid biosynthesis
1.6 Metabolism of other amino acids				Tyrosine metabolism	Glycosphingolip biosynt (globo series)
1.7 Glycan biosynthesis and metabolism				Retinol metabolism	One carbon pool by folate
1.8 Metabolism of cofactors and vitamins				Drug metab - cytochrome P450	Inositol phosphate metabolism
1.9 Metabolism of terpenoids and polyketides				Metab of xenobiotics by cyt P450	O-Mannosyl glycan biosynthesis
1.10 Biosynthesis of other secondary metabolism				Arachidonic acid metabolism	Starch & sucrose metabolism
1.11 Xenobiotics biodegradation and metabolism				Nitrogen metabolism	Sphingolipid metabolism
2. Genetic information processing				Homologous recombination	RNA polymerase
2.1 Transcription				RNA polymerase	Aminoacyl-tRNA biosynthesis
2.2 Translation				Protein export	Protein export
2.3 Folding, sorting, and degradation				RNA degradation	Mismatch repair
2.4 Replication and repair				Mismatch repair	Spliceosome
3. Environmental information processing				mTOR sign pathway	Calcium sign pathway
3.1 Membrane transport				Calcium sign pathway	Phosphatidylinositol sign system
3.2 Signal transduction				Phosphatidylinositol sign. system	ABC transporters
3.3 Signaling molecules and interaction				ECM-receptor interaction	mTOR sign pathway
4. Cellular processes				Regulation of autophagy	Lysosome
4.1 Transport and catabolism				Peroxisome	Regulation of autophagy
4.2 Cell motility				p53 sign pathway	p53 sign pathway
4.3 Cell growth and death				Gap junction	Adherens junction
4.4 Cell communication				Oocyte meiosis	Tight junction
5. Organismal systems				Aldosterone-regulated Na reabsorpt	Complement & coagulation cascades
5.1 Immune system				Cytosolic DNA-sensing pathway	Hematopoietic cell lineage
5.2 Endocrine system				PPAR sign pathway	Pancreatic secretion
5.3 Circulatory system				Complement & coagulation cascades	Cytosolic DNA-sensing pathway
5.4 Digestive system				Proximal tubule bicarb reclamation	Toll-like receptor sign pathway
5.5 Excretory system				Cardiac muscle contraction	Aldosterone-regulated Na reabsorpt
5.6 Nervous system				Long-term potentiation	RIG-I-like receptor sign pathway
5.7 Sensory system				NOD-like receptor sign pathway	Circadian rhythm - mammal
5.8 Development				Gastric acid secretion	NOD-like receptor sign pathway
5.9 Environmental adaptation				Pancreatic secretion	Insulin sign pathway

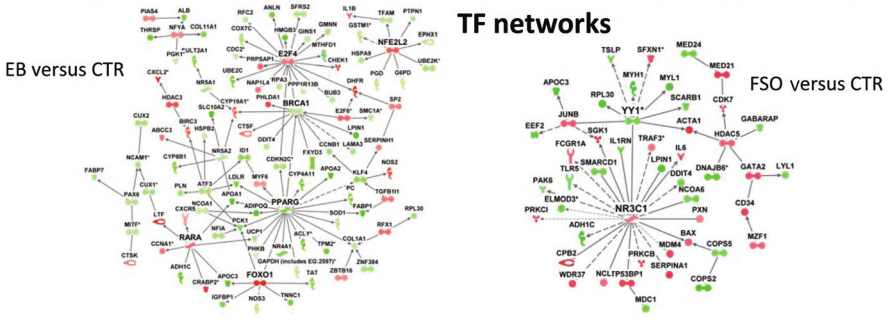


Figure 9.4 Overall calculated dynamic impact (DIA) in KEGG pathways of DEG in bovine mammary tissue from midlactation cows fed a control diet (CTR), the control diet supplemented with saturated lipid (EB), or the control diet supplemented with a blend of fish oil and soybean oil (FSO) for 3 weeks. Shown are the main pathway classification groups (left column) and corresponding subgroups. The heat map denotes potential increase (red shade) or decrease (green shade) of metabolic flux or signaling through the pathway for each treatment comparison. The overall impact is denoted by the size of the blue bar (the larger the bar, the greater the impact of DEG on the category of pathways). The most impacted pathways with the overall flux (red shade denotes increases and green shade denotes decreases) for the comparison EB versus CTR and FSO versus CTR are shown in the right column. The transcription factor (TF) networks produced by DEG in EB versus CTR and FSO versus CTR are reported in the bottom panel. The TF in each network are highlighted by larger font. Details of the animal experiment and portions of the microarray analysis have been reported previously (Invernizzi et al., 2009, 2010).

Chapter 1
Introduction to Systems Biology for Animal Scientists

Christopher K. Tuggle, Fadi Towfic, and Vasant G. Honavar

Why Should Animal Scientists Be Interested in Learning About Systems Biology?

This section describes three arguments to justify why animal scientists should explore the field of Systems Biology.

The Goals of Systems Biology Are Directly Aligned with Those of Animal Science

Systems Biology aims to produce information to make biology predictive, and (at least as a final outcome) to do so at the organism level. Such prediction at the animal level would be very useful for practical goals in animal agriculture such as improving phenotypic traits, especially those with low heritability, that are otherwise recalcitrant to such improvement and for which genomic approaches have been highlighted (Green et al., 2007; Sellner et al., 2007). These include traits that are difficult or expensive to measure such as female reproduction, efficiency of feed utilization, and resistance to disease. However, such predictive power would be extremely useful in studies in animal nutrition, physiology, immunology, and reproductive biology, where the goal is to understand the effect of altered feeds and feeding regimens, feed additives, hormones or other drugs, as well as effects of changes in management, on the whole animal. Of course, more intermediate goals of Systems Biology, such as a deeper understanding of specific important pathways and pathway interactions in relevant cells, tissues and organ systems, will more immediately show the value of the Systems Biology approach in the livestock field as well as in biomedicine.

Systems Biology and Livestock Science, First Edition. Edited by Marinus F.W. te Pas, Henri Woelders, and André Bannink.

Translation of Systems Biology Understanding of Human and Model Vertebrates to livestock Species Will Be Highly Instructive

Significant funding is being devoted toward developing the required substantial data and analytical resources for Systems Biology studies of several species, including humans, mice, zebrafish, and a number of invertebrates and microorganisms. Much of these data, and the models and hypotheses generated, will be applicable to modeling the biology of livestock species of interest to animal scientists, especially for data collected on the vertebrate species where the majority of this effort is focused. While the methods necessary to globally compare biological networks need to be developed and refined continually in order to best perform such comparisons, animal scientists have used comparative biology for many years to take advantage of biological data and insights from the biomedical and fundamental biology fields. As discussed in the next section, comparative efforts and value can also work in the reverse, improving the modeling of human and model species biology as well.

Systems Biology Will Best Utilize New Genome Information for All Species

The first two points taken together point to the promise of Systems Biology for making optimal use of the new genomic information from species of interest to address specific questions in animal science. According to some authors, biology has seen a paradigm shift in the past 10–15 years, since the initial fruits of the Human Genome Project (HGP) began to be harvested (Schena et al., 1995; Lander et al., 2001; IHGSC, 2004). The age in which biological molecules are studied using a reductionist approach—in isolation—is fast drawing to a close. This is especially true in very well-funded areas of biology such as human medicine. Funding agencies, journals, and public stakeholders are increasingly interested in how molecular studies and their conclusions are integrated within larger systems, e.g., organs, organisms, species, and ecosystems. This shift is partly due to the expectation that research, even research whose goal is to describe fundamental biological processes, must have a realizable and practical benefit. A "selling point" in the 1980s put forth to encourage funding of the HGP was that a global understanding of the human and model organism genomes would result in practical benefits as yet unknown. However, the shift was greatly accelerated by the very success of the HGP in creating methods for global measurements of the "parts list" of the genome, the variation of parts structure (the genes), and the interactions between these parts. Happily, the animal science community has always had an integrative and practical view of research, developing new knowledge with a focus on applying such information as quickly as possible. Now that most species of interest to animal scientists have substantial genome tools, the concurrent and future development of bioinformatics tools to explore, interpret and integrate molecular-level data across interaction networks, tissues, and organs to predict biological states (read: phenotypes) will be highly beneficial to livestock interests. Conversely, the development of a predictive biology across multiple vertebrate species, including the livestock species, will deepen the understanding of the human species as well. The value of comparative models for human biology will be greatly increased if multilevel, detailed modeling of the same processes can be manipulated, and if perturbation effects can be iteratively predicted,

tested, and new predictions be made and retested. Many such perturbations cannot be ethically performed in humans, but may be justified in animals. Thus, improvement in the systems-level modeling of several livestock species, already validated as excellent models for human physiology such as the pig (Dehoux and Gianello, 2007; Lunney, 2007) and sheep (Scheerlinck et al., 2008), will improve biomedical understanding as well.

Thus, for the above reasons, it is important for animal scientists to understand the current state and future promise of Systems Biology. In the following pages, we first provide a general description of Systems Biology. Then, we describe some current applications of systems approaches in livestock biology. Finally, we speculate on aspects of Systems Biology that are likely to find use in the near future by scientists interested in livestock and other areas of animal sciences.

What Is Systems Biology?

Over the past several decades, biologists have been accumulating detailed knowledge of the building blocks of biological systems, e.g., DNA, RNA, proteins, cells, tissues, organs, organisms, and ecologies. Anatomical, physiological, molecular, cellular, and structural approaches to biology have revolutionized our understanding of how living organisms function. However, biological systems are more than simply a collection of molecules, cells, or organs. We need to understand how the parts work together to form dynamic functional units, e.g., how genetic and regulatory interactions and environmental factors orchestrate development, aging, and response to disease. The emergence of high-throughput techniques has made possible system-wide measurements of biological variables, e.g., the expression of thousands of genes under different conditions or perturbations, the interactions between proteins, and between proteins, genes, regulatory RNAs, small ligands, and other signaling agents.

The term "Systems Biology" refers to a collection of methods and tools that attempt to understand complex biological systems by leveraging diverse datasets generated using disparate instruments of observation, and by modeling the interactions among the large numbers of constituent components (Jeong et al., 2000; Kitano, 2002; de Jong, 2002; Auffray et al., 2003; Ge et al., 2003; Ideker, 2004; Klipp et al., 2005; Liu, 2005; Baitaluk et al., 2006; Bruggeman and Westerhoff, 2007). Examples of Systems Biology research include studies of genetic interaction and regulatory networks (de Jong, 2002), metabolic networks (Jeong et al., 2000), and their combinations (Auffray et al., 2003; Baitaluk et al., 2006). Recent advances in Systems Biology have led to substantial progress on problems such as uncovering the essential macromolecular sequence and structural features of molecular interactions (Walhout, 2006); extracting signaling pathways from gene and protein interaction networks (Scott et al., 2006; Hecker et al., 2008); discovering topological and other characteristics of these networks (Ravasz et al., 2002; Farkas et al., 2003; Yook et al., 2004; Basso et al., 2005; Khanin and Wit, 2006); integration of disparate types of data (microarray, proteomics, physical interaction, subcellular localization, etc. (Bernard and Hartemink, 2005; Sharan and Ideker, 2006; Hecker et al., 2008)); prediction of the most important nodes in large genetic and protein networks (Jeong et al., 2001); finding gene modules that orchestrate metabolic or cellular functions (Milo et al., 2002; Segal et al., 2003; Sen et al., 2006).

Because a major effort in animal genomics has recently created large datasets (primarily RNA based but also some proteomic analyses have been reported (Bendixen et al., 2010; Picard et al., 2010)) to uncover systematic interactions and outcomes in animal systems, the rest of this section focuses on such network modeling. Progress in modeling higher level systems that control immunological or physiological systems will depend on not only the identification of the molecular building blocks of those systems, but also data on the interactions among the parts, the behaviors of such systems under different conditions. Hence, significant efforts have focused on developing several broad classes of network models in Systems Biology, as models need to describe the underlying gene and protein networks at the appropriate levels of abstraction for exploring different types of questions (see Figure 1.1A for an example of a network model describing biological events). Further, methods are needed to construct such models using data that is variable in quantity, quality, and granularity. In what follows we discuss several forms of the suggested models (Figures 1.1A–D):

- **Undirected graphs** in which nodes represent genes or proteins and links between nodes represent interactions, e.g., protein–protein interaction networks (Uetz et al., 2000; Ito et al., 2001; Stelzl et al., 2005) or weighted graphs in which the weights on links model the strength of interaction (e.g., gene expression correlation networks; see Stuart et al., 2003); see Figure 1.1B. Such networks provide a global picture of gene–gene or protein–protein interactions that can further be analyzed to identify putative functional modules (Uetz et al., 2000; Ito et al., 2001; Stelzl et al., 2005) or nodes that play important roles (e.g., hubs; Jeong et al., 2001); or to determine topological features (degree distribution, hierarchical structure, modularity, etc. (Ravasz et al., 2002; Farkas et al., 2003; Yook et al., 2004; Khanin and Wit, 2006)). Comparative analysis of two or more networks of the same type from different species can help identify conserved functional modules (Ogata et al., 2000; Matthews et al., 2001; Stuart et al., 2003; Yu et al., 2004a; Sharan and Ideker, 2006).
- **Directed graphs** that model influences between genes where nodes represent genes and unlabeled or labeled edges denote regulatory interactions. Pathway databases such as TRANSPATH (Krull et al., 2006), KEGG (Kanehisa et al., 2008) are examples of richly annotated directed graphs. Tracing of directed paths in such graphs can uncover sequences of regulatory events, and for example redundant regulatory mechanisms; directed cycles indicate feedback regulation (see Figure 1.1C). Comparison of pathways can reveal common subgraphs and putative evolutionary relationships. Topological analysis can reveal the distributions and average numbers of regulators per gene.
- **Boolean networks** (Thomas, 1973; Kauffman, 1993; Silvescu and Honavar, 2001; Lähdesmäki et al., 2003; Kim et al., 2007) that model influences between genes or proteins using Boolean functions. Nodes represent genes or proteins whose states are modeled by binary variables (0, 1); see Figure 1.1D. States of genes are updated in discrete time steps; the state of a node at time $t + 1$ is a Boolean function of the states at time t (or more generally, the states of the nodes/genes at $t, t - 1, t - 2, t - 3 \ldots t - T$ time steps, where T is a parameter representing the maximum number of previous time steps being considered) of the genes that influence it. An N gene Boolean network can in principle be in one of 2^N states. Boolean functions can model complex nonlinear regulatory influences between genes; Boolean networks

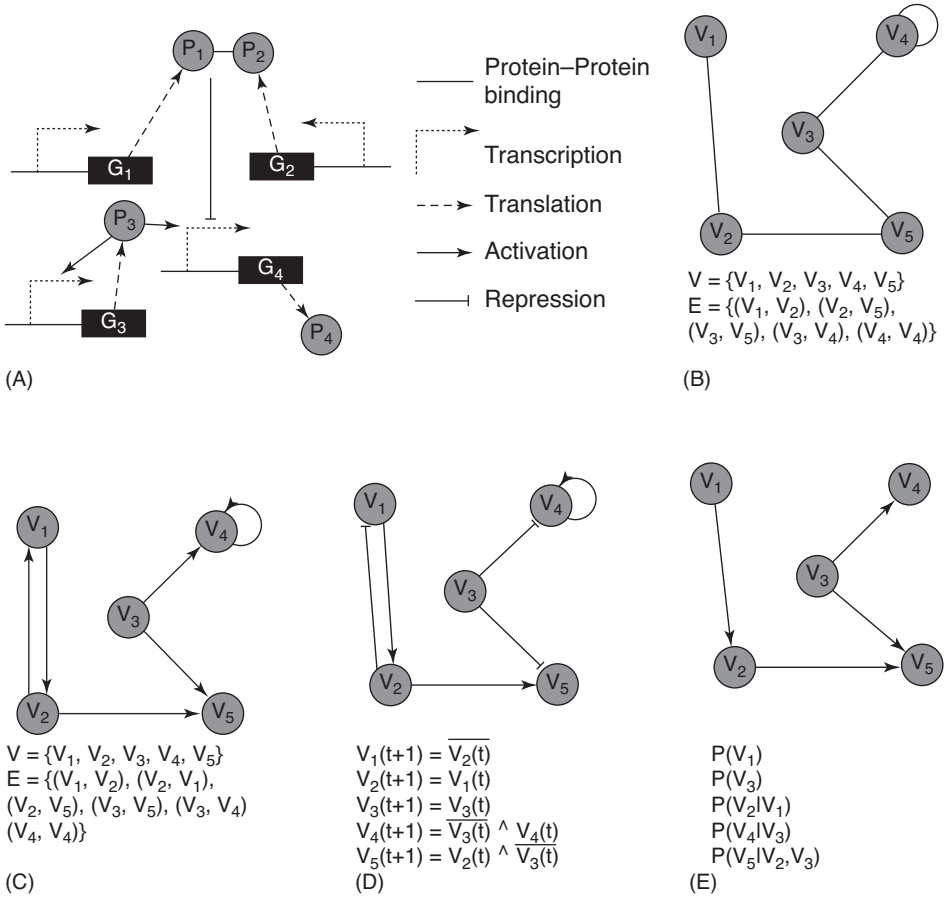


Figure 1.1 (A) A regulatory network schematic. This figure shows the transcription, followed by translation events of Protein 1 (P1) and Protein 2 (P2) from genes 1 (G1) and 2 (G2), respectively. P1 and P2 then form a protein complex (P1–P2) by binding to each other. The P1–P2 complex then acts as a repressor of Gene 4 (G4), which codes for Protein 4 (P4). G4 is activated by Protein 3 (P3), which also acts as the activator for its own gene (G3). (B) An example of an undirected graph representing interactions between nodes in a network. Here, edges do not have directions (i.e., the elements in the edge set are unordered) and self-edges indicate self-interactions (e.g., formation of a protein dimer in a protein–protein interaction network). (C) An example of a directed graph representing interactions between nodes in a network. Here, edges have one-way interactions (i.e., the elements in the edge set are ordered). Feedback is indicated by self-loop in case of self-regulation (e.g., V4) or cycles (e.g., between V1 and V2). (D) An example of a Boolean network representing qualitative interactions between nodes. Here, edges can indicate more than one type of relationship (e.g., activation in case of arrows and inhibition indicated by the perpendicular terminator edge). The relationships between the nodes can be qualitatively described using Boolean notation as indicated in the figure. (E) An example of a Bayesian network representing quantitative interactions between nodes. Here, edges can represent probabilistic relationships between nodes as indicated in the figure.

can only model simplified (discrete time, discrete state) dynamics. Simulation can be used to examine the effect of specific perturbations (e.g., gene knockouts). Reachability analysis over the state space of the network can be used to answer the question whether a network can get from one state *s* (e.g., healthy) to a different state *g* (e.g., diseased).

- **Probabilistic models,** e.g., **Bayesian networks** (Ghahramani, 1998; Friedman et al., 2000; Hartemink et al., 2002; Pe'er, 2005; Myers and Troyanskaya, 2007; Lähdesmäki and Shmulevich, 2008) and related probabilistic approaches (Datta et al., 2004; Lähdesmäki et al., 2006) that capture stochastic aspects of interactions. A Bayesian network is a directed acyclic graph (DAG) wherein each node corresponds to a random variable (RV) that models the state of a node (Boolean, discrete, or continuous). Directed links represent direct dependencies (not to be confused with causal influences). Each node X_i has associated with it, a probability distribution for that RV conditioned on its parents in the graph *G*. A Bayesian network (BN) specifies a factorization of the joint probability distribution as a product of the probability distributions of each node conditioned on its parents in the BN. That is, Bayesian networks can summarize observed dependencies in gene expression measurements and predict effects of interventions on the conditional distributions; see Figure 1.1E.
- **Dynamic models** (not shown) based on differential equations or partial differential equations (Conrad and Tyson, 2006; Del Vecchio and Sontag, 2007) or stochastic differential equations (Chen et al., 2005) that capture detailed information about quantities and rates of specific biochemical reactions.

Increasing the Reliability of Gene and Protein Networks

The lack of complete experimental data on the interaction networks presents an important challenge for construction and analysis of such networks (Bader, 2003; Bader et al., 2004). Vidal's group (Han et al., 2004), for example, observed that the incompleteness of protein–protein interaction networks can lead to misleading conclusions from topological analysis of the networks. The presence of false positives in protein–protein interaction datasets poses serious challenges in utilizing such data in generating or validating specific hypotheses. The false positives are due to two reasons: (i) The biochemical methods such as yeast two-hybrid analyses test only if the proteins can bind sufficiently well so that a reporter system read-out is positive and do not take account of physiological concentrations of the tested molecules; (ii) Two proteins may never be present in the same cell at the same time, and thus never interact, even though they may physically be able to do so. It has been shown that some reported protein interactions could not be reconciled with known protein complexes (Edwards et al., 2002). Likewise, there are similar problems with the gene networks (de Jong, 2002). It is well known that gene expression measurements obtained using microarray experiments are susceptible to many potential sources of variation/error (Churchill, 2002; Novak et al., 2002). Posttranslational modifications (e.g., phosphorylation, glycosylation) can result in multiple variants of proteins being produced from a mRNA. The task of modeling the underlying networks is greatly complicated by the fact that the "function" of a gene, RNA, or protein in particular pathways depends on many

variables, including alternative splicing, miRNA modulation, posttranslational modification, and subcellular localization. Combining data from multiple sources presents several computational and statistical challenges. Such efforts are further complicated by the fact that the data typically come from different studies that are driven by different objectives. Despite these challenges, graph representations of networks provide an attractive framework for combining data from multiple sources (e.g., networks generated from data gathered from different organisms; (Sharan and Ideker, 2006)). The knowledge of functional orthologs of proteins in networks based on data from different species (e.g., human, mouse) is often incomplete. Moreover, the problem of aligning multiple networks in general requires solution of the subgraph isomorphism problem. This problem is known to be computationally intractable (NP-hard) (Garey and Johnson, 1990) and hence requires either the use of heuristic methods (Berg et al., 2004) or forcing consistent one-to-one mappings between nodes in different networks (Stuart et al., 2003), which is unrealistic when dealing with data from different species (e.g., due to gene duplication events or incomplete knowledge of orthologs). Towfic et al. (2010) have developed novel, efficient, and modular approaches to aligning multiple protein–protein interaction networks or gene coexpression networks. The resulting algorithms have been successfully applied for distinguishing orthologs (gene sequences predicted to have a direct evolutionary descendant relationship) from paralogs (gene sequences predicted to arise from a duplication within a species yet often compared across species), and grouping tissues, species, etc. on the basis of similarity of networks being aligned.

Leveraging High-Throughput Data to Refine Network Models and to Prioritize Experiments

Many current approaches to infer relationships and build gene network models rely directly on gene expression data. Genes that show correlated expression patterns are further examined for evidence that they might be coregulated. Such analysis involves the identification of genes or proteins that are differentially expressed under different conditions (Rockman and Kruglyak, 2006), then the clustering and grouping of genes based on similarity of expression profiles measured at different times or under diverse conditions (Eisen et al., 1998; Alon et al., 1999; Bar-Joseph, 2004; Jiang et al., 2004; Lonosky et al., 2004). A recognition of coexpression of such gene clusters can identify common aspects such as functions through Gene Ontology (GO) annotation classification (Abba et al., 2004; Fraser et al., 2005; Sievertzon et al., 2005), shared transcription factor binding sites (Cole et al., 2005; Liu and Agarwal, 2005; Ho Sui et al., 2007) and association to known phenotypes (Brown et al., 2000; Carter et al., 2004). Typically, gene expression studies measure the level of expression of hundreds or thousands of genes under a small number of perturbations or at a few time points. Clustering of gene expression profiles and analyses of the resulting clusters in terms of functional annotations or phenotypes provides an insufficient basis for asserting functional dependencies with certainty. Data from experimental interventions (e.g., gene knockouts) designed to uncover how the activity of a specific gene is influenced by the activities of other genes, are, at present, quite expensive to gather in complex organisms. Such data are also typically rather sparse, due, in part,

to the expense of creating such new genetic mutants. Hence, the building, from the bottom up, of detailed models of interactions between genes that include directed links denoting functional dependencies, or Boolean functions that describe how the expression of specific genes of interest are influenced by other genes, is beyond the reach of current experimental and computational methods. However, experimental biologists often have detailed knowledge of the role of a small number of genes within specific pathways (e.g., a pathway that is responsible for the differentiation of retinal stem cells into rods and cones, or a pathway that governs the response of yeast cells to different types of stress) based on detailed molecular, biochemical, and genetic manipulations; see, for example, the integrated NCBI databases on specific genes (http://www.ncbi.nlm.nih.gov/gene). A more practical approach is to leverage high-throughput data to expand detailed models of specific fragments on a pathway of interest. Hence, there has been significant interest in methods for iterative refinement of gene networks. In the foreseeable future, we anticipate such iterative approaches, coupled with techniques for prioritizing experiments, to find use in building predictive models of complex traits, e.g., disease resistance, in animals. In the next sections, we describe the details of examples of work in this area, which is exemplified by recent progress in one of our groups (VH).

Expanding a Seed Network of Genes Involved in Retinal Development by Querying Multiple Gene Expression Datasets

A number of published studies (Blackshaw et al., 2004; Dorrell et al., 2004; Akimoto et al., 2006; Liu et al., 2006) have profiled changes in gene expression during normal retinal development. However, successful extraction of gene networks controlling retinal development remains elusive. Incompleteness of the datasets and differences in observed correlations in gene expression across different experiments, along with low (or incomplete) temporal coverage (hundreds or thousands of genes measured at only a handful of time points), present significant challenges in reliable reconstruction of the underlying network. We explored an approach to query data from five previously published high-throughput gene expression datasets for the developing retina (Blackshaw et al., 2004; Dorrell et al., 2004; Akimoto et al., 2006; Liu et al., 2006), using a directed graph "seed network" (Figure 1.2) of genes that have been shown

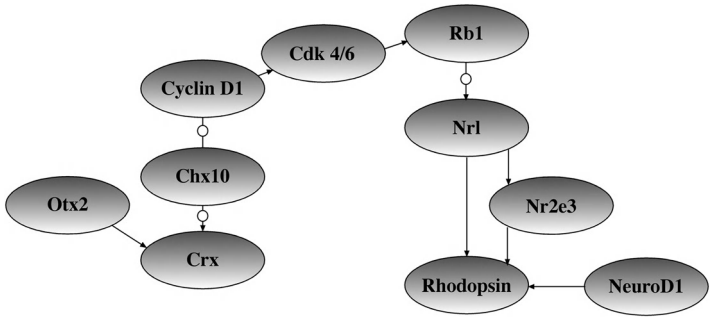


Figure 1.2 Initial Seed Network of gene interactions involved in rod cell differentiation in the eye (Hecker et al., 2008).

to govern rod development, constructed manually based on data from a large number of detailed molecular and genetic experiments (Chen et al., 1997; Ahmad et al., 1998; Mears et al., 2001; Green et al., 2003; Nishida et al., 2003; Pennesi et al., 2003; Cheng et al., 2004; Rutherford et al., 2004; Zhang et al., 2004). Despite the low level of concordance across the different datasets, we showed that by integrating multiple datasets, we could reconstruct the links between the seed-network genes simply on the basis of observed correlations between genes in multiple (at least two out of five) gene expression datasets. The results are illustrated in Figure 1.2, where red lines indicate negative correlations and blue lines indicate positive correlations.

In this network, there are positive correlations between several genes known to be expressed by dividing cells and positive correlations between genes known to be expressed by mature photoreceptors, with negative correlations between the two groups. Based on the premise that genes that are likely to play important roles in rod photoreceptor development are likely to be correlated with more than one seed-network gene, we queried the composite dataset to identify genes that were correlated with multiple seed-network members. Cell signaling pathway data (KEGG; www.genome.jp/kegg/pathway.html; Kanehisa et al., 2008) were then retrieved for each gene that was correlated with multiple members of the seed network. Using this procedure, ten such genes were identified as part of the BMP/SMAD signaling pathway. The BMP/SMAD signaling has been implicated in rod photoreceptor development (Murali et al., 2005); 22 proteins were identified as members of WNT/Frizzled signaling, which has been implicated in rod photoreceptor differentiation (Yu et al., 2004b); and finally, 24 genes were identified as members of the insulin/IGF-1 signaling (which are not distinguished from one another in KEGG). From the list of genes correlated with multiple seed-network members, we identified eight additional hypothesized candidates for addition to the seed network; (Figure 1.3; Hecker et al., 2008). For example, Figure 1.4 shows several genes involved in heparin sulfate biosynthesis and cystatin C, which are predicted to be linked with "rod-specific genes" (enclosed in the box). Solid arrows indicate links that are consistent with published experimental results (not used to establish the connections). Dashed arrows indicate links for which no experimental evidence is currently available. These results demonstrate that while extraction of gene networks *de novo* from gene expression studies has been difficult, our approach of using multiple datasets, and a seed network to query a combined dataset has successfully identified candidate genes and pathways

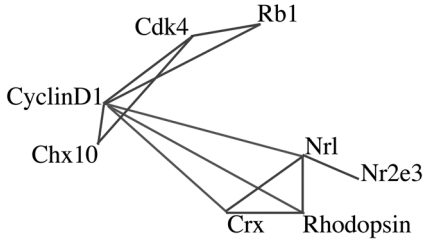


Figure 1.3 Seed network reconstructed from multiple gene expression datasets. Blue lines indicate positive correlations and red lines indicate negative correlations (Hecker et al., 2008). (*See insert for color representation of this figure*.)

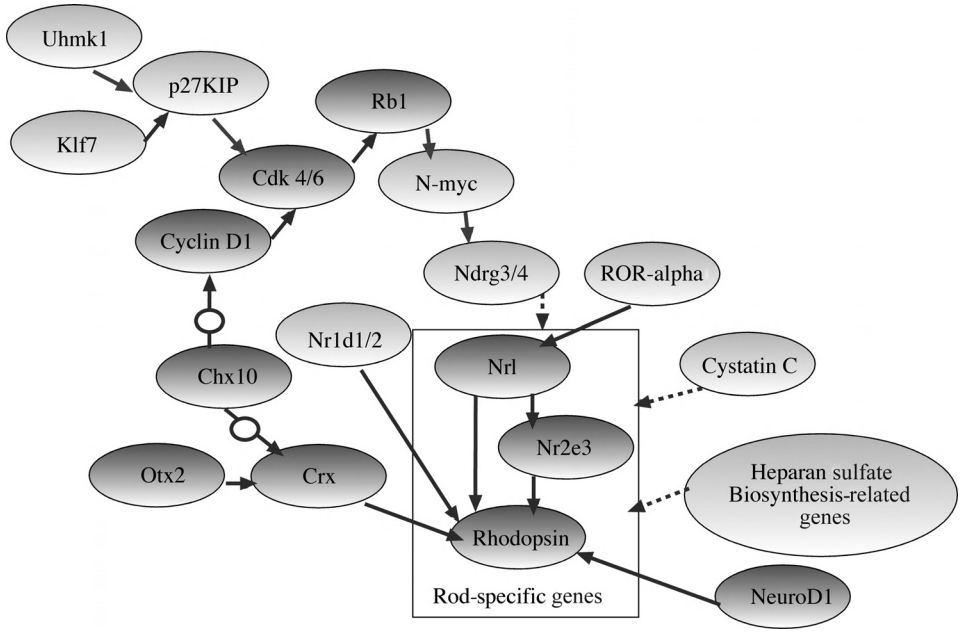


Figure 1.4 Expanded network. Original seed network nodes are shown in gray, and added network nodes are shown in blue. Blue arrows correspond to excitatory interactions, and red arrows correspond to negative interactions. Direct influences are shown using solid lines and indirect ones (through possible intermediaries) using dotted lines (Hecker et al., 2008). (*See insert for color representation of this figure*.)

important for photoreceptor development. Importantly, two pathways identified through this seed-network analysis have previously been shown to be involved in rod development (Hecker et al., 2008 and references therein). To facilitate analysis of the sort described above, we have implemented a prototype of an open source tool, the *Retina Workbench* (Kohutyuk, 2007) [www.cs.iastate.edu/∼retinaworkbench/], for constructing, querying, and analyzing multiple gene expression datasets using a seed network. We have made it available as a plug-in for *Cytoscape*, [www.cytoscape.org/] (Shannon et al., 2003), a widely used suite of software for visualization and analysis of networks. The *Retina workbench* allows sharing of datasets, networks, and analysis results among members of a research group or across research groups. Although initially aimed at a community of biologists interested in the retina, the tool can be adapted to work with other datasets simply by populating the associated database with the relevant datasets. This example shows how integration of data in a Systems Biology way can yield new biological insights.

Topological Analysis of Interaction Networks

Several groups have suggested, based on topological analysis of networks, that gene, protein, and metabolic networks display special topological characteristics, e.g., they are approximately scale-free, i.e., the degree distribution $P(k) \sim k^{(-\gamma)}$. In other words, the number of nodes that have k links approximately follows a power law with decay

constant, gamma, describing the decrease of the frequency of nodes with k links as k increases (Ravasz et al., 2002; Farkas et al., 2003; Yook et al., 2004; Khanin and Wit, 2006). Scale-free networks are dominated by a few highly connected nodes or hubs that have been shown to be critical to the overall functioning of the network (Jeong et al., 2000). Therefore, these networks are likely to be conserved through evolution. Gene and protein networks have also been reported to be modular, i.e., they have a high average clustering coefficient with the clustering coefficient C_i of a node i being the ratio of links between the nodes within its neighborhood (the direct neighbors of node i) relative to the maximum number of links that could exist between them. Protein and metabolic networks have been shown to display *hierarchical modularity* arising from a scale-free topology combined with a modular structure (Ihmels et al., 2002; Ravasz et al., 2002; Ravasz and Barabasi, 2003; Yook et al., 2004).

Spectral Analysis of Protein–Protein Interaction Networks

Spectral clustering (Shi and Malik, 2000; Ding et al., 2005; Dhillon et al., 2007; von Luxburg, 2007) of protein–protein interaction networks has been used with success to identify connected components and putative functional modules in protein networks. Sen et al. (2006) have recently published an analysis of the General Repository for Interaction Datasets (GRID) (Breitkreutz et al., 2003) database, a curated compilation of data from several yeast protein–protein interaction datasets (Ito et al., 2000, 2001; Uetz et al., 2000; Tong et al., 2001; Gavin et al., 2002; Ho et al., 2002; Breitkreutz et al., 2003). This analysis used singular value decomposition (SVD) of a matrix representation of a yeast protein–protein interaction network consisting of 4906 proteins, and 19,037 pair-wise interactions from the GRID database to identify clusters. The resulting clusters were shown to have related functions (as suggested by their GO functional annotations). This allowed them to use the GO annotations to predict cluster membership of novel proteins (and hence their interacting partners) and to correctly predict several new protein–protein interactions. These have since been confirmed experimentally by Krogan and colleagues (Krogan et al., 2004). Honavar's group used spectral analysis to identify modules from the mouse protein–protein interaction network constructed from BIND, DIP, and MINT protein–protein interaction databases (Bader et al., 2003; Salwinski et al., 2004; Mishra et al., 2006; Chatr-aryamontri et al., 2007). The results (Figure 1.5) show that several of the seed-network genes are associated with clusters of proteins within the protein–protein interaction network.

Identifying Novel Interactions

As noted earlier, incompleteness of protein–protein interaction networks can lead to misleading conclusions based on analysis of the networks. The mouse protein–protein interaction dataset (2736 pair-wise interactions involving 2134 proteins) is significantly less complete compared to its human counterpart (35,021 pair-wise interactions involving 9305 human proteins). However, many of the mouse proteins have human orthologs (a gene g_1 in species A is said to be an ortholog of gene g_2 in a species B if g_1 and g_2 share a common ancestral gene g in a species C that is a shared evolutionary ancestor of the species A and B. The strongest evidence that two similar genes are

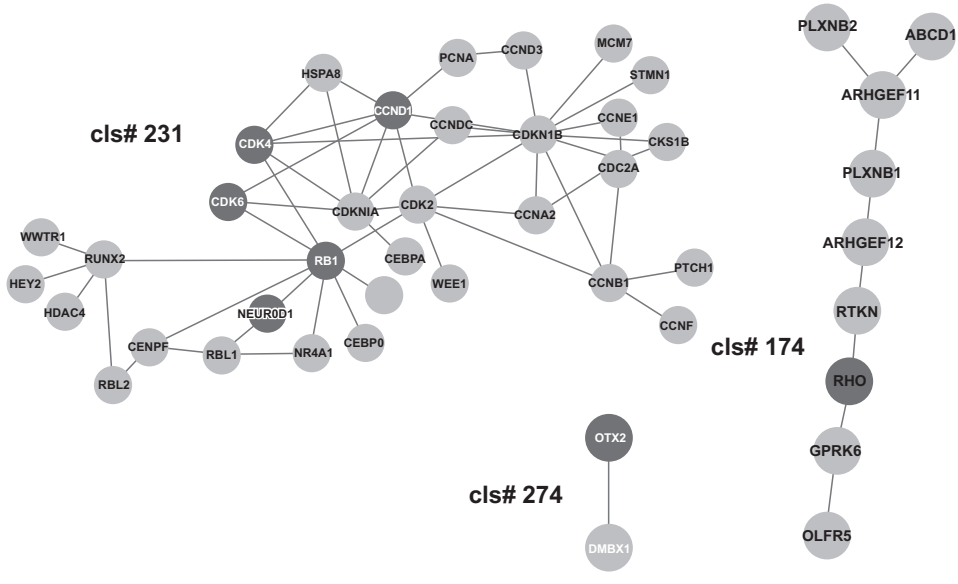


Figure 1.5 Some representative clusters within the mouse protein–protein interaction network obtained by spectral clustering. Seed network genes are shown in red. (*See insert for color representation of this figure*.)

orthologous is the result of a phylogenetic analysis of the gene lineage). This suggests the possibility of using mouse *orthologs* of human proteins be used to infer missing interactions in the mouse protein–protein interaction network.

Beginning with a fragment of the mouse protein–protein interaction network based on the seed network described in section A (Figure 1.2), the Honavar group identified the human orthologs of mouse proteins in this network using the Jackson Labs' mammalian orthology resource: www.informatics.jax.org/orthology.shtml. Additional links in the mouse–protein interaction network were then inferred based on links between the corresponding orthologs in the human protein–protein interaction network (Figure 1.6).

A Systems Biology Paradigm: The Progress in Analysis of the Mammalian Immune Response Network

Excellent examples of Systems Biology analyses that are immediately approachable for animal scientists can be taken from the field of systems analysis of immunology, which has begun to generate the datasets required for detailed systems biological modeling (Hyatt et al., 2006; Heng and Painter, 2008; Kleinstein et al., 2008; Amit et al., 2009; Gardy et al., 2009; Zak and Aderem, 2009). A superb early example in this field is that published by Gilchrist et al. (2006). The purpose of this research was identification of regulatory factors controlling expression of genes responding to lipopolysaccharide (LPS), a component of gram-negative bacteria such as *Escherichia* and *Salmonella* spp. This group collected high-dimensional gene expression profiles of murine macrophages in culture after LPS stimulation, and through expression

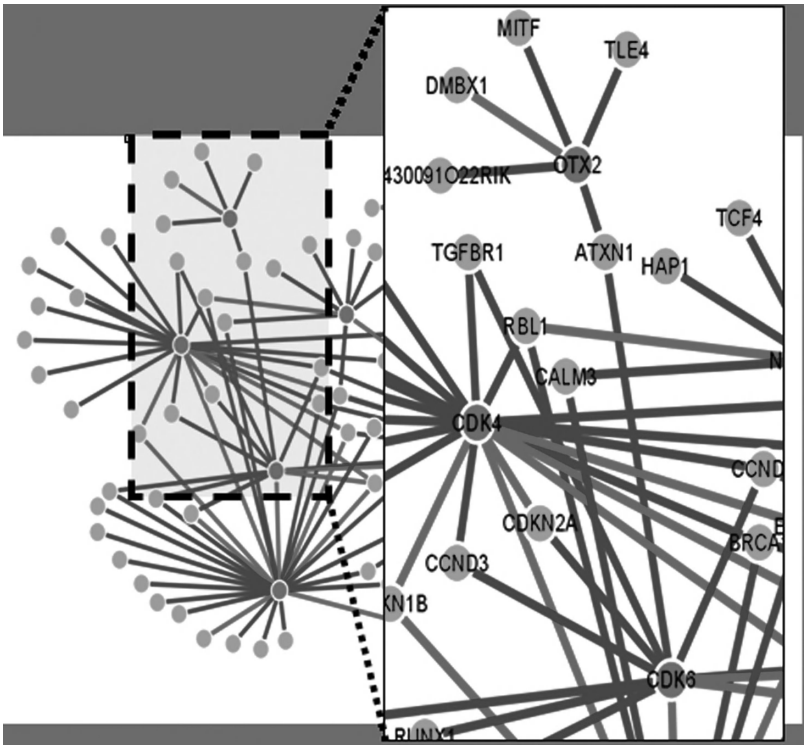


Figure 1.6 A fragment of the mouse protein-protein interaction network expanded on the basis of the human counterpart. Seed network mouse proteins and the links in the mouse protein–protein interaction network are shown in red. Nodes and links added based on orthologs in the human network are shown in green and blue, respectively. Note that two proteins that were not initially linked in the mouse Otx2 and CDK6 are now connected through ATXN1. (*See insert for color representation of this figure*.)

clustering they noticed that some clusters of genes (cluster 1) had peak levels of expression at about 4 hours post-LPS, which was after the 1-hour peak of other gene clusters (cluster 2). These peaks were transitory, with expression levels dropping rapidly after the peak was reached. Their hypothesis was that transcription factors (TF) in cluster 2 might be regulating the set of genes in cluster 1, and thus would be coregulated themselves to be induced early in the process. Examining the proximal promoter sequences of the genes in both groups, and combining that information with known TF protein–protein interaction data, they found that ATF3 TF motifs were common in these promoters, and that ATF3 was part of a complex of factors including the well-described NF-κB inflammatory master regulatory TF. To determine the role of ATF3 in regulating gene expression during early LPS response, they collected additional data on the occupancy of Rel (an activator subunit of the NF-κB TF) and ATF3 on the predicted regulatory motifs at two genes in cluster 1, Il6, and Il12, during the response to LPS. At all time points, they also determined the nuclear and cytoplasmic levels of these two TF. Using all these data, they were able to relate the levels and location of each TFs on the promoter of each gene with the activation

state of the gene and develop a kinetic model to predict the expression of a gene like Il6/Il12 given the estimates of the two TF. They found a negative correlation of expression with ATF3 promoter occupancy, indicating ATF3 was a negative regulatory factor, causing the anti-inflammatory response that brings the gene back to baseline soon after stimulation. With such modeling, they could also predict the expression response of the target genes given previously untested levels of ATF3, such as the complete absence of ATF3 function, which would be seen in an ATF3 genetic mutant. They then tested their gene expression model in macrophages from wild-type and ATF3 mutant mice, and showed that the model correctly predicted that the IL6 gene expression would rise but not return to baseline, staying maximally expressed to 6 hours post-LPS. In summary, this group's work is a paradigm for a Systems Biology approach, as they created and integrated several orthogonal but interacting datasets, they developed a model to explain the behavior of the parts of the system, and, most importantly, they tested the predictions of their model. Since 2006, this group has refined the model through similar analysis of other TF, including CEBP TFs, to further define this regulatory control (Litvak et al., 2009).

A real tour de force in systems immunology was published more recently, and exemplifies the progress in predictive regulatory control models that has been made in this field by using a systems approach (Amit et al., 2009). This group focused on the response of mouse dendritic cells to five different microbial pattern molecules. Dendritic cells are among the most important cells in the early immune response, due to their role in instructing the adaptive immune system on the identity of invading pathogens through antigen presentation. The main purposes were to collect system-wide data on a variety of responses in a model cell, to model these interactions, to perturb the defined system in a unbiased and comprehensive manner, to model the system responses to perturbation, and to integrate these data to define the interactions of genes in specific network-based responses to different pathogens. The overall scheme of the project is shown in Figure 1.7. Amit and coworkers measured RNA expression levels genome-wide over 24 hours at nine different time points after exposure to LPS, poly(I-C), and other molecules mimicking microbial molecular patterns. Cluster analysis of these gene expression patterns indicated that the response to LPS significantly represented the response to the other molecules, for which 80 different clusters of gene with similar expression patterns were identified. Thus, LPS was used in the rest of the study where selected candidate regulators were perturbed to explore the regulatory network. They made two important choices at this step. First, they used their expression data to identify 117 candidate important regulators, through assuming that regulators controlling the expression clusters would have correlated expression patterns with these clusters. Second, to simplify the evaluation of perturbing the system by knocking down expression of these regulators, they identified a set of 118 genes that would best model the entire dataset. They used an approach from information theory and minimized the conditional entropy accounted for by the expression data in sets of tester genes, incrementally adding genes and retesting the new set until the expression responses best represented the available global data. They then used shRNA technology to shut off at least 75% of each regulator's RNA level in independent experiments, and measured the effect on expression for each of the testers. A model was developed using the changes in these "signature" genes after each perturbation that connected regulators and specific targets that showed

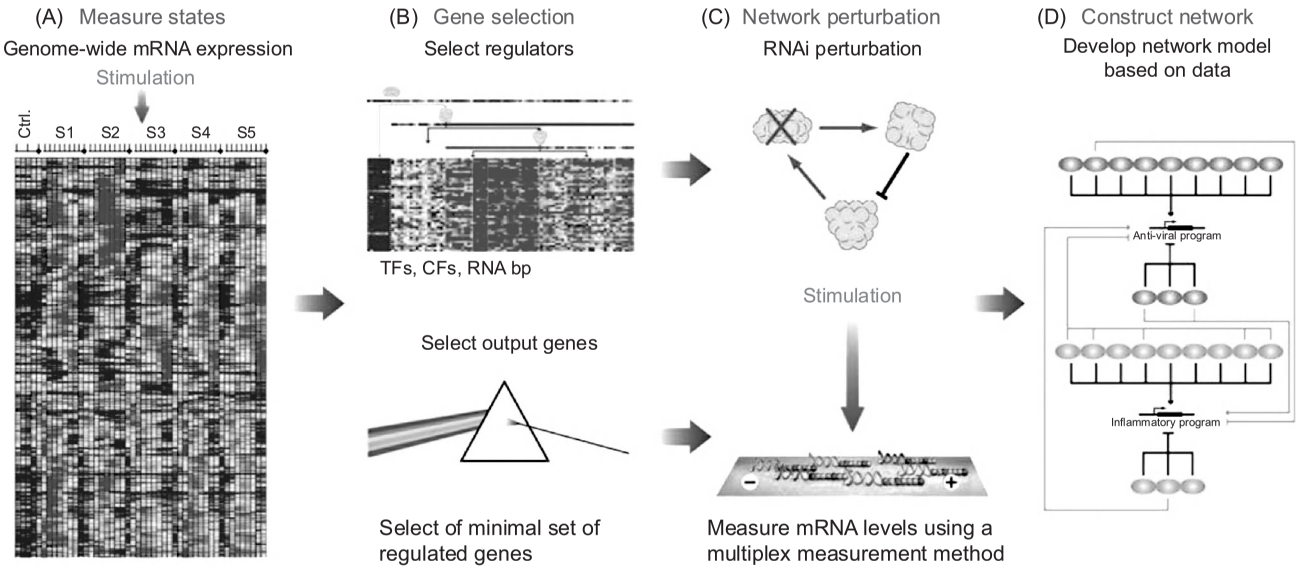


Figure 1.7 An example of an unbiased and systematic strategy for regulatory network reconstruction. Amit et al. (2009) strategy consists of four steps. (A) State measurement. They used genome-wide expression profiles under different stimuli (S1–S5), at different time points (tick marks). Rows—genes; columns—experiments; red—induced; blue—repressed; white—unchanged. (B) Gene selection. Amit et al. (2009) then identify candidate regulators that are transcriptionally regulated and predictive of the expression of gene modules (top) and select a signature of target genes that maximally represents the full expression profile (bottom). (C) Network perturbation. They then generated a functionally validated shRNA library for all potential regulators and used it to knockdown each regulator (top). Following stimulation of genetically perturbed cells (red arrow), Amit et al. (2009) measured the expression of the signature genes using the nCounter multiplex mRNA detection system (bottom). (D) Network reconstruction. Combining genome-wide expression profiles and perturbed multiplex measurements was performed to reconstruct a regulatory network associating regulators with individual targets and overall responses. (Adapted from Amit et al. (2009) *Science* **326**, 257–263. Reprinted with permission from AAAS.) (*See insert for color representation of this figure*.)

evidence of significant response to that regulator. Cumulative analysis then produced a global network with 2322 significant interactions that were approximately 75% activation and 25% repression. This network contained many known interactions, which therefore provided corroborative evidence in support of the network, but also helped uncover many novel regulatory connections, showing the value of this unbiased perturbation approach. Within the global network they recognized two large subnetworks (inflammatory and antiviral) as well as several smaller networks that controlled more specific aspects of the response. Exploring the interactions within these networks, they described coherent and incoherent interactions within small networks. The former describes multiple interactions between three or more genes in which regulatory actions are reinforcing, i.e., regulator A activates B, and both A and B activate C, creating in this example a feed-forward regulatory loop. Such coherence was found in the majority of the feed-forward regulatory loops. Importantly, through subnetwork analysis they discovered that the inflammatory network is controlled through dominant activators (defined as genes that activate more genes in inflammatory networks than in antiviral), cross-inhibitory regulators (genes that activate inflammatory target but suppress antiviral genes), and specific activators, which stimulate expression of only inflammatory network genes. In the antiviral network, they found Stat1 and Stat2 regulate many network members, which are also in turn regulated by a second tier of "middle-manager" regulators. Importantly, these analyses culminated in the development of a core network of 24 regulators, which is proposed by the authors to be coordinating responses to various pathogenic stimuli through both dominantly activating genes and cross-repression of regulatory factors not needed for a specific response. Of importance is the clear demonstration that many unknown regulatory factors were implicated in these analyses; for even the 24 core regulators, which one would expect to have been studied in innate immune response previously, nearly half (11 of the 24) were novel in such regulatory roles. Many of the downstream effectors were also newly implicated in the immune response. Amit et al. (2009) explicitly compared their perturbation approach with traditional observational approaches that rely only on correlation of regulatory expression with predicted target gene expression. In the latter, many false positives are found due to the fact that with a finite number of conditions (often partially confounded), many regulators and targets simply have an expression patterns that are too similar to be sufficiently differentiated. One caveat important to note is that the responses to the shRNA perturbations may be indirect responses and, thus, the assumption that expression changes due to treatment with a specific shRNA are due to direct effects on the target gene with altered expression is probably inaccurate for some specific regulatory–target interactions. However, it is also likely that the systems approach also provided robustness with respect to any one specific interaction proposed.

A major value of the broad-scale perturbation approach to developing interaction networks is the expectation that such approaches can be generalized to other systems. This can be done using a training dataset obtained using expression profiling and then testing the predicted relationships through large-scale perturbation experiments. While the authors indicate the costs are modest, this state-of-the-art biomedical project is an example of the goal to be aspired to by animal scientists but would be quite difficult to develop in the current agricultural research funding environment.

What Parts of Systems Biology Are in Use in Animal Science Today?

Systems Biology is a young field, especially for those interested in application to livestock biological systems. A prerequisite for molecular Systems Biology clearly is a completed genome sequence for the organism of choice. Many reports have been published over the past decade (reviewed in Dekkers and Hospital, 2002; Pomp et al., 2004; Georges, 2007; Hu et al., 2009) using whole-genome genotyping methods to find genome regions and quantitative trait nucleotides in livestock through statistical analysis of genotype and phenotype through QTL mapping, and, more recently, SNP association mapping (e.g., see Kolbehdari et al., 2009). However, in this review, we will not discuss papers analyzing structural genomics and phenotype that do not also integrate additional data such as gene expression. A search of the PubMed-indexed literature identified a number of publications describing approaches that, in the future, aspire to develop into systems biological analyses. Several do incorporate elements of integrative data modeling and analyses. In the sections below, we briefly describe those papers that report such analyses, as well as those genomics papers that have elements of integrative genomic analyses. Clearly, the availability of a draft genome sequence provides the opportunity to apply comprehensive tools and analyses. Thus, the chicken and bovine scientific communities have made progress in applying Systems Biology approaches. As well, possibly due to the closer association to genomics, animal breeding was one of the first research areas to begin to use Systems Biology methods and to discuss the value of such global approaches to animal agriculture (Green, 2009). Other disciplines of importance to animal scientists, such as physiology (Dow, 2007) and nutrition (Panagiotou and Nielsen, 2009), are also adopting the methods and tools of Systems Biology to answer questions of interest to those communities.

Chicken and Turkey

The publication of the chicken genome sequence (ICGSC, 2004) and a companion paper describing millions of genetic variants (Wong et al., 2004) ushered in the beginnings of System Biology in this species (Burt, 2005, 2007; Lamont, 2006; Cogburn et al., 2007). Associated with the chicken's status as the first livestock species to have its genome sequenced, the datasets and bioinformatics tools available for this species are advanced compared to other livestock species. For example, the chicken genome sequence was utilized at multiple levels in creating a draft sequence assembly of the turkey genome reported in 2010 (Dalloul et al., 2010). Because the chicken has been studied both as an agriculturally important species as well as a model organism for developmental biology studies, multiple communities have developed tools for analyzing and integrating genomic data that often can be applied to the chicken genome (McCarthy et al., 2007; Jupiter et al., 2009; Konieczka et al., 2009; van den Berg et al., 2009). For example, Starnet is a bioinformatics tools designed to create gene regulatory networks using user-defined gene sets. In addition to generating Pearson correlation-based networks of specific chicken datasets, Starnet has data stored for a total of ten species and can also produce interaction networks (Jupiter et al., 2009). Furthermore, a molecular interaction prediction and analysis tool named BioNetBuilder is available for analyzing data from chicken and other species

(Konieczka et al., 2009). Konieczka and colleagues have used a microarray dataset from chicken embryos as well as interaction data from other data sources (including KEGG, MINT, Biogrid, and others) to create a chicken interactome, which contains 72,000 predicted interactions among 8140 genes. Further, interactome mining and subnetwork extraction can be accomplished by the user via a Web interface (Konieczka et al., 2009).

Identification and analysis of the genetic lesion responsible for the chicken phenotype *talpid*^{3} exemplifies the integration of biological and computational approaches in chicken genomics and developmental biology (Davey et al., 2006; Bangs et al., 2010). Using classical mapping of this recessive trait, five candidate genes in the critical region were found using the annotated chicken genome sequence, and a frameshift mutation was identified in a gene without biological function, KIAA0586. Electroporation of a wildtype KIAA0585 cDNA, along with a construct expressing *Shh*, rescued gene expression patterns for markers of correct dorsoventral limb patterning affected in the *talpid*^{3} mutant. In a recent expansion of this investigation, this group has clustered gene expression data from spatial regions of wildtype and *talpid*^{3} limb tissue to find genes dependent on *talpid*^{3} function. Correlation clustering of these data identified a Hoxd13-containing cluster, which was of interest because of the known role of *HoxD* genes in limb patterning. Because overlapping spatial patterns of expression control combinatorial specification of limb structure, the authors more precisely determine three-dimensional expression patterns for many of the HoxD13 cluster genes. Using computational analysis to quantitatively compare and cluster these expression patterns, they could identify seven such spatial clusters. Six genes clustered with HoxD13, and these new relationships were used to develop a precise (time and space-specific) gene regulatory network that acts downstream of *Shh* in the chicken limb (Bangs et al., 2010).

Cattle

Several authors have described current efforts to develop datasets and tools so that their field of interest can take advantage of the promise of Systems Biology, including bovine genomics in lactation (Loor and Cohick, 2009), reproduction (Adjaye, 2005; Fazeli and Pewsey, 2008), and in animal breeding and genetics (Green, 2009). As an example of recent bioinformatics work using the cattle genome sequence for comparative analysis, Seo and Lewin (2009) have used Pathway Tools and MetaCyc to create a cattle-specific metabolic pathway database. Comparison to other mammalian metabolic pathways showed that genes for 22 metabolic enzymes with evidence for activity in mammalian systems were not present in the cattle genome sequence.

The groups of Reverter and Dalrymple have used methods and analyses that approach Systems Biology in recent publications on understanding gene regulatory networks in beef cattle (Hudson et al., 2009; Reverter et al., 2010). Indeed, their method could be viewed as a "holistic" analysis of global "differential wiring" of gene expression networks between two phenotypic states to find causative mutations that account for those differing phenotypes. The first publication using this differential wiring (DW) method compared muscle RNA expression at several fetal and postnatal stages of development, and contrasted the Piedmontese and Wagyu phenotypic differences.

These phenotypes are known to be caused by a mutation in the myostatin gene in the Piedmontese breed (Kambadur et al., 1997). The key insight here was to develop a question that would provide "myostatin" as the answer, as this gene—node in the expression correlation network—was clearly controlling the phenotypic differences in muscle seen in the these breeds. To develop a metric that would ask such a question, they first calculated a gene expression correlation network and identified all nodes whose links with other nodes changed significantly between breeds (i.e., significant DW). To find the most influential factors with DW, they then identified all nodes with annotation as a regulatory factor and connected these regulators with high DW to these genes with high phenotypic impact, measured by a so-called phenotypic impact factor (PIF). The PIF is based on both the differential expression and the absolute expression level of the gene. A highly expressed gene that is also differentially expressed is predicted to have a high phenotypic impact. Those genes calculated to have DW to high PIF genes are given a high regulatory impact factor (RIF). The gene with highest RIF in their dataset was myostatin. This work is interesting for developing the idea to ask the right question, which was "what relationship measure gives the answer 'myostatin'?" As important was the fact that this analysis of microarray data, along with known regulatory annotation of the network nodes, was performed using data on only 27 animals and a similar number of arrays. This indicates that, given the right analysis, microarray data from a very reasonably sized project can give relevant and precise answers to complex questions. Recently, Reverter et al. (2010) tested their RIF approach on several additional datasets. They showed that the RIF analysis appears universally applicable, and identified likely or known regulators controlling phenotypic differences ranging from breast cancer survival to adipocyte differentiation.

Pig

There are limited reports describing the integration of large-scale data types that utilize pig-specific data. As stated above, this is primarily due to the lack of a completed genome sequence. Many examples of transcriptional profiling have been published (reviewed in Tuggle et al., 2007), although nearly all primarily report on functional genomics level analyses rather than on significant integration of data to develop predictions of larger systems. Our group has shown that clustering of porcine gene expression data can provide information for the prediction of common regulatory control, as well as common function, through GO (Wang et al., 2008). A basic assumption for such work is that genes with correlated responses to stimulus ("coexpressed") may be functioning together to provide a cellular function important for appropriate response to a specific stimulus. The mechanisms causing such coexpression can be hypothesized to be a common regulatory factor or factors. We have used lymph node response to infection by *Salmonella* over time to identify a set of genes with the common response of activation within the first 8–48 hours post infection (Wang et al., 2008). The upregulated genes within 8–24 hours post infection were richly annotated (about 25% of the total gene list) as NF-κB target genes. We hypothesized that the remainder of the genes with this response behavior might be regulated as well by NF-κB. Analyzing for overrepresentation of NF-κB motif sequences in the promoter region of the human orthologues of these set of genes, we found many of the 75% unknown target genes

in the coexpressed list were overrepresented for NF-κB motifs. We are testing these predictions using a number of biochemical techniques, and have shown that these motifs are bound by NF-κB protein *in vitro* (O. Couture, C. Tuggle et al., unpublished observations).

Some papers report on some initial dual measurement and comparison of multiple high-dimension datasets. Hornshoj et al. (2009) showed transcriptomic data (cDNA microarrays and 454-based sequencing) and proteomics data of muscle and heart tissue. While only 148 RNAs and proteins could be identified in both technologies, this was nearly half of the proteins measured (354), and a global analysis indicated the measured levels were positively correlated (Pearson's correlation coefficient was 0.49 and 0.53 for protein level compared to microarray and 454 sequencing, respectively). Other papers have used porcine tissue in projects to understand the systems that control tissue remodeling (Popovic et al., 2009). Another report reviewed how a mathematical modeling approach (indicial response function, IRF) analysis could be used to measure the ratio of change in a specific system feature due to a specified amount of stimulus, and that IRF can be used to integrate observed complexity using convolution (Kassab, 2009).

Few reports have appeared describing the integration of porcine expression data with genetic segregation data in a population, the eQTL or "genetical genomics" approach (de Koning et al., 2005; de Koning and Haley, 2005). One report has described a "comparative systems genetics" approach to compare such data across species. Kadarmideen and Janss (2007) used a QTL mapping approach in mice to find eQTLs for cortisol levels, and a population genetic analysis of the cortisol levels in a population of pigs selected for stress that showed cortisol levels are highly heritable and that a major segregating gene is present. They then suggest using the mouse eQTL localizations to narrow the search for the equivalent controlling genes in pigs. Steibel et al. (2010) reported a preliminary analysis from an eQTL study of carcass phenotypes and muscle gene expression in F2 animals from a Duroc x Pietrain cross. Using microarray data on loin muscle tissue for more than 400 F2 animals, they identified 263 putative eQTLs associated with an oligonucleotide with a known genome map position. Three pathways (with some overlapping GO terms) associated with these eQTLs were highlighted; these included terms for lipid metabolism, posttranslational protein modification, cell cycle, DNA replication/repair/recombination, and cell death. A joint analysis found 12 genomic regions coincident for pQTL and eQTLs, which indicates a high probability that a cis-acting eQTL is present in these regions. A cis-acting eQTL is predicted when a QTL that controls expression of a specific gene maps to the physical location of that gene. A number of candidate genes for these eQTLs were identified, based on both their correlated expression levels and pQTL mapping results.

Aquaculture

While a number of papers on fish species have discussed Systems Biology approaches, such work is primarily performed in the context of ecotoxicogenomics, as fathead minnows (FHM) are an established model for measuring whole-animal effects of environmental toxins (Miracle and Ankley, 2005). Shoemaker et al. (2010) developed an *in silico* metabolic model of toxicology using FHM steroidogenesis as a paradigm,

which may be useful in reproductive physiology in land vertebrates. Interestingly, they found that not only the interactions identified as most sensitive were important in modeling network response to stimuli (which is often assumed to be the most informative interactions), but that also the robustness of the signal receiver to system noise is important. In other modeling work, Rajasingh et al. (2006) have modeled whole-animal metabolism of a specific, economically important carotenoid, astaxanthine, in Atlantic salmon using ordinary differential equations. In integrative genomics work, Nilsson et al. (2009) used correlation of expression of all genes in microarray data with known heme biosynthesis pathway genes. They found five additional genes responding specifically in the mitochondria. Using specific knockdowns of these genes in Zebrafish to test this prediction resulted in all modified fish having severe anemia (Nilsson et al., 2009).

Further Reading

Partial Listing of Online Resources for Systems Biology

Systems Biology Resources (all URLs checked October 2, 2010)

Books

1. Klipp, E., Liebermeister, W., Wierling, C., Kowald, A., Lehrach, H., & R. Herwig (2009) *Systems Biology: A Textbook,* 1st Edition. Wiley-VCH, Weinheim, Germany.
2. Alon, U. (2006) *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC, Boca Raton, Florida, USA.
3. Kaneko, K. (2006) *Life: An Introduction to Complex Systems Biology*. Springer, Berlin, Germany.
4. Palsson, B. (2006) *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, Cambridge, UK.

Books 2–4 were reviewed at:

http://www.nature.com/nature/journal/v446/n7135/full/446493a.html
See other similar books at http://systems-biology.org/resources/books/.
See also Journal Special Issues devoted to Systems Biology or network analyses:
Science July 24, 2009, 325: 405–432. Articles by Kim and Barabasi are most relevant.
Science March 1, 2002, 295: 1661–1682.

Journals

Molecular Systems Biology (http://www.nature.com/msb/index.html)
BMC Systems Biology (http://www.biomedcentral.com/bmcsystbiol/)
BMC Bioinformatics (http://www.biomedcentral.com/bmcbioinformatics/)
Bioinformatics (http://bioinformatics.oxfordjournals.org/); see also Briefings in Bioinformatics (http://bib.oxfordjournals.org/).
PLOS Computational Biology (http://www.ploscompbiol.org/home.action)

Web sites

- ***General/Software Collections***
  1. http://www.systems-biology.org/. A comprehensive portal for resources, software conferences and jobs in the Systems Biology field.
  2. http://www.bioinformatics.org/. A portal for the Bioinformatics.org user group which has the following purpose: "*We develop and maintain computational resources to facilitate world-wide communications and collaborations between people of all educational and professional levels.*"
  3. http://www.bioconductor.org/. A collection of R-based software for analyzing high-dimension biological data.
  4. http://sbml.org/Main_Page. The portal for the group that developed and uses the Systems Biology Mark-up Language, which is a machine-readable language that can be used by different software tools to represent and visualize biological models.
- ***Pathways and Data Analyses Resources***

  The Kyoto Encyclopedia of Genes and Genomes is a comprehensive database collecting pathways information for many organisms (http://www.genome.jp/kegg/)

  The Reactome project curates a large number of biological pathways, emphasizing human systems (http://www.reactome.org/)

  The GO database assigns and stores functional terms using a defined vocabulary to gene or protein entities in genomes; often using cross-species information based on structural similarity (http://www.geneontology.org/)

  Pathguide contains information about **325** biological pathway related resources and molecular interaction related resources (http://www.pathguide.org/)
- ***Courses and Educational Resources/Societies/Meetings***
  1. http://www.bioinformatics.org/wiki/Educational_services
  2. http://www.systemsbiology.org/Intro_to_ISB_and_Systems_Biology. An introduction to Systems Biology from the Institute's biomedicine perspective.
  3. The International Society for Computational Biology (http://www.iscb.org/) is a professional society of scientists interested in Computational Biology. The Society organizes several bioinformatics/computational biology meetings relevant to Systems Biology, the largest is called Intelligent Systems for Molecular Biology (ISMB).

Acknowledgments

This article is based in part on the work of C.K. Tuggle supported by the US Department of Agriculture-NRICGP (2009-35205-05192) and the Iowa State University Agricultural Experiment Station, and on the work of V. Honavar supported by the National Science Foundation, while working at the Foundation, and in part by the above grant from the United States Department of Agriculture (USDA). This article was also supported in part by an Integrative Graduate Education and Research Training (IGERT) fellowship to Fadi Towfic, funded by the National Science Foundation (NSF) grant (DGE 0504304) to Iowa State University. Any opinion, finding,

and conclusions contained in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation or the US Department of Agriculture.

References

Abba, M.C., Drake, J.A., Hawkins, K.A., et al. (2004) Transcriptomic changes in human breast cancer progression as determined by serial analysis of gene expression. *Breast Cancer Research* **6**, R499–R513.

Adjaye, J. (2005) Whole-genome approaches for large-scale gene identification and expression analysis in mammalian preimplantation embryos. *Reproduction, Fertility and Development* **17**, 37–45.

Ahmad, I., Acharya, H.R., Rogers, J.A., et al. (1998) The role of NeuroD as a differentiation factor in the mammalian retina. *Journal of Molecular Neuroscience* **11**, 165–178.

Akimoto, M., Cheng, H., Zhu, D., et al. (2006) Targeting of GFP to newborn rods by Nrl promoter and temporal expression profiling of flow-sorted photoreceptors. *Proceedings of the National Acadademy of Sciences of the USA* **103**, 3890–3895.

Alon, U., Barkai, N., Notterman, D.A., et al. (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Acadademy of Sciences of the USA* **96**, 6745–6750.

Amit, I., Garber, M., Chevrier, N., et al. (2009) Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326**(5950), 257–263.

Auffray, C., Imbeaud, S., Roux-Rouquie, M., et al. (2003) From functional genomics to systems biology: concepts and practices. *Comptes Rendus Biologies* **326**, 879–892.

Bader, G.D., Betel, D., & Hogue, C.W. (2003) BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Research* **31**, 248–250.

Bader, J.S. (2003) Greedily building protein networks with confidence. *Bioinformatics* **19**, 1869–1874.

Bader, J.S., Chaudhuri, A., Rothberg, J.M., et al. (2004) Gaining confidence in high-throughput protein interaction networks. *Nature Biotechnology* **22**, 78–85.

Baitaluk, M., Qian, X., Godbole, S., et al. (2006) PathSys: integrating molecular interaction graphs for systems biology. *BMC Bioinformatics* **7**, 55.

Bangs, F., Welten, M., Davey, M.G., et al. (2010) Identification of genes downstream of the Shh signalling in the developing chick wing and syn-expressed with Hoxd13 using microarray and 3D computational analysis. *Mechanisms of Development* **127**, 428–441.

Bar-Joseph, Z. (2004) Analyzing time series gene expression data. *Bioinformatics* **20**, 2493–2503.

Basso, K., Margolin, A.A., Stolovitzky, G., et al. (2005) Reverse engineering of regulatory networks in human B cells. *Nature Genetics* **37**, 382–390.

Bendixen, E., Danielsen, M., Larsen, K., et al. (2010) Advances in porcine genomics and proteomics—a toolbox for developing the pig as a model organism for molecular biomedical research. *Briefings in Functional Genomics* **9**, 208–219.

Berg, J., Lassig, M., & Wagner, A. (2004) Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications. *BMC Evolutionary Biology* **4**, 51.

Bernard, A. & Hartemink, A.J. (2005) Informative structure priors: joint learning of dynamic regulatory networks from multiple types of data. *Pacific Symposium on Biocomputing*, 459–470.

Blackshaw, S., Harpavat, S., Trimarchi, J., et al. (2004) Genomic analysis of mouse retinal development. *PLoS Biology* **2**, E247.

Breitkreutz, B.J., Stark, C., & Tyers, M. (2003) The GRID: the General Repository for Interaction Datasets. *Genome Biology* **4**, R23.

Brown, M.P., Grundy, W.N., Lin, D., et al. (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proceedings of the National Academy of Sciences of the USA* **97**, 262–267.

Bruggeman, F.J. & Westerhoff, H.V. (2007) The nature of systems biology. *Trends in Microbiology* **15**, 45–50.

Burt, D.W. (2005) Chicken genome: current status and future opportunities. *Genome Research* **15**, 1692–1698.

Burt, D.W. (2007) Emergence of the chicken as a model organism: implications for agriculture and biology. *Poultry Science* **86**, 1460–1471.

Carter, S.L., Brechbuhler, C.M., Griffin, M., et al. (2004) Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* **20**, 2242–2250.

Chatr-aryamontri, A., Ceol, A., Palazzi, L.M., et al. (2007) MINT: the Molecular Interaction database. *Nucleic Acids Research* **35**(Database issue), D572–D574.

Chen, K.C., Wang, T.Y., Tseng, H.H., et al. (2005) A stochastic differential equation model for quantifying transcriptional regulatory network in Saccharomyces cerevisiae. *Bioinformatics* **21**, 2883–2890.

Chen, S., Wang, Q.L., Nie, Z., et al. (1997) Crx, a novel Otx-like paired-homeodomain protein, binds to and transactivates photoreceptor cell-specific genes. *Neuron* **19**, 1017–1030.

Cheng, H., Khanna, H., Oh, E.C., et al. (2004) Photoreceptor-specific nuclear receptor NR2E3 functions as a transcriptional activator in rod photoreceptors. *Human Molecular Genetics* **13**, 1563–1575.

Churchill, G.A. (2002) Fundamentals of experimental design for cDNA microarrays. *Nature Genetics* **32**(Suppl), 490–495.

Cogburn, L.A., Porter, T.E., Duclos, M.J., et al. (2007) Functional genomics of the chicken-a model organism. *Poultry Science* **86**, 2059–2094.

Cole, S.W., Yan, W., Galic, Z., et al. (2005) Expression-based monitoring of transcription factor activity: the TELiS database. *Bioinformatics* **21**, 803–810.

Conrad, E. & Tyson, J. (2006) Modeling molecular interaction networks with nonlinear ordinary differential equations. In: *System Modeling in Cellular Biology* (Eds. Z. Szallasi, J. Stelling, & V. Periwal), pp. 97–124. MIT Press, Cambridge, MA, USA.

Dalloul, R.A., Long, J.A., Zimin, A.V., et al. (2010) Multi-platform next-generation sequencing of the domestic turkey (Meleagris gallopavo): genome assembly and analysis. *PLoS Biology* **8**(9), pii: e1000475.

Datta, A., Choudhary, A., Bittner, M.L., et al. (2004) External control in Markovian genetic regulatory networks: the imperfect information case. *Bioinformatics* **20**, 924–930.

Davey, M.G., Paton, I.R., Yin, Y., et al. (2006) The chicken talpid3 gene encodes a novel protein essential for Hedgehog signaling. *Genes & Development* **20**, 1365–1377.

de Jong, H. (2002) Modeling and simulation of genetic regulatory systems: a literature review. *Journal of Computational Biology* **9**, 67–103.

de Koning, D.J., Carlborg, O., & Haley, C.S. (2005) The genetic dissection of immune response using gene-expression studies and genome mapping. *Veterinary Immunology and Immunopathology* **105**, 343–352.

de Koning, D.J. & Haley, C.S. (2005) Genetical genomics in humans and model organisms. *Trends in Genetics* **21**, 377–381.

Dehoux J.P. & Gianello P. (2007) The importance of large animal models in transplantation. *Frontiers in Bioscience* **12**, 4864–4880.

Dekkers, J.C. & Hospital, F. (2002) The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics* **3**, 22–32.

Del Vecchio, D. & Sontag, E.D. (2007) Dynamics and control of synthetic bio-molecular networks. In: *American Control Conference*, pp. 1577–1588. New York. Available on: http://www.mit.edu/~esontag/PUBDIR/Category/conferences.html.

Dhillon, I.S., Guan, Y., & Kulis, B. (2007) Weighted graph cuts without eigenvectors a multilevel approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**, 1944–1957.

Ding, C., He, X., & Simon, H.D. (2005) On the Equivalence of Nonnegative Matrix Factorization and Spectral Clustering. pp. 606–610. *In: Proceedings SIAM Data Mining Conf, Society for Industrial and Applied Mathematics*, Philadelphia, USA.

Dorrell, M.I., Aguilar, E., Weber, C., et al. (2004) Global gene expression analysis of the developing postnatal mouse retina. *Investigative Ophthalmology & Visual Sciece* **45**, 1009–1019.

Dow, J.A. (2007) Integrative physiology, functional genomics and the phenotype gap: a guide for comparative physiologists. *Journal of Experimental Biology* **210**(Pt 9), 1632–1640.

Edwards, A.M., Kus, B., Jansen, R., et al. (2002) Bridging structural biology and genomics: assessing protein interaction data with known complexes. *Trends in Genetics* **18**, 529–536.

Eisen, M.B., Spellman, P.T., Brown, P.O., et al. (1998) Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Acadademy of Sciences of the USA* **95**, 14863–14868.

Farkas, I., Jeong, H., Vicsek, T., et al. (2003) The topology of the transcription regulatory network in the yeast, Saccharomyces cerevisiae. *Physica A: Statistical Mechanics and its Applications, Elsevier* **318**, 601–612.

Fazeli, A. (2008) Maternal communication with gametes and embryos. *Theriogenology* **70**, 1182–1187.

Fazeli, A. & Pewsey, E. (2008) Maternal communication with gametes and embryos: a complex interactome. *Briefings in Functional Genomics and Proteomics* **7**, 111–118.

Fraser, C.M., Rider, L.W., & Chapple, C. (2005) An expression and bioinformatics analysis of the Arabidopsis serine carboxypeptidase-like gene family. *Plant Physiology* **138**, 1136–1148.

Friedman, N., Linial, M., Nachman, I., et al. (2000) Using Bayesian networks to analyze expression data. *Journal of Computational Biology* **7**, 601–620.

Gardy, J.L., Lynn, D.J., Brinkman, F.S., et al. (2009) Enabling a systems biology approach to immunology: focus on innate immunity. *Trends in Immunology* **30**, 249–262.

Garey, M.R. & Johnson, D.S. (1990) *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman & Co, New York, USA.

Gavin, A.C., Bosche, M., Krause, R., et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**(6868), 141–147.

Ge, H., Walhout, A.J., & Vidal, M. (2003) Integrating 'omic' information: a bridge between genomics and systems biology. *Trends in Genetics* **19**, 551–560.

Georges, M. (2007) Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. *Annual Review of Genomics and Human Genetics* **8**, 131–162.

Ghahramani, Z. (1998) Learning Dynamic Bayesian Networks, *Lecture Notes in Artificial Intelligence*, pp. 168–197. Springer-Verlag, Berlin, Germany.

Gilchrist, M., Thorsson, V., Li, B., et al. (2006) Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4. *Nature* **441**(7090), 173–178.

Green, E.S., Stubbs, J.L., & Levine, E.M. (2003) Genetic rescue of cell number in a mouse model of microphthalmia: interactions between Chx10 and G1-phase cell cycle regulators. *Development* **130**, 539–552.

Green, R.D. (2009) ASAS centennial paper: Future needs in animal breeding and genetics. *Journal of Animal Science* **87**, 793–800.

Green, R.D., Qureshi, M.A., Long, J.A., et al. (2007) Identifying the future needs for long-term USDA efforts in agricultural animal genomics. *International Journal of Biological Science* **3**, 185–191.

Han, J.D., Bertin, N., Hao, T., et al. (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, **430**(6995), 88–93.

Hartemink, A.J., Gifford, D., Jaakkola, T., et al. (2002) Bayesian Methods for Elucidating Genetic Regulatory Networks. *IEEE Intelligent Systems* **17**, 37–43.

Hecker, L.A., Alcon, T.A., Honavar, v., et al. (2008) Using a Seed-Network to Query Multiple Large-Scale Gene Expression Datasets from the Developing Retina in Order to Identify and Prioritize Experimental Targets. *Bioinformatics and Biology Insights* **2**, 91–102.

Heng, T.S. & Painter, M.W. (2008) The Immunological Genome Project: networks of gene expression in immune cells. *Nature Immunology* **9**, 1091–1094.

Ho Sui, S.J., Fulton, D.L., Arenillas, D.J., et al. (2007) oPOSSUM: integrated tools for analysis of regulatory motif over-representation. *Nucleic Acids Research* **35**(Web Server issue), W245–W252.

Ho, Y., Gruhler, A., Heilbut, A., et al. (2002) Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature*, **415**(6868), 180–183.

Hornshoj, H., Bendixen, E., Conley, L.N., et al. (2009) Transcriptomic and proteomic profiling of two porcine tissues using high-throughput technologies. *BMC Genomics*, **10**, 30.

Hu, X., Gao, Y., Feng, C., et al. (2009) Advanced technologies for genomic analysis in farm animals and its application for QTL mapping. *Genetica* **136**, 371–386.

Hudson, N.J., Reverter, A., & Dalrymple, B.P. (2009) A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. *PLoS Computational Biology* **5**, e1000382.

Hyatt, G., Melamed, R., Park, R., et al. (2006) Gene expression microarrays: glimpses of the immunological genome. *Nature Immunology* **7**, 686–691.

ICGSC (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, **432**(7018), 695–716.

Ideker, T. (2004) Systems biology 101-what you need to know. *Nature Biotechnology* **22**, 473–475.

IHGSC (2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**(7011), 931–945.

Ihmels, J., Friedlander, G., Bergmann, S., et al. (2002) Revealing modular organization in the yeast transcriptional network. *Nature Genetics* **31**, 370–377.

Ito, T., Chiba, T., Ozawa, R., et al. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Acaddemy of Science of the USA*, **98**, 4569–4574.

Ito, T., Tashiro, K., Muta, S., et al. (2000) Toward a protein-protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proceedings of the National Acaddemy of Science of the USA*, **97**, 1143–1147.

Jeong, H., Mason, S.P., Barabasi, A.L., et al. (2001) Lethality and centrality in protein networks. *Nature*, **411**(6833), 41–42.

Jeong, H., Tombor, B., Albert, R., et al. (2000) The large-scale organization of metabolic networks. *Nature*, **407**(6804), 651–654.

Jiang, D., Tang, C., & Zhang, A. (2004) Cluster Analysis for Gene Expression Data: A Survey. *IEEE Transactions on Knowledge and Data Engineering (TKDE)* **22**, 1370–1386.

Jupiter, D., Chen, H., & VanBuren, V. (2009) STARNET 2: a web-based tool for accelerating discovery of gene regulatory networks using microarray co-expression data. *BMC Bioinformatics*, **10**, 332.

Kadarmideen, H.N. & Janss, L.L. (2007) Population and systems genetics analyses of cortisol in pigs divergently selected for stress. *Physiological Genomics* **29**, 57–65.

Kambadur, R., Sharma, M., Smith, T.P., et al. (1997) Mutations in myostatin (*GDF8*) in double-muscled Belgian Blue and Piedmontese cattle. *Genome Research* **7**, 910–916.

Kanehisa, M., Araki, M., Goto, S., et al. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Research* **36**(Database issue), D480–D484.
Kassab, G.S. (2009) A systems approach to tissue remodeling. *Journal of Biomechanical Engineering* **131**, 101008.
Kauffman, S.A. (1993) *The origins of order: self-organization and selection in evolution.* Oxford University Press, New York, USA.
Khanin, R. & Wit, E. (2006) How scale-free are biological networks. *Journal of Computational Biology* **13**, 810–818.
Kim, H., Lee, J.K., & Park, T. (2007) Boolean networks using the chi-square test for inferring large-scale gene regulatory networks. *BMC Bioinformatics* **8**, 37.
Kitano, H. (2002) Systems biology: a brief overview. *Science* **295**(5560), 1662–1664.
Kleinstein, S.H. (2008) Getting started in computational immunology. *PLoS Computational Biology* **4**, e1000128.
Klipp, E., Herwig, R., Kowald, A., et al. (2005) *Systems Biology in Practice*. Wiley-VCH, Weinheim, Germany.
Kohutyuk, O. (2007) Retina Workbench: A Flexible Database System for Manipulating and Mining Expression Data and Genetic Regulatory Networks. In: *Computer Science*, pp. (In preparation). Iowa State University, Ames, IA, USA.
Kolbehdari, D., Wang, Z., Grant, J.R., et al. (2009) A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. *Journal of Animal Breeding and Genetics* **126**, 216–227.
Konieczka, J.H., Drew, K., Pine, A., et al. (2009) BioNetBuilder2.0: bringing systems biology to chicken and other model organisms. *BMC Genomics* **10**(Suppl 2), S6.
Krogan, N.J., Peng, W.T., Cagney, G., et al. (2004) High-definition macromolecular composition of yeast RNA-processing complexes. *Molecular Cell* **13**, 225–239.
Krull, M., Pistor, S., Voss, N., et al. (2006) TRANSPATH: an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Research*, **34**(Database issue), D546–D551.
Lähdesmäki, H., Hautaniemi, S., Shmulevich, I., et al. (2006) Relationships between probabilistic Boolean networks and dynamic Bayesian networks as models of gene regulatory networks. *Signal Processing* **86**, 814–834.
Lähdesmäki, H. & Shmulevich, I. (2008) Learning the structure of dynamic Bayesian networks from time series and steady state measurements. *Machine Learning, Springer* **71**, 185–217.
Lahdesmaki, H., Shmulevich, I., & O., Y.-H. (2003) On learning gene regulatory networks under the boolean network model. *Machine Learnings* **52**, 147–167.
Lamont, S.J. (2006) Perspectives in chicken genetics and genomics. *Poultry Science* **85**, 2048–2049.
Lander, E.S., Linton, L.M., Birren, B., et al. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**(6822), 860–921.
Litvak, V., Ramsey, S.A., Rust, A.G., et al. (2009) Function of C/EBPdelta in a regulatory circuit that discriminates between transient and persistent TLR4-induced signals. *Nature Immunology* **10**, 437–443.
Liu, E.T. (2005) Systems biology, integrative biology, predictive biology. *Cell* **121**, 505–506.
Liu, J., Wang, J., Huang, Q., et al. (2006) Gene expression profiles of mouse retinas during the second and third postnatal weeks. *Brain Research* **1098**, 113–125.
Liu, R. & Agarwal, P. (2005) Computational identification of transcription factors involved in early cellular response to a stimulus. *Journal of Bioinformatics and Computational Biology* **3**, 949–964.
Lonosky, P.M., Zhang, X., Honavar, V.G., et al. (2004) A proteomic analysis of maize chloroplast biogenesis. *Plant Physiology* **134**, 560–574.

Loor, J.J. & Cohick, W.S. (2009) ASAS centennial paper: Lactation biology for the twenty-first century. *Journal of Animal Science* **87**, 813–824.

Lunney, J.K. (2007) Advances in swine biomedical model genomics. *International Journal of Biological Sciences* **3**, 179–184.

Matthews, L.R., Vaglio, P., Reboul, J., et al. (2001) Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs". *Genome Research* **11**, 2120–2126.

McCarthy, F.M., Bridges, S.M., Wang, N., et al. (2007) AgBase: a unified resource for functional analysis in agriculture. *Nucleic Acids Research* **35**(Database issue), D599–D603.

Mears, A.J., Kondo, M., Swain, P.K., et al. (2001) Nrl is required for rod photoreceptor development. *Nature Genetics* **29**, 447–452.

Milo, R., Shen-Orr, S., Itzkovitz, S., et al. (2002) Network motifs: simple building blocks of complex networks. *Science* **298**(5594), 824–827.

Miracle, A.L. & Ankley, G.T. (2005) Ecotoxicogenomics: linkages between exposure and effects in assessing risks of aquatic contaminants to fish. *Reproductive Toxicology* **19**, 321–326.

Mishra, G.R., Suresh, M., Kumaran, K., et al. (2006) Human protein reference database–2006 update. *Nucleic Acids Research* **34**(Database issue), D411–D414.

Murali, D., Yoshikawa, S., Corrigan, R.R., et al. (2005) Distinct developmental programs require different levels of Bmp signaling during mouse retinal development. *Development* **132**, 913–923.

Myers, C.L. & Troyanskaya, O.G. (2007) Context-sensitive data integration and prediction of biological networks. *Bioinformatics* **23**, 2322–2330.

Nilsson, R., Schultz, I.J., Pierce, E.L., et al. (2009) Discovery of genes essential for heme biosynthesis through large-scale gene expression analysis. *Cell Metabolism* **10**, 119–130.

Nishida, A., Furukawa, A., Koike, C., et al. (2003) Otx2 homeobox gene controls retinal photoreceptor cell fate and pineal gland development. *Nature Neuroscience* **6**, 1255–1263.

Novak, J.P., Sladek, R., & Hudson, T.J. (2002) Characterization of variability in large-scale gene expression data: implications for study design. *Genomics* **79**, 104–113.

Ogata, H., Audic, S., Barbe, V., et al. (2000) Selfish DNA in protein-coding genes of Rickettsia. *Science* **290**(5490), 347–350.

Panagiotou G. & Nielsen J. (2009) Nutritional systems biology: definitions and approaches. *Annual Review of Nutrition* **29**, 329–339.

Pe'er, D. (2005) Bayesian network analysis of signaling networks: a primer. *Science signaling Signal Transduction Knowledge Environment* **281**, pl4.

Pennesi, M.E., Cho, J.H., Yang, Z., et al. (2003) BETA2/NeuroD1 null mice: a new model for transcription factor-dependent photoreceptor degeneration. *Journal of Neuroscience* **23**, 453–461.

Picard, B., Berri, C., Lefaucheur, L., et al. (2010) Skeletal muscle proteomics in livestock production. *Briefings in Functional Genomics* **9**, 259–278.

Pomp, D., Allan, M.F., & Wesolowski, S.R. (2004) Quantitative genomics: exploring the genetic architecture of complex trait predisposition. *Journal of Animal Science* **82**(E-Suppl), E300–E312.

Popovic, N., Bridenbaugh, E.A., Neiger, J.D., et al. (2009) Transforming growth factor-beta signaling in hypertensive remodeling of porcine aorta. *American Journal of Physiology – Heart and Circulatory Physiology* **297**, H2044–H2053.

Rajasingh, H., Oyehaug, L., Vage, D.I., et al. (2006) Carotenoid dynamics in Atlantic salmon. *BMC Biology* **4**, 10.

Ravasz, E. & Barabasi, A.L. (2003) Hierarchical organization in complex networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics* **67**(2 Pt 2), 026112.

Ravasz, E., Somera, A.L., Mongru, D.A., et al. (2002) Hierarchical organization of modularity in metabolic networks. *Science* **297**(5586), 1551–1555.

Reverter, A., Hudson, N.J., Nagaraj, S.H., et al. (2010) Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. *Bioinformatics* **26**, 896–904.

Rockman, M.V. & Kruglyak, L. (2006) Genetics of global gene expression. *Nature Reviews Genetics* **7**, 862–872.

Rutherford, A.D., Dhomen, N., Smith, H.K., et al. (2004) Delayed expression of the Crx gene and photoreceptor development in the Chx10-deficient retina. *Investigative Ophthalmology & Visual Science* **45**, 375–384.

Sachs, K., Gifford, D., Jaakkola, T., et al. (2002) Bayesian network approach to cell signaling pathway modeling. *Science signaling Signal Transduction Knowledge Environment* **148**, pe38.

Salwinski, L., Miller, C.S., Smith, A.J., et al. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Research* **32**(Database issue), D449–D451.

Scheerlinck, J.P., Snibson, K.J., Bowles, V.M., et al. (2008) Biomedical applications of sheep models: from asthma to vaccines. *Trends Biotechnol* **26**(5): 259–266.

Schena, M., Shalon, D., Davis, R.W., et al. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**(5235), 467–470.

Scott, J., Ideker, T., Karp, R.M., et al. (2006) Efficient algorithms for detecting signaling pathways in protein interaction networks. *Journal of Computational Biology* **13**, 133–144.

Segal, E., Shapira, M., Regev, A., et al. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature Genetics* **34**, 166–176.

Sellner, E.M., Kim, J.W., McClure, M.C., et al. (2007) Board-invited review: Applications of genomic information in livestock. *Journal of Animal Science* **85**, 3148–3158.

Sen, T.Z., Kloczkowski, A., & Jernigan, R.L. (2006) Functional clustering of yeast proteins from the protein-protein interaction network. *BMC Bioinformatics* **7**, 355.

Seo, S. & Lewin, H.A. (2009) Reconstruction of metabolic pathways for the cattle genome. *BMC Systems Biology* **3**, 33.

Shannon, P., Markiel, A., Ozier, O., et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**, 2498–2504.

Sharan, R. & Ideker, T. (2006) Modeling cellular machinery through biological network comparison. *Nature Biotechnology* **24**, 427–433.

Shi, J. & Malik, J. (2000) Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**, 888–905.

Shoemaker, J.E., Gayen, K., Garcia-Reyero, N., et al. (2010) Fathead minnow steroidogenesis: in silico analyses reveals tradeoffs between nominal target efficacy and robustness to cross-talk. *BMC Systems Biology* **4**, 89.

Sievertzon, M., Wirta, V., Mercer, A., et al. (2005) Transcriptome analysis in primary neural stem cells using a tag cDNA amplification method. *BMC Neuroscience* **6**, 28.

Silvescu, A. & Honavar, V. (2001) Temporal boolean network models of genetic networks and their interference from gene expression time series. *Complex Systems* **13**, 54–75.

Smith, G.W. & Rosa, G.J. (2007) Interpretation of microarray data: trudging out of the abyss towards elucidation of biological significance. *Journal of Animal Science* **85**(13 Suppl), E20–E23.

Steibel, J.P., Bates, R.O., Rosa, G.J.M., et al. (2010) Global Linkage Analysis of Gene Expression of Loin Muscle Tissue Identifies Candidate Genes in Pigs. In: *9th World Congress on Genetics Applied to Livestock Production, CD-ROM Communication 0139*, Leipzig.

Stelzl, U., Worm, U., Lalowski, M., et al. (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122**, 957–968.

Stuart, J.M., Segal, E., Koller, D., et al. (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**(5643), 249–255.

Thomas, R. (1973) Boolean formalization of genetic control circuits. *Journal of Theoretical Biology* **42**, 563–585.

Tong, A.H., Evangelista, M., Parsons, A.B., et al. (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**(5550), 2364–2368.

Towfic, F., VanderPlas, S., Oliver, C.A., et al. (2010) Detection of gene orthology from gene co-expression and protein interaction networks. *BMC Bioinformatics* **11**(Suppl. 3), S7.

Tuggle, C.K., Wang, Y., & Couture, O. (2007) Advances in swine transcriptomics. *International Journal of Biological Sciences* **3**, 132–152.

Uetz, P., Giot, L., Cagney, G., et al. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**(6770), 623–627.

van den Berg, B.H., Konieczka, J.H., McCarthy, F.M., et al. (2009) ArrayIDer: automated structural re-annotation pipeline for DNA microarrays. *BMC Bioinformatics* **10**, 30.

von Luxburg, U. (2007) A tutorial on spectral clustering. *Statistics and Computing* **17**, 395–416.

Walhout, A.J. (2006) Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. *Genome Research* **16**, 1445–1454.

Wang, Y., Couture, O.P., Qu, L., et al. (2008) Analysis of porcine transcriptional response to Salmonella enterica serovar Choleraesuis suggests novel targets of NFkappaB are activated in the mesenteric lymph node. *BMC Genomics* **9**, 437.

Wong, G.K., Liu, B., Wang, J., et al. (2004) A genetic variation map for chicken with 2. *8 million single-nucleotide polymorphisms*. *Nature* **432**(7018), 717–722.

Yook, S.H., Oltvai, Z.N., & Barabasi, A.L. (2004) Functional and topological characterization of protein interaction networks. *Proteomics* **4**, 928–942.

Yu, H., Zhu, X., Greenbaum, D., et al. (2004a) TopNet: a tool for comparing biological subnetworks, correlating protein properties with topological statistics. *Nucleic Acids Research* **32**, 328–337.

Yu, J., He, S., Friedman, J.S., et al. (2004b) Altered expression of genes of the Bmp/Smad and Wnt/calcium signaling pathways in the cone-only Nrl-/- mouse retina, revealed by gene profiling using custom cDNA microarrays. *Journal of Biological Chemistry* **279**, 42211–42220.

Zak, D.E. & Aderem, A. (2009) Systems biology of innate immunity. *Immunological Reviews* **227**, 264–282.

Zhang, J., Gray, J., Wu, L., et al. (2004) Rb regulates proliferation and rod photoreceptor development in the mouse retina. *Nature Genetics* **36**, 351–360.

Chapter 2
Modeling Approaches in Systems Biology, Including Silicon Cell Models

Alexey N. Kolodkin, Fred C. Boogerd, Frank J. Bruggeman, and Hans V. Westerhoff

What Is Systems Biology

Sometimes it is understood as holism and opposed to reductionism. However, we would like to present here another point of view and show that it is reductionism and holism at the same time.

Raising the Question

Biology, by definition, is the science that focuses on the study of life and living organisms. There are many, sometimes contradictory, definitions of life. Now, at the time when artificial intelligence is no longer just a subject of fantasy literature, when artificial systems can imitate many properties of living systems, including the ability to grow and to adapt to changes in the environment, to reproduce themselves, to metabolize external substrates, the question of what is life is especially burning. As an example, vast debates around this topic can be found in a special issue of the journal *Origins of Life and Evolution of Biospheres* (Gayon et al., 2010). Although there is still no consensus and strict definition of what is life, we can say that life is something that is not present in the particles of inorganic matter (biomolecules) when they are in isolation, but something that appears when these biomolecules interact with each other in a whole cell. We would call this property *systemic*, because it is not present in the parts, but appears only in an entire system (Alberghina and Westerhoff, 2005).

In fact, what is considered as a system and is studied by biology is not necessarily limited to an entire organism. The main object of a particular study can be a small part of a living organism, e.g., a particular metabolic network. On the other hand, it can also be the biosphere as a whole, as a bigger system with many interacting organisms. In more general terms, biology studies semiopen or "metabolic systems" that selectively interact with their environment by way of mass and energy exchange, where the decrease of free energy in the environment is coupled with the increase of

Systems Biology and Livestock Science, First Edition. Edited by Marinus F.W. te Pas, Henri Woelders, and André Bannink.

the order of the biosystem itself (decreasing its own entropy) (Westerhoff and Dam, 1987). An important feature of a living system is its modular structure. For example, livestock consists of separate animals, the body of an animal consists of organs, organs consist of tissues, tissues consist of cells, and so on, down to metabolites, enzymes, proteins, lipids, DNA, RNA, and other biomolecules. Biology halts at the level of biomolecules; we can indeed stop at the level of the molecular world, where properties stemming from quantum mechanics are understood as such, or have been established empirically.

In the cascade of biological levels, one encounters several areas where systemic properties appear. For example, some systemic properties characterizing life (e.g., homeostasis, response to stimuli) appear already on the level of a single intracellular network. From there, several intracellular networks form a next level of complexity, e.g., the cell, with new systemic properties (e.g., metabolism, growth, adaptation, reproduction). Then, combination of living cells and their interactions can lead to emergence of organisms and so on.

To summarize, biology has to deal with systems and systemic properties; consequently, biology has to use a systemic approach. Then, why is Systems Biology a new science? How could biology be something different from Systems Biology (Boogerd, 2007)?

One could say that Systems Biology is a kind of biology, where experiments are accompanied by modeling. However, it is not fully true. First of all, because building and analyzing models is a fundamental component of any science. Moreover, in a very broad definition, *model* simply means a "representation of a limited part of reality with related elements," a projection of one system, e.g., the real world, to another system. Following that definition, the formation of a conditional reflex to a stimulus could also be considered as a kind of "modeling" of reality; Pavlov's dog began to salivate in response to a neutral stimulus preceding the feeding, because the real world was reflected in its nervous system. In this broader understanding of what is modeling, unconditional reflexes could also be seen as models, but models written by evolution on a "hard memory" of species. If we now narrow down the definition of modeling with respect to science only, then a model can be defined as "a way by which the real object is connected to the rationale of a scientist" and modeling as "the construction of physical, conceptual, or mathematical simulations of the real world."The essence here is that what we call scientific reality, i.e., the way we see the real object (world), is anyway just a model of it; it is the interpretation based on our theories. Consequently, either biology or Systems Biology or any other science always has to deal with modeling. If we narrow down the term modeling further to just mathematical modeling or even to exclusively computer modeling, then there is still a problem to define Systems Biology as modeling plus experimentation. There are other areas in biology, e.g., mathematical biology, that aim to use both experiments and mathematical models. However, Systems Biology means more. It is not just something plus something. Systems Biology is a conceptual approach, a new scientific paradigm. We will try to show this in the following sections. Later, after clarifying what Systems Biology is, we will be back to modeling and discuss how mathematical modeling serves Systems Biology, what the top-down, bottom-up, and middle-out approaches mean, what the silicon cell model is, and what the perspectives and the practical applications of Systems Biology are.

Emergence

Concept of Emergence

In the previous section, we have used the term *systemic properties* to describe the properties shown by the system as a whole, properties that elements lack in isolation. Now, we want to be stricter and specify the relation between the properties of a system and those of its elements. So we should introduce the term *emergent property.* In colloquial language, *emergent* simply means the act or process of rising or appearing. However, in a scientific setting, an emergent property means a property of a system that satisfies three criteria: not only (i) the thesis of systemic (organizational) property (property that should not be exhibited by elements in isolation) that restricts the type of property that may be considered emergent, but also (ii) the thesis of physical monism, and (iii) the thesis of synchronous determinism. The thesis of physical monism restricts the nature of the system's elements. It states that the system consists of only physical entities and denies any supernatural influences. The thesis of synchronous determinism restricts the way systemic properties and the system's microstructure are related to each other; it states that there can be no difference in systemic properties without changes in the structure of the system or in the properties of the components (Stephan, 2006). If all of the three theses are satisfied at the same time, then the property may be called an *emergent property*. In other words, together they constitute the minimal criteria for (weak) emergence. All other, more sophisticated, notions of emergence have their base in weak emergence.

However, if we use the definition of emergence presented above, almost all properties of a system could be considered emergent. For example, the hardness of diamond emerges from the interactions between its carbon atoms. The human mind and self-consciousness emerge from the interactions between neurons. However, we can already intuitively note the difference between the emergence of self-consciousness and the emergence of hardness in diamonds. This problem has been amply discussed in the philosophy of mind, where it has been suggested to make a distinction between *strong emergence* (self-consciousness) and *weak emergence* (hardness of diamond) by the criterion of *irreducibility* (Stephan, 2006). Weak emergence then satisfies the three theses stated above. Strong emergence would satisfy all criteria of weak emergence plus an additional one—*irreducibility*. In the words of the British emergentist philosopher C.D. Broad, irreducibility means that "the characteristic behavior of the whole could not, even in theory, be deduced from the most complete knowledge of the behavior of its components, taken separately or in other combinations, and of their proportions and arrangements in this whole" (Broad, 1925). This irreducibility would then mean that there is (strong) emergence, although Broad did not distinguish between strong and weak emergence.

Three Varieties of Irreducibility

According to the contemporary philosopher A. Stephan, a systemic property is irreducible "if (i) it is not functionally construable or reconstruable; if (ii) it cannot be shown that the interactions between the system's parts fill the systemic property's specified functional role; or if (iii) the specific behavior of the system's components,

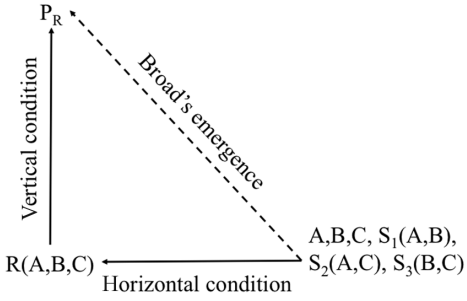


Figure 2.1 A graphic illustration of two conditions for emergence. A, B, and C are the parts making up the system. S_1(A,B), S_2(A,C), and S_3(B,C) are simpler subsystems. R(A,B,C) represents the behavior of A, B, and C *within* the system and P_R is a systemic property. The diagonal arrow represents Broad's idea of emergence. The horizontal and vertical arrows capture the two conditions implicit in Broad. (Adapted from Boogerd et al., 2005.)

over which the systemic property supervenes, does not follow from the component's behavior in isolation or in simpler configurations" (Stephan, 2006). Importantly, these varieties of irreducibility are independent of each other.

So each of these three conditions would lead to irreducibility. If condition (i) applies, properties are strongly emergent, but nothing useful can be done with this statement. Such properties are not functionally construable, because there is a lack of knowledge about the underlying mechanism. From a biological point of view, this is not an interesting case of strong (epistemological) emergence.

The other two conditions give rise to strong ontological forms of emergence. We will explain this using Figure 2.1. The triangle connects three kinds of properties. P_R represents a systemic property, which might be strongly emergent or not. A, B, and C denote parts of the system and the properties they have in isolation. S_1(A,B), S_2(A,C), and S_3(B,C) are subsystems and their properties. R(A,B,C) stands for the behavior and interactions of the parts within the system. The way Broad (1925) defined emergence can now be visualized by following the diagonal arrow: if P_R cannot be deduced from the properties of isolated elements/subsystems, it is an emergent property.

The direct relation between a systemic property and the properties of isolated elements/subsystems (the diagonal arrow) can be subdivided in two separate relationships (the horizontal and the vertical arrows). The vertical arrow describes whether or not the systemic property (P_R) can be deduced from the behavior and interactions of the parts within the system, R(A,B,C) (i.e., irreducibility (ii)). The horizontal arrow describes whether the behavior and interactions of components when in the system (R(A,B,C)) can be predicted from the behavior of components (or subsystems) in isolation (A,B,C,S_1,S_2,S_3) (i.e., irreducibility (iii)). In our opinion, the concept of strong emergence as it can be found in the philosophy of mind is related to the vertical arrow. This would imply that the specified functional role cannot be reductively explained even when there is full knowledge about the behavior of the parts within the system, represented by R(A,B,C). In other words, the property P_R is not reducible, even though in this case the underlying mechanism is completely clear in a systemic context. For metaphysicians, this then represents a case of strong emergence. However, we think that such a notion of strong emergence would introduce some vitalistic force,

and therefore, we cannot agree with this concept. In contrast, we propose that strong emergence arises along the line of the horizontal arrow (irreducibility (iii)), i.e., when R(A,B,C) cannot be deduced from the full knowledge of the behavior of the parts and their subsystems in isolation. Because complete knowledge of the subsystems—of the modularity of complexity—is allowed in the prediction base for R(A,B,C), a systemic property P_R does not become strongly emergent easily. It forms a heavy constraint on which systemic properties may be called strongly emergent. In this way, trivialization of the concept, as in the case of weak emergence, is avoided and only the biologically interesting cases of emergence remain.

Reconstruction of Emergent Properties

At this point, it is worthwhile to pay attention to the distinction between a systemic property being strongly emergent and whether or not that property can be reconstructed. If a systemic property P_R turns out to be a strongly emergent property, it does not imply that P_R cannot be reconstructed in a mechanistic model. Our claim (see the section above) is that if R(A,B,C) is known, then P_R is explainable from R(A,B,C). Consequently, if we are able to reconstruct R(A,B,C) in a mechanistic model, in which not only the complete knowledge of the parts and the subsystems is employed but also knowledge pertaining to R(A,B,C), a full mechanistic explanation of a strongly emergent P_R can be given. Thus, for reconstructing P_R, we do *not* limit ourselves in choosing the resources we have available. Here, all kinds of knowledge are allowed to be used. The behavior of part A within the system is dependent on its own relational properties, which can be determined in isolation, and also on the state of the system, i.e., the concentration of B and C. We define this (state-dependent) behavior of component A within the system as a component property of A. The same applies to parts B and C. These (state-dependent) component properties and the state-independent relational properties together constitute R(A,B,C).

We will now explain the difference in more detail in the following example. Let us consider a network of reactions as presented in Figure 2.2A. We can predict from knowledge of enzymes 1–6 in isolation that the entire system will establish a stationary state for the flux through the pathway and for the individual metabolite concentrations. For a wide range of parameter values, a steady state will indeed be established. However, for some sets of parameter values, an oscillatory state will show up instead (Figure 2.2B). This oscillatory behavior would count as a strongly emergent property of the system. Nevertheless, the occurrence of oscillations could be easily reconstructed by a simple mechanism in mathematical models, provided that at least some systemic knowledge is also available, that is, knowledge of the component properties, i.e., the actual synthesis and degradation rates of X, Y, and Z that would determine actual X, Y, and Z concentrations in the system.

Toward a Hierarchy of Strong Emergence

After having made the distinction between weak and strong emergence and having applied a strict criterion for strong emergence, we would like to elaborate on the question of how to discriminate among strongly emergent properties. How to implement

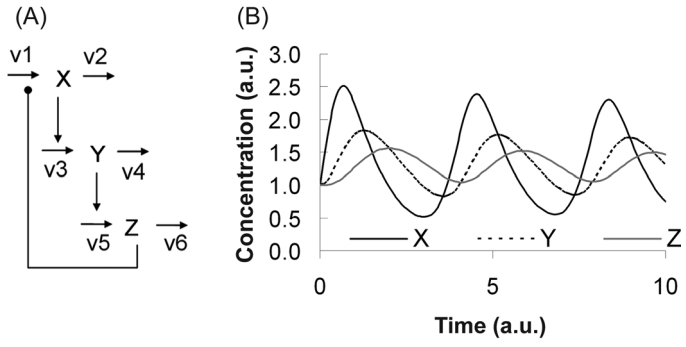


Figure 2.2 A network of enzymes and metabolites featuring oscillatory behavior. (A) The numbers 1–6 denote different enzymes that operate at a certain rate v. Metabolite X is produced in reaction 1 and degraded in reaction 2; metabolite Y is produced in reaction 3 and degraded in reaction 4; metabolite Z is produced in reaction 5 and degraded in reaction 6. Metabolite X activates the synthesis of Y; Y activates the synthesis of Z and Z inhibits the synthesis of X. (B) For certain parameter ranges the concentrations of X, Y, and Z will oscillate.

a hierarchy of strong emergence? In the following, we suggest three possible answers to this question.

As stated before, strong emergence can be reconstructed, e.g., in a mechanistic, kinetic model, if we know all relevant components and relational properties of the system's parts; the former properties are determined in part by the relational properties of the parts and in part by the state of the entire system. The state dependency of some properties provides a possible criterion for the strength of emergence. If a cellular system shows a strongly emergent property that depends on a number of discernable state-dependent properties of the system, then, in principle, the strength of emergence might be perceived as being proportional to the number of state-dependent properties needed to reconstruct the emergent property.

For the simple oscillatory behavior of the enzyme network of Figure 2.1 only a few state-dependent properties are required, and therefore, it would be positioned at the low end of a strength scale of emergent properties. We could identify four different criteria for evaluating how strong the emergence is. Each of these criteria would contain its own strength scale of emergent properties.

The thermodynamic criterion connected with the flux of energy through a cell could be one important factor contributing to the strength of emergence. When a cell grows, it requires free energy and consequently it requires a high flux through the pathway converting glucose to pyruvate; we cannot reconstruct the ability of a cell to grow without qualitative information regarding this flux. The knowledge of steady-state concentrations of intermediates is just not enough, because steady-state concentrations could be the same for different values of the flux.

A second criterion for determining the "strongness" of emergence may be the number of interactions leading to the emergence. For example, the proliferation of a tumor cell could be considered less strongly emergent than the proliferation of a normal cell, because proliferation of normal cells is determined by more regulatory processes than tumor cells.

A third criterion may be connected with the occurrence of hysteresis in a system, which makes it impossible to predict the system's state without looking at the history of that system.

The subsystemic level that functions as the prediction base for the emergent property might be a fourth criterion. If the system featuring P_R has a complex and diverse organization, often the modularity of complexity allows the discrimination of several operational levels of organization, each with their own set of characteristic subsystems. The strength of emergence might then be considered highest when knowledge of the behavior of the subsystems in the layer just below the system itself is not sufficient to deduce P_R. If for another P_R, not the first but the second level of organization harbors subsystems of which the available full knowledge is not sufficient to deduce this P_R, it constitutes a less strong form of strong emergence. Yet another P_R might be considered only weakly emergent if the prediction base is positioned at the lowest level, i.e., at the level of the isolated parts.

For the time being, the above criteria for the evaluation of the strength of emergence are just suggestions for further thinking. We do not have a firm theory on how to estimate quantitatively that to what extent strong emergence is strong. The ultimate goal would be to arrive at an overarching theory on how to integrate the above-mentioned, and most probably other, criteria into a measure for the overall strength of emergence.

Deeming the Emergence To Be Less Strong

In the case of strong emergence, we need to know much about a high number of interactions, and we need a lot of information concerning state-dependent properties of system's elements. However, in some cases, to some extent, this task can be simplified. We can consider the emergence to be less strong than it is in reality. Let us consider a hypothetical biochemical network in which a certain emergent property, e.g., homeostasis, appears from the interactions between a certain number of proteins. If one of these proteins is absent, this may result in the complete disappearance of the emergent property that we are interested in. On the level of the organism, this would imply different functioning and may result in a disease. In the simplest case, the absence of one functional protein could be caused by the mutation of a single gene. We can then immediately make the link: defect in gene—defect in protein—difference in emergent property—defect in functioning. This is what biology has been trying to do for a long time, to reduce the strength of emergence. However, we know that reality is more complicated. The majority of diseases are caused by multiple factors; they cannot be explained on the basis of a single gene—one disease paradigm. In another example, if we aim to increase milk production in a cow, we are unlikely to find a single gene completely responsible for the whole process of lactation. Again, it would be better to consider that many genes work together in the same network, many proteins interact with each other, and the phenotype emerges from all these interactions (Westerhoff et al., 2009).

So, we need to consider the higher level of emergence, we need a systemic view. In fact, the systemic view of biology has been present all the time and in fact dominant initially, as physiology. However, due to the lack of predictive power of physiology for

phenomena that we now know depend on molecular changes (e.g., mutations, cancer, AIDS), it has become an intentional and successful drive in the ages of biochemistry, molecular biology, and biophysics to get rid of the complexity of considering a system as a whole and to proceed to simpler systems completely definable in terms of one or just a few molecules. Later, thanks to the tremendous progress in genomics and molecular biology enabling the identification of all the individual macromolecules and their inherent activities, there seemed no limit to the information that could be obtained about the parts of the system. All that information could then well be added, and might in some way contribute to understanding of the whole. Indeed, cell biology has drawn schemes of macromolecular networks allowing us to bring everything into the context of a system. However, this is only a first step beyond adding the interactions; it does not reflect the dynamics and nonlinearities of the interactions, which matter much if not most.

The genome-wide dynamic network analyses have remained mostly qualitative and thereby speculative. This is due not only to the difficulty of measuring and quantitatively characterizing interactions but also to the difficulty of operating with the enormous amount of information in the genome and its expressions, amounts that cannot fit in any, even the biggest, head.

Empowered by Mathematics and Computers—the Right Moment for Systems Biology to Take Strong Emergence as Strong as It Is

Interactions between biomolecules are mostly physical and chemical reactions; they are measurable processes and it should, therefore, be possible to describe them through mathematical equations. These equations can be integrated into a mathematical model and simulated on a computer (Westerhoff et al., 2009). This can greatly empower the capacity to use genomic and molecular biological data, break through the limitations of any single human mind to operate with the large number of interactions and parameters, and in fact serve to integrate nonlinearly the activities of a great many such human brains. However, the route taken by mathematical biology has been a bit different. Mathematical biology has had the tendency to abstract away from the detail and the actual, because it aimed mostly for generic principles (Peter and Davidson, 2009).

The integration between mathematics and biology deserves higher expectations than this: on the one hand, to consider biological organism as a complex, mathematically describable system, and, on the other hand, to explain the functioning of that system in terms of specific quantitative data of interaction between its molecules; to understand how biological function, absent from macromolecules in isolation, arises when they interact in the system; to consider the emergence as strong as it is. The science aiming to fit these expectations was born about 10 years ago and named *Systems Biology* (Westerhoff et al., 2009). Systems Biology has managed to integrate historical paradigms of mathematical biology and molecular genetics (Westerhoff and Palsson, 2004). It is in this integration that Systems Biology differs from both mathematical biology and molecular genetics. Systems Biology also differs from physiology, which describes the functioning of biological systems in their entirety, without complete

reference to the components. For example, cell physiology helps describe qualitatively how ATP levels change when muscle is innervated and why this leads to contraction. It does not explain this in a mode that predicts on the basis of changes in molecular processes.

Systems Biology aims to take this step. It aims to put the interactions together into a total picture, and thereby mediate between molecular biology and physiology. Systems Biology thereby is neither holism nor reductionism, or it is perhaps both at the same time; it connects the two.

We would also like to emphasize that Systems Biology is not just computation plus experimentation and not mathematics plus biology either. Systems Biology is a conceptual approach for understanding biological complexity as such in terms of interactions between macromolecules. Mathematical descriptions and computer simulations are tools for that approach. Systems Biology is a new science with new paradigms (Westerhoff et al., 2009) not identical to any of, but arising from the integration of, physics, chemistry, and the life sciences, with the help of mathematics in order to consider the emergence as strong as it is.

Various Systems Biological Models

There are various approaches to the mathematical description of interactions in the biological systems; the simplest one may be based on graph theory. A graph is a set of objects called nodes or vertices connected by links called lines or edges. Nodes are usually attributed to different species of biomolecules and edges to interactions between these biomolecules. In an undirected graph, a line from node A to node B is considered to be equal to a line from B to A. In a directed graph, the two directions are counted as being distinct arcs or directed edges (Kestler et al., 2008).

In contrast to above qualitative models, in quantitative mechanistic models one would need to describe interactions between species of biomolecules, e.g., in terms of mass action or Michaelis–Menten kinetics. The advantage of such an approach is that the model describes the kinetics of the system, e.g., changes in the concentrations of the variable intermediates as functions of time. If reaction rates in a kinetic model are based on the real thermodynamics, this kinetic model could also be called a (nonequilibrium) thermodynamic model (Bruggeman and Westerhoff, 2006).

A model can be either simple (network of subsystems such as organs) or detailed (network of molecules). If detailed on the level of molecules, we can further distinguish macro-, meso-, and microscopic modeling. If we can neglect the limitations in the diffusion of molecules on the reaction rate and consider each species of biomolecules as a single pool, then the model would be called macroscopic and could be described as a system of ordinary differential equations in terms of ensemble averages. In fact, this is the most popular approach in modeling. However, models can also be mesoscopic (stochastic simulations of the behavior of populations of molecules) or even microscopic (tracing every molecule individually).

Hence, the classification in terms of macro-, meso-, and microscopic models is different from the classification of models *as detailed* versus *simplified.* The classification of models as *detailed* and *simplified* is based not on the way we treat the interactions between elements in a system, but on what we consider as a single element. For

example, if the organism is described as a network of molecules, the model would be called detailed. However, the same organism could also be described as a network of subsystems such as organs. Let us consider an abstract example. To make a detailed model of a house, we have to measure the properties of every brick and the interactions of each brick with all other elements. However, we could also consider a house as a structurally simplified model with just five elements: four walls and a roof. The latter model would allow manipulating the location and properties of each wall as a whole subunit and would give us an insightful understanding of how the house is built in general and what is the function of each. At the same time, we can make another model, which would describe how properties of a certain wall appear from the interactions between bricks. This gives an example of the modularity of complexity, which was discussed earlier in the context of emergence. Indeed, functional properties of a whole system emerge from interactions between subsystems and each of these subsystems is also a complex system emerging from subsystems of lower complexity.

There are many biological examples concerning this type of modularity. For example, the transport of cargo proteins between nucleus and cytoplasm is provided by a large network (about 100 reactions) involving different types of importins, exportins, and other components of the transport machinery (Macara, 2001). The effectiveness of nucleocytoplasmic transport would depend on a number of different thermokinetic aspects, including affinity of cargo for its transport protein, the quality and the state of the nuclear localization signal and the nuclear export signal of the cargo proteins, the saturation of the transport machinery with other cargoes, and the energetic efficiency of the entire process. However, a whole transport network could also be considered as a single module. The transport, being in fact the most relevant emergent property of the whole network, could be described as a single reaction with measurable rate characteristics.

Now, let us look at different cargoes transported by the same nucleocytoplasmic transport systems. Transport of each cargo can be considered as a separate module. However, these modules will be connected through the competition of binding of different cargoes to several common proteins. To find out how they control each other, we can calculate control (flux control and concentration control) and elasticity coefficients (Burns et al., 1985). The model addressing the control of fluxes and concentrations, not their magnitudes, in the relation to metabolic networks would be called a *metabolic control analysis model* (Westerhoff and Kell, 1987; van der Gugten and Westerhoff, 1997; Westerhoff et al., 2009).

Three Strategies to Build a Model: Top-Down, Middle-Out, and Bottom-Up

We have discussed that systems biological models link the layer of interacting biomolecules with the systemic functioning of the organism emerging from these interactions. There are three different strategies to build this link (Figure 2.3).

One way is the bottom-up, mechanisms-based strategy: first, one describes the actual mechanism in terms of mathematical equations, then one assigns model parameters with experimentally determined values and verifies the model by comparing

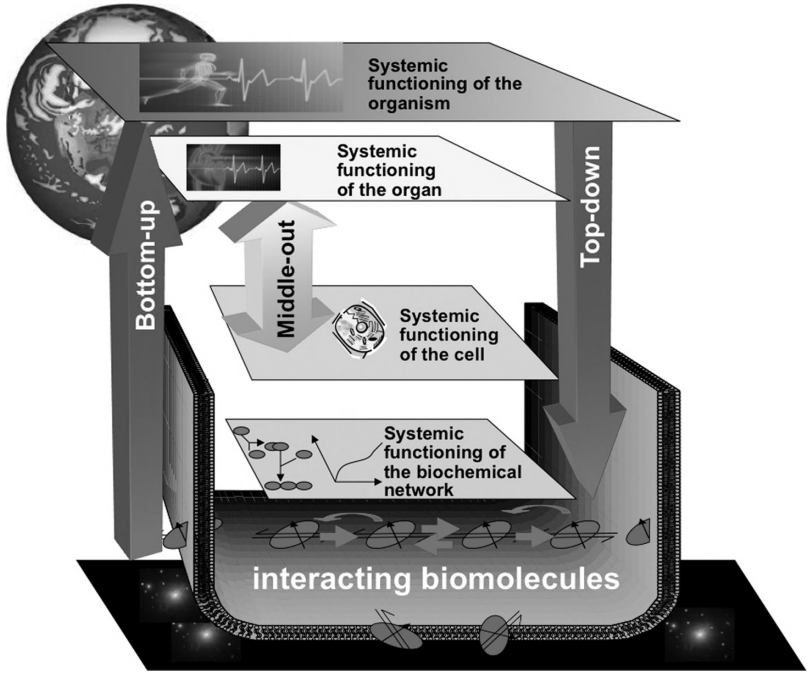


Figure 2.3 Three approaches to link the layer of interacting biomolecules to the systemic functioning: the bottom-up approach starts with experimentally characterized interactions between biomolecules; the top-down approach, in contrast, starts with the systemic functioning; the middle-out approach starts with an entity of an intermediate complexity, e.g., the cell, and goes up to the functioning of an organism and down to the interacting biomolecules.

its systemic behavior with the behavior of a real system (Bakker et al., 1997, 2000a; Rohwer et al., 2000; Westerhoff, 2001).

The term bottom-up refers to the direction chosen: from known or assumed properties of the components one deduces system functions. When bottom-up approach is applied to cell (organelle) metabolism, then the model is often called a silicon cell model (Westerhoff, 2001; Reijenga et al., 2005).

In the ideal silicon cell model, all parameter values are obtained from experimental measurements. In fact, there are only few examples of pure silicon cell models. They are still rather small and are devoted to only separate metabolic modules, e.g., glycolysis (Bakker et al., 1997, 2000a; Teusink et al., 1998; Olivier and Snoep, 2004).

Another way is to start with the systemic behavior (top-down modeling): first, one determines how the often complicated systemic function of interest varies with conditions, or in time, and from the observations one induces hypothetical structures that can be responsible for this function. This is a data-driven, "digital" approach (Lauffenburger, 2000). The system behavior is influenced (perturbed) and one then takes a top-down, bird's eye view, looking down toward system components, on a genome-wide, proteome-wide, or metabolome-wide scale. For example, when an organism is exposed to different environments, different metabolic pathways will be active; the

activity of a metabolic pathway is reflected in the phenotypic data, such as the flux through this pathway. The activity of the metabolic pathway is connected with the concentration level of enzymes that is reflected in proteomic data. Furthermore, the change of enzyme concentrations takes place due to regulation of gene expression. The latter process may be estimated from the relative mRNA levels measured with the use of microarray assays. Then, phenotypic, proteomic, and genomic date could be integrated in a model, called constraint-based reconstruction and analysis (COBRA) (Raghunathan et al., 2009). This type of model is not mechanism based, because the particular parameters of the processes (e.g., transcription or translation rates of enzymes or catalytic activities of enzymes) are not measured. On the contrary, the parameterization is obtained through fitting of the behavior of the entire model to the behavior of the real system.

A complex small system, say, a metabolic network consisting of biomolecules and exhibiting its own emergent properties, can be considered at the same time as a subsystem that is part of a larger system, like the cell. In their turn, cells interact with each other and form an even more complex system, like an organ, and so on. Consequently, the fragmentary knowledge can also be integrated in a middle-out strategy that allows modeling the behavior of a single organ or a single functional system in terms of interactions between entities of the lower, but not necessarily molecular levels of organization. For example, the emergence of electromechanical activity of the heart can be modeled in terms of interactions between complete cells as electrochemical entities, an approach that has been used in the "Physiome" project (Kohl and Noble, 2009).

The availability of data concerning single interactions and the knowledge of hypothetical mechanisms would drive the bottom-up strategy. The development of bioinformatics and the availability of large sets of measured variables drive the top-down strategy. At least one final goal of these three approaches is the same: to link physiological behavior with the underlying layer of interacting molecules. For example, when the underlying level of interacting elements in the middle-out approach reaches the level of physicochemical interactions between biomolecules and when the systemic function is extended to the whole organism, the model should be equivalent to one obtained by use of the bottom-up or top-down strategy. Analogously, the perfect top-down parameterization would make a model with the same functionality as a model built using the bottom-up approach. In the unreachable limit, it does not matter what approach is used; the final aim is a unique computer replica of the living organism for computing life on the basis of the complete biochemical, genomic, transcriptomic, proteomic, metabolomic, and cell-physiomic information (see http://www.bio.vu.nl/hwconf/Silicon/).

Perspectives of Silicon Cell Models: Advantages and Concerns

However, while we are still far from this ultimate and precise model that we may or may not achieve in our lifetime, the three types of approaches are quite different. Which would deliver most spin-offs early on, *en route* to that ideal model? Models built with top-down approaches are phenomenological; consequently, for every new experiment the entire model should be refitted to all existing experiments, allowing all

parameter values to be adjusted so as to make the fit optimal. For large models, this can become increasingly bothersome. The model based on the bottom-up approach, e.g., silicon cell model, is free from this drawback. Its parameters are "hard" in the sense that they correspond to known properties of system components (e.g., molecules). Once known, the parameter values should not change anymore unless the model is wrong. For example, once the glycolysis in human erythrocytes has been modeled correctly and if "hard" parameters have been measured adequately, we could then expect that this model could be always incorporated as one brick in a bigger, e.g., silicon human model. It is not a theoretically unsolvable problem that some kinetic properties of enzymes may be regulated or otherwise be influenced by conditions in the cell and expression levels of other enzymes (what we have called earlier state-dependent properties). At least theoretically, the "hard" parameters in the model can be expressed as a function of concentrations of other elements and of environmental factors, such as pH and temperature. Practice is not yet quite the same. Often, we do not know the values of all "hard" parameters, but we may try to run a mechanistic bottom-up model using "assumed" parameter values, because it may lead to new insights and hypotheses that can be tested empirically.

Another possible advantage of the silicon cell models is that if adjacent parts of cell function are modeled in the same terms, or in terms that can be readily translated into one another, their models can be integrated into the larger model. Building of the final model would merely mean the adequate interconnection of many existing silicon cell models, many of which may then be already available and, even more importantly, will have been validated independently. The final, "ideal" model will be just a bigger and more complex silicon cell model.

We would like to stress that, when we talk about a silicon cell model, it implies that (i) the model of the cell functioning is built with the bottom-up approach, (ii) the interacting elements of that model are placed at the level of biomolecules, and (iii) the "hard" properties of those biomolecules are based on experimentally measured parameters. However, it does not imply that the system we model is always an entire cell. We can talk as well about the silicon cell model of a particular metabolic pathway, such as glycolysis. On the other hand, the modeled system can be bigger than one cell. For example, it can be a silicon organism model.

Concerns with the building of a silicon model of an entire organism (e.g., a silicon human or silicon animal) usually refer to the "astronomical" number of interactions involved in the complete body (Noble, 2006). However, because of the modular organization of the organism, the number of interactions may be large, but not quite "astronomical." Let us show what difference modularity makes for the numbers. If we talk about a human being, and think about the interactions between the 25,000 genes in each of the 10^{14} cells of the whole body (2.5×10^{18} genes per body), then the number is pretty high, i.e., $2.5 \times 10^{18}\,!/2 \approx 10^{2 \times 10\hat{}19}$, i.e., 1 with 2×10^{19} zeros, much more than the number of atoms in the universe ($\approx 10^{80}$). If we only envisage binary interactions, the number is smaller ($2.5 \times 10^{18} \times (2.5 \times 10^{18} - 1))/2 \approx 3 \times 10^{36}$) but still enormous. However, taking into account the modular organization of the body and the fact that not everything may interact with everything else, the number of interactions becomes much smaller. Let us start with a single cell. If a cell contains about 1000 metabolic enzymes ("enzyme types" really but we assume that all enzymes defined by the same gene(s) behave as a single ensemble) and about 500 metabolites

("metabolite types" really but we again assume ensemble behavior), maximally 5×10^5 binary enzyme–metabolite interactions are possible. These are the current numbers for yeast (Herrgard et al., 2008), but although the yeast genome is approximately five times smaller than, e.g., the human genome, we do not expect high difference between organisms in terms of the number of catalyzed reactions in a single cell. Besides, 5×10^5 interactions are an overestimation since in reality not every enzyme can interact with every metabolite. It is much more likely that an enzyme interacts on average with at most five metabolites, bringing down the number of metabolic interactions to only 5000. Continuing this line of thought, there are about 3000 human transcription factor genes. If every transcription factor binds to 100 different genes, then there are about $3000 \times 100 = 3 \times 10^5$ interactions. If the average factor is much more specific, then this number could be only 10,000. Together with metabolic interactions, we approach the order of 10^4. The addition of tens of thousands of interactions on the level of transporters, receptors, and so on would not change this order of magnitude of the number of interactions in a cell substantially. Now, let us go to the intercellular level where 10^{14} cells are organized in tissues and organs, five cell types per organ. Let us say that each cell type interacts with 100 neighbors via maximally 50 metabolites (25,000 interactions), and that one organ interacts with all other 71 organs via another 50 metabolites (a little more than 3500 interactions). If we sum all interactions mentioned above, we would be still in the order of 10^5. This is indeed not a small number, but taking into account the increasing computational power, we do not see that it should cause any principle limitations. The essence of these calculations is that, if one foregoes the natural organization of living systems, the numbers of interactions appears astronomical, but with a bit of realism, these numbers turn out to become manageable within a few decades

Secondly, various strategies, such as "blueprint modeling" and the "domino approach," could serve to make the silicon cell approach more affordable. In fact, both of these tricks are two sides of the same coin—the modularity. "Blueprint" modeling is based on the fact that many modules are organized in a similar way. For example, many biochemical networks are so similar that they may be considered instantiations of the same master scheme. Consequently, only one "blueprint" master model should be built. Later, this master model could be "adjusted" for each particular system, e.g., by merely adjusting the expression levels of the enzymes. Let us discuss this by taking the example of nuclear receptor signaling. Nuclear receptors (NRs) are widely involved in the regulation of development, inflammation, and metabolism (El-Sankary et al., 2002; Carlberg and Dunlop, 2006; Ebert et al., 2006). NRs are transcription factors, shuttling between the nucleus and cytoplasm. When an NR has its ligand (e.g., a steroid hormone) bound (activated NR), it can bind to its so-called response element (a specific nucleotide sequence on the DNA) and participate in transcription initiation. To a considerable extent, the amplitude of the transcriptional response depends on the concentration of activated receptor near its response elements in the nucleus. In this way, the activation of the receptor is often connected with the increase of the concentration of NR in the nucleus resulting from ligand-dependent regulation of its nuclear import and export. For example, a glucocorticoid receptor (GR), in the absence of ligand has an almost exclusively cytoplasmic localization and almost completely shifts to the nucleus after the addition of ligand. Although to a smaller extent, the ligand-dependent shift of the intracellular NR localization is also

relevant for the vitamin D receptor (VDR), only 10% of which is located in the nucleus in the absence of ligand, and 50% in the presence of ligand. In contrast, the intracellular localization of the pregnane X receptor (PXR) hardly depends on the presence of ligand; the largest fraction of PXR is always present in the nucleus, both in the absence and in the presence of ligand. Indeed, both the fraction of NR that resides in the nucleus and the ligand-dependent regulation of that fraction are highly variable among the 48 members of the NR family. However, this does not implicate that 48 entirely different NR nucleocytoplasmic transport silicon cell models would be required. We have already stated above that nucleocytoplasmic transport involves about 100 reactions between proteins involved in the transport machinery (Macara, 2001; Pemberton and Paschal, 2005). However, most of these reactions would be the same for all cargoes (including liganded and unliganded NR). For example, the transport of importin-cargo complex through the nuclear membrane is based on the interactions of importin with the filaments of the nuclear pore complex; it does not matter which cargo is attached at that moment to the importin. All the difference between nucleocytoplasmic transport of different cargoes, including all NRs, resides in the difference in the affinity of cargo to importins and exportins. Consequently, if the nucleocytoplasmic transport model is built only for one cargo (a particular NR), changing only two parameters (the affinities of the cargo for importins and exportins) may tune the model to a different nuclear receptor.

Another strategy capable of simplifying silicon cell modeling is also connected with the view of the organism as a complex system consisting of many interacting modules. It is a "domino approach." We have already discussed that not every element of each biological module should necessarily interact with all elements of other modules. There are usually only "key" metabolites or proteins that interconnect different modules. For example, the most exchangeable metabolite is ATP (Fell and Wagner, 2000). First, we can distinguish between processes involved in ATP synthesis and ATP consumption. Then, using *in vitro* enzyme kinetic assays or modular kinetic analysis (Ciapaite et al., 2005) we can identify how these processes depend on ATP concentration and formulate a first model with the intermediate in the middle and the several processes around it. This model would predict how activation of the processes affects the concentration of the intermediates and the fluxes at steady state. We can compare our model predictions with corresponding experiments. Failure of the model to predict the observations is used to invoke either an additional process or additional metabolic intermediates. By incorporating a next additional process or metabolite one adds the next domino stone. The aim is to keep the model as simple as possible and add additional elements if the model does not yet describe the system adequately. This domino approach is currently used by the MOSES project (Micro Organism Systems Biology: Energy and *Saccharomyces cerevisiae*) (www.moses.sys-bio.net).

Both the "blueprint modeling" and the "domino approach" could be used not only as strategies to make modeling easier in the terms of the system of mathematical equations but also being very useful in organizing the logistics of building large silicon cell models. In this way, a "domino approach," is also a strategy how the results from different research groups can be integrated into the same model. "Blueprint modeling" can be highly useful in the development of online silicon cell projects. For example, it has already been successfully implemented in the JWS site. The JWS site contains several silicon cell models collected by Jacky Snoep and Brett Olivier and devoted to separate

metabolic modules. Users can run these models in a Web browser via an easy-to-use interface (http://www.jjj.bio.vu.nl/ and http://jjj.biochem.sun.ac.za/info.html) (Snoep and Olivier, 2002). The rate laws of every model are fixed, but kinetic parameters can be changed locally by the user without affecting the default values stored in the curated database on the server. What is stored is a "blueprint model." Upon obtaining either better kinetic values or values for different physiological conditions, any experimentalist can easily reparameterize the model and run simulations online. In the case of improving the model to a version to be published in any of a selected number of journals, the investigator can send results to the model curating team, so that the default "blueprint model" could be improved.

As we have already noticed, the JWS site contains separate silicon cell models; often one model corresponds to a single module of intracellular networks. For example, one model is built for yeast glycolysis, another for EGF-induced signal transduction, another one for histone-gene expression in early development, and so on. The next step is to join all these modules together into a bigger model capable of reconstructing *in silico* hierarchically higher level of strong emergence—the emergence of a whole cell, and, ideally, even further, the emergence of a whole body, an organism. However, the integration of different models, at least for the time when the authors are writing these lines, is associated with several additional challenges. We have already discussed one of them, concerning experimental measuring of high (but not astronomical), number of parameters required to determine all component properties. Another challenge is how to deal with multiple space and time scales in the same model and, finally, how to build an infrastructure most suitable for creating uniform, easily (ideally automatically) integrable models. One potential solution is related to further development of the JWS site.

Another initiative is WikiPathways (www.wikipathways.org) that is focused on integrating models (for the type being, mostly graphical representation of biochemical silicon cell networks) form different pathways into one "Google map"-looking like frame, with possibility for user-made Wiki-type changes. Another well-known initiative is E-Cell (http://www.e-cell.org/) that hosts databases for various (including silicon cell) models and provides modeling environment for simulations. The principle difference between JWS and E-Cell is that E-Cell requires downloading model to one's own computer that may become more problematic with the increase of the model's size. Another project is called Virtual Cell (www.nrcam.uchc.edu). It is focused on building Web-based modeling environment for cell biology that can be used to calculate what happens in cells (and is more powerful for spatial aspects). Virtual Cell has also some actual models, but mainly for the purpose of demonstrating the procedure. The main focus of Virtual Cell is not the model, but software packages, such as Virtual FRAP (fluorescent recovery after photobleaching), where biologist can upload their own experimental data and run virtual experiments, calculate statistics or play *in silico* with different values of experimental parameters.

Apart from public (JWS, WikiPathways, E-Cell, and Virtual Cell), there are also various commercial initiatives, for example, Symphony of Genomatica (http://www.genomatica.com/technology/platform/), which has organism- and pathway-specific information) and may also be used for kinetics simulations. Why there is so high interest in biological modeling, in spite of the fact that there is still a long way to a complete silicon cell model? We will try to answer this in the next chapter.

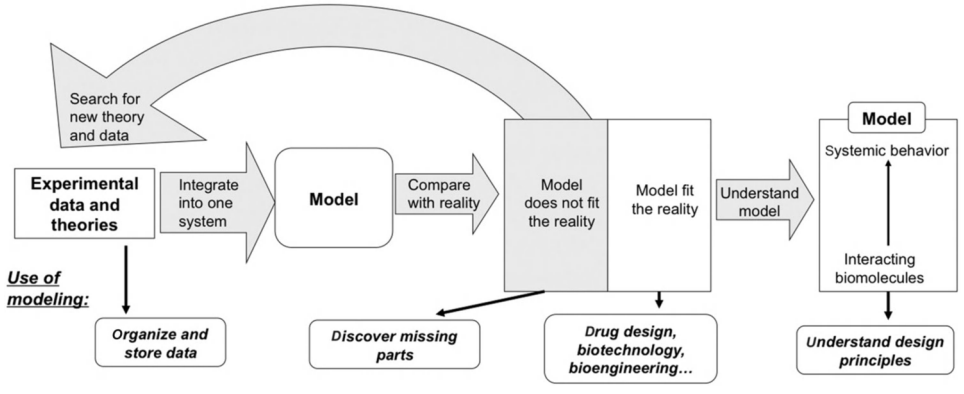


Figure 2.4 The usefulness of modeling. Modeling starts with the integration of different experimental data into one system. In this way, modeling helps to organize and store data. Mathematical descriptions of the interaction between model elements allows simulating the emergent behavior of a system as a whole; comparing the systems behavior simulated in the model with reality as observed experimentally may provoke to search for new theories and new data (especially if the model does not fit reality). If the model fits reality, it may be used in drug design, (e.g., network-based drug design), biotechnology, bioengineering, or other applications discussed in the current chapter. Studying the role of particular design features in the model may lead to understanding of system functioning and to the discovery of design principles.

Use of Systems Biological Models, Including Silicon Cell Models

An important question is the usefulness of systems biological models, or more precisely, the usefulness of modeling, because not only the final model may be useful but also the process of its construction (Figure 2.4). We will discuss this using the building of silicon cell models as an example.

The building of the model starts with the collection of experimental data and with organizing these data according to existing theories (Figure 2.4). In fact, it helps to collect and store kinetic and other information about the biological system. A good example here is the BioModels database constructed in the framework of BioModels project (Le Novere et al., 2006). BioModels database provides an access to quantitative models of biochemical and cellular systems in the SBML format (Systems Biology markup language). Each model has a link to the reference publication and gives the proper numerical results.

When the data are integrated into a single model, the model should be validated. To validate the model, the simulated systemic behavior should be compared with the behavior of a real system; this comparison is especially powerful if both the real object and its model are challenged with a series of perturbations that were not considered while the model was under construction and that were not used for parameter fitting. In many cases the predicted behavior does not fit the reality. This may then lead to the discovery of parts or mechanisms missing in the model. For example, when the detailed model of yeast glycolysis failed to bring about a steady state, an interesting discovery was made, with the concomitant identification of regulatory mechanisms to prevent the turbo explosion (Teusink et al., 1998).

Let us discuss this example in more detail. Glycolysis produces net ATP. Nonetheless, it starts with two ATP-consuming reactions, resembling a turbo charger-engine, in which exhaust gases are fed back to enhance the air-input step. One feasible reason for these activation steps is that they help to make these and the subsequent steps in the pathway thermodynamically downhill. In that sense it is a useful process, but it entails a risk of investing too much of ATP in the first reactions, so much that yeast can overaccelerate the phosphorylation of the input hexoses and die from accumulation of these compounds or from a lack of phosphate. This happened in a detailed yeast model made by Teusink and coauthors (Teusink et al., 1998). However, real yeast was found to be more robust. It led to rethinking of the model and to discovering that in reality, this turbo explosion is prevented because the first phosphorylation step (catalyzed by hexokinase) is regulated by the negative feedback loop via trehalose-6-phosphate, which is produced from glucose-phosphate by trehalose-6-phosphate synthase. Experiments on trehalose-6-phosphate synthase mutants confirmed the role of this negative feedback regulation. A variation on this theme is the discovery of possible functions of structures that are already known, for example, the role of the glycosome in preventing the glycolytic turbo explosion in the parasite *Trypanosoma brucei* (Bakker et al., 2000a).

Finally, once a model has been improved and fits the reality, it can be used for various applications. For example, using a silicon model of a metabolic network one can design how to change the metabolism of the organism in the desired direction. A fine example is the design of an *Escherichia coli* strain producing polylactic acid, a biopolymer analogous to petroleum-based polymers (Jung et al., 2010). Analogously, systems biological models could be used to change the metabolism of other organisms that could be applied in the area of livestock production, e.g., changing the metabolism of insects could make them more suitable to meet the challenge of providing protein for more than 9 billion people in the nearest future (Vogel, 2010).

Models may also become increasingly useful in drug design and especially in differential network-based drug design. For example, a silicon cell model of the known metabolic network may be used to find proper target enzymes for drugs correcting some malfunctioning of human cells. The aim could also be to kill a cancer cell (Hornberg et al., 2006) or the cell of a parasite (Bakker et al., 2000b). Vis-à-vis the development of a drug for the treatment of African trypanosomiasis (the disease caused by *T. brucei*), a silicon cell model has been developed for the glycolysis of the parasite and was compared with a model of the glycolysis of human erythrocytes. On the basis of metabolic control analysis, network targets were found where the glycolysis of the parasite was more sensitive to the inhibition than human erythrocyte glycolysis. This information remains to be used for developing drugs killing *T. brucei* with reduced side effects.

Finally, a model may help understanding the role the organization of the system plays in bringing about global functional properties. Because, apart from giving a functional explanation of how a systemic property is accomplished through the working of an underlying mechanism, one might also want to understand why the mechanism is organized the way it is, and not differently (Wouters, 2007). In other words, we are referring to the design of the system. Design can be defined as "the constellation of system components, their specific properties, and their pattern of interactions that together determine the integrated behavior of the system" (Wall et al., 2004). We can study in the model the role of certain design features in obtaining the function

of interest. In this way, *design principles* of the system may be revealed—general concepts that summarize our understanding of how the design of a system is related to its function.

If we review the examples listed above, it is clear that they deal with only a small metabolic part of an organism modeled in terms of interactions between biomolecules. However, as we have also discussed above, there are no principle limitations to build the model of a whole organism in terms of interacting biomolecules. In order to consolidate efforts in that direction, the Tokyo Declaration has been signed by leading systems biologists in February 2008 (News in Brief, 2008). The declaration aims for a computer replica of a whole human body (silicon human) to be complete for 90% by 2038. This whole-body mechanistic model of the human is highly welcome; it is difficult to overestimate its potential use for the comprehensive understanding of body functioning, for intensification of drug discovery, and for the development of patient-specific treatments.

One may expect that the development of a silicon human will be preceded by the development of silicon animals, including objects of relevance for livestock science. Indeed, due to biomolecular and functional similarity, the "silicon pig," for example, should be a most welcome "blueprint model" for a silicon human or *vice versa*. Most processes and qualitative descriptions of biomolecular interactions should be similar. They would differ in kinetic parameter values and expression levels. Reparameterization would allow us to switch the model from one organism to another, e.g., reparameterization of just a few genes could "turn a cow into a pig," in the sense that it would ovulate a multitude of oocytes per cycle (pig) or a limited number (one or two in cattle). On the one hand, this means that already early phases of a silicon human project could revolutionize livestock science, allowing for better understanding of animal functioning and opening new approaches to the intervention in animal organisms, adjusting it to human needs. On the other hand, progress in the Systems Biology of livestock animals could contribute greatly to the progress of the silicon human. Due to the virtual impossibility of experimenting on the human, at some stage the silicon human project would need to pass through a silicon animal. And here we may make an analogy with the turbo effect in glycolysis discussed earlier. At some stage, human Systems Biology should invest its "ATP" (theories, techniques, funds) into livestock animal Systems Biology, in order to receive later more "ATP" to drive the silicon human. Evolution has designed yeast. Can we design the road map for the silicon livestock?

Acknowledgments

We thank the BBSRC, EPSRC (BBD0190791, BBC0082191, BBF0035281, BBF0035521, BBF0035521, BBF0035361, BBG5302251, SySMO P49), EU-FP7 (BioSim, NucSys, EC-MOAN), and other funders for support (http://www.systembiology.net/support).

References

Alberghina, L. & Westerhoff, H.V. (2005) Systems biology: definitions and perspectives. In: *Topics in Current Genetics*, Vol. 13, (Ed. S. Hohmann). Springer, Berlin.

Bakker, B.M., Mensonides, F.I., Teusink, B., et al. (2000a) Compartmentation protects trypanosomes from the dangerous design of glycolysis. *Proceedings of Natural Academy of Science of the United States of America* **97**, 2087–2092.

Bakker, B.M., Michels, P.A.M., Opperdoes, F.R., et al. (1997) Glycolysis in bloodstream form Trypanosoma brucei can be understood in terms of the kinetics of the glycolytic enzymes. *Journal of Biological Chemistry* **272**, 3207–3215.

Bakker, B.M., Westerhoff, H.V., Opperdoes, F.R., et al. (2000b) Metabolic control analysis of glycolysis in trypanosomes as an approach to improve selectivity and effectiveness of drugs. *Molecular and Biochemical Parasitology* **106**, 1–10.

Boogerd, F.C., Bruggeman, F., Hofmeyr, J.H.S., et al. (2007) *Systems Biology Philosophical Foundations*. 1st edn. Elsevier, Amsterdam.

Broad, C.D. (1925) *The Mind and Its Place in Nature.* Routledge & Kegan Paul, London, UK.

Bruggeman, F.J. & Westerhoff, H.V. (2006) Approaches to biosimulation of cellular processes. *Journal of Biological Physics* **32**, 273–288.

Burns, J.A., Cornish-Bowden, A., Groen, A.K., et al. (1985) Control analysis of metabolic systems. *Trends in Biochemical Sciences* **10**, 16.

Carlberg, C. & Dunlop, T.W. (2006) An integrated biological approach to nuclear receptor signaling in physiological control and disease. *Critical Reviews in Eukaryotic Gene Expression* **16**, 1–22.

Ciapaite, J., Van Eikenhorst, G., Bakker, B.M., et al. (2005) Modular kinetic analysis of the adenine nucleotide translocator-mediated effects of palmitoyl-CoA on the oxidative phosphorylation in isolated rat liver mitochondria. *Diabetes* **54**, 944–951.

Ebert, R., Schutze, N., Adamski, J., et al. (2006) Vitamin D signaling is modulated on multiple levels in health and disease. *Molecular and Cellular Endocrinology* **248**, 149–159.

El-Sankary, W., Bombail, V., Gibson, G.G., et al. (2002) Glucocorticoid-mediated induction of CYP3A4 is decreased by disruption of a protein: DNA interaction distinct from the pregnane X receptor response element. *Drug Metabolism and Disposition* **30**, 1029–1034.

Fell, D.A. & Wagner, A. (2000) The small world of metabolism. *Nature Biotechnology* **18**, 1121–1122.

Gayon, J., Malaterre, C., Morange, M., et al. (2010) Defining life: conference proceedings. *Origins of Life and Evolution of Biospheres* **40**, 119–120.

Herrgard, M. J., Swainston, N., Dobson, P., et al. (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nature Biotechnology* **26**, 1155–1160.

Hornberg, J.J., Bruggeman, F.J., Westerhoff H.V., et al. (2006) Cancer: A systems biology disease. *Biosystems* **83**, 81–90.

Jung, Y.K., Kim, T.Y., Park, S.J., et al. (2010) Metabolic engineering of escherichia coli for the production of polylactic acid and its copolymers. *Biotechnology and Bioengineering* **105**, 161–171.

Kestler, H.A., Wawra, C., Kracher, B., et al. (2008). Network modeling of signal transduction: establishing the global view. *Bioessays* **30**, 1110–1125.

Kohl, P. & Noble, D. (2009) Systems biology and the virtual physiological human. *Molecular Systems Biology* **5**, 292. doi:10.1038/msb.2009.51.

Lauffenburger, D.A. (2000) Cell signaling pathways as control modules: complexity for simplicity? *Proceedings of Natural Academy of Science of the United States of America* **97**, 5031–5033.

Le Novere, N., Bornstein, B., Broicher, A., et al. (2006) BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research* **34**, D689–D691.

Macara, I.G. (2001) Transport into and out of the nucleus. *Microbiology and Molecular Biology Reviews* **65**, 570–594.

News in Brief (2008) Systems biologists hatch plan for virtual human. *Nature* **451**, 879.

Noble, D. (2006). *The Music of Life: Biology Beyond Genes*. Oxford University Press, Oxford.
Olivier, B.G. & Snoep, J.L. (2004) Web-based kinetic modelling using JWS Online. *Bioinformatics* **20**, 2143–2144.
Pemberton, L.F. & Paschal, B.M. (2005) Mechanisms of receptor-mediated nuclear import and nuclear export. *Traffic* **6**, 187–198.
Peter, I.S. & Davidson, E.H. (2009) Modularity and design principles in the sea urchin embryo gene regulatory network. *FEBS Letters* **583**, 3948–3958.
Raghunathan, A., Reed, J., Shin, S., et al. (2009) Constraint-based analysis of metabolic capacity of Salmonella typhimurium during host-pathogen interaction. *BMC Systems Biology* **3**, 38. doi:10.1186/1752-0509-3-38.
Reijenga, K.A., Van Megen, Y., Kooi, B.W., et al. (2005) Yeast glycolytic oscillations that are not controlled by a single oscillophore: a new definition of oscillophore strength. *Journal of Theoretical Biology* **232**, 385–398.
Rohwer, J.M., Meadow, N.D., Roseman, S., et al. (2000) Understanding glucose transport by the bacterial phosphoenolpyruvate: glycose phosphotransferase system on the basis of kinetic measurements in vitro. *Journal of Biological Chemistry* **275**, 34909–34921.
Snoep, J.L. & Olivier, B.G. (2002) Java Web Simulation (JWS); a web based database of kinetic models. *Molecular Biology Reports* **29**, 259–263.
Stephan, A. (2006) The dual role of 'emergence' in the philosophy of mind and in cognitive science. *Synthese* **151**, 485–498.
Teusink, B., Walsh, M.C., Van Dam, K., et al. (1998) The danger of metabolic pathways with turbo design. *Trends in Biochemical Sciences* **23**, 162–169.
Van Der Gugten, A.A. & Westerhoff, H.V. (1997) Internal regulation of a modular system: the different faces of internal control. *Biosystems* **44**, 79–106.
Vogel, G. (2010) For more protein, filet of cricket. *Science* **327**, 811.
Wall, M.E., Hlavacek, W.S., & Savageau, M.A. (2004) Design of gene circuits: Lessons from bacteria. *Nature Reviews Genetics* **5**, 34–42.
Westerhoff, H.V. (2001) The silicon cell, not dead but live! *Metabolic Engineering* **3**, 207–210.
Westerhoff, H.V. & Van Dam, K. (1987) *Thermodynamics and Control of Biological Free Energy Transduction*. Elsevier, Amsterdam.
Westerhoff, H.V. & Kell, D.B. (1987) Matrix-method for determining steps most rate-limiting to metabolic fluxes in biotechnological processes. *Biotechnology and Bioengineering* **30**, 101–107.
Westerhoff, H.V., Kolodkin, A., Conradie, R., et al. (2009) Systems biology towards life in silico: mathematics of the control of living cells. *Journal of Mathematical Biology* **58**, 7–34.
Westerhoff, H.V. & Palsson, B.O. (2004) The evolution of molecular biology into systems biology. *Nature Biotechnology* **22**, 1249–1252.
Wouters, A.G. (2007) Design explanation: determining the constraints on what can be alive. *Erkenntnis* **67**, 65–80.

Chapter 3
The IUPS Physiome Project: A Worldwide Systems Biology Initiative

Gregory O'Grady, Peng Du, and Leo K. Cheng

Introduction

The modern biological era has yielded a rapid expansion of scientific knowledge. With new technologies and research techniques constantly being developed, information is continuing to amass at an ever-increasing rate. A landmark event in this era of progress was the announcement of the initial sequencing of the human genome, which was the subject of dedicated issues of Nature and Science in 2001 (Nature, 2001; Science, 2001). The Human Genome Project was an international effort of impressive scope, ultimately culminating in the collation and distribution of complete human DNA sequence databases that are freely available via the Internet (Lander et al., 2001; Venter et al., 2001). The outcomes of the Human Genome Project are widely anticipated to underpin major advances in clinical medicine during the twenty-first century, and its success has been a testimony to the power of broad interdisciplinary and international collaboration and "think-big" science.

Following the completion of the Human Genome Project in 2003, much focus has shifted to the substantially more challenging task of integrating and interpreting the functions of the many thousands of genes that have been sequenced. As stated by the authors of the initial human genome sequence paper: "In principle, the string of genetic bits holds long-sought secrets of human development, physiology and medicine. In practice, our ability to transform such information into understanding remains woefully inadequate" (Lander et al., 2001). As a result, large-scale international efforts are now beginning to focus on convoluted tasks such as defining the regulatory mechanisms that coordinate gene expression, unravelling the RNA "transcriptome," and identifying the vast array of encoded proteins and their functions (the "proteome").

The accelerating pace of progress of genome science and molecular biology has brought another daunting challenge into sharp focus: how to reintegrate the vast mass of accumulating scientific knowledge into a coherent understanding of the structure and function of whole cells, tissues, organs, and organisms. To realize the full potential of the vast expanse of biological knowledge that has been uncovered, it is now critical to

Systems Biology and Livestock Science, First Edition. Edited by Marinus F.W. te Pas, Henri Woelders, and André Bannink.

interpret, understand, and apply this reservoir of knowledge in its complete biological context. In the words of physiome pioneer James Bassingthwaighte, "The time has arrived in biological science to put it all back together again" (Bassingthwaighte 2000a).

It is from this background that the International Union of Physiological Sciences (IUPS) Physiome Project has grown and prospered. The term physiome stems from "physio-" (life) and "-ome" (whole) (Bassingthwaighte, 1995). In its broadest expression, the IUPS Physiome Project is an international collaborative effort that seeks to provide a "quantitative description of the physiological dynamics and functional behavior of the intact organism," and moreover, of the organism in states of disease (Bassingthwaighte, 2000a, Hunter and Borg, 2003). As outlined by Bassingthwaighte: "(The Physiome) is built upon the morphome, the quantitative description of anatomical structure, chemical and biochemical composition and material properties of an intact organism, including its genome, proteome, cell, tissue, and organ structures up to those of the whole intact being. The Physiome Project is a multicenter, integrated program to design, develop, implement, test and document, archive, and disseminate quantitative information and integrative models of the functional behavior of molecules, organelles, cells, tissues organs, and intact organisms from bacteria to man" (Bassingthwaighte, 2000a).

The concept of a worldwide Physiome Project was originally presented at the 32nd IUPS World Congress in Glasgow, Scotland, in 1993, in a report to the IUPS from its Commission on Bioengineering in Physiology. The Project was officially launched after the 33rd IUPS World Congress in 1997 in Petrodvorets, Russia, following a Satellite Symposium entitled "On Designing the Physiome Project." The 34th IUPS World Congress in Christchurch, New Zealand, in 2001, established the Physiome Project. It was then designated a major focus of the IUPS for the subsequent decade. Professor Peter Hunter (Auckland Bioengineering Institute, The University of Auckland, New Zealand) was appointed Chair of the Physiome Commission of the IUPS in 2000 and, at the time of writing, was cochair with Professor Aleksander Popel (Johns Hopkins University, MD) of the recently combined IUPS Physiome and Bioengineering Committee (Hunter and Borg, 2003).

The blueprint and inspiration for the IUPS Physiome Project is derived partly from the Human Genome Project, on which the Physiome Project was intentionally modeled by its architects (Bassingthwaighte, 1995). A central Physiome Project aim has, therefore, been the development of a robust, readily assessable infrastructure, which will act as the practical framework for the mathematical modeling of molecular pathways, cells, tissues, organs, and whole-organism functions. It is envisaged that Physiome researchers worldwide will increasingly work and interact with each other through this infrastructure, such that their pooled work ultimately spans and quantitatively integrates information across all levels of biological organization from genes to the whole organism. As with the Human Genome Project, one ultimate goal is the free availability of these quantitative models via Internet repositories, in order to maximize their potential to underpin and drive future advances in clinical medicine, livestock science, and all other biological fields. The deep complexity of biological systems makes the Physiome vision an ambitious one, and its success will require unprecedented international and interdisciplinary cooperation on a grand scale (Hunter et al., 2002).

As in human biology, the pace of molecular advances has also been dramatic in livestock science, and much of the technology used for investigating human biology has been rapidly adapted to the challenges of optimizing livestock health and production. In genomics, efforts to sequence all significant large animal genomes have been greatly aided by the dramatic reduction in sequencing costs, which have reduced on average by roughly twofold per 18 months over the recent decade. The bovine genome sequencing has already been completed by the Bovine Genome Sequencing and Analysis Consortium (Elsik et al., 2009). This was obtained by more than 300 scientists in 25 countries after 6 years of effort. At the time of writing, the ovine genome has not been fully sequenced, although a detailed genetic map has been published (de Gortari et al., 1998), and a draft version of the complete genome has been produced by assembling sheep DNA sequences from information given by the genomes of other mammals (Dalrymple et al., 2007). As detailed knowledge in the livestock molecular sciences continues to unfold, the development of livestock Physiome Projects will become an essential adjunct to make full use of this information.

The purpose of this chapter is to review the origins, aims, and current status of the IUPS Physiome Project. A central focus will be to explain the core principles of the Physiome Project and the infrastructure, in order to enable researchers in the nascent field of livestock Systems Biology to use and adapt the existing infrastructure to their individual research efforts. Examples are sourced from current progress in several sections of the Physiome Project, and particularly from the Digestive Physiome, and future directions are considered.

Fundamental Principles of the Physiome Project

Systems Biology

A key concept of the Physiome Project is to promote an integrated understanding of biological entities through quantifications of Systems Biology. One definition of Systems Biology is as "a field of study that takes into account complex interactions in biological systems at different scales of biological organization, from the molecular to cellular, organ, organism, and even societal and ecosystem levels" (Popel and Hunter, 2009). Other times, a more narrow view of Systems Biology is taken, with the focus being on computational biology at the level of the kinetics of individual ion channels and biochemical pathways. By either definition, the study of Systems Biology is conceptually a holistic approach rather than the conventional reductionist approach that is employed by the majority of biological researchers, and which has been the dominant scientific approach to date. In a reductionist approach, components of a system are categorically broken down into the most basic building blocks, and it is often assumed that a system can be reconstructed by studying extensively the functions and behaviors of each block in isolation (so-called "bottom-up" thinking).

Inevitably, the interactions between different components in a biological system are complex, and the Systems Biology conception is that the functions of whole organism cannot be understood simply by studying each component individually and in isolation. A key question after the completion of the Human Genome Project is, "how do these genes contribute to the structure and function of a biological entity?" The answer to

this question cannot be explained simply by studying the effects of each gene alone in isolation, but rather, the integrated functions of these genes must be considered together at each major level of organization (e.g., cell, tissue, organ systems, and whole body). As such, biological events are known to be much more than the sum of their parts.

Studies into the lethal cardiac condition "Long QT syndrome" provide an example of the potential for mathematical Physiome-type models to effectively integrate and expand on genetic knowledge (Clancy and Rudy, 1999). Long QT syndrome is a rare inheritable condition, characterized by palpitations, syncope (collapse), and possibly sudden death. The condition is named after a diagnostic feature in patients' electrocardiograms (ECGs), i.e., lengthening of the QT interval, typically related to an underlying prolongation of the ventricular action potential. As in many other biological areas, the mechanistic understanding of Long QT syndrome underwent a revolution in the late twentieth century, due to advances in molecular medicine, in combination with an effective international registry of affected families (Moss and Kass, 2005). As a result, several specific causative genetic mutations underlying Long QT syndrome were identified, together with their consequent deleterious effects on ion channels. To model the functional consequences of these mutations, Clancy and Rudy (1999) focused on a distinct genetic class of the disease, occurring on the *SCN5A* gene (coding an element of a sodium channel in a cardiac cell), and integrated its biological effects into a biophysically based mathematical model of a cardiac cell action potential (i.e., a model that comprehensively incorporates experimentally derived details of ion channel and intracellular functions) (Luo and Rudy, 1994). Using this modeling framework, these investigators were able to explicitly link genetics to cellular pathophysiology, by quantitatively demonstrating how the *SCN5A* mutation induced its arrhythmogenic consequences at the level of whole-cell behavior (Clancy and Rudy, 1999). As pointed out by the authors: "As more idiopathic diseases are linked to congenital abnormalities, modeling strategies of this type can help bridge the gap between genetic molecular defects and their phenotypic consequences." (Clancy and Rudy, 1999) Indeed, all genetics may ultimately be linked to phenotypes through important strategies such as this, to yield integrated insights into their functions.

Another fact in favor of the Systems Biology approach is that biological functions, such as digestion, are regulated by inputs from across multiple organ systems within an organism. For example, normal gastrointestinal (GI) motility is a result of intrinsic electrophysiological, hormonal, and neuronal influences (Cheng et al., 2010). Therefore, while a reductionist approach is sufficient in the conventional sense for studying engineering mechanical systems and the "building blocks" of biological systems, a more global systems-oriented approach must be considered when quantifying integrated tissue and organ functions. This fundamental concept has provided significant strategic direction to the Physiome Project.

As demonstrated in the Long QT example above, it is of central importance in the Physiome Project's philosophy that the Systems Biology approach should be rigorously quantitative, as can only be achieved by the construction of mathematical models. As stated by Denis Noble, "Beyond a certain level of complexity (and in mathematical terms this could mean something as simple as a pair of differential equations), qualitative thinking fails us" (Noble, 2002). The laws of physical conservation are integral principles that guide the modeling of organ systems, for example, as conservation of

mass and momentum, of electrical current and charge, of gas flow in the airways, of deformation mechanics in the circulatory system, and so on. These laws inform the foundations of "continuum models" that detail organic relationships, without having to explicitly specify the detailed behaviors of all of the involved components (Hunter and Borg, 2003). Fortunately, mathematical modeling approaches have been greatly enabled in recent years by computational technical advances, such that the highly complex task of quantitatively defining and applying these models can now be efficiently performed *in silico*.

Multiscale

Knowledge in the biological sciences now encompasses vast spatiotemporal horizons: time scales span from sub-microseconds for chemical reactions to months and years for growth and development; and spatial scales span from nanometers for protein size, up to meters for the size of whole organisms. Traditional physiological research methods have been unable to quantifiably address such multiscale problems under one framework, restricting them to qualitative comparisons and evaluations between different physiological activities in the same species—effectively limiting their ability to develop and apply quantitative insights of whole-systems function, and narrowing their scopes of practice.

Modern computational hardware and software advances, in conjunction with mathematical modeling, have been instrumental catalysts in overcoming this major research constraint in modern physiology. As a result, mathematical multiscale modeling representations have now grown to become a central approach adopted in the Physiome Project to advance the understanding and applications of complex biological interactions. It should be clear that no single model could encompass the vast spatiotemporal horizons discussed above; rather, a multiscale model is typically the combination of several related models applied in unison. In addition, different information is relevant at each level of biological function. For example, at the cell level, models are dominated by the complex biochemistry of cell proteins; in contrast to the physical laws and continuum models that dominate at the tissue and organ levels.

One of the key advantages of the multiscale representation is that it mimics the biological hierarchy of the fundamental cellular contributions to tissue and organ functions. Therefore, in a multiscale model, implementations to the model framework could be imposed at any biophysical scale to represent a realistic intervention to the biological hierarchy that the particular biophysical scale represents, such that effects on other scales could then be quantified. Thus, a multiscale model is much more than a one-way framework. If we conceptually treat the model components at each scale collectively as a "module," which could be studied as a self-contained "functional unit" in a system (explained in detail in the following section: modularity and compatibility), then a multiscale representation can be thought as a collage of modules, with modules interacting within as well as in-between different scales to represent the behaviors of the system as a whole.

Figure 3.1 provides an example of the vast temporal and spatial horizons encountered by researchers in the digestive system, with specific reference to evaluating the basis for GI motility. The same Physiome modeling principles apply to all other

organ systems: a multiscale approach is needed to accurately capture the physiological activity of the system, with module models at each scale adapted and linked under mathematical frameworks to generate physiologically quantifiable results. New information "emerges" or is relevant at each level of biological organization, and this detail must be accounted for if the models are ultimately to be physiologically realistic and predictive. In parallel, it is also ideal that models should be simplified at each new level of biological organization, such that they retain mainly the essential information to explain the phenomena relevant to that scale.

Modularity and Compatibility

The idea of widely compatible modularity is a further major principle adopted in the Physiome Project (Bassingthwaighte et al., 2009). A capable multiscale model should be structured in such a fashion that as knowledge improves, model components in modules can be updated, corrected, and augmented.

The separation and definition of module can be somewhat arbitrary, but a general rule is that each module should contain within it a self-sufficient set of functions that represent a biological process that can be studied in isolation in a system (Bassingthwaighte et al., 2009). While such a criterion may seem a reductionist approach rather than a Systems Biology approach, one must also accept the reality that the accrual of modern knowledge regarding the majority of physiological functions has been established from centuries of empirical observations in *in vitro* specimens isolated from their *in vivo* states. Therefore, it is imperative that when discussing a complex biological system such as the digestive system, we should not presume that the behaviors of a module and its impacts on the whole system can be extrapolated from its self-contained functions alone.

As an example of modularity, many smooth muscle cells contain within their cell membranes a type of L-type (named after its relatively long activation time) calcium channel, which is thought to contribute significantly to the whole-cell electrical activity (Corrias and Buist, 2007). This type of L-type calcium channel can be treated as a module in a cell model of the smooth muscle cell. Each L-type module does not necessarily represent a physical L-type ion channel, but rather the collective actions of all of the L-type channels in a smooth muscle cell form a module of L-type conductance (Corrias and Buist, 2007). The functions of this L-type conductance are studied in experiments by chemically inhibiting other types of ion conductances in the cell membrane, while subjecting the cell to extraneous sources of current, in a technique known as the patch-clamping technique (Hodgkin and Huxley, 1952). There are many other types of ion conductances in smooth muscle cells, and if each is quantified and conceptualized as an independent module, then the whole-cell electrical activity can be represented as a system of these modules, as has been achieved for gastric smooth muscle cells (Corrias and Buist, 2007).

Expanding on this example of gastric electrical activity is also useful when considering modularity across multiple scales. At the subcellular level, a module may represent the intracellular reactions that lead to the opening and closing of individual gating channels, or of a protein kinase that mediates intracellular calcium cycling. In the normal state, the kinetics of these subcellular level modules are inextricably linked

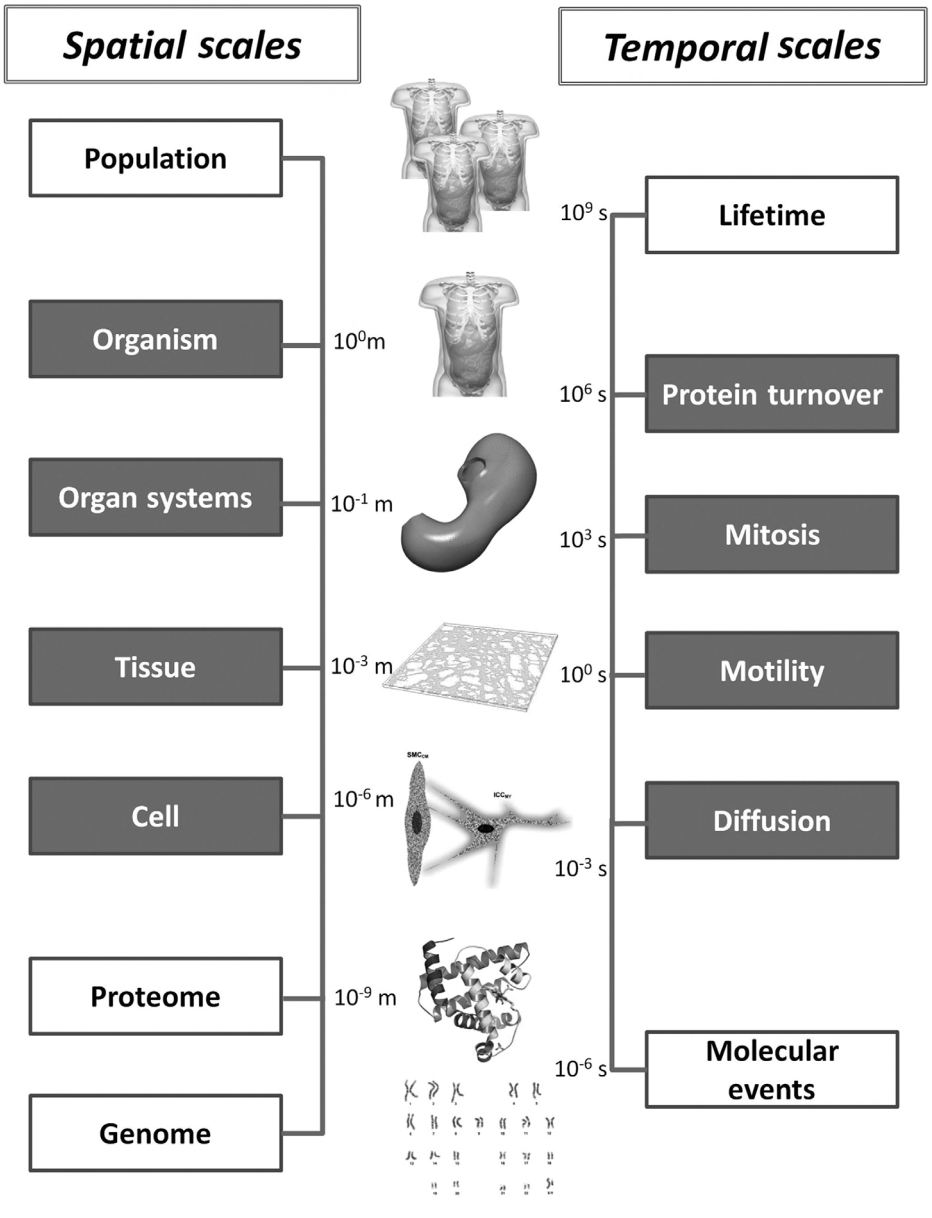


Figure 3.1 Illustration of the vast temporal and spatial scales modeled within the Physiome Project, with examples from the Digestive Physiome. The scope of the Physiome Project spans the shaded boxes and complements other key large-scale research trends in modern biology, such as genomics, the transcriptome, and proteomics. Inputs from these fields routinely inform Physiome modeling, for example, the "modules" of genome and mRNA sequences being created as functional models in the synthetic biology community can map directly into cell-components in Physiome models (Popel and Hunter, 2009). The large variation in both spatial and temporal scales is evident with ranges of 10^9 m and 10^{15} s. The developing field of "bioinformatics," which refers to computational analysis of complex protein, signaling and other metabolic pathways, will also be instrumental in facilitating the growth of the Physiome Project. (Adapted from Hunter and Borg, 2003.)

to each other, such that together they give rise to the manifestation of whole-cell electrical events. At the cellular level, a module may represent the electrical event of a smooth muscle cell, as previously detailed. At the tissue level, a module may represent the regional variations in the distributions of electrically active cellular networks. There are important differences in the population of these electrically active cells and functions in the stomach, for example, experimental recordings have suggested that in several mammalian stomachs, the gastric fundus (superior portion) is largely electrically quiescent, reflecting its role in food storage rather than food mixing. A whole-organ model of the stomach must incorporate modular components from all of these lower level details in order to accurately represent the total functions of the organ.

In order to represent the contractions that are induced by the electrical activity, another system of modules would need to be further incorporated, again including cellular-level components (e.g., contractile elements), tissue components (e.g., muscle fiber orientations), and organ-level components (e.g., regional differences in muscle thickness). As the stomach is filled with food, the mechanical properties of the fibers may change, and the module representing muscle fiber may be substituted or updated by another module suited to the role of changing fiber mechanical properties during filling. Therefore, it can be seen that effectively representing higher order functions must generally involve integrating a vast amount of relevant deeper knowledge of distinct categories and types.

In practice, modules are often viewed as computational units, which must be easily accessible and modifiable and readily archived in computational databases of biological functions. A given module, e.g., the L-type conductance module, contains a set of mathematical equations that quantitatively describe the response of the L-type conductance under experimental conditions. An equation may represent the activation (opening) or the inactivation (closing) response of a channel to voltage, and a system of those equations results in a complicated gating kinetic (with a mixture of opening and closing responses) of a channel to cell membrane potential. Should the understanding of the response of the L-type channel change, these mathematical equations can be updated to reflect this change, while retaining the overall integrated functions of the L-type module in the whole-cell system. Furthermore, the parameters in the modules can be related to realistic physical quantities such as temperature and chemical concentrations, and hence the environment in which the module is subjected to can be represented accurately by updating the values of these parameters.

Multiscale models are organized in an inherently hierarchical fashion and the boundary of a hierarchy is typically given according to the physical anatomy of the system (Bassingthwaighte et al., 2009). For example, the highest hierarchy with the largest physical scales may be the body scale for an individual animal. The modules at this level may represent the overall physical attributes of the body mass. The body level module can also comprise a set of tissue level modules such as fat and muscle compositions, and structural information such as bone densities. While higher level modules are somewhat generic in their specific functional descriptions, they nevertheless are necessary for representing an overall integrated function that reflects and integrates the contributions of lower level modules. By contrast, most basic level modules are highly specific in the descriptions of their particular functions. An example of a basic level module in a multiscale model of gastric electrical event would be a particular

intracellular receptor mechanism responsible for cycling calcium ions from intracellular stores into the bulk cytoplasm space. If this receptor pathway was blocked in a multiscale model, then it imposes a perturbation in the calcium cycling at the cell-level modules, the effect of which can eventually be an aberrant generation of gastric electrical events at the organ level.

There are a number of challenges facing a module-based implementation of multiscale modeling framework. Some of these challenges are theoretical but most are due to pragmatic considerations concerning with the limitations of computation power and implementation. Technically, as Bassingthwaighte et al. (2009) pointed out, module compatibility within the same level and between levels require standardization in the design and coding implementation. The ability to reproduce published equations and formula from scientific papers into computer code is essential for communication of modules. Units on either side of the equation must be dimensionally consistent and adhere to the Système International (SI) standard. In practice, such standardization is difficult to achieve without a curator of model databases. One such example is the CellML initiative, which aims to store and exchange computer-based mathematical modules (more detail covered under section: The framework and strategies of the Physiome Project). Ultimately, the success of this facet of the Physiome Project relies on the willingness of researchers and scientists around the world to commit to modular standards that will allow the creation of individual models that can be integrated into multiscale frameworks as widely and efficiently as possible.

Another major consideration of multiscale model is module reduction. A major impetus for the reduction of modules in a system is that it allows more efficient computation. In addition, a more explicit conclusion can often be gained from a simpler system. As stated by Dennis Noble, "Models are partial representations. Their aim is explanation: to show which features of a system are necessary and sufficient to understand it . . . The power of a model lies in identifying what is essential, whereas a complete representation would leave us just as wise, or ignorant, as before." (Noble, 2002)

However, what is always not so clear is what degree of simplification is appropriate. For example, if the core temperature of the body is presumed to change very slightly and slowly over a long temporal scale, then is it necessary to include a module of temperature in every system in question, or can temperature often be simply treated as a constant value? More importantly, if we were to include a modular description of temperature, how do we compute such a slow-changing module (hours) in a system that may also contain very fast-changing submodules (milliseconds)? The same principle applies to many spatial considerations as well. Considering the length of the GI tract, is it computationally feasible to simulate the digestive process along the entire GI tract while also including modules that describe activities at the molecular scale? Inevitably, judicious design and application of modules are essential, and this is part of the "art" of modeling. As pointed out by Hunter and Borg (2003), model detail may be best guided by the principle of Occam's Razor ("entities must not be multiplied beyond what is necessary"), and therefore, in general, model detail should be kept to the minimum level that is necessary to convey the functions of the system (Bassingthwaighte et al., 2009).

In practice, the level of detail included in any particular module is obviously also highly dependent on the span of biological scales that the overall system embodies. An example of a generic cell model that consists of many submodules is illustrated in

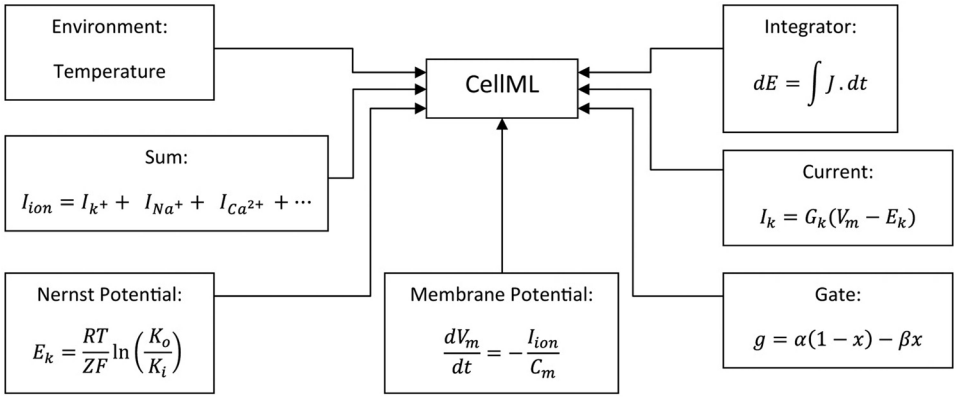


Figure 3.2 Illustration of an example of modular representation of a generic biophysical cell model. There are a number of modules in this generic cell model, and they provide an example of mapping of lower to higher level modules. One module is used to quantify the Nernst potential of the potassium ion (E_k) while another quantifies the gating variable (g). The ion conductance (I_k) module is quantified through relating the Nernst potential module, the membrane potential (V_m), and the gating variable modules; the membrane potential module relates to the sum of a number of ion current modules (I_{ion}). The model is solved using the integrator module. (Adapted from Wimalaratne et al., 2009.)

Figure 3.2. At higher levels, a hierarchical arrangement is generally used, such that the lowest level modules are complete quantitative descriptions, while the higher level modules provide correct information for the relevant physiological question, but are substantially reduced in detail to be more efficient representations for larger scale computational efficiency.

Finally, while it may be necessary for us to conceptually separate and divide the functions of a biological system into modules to understand and apply them, we should always remember that in reality the blocks we view as modules are an ensemble of a large system that can never really function properly without the support and interactions of its constituents as a whole.

Anatomically Based Modeling

In order to be physiologically realistic, it is critical that Physiome models include relevant modular information on anatomical form. Bassingthwaighte (2000b) has likened this need to a "return to the study of anatomy as the basis for physiology"—the two disciplines once being deeply interconnected, but having subsequently drifted apart as a result of the dramatic progress of *in vitro* research approaches.

Without anatomical detail, physiological models often risk being overly simplistic, because the structural orientations and relationships between biological elements substantially dictate their physiological functions. When attempting to model any particular function, it is often found that major gaps in current scientific knowledge regarding biological structure are brought into sharp focus. While the reductionist inquiries into molecular systems have often been rigorous, the structural relationships between these components are often less well quantified. One important by-product

of modeling is that the process serves to define these knowledge gaps and to direct experimental efforts toward them.

The need to incorporate structural detail may be as relevant to a particular function at the cellular levels as it is at the tissue level. For example, the spatial variation of material properties (e.g., distributions of collagen, gap junctions, and ion channels) are critical determinants of cellular function. In future, work toward anatomically based cell models, whereby all physiological functions are largely quantified by empirically derived structural considerations, is likely to be a productive research direction. As stated by Bassingthwaighte (2000b), "no longer can the cell be treated as a well-stirred chemical tank."

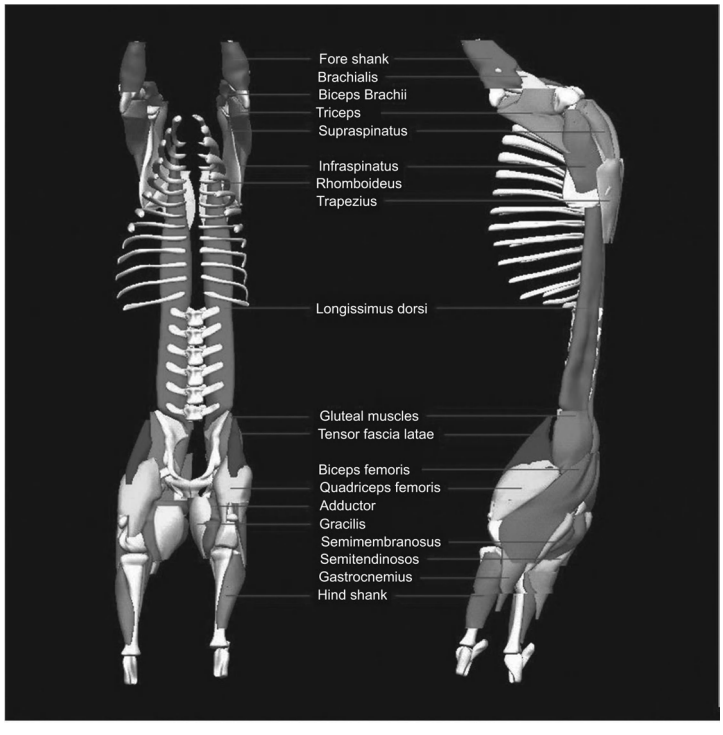
Advances in medical physics have been a critically important development for enabling the Physiome Project (Hunter and Borg, 2003). In both clinical settings and the lab, it is now possible to rapidly obtain high-resolution (millimeter to micrometer range) structural information with techniques such as magnetic resonance imaging (MRI), computed tomography (CT), ultrasound, and positron-emission tomography (PET) scanning. The Visible Human Project (see Table 3.1), which provides macroscopic datasets of male and female human anatomical structures in freely available online databases, presents the most readily available high-resolution resource, and has already been widely applied in many simulations (Spitzer and Whitlock, 1998). An example of the application of CT and Visible Human Data is in the creation of geometric models of the GI-tract organs, which have been used to interpret and study electrical and magnetic fields (Pullan et al., 2004; Cheng et al., 2007, 2010). In another example, a pulmonary research group have used CT images of the ovine lung as a framework to inform physiologically accurate solutions of soft-tissue mechanical deformations and airflow dynamics during ventilation (Tawhai et al., 2006).

Fewer imaging and anatomical resources are readily available to livestock scientists. Practical difficulties are encountered, such as access to high-resolution medical imaging devices for animal work, and the size and capabilities of these devices often being incompatible with livestock. However, the options, availability, and use of anatomical datasets in livestock science are currently expanding (Sandu et al., 2010).

Figure 3.3 shows an example of anatomically realistic models of sheep and pig carcasses that were derived from CT images. These models provide quantitative information about the anatomy of the bone and muscle surfaces from a model carcass as well as the connectivity between each group. Such "virtual carcasses" present the opportunity to develop improved methods for meat dissection and for marketing and training purposes (Crocombe et al., 1999; Bodley, 2000).

The need to define tissue structures in adequately high-resolution to provide a useful basis for physiological simulations has driven the uptake and development of innovative technological solutions to fill existing knowledge gaps. Micro-CT and MR microscopy have proved highly useful advances in this regard, achieving resolutions down to almost the cellular level (Sandu et al., 2010). PET scanning offers sensitivities in the picomolar range. Meanwhile, optical imaging techniques and multimodal approaches offer significant future potential.

Another interesting example of research in this area is the recent development of an automated imaging rig that allows serial confocal imaging and progressive thin sectioning of extended stained tissue volumes, with automated reconstruction of the tissue block achieved via a computer linked to the robotic translation stage (Sands et al., 2005). Simulations run on cardiac tissue blocks imaged in this way have been



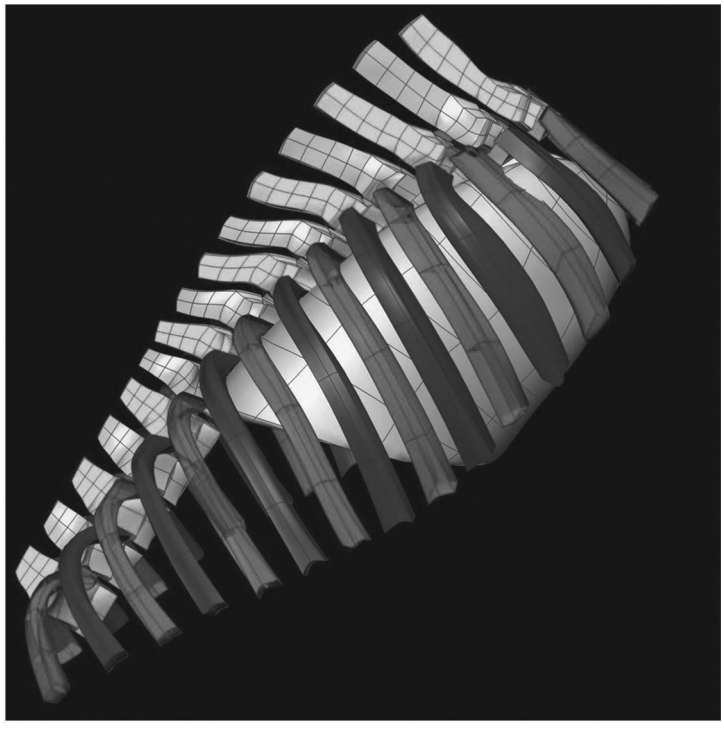


Figure 3.3 Anatomically realistic finite element models of lamb carcass and pig rib cage constructed from CT images. Shown is (upper) a lamb model consisting of 128 muscle and bone groups and (lower) the rib cage of a pig containing 45 vertebrae and ribs. (Lamb carcass image reproduced from Bodley, 2000.)

Table 3.1 Physiome-type infrastructure and resources currently available.

Name	Description	URL
NCBI	National Center for Biotechnology Information	http://www.ncbi.nlm.nih.gov/
GenBank	NIH genetic sequence database	http://www.ncbi.nlm.nih.gov/Genbank/
Visible Human	The Visible Human Project	http://www.nlm.nih.gov/research/visible/
KEGG	Kyoto Encyclopedia of Genes and Genomes	http://www.genome.jp/kegg/
Pathway-KEGG	Searches KEGG for pathways of interest	http://www.asgbioinformatics.wur.nl/
Gene Ontology	Gene Ontology database	http://www.geneontology.org/
CellML	Biophysical model exchange language and model repository	http://www.cellml.org/
SBML	Systems Biology Markup Language for representing biological processes	http://sbml.org/
BioModels	Community standards for biological models and SBML model repository	http://biomodels.net/
MML	Mathematical Modeling Language	http://nsr.bioeng.washington.edu/jsim/docs/MML.html
FieldML	Exchange language for defining generalized mathematical fields	http://fieldml.org/
JSim	Simulation environment for physiology	http://nsr.bioeng.washington.edu/jsim/
Chaste	Cancer, Heart, and Soft-tissue environment	http://web.comlab.ox.ac.uk/chaste/
Continuity	Simulation environment for multiscale modeling.	http://www.continuity.ucsd.edu/
CMISS	Mathematical modeling and visualization environment	http://www.cmiss.org/
SimTK	Software and model repository for physics-based simulations of biological structures	http://simtk.org/
OpenCell	Simulation environment for CellML models	http://www.cellml.org/tools/opencell/

used to understand the effects of complex tissue architectures on the propagation of electrical activity, providing an excellent example of multiscale biophysically and anatomically applied modeling to "bridge scale gaps" to provide original physiological insights (Trew et al., 2006). For example, such work is now being applied to understand cardiac defibrillation strategies. Large-scale structures, such as a total human gastroesophageal junction segment have also been imaged in this way

with unprecedented high definition, using a digital camera instead of a microscope (Yassi et al., 2011).

Ultimately, the key to successful anatomically based modeling is "the ability to understand and represent structure–function relationships," and this is true at all levels of biological organization from proteins to whole organs (Hunter and Borg, 2003). Repositories of mathematical models that couple detailed structural and functional relationships will be one the most valuable outcomes of the Physiome Project.

Experimental Evidence, Guidance, and Validation

Another fundamental principle of the Physiome Project is that each mathematical parameter included in each model should be informed, where possible, by empirical experimental evidence. This need has already been emphasized in the above sections: in order for a physiological model to be accurate, reliable, and predictable, it must be appropriately informed by reliable high-quality functional and structural experimentally derived detail. Models cannot simply be built from engineering or mathematical first-principles under the assumption that the result will approximate reality.

In practice, it is not always possible to inform or verify models from experimental evidence alone, because often that evidence is not available or has not been discovered. In fact, as pointed out above, it is often the case that modeling aims will lead to the direct identification of knowledge gaps, leading to new experiments and driving new technologies.

In other cases, applied modeling is able to inform experimental work by presenting new insights and theories, and by specifically identifying likely experimental targets. For example, *in silico* studies are now being used to guide therapeutic drug development. This work is receiving significant interest from pharmaceutical companies hoping to dramatically reduce the massive costs associated with bringing each new drug to market. As another example of this type of application, models are also being applied in the experimental field of gastric electrical stimulation (applied to treat gastric motility disorders and obesity), in order to refine experimental protocols and reduce the total experimental animal burden needed to achieve scientific outcomes (Cheng et al., 2010). Like drug development, gastric electrical stimulation is an experimentally demanding field of research, because there a wide range of variables such as stimulus pulse width, pulse amplitude, pulse on-off timing, location, and number of stimulation leads. By using biophysical cell models, and incorporating these models into tissue frameworks, combined modeling and experimental teams have been able to closely replicate experimental stimulation outcomes *in silico*. Subsequent work is now underway to apply the models to intelligently reduce the vast field of possible stimulation protocols to a smaller number of theoretically promising protocols for targeted experimental studies (Du et al., 2009a, 2009b).

For a model to be regarded as an accurate and reliable representation of a physiological function, it must be carefully validated in experimental studies. Thus, the experimental guidance of modeling links in at multiple stages in model creation: in reliably informing all integral elements of a model where possible, and then once the description of the model is thought to be complete, predictive simulations can be carried out and compared against objective experimental evidence.

While current research in the Physiome Project and Systems Biology has been largely focused on outputs related to human clinical medicine, the knowledge of these studies are certainly relevant and transferable to livestock science. There is also a volume of research being performed on livestock animals (e.g., porcine and ovine models) as part of the translation process from *in vitro* investigations to medical science, and much of this work will be instrumental in accelerating progress in various parts of livestock Physiome Projects.

Open, Readily Available Access

The Physiome Project encompasses more than simply research and science. It includes the archiving, dissemination, and organization or information. Of particular relevance, in our view, is the need to establish public access to data and models derived as part of ongoing research. One problem in medicine and biology is that much relevant information is simply difficult to retrieve.

A fundamental cornerstone of the Physiome Project involves enabling collaborative developments between interested research groups. Rapidly developing technologies such as high-speed Internet, grid and cloud computing, video conferencing, and collaborative document sharing are facilitating effective collaborative research. To avoid continually, "reinventing the wheel," the ability to review, (re)use, develop, and efficiently apply prior work is highly desirable. The successful decision for the majority of Human Genome Project investigators to publish and share information on annotated gene sequences, obtained through publicly funded research, should be followed in the Physiome Project. The alternative of patenting and therefore "locking up" and restricting the application of outputs would invoke a whole host of ethical, scientific, and socioeconomic dilemmas, as would have occurred with regard to the data derived from the Human Genome Project.

However, the public release of these data also raises its own range of ethical and logistical issues. One key issue is the importance of scientific integrity and social responsibility. Existing methods of publishing experimental observations and models are notorious for populating the literature with partly outdated or incomplete representations, arising in part from the translation into the manuscript form, as part of typesetting process, or as a result of the long periods between initial submission and final publication. A globally accepted standard solution to these challenges may be regarded as a key factor for promoting widespread use and trust in online Physiome resources. For these resources to be equally available to users from all around the world, reliability is also a key issue, such that mirrored dissemination sites located around the world may need to be considered.

The Framework and Strategies of the Physiome Project

In the foundation years of the Physiome Project, significant effort has been devoted to establishing a comprehensive modeling infrastructure to support effective simulation, model sharing, and international collaboration. This effort has specifically focused on the building of model databases and computational tools (see Table 3.1) (Bassingthwaighte, 2000a). In a task as grand and complex as the Physiome Project,

any particular research group can only be expert in a small portion of the work, and a framework of common working languages is essential to pull together, assimilate, and disseminate the tools and results from and to the many groups focused on different parts of the whole.

Fundamental to implementing this framework is the standardization of scientific and programmatic vocabularies to reduce the growing heterogeneity of terms used to formalize the description of both experimental data and mathematical models of physiological processes from different organ systems presented by different research groups. The ultimate demonstration of the importance of standardization of units between groups was when NASA lost a $125 million Mars Climate Orbiter because one part of the engineering team used metric units while another part used imperial units to prescribe the power provided by the thrusters of the Orbiter. Another example is when incompatible software caused significant delays in the production of the Airbus A380. In this case, it seemed the incompatibilities formed when German and Spanish Airbus facilities continued to use version 4 of the design package CATAI, while British and French sites had migrated to version 5.

In physiological terms, different schools of thoughts have argued over decades over the terminology of a great number of variables in many specialist fields. As writer Felipe Fernandez-Armesto pointed out, "the same things are called by different names by rival schools and the same terms are assigned conflicting meanings." (Fernandez-Armesto, 1997) A key aspect of the Physiome Project is to standardize the naming conventions and unit measurement of biological functions, because only then can the fragmented knowledge produced by many different research groups be truly applied in an integrated fashion.

To facilitate the understanding of Systems Biology, the Physiome Project community has developed and maintained a number of key databases and programming environments (Popel and Hunter, 2009). Significant current efforts focus on developing markup languages for encoding models, including metadata and data, application programming interfaces based on the markup languages, libraries of open source tools that can read and write markup encoded files, data and model repositories based on markups, and implementing work flows that enable model results to be reproduced (Popel and Hunter, 2009). Examples of Physiome model encoding standards and infrastructures are listed in Table 3.1. Additional information about each resource can be obtained via the home page for each project.

Two major model repositories have been developed: Biomodels (primarily containing SBML pathways models) and the CellML repository. In particular, the CellML repository serves an important purpose of reproducing results of published cell models in a consistent manner. As published biophysically based cell models are beginning to encompass a system many ordinary differential equations (ODEs), it is becoming difficult to represent the results within a paper version of the work. For example, a modified version of the Corrias and Buist interstitial cell of Cajal (ICC) model contains a system of 18 differential equations and as many as 67 cell parameters (Corrias and Buist, 2008; Du et al., 2010).

To further complicate the issue, cell models are often programmed in different programming languages, and interpreted and solved in different software environments. Therefore, there is also a significant need for a standard coding environment and curated databases for published cell models and results. Accordingly, the CellML

repository also allows scientists to share and reuse models even if they are using different model-building software, ensuring the reproducibility of the model and accelerating model building. Ultimately, the aim is that curated models should be made available both at the time of peer review and after publication of an associated peer-reviewed manuscript.

Recently, there has been a drive to also develop common metadata standards for annotating the deposited models with detailed biophysical information, and the SBML and CellML groups are now working together to develop such standards. This metadata serves several purposes. Firstly, it provides a readily accessible means to establish the sources and correctness of the model's derivation. Secondly, it smoothes the process of combining models into composite models, in order to achieve vertical, i.e., multiscale, or horizontal integration of the models in further applications. Another metadata standard presently under development is SED-ML (the Simulation Experiment Description Markup Language) (MIASE, 2009.), which seeks to encode the numerical algorithms and associated parameters for running a model simulation. One drawback of the metadata process is that it can be a time-consuming task for the developer of the models, presenting a disincentive to complete the process in ideal depth.

In addition to the markup coding standards and model repositories, software frameworks are also currently being developed by the Physiome Project for solving the equations encoded by the markup languages, particularly in multiscale models that incorporate structural and anatomical information (Popel and Hunter, 2009). As per the Physiome Project principles, these frameworks are also open source and based on internationally collaborative efforts (refer to Table 3.2; readers are referred to each projects' Web site for more specific and up-to-date details).

Future work in Physiome framework development includes developing appropriate standards and databases for model parameter sets, and including the ability to record the provenance of those parameters. There are currently no mechanisms for annotating the experimental origin of these parameters, and their dependence on species, temperature, pH, etc., is often obscure. Having the models and data available in standardized formats, with clearly stated dependencies, will improve the utility of

Table 3.2 Example Physiome-style projects based around the world from different major organ systems.

X-ome	URL
IUPS Physiome Project	http://www.physiome.org.nz/
NSR Physiome Project	http://www.physiome.org/
Virtual Physiological Human	http://www.vph-noe.eu/
Heart Physiome	http://www.physiome.ox.ac.uk/ http://www.physiome.org.nz/heart/
Digestive Physiome	http://www.physiome.org.nz/projects/digestive/ http://www.abi.auckland.ac.nz/uoa/gi-current-projects/
Lung Physiome	http://www.physiome.org.nz/projects/respiratory/
EuroPhysiome	http://www.europhysiome.org/
Kidney Physiome	http://physiome.ibisc.fr/qkdb/

the models and facilitate the creation of work flows that can generate model results from parameter sets and input data, in order to compare model predictions with experimental data in an automated fashion (Popel and Hunter, 2009).

The Current Status of Physiome Modeling

The Physiome Project can be considered to be a collation of various sub-"physiomes," e.g., of organ systems (the cardiome, the digestive physiome) or of molecular, cell, or tissue systems (e.g., the "metabolome" and the "epitheliome") (Popel and Hunter, 2009).

By far the most mature and advanced branch of the Physiome Project is the Cardiome. The Cardiome is an international effort to build a biophysically based multiscale mathematical model of the heart, and has advanced on several biophysical fronts that together represent a sophisticated understanding of the cardiac structure and function (Bassingthwaighte et al., 2009). Advanced Cardiome examples include models describing the cardiac electrical conduction (Hunter et al., 2003), mechanics of ventricular contractions (Nash and Hunter, 2000), the fluid dynamics of ventricular blood flow (Nordsletten et al., 2007), and the perfusion of myocardial tissues (Lee et al., 2009). These models span from ion channel mechanisms to whole-organ function and are now being applied to clinical applications such as drug discovery, medical device development, and the diagnosis of coronary artery disease (Popel and Hunter, 2009).

Many of the other "–omes" related to major biological systems are rapidly developing as shown in Table 3.2. For example, in the Lung Physiome, mathematical models have been applied to calculate the airway transport mechanisms for airflow and gas exchange. Soft-tissue mechanics techniques have also been applied to investigate the relationship between diminished lung capacity in diseases such as emphysema and the functional residual capacity. In the musculoskeletal system, mathematical models of the connective tissues such as muscles, tendons, ligaments and cartilage, and bones have been developed. In another applied field, sophisticated imaging and mapping techniques are being applied to predict the spread of cancers such as melanoma via lymph nodes (Popel and Hunter, 2009).

Another rapidly developing Physiome area, with significant potential for applications to livestock science, is the Digestive Physiome. The work developing from this branch of science is likely to eventually find application in multiple important livestock research areas, including optimizing animal nutrition and growth, minimizing methane production, optimizing milk production, and maintaining digestive health. The current status of this Physiome is discussed in some depth in the next section, to provide a more complete example of Physiome principles and frameworks in action.

The Digestive Physiome

The Digestive Physiome is a category of the Physiome Project, which deals with the physiological processes involved in the digestion of food particle and absorption of nutrients in the GI tract. The GI tract is a continuous tube comprising several distinct

organs that runs over several meters in length from the mouth to the anus. While these GI tract organs are functionally discrete, their behavior is also tightly coordinated and coregulated, affording a sophisticated overall level of integrated function (Du et al., 2010).

Significant recent attention in the Digestive Physiome has focused on the electrophysiological basis for motility. Like the heart, the GI tract has intrinsic pacemakers and propagating electrical rhythms that underlie contractile function. These electrical rhythms are known as slow waves, and it is now known that they are generated and propagated by a specialized network of cells, termed the interstitial cells of Cajal (ICCs), which lie in the wall of the GI tract. The slow waves spread from the ICCs to the surrounding smooth muscle cells, causing depolarization and gut contractions, when other coregulatory conditions are met, such as occurs after a meal.

Multiple Scales in the Digestive Physiome

Recent work on modeling slow wave function provides an excellent example of a multiscale analysis at work, whereby mathematical modeling techniques have been productively applied to bridge the gaps between vast biological scales (Cheng et al., 2010).

An extensive literature of cellular and subcellular studies has now grown to explain the detailed molecular basis of the slow wave in the ICCs, and this remains a focus of substantial ongoing investigation (Sanders et al., 2006). While these subcellular functions occur over the scale of milliseconds and nanometers, the whole-cell electrical events occur over the course of seconds. Meanwhile, at the tissue level, GI electrical propagation is actively mediated over many centimeters of the gut via a process termed entrainment, in which ICCs with different intrinsic frequencies operate at a single common frequency in the intact tissue (Sanders et al., 2006). It takes over a minute for this entrained behavior to propagate the length of a large animal stomach (Lammers et al., 2009). At the organ level, significant regional differences in slow wave propagation characteristics (velocity and amplitude) have been observed, and the geometry of the gut organs also fundamentally dictates the properties of slow wave propagation. At the whole-body level, interest in the GI field has focused on noninvasively evaluating the dispersion of the resultant electrical potential across the torso in normal and disease states (akin to ECGs), in order to inform the diagnosis such disorders, (Cheng et al., 2010). Therefore, it can be readily appreciated that a multiscale approach is required to encapsulate the range of activities of this system.

Mathematical Modeling of GI Electrophysiology

Mathematical modeling has now been applied to multiple levels of GI electrical activity to understand, interpret, and extend knowledge of slow wave function in health and disease (Cheng et al., 2010). As mentioned above, these validated mathematical models essentially provide an alternative virtual medium in which the hypothesis of normal and abnormal physiology can be exhaustively investigated, and the effects for treatment strategies predicted, without explicitly relying solely on animal and human experimental models.

Mathematical models of intestinal electrical activity have been formulated as early as the 1960s, when Nelsen and Becker (1968) suggested that a chain of relaxation oscillators could simulate the electromechanical activity in the small intestine. During the early 1970s, Sarna et al. (1971) developed this idea further by using an array of bidirectionally coupled oscillators to simulate different aspects of the GI-tract activity. For the next two decades, the pioneering work by these authors had a significant influence on the terminology and experimental methods used in investigations of GI motility.

As in all areas of Physiome science, the ongoing development of mathematical models of gut electrical activity have been steadily growing in complexity as more experimental evidence regarding the electrophysiological roles of the ICCs and gastric smooth muscle cells is progressively uncovered. In the GI field, it was not until 2004 that the gastric electrical events were finally simulated under the multiscale framework, almost a decade after the inception of the Physiome Project (Pullan et al., 2004). This early work strongly motivated the creation of more sophisticated cell models containing biophysical subcellular components based on real empirical data, rather than being based on purely theoretical considerations such as chains of relaxation oscillators (Corrias and Buist, 2007, 2008; Faville et al., 2008, 2009).

Even though the concept of a multiscale framework to gut modeling has been adopted relatively recently compared to the cardiac field, significant process is now being made, and integrated understanding of the gut electrophysiology is now appearing across multiple biophysical scales (Cheng et al., 2010). Multiscale GI mathematical models have now been developed that span from the cell to tissue to organ and finally through to the body surface levels. As illustrated in Figure 3.4, these multiscale GI models have been able to show how cellular events coupled with tissue, organ, and torso geometries, ultimately give rise to signals that can be recorded from the body

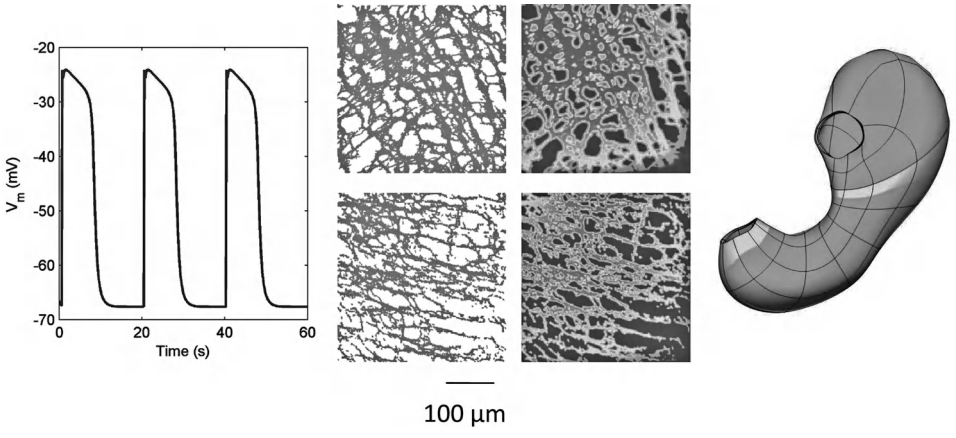


Figure 3.4 Multiscale gastric electrical models. Left, simulated autonomous electrical activity of the interstitial cells of Cajal (ICC) using the Corrias and Buist (2007) ICC model; Middle, simulated propagation of ICC electrical events in (top) normal ICC network and (bottom), diminished ICC network induced by altering the genetics of $H5T_{2B}$ receptors in mouse subjects (Du et al., 2010); Right, simulated gastric electrical propagation in an anatomical model of the human stomach (Cheng et al., 2007).

surface for diagnostic purposes (Buist et al., 2006; Cheng et al., 2010), including in states of gastric dysrhythmias (Cheng et al., 2007).

Anatomical Structure and Physiological Function in the Digestive Physiome

Recent work in this branch of the Digestive Physiome also illustrates the potential of detailed anatomical modeling to quantifiably link structure and function. Degradation in the ICCs network has been associated with several disorders of GI motility and transit, most notably diabetic gastroparesis and slow transit constipation (Farrugia, 2008). However, tissue preparations and quantifications of loss of ICCs are typically carried out after the tissue is processed and fixed by chemicals. As a result, the effects of structural degradation on electrophysiology and mechanical contraction are generally qualitative in nature, because the cellular and tissue electrophysiological and structural properties in these fixed states cannot be quantitatively inferred in the context to their *in vivo* states.

In a recent study, mathematical modeling was used to quantitatively relate the anatomical structure of ICCs to electrical propagation in health (mouse intestines at normal state), and disease (a model in which the ICC population was physically depleted, by knockout of the 5-HT_{2B} serotonin receptors) (Du et al., 2010; Tharayil et al., 2010). To achieve this, a modified version of a biophysically based ICC model (Corrias and Buist, 2008) was used to simulate the electrical activity of ICCs, and was embedded into a virtual network created from images of ICC networks taken from the healthy and diseased mouse intestines. As this cell model has previously been extensively validated to simulate realistic cellular electrical activity, the model was capable of rendering realistic propagation of intestinal electrical events through the virtual ICC networks. To model these emergent tissue level properties, investigators recently incorporated new molecular elements (i.e., introducing a voltage-dependent IP3-related mechanism) into an ICC cell model to enable virtual entrainment, then simplified the model by removing many components unessential for basic entrainment activity, such that larger scale simulations could be handled with computational efficiency (Du et al., 2010). The results showed that there was a substantially reduced current density generated by the abnormal ICC network, helping to explain why a motility difference might occur clinically in diseases in which ICCs are lost (Figure 3.4 middle) (Du et al., 2010).

While the example of detailed structural modeling of ICC networks focuses on a highly specialized area of digestive science, some valuable principles are demonstrated by this work that are generally applicable to Physiome research as a whole. A multiscale model was effectively developed and applied by "linking" the activities of different cell models to tissue and anatomical structures—to generate a predictive physiological outcome that could not easily be quantified through traditional experimental methods. Secondly, this example illustrates the general pattern through which the Physiome progresses: interdisciplinary teams of modelers and experimentalists working together on a small problem of physiological importance, giving rise to novel methods and models/modules that can be archived such that they are available for groups in future.

Mechanical Models in the Digestive Physiome

To date, mechanical modeling in the GI tract has mostly been performed independent of electrical activity. One reason for the lack of coupling between electrical and

mechanical models can be attributed to the limited availability of GI experimental data regarding the underlying molecular events. By contrast, in the cardiac field, electromechanical coupling models are more advanced (e.g., Nickerson et al., 2006), and are being applied to address a number of important clinical questions, including quantifying the effects of cardiac arrhythmias on cardiac mechanical functions.

Some of the mechanical models of the GI tract that are now being developed have focused on modeling the flow of luminal contents within the stomach and intestines. The main aim of these models has been to establish how the physical contractions of the gut wall contribute to the mixing and transportation of luminal contents. This work is highly relevant to livestock science, for example, in order to optimize the constituency and bioavailability of feed, and to minimize the production of waste products such as methane. Drug companies are also becoming interested in such models, in order to predict and optimize the breakdown and bioavailability of pharmaceutical agents that are delivered via the gut.

One of the earlier examples of this field of modeling represented the GI tract as a hydraulic system, in which the profile of luminal flow is studied using fluid dynamics over a cylindrical control volume with major assumptions such as incompressibility, Newtonian fluid, axial symmetry (Melville et al., 1975). The assumptions made via this approach were relatively crude, in the sense that the geometry of the intestines is complex, so the luminal flow is most likely not axial symmetric. Furthermore, luminal contents are a heterogenous mixture of solid and liquid particles, therefore, not a Newtonian fluid (Stavitsky et al., 1981). However, as an initial model these assumptions are useful. The consequences of these assumptions is such that the luminal flow velocity can be expressed as a function of the contraction velocity of the gut wall, the cavity radius, and the maximum radius of the luminal wall (Melville et al., 1975)

As more advanced imaging modalities such as the MRI have become available, the *in vivo* movements of the stomach have been captured in real time and used as boundary conditions to the mechanical models (Pal et al., 2004). From simultaneous recordings using MRI and gastric manometry, the changes in shape and pressure of the stomach during gastric emptying have been characterized. Subsequently, computational simulations of stomach emptying have been performed (Pal et al., 2004). This model used the lattice-Boltzmann method to quantify the distribution of food particles in the stomach by calculating the spread of particles by the root-mean-square radius of the particles from their collective center of mass. This model demonstrated interesting theories regarding gastric fluid motion during contractions, such as describing a jet-like retrograde stream that generates large stresses on the gastric contents, and a circulatory motion that facilitates the mixing of gastric contents (Pal et al., 2004, 2007).

Similar studies have also been carried out by reconstructing the three-dimensional (3D) anatomy of the stomach from ultrasound scans of a rat stomach (Liao et al., 2005). In this case, a set of algorithms was used to calculate the geometric characteristics of the 3D model, such as the volume, surface area, and curvatures. The relationships between the curvatures of the stomach in the longitudinal and circular directions with respect to gastric distension pressures were investigated. The model demonstrated that the relative radii of the curvatures were dependent on the magnitude of gastric distension pressure. It is known that abnormal gastric accommodation found in some conditions, such as functional dyspepsia and after vagotomy (dissection of the vagus

nerve), is expected to alter the regional gastric curvatures. Hence, the algorithms used in the rat stomach model offer a methodological way of characterizing the geometry of the stomach in order to identify abnormal gastric accommodations.

In summary, progress in the Digestive Physiome is beginning to link multiple scales and quantifiably link structural and functional relationships. In future, work in fields that are currently distinct, such as GI electrophysiology and mechanics will be integrated into more sophisticated models. Importantly, much of this work is applicable and transferrable to mammalian physiology as a whole, or is readily able to be translated to livestock science.

Present Challenges in Physiome Project

The use of modeling the biological fields has lagged behind the fields of physics and chemistry, where many fundamental modeling advances were made in the eighteenth and nineteenth centuries. This lag is partly due to the high levels of complexity within biological systems, which is due to the large number of variables defining these systems, the huge number of interactions between these variables, the many inbuilt redundancies, and the fact that these events occur over vast spatial and temporal scales. The use of mathematical models as descriptions of these complicated biological systems is inevitably incomplete or inaccurate to some degree. The process is, by necessity, also one of iterative improvement and refinement.

Although mathematical modeling is increasingly able to meet these complex requirements, and continues to grow in importance in today's multidisciplinary research environment, many other significant challenges remain, principally related to the creation of models, computational expenditure, experimental methodologies, and ethics. The creation of reliable models is largely dependent on the existing knowledge of molecular biology, signal-pathways, and imaging techniques—and in many instances the components and interactions of these complex pathways are far from being exhaustively quantified. Where the functions and behaviors of a component or a number of key components in a system are unknown, assumptions made during the modeling process could sometimes lead to an erroneous analysis of the system.

Furthermore, in practice, experimental evidence of a system generally comes from studies of many different species due to model suitability and cost and ethical constraints. Here, the models face the problem of presenting an idealized mammalian description of an integrated system based on data from different species, while being accurately representative of no single species. Indeed, physiological functions within the same biological pathways will often differ markedly between different species. Therefore, accuracy of inference using a model is dependent on how well the parameters in that model are tailored to a particular species. Transference of model and information should be considered in the context of careful parameterization and application of a model in a particular species.

Computational limitation is another area of major practical concern. In general, computational power is progressively increasing at a rapid rate, and large and more complex problems are continually being tackled. However, in many cases, the extent of analysis for a multiscale problem is limited by the amount of computer power at one's disposal. This is particularly important in instances where a very high spatial

resolution of computational points is required to gain a numerically converged solution. If a multiscale model was to be "truly multiscale," then it would ideally contain both extremes of spatiotemporal resolutions, and consequently the model must be solved at a time step lower than the finest temporal scale. This would incur a huge computation expense. In practice, this is generally avoided by dividing the larger multiscale models into smaller models that include similar spatiotemporal scales, to be solved separately, and then using appropriate coupling conditions to ensure continuity. This is a technique readily being employed when coupling cardiac fluid flow and mechanics simulations as well as simulation of body surface electrical potentials from far field sources (Pullan et al., 2005). However, with the continual development of numerical algorithms and solvers and computer hardware, the solution times of large-scale models will be reduced in future, and the applications of multiscale models will continue to grow.

Applications to Livestock Science

The complexity of modern livestock science presents many opportunities and challenges for applying Physiome-type strategies. As Woelders et al. (2011) have described the growing appreciation of the high complexity of biological organization in livestock analyses has yielded a huge reservoir of empirical data now available for integration into models. As a result, livestock Physiomes will undoubtedly grow to inform areas of current livestock research focus, across such research domains as increased productivity, product quality, disease resistance, fertility, behaviors, animal welfare, and reduced ecological footprints.

Of central importance is the power of Physiome models to define and apply quantifiable relationships between the genotype and phenotype, in order to gain a finer control over optimizing and expressing the desired traits in livestock—whether researching improved salmonella resistance in chickens or manipulations of the bovine oestrous cycle.

One simple applied example for livestock science that has grown out of the Physiome approach is in the identification and optimization of lamb cuts via video-imaging analyses. In this example, a laser-scanning system captures a 3D image of as a meat carcass is cut along the conveyor belt (Hilton et al., 1992). The snapshots of the cuts are compared to a computer database of different cuts. The type of cut is automatically classified using mathematical equations at a rate of 1 second per cut, leading to an optimized result specific to each particular animal. This modeling technology holds significant potential to decrease labor costs and increase food safety, product quality, and processing efficiency (Bowman et al., 1993). One of the aims of this particular project is to develop means of "virtual dissection" using computational modeling techniques on "virtual carcasses" (discussed above on page 63), in order to maximize yield through objective-orientated optimizing techniques. Soft-tissue mechanics techniques could then be applied to the virtual carcass to simulate virtual deformation or dissection. The sophisticated anatomical lamb and pig models previously shown in Figure 3.3 already provide the foundations on which this Physiome-output research can readily be based.

In general, the most developed analysis of Systems Biology in livestock science is genome network pathway informatics. Recent studies have related microarray data

to larger biochemical pathways from the KEGG database (te Pas et al., 2007, 2008). The key information of Systems Biology approach targets the different physiological pathways that may be involved in the regulation and proliferation in different physiological features, thereby relating genome functions to developmental process (te Pas et al., 2007). Another important application of the Physiome Project to livestock science is the categorization and modeling of critical molecular pathway information (te Pas et al., 2008). Pathway information provides insight into the biological processes underlying microarray data.

Much less detailed pathway information is currently available in the livestock sciences, than in human biology, but this volume is now rapidly expanding. Research also suggests that the genetics of many livestock species will be similar to those humans' genes, which have already been sequenced and stored in databases such as the KEGG (Kanehisa et al., 2006). However, one issue that needs to be addressed is that many software packages use species-specific gene IDs that cannot yet handle genomics data from other species (te Pas et al., 2008). Therefore, bioinformatics were combined with the programming language PERL to create a software tool (PathwayKegg in Table 3.1) that uses species-independent gene IDs to streamline the process of searching for pathways information in online databases, using microarrays data followed by combining pathway information with microarray data. An example analysis using the software provided insight into differential line-specific biological processes that may explain the variations in the host chicken's response to *Salmonella* (te Pas et al., 2008).

The selection and progressive improvement of genetic makeup in order to accentuate certain desirable phenotypes has obviously long been practiced, such as is exemplified in the domestication and resultant varieties of the common cat and dog (Kadarmideen et al., 2006). The discovery of DNA and advances in both experimental and modeling tools now presents a highly exciting time in livestock science, allowing mankind the opportunity of selectively manipulating genetics in our favor, to greater degrees and with substantially more precision than ever before. The latest incorporation of "-omics" technologies and the Physiome Project could "fundamentally change the practice of animal breeding, moving away from a basically "black box" approach toward an approach that considers the regulatory network and pathways underlying the expression of important phenotypes." (Kadarmideen et al., 2006).

The key issue in moving toward this exciting future prospect is no longer so much the lack of basic information, but the ability to understand the interactions and polymorphisms of these fundamental modules in higher biological scales. There is hope that the various "-omics" technologies and broader multiscale principle of the Physiome Project will ultimately initiate a new era of information utilization and ultimately bring many innovative applications of these new insights to livestock science research.

Conclusions and Future Directions

The Physiome Project is an ambitious, collaborative project with a long horizon. Work involving more complex levels of structural and functional multiscale integration is essential, if the Physiome Project is to significantly improve the understanding of whole-organ physiology. In addition to this vertical integration across multiple scales,

it is important to include horizontal integration across boundaries such as between organ systems and organisms (Kohl and Noble, 2009).

The paradigm whereby theoretical work is pursued in close and continuous iteration with experimental and/or clinical investigations often produces the best results. Therefore, an essential aim for the Physiome Project is the quantitative integration of *in silico*, *in vitro,* and *in vivo* research from multiple species. In addition, as new types of data, methods, and technologies emerge and computation power continues to rapidly advance, the Physiome Project must be capable of rapidly and efficiently adapting to these changes.

Although one of its principal focuses is on improving the understanding of human health, many of the techniques and new knowledge growing out of the Physiome Project community is directly applicable to other organisms and to the livestock industry. As stated by Denis Noble, "Biology is set to become a highly quantitative science. In the present century, it will also become the most computer-intensive science" (Noble, 2002). The Physiome Project is underway, the goals are set, and the frameworks are in place, but in reality it is still very early days for this visionary enterprise. Ultimately, livestock science stands to gain a great deal from the advances that the Project will provide, and livestock scientists who embrace this field while it is still developing will have much to add and gain.

Acknowledgments

The authors would like to acknowledge the contributions and support of Profs. Peter Hunter and Andrew Pullan. The authors are grateful for financial support via the NZ Health Research Council, the NIH (R01 DK64775), and a University of Auckland Doctoral Scholarship.

References

Bassingthwaighte, J.B. (1995) Toward modeling the human physionome. *Advances in Medicine and Biology* **382**, 331–339.

Bassingthwaighte, J.B. (2000a) Strategies for the Physiome Project. *Annals of Biomedical Engineering* **28**, 1043–1058.

Bassingthwaighte, J.B. (2000b) Back to fundamentals: anatomy-based physiological bioengineering. *Annals of Biomedical Engineering* **28**, 701–703.

Bassingthwaighte, J., Hunter, P., & Noble D. (2009) The Cardiac Physiome: perspectives for the future. *Experimental Physiology* **94**(5), 597–605. PMID: 19098089

Bodley, J.M. (2000) Modelling meat carcasses. Technical Report, Department of Engineering Science, The University of Auckland, 1–106.

Bowman, C., Beach, D., Hilton, P.J., et al. (1993) Machine Vision Sensing for Meat Processing Automation. Meat '93, The Australian Meat Industry Research Conference.

Buist, M.L., Cheng, L.K., Sanders, K.M., et al. (2006) Multi-scale modelling of human gastric electric activity: can the electrogastrogram detect functional electrical uncoupling? *Experimental Physiology* **91**, 383–390.

Cheng, L.K., Komuro, R., Austin, T.M., et al. (2007) Anatomically realistic multi-scale models of normal and abnormal gastrointestinal electrical activity. *World Journal Gastroenterology* **13**, 1378–1383.

Cheng, L.K., O'Grady, G., Du, P., et al. (2010) Gastrointestinal system. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **2**(1), 65–79. doi:10.002/wsbm.019.

Clancy, C.E. & Rudy, Y. (1999) Linking a genetic defect to its cellular phenotype in a cardiac arrhythmia. *Nature* **400**, 566–569.

Corrias, A. & Buist, M.L. (2007) A quantitative model of gastric smooth muscle cellular activation. *Annals of Biomedical Engineering* **35**, 1595–1607.

Corrias, A. & Buist, M.L. (2008) Quantitative cellular description of gastric slow wave activity. *American Journal Physiology Gastrointestinal Liver Physiology* **294**(4), G989–G995.

Crocombe, J., Clarke, R., & Pullan, A.J. (1999) *Modelling a Lamb Hind Leg, The Simulation Technology and Training (SimTecT 99) Conference*, Melbourne, Australia, March, 29—April.

Dalrymple, B.P., Kirkness, E.F., Nefedov, M., et al. (2007). Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome Biology* **8**(7), R152. doi:10.1186/gb-2007–8-7-r152. PMID 17663790.

de Gortari, M.J., Freking, B.A., Cuthbertson, R.P., et al. (1998) A second-generation linkage map of the sheep genome. *Mammalian Genome* **9**(3), 204–209.

Du, P., Li, S., O'Grady, G., et al. (2009b) Effects of electrical stimulation on isolated rodent gastric smooth muscle cells evaluated via a joint computational simulation and experimental approach. *American Journal Gastrointestinal Liver Physiology* **297**, 672–680.

Du, P., O'Grady, G., Davidson, J.B., et al. (2010) Multi-scale modeling of the gastrointestinal electrophysiology and experimental validation. *Critical Reviews in Biomedical Engineering* **38**(3), 225–254.

Du, P., O'Grady, G., Windsor, J.A., et al. (2009a) A tissue framework for simulating the effects of gastric electrical stimulation and in-vivo validation. *IEEE Transactions on Biomedical Engineering* **56**, 2755–2761.

Elsik, C.G., Tellam, R.L., Worley, K.C., et al. (Bovine Genome Sequencing and Analysis Consortium) (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**(5926), 522–528

Farrugia, G. (2008). Interstitial cells of Cajal in health and disease. *Neurogastroenterology and Motility* **20**(Suppl 1), 54–63.

Faville, R.A., Pullan, A.J., Sanders, K.M., et al. (2008) A biophysically based mathematical model of unitary potential activity in interstitial cells of Cajal. *Biophysical Journal* **95**(1), 88–104. PMID: 18339738.

Faville, R.A., Pullan, A.J., Sanders, K.M., et al. (2009). Biophysically based mathematical modeling of interstitial cells of Cajal slow wave activity generated from a discrete unitary potential basis. *Biophysical Journal* **96**(12), 4834–4852. PMID: 19527643.

Fernandez-Armesto, F. (1997) *Truth A History and a Guide for the Perplexed*. Transworld Publishers Ltd, London.

Hilton, P.J., Gabric, R.P., & Waltenberg, P.T. (1992). 3-D imaging using laser point triangulation. *Proceedings Seventh New Zealand Image Processing Workshop*, pp. 123–128, Christchurch.

Hodgkin, A.L. & Huxley, A.F. (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal Physiology* **117**, 500–544.

Hunter, P.J. & Borg T.K. (2003) Integration from proteins to organs: the Physiome Project. *Nature Reviews Molecular Cell Biology* **4**, 237–243.

Hunter, P.J., Pullan, A.J., & Smaill B.H. (2003) Modeling total heart function. *Annual Reviews in Biomedical Engineering* **5**, 147–177.

Hunter, P.J., Robbins, P., & Noble D. (2002) The IUPS human Physiome Project. *Plügers Archive European Journal of Physiology* **445**, 1–9.

Kadarmideen, H.N., von Rohr, P., & Janss L.L.G. (2006) From genetical genomics to systems genetics: potential applications in quantitative genomics and animal breeding. *Mammalian Genome* **17**(6), 548–564.

Kanehisa, M., Goto, S., Hattori, M., et al. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research* **34**(database issue), D354–D357.

Kohl, P.& Noble, D. (2009) Systems biology and the virtual physiological human. *Molecular Systems Biology* **5**, 292. PMID: 19638973.

Lammers, W.J., VerDonck, L., Stephen, B., et al. (2009) Origin and propagation of the slow wave in the canine stomach: the outlines of a gastric conduction system. *American Journal of Physiology Gastrointestinal Liver Physiology* **296**, 1200–1210.

Lander, E.S., Linton, L.M., Birren, B., et al. (2001) (The International Human Genome Mapping Consortium). Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.

Lee, J., Niederer, S., Nordsletten, D., et al. (2009) Coupling contraction, excitation, ventricular and coronary blood flow across scale and physics in the heart. *Philosophical Transactions of The Royal Society A: Mathematical Physical & Engineering Sciences* **367**, 2311–2331.

Liao, D., Zhao, J., & Gregersen, H. (2005) Regional surface geometry of the rat stomach based on three-dimensional curvature analysis. *Physics in Medicine and Biology* **50**, 231–246.

Luo, C.H. & Rudy, Y. (1994) A dynamic model of the cardiac ventricular action potential. I. Simulations of ionic currents and concentration changes. *Circulation Research* **74**, 1071–1096.

Melville, J., Macagno, E., & Christensen, J. (1975) Longitudinal contractions in the duodenum: their fluid-mechanical function. *American Journal of Physiology* **228**, 1887–1892.

MIASE. Minimal Simulation Information about an Experiment Project. SED-ML the Simulation Experiment Description Markup Language. Available on: http://www.ebi.ac.uk/compneur-srv/sed-ml/. Accessed October, 2009.

Moss, A.J. & Kass, R.S. (2005) Long QT syndrome: from channels to cardiac arrhythmias. *Journal of Clinical Investigation* **115**(8), 2018–2024.

Nash, M.P. & Hunter, P.J. (2000) Computational mechanics of the heart: from tissue structure to ventricular function, *Journal of Elasticity* **61**, 113–141.

Nature (2001) The human genome special issue. *Nature* **409**, 745–964.

Nelsen, T.S. & Becker, J.C. (*1968*) Simulation of the electrical and mechanical gradient of the small intestine. *American Journal of Physiology* **214**, 749–757.

Nickerson, D.P., Nash, M.P., Nielsen, P.F., et al. (2006) Computational multi-scale modeling in the IUPS Physiome Project: modeling cardiac electromechanics. *IBM Journal of Research and Development* **50**(6), 617–630.

Noble, D. (2002) The rise of computational biology. *Nature Reviews Molecular Cell Biology* **3**, 460–463.

Nordsletten, D.A., Hunter, P.J., & Smith N.P. (2007) Conservative arbitrary lagrangian-eulerian forms for boundary driven and ventricular flows. *International Journal of Numerical Methods in Fluids* **56**, 1457–1463.

Pal, A., Brasseur, J.G., & Abrahamsson B. (2007) A stomach road or "Magenstrasse" for gastric emptying. *Journal of Biomechanics* **40**, 1202–1210.

Pal, A., Indireshkumar, K., Schwizer, W., et al. (2004) Gastric flow and mixing studied using computer simulation. *Proceedings of the Royal Society B: Biological Sciences* **271**, 2587–2594.

Popel, A.S. & Hunter, P.J. (2009) Systems biology and Physiome Projects. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **1**, 153–158.

Pullan, A.J., Buist, M.L., & Cheng L.K. (2005) *Mathematically Modelling the Electrical Activity of the Heart: From Cell to Body Surface and Back Again*. World Scientific Publishing Company, Singapore.

Pullan, A.J., Cheng, L.K., Yassi, R., et al. (2004) Modelling gastrointestinal bioelectrical activity. *Progress in Biophysics and Molecular Biology*. **85**, 523–550.

Sanders, K.M., Koh, S.D., & Ward, S.M. (2006) Interstitial cells of Cajal as pacemakers in the gastrointestinal tract. *Annual Reviews in Physiology* **68**, 307–343.

Sands, G.B., Gerneke, D.A., Hooks, D.A., et al. (2005) Automated imaging of extended tissue volumes using confocal microscopy. *Microscopy Research and Technique* **67**, 227–239.

Sandu, G.S., Solorio, L., & Broome, A. (2010) Whole animal imaging. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **2**, 398–421.

Sarna, S.K., Daniel, E.E., & Kingma Y.J. (1971) Simulation of slow-wave electrical activity of small intestine. *American Journal of Physiology* **221**, 166–175.

Science (2001) Human genome special issue. *Science* **291**, 1145–1434.

Spitzer, V.M. & Whitlock, D.G. (1998) The Visible Human Dataset: the anatomical platform for human simulation. *Anatomical Record* **253**, 49–57.

Stavitsky, D., Macagno, E.O., & Christensen, J. (1981) Finite-element analysis of flow induced by contractions like those of the intestine. *Journal of Biomechanics* **14**, 183–193.

Tawhai, M.H., Nash, M.P., & Hoffman, E.A. (2006) An imaging-based computational approach to model ventilation distribution and soft-tissue deformation in the ovine lung. *Academic Radiology* **13**, 113–120.

te Pas, M.F.W., Hulsegge, I., Coster, A., et al. (2007) Biochemical pathways analysis of microarray results: regulation of myogenesis in pigs. *BMC Developmental Biology* **7**, 1–15.

te Pas, M.F.W., van Hemert, S., Hulsegge, B., et al. (2008) A pathway analysis tool for analyzing microarray data of species with low physiological information. *Advances in Bioinformatics* 719468. PMID: 19920988.

Tharayil, V.S., Wouters, M.M., Stanich, J.E., et al. (2010) Lack of serotonin 5-HT(2B) receptor alters proliferation and network volume of interstitial cells of Cajal in vivo. *Neurogastroenterology and Motility* **22**, 462–469; e109–e110.

Trew, M.L., Caldwell, B.J., Sands, G.B., et al. (2006) Cardiac electrophysiology and tissue structure: bridging the scale gap with a joint measurement and modelling paradigm. *Experimental Physiology* **91**, 355–370.

Venter, C., Adams, M.D., Myers, E.W., et al. (2001) The sequence of the human genome. *Science* **291**, 1304–1351.

Wimalaratne, S.M., Halstead, M.D.B., Lloyd, C.M., et al. (2009) Facilitating modularity and reuse: guidelines for structuring CellML1.1 models by isolating common biophysical concepts. *Experimental Physiology* **94**, 472–485. doi:10.1113/expphysiol.2008.045161.

Woelders, H., Te Pas, M.F.W., Bannink, A., et al. (2011) Systems biology in animal sciences. *Animal*, doi:10.1017/S1751731111000036.

Yassi, R., Cheng, L.K., & Al-Ali, S. (2010) Three-dimensional high-resolution reconstruction of the human gastro-oesophageal junction. *Clinical Anatomy* **23**(3), 287–296. PMID: 20169612.

Chapter 4
Systems Biology in Livestock Health and Disease

Gordon M. Kirby

Introduction

This chapter focuses on the impact of Systems Biology in livestock health and disease. The application of Systems Biology concepts and technologies in personalized human medicine is discussed and contrasted to the present use of Systems Biology approaches in the management of livestock diseases at the population level. Examples of current or potential applications of Systems Biology in the prevention of disease and in the maintenance of health and productivity of food-producing ruminants, swine, and poultry are presented. Discussions focus on the impact of Systems Biology in the identification of pathogens, diagnosis of disease, understanding the genetic determinants of disease susceptibility or resistance, host–pathogen interactions, and the mechanisms of development and spread of important diseases in livestock production systems.

Defining Systems Biology in the Medical Context

Systems Biology constitutes the comprehensive and integrative analysis of the structural and functional properties of all the components of biological organisms including their identity, correlations, and dynamic interactivity (Peng et al., 2009). This includes analysis of the hierarchical complexities of biological information from DNA to RNA, proteins, metabolites, macromolecular complexes, signaling networks, cells, organs, organisms, and species and their responses to environmental stimuli. The reductionist approach advocated by molecular biology has made notable advances in defining the mechanisms of disease including specific host–agent interactions, identification of targets for therapy and prevention of disease by studying a limited number of individual components of biological systems (Feng et al., 2009). However, this approach is not conducive to generating a complete understanding the entire biological system. With the realization that organisms do not consist of isolated subsets of genes, proteins, and metabolites, Systems Biology constitutes a more rational approach to the study of the mechanisms that underlie health and disease. Systems Biology examines disease

Systems Biology and Livestock Science, First Edition. Edited by Marinus F.W. te Pas, Henri Woelders, and André Bannink.

Box 4.1 System Biology and Disease

"Systems Biology examines disease through qualitative and quantitative analyses of the complex relationships between multiple components of a biological system as a whole and the responses to genetic and environmental perturbations. This new paradigm is changing the way we understand, prevent, diagnose, and treat disease."

through qualitative and quantitative analysis of the complex relationships between multiple components of a system and the responses to genetic and environmental perturbations (see Box 4.1).

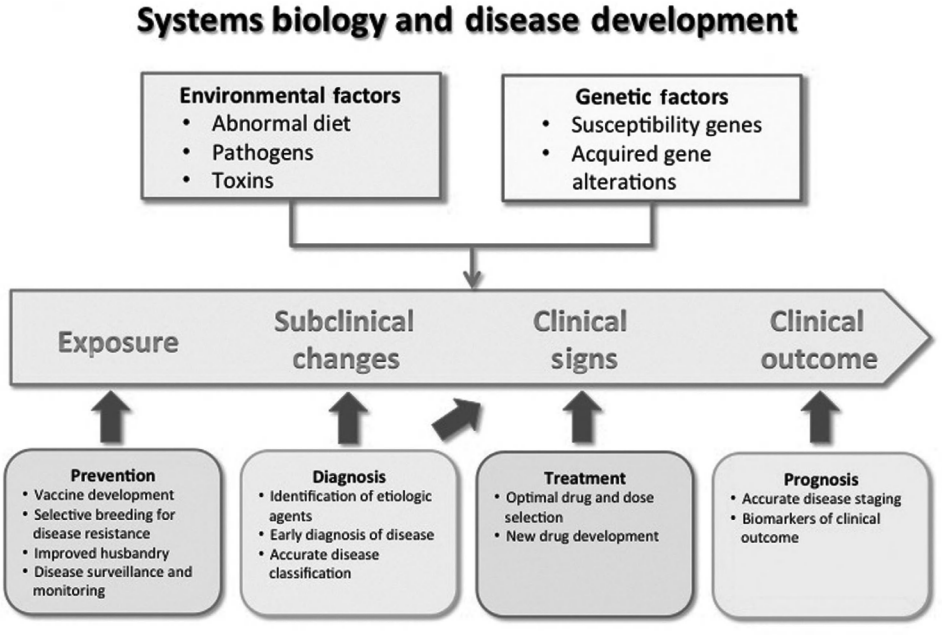
Measurement of the genomic, proteomic, and metabolic data that reflect the interactivity of the system's components is accomplished through the use of various high-throughput technologies (Peng et al., 2009). Because of the potential to generate vast amounts of data, bioinformatics and computational resources are key elements that enable the integration and interpretation of these data.

From many perspectives, Systems Biology represents a paradigm shift in biology and medicine by embracing a new culture that recognizes the importance of systematic interpretation of biology, health, and disease through the study of the dynamic and interdependent interactions of the complex network of genes and their protein products (Feng et al., 2009). Application of this new holistic philosophy to human and veterinary medicine has the potential to evolve healthcare to new standards that will substantially enhance our knowledge of the pathogenesis and improve the diagnosis, prevention, and treatment of disease (Olden, 2006).

Establishing the Need for Systems Biology Approaches in Human and Veterinary Medicine

Imbalances in the interconnected networks of proteins, cells, and tissues result in perturbations of the specialized structural and functional roles of components that are manifested as disease. There is a need to develop more predictive tools to assist in the early diagnosis of disease, to identify and assess new candidates for drug development, and to improve our understanding of the complex mechanisms and causes of disease development. Currently, many of the existing biomarkers lack the necessary sensitivity or specificity for the early detection of human and animal diseases. Systems Biology technologies have the potential capacity to identify and characterize biomarkers that reflect the exposure of the system to environmental insults (e.g., infectious agents, toxins, abnormal diets) and mirror the associated biological consequences that are manifested as disease.

Such biomarkers may represent indicators of disease states that could be validated as new diagnostic tests, but they may also represent new targets for novel therapies, opportunities for vaccine development, or reflect patient responses to therapeutic interventions to be used for prognostication. As a result, implementing Systems Biology concepts and approaches has the potential to substantially expand the range of options for patient treatment and to improve treatment outcomes (Elrick et al., 2006). Following this strategy represents a shift from the current situation of intensive

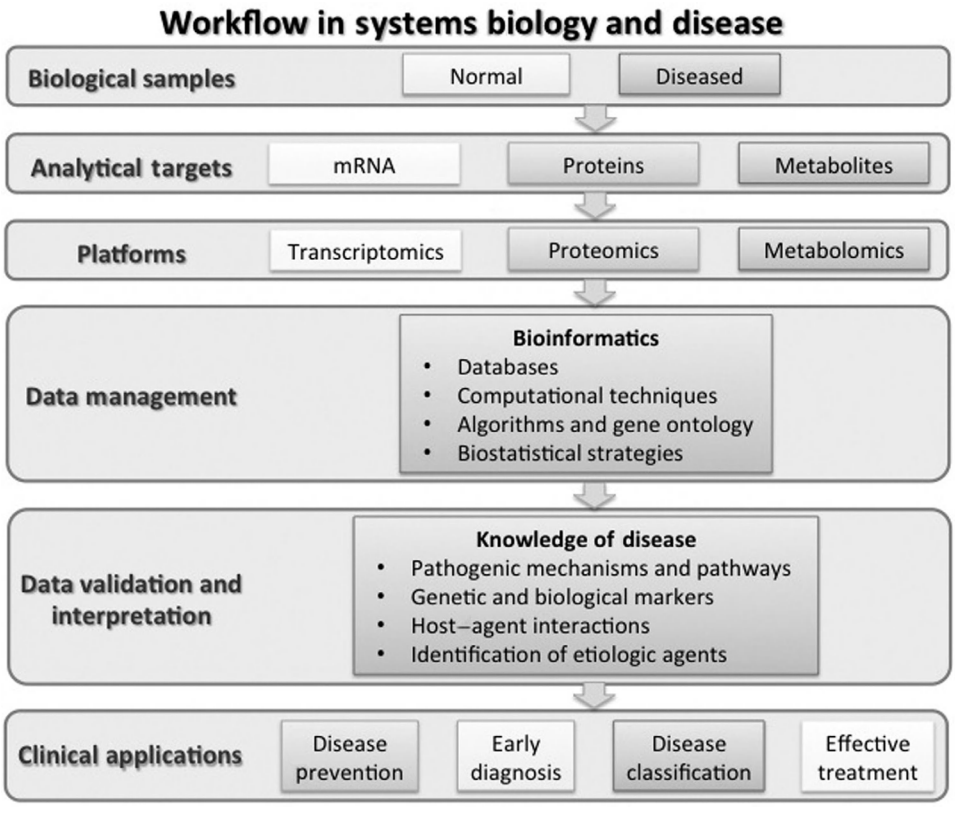


reaction at the later stages of clinical disease toward a more predictive, preemptive, and preventive approach to maintaining health, whereby disease susceptibility and predisposing factors are assessed and treatments are optimized for individuals before clinical disease becomes apparent (Olden, 2006; Downing, 2009).

In order to adopt Systems Biology concepts in medical research and clinical practice, there is a need to derive the capacity to process the massive amounts of information. This can be accomplished through the development and routine use of "omics" technologies (i.e., genomics, transcriptomics, proteomics, metabolomics) with high-throughput capabilities and bioinformatics tools that have the capacity to store, analyze, integrate, and interpret large amounts of complex data.

More recently, the development of new technologies such as microfluidics and nanotechnologies is providing the capacity to efficiently deliver, multiple and more precise biological measurements in real time at the point-of-care (i.e., patient-side).

The fallout of the human genome project has revealed that the complex phenotype of chronic multifactorial diseases is not a result of single genes and their associated proteins, but is more likely due to the interactions of multiple genes and environmental factors. Thus, the past reliance on studying a few genes or proteins in single pathways is inadequate for predicting and preventing the occurrence of multifactorial diseases with complex etiologies. Using a Systems Biology approach, the dynamic alterations in activity and amount of molecular components of cells and tissues can be monitored in animals whose genetic makeup has been characterized. The challenge of biomedical research is to identify the relevant genes and environmental factors associated with disease development, to elucidate the underlying mechanisms by which they interact to the cause disease, and to delineate the inherited capacity of individuals to respond to and resist environmental influences (Olden, 2006; Borden and Raghavan, 2010). By



monitoring the influences of environmental factors on gene expression at the mRNA and protein level and on the associated metabolite profiles in body fluids and tissues, biological indicators or biomarkers can be developed and validated that reflect the activity of all relevant genes and related products. In this way, events that have an adverse impact on health can be identified and predicted, and specific intervention strategies can be implemented, long before clinical disease is detected by the diagnostic approaches currently used in medical practice (Olden, 2006).

Systems Biology and Personalized Healthcare in Human Medicine

The concept of personalized healthcare has evolved from the need to develop novel and effective approaches to managing human health and disease of individuals. Systems Biology is beginning to have a profound impact on human medicine by enabling clinicians and biomedical scientists to comprehend and model the biological system in a more holistic and comprehensive manner. A more precise and thorough understanding of the molecular pathogenesis of disease enables the development of a more effective healthcare system through the design of optimal therapies that can be initiated earlier, or the design of strategies that can detect and prevent disease prior to onset. (Joshi and Kucherlapati, 2008). Thus, Systems Biology approaches have the

potential to develop personalized healthcare and to guide medical decision-making by integrating information regarding individual genetic profiles and environmental perturbations with patient-specific preventative, diagnostic, and therapeutic interventions with minimal side effects (Weston and Hood, 2004; Joshi and Kucherlapati, 2008; Borden and Raghavan, 2010). The introduction of analytical tools with increasing sophistication will allow medical practitioners to assess the risk of disease development in individual patients, to use "molecular signatures" for disease diagnosis and prognosis, and to make informed decisions regarding the management of their patients' health by selecting the appropriate drug and dose for therapy.

Therefore, personalized medicine is *preemptive* in that disease is identified and managed early before clinical signs are apparent (Downing, 2009); *predictive* in identifying the risk of disease development by factors such as pathogens or toxins through the presence of protein or metabolite markers in body fluids, and *preventive* by identifying genes that determine the susceptibility to the development of disease or the response to drug therapy (Weston and Hood, 2004).

Areas of Application of Systems Biology to Human Medicine

Biological systems can be explored to acquire the relevant medical information regarding system malfunction. This information includes the environmental or genetic factors that contribute to the etiology of the system malfunction, the specific underlying pathogenic mechanisms, the diagnostic evidence that specific components of the system are defective and the identity of the therapeutic targets that may result in a functional restoration of the system (Ginsburg and Willard, 2009). Some areas where Systems Biology and personalized medicine are expected to have potential impact are outlined in the following sections.

Understanding Disease

Disruption of the interactivity of the components of a complex system and the subsequent functional impairment is manifested clinically as disease. Systems Biology uses a holistic approach to elucidate the complex pathophysiological mechanisms of disease by examining gene expression and the connectivity of their protein products (see Box 4.2).

Box 4.2 System Biology in Clinical Medicine and Pathology

Assists in:

- understanding *mechanisms* of disease and host–pathogen interactions;
- identifying and characterizing *biological roles* for proteins with abnormal expression;
- identifying potential *targets* for clinical therapy;
- identifying predictive *biomarkers* for diagnosis, disease susceptibility, and prognosis.

Bioinformatic tools can help with the elucidation of the mechanisms of action by associating sets of differentially expressed genes with a particular pathway (Harrill and Rusyn, 2008). Harrill and Rusyn (2008) assembled a list of public databases that aid in the annotation and interpretation of gene expression data in terms of cellular process, functions, and pathways. These databases include Gene Ontology (GO; www.geneontology.org), Gene Map Annotator and Pathway Profiler (www.genmapp.org), the Science Signaling Connections Map (stke.sciencemag.org/cm/), BioCarta (www.biocarta.com/genes/index.asp), Reactome (www.genome knowledge.org), and KEGG (www.genome.jp/kegg/pathway.html). By associating changes in gene expression with a particular pathway, these databases can provide biological relevance to gene expression data by identifying perturbations in molecular mechanisms.

Screening for Disease Susceptibility

Disease susceptibility can be characterized by genomics assessments of inactivating mutations or gene deletions that remain stable through an individual's lifetime, or by assessing dynamic alterations in gene expression (e.g., transcriptional profiles, patterns of protein expression and metabolite levels) that change in response to environmental stimuli (Ginsburg and Willard, 2009). Individuals can then be screened for disease susceptibility using genomics or gene expression techniques. The relationship between genomics and disease phenotypes can be investigated by examining quantitative trait loci (QTL) that identify specific genotypes or polymorphic variations of a phenotype (e.g., single nucleotide polymorphisms or single nucleotide polymorphisms (SNP)) that characterize a disease in a population of individuals (Harrill and Rusyn, 2008). For example, the risk of developing cancer can be determined by assessing mutations in susceptibility genes such as *BRCA1* or *BRCA2* in hereditary forms of breast cancer, and in the mismatch repair genes *MLH1* and *MSH2* in hereditary nonpolyposis colorectal cancer (HNPCC) (Borden and Raghavan, 2010).

Optimizing Diagnosis and Prognosis

Genome expression profiling can be used to correlate gene expression patterns with disease classification in individual patients and to predict their response to therapy. For example, genomic signatures are currently being used to identify specific subtypes of cancer. This approach is useful in determining patient prognosis, in assisting with decisions of whether or not to implement therapy or in delineating specific therapeutic options. Indeed, expression of drug resistance genes in specific tumors can influence patient survival following chemotherapy. For example, *glutathione S-transferase Pi (GSTP)* overexpression is a marker of resistance to the chemotherapeutic agent cisplatin. Expression of alleles that reduce GSTP activity can result in a higher susceptibility to drug treatment and a prolonged survival for lung, ovarian, breast, metastatic colon cancers, and multiple myeloma (Ginsburg and Willard, 2009; Calvo et al., 2005).

Box 4.3 System Biology in Pharmacology and Toxicology

Assists in:

- understanding *mechanisms of action* of pharmaceutical agents;
- screening for therapeutic *efficacy and safety* in drug discovery;
- *monitoring* of drug therapy in clinical practice;
- characterizing *drug resistance;*
- understanding the *genetic basis for variable responses* to drugs or toxicants.

Improving Treatment *(See Box 4.3)*

Targeted Therapies

Systems Biology can personalize medicine by assisting in the selection of therapies based on the genetic characteristics of the patient as well as the molecular features of the disease (Ginsburg and Willard, 2009). Some drugs have been developed that target specific sites in molecular pathways to interrupt signaling pathways. For example, imatinib is a kinase inhibitor drug used in the treatment of chronic myelogenous leukemia that inactivates signaling by inappropriately activated human epidermal growth factor receptor 2 (HER2) receptors. The molecular profile of tumors can also provide useful information regarding patient response to therapy. For example, characterizing the expression of the *HER2* gene in tumors can identify breast cancer patients who are likely to respond to either imatinib or trastuzumab, a monoclonal antibody that specifically targets HER2 (Schilsky, 2010).

Dose Calculations

There is increasing awareness in the human medical community of the impact of genetics on drug metabolism, and it is acknowledged that not all patients benefit equally from the same dosage. Pharmacogenomics examines the influence of specific genes on patient responses to relevant drugs as well as the effects of drugs on gene expression. Molecular biomarkers can provide information on drug distribution and drug targets as well as patient response to therapy (Ginsburg and Willard, 2009). Some drug metabolizing enzymes are polymorphic with variable expression, resulting in a limited or an extensive degree of metabolism of specific drugs. For example, cytochrome P450 2D6 (CYP2D6) is a polymorphic enzyme that metabolizes tamoxifen (an estrogen receptor antagonist) to active metabolites that are effective in targeting estrogen receptors in some breast cancers. Individuals with limited CYP2D6 metabolizing capacity are unlikely to benefit from tamoxifen therapy, leading to an increased risk of cancer progression. There is a need to identify and validate clinically useful biomarkers that can identify drug-metabolizing capacity and thereby predict the clinical benefit in individual patients (Schilsky, 2010).

Drug Development

Systems Biology can increase the efficiency of the process of drug discovery by identifying new drug targets or by streamlining clinical trials by selecting participants based on their drug metabolizing phenotype. For example, novel antimicrobial agents can be identified using DNA microarrays and genome sequencing data and the effect of putative anti-inflammatory compounds on proinflammatory mediators can be assessed in cell lines using two-dimensional electrophoresis or other proteomics technologies (Witkamp, 2005).

Maximizing Drug Safety

Systems Biology concepts and high-throughput "omics" technologies are also being applied to toxicology to enhance understanding of toxicity mechanisms and to facilitate the early detection and prediction of toxicities (Joshi and Kucherlapati, 2008). A personalized medicine approach may improve the safety of drug therapy by using toxicogenomics and toxicoproteomics approaches to discover new noninvasive biomarkers and gene expression patterns that reflect subclinical drug toxicity. These biomarkers serve as molecular fingerprints that could subsequently be developed into diagnostic tests that identify individuals who are susceptible to developing adverse drug reactions. Such applications would be very useful in minimizing adverse drug reactions and facilitate drug development, clinical trials, and postmarket surveillance (Elrick et al., 2006). Indeed, pharmacogenomic tests for polymorphisms of drug metabolizing enzymes have resulted in regulatory changes affecting labeling and prescribing conditions for certain medications (Downing, 2009).

Barriers to Implementing Personalized Medicine in Human Medical Practice

For personalized medicine to be fully accepted, the validity and utility of new treatments and the diagnostic and prognostic tests developed using "omics" technologies must be established. Typically, clinical utility is evaluated by assessing outcomes in expensive, randomized clinical trials. Because personalized medicine is an evolving field, the underlying knowledge base is often incomplete. This prevents universal acceptance and creates difficulties in establishing specific regulatory requirements that guide the use of genetic information in clinical practice. While advances in sequencing technologies will make the complete genomic sequence of individuals a financially viable reality, assessment tools are necessary to allow physicians to extract the relevant information for knowledge-based clinical decision-making. In addition, complex genetic tests are expensive from the perspective of their validation and use, but they have the potential for significant savings to healthcare systems. Nonetheless, broad access to genetic information has raised concerns regarding the possibility of inappropriate discrimination of individuals by employers, insurance companies, or society in general (Joshi and Kucherlapati, 2008).

Box 4.4 Characteristics of "omics" technologies

- Highly selective, sensitive, and specific
- High throughput and easily reproduced
- Enable simultaneous analysis of thousands of transcripts, proteins, or metabolites
- Enable comparisons of gene expression in different cell types, normal and diseased tissues, treated and untreated samples, etc.

Systems Biology Techniques

A number of sophisticated, high-throughput, analytical tools have been developed with considerable capacity for collection, analysis, processing, and management of large amounts of data that are relevant to the understanding of the likelihood of occurrence, progression, or management of disease (Feng et al., 2009; Ginsburg and Willard, 2009) (see Box 4.4).

The following discussion focuses on the new technological developments in high-throughput assessment tools that are currently in use in biomedical research and clinical practice.

Polymerase Chain Reaction

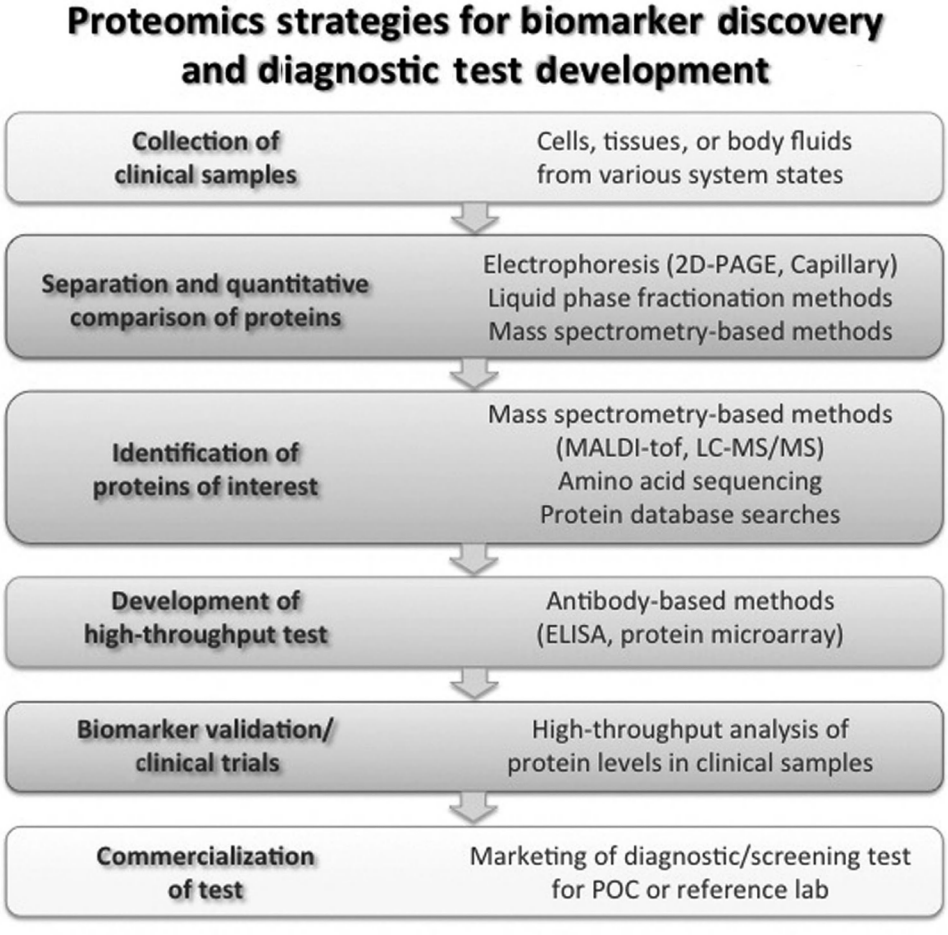
The polymerase chain reaction (PCR) is a highly specific, sensitive, and rapid technology that is capable of identifying DNA or RNA in a wide range of samples. The qualities of high specificity, sensitivity, and speed make PCR an attractive technology for the analysis of DNA and RNA in a broad range of samples that includes tissues and body fluids. In this respect, PCR represents an efficient alternative to *in vitro* culture of infectious agents or immunoassays for the diagnosis of infectious disease. Fluorescence-based real-time PCR has the added attraction that it allows for the detection and the simultaneous quantitative analysis of known genetic targets with minimal cross-contamination and a broad dynamic range. Examples of applications of real-time PCR include comparisons of gene expression profiles, analysis of SNP and specific genetic defects, viral load determination, and allelic discrimination. Multiplex PCR has the capacity to simultaneously amplify two or more genetic targets using sequence-specific primers that are unique to each target, allowing for multiple tests to be performed on a single diagnostic sample. Real-time PCR can also be used to confirm the expression of genes identified in large sample populations that were originally identified in gene expression patterns by cDNA microarray technology. A disadvantage of real-time PCR is the cost of reagents, equipment, and requirement of specialized technical expertise (Coussens and Nobis, 2002; Plummer, 2007; Hoffmann et al., 2009).

Restriction fragment length polymorphism (RFLP) is a PCR-based assay that identifies genotype by evaluating specific DNA sequences that are recognized by restriction enzymes. Digestion by restriction enzymes recognizes nucleotide differences in the amplified sites allowing for identification of allelic variants (Plummer, 2007).

Proteomics

While DNA microarray analysis identifies gene expression patterns, it provides limited understanding of changes in protein levels, protein–protein interactions, stability of proteins, and posttranslational modifications (Calvo et al., 2005). Proteomics is the study of the entire complement of proteins produced by an organism with a particular focus on structural and functional aspects and the temporal and quantitative alterations that occur in association with disease, cellular stress, toxicity, or specific physiological conditions. Proteomics approaches include various techniques that separate complex mixtures of proteins and mass spectrometry (MS) that identifies the individual separated proteins.

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) separates proteins electrophoretically; first, by isoelectric point (the pH at which the protein carries a neutral charge) and second, by molecular weight. Capillary electrophoresis (CE) uses high voltages to electrophoretically separate proteins and peptides in microcapillary



tubes. Known or novel proteins of interest can then be identified by MS techniques such as matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) and surface enhanced laser desorption ionization (SELDI). With these techniques, proteins are mixed with an energy-absorbing matrix and separated by TOF-MS by applying laser beam energy that causes desorption/ionization of proteins followed by measurement of their mass/charge ratio. Liquid chromatography-tandem mass spectrometry (LC-MS-MS) combines the physical separation of proteins by liquid chromatography with multiple steps of mass analysis and fragmentation of peptides by mass spectrometry. The LC-MS-MS can detect protein components of a complex mixture and derive the amino acid sequence of individual peptides (Lescuyer et al., 2010).

Proteomics and Biomarker Discovery

Proteomics allows for the elucidation of the pathophysiological mechanisms of disease and the identification of disease biomarkers (Lescuyer et al., 2007). Patterns of proteins and differences in protein expression in healthy and diseased patients provide proteomic signatures that characterize disease-specific profiles. While single protein biomarkers may be indicative of a specific disease, recent evidence has demonstrated that multiple serum biomarkers measured simultaneously is a more effective diagnostic approach. Other approaches suggest that the pattern of altered protein levels is diagnostic rather than the identity of biomarker proteins (Calvo et al., 2005). The identification of detected proteins allows for the evaluation of biological relevance by determining the relationship between the biomarkers and the disease process, a key factor in clinical decision-making.

Various tissues, body fluids, cells, and subcellular components have been analyzed using proteomics techniques to identify biomarkers of disease (Calvo et al., 2005; Elrick et al., 2006; Lescuyer et al., 2007, 2010). The biomarkers can then be developed into screening tools for early detection of disease, disease classification, indicators of drug toxicity, indicators of prognosis, or response to therapy. Biological fluids are readily accessible and are frequently used to assess the presence of biomarkers. However, the analysis of the serum or plasma proteome, considered to be a universal source of biomarker for many tissues, has proven to be challenging due to the large number of proteins and the predominance of high-abundance proteins such as albumin, immunoglobulins, and transferrin. Typically, abundant proteins are depleted from serum or plasma to increase the likelihood of detecting proteins of low abundance that often represent biomarkers. Unfortunately, even with efficient removal of abundant proteins, current proteomics technologies are still unable to assess the full dynamic range of protein concentrations in serum or plasma (Elrick et al., 2006; Lescuyer et al., 2007).

The cerebral spinal fluid constitutes a useful biological fluid for the diagnosis of neurological diseases as this fluid has a lower volume and it contains fewer proteins than plasma and accumulates proteins from a single organ, i.e., the brain and spinal cord (Lescuyer et al., 2007). While urine is readily obtainable, the disadvantages of analyzing this fluid are that urine collects proteins from plasma and the urogenital system, protein concentrations vary considerably and the presence of excreted compounds such as salts and urea makes comparative analysis difficult. Current thinking

is that the proteome of serum or plasma, or of other biological fluids, is altered by proteins that are liberated by diseased tissue or released by enzymatic cleavage. For example, increased serum levels of liver transaminases such as alanine aminotransferase reflects liver damage, prostate-specific antigen is a biomarker for prostate cancer and troponin I and T are indicative of acute myocardial infarction (Lescuyer et al., 2007). However, comparison of serum or plasma from cancer patients and healthy individuals has frequently revealed increases in inflammation-related proteins (e.g., haptoglobins or serum amyloid protein) that do not necessarily represent cancer-specific biomarkers.

In order to be clinically useful, biomarkers must be validated in clinical trials by demonstrating a significant and consistent difference in biomarker concentration in the diseased compared to the control state. In clinical trials, the precision and accuracy of the diagnostic test are evaluated by determining the specificity, sensitivity, and predictive value. If preliminary evaluation of sensitivity looks promising, then the biomarker may be developed into a diagnostic test using a technique such as the enzyme-linked immunosorbent assay (ELISA) that allows for quantitative measurements in larger cohort studies. Ideal biomarkers should be (a) highly specific, i.e., have a low false-positive rate such that a positive result rules a disease in (SpPin), (b) highly sensitive, i.e., have a low false-negative rate such that a negative result rules a disease out (SnNout), and (c) have a high predictive value, i.e., the probability that a positive test actually indicates the presence of the disease. In addition, the test should be clinically useful and represent a significant improvement over existing diagnostic tests. Thus, biomarker discovery is divided into a discovery and a validation phase, both of which involve close collaboration between clinical chemists and clinicians. (Lescuyer et al., 2007).

Metabolomics

One of the most recent "omics" technologies to be applied to Systems Biology and clinical medicine and research is metabolomics. Metabolomics involves the identification and quantification of all metabolites in a cell, tissue, or whole organism under specific states or conditions (Weckwerth, 2010). As with other Systems Biology techniques that contribute to the characterization of molecular phenotypes, metabolomics has evolved with technological advancements. In particular, nuclear magnetic resonance, mass spectrometry, bioinformatics, and advanced statistical analysis allow for high-throughput analysis of the physicochemical properties of a metabolome. However, compared to proteins and nucleic acids, the chemical diversity of metabolites makes them more difficult to extract and analyze with the use of a universal technique (Hocquette et al., 2009). Moreover, analytical sensitivity for the detection of metabolites is critical, as there are no amplification techniques that are available for nucleic acids. The separation of organic metabolites by gas or liquid chromatography and subsequent ionization and separation of ions according to their mass/charge ratio by MS allows for the sensitive detection of metabolites in tissues or body fluids in micromolar quantities (Dettmer et al., 2007). While biomarkers of metabolism can be identified through comparisons of mass information with existing libraries, a major limitation with metabolomics is the paucity of extensive databases and

bioinformatics tools for Systems Biology-based analysis of metabolites (Hocquette et al., 2009).

In human medicine, metabolomics has typically been used to characterize dietary and therapeutic interventions and disorders of metabolism (Oresic, 2009) with applications in nutrition research and biomarker studies in diagnostics and drug research. These metabolomics studies have focused on age- and obesity-related disorders, such as diabetes mellitus, metabolic syndrome, and atherosclerosis. However, metabolomics is increasingly applied to the study of pathogenic mechanisms of complex diseases, such as cancer, cardiovascular diseases as well as disorders of the nervous, endocrine, and digestive systems (Griffiths et al., 2010).

Novel Technologies

Several new technologies have recently been developed that have some considerable advantages over conventional analytical systems. Our discussion focuses on those based on microfluidic systems and nanotechnology.

Microfluidic Systems

By miniaturizing components of existing analytical techniques, microfluidic technology has the potential to improve existing Systems Biology tools by reducing sample consumption and analysis time and maximizing throughput. However, the requirements of micropumps and miniaturized detection systems constitute disadvantages that complicate the mechanics of microfluidic systems (Feng et al., 2009). Microfluidic methods based on CE have been used to analyze DNA, protein, and metabolites. For example, chip-based CE with multiple microfluidic channel arrays markedly improves DNA sequencing performance compared to conventional technology. In addition, pyrosequencing, a DNA sequencing method that monitors pyrophosphate release on nucleotide incorporation, has been adapted to microfluidics by trapping DNA on microbeads in a filter chamber while monitoring the flow of nanoliter quantities of pyrosequencing reagents. Other applications of microfluidics include cell culture, cell sorting, and cell–cell signaling (Feng et al., 2009). Moreover, microreactors for DNA amplification have been developed that involve submicroliter PCR chambers with microfabricated CE systems for ultrasensitive PCR of single DNA template molecules and real-time detection of genomic samples (Feng et al., 2009).

As mentioned above, the detection of low abundance target proteins by proteomics analysis requires a preconcentration step through the elimination of high-abundance proteins. Enrichment of low-abundance proteins has been facilitated by microfluidic technology using miniaturized isoelectric focusing, immunoaffinity CE or membrane-based approaches. On-chip protein separation involving one-dimensional and 2D CE separation systems is also achieved via microfluidics. Microfluidic chips have also been coupled to MS analysis to facilitate peptide mapping, to identify posttranslational modifications, protein–protein interactions, and amino acid sequences of biomarkers thereby addressing issues such as speed, throughput, and cost efficiency (Feng et al., 2009).

Cell-Based Biosensors (CBBs)

Current pathogen detection methods rely on the culture of microorganisms, PCR, or immunochemical approaches that are time-consuming and often inconclusive or fail to detect unknown or emerging pathogens. There is a need to develop broad-spectrum cell-based screening tools for rapid biosensing of pathogenic microorganisms that can be used on patients at point-of-care (i.e., bedside or stall-side) or for environmental monitoring. Cell-based biosensors (CBBs) are powerful tools that rapidly assess functional responses of cells to stressors such as pathogens and toxins, and other hazards relevant to clinical, environmental, agricultural, and pharmaceutical settings. By combining methods pertaining to microbiology, molecular biology, physics, and engineering, CBBs comprise of cell culture systems (prokaryotic or eukaryotic cells) that detect cell–analyte interactions with the capacity for high-throughput optical and electrical screening (Banerjee and Bhunia, 2009). Disadvantages of CBBs are a lack of specificity, a limited long-term stability, fragility, and high cost (Banerjee and Bhunia, 2009). However, refinements to CBB systems, such as three-dimensional cell culture systems, modified growth media, genetically engineered cell surface antibodies or receptors and fluorescence probes, have improved specificity and sensitivity of analyte detection (Banerjee and Bhunia, 2009). For example, B lymphocyte-based biosensors that express immunoglobulins to specific pathogens produce bioluminescence that can be detected by a luminometer when specific bacteria or viruses bind to the antibodies. Other cell-based assays detect cytotoxicity resulting from cellular membrane damage induced by the bacterium or toxin. Membrane damage can then be measured by intracellular incorporation of dyes or release of cellular enzymes such as lactate dehyrogenase that are detected by secondary transducers such as microelectrodes or optical detectors (Banerjee and Bhunia, 2009).

Personalized Medicine Versus Livestock Population Health

The intrinsic differences in human and veterinary medicine precludes their direct comparison with respect to the application of Systems Biology approaches in healthcare. For example, personalized or individualized medicine is unlikely to be of major interest in the treatment of livestock diseases, as there is limited emphasis on customizing healthcare for individual food-producing animals and the major focus is on herd or population health. Notable exceptions are individual incidences of health problems in high-yielding dairy cows, where certain individual animals are more sensitive to diseases such as subclinical rumen acidosis, mastitis or Johne's disease, and other problems during early lactation. A current challenge is to use Systems Biology techniques to determine the underlying determinants of individual susceptibility to disease and to implement prevention and selective breeding strategies to reduce these incidences. The prevalence of chronic degenerative diseases such as cancer, diabetes, and heart disease in aging human beings has spawned interest in individualized medicine due to the need for increased understanding of molecular pathogenesis and more precise and effective diagnostic tests and therapies. Factors such as the economics of modern agriculture and intensive livestock production, the limited life span of food animals, and differences in environmental influences on disease

Box 4.5 Impact of Systems Biology in Livestock Management

- Diagnosis of disease and determining etiology
- Identification of genetic determinants of disease susceptibility or resistance
- Understanding pathogenic mechanisms and host–pathogen interactions
- Disease surveillance and identification of factors affecting spread of diseases
- Development of vaccines, antimicrobial drugs, and selective breeding programs

development tend to shift the focus of healthcare of livestock from chronic degenerative diseases to infectious diseases. Again, dairy cattle is the obvious exception to this trend in livestock health management due to their comparatively longer life spans that lead to the development of chronic problems such as infertility, metabolic diseases, lameness and mastitis, and other factors associated with suboptimal production. As a result, current interest is centered on the use of Systems Biology concepts and technologies in the area of molecular diagnostics of infectious agents as well as the identification of factors associated with susceptibility or resistance to infectious disease. While there is less investment in the development of new pharmaceuticals for animal health than in human medicine, Systems Biology may have a large impact in the development of vaccines and new antimicrobial drugs for livestock (see Box 4.5).

Until recently, limitations in the availability of reagents and genomic information have limited our understanding of the molecular pathogenesis of infectious diseases of livestock species. However, the sequencing of the genomes of major domestic species and postsequence activities such as proteomics and metabolomics promise to provide significant benefits for livestock species. Following the sequencing of the human genome in 2001, several projects aimed at sequencing the genomes of food-producing animals including cattle, chickens, pigs, sheep, and several aquatic species are either completed or ongoing (Plummer, 2007). In April 2009, domestic cattle were the first livestock species to have their genome sequence published, followed by the pig genome in November, 2009. The US Livestock Species Genome Projects Web page (http://www.animalgenome.org/), supported by the National Animal Genome Research Program provides information on the status of these projects and information on livestock species. The National Institute of Food and Agriculture, US Department of Agriculture in conjunction with the governments of Australia and New Zealand are currently working to provide a road map of the sheep genome. In addition, public data from prokaryotic genome sequencing projects is available at the following Web site: http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html. Finally, introduction of new generation DNA sequencing technologies should allow for the rapid and inexpensive sequencing of individual genomes in under 4 weeks for less than US$ 50,000 (Pushkarev et al., 2009).

At the moment, the challenge is to not to gather as much information as possible but to understand the functional and operational implications of genomic information. A holistic approach is required that integrates the various elements (DNA, RNA, proteins, metabolites, cells tissues) and biological information in order to unravel the complexities of environment–genotype interactions and the underlying biology

of the resultant phenotypes of livestock disease. In the postsequencing era, use of increasingly sophisticated quantification platforms such as microarrays, proteomics, metabolomics, EST libraries, and QTL markers will help to elucidate the influence of environmental factors on gene expression and phenotype and the underlying mechanisms that control functionality (Green et al., 2007). This process will be facilitated by the already well-characterized phenotypes of agricultural species that have been closely monitored and modified through selective breeding (Green et al., 2007). The principal benefit for food-producing animals will be in translating genomic data to identify those phenotypic characteristics with added value such as disease resistance, optimal reproductive efficiency, or enhanced growth rate and to understand the underlying mechanisms. In addition, genomics can create opportunities to develop and validate new tests and screening tools to diagnose disease and detect emerging pathogens, to increase understanding of disease pathogenesis or mechanisms of antimicrobial resistance, to discover and evaluate new therapies, and to identify biomarkers of drug efficacy and toxicity, all of which could contribute to maximizing the health and productivity of food animals (Green et al., 2007).

Molecular Diagnostics

Identification of Pathogens

In veterinary medicine of livestock, the most frequent and extensive application of Systems Biology techniques is in the area of molecular diagnostics. The emergence of PCR enabled the development of new testing modalities that can identify and amplify specific DNA or RNA sequences from infectious agents (e.g., bacteria and viruses) and specific genomic sequences responsible for disease susceptibility and resistance in production animals (Plummer, 2007). PCR-based molecular tests are routinely performed by veterinary diagnostic lab to provide definitive diagnosis of infection based on the presence of specific pathogens. For example, PCR assays can differentiate toxin types of *Clostridium perfringens* and identify strains of *Escherichia coli* and *Pasteurella multocida* based on PCR amplification of toxin genes. In many diagnostic lab, the cost of diagnostic testing is reduced through the use of multiplex PCR technology that can detect multiple pathogens or genetic toxin typing in a single reaction. As was mentioned previously, real-time PCR may eventually replace conventional PCR in molecular diagnosis due to the advantages of quantification of DNA or RNA copy numbers and the reduced risk of sample cross-contamination (Plummer, 2007). The challenges of reliable real-time PCR detection of RNA viruses with mutation-induced variable genomes can be circumvented by running multiplex PCR with primers to different regions of the viral genome (Hoffmann et al., 2009).

The impact of real-time PCR on molecular diagnostics is most apparent in the diagnosis of viral pathogens of livestock that can result in diseases with serious impact and are notifiable to the World Organisation for Animal Health (OIE) or other regulatory agencies. Examples of these diseases include foot-and-mouth disease (FMD), classical swine fever virus (CSF), bluetongue virus (BTD), avian influenza (AI), and Newcastle disease (ND) (Hoffmann et al., 2009).

Foot-and-Mouth Disease (FMD)

Foot-and-mouth disease is a severe, highly contagious viral disease of cattle, sheep, goats, pigs, and cloven-hoofed wild ruminants. While there are no public health or food safety risks associated with FMD, the devastating impact of this disease on the food industry and entire economies due to rapid spread, high morbidity, and reduced animal productivity necessitates rapid testing and disease control actions such as widespread culling of infected and at-risk animals in response to FMD. In Canada, confirmatory testing for FMD is done by the Canadian Food Inspection Agency at the National Centre for Foreign Animal Diseases by either tissue culture, virus isolation, ELISA, or RT-PCR. In the 2001 outbreak of FMD in the United Kingdom, control strategies required that infected animals be slaughtered within 24 hours of detection. Given the requirement for speedy diagnosis, portable RT-PCR machines were required in the field. This reflects the need for simple, easy to use, point-of-care technologies for pathogen diagnosis and robust protocols that eliminate tissue-derived inhibitors of the PCR reaction. In addition, current studies are investigating the use of RNA interference (RNAi) techniques to identify regions of the FMD virus that are least variable, information that would be useful in the generation of disease-resistant livestock, and the strategic development of new drugs and vaccines (Pengyan et al., 2008).

Bluetongue Disease (BTD)

Bluetongue disease is a noncontagious viral disease of domestic and wild ruminants that is transmitted by insects, particularly biting midges of the *Culicoides* species. BTD virus causes serious debilitating disease in domestic ruminants and significant mortality, requiring rapid diagnosis and control measures. The recent spread of BTD into northern Europe has necessitated the development of high-throughput protocols for rapid and sensitive detection of BTD virus by real-time PCR.

Avian Influenza (AI)

Avian influenza is a contagious viral infection caused by the influenza virus Type "A," family Orthomyxoviridae that can affect several species of poultry as well as wild birds. The AI viruses can be classified into two categories: low pathogenic AI (LPAI) and high pathogenic AI (HPAI) forms based on the severity of the illness. The HPAI form is extremely infectious in poultry resulting in rapid spread and high mortality within 48 hours. Some HPAI strains such as H5N1 may, on rare occasions, cause disease in humans. While virus isolation is the gold standard for AI diagnosis, this method and subsequent testing can take several days to reach a diagnosis. Real-time PCR is also a valuable and rapid diagnostic tool for clinical specimens particularly in LPAI outbreaks, where clinical signs may be vague, and in HPAI epidemics, where emergency management is dependent upon rapid and definitive diagnosis. Primer sets specific for the hemagglutinin gene are currently used; however, high sequence variability may affect diagnostic sensitivity of validated RT-PCR protocols (Hoffmann et al., 2009).

Newcastle Disease (ND)

Newcastle disease is a highly contagious viral disease of wild birds and domestic fowl that causes a high incidence of mortality and serious economic impact to the poultry industry. Conventional diagnostic methods are labor-intensive, time-consuming, expensive, and can significantly delay disease control measures. While RT-PCR has the potential for rapid diagnosis, the variability of the F (fusion protein) gene precludes reliable amplification of all virulent ND viruses. Due to the devastating consequences of misdiagnosis resulting from false-negative or false-positive data, RT-PCR is currently not approved by the World Organisation for Animal Health (OIE) for the diagnosis of ND (Hoffmann et al., 2009).

Classical Swine Fever (CSF)

Classical swine fever is a highly contagious viral disease that causes high morbidity and frequent fatalities in domestic and wild pigs. Because of variable and nonspecific clinical and pathological signs, laboratory testing is essential for the confirmation of diagnosis. While virus isolation and immunochemical assays for virus or antibody detection are standard, PCR techniques including RT-PCR are becoming increasingly important and acceptable approaches for rapid and accurate identification of CSF (Hoffmann et al., 2009).

Identification of Genetic Disease

The sequencing of the genomes of several food-producing animals has facilitated the development of molecular diagnostics for genetic diseases and the identification of genes responsible for susceptibility to specific diseases of economic importance (Plummer, 2007). While diagnostic testing for genetic disease in livestock is rare, an increasing number of tests is commercially available that can be used to detect the genetic variations associated with these diseases. For example, ovine hereditary chondrodysplasia (spider lamb syndrome) characterized by musculoskeletal deformities in lambs is caused by a SNP in the *fibroblast growth factor receptor 3* gene that results in loss of tyrosine kinase function. A commercial test is currently available to screen breeding animals and detect carriers of this gene (Plummer, 2007). Some of the commercial tests for genetic diseases of cattle currently available in North America include deficiency of uridine monophosphate synthetase (*UMPS* gene), complex vertebral malformation (*SLC35A3* gene), and Factor IX deficiency (*factor IX* gene) in Holstein cows, protoporphyria (*ferrochelatase* gene) in Limousin cattle, bovine hereditary zinc deficiency (*SLC39A4* gene) in Shorthorn cattle, alpha-mannosidosis (MANNA) in Angus cattle, and platelet bleeding disorder (CalDAG-GEFI) in Simmental cattle.

The main utility of these molecular diagnostics tests is to screen breeding animals in order to identify carriers and limit their use with the eventual goal of genetic disease elimination. As a result, some breed associations now require the inclusion of genetic disease test results on official documents of registered animals. The use of this type of

molecular diagnostic testing will undoubtedly become routine as genetic diseases are characterized and the associated genes are identified.

Determining Disease Susceptibility or Resistance

Susceptibility to disease may result from a limited functional capacity in one or more of the components of host defense mechanisms. For example, bovine leukocyte adhesion deficiency (BLAD) is a disease of cattle characterized by an increased susceptibility to infection due to the dysfunction of leukocytes. Neutrophil recruitment to sites of inflammation requires that neutrophils adhere to, and migrate through, the capillary endothelium. This is an autosomal recessive disease due to a genetic polymorphism in the *CD18* gene that codes for adhesion molecules in leukocytes reducing their adhesion to the capillary endothelium (Nagahata, 2004). Carriers of BLAD can be readily detected by PCR followed by restriction enzyme analysis of the amplicons (i.e., RFLP). Because the use of BLAD carrier bulls in artificial insemination programs increases the prevalence of BLAD in dairy herds, control programs are focusing on eliminating the BLAD carriers via PCR screening (Norouzy et al., 2005).

Genetic polymorphisms can also affect disease susceptibility in the presence of a pathogen or toxin. Malignant hyperthermia (MH) is an inherited myopathy and hypermetabolic syndrome involving skeletal muscle. Clinical signs are characterized by hyperthermia, tachycardia, hypoxemia, metabolic and respiratory acidosis, muscle rigidity, and death. Researchers at the University of Guelph and the University of Toronto determined the cause of MH to be a mutation in the ryanodine receptor gene that controls Ca^{2+} release from the sarcoplasmic reticulum in skeletal muscle (Fujii et al., 1991). Malignant hyperthermia is typically triggered by succinylcholine or volatile anesthetics in genetically susceptible pigs and humans, but the porcine disease may also be triggered by environmental stresses (e.g., rough handling, transportation, fighting, or hot weather) and hence it is also called porcine stress syndrome. A PCR-based test that detects the mutation in the ryanodine receptor gene can identify homozygous MH-resistant and MH-susceptible animals as well as heterozygous carriers. Pigs carrying this mutation are more prone to producing pale, soft, and exudative pork that is unattractive to consumers due the gray, soft, and watery appearance.

Systems Biology approaches have the potential to elucidate the relative contributions of genetics and the environment to the susceptibility of food-producing animals to more common problems such as gastrointestinal infections in poultry and pigs and rumen malfunction in ruminants. An integrated Systems Biology approach would benefit livestock producers by providing critical information and guidance as to whether specific production-limiting problems should be handled by farm management or by genetics/breeding programs or a combination of both.

Using Systems Biology to Understand Host–Pathogen Interactions

Study of the interactions between pathogens or toxins and host gene products is an important requirement in understanding the pathogenesis of disease. This knowledge

is fundamental to the implementation of intervention strategies that are intended to prevent or treat disease. However, use of Systems Biology concepts and high-throughput technologies is often hindered by a lack understanding of gene function. One of the main challenges of Systems Biology is to convert large datasets derived from high-throughput technologies into meaningful information related to structure and function. Gene Ontology (GO) is a well-developed and commonly used bioontology process for applying Systems Biology to livestock and poultry health and productivity and the interrelationship of host and pathogen. The GO provides "functional annotations based on molecular function, biological process, and cellular component" (Mccarthy et al., 2009). In this way, GO can impart relevance to data derived from microarray and proteomics technologies allowing researchers to understand the function and interactive relationships of pathogens and various host gene products in order to elucidate disease mechanisms (Mccarthy et al., 2009). Recently, a Web-accessible database called "AgBase" (www.agbase.msstate.edu) has been established to provide assistance with analysis of functional genomics datasets and interpretation of Systems Biology in agricultural species (Mccarthy et al., 2007).

Molecular Epidemiology

The field of molecular epidemiology uses molecular techniques to examine the causes and spread of disease. The objective of molecular epidemiology is to identify the etiologic agents (e.g., viruses, bacteria, parasites, toxins, etc.) and detect their source, the dynamics and routes of disease transmission, the determinants of disease spread (e.g., virulence genes), the factors that affect disease control such as drug resistance (e.g., against antimicrobials), and antigens relevant to vaccine efficacy (e.g., immunogenic epitopes) (Zadoks and Schukken, 2006). In addition, molecular epidemiology can detect environmental or host factors, or conditions that influence the introduction and spread of disease-causing agents. For example, molecular biology methods can be used to identify and monitor the global spread of highly contagious and pathogenic strains of microorganisms (Guan et al., 2009). These methods distinguish between persistent, emerging, or reemerging diseases, and they can differentiate vaccinated farm animals from those with natural infections (Klein, 2009). Molecular epidemiology is becoming an increasingly important resource for veterinary public health and the maintenance of health of food-producing animals (Zadoks and Schukken, 2006).

Comparative Genotyping

The majority of comparative genotyping methods used in molecular epidemiology, in particular the applications used in the identification of different strains of pathogens, involve PCR (Zadoks and Schukken, 2006). For example, PCR specifically identifies zoonotic pathogens in food products such as *Listeria monocytogenes* in meat or *Salmonella* sp. in contaminated milk. In addition, PCR-based strain typing methods and sequence-specific primers can be used to identify specific subspecies of bacteria. Alternatively, antimicrobial resistance or virulence genes can be identified and their

horizontal transmission tracked through a process called multilocus sequence typing (MLST). With MLST, the identification of strains is achieved by sequencing multiple distinctive genes or loci within the genome of an organism (Zadoks and Schukken, 2006). Thus, molecular epidemiology approaches can monitor the evolution of virulent strains of pathogens (e.g., *E. coli* 0157:H7) and the acquisition and spread of antimicrobial resistance (e.g., methicillin-resistant *Staphylococcus aureus*). In addition, the geographic site of origin and transmission of RNA viruses, such as FMD that can have potentially devastating effects on food security and those viruses such as HPAI with severe public health implications, can be monitored by RT-PCR (Zadoks and Schukken, 2006).

Source Tracing

Molecular epidemiology plays an important role in the identification of the source of infection or contamination. For example, the source of contamination of drinking water by *E. coli* 0157:H7 in the town of Walkerton, Ontario, Canada, that resulted in the death of seven people and 2500 illnesses in May, 2000, was traced by the genetic testing of well water contaminated by healthy carrier cattle. Similarly, the origin of the pandemic H1N1 strain of influenza that emerged in Mexico and the United States in 2009 probably originated in swine, since DNA sequencing determined that the closest ancestral gene of the novel segments found in this strain were of swine origin (Forrest and Webster, 2010). Moreover, molecular epidemiological studies have determined that the highly pathogenic H5N1 strain of avian influenza crossed the host species barrier from birds to humans in 1997 in southern China and was eventually spread globally via migrating birds (Forrest and Webster, 2010). Thus, techniques used in molecular epidemiology have generated important information concerning the point of origin of pathogens, and they are instrumental in monitoring the spread of zoonotic disease from animals to humans. Since many livestock diseases have profound implications on the security and safety of our food supply, molecular epidemiology has considerable impact on both animal and public health emergency management. A clear understanding of the transmission dynamics of highly contagious viral diseases such as FMD, AI, or ND can influence international trade through the implementation and enforcement of severe regulatory measures such as trade embargoes, quarantine of animals and food products, strict sanitary procedures, and import controls (Zadoks and Schukken, 2006).

Pathogen Adaptation to Host Species

The ability of pathogenic strains to adapt to specific host species has been characterized by molecular epidemiology studies. For example, different strains of *Streptococcus agalactiae* cause mastitis in dairy cows and clinical infection in humans. Moreover, strains of *Mycobacterium avium subspecies paratuberculosis* found in sheep are different from those isolated from goats and cattle suggesting that sheep do not represent a risk of transmission to cattle.

Vaccine Development

Molecular epidemiology can also be useful in vaccine development and predicting the efficacy of a vaccine. Evaluating whether a disease is due to a single or multiple strains of a pathogen determines whether a multivalent vaccine will be required. In addition, characterizing the geographic distribution of serotypes will determine vaccine efficacy in different countries (Terregino et al., 2008; Reeve et al., 2010).

Example of Systems Biology Applications in Livestock Health: Mastitis in Dairy Cattle

Mastitis is the most frequent and costly disease in dairy herds resulting in reduced productivity and negatively affects yield, composition, and technological properties of milk (Seegers et al., 2003). The overall mean incidence rate of clinical mastitis is 23 cases per 100 cow-years in Canadian dairy herds (Olde Riekerink et al., 2008). While there are no current statistics for the incidence rate of subclinical mastitis on Canadian dairy farms, it has been estimated that for every case of clinical mastitis in the herd, there are between 15 and 40 subclinical cases (Gill et al., 1990).

Immunogenomics research may provide molecular explanations for genetic variation in susceptibility to mastitis, and it may identify potential new therapeutic targets and strategies to protect against mastitis-causing bacteria by enhancing mammary immunity in susceptible cows. Analysis of gene expression profiles in neutrophils and other leukocytes in early nonlactating and periparturient dairy cows may identify genes that affect susceptibility to clinical mastitis. Moreover, knowledge of the genomics of bacterial pathogens may allow for the production of DNA vaccines that express genes involved in antigen presentation to T helper-1 lymphocytes. In this way, DNA vaccines or novel adjuvant drugs could stimulate B lymphocytes to produce antibodies that target pathogens involved in mastitis (Burton and Erskine, 2003).

The current methods of diagnosing clinical mastitis, e.g., somatic cell counting, culture of causative microorganisms, measurement of biomarker enzymes (e.g., N-acetyl-ß-D-glucosaminidase and lactate dehydrogenase) and California mastitis testing, have disadvantages of being either labor-intensive, costly, or ineffective in detecting subclinical mastitis. Therefore, there is a need to develop rapid and sensitive diagnostic assays for early detection of mastitis based on novel molecular biomarkers and sensor-based platforms in order to reduce the economic effects of this disease (Viguier et al., 2009). Proteomics techniques have identified proteins found at varying concentrations in milk from cows with mastitis. Haptoglobin and serum amyloid A (SAA) are two sensitive acute phase proteins that have been assessed in serum and in milk during acute clinical mastitis and in milk from cows with experimentally induced chronic sub-clinical *S. aureus* mastitis (Gronlund et al., 2005). Molecular chaperone proteins involved in protein folding, proteins associated with neutrophils and macrophage function and serum proteins such as transferrin and bovine serum albumin have also been found in mastitic milk (Viguier et al., 2009). Other researchers have identified bacterial proteins associated with pathogen resistance in milk with mastitis. While

some immunoassays have been developed to these biomarkers, validation in clinical trials is required before these tests are routinely accepted as screening tools for mastitis detection (Viguier et al., 2009).

In summary, Systems Biology principles and technologies have the capacity to reduce the impact of mastitis on the dairy industry by identifying new diagnostic biomarkers and by elucidating the immunopathogenic mechanisms of mastitis. The latter in particular may reveal targets for prevention and therapy. For example, metabolic adaptations such as altered lipid metabolism and oxidative stress during the periparturient period, may result in changes in the bovine immune system that contribute to uncontrolled inflammation and the development of disorders of economic importance such as metritis and mastitis. (Goff, 2006; Sordillo et al., 2009)

Conclusion: Challenges of Applying Systems Biology Concepts and Techniques to Livestock Health Management

The application of Systems Biology concepts and techniques to veterinary medicine and livestock health and disease has made significant advances in recent years, but still lags behind human medicine. Indeed, there are very few examples of gene-based healthcare currently in use in livestock veterinary medicine. Reasons for the delay in the widespread implementation of new molecular techniques in the study of livestock health and disease include a lack of molecular reagents for the target animals, minimal characterization of drug metabolism, limited capacity for bioinformatics, the complexity of data required to support System Biology applications for herd health, and a paucity of commercial tests for genotyping diseases and genes related to disease resistance (Coussens and Nobis, 2002). Newly generated data must be rigorously assessed with reproducible assays so that outcomes are evidence-based. Thus, genes and protein products that are responsible for disease resistance and susceptibility must be characterized and investigated as potential opportunities for therapeutic intervention. Biomarkers of early detection, prognostication, and response of disease to therapy must be characterized and validated as tests for use in selective breeding and disease prevention programs. This approach will reduce false associations between gene expression and health that may be costly to the food animal industry.

There is a need for increased funding for basic and clinical research in order to gain a full understanding of the mechanisms and factors that influence production-limiting diseases, to identify novel targets for therapy, and to formulate strategies for prevention and control of diseases of livestock. Government funding is limited, private veterinary diagnostics labs and pharmaceutical companies have typically not invested substantially in research and development, and livestock organizations have minimal resources and impact. On the other hand, researchers must develop reliable diagnostic tests that will be perceived as useful and acceptable by regulatory authorities, diagnostic companies, and clinical practitioners (Ginsburg and Willard, 2009). National agencies such as the Canadian Food Inspection Agency and the US Department of Agriculture and international organizations such as the OIE and the FAO must be supportive of the concepts and benefits of Systems Biology in livestock health. These organizations must promote the integration of the Systems Biology paradigm and related technologies into the culture of livestock production so that the benefits are

clear to food animal producers and veterinarians. Livestock veterinary medicine needs to build on the momentum of preventive herd health that has developed over the past 25 years by replacing the reactive approach to disease and management and maintenance of livestock health to one that focuses on disease prediction and prevention in food-producing animals (Borden and Raghavan, 2010; Schilsky, 2010).

In conclusion, Systems Biology approaches and technologies will undoubtedly contribute significantly to improving health and productivity of livestock. Moreover, implementation of Systems Biology concepts and practices has the potential to provide enormous benefits for livestock producers and consumers of animal products. For example, there is an opportunity to distinguish health from disease using multiparameter analysis of tissues and body fluids (e.g., disease diagnosis), to determine the probability of disease in individual animals (e.g., immune resistance), to monitor exposure to environmental signals (e.g., identification of pathogens and toxins), and to facilitate the integration the various livestock research disciplines (e.g., animal nutrition, management, breeding, and veterinary sciences). Handheld microfluidic devices have the capacity for on-site analysis of protein or mRNA levels, detection of emerging gene mutations, and exposure to infectious agents. These technologies are currently in development and could contribute considerably to assuring a safe and secure food supply and minimizing the impact of production-limiting disease (Downing, 2009). In view of this potential positive impact, governments and the livestock industry should develop a long-term vision and strategy to support research on Systems Biology and the development of high-throughput technologies that could significantly influence health, performance, and productivity of food-producing animals (Green et al., 2007).

References

Banerjee, P. & Bhunia, A.K. (2009) Mammalian cell-based biosensors for pathogens and toxins. *Trends in Biotechnology*, **27**, 179–188.

Borden, E.C. & Raghavan, D. (2010) Personalizing medicine for cancer: the next decade. *Nature Reviews Drug Discovery* **9**, 343–344.

Burton, J.L. & Erskine, R.J. (2003) Immunity and mastitis. Some new ideas for an old disease. *Veterinary Clinics of North America: Food Animal Practice* **19**, 1–45.

Calvo, K.R., Liotta, L.A., & Petricoin, E.F. (2005) Clinical proteomics: from biomarker discovery and cell signaling profiles to individualized personal therapy. *Bioscience Reports* **25**, 107–125.

Coussens, P.M. & Nobis, W. (2002) Bioinformatics and high throughput approach to create genomic resources for the study of bovine immunobiology. *Veterinary Immunology and Immunopathology* **86**, 229–244.

Dettmer, K., Aronov, P.A., & Hammock, B.D. (2007) Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews* **26**, 51–78.

Downing, G.J. (2009) Policy perspectives on the emerging pathways of personalized medicine. *Dialogues in Clinical Neuroscience* **11**, 377–387.

Elrick, M.M., Walgren, J.L., Mitchell, M.D., et al. (2006) Proteomics: recent applications and new technologies. *Basic & Clinical Pharmacology & Toxicology* **98**, 432–441.

Feng, X., Du, W., Luo, Q., et al. (2009) Microfluidic chip: next-generation platform for systems biology. *Analytica Chimica Acta* **650**, 83–97.

Forrest, H.L. & Webster, R.G. (2010) Perspectives on influenza evolution and the role of research. *Animal Health Research Reviews* **11**, 3–18.

Fujii, J., Otsu, K., Zorzato, F., et al. (1991) Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* **253**, 448–451.

Gill, R., Howard, W.H., Leslie, K.E., et al. (1990) Economics of mastitis control. *Journal of Dairy Science* **73**, 3340–3348.

Ginsburg, G.S. & Willard, H.F. (2009) Genomic and personalized medicine: foundations and applications. *Translational Research* **154**, 277–287.

Goff, J.P. (2006) Major advances in our understanding of nutritional influences on bovine health. *Journal of Dairy Science* **89**, 1292–1301.

Green, R.D., Qureshi, M.A., Long, J.A., et al. (2007) Identifying the future needs for long-term USDA efforts in agricultural animal genomics. *International Journal of Biological Science* **3**, 185–191.

Griffiths, W.J., Koal, T., Wang, Y., et al. (2010) Targeted metabolomics for biomarker discovery. *Angewandte Chemie International Edition* **49**, 5426–5445.

Gronlund, U., Hallen Sandgren, C., & Persson Waller, K. (2005) Haptoglobin and serum amyloid A in milk from dairy cows with chronic sub-clinical mastitis. *Veterinary Research* **36**, 191–198.

Guan, Y., Smith, G. J., Webby, R., et al. (2009) Molecular epidemiology of H5N1 avian influenza. *Revue Scientific Technology* **28**, 39–47.

Harrill, A.H. & Rusyn, I. (2008) Systems biology and functional genomics approaches for the identification of cellular responses to drug toxicity. *Expert Opinion on Drug Metabolism & Toxicology* **4**, 1379–1389.

Hocquette, J.F., Cassar-Malek, I., Scalbert, A., et al. (2009) Contribution of genomics to the understanding of physiological functions. *Journal of Physiology & Pharmacology* **60**(Suppl 3), 5–16.

Hoffmann, B., Beer, M., Reid, S.M., et al. (2009) A review of RT-PCR technologies used in veterinary virology and disease control: sensitive and specific diagnosis of five livestock diseases notifiable to the World Organisation for Animal Health. *Veterinary Microbiology* **139**, 1–23.

Joshi, V.A. & Kucherlapati, R. (2008) Genetics and genomics in the practice of medicine. *Gastroenterology* **134**, 1284–1288.

Klein, J. (2009) Understanding the molecular epidemiology of foot-and-mouth-disease virus. *Infection, Genetics and Evolution* **9**, 153–161.

Lescuyer, P., Farina, A., & Hochstrasser, D.F. (2010) Proteomics in clinical chemistry: will it be long? *Trends in Biotechnology* **28**, 225–229.

Lescuyer, P., Hochstrasser, D., & Rabilloud, T. (2007) How shall we use the proteomics toolbox for biomarker discovery? *Journal of Proteome Research* **6**, 3371–3376.

Mccarthy, F.M., Bridges, S.M., Wang, N., et al. (2007) AgBase: a unified resource for functional analysis in agriculture. *Nucleic Acids Research* **35**, D599–D603.

Mccarthy, F.M., Mahony, T.J., Parcells, M.S., et al. (2009) Understanding animal viruses using the Gene Ontology. *Trends in Microbiology* **17**, 328–335.

Nagahata, H. (2004) Bovine leukocyte adhesion deficiency (BLAD): a review. *Journal of Veterinary Medical Science* **66**, 1475–1482.

Norouzy, A., Nassiry, M.R., Eftekhari Shahrody, F., et al. (2005) Identification of bovine leucocyte adhesion deficiency (BLAD) carriers in Holstein and Brown Swiss AI bulls in Iran. *Genetika* **41**, 1697–1701.

Olde Riekerink, R.G., Barkema, H.W., Kelton, D.F., et al. (2008) Incidence rate of clinical mastitis on Canadian dairy farms. *Journal of Dairy Science* **91**, 1366–1377.

Olden, K. (2006) Toxicogenomics—a new systems toxicology approach to understanding of gene-environment interactions. *Annals of the New York Acadmy of Sciences* **1076**, 703–706.

Oresic, M. (2009) Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction. *Nutrition, Metabolism and Cardiovascular Diseases* **19**, 816–824.

Peng, X., Chan, E.Y., Li, Y., et al. (2009) Virus-host interactions: from systems biology to translational research. *Current Opinion in Microbiology* **12**, 432–438.

Pengyan, W., Yan, R., Zhiru, G., et al. (2008) Inhibition of foot-and-mouth disease virus replication in vitro and in vivo by small interfering RNA. *Virology Journal* **5**, 86.

Plummer, P.J. (2007) Molecular diagnostics for the food animal practitioner. *Veterinary Clinics of North America: Food Animal Practice* **23**, 481–501.

Pushkarev, D., Neff, N.F., & Quake, S.R. (2009) Single-molecule sequencing of an individual human genome. *Nature Biotechnology* **27**, 847–852.

Reeve, R., Blignaut, B., Esterhuysen, J.J., et al. (2010) Sequence-Based Prediction for Vaccine Strain Selection and Identification of Antigenic Variability in Foot-and-Mouth Disease Virus. *PLoS Computational Biology* **6**, e1001027.

Schilsky, R.L. (2010) Personalized medicine in oncology: the future is now. *Nature Reviews Drug Discovery* **9**, 363–366.

Seegers, H., Fourichon, C., & Beaudeau, F. (2003) Production effects related to mastitis and mastitis economics in dairy cattle herds. *Veterinary Research* **34**, 475–491.

Sordillo, L.M., Contreras, G.A., & Aitken, S.L. (2009) Metabolic factors affecting the inflammatory response of periparturient dairy cows. *Animal Health Research Reviews* **10**, 53–63.

Terregino, C., Toffan, A., Beato, M.S., et al. (2008) Pathogenicity of a QX strain of infectious bronchitis virus in specific pathogen free and commercial broiler chickens, and evaluation of protection induced by a vaccination programme based on the Ma5 and 4/91 serotypes. *Avian Patholpgy* **37**, 487–493.

Viguier, C., Arora, S., Gilmartin, N., et al. (2009) Mastitis detection: current trends and future perspectives. *Trends in Biotechnology* **27**, 486–493.

Weckwerth, W. (2010) Metabolomics: an integral technique in systems biology. *Bioanalysis* **2**, 829–836.

Weston, A.D. & Hood, L. (2004) Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *Journal of Proteome Research* **3**, 179–196.

Witkamp, R.F. (2005) Genomics and systems biology–how relevant are the developments to veterinary pharmacology, toxicology and therapeutics? *Journal of Veterinary Pharmacology and Therapeutics* **28**, 235–245.

Zadoks, R.N. & Schukken, Y.H. (2006) Use of molecular epidemiology in veterinary practice. *Veterinary Clinics of North America: Food Animal Practice* **22**, 229–261.

Chapter 5
Systems Biology of Host–Food–Microbe Interactions in the Mammalian Gut

Vitor A.P. Martins dos Santos, Michael Müller, Willem M. de Vos, Jerry Wells, Marinus F.W. te Pas, Guido Hooiveld, Peter van Baarlen, Mari. A. Smits, and Jaap Keijer

The Gastrointestinal Tract and Body Homeostasis

The gastrointestinal tract (GI-tract) is the major site where food and feed meets the body, is converted into nutrients and metabolites that serve as fuel, components or signaling molecules for our cells, and is scrutinized for the presence of toxic or pathogenic components. It contains the largest repertoire of immune functions, it has intense metabolic activity, and is colonized since birth by microbes (collectively called microbiota) that have developed intimate relations with the host in many different ways (Bäckhed et al., 2005; O'Hara and Shanahan, 2006; Zoetendal et al., 2008; Camp et al., 2009). The intestinal microbiota consists for a large part of bacteria that belong to the phyla of *Firmicutes*, *Bacteroides*, *Actinobacteria*, *Proteobacteria*, and *Verrucomicrobia*. These contribute to nutrient processing and signaling, and produce metabolites with essential functions, such as vitamins and short-chain fatty acids (SCFAs) (Zoetendal et al., 2008).

The GI-tract not only serves as the entry point for nutrients and energy in the body, but is also an important metabolic and endocrine organ as such, being a target for balancing the body energy status. The metabolic capacity of the intestine affects whole-body physiology, directly, by altering its energy use, but also indirectly, by secreting key metabolic (endocrine) hormones (such as glucagon-like peptide-1 (GLP-1), gastric inhibitory polypeptide (GIP), cholecystokinin (CCK), and oxyntomodulin). Hormones that regulate long-term energy balance (leptin, adiponectin, insulin) also affect the intestine. Malfunctions in this system can cause obesity and a cluster of associated metabolic disorders, including insulin resistance, type-2 diabetes, endothelial dysfunction, complex dyslipidemia, hypertension, and atherosclerosis. A causal link to low-grade inflammation has been well established, but the mechanisms by which high-fat diets and high energy feeding promote these disorders are not fully understood. Furthermore, growing evidence suggests that the gut microbiota plays a key role in the

Systems Biology and Livestock Science, First Edition. Edited by Marinus F.W. te Pas, Henri Woelders, and André Bannink.

development of metabolic disease. The microbiota ferment dietary compounds not digested by the host thereby contributing to our energy harvest and provide lipogenic substrates leading to modulation of lipid metabolism and fat deposition. The ability of the intestine for energy production and energy signaling is affected by the body energy status and can be modulated by specific nutrients. Quantitative modeling of these processes will provide fundamental physiological understanding of the regulation of intestinal energy metabolism in the context of the whole body, and will be key to the development of nutritional strategies to improve metabolic fitness and resistance to disease.

The Mammalian Gut as Gatekeeper of Homeostasis

The mucosal tissues along the intestine are responsible for sensing luminal contents. Moreover, intestinal cells secrete signaling molecules, such as gut hormones and pro- and anti-inflammatory chemokines and cytokines, to which the liver, pancreas, muscle, adipose tissue, and the immune system respond by modulating their functionality to maintain homeostatic control. Compromised functionality of the intestine has been linked to various diseases such as obesity, cardiovascular and inflammatory intestinal disorders, metabolic syndrome, diabetes, septic shock, infections, and brain diseases (Ley et al., 2005; Dumas et al., 2006; Cani et al., 2008; Membrez et al., 2008; Hildebrandt et al., 2009; Tilg et al., 2009; Wall et al., 2009; Vijay-Kumar et al., 2010). This highlights importance of the interplay between food, microbes, and host, and the role of the gatekeeper of body homeostasis, as depicted in Figure 5.1.

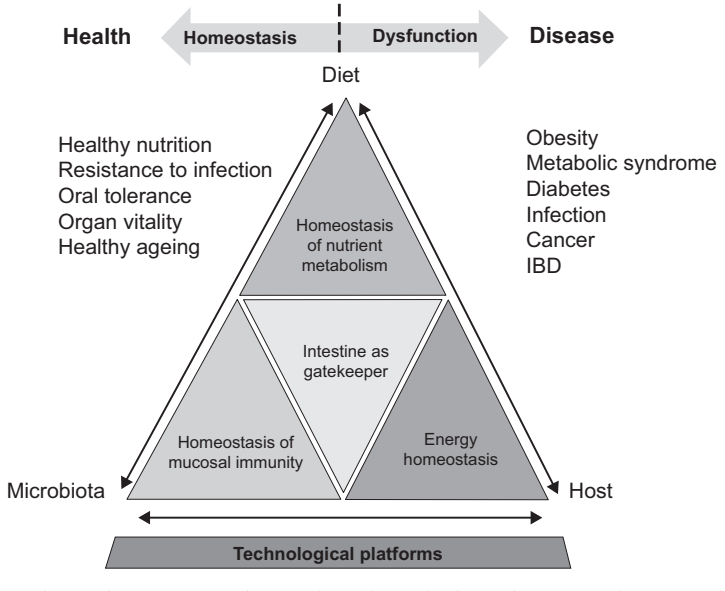


Figure 5.1 Schematic representation of the role of the intestine as gatekeeper of homeostasis and in preventing dysfunction (disease).

Although the relations above described between the intestine and chronic diseases refer to humans, they are equally applicable to pets, horses, and zoo animals. A proper functioning GI-tract is equally important for the health of production animals and directly relates to production efficiency. Several different disturbances of the gut usually occur during the lifetime of food-producing animals, especially during stressful periods like weaning. Intestinal disturbances may be the result of the presence and outgrowth of pathogens, the presence of contaminants, unbalanced nutrition, and an imbalance in the composition of microbiota, e.g., by the frequent use of antibiotics. In an unbalanced state of the gut, processing, conversion, and absorption of nutrients are negatively affected, resulting in decreased feed conversions and growth rates and higher susceptibilities to infections with residing or new pathogens. Intestinal disturbances also affect the immunological development and responsiveness of the gut, thereby further promoting the animal's susceptibility to infections with pathogens. Likewise, large amounts of antibiotics are frequently supplied to farm animals, especially during or immediately after stressful transition periods. However, human health issues urge for a rapid and significant decline in the use of antibiotics by the animal production sector. Therefore, a major challenge is to develop animal production systems that do not depend on the large-scale use of antibiotics without compromising health parameters while simultaneously improving production efficiency and reducing its ecological footprint. One important way to accomplish this is to unravel the mechanisms involved in host–feed–microbe/pathogen interactions in the GI-tract and to fully exploit the intrinsic biological potential of these interactions. Since it is known that there is a significant variation in intestinal functionality and health between livestock animals in and between breeds, there is much to gain in this respect.

The Need for Systems Approaches to Study Diet–Host–Microbiota Interactions

To exploit the ever-increasing wealth of information in an intelligent way, mathematical models need to be constructed to predict how the host, food, and microbes react to one another and, collectively, determine the gut functions and body homeostasis. This requires the application of Systems Biology approaches to the intestinal tract. While this is a challenging task, it is a timely moment to consider these approaches not only to capitalize on the wealth of the accumulating data but also to generate hypotheses and experimental approaches that lead to a better understanding of the gut function. Hence, here we discuss the basic elements for initiating such a (post-) genomics-based Systems Biology approach with specific focus on the spatiotemporal processes along the intestinal tract, the effect of diet, and the contribution of interacting microbes. We argue that such a mathematical framework should cross-link, top-down, and bottom-up modeling approaches as exemplified in Figure 5.2 below. For the animal sciences such models would be valuable tools for simulation purposes in order to optimize and "customize" animal feeds, to improve health and efficiency traits of the gut by genetic selection, and for the development of prevention and/or intervention schemes.

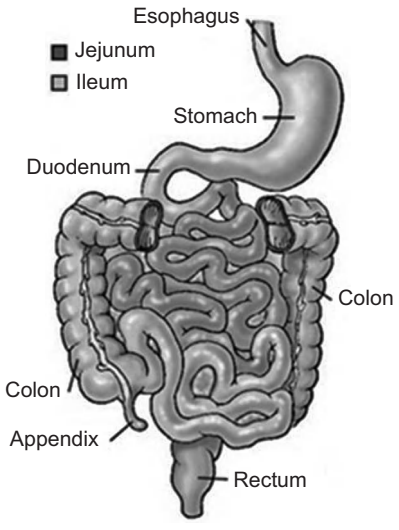


Figure 5.2 Anatomy of the digestive tract showing the relationship of the small bowel (stomach, duodenum, jejunum, and ileum) to the stomach and colon. (Adapted from http://www.yoursurgery.com/ProcedureDetails.cfm?BR=1&Proc=49.)

The Gastrointestinal (GI) Tract

The mammalian GI-tract is differentiated into various anatomical regions with different dimensions that are linked to digestive processing of the food (Figure 5.2). The functions of the upper part (stomach, duodenum, jejunum) are mainly digestion of carbohydrates, proteins, and fat; acid secretion; and absorption of monosaccharides, fatty acids (FAs), cholesterol, amino acids, di- and tripeptides, vitamins, and minerals. In the rumen of ruminants, digestion and fermentation processes (biohydrogenation of FAs) occur that play a central role in the conversion of forages and fibers into valuable animal products, like milk FAs but also in the emission of green house gases.

The longest in size is the ileum, which has also has the largest surface allowing primary uptake of nutrients and other food components. The colon in contrast is relatively short, but the food components that have escaped earlier digestion are exposed for a considerable amount of time to a large consortium of anaerobic microbes allowing for final digestion. The function of this lower part of the intestinal tract (cecum and colon) is mainly absorption of water, electrolytes, and SCFAs produced by the microbes. The majority population of bacteria in the lower part is anaerobic and the amount of bacteria is much higher compared to the upper part of the GI-tract (Table 5.1).

While at all sites there is a gradient of enzymes, pH and oxygen, and microbes, they share a common architecture consisting of a single layer of epithelial intestinal cells that are producing large amounts of mucus, heavily glycosylated proteins

Table 5.1 Features of the various anatomic sites of the average gastrointestinal tract in humans indicating the different food transit times and progressive microbial colonization.

Intestinal site	Length (cm)	Transit (h)	Density microbiota (cells/g)
Stomach	12	2–6	10^0–10^4
Duodenum	25	3–5	10^4–10^5
Jejunum	160	3–5	10^5–10^7
Ileum	215	3–5	10^7–10^8
Cecum	6	10–20	10^{10}–10^{11}
Colon	130	10–20	10^{10}–10^{11}
Rectum	18	1	10^{10}–10^{11}

(Derrien et al., 2010). This constitutes the interface that separates the intestinal food components and microbiota from the body.

Energy Homeostasis

Both the handling of food and maintenance of the barrier function require substantial energy. This is exemplified by the fact that the intestine uses 5–10% of body energy in the postprandial/fasting state (Rolfe and Brown, 1997), while its energy use increases to 15–30% in the prandial/fed state. Clearly, this requires major adaptive responses that have to take place in an organized manner. At present, no models, let alone quantitative models, are available for adaptive changes in intestinal energy metabolism. A sufficient ATP-generating capacity is essential for uptake and secretion, but also to maintain its barrier function. Indeed, both synthetic and natural compounds, including the cytokine TNF-alpha, which can uncouple mitochondrial electron transfer from ATP production, increase intestinal permeability (Somasundaram et al., 2000; Baregamian et al., 2009).

Signaling and Hormone Homeostasis

Obviously, the intestine does not operate in isolation in the body. It is a major endocrine organ that produces a large number of hormones, including CCK, peptide YY (PYY), and glucose-dependent insulinotropic peptide (GIP) and the preproglucagon-derived peptides such as Glucagon like peptides 1 and 2 (GLP-1, GLP-2) and oxyntomodulin. GLP-1 and GIP are incretins that evoke an insulin response. These hormones have a role in signaling energy use, energy demand, and energy uptake in the body and as such (or directly) have a role in hunger and satiety. In return, the intestine responds to endocrine hormones that are involved in signaling of the body energy status, including leptin, adiponectin, and insulin. Profound quantitative understanding of the relationship between intestinal energy metabolism, body energy status, and endocrine signaling is essential to develop (nutritional) strategies, not only to improve intestinal function but also to prevent metabolic complications arising from dysfunctional energy

homeostasis. This is especially true for production animals that encounter periods of negative energy balance, for example, milk cows at the start of lactation.

Homeostasis of Tolerance and Immunity

Epithelial cells form a single cell layer that is in continuous contact with the intestinal lumen content. Together with immune cells, epithelial cells regulate immune homeostasis and tolerance that ensure that antigens from the food and microbiota do not lead to inflammation. Tolerance and homeostasis are comediated by the microbiota, which can influence inflammation through modulation of NF-κB and PPAR activation. Moreover, SCFAs produced by the intestinal microbiota have recently been shown to affect inflammation via GPR43 (Oberhardt et al., 2009). Homeostasis is also linked to dietary factors, such as high fat, which can exacerbate inflammatory signaling and ultimately negatively impacts epithelial integrity under conditions of stress or mild inflammation. Immune homeostasis is tightly linked to metabolic (lipid) homeostasis through antagonistic activity of PPAR signaling and NF-κB signaling.

Our current model for the homeostasis of tolerance and immunity at mucosal surfaces highlights a regulatory role for the epithelium (Figure 5.3A). Under steady-state conditions, the epithelia produce retinoic acid, TGF-β, and TSLP, factors that "condition" immune cells toward a "tolerogenic" phenotype. Moderate stimulation of epithelial cells leads to NF-κB-mediated increased secretion of additional factors such as IL-10, BAFF, APRIL, and SLPI (Figure 5.3). These factors serve to promote and regulate immune cell activation and maturation such that homeostasis and immune tolerance are maintained.

Nutritional Challenges

The intestine is the major organ for the uptake of nutrients and other food components. Whereas much is known about the absorption and transport of nutrients and food components, only recently work has been done focusing on the regulation of intestinal genes and functional properties upon adaptive response to nutritional challenges. One of the most important challenges that has received significant attention is dietary fat, and in particular in its recently recognized interaction with the gut microbiota (Cani et al., 2008; De Wit et al., 2008; Hildebrandt et al., 2009). Below, we discuss recent aspects on the role of dietary lipids on the regulation of gene expression in the different regions of the small intestine with a particular focus on the role of nutrient-sensing transcription factor PPARα. Up to now, in most of these studies the mouse has been applied as a model.

Dietary Lipids

On a daily basis, the human small intestine metabolizes an average of 100 g of dietary fat, more than 90% of which is composed of triacylglycerols (TGs). The absorption of lipids from the lumen is generally highly efficient, whereby approximately 4% of the

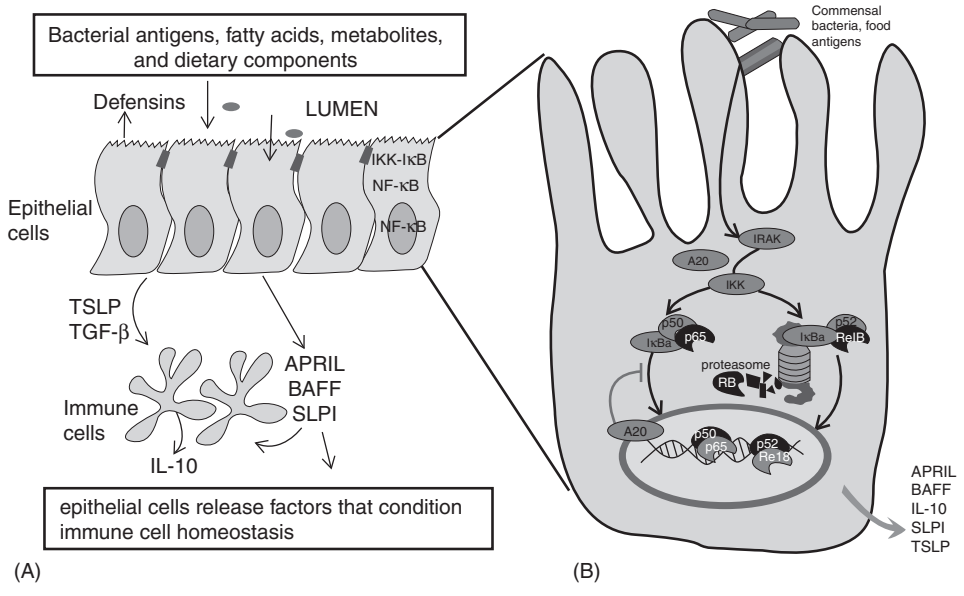


Figure 5.3 (A) Schematic depiction of the main epithelial factors that condition immune cells to maintain homeostasis at mucosal surfaces. (B) Stimulation of TLR and other receptors in epithelial cells leads to intracellular signaling, degradation of IκB, and transcriptional activity of NF-κB, including release of factors promoting immune cell maturation and PPAR-mediated conditioning of immune cells to non-inflammatory cell type.

ingested fat escapes into feces. Prior to absorption, TGs are hydrolyzed by gastric and pancreatic lipases to free FAs and 2-monoacylglycerols (2MAGs), both of which are taken up by enterocytes.

Because both products are potentially toxic—especially free FAs at higher concentrations—they must be rapidly neutralized. This can be accomplished through multiple mechanisms. The dietary fat is metabolized or converted to TGs. Once taken up by cells in the enterocyte they have diverse metabolic and cellular fates. The incoming FAs are activated and may be sequestered by binding to FA-binding proteins (FABPs), which also regulate intracellular trafficking of FAs. The absorbed FAs and 2MAGs are quickly resynthesized into TG, which temporarily resides in intracellular lipid droplets and is ultimately transported out of the cell as TG-rich lipoprotein particles called chylomicrons. Because the intestine expresses many enzymes that control mitochondrial, peroxisomal, and microsomal oxidation, FAs can also be catabolized. The amount of TGs that are absorbed by the intestine depends on several physiological and nutritional factors, including the amount and composition of TGs and phosphatidylcholine (PC) in the intestinal lumen. However, the molecular mechanisms that underlie these phenomena are not fully understood.

FAs play an important role as signaling molecules in the regulation of their own metabolism. The transcription factor PPARα, a member of the superfamily of nuclear receptors, plays an essential role in this regulation. PPARα (NR1C1) is a member of the superfamily of nuclear receptors and is closely related to the PPAR isoforms β/δ

(NR1C2) and γ (NR1C3). Bunger et al. (2007) demonstrated that PPARα is highly expressed in small-intestine cells, coinciding with the primary anatomical location where FAs are digested, absorbed, and transported into the body as chylomicrons. Moreover, this genome-wide analysis and preliminary data (Hooiveld et al., unpublished) suggest that all major steps in enterocyte lipid handling are regulated by PPARα, pointing to a role of this nuclear receptor as a major regulator of intestinal lipid absorption.

The capacity for fat absorption in the intestine can be adapted to the fat content of the diet in the mouse, and this adaptive capacity is apparently much more pronounced for lipids than for other nutrients. Fat-mediated adaptation takes place through two complementary events (Niot et al., 2009). A switch from a low fat and high-carbohydrate diet to a high-fat and low-carbohydrate diet results in the induction of intestinal cell proliferation with the apparent goal to increase the absorptive area. Furthermore, chronic high-fat diet results in the coordinate regulation of a large array of genes many of them contributing to an increased metabolic capacity related to lipid homeostasis. To gain insights into the potential role of the small intestine in development of obesity, De Wit et al. (2008) and Bunger et al. (unpublished) studied dietary fat-induced differential gene expression along the longitudinal axis of small intestines of C57BL/6J mice (Figure 5.4).

Dietary fat (45 energy%) had the most pronounced effect on differential gene expression in the middle part of the small intestine mainly on biological processes related to lipid metabolism, cell cycle, and inflammation. In-depth network analysis—see modeling section below for the methods—revealed that the lipid-sensing nuclear receptors such as peroxisome proliferator-activated receptors (PPARs), liver X receptors (LXRs), and farnesoid X receptor (FXR) play an important regulatory role in the response of the small intestine to high-fat diet. A hypothesis derived from this is that dietary fat-induced development of obesity and insulin resistance in mice with substantial changes in gene expression in the small intestine over time are causally linked.

Short-Chain Fatty Acids (SCFA) and Microbiota

SCFA, along with middle- and long-chain FAs, are derived from dietary lipids and have important nutritional implications. They are major products from fermentation processes from gut microbiota of nondigestible carbohydrates. SCFAs consist mainly of butyrate, acetate, and propionate, which are almost completely absorbed along the GI-tract and in the rumen of cattle. Many metabolic properties are shared between microbial taxa and hence it is difficult to link the capacity of producing specific SCFAs to phylogenetic information (Zoetendal et al., 2008). However, there is great interest in doing so following the hallmark observation that the ratio of bacteria belonging to the *Bacteroidetes* and *Firmicutes* changes in obese subjects that are losing weight (Ley et al., 2006). This and following studies have shifted the attention to the impact of the intestinal microbiota on nutrient processing and have been reviewed recently (Ley, 2010; Vrieze et al., 2010). Bacteria of the *Bacteroidetes* phylum produce high levels of acetate and propionate, whereas several members of the *Firmicutes* phylum produce high amounts of butyrate (Maslowski et al., 2009). Normal colonic epithelia derive 60–70% of their energy supply from SCFAs, particularly butyrate. Propionate is largely

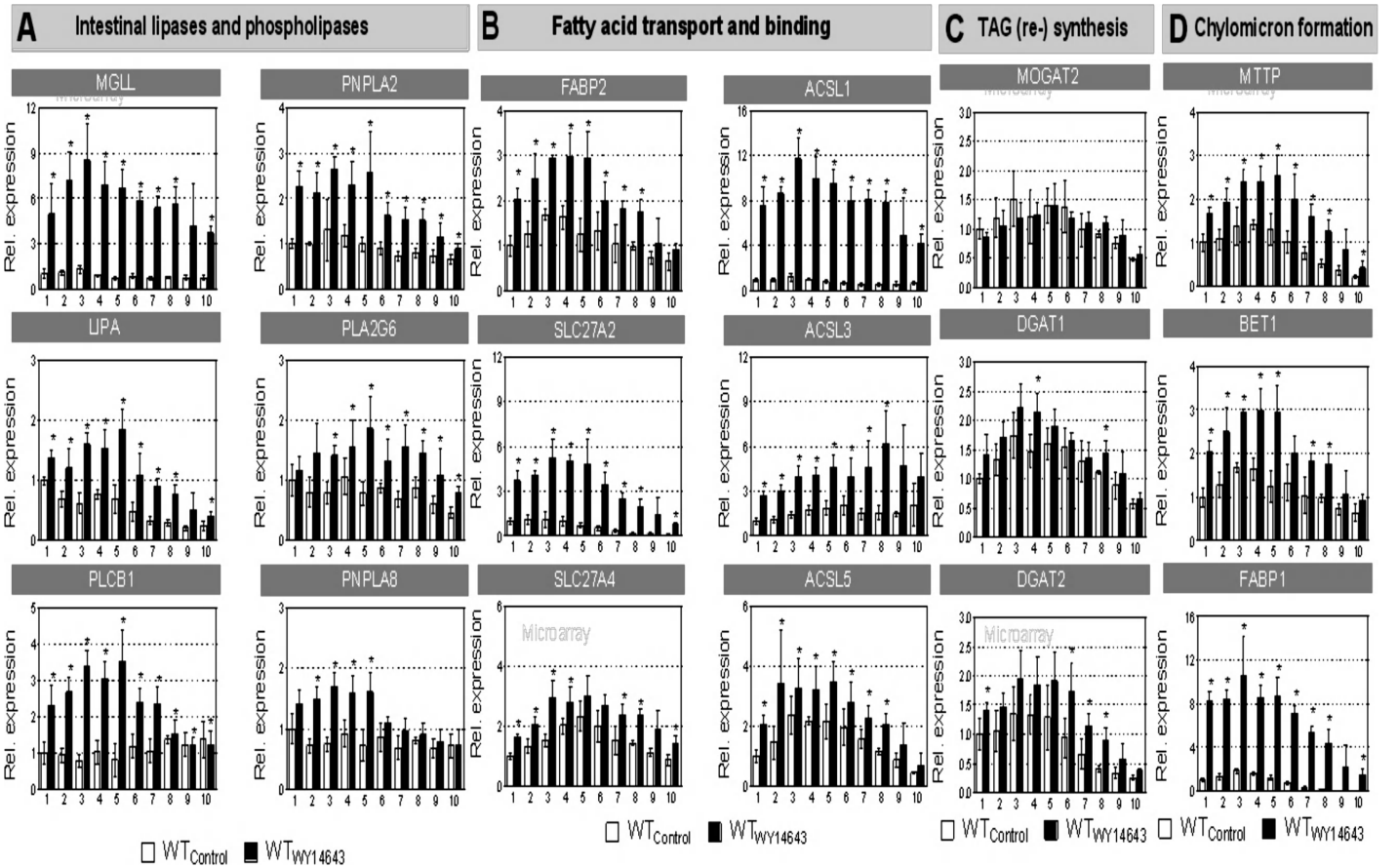


Figure 5.4 Longitudinal distribution (from C1–C10) and compensatory regulation of several genes involved in crucial steps of dietary lipid metabolism in the small intestine in wild-type mice on PPARα activation by the specific ligand WY14643 (white columns, control; black columns, ligand feeding).

taken up by the liver and is a good precursor for gluconeogenesis, liponeogenesis, and protein synthesis. Acetate enters the peripheral circulation to be metabolized by peripheral tissues and is a substrate for lipid and cholesterol synthesis (Schwiertz et al., 2010).

SCFA can affect the intestinal barrier function. Dietary intervention in rats with a diet high in fructo-oligosaccharides (FOS), which produces high levels of SCFA in cecum and colon, resulted in increased permeability to pathogenic bacteria. Transcriptome analysis showed a concomitant upregulation of intestinal peptides with a role in satiety signaling, such as proglucagon, CCK, peptide YY (PYY), and pancreatic polypeptide (PPY), as well as up-regulation of mitochondrial gene expression (Rodenburg et al., 2008a). These results were explained by exhaustion of mitochondrial capacity for ATP production, which was used to maximal capacity for maintenance of cellular pH. In more detail, high levels of FOS fermentation products cause intracellular acidification of epithelial cells, when protonated SCFA diffuse from the gut lumen into epithelial cells. These SCFA cause intracellular acidification and induce proton pump activity (NHE and NBC transporters), which may lead to ATP depletion. Reduced ATP levels are compensated by increased mitochondrial gene expression. The consequential disturbed energy metabolism causes increased permeability. These results show the importance of homeostasis of intestinal epithelial energy metabolism for maintenance of the intestinal barrier function. Indeed, it has been shown that uncoupling of intestinal mitochondria leads to increased bacterial translocation, immune cell infiltration, and ulceration in rats (Somasundaram et al., 1997; Nazli et al., 2004). These data further raise the question how the very high production of SCFA is dealt with in ruminants at a mechanistic level.

SCFA and Energy Metabolism

Also, other interventions can change SCFA production. In a recent human study (Duncan et al., 2007), it was discovered that when carbohydrate intake was lower the acetate production increased, the butyrate production decreased, and that of propionate did not change. Subsequently, leaner people had a higher ratio of acetate to butyrate and propionate. This respective ratio was also higher, though only slightly, in lean volunteers (Schwiertz et al., 2010). Here, the *Bacteroides* species, the known propionate producers, are also higher in the overweight volunteers. The colonic mucosa also draws on butyrate as energy source. In addition, butyrate functions as histone deacetylation inhibitor and has significant effects on chromatin structure and increased levels will have effects on increased DNA transcription, including potentially increased release of gut hormones. From the SCFA produced in the gut, butyrate will have the most significant local effects on functional properties of the gut.

SCFA and Signaling

Beside their role as fuels or metabolic precursor, SCFAs can also serve as ligands for G-protein-coupled receptors (GPCRs): GPR43 and GPR41. The GPR41 is activated

equally by propionate or butyrate, whereas GPR43 prefers propionate to other SCFAs (Hirasawa et al., 2008). GPR41 is associated with the excretion of the incretin PYY. Samuel et al. (2008) found that GPR41 is expressed in enteroendocrine cells, mainly in the distal small intestine (ileum) and colon. They found that a deficiency of GPR41 is associated with reduced expression of PYY, increased intestinal transit rate, and reduced extraction of energy from SCFAs that are made by the microbiota. These results indicate that GPR41 regulates host energy balance through mechanisms that are dependent upon the gut microbiota (Ichimura et al., 2009). GPR41 mediates a key microbial–host communication circuit via the SCFAs produced by microbiota. One feature of this interaction is that activating GPR41 increases the amount of circulating enteroendocrine hormones such as PYY, which reduces gut motility and thus increases production/absorption of SCFAs. Analysis of the deficient mice suggests that inhibition of GPR41 could mean that less energy is extracted from the indigestible fibers and thus could contribute to weight loss (Samuel et al., 2008).

The Intestinal Microbiota

As a consequence of developments in high-throughput, functional genomics, and metagenomics-based approaches, recent years have seen a renaissance in the interest in the microbial diversity and activity in the intestinal tract (Zoetendal et al., 2008). Since the vast majority of the intestinal microbes have not yet been cultured (Rajilić-Stojanović et al., 2007), recent progress in determining the microbial diversity in the intestinal tract has largely been based on the application of 16S rRNA-based methods that include high-throughput methods, such as next-generation sequencing and microarray analysis or both (Zoetendal et al., 2008; Claesson et al., 2009). An ultimate exponent is the recent study providing the sequence of 3 million genes that provide a baseline intestinal coding capacity of the intestinal microbes (Qin et al., 2010). These and earlier lower throughput methods have shown that the intestinal colonization process starts immediately after birth resulting in adult-like complexity developing after the weaning process (Palmer et al., 2007). The intestine microbial community in adult animals is high individual and consists of over 1000 microbial phylotypes (Rajilić-Stojanović et al., 2007; Zoetendal et al., 2008). This indicates that the intestinal microbiota may be an important factor contributing to the subject's health status. Within our life span, the average composition of the adult communities appears to be rather similar and only subjects with an age above 100 years are characterized by a significantly different microbiota, possibly to their changed immune system (Biagi et al., 2010). Moreover, the adult community is remarkably stable and resilient, even against major perturbations such as the use of antibiotics (Dethlefsen et al., 2008). Long-term studies have even shown significant stability in a period over 10 years and this led to the concept of an individual core of microbes that reflect specific interactions with the host (Zoetendal et al., 2008). This individual core differs from the common core of microbes that include the phylotypes that are shared between individuals as discussed below. This indicates that the composition of gut microbiota can be regarded as a complex polygenic trait that can be modulated by specific environmental factors as well as by specific host genetic factors (Bensona et al., 2010). Comparisons of microbiota composition between humans,

chickens, cows, pigs, and geese have also been described recently (Lee et al., 2010). Moreover, a wide array of studies in germ-free and other model animal systems has been providing considerable insight in the impact of microbes on the host (Hooper et al., 2010). Finally, host–microbe interactions are now being studied in human and livestock systems with both attentions for the host as well the microbe expression (Gross et al., 2008; van Baarlen et al., 2009; Marco et al., 2010). Similarly, the effects of different dietary conditions diet on intestinal gene expression have been studied extensively in various nutrigenomics studies (Bunger et al., 2007; De Vogel–van den Bosch et al., 2008; De Wit et al., 2008; Van Den Bosch et al., 2008; Mair et al., 2010). Impaired barrier function by dietary FOS in rats is accompanied by increased colonic mitochondrial gene expression (Rodenburg et al., 2007a, 2007b, 2008a). These studies have led to a framework to analyze datasets that can be applied to many species (Rodenburg et al., 2008b). Together, enormous amounts of the datasets of different types are being continuously generated that need to be considered in a higher context framework.

Defining Scales in Microbial Ecology of the Intestinal Tract

The GI-tract, as any ecosystem, exists along temporal, spatial, and environmental scales; these scales define three conceptual dimensions within which the GI ecosystem exists. However, any single analysis of an ecosystem is performed using a limited range of scales: "a low-dimensional slice through a high-dimensional cake" (Zengler et al., 2002; Camp et al., 2009). This restricted range of scale is imposed by our experimental design as well as limitations in our perceptual capabilities. This is important because different scales might be subject to different selective processes. Fine temporal and spatial scales can generally provide greater detail yet be more susceptible to stochastic events, whereas coarser scales can be more regular and predictable. Therefore, the biases associated with each scale must be recognized and an understanding of the interaction among phenomena on different scales has to be developed. Figure 5.5 depicts schematically (after Camp et al., 2009) how temporal and spatial scales apply to the GI ecosystem for three mammalian "levels."

Microbiota and Systems Approaches

The enormous complexity and individuality of the intestinal microbiota complicates the application of Systems Biology approaches. Hence, it is of interest to determine the microbial phylotypes that are shared by different individuals. This so-called common core has attracted considerable attention and recent studies on the adult core microbiota in humans revealed a varying number of phylotypes (Rajilić-Stojanović et al., 2009; Tap et al., 2009; Turnbaugh et al., 2009). This can be attributed to the absence of a common definition, the application of different methodologies, and the analysis of subjects with a widely different genetic background. Hence, standardized approaches are required to further define the core microbiota. This was initiated in a recent meta-analysis of the intestinal microbiota of over 1000 subjects that were all characterized in the same way by deep analysis with a phylogenetic an intestinal

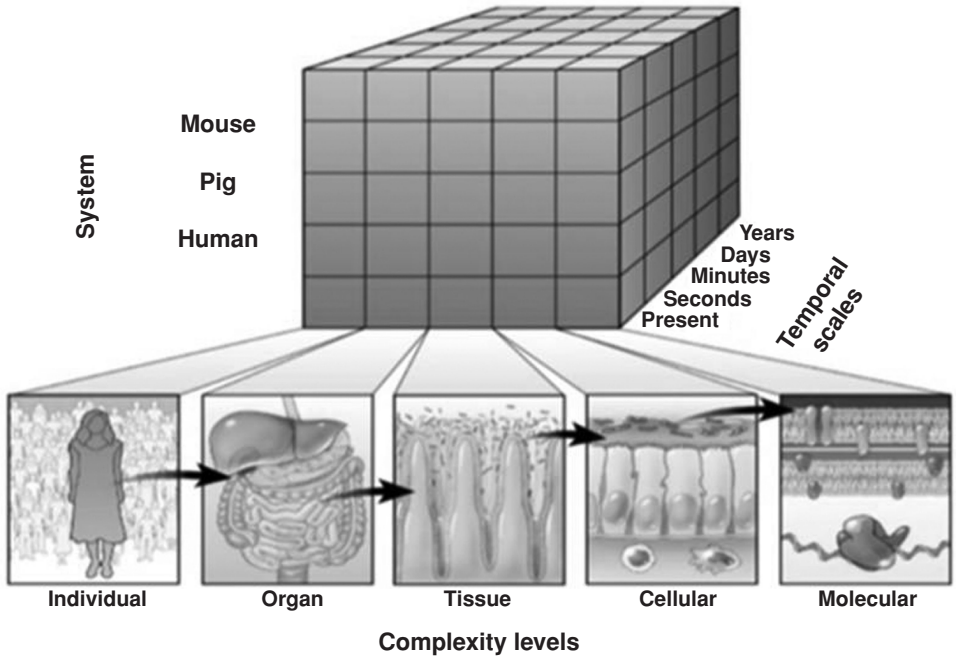


Figure 5.5 Schematic representation of the embedding of the gastrointestinal tract (GI-tract) and the levels and complexity handled for different mammalian systems. The y-axis defines the increasing complexity of the model system, whereas the x-axis defines various complexity levels at which the GI ecosystem is perceived. The upper macroscopic level consists of the individual host and progresses down through the levels of organ system, tissue, epithelial cell, microbial cell, and molecule. Temporal scales (z-axis) are defined by the time over which variation in the GI ecosystem is perceived, beginning with the present and progressing into seconds, minutes, days, and years. (Adapted from Camp et al., 2009.)

microarray (Nikkila and de Vos, 2010). Similar studies are underway for major livestock species.

Another factor complicating the application of systems approaches is the fact that the vast majority of studies of the intestinal microbiota draw on fecal samples. While these may reflect the overall microbiota, the colonic microbes are the most numerically dominant (see Table 5.1). This prevents the analysis of the upper intestinal microbes, and these can only be detected by ultra deep analysis methods assuming that they reach the colon in an intact form. Other avenues to study the ileum intestinal microbiota have recently been developed and rely on direct sampling or analysis of ileostoma subjects (Hartman et al., 2009; Booijink et al., 2010).

A last aspect that affects the use of systems approaches is the absence of genomic data for all intestinal microbes. In this regard, the Human Microbiome Project has made substantial progress. Recently, the draft genome sequences of 178 genomes of partly related microbial species were determined, revealing over 30,000 unique genes (Nelson et al., 2010). However, the majority of the intestinal microbiota has not yet been cultured and the number of new phylotypes grows faster than the cultured

intestinal species (Rajilić-Stojanović et al., 2007). Hence, much progress can be expected from metagenomic approaches and from the thorough analyses of the results thereof. Following the first metagenomic sequences of 3 US and 12 Japanese subjects, now the metagenomes of 121 EU subjects have been reported by the MetaHit consortium (Gill et al., 2006; Kurokawa et al., 2007; Qin et al., 2010, respectively). This has led to an inventory of over 3,000,000 unique genes that serve as a baseline for future studies. In the latter study, use was made of new, generation sequencing technologies and with the advent of even higher throughput machines significant progress is to be expected, even if the complete metagenome exceeds the nine million genes presently estimated (Xie et al., 2009). The major bottleneck is now how to use this large amount of data and provide predictive models based on relevant sets of microbial genes.

Host–Microbiota Interactions

A variety of interactions with the host have been described. However, most have been derived from gnotobiotic animal model studies. A high level of sophistication has been reached since the first pioneering studies with *Bacteroides thetaiotamicron* colonizing germ-free mice (Hooper et al., 2010). Recent reports describe the use of knockout mice, microbial communities of varying complexity, and global approaches that not only addressed the host but also the microbiota response (Wen et al., 2008; Mahowald et al., 2009; Rey et al., 2010). Moreover, various labs are involved in the design and construction of germ-free mice that stably harbor microbial communities exclusively consisting of relevant human species. These could be of great interest as initiating points for model studies aiming at systems approaches. However, it remains to be seen whether these interactions are stable and representative, notably in view of the observation that the human microbiota is highly subject-specific. Moreover, this approach relies on the cultivation of human microbes in gnotobiotic animals. Finally, the number of combinations of microbes increases exponentially with the amount of species and hence may provide only information on a small part of the microbiota.

An alternative approach is the use of human and livestock species as models to test the effect of added microbiota. This has now become technically feasible and the first reports have described the effect of orally administered *Lactobacillus plantarum* on the gene expression in the human duodenum (Troost et al., 2008; van Baarlen et al., 2010) and the gene expression in duodenum, ileum, and colon of pigs (Gross et al., 2008). This has revealed a large impact on the immune signaling network providing support for a probiotic function. This is to be expected in view of the high oral dose (over 10^9 cells per mL) and the low degree of colonization in the human duodenum (see Table 5.1). In the pig study, gene expression was most affected in the ileum. In a recent comparative study with three different *Lactobacillus* species, it was found that the host gene expression is determined by the specific strain that is applied (van Baarlen et al., 2010). Obviously, this approach with human volunteers can be used only in a limited number of cases, in livestock species these limitations are less. It illustrates the power and potential of performing studies in humans and livestock animals. Hence, holistic studies with changing microbiota or diet can be designed where the host gene expression is addressed and can be analyzed with systems tools designed for that purpose. A major challenge is now to integrate, manage, and interpret this wealth of

generated information so to translate it to testable hypotheses and insights that can contribute to the understanding of gut function.

Integrated Modeling Approaches

Understanding that microbiota is pivotal in our intestinal function has revolutionized our insights into the intestinal tract. The triangle of food/feed–host–microbes (Figure 5.1) has become the pinnacle of advanced global initiatives aiming at characterizing the role of the intestine as pivot of body homeostasis. Currently, the generation of "omics" data is no longer the main bottleneck toward the understanding of gut functions. What is urgently needed are systems approaches to position the accumulating data on the interplay between food/feed–microbiota–intestine with spatial, temporal, and environmental dimensions into a conceptual framework, provide mechanistic models, and allow predictions. Figure 5.6 frames the various modeling approaches herein discussed.

Top-Down Modeling

Multivariate Statistics

The generation of increasingly more—omics and other—data in wider sets of experiments over many conditions has enabled the application of robust statistical analysis and modeling procedures to—possibly—ascertain causality and links between different factors. In a recent study, Martin et al. (2008) used multivariate statistical analysis to assess the correlations between changes in host parameters, bacterial populations,

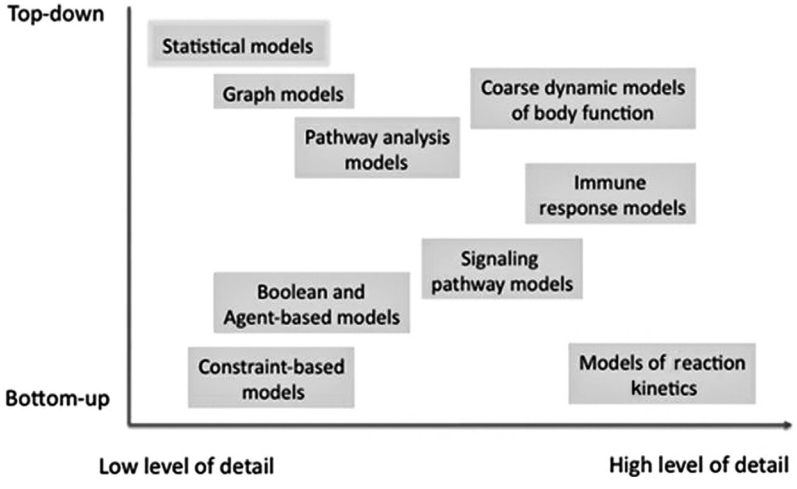


Figure 5.6 Simplified schematic representation of the various modeling approaches herein described. These approaches run for different levels of biological organization and time scales. These would be then a third and fourth dimensions to be taken into account.

and in the fecal composition of SCFAs in a humanized microbiome mouse model. In that same study, the authors represented the intercompartment metabolic correlation through pixel maps and used bipartite graphs to display the correlation matrix derived from focal SCFAs and respective microbial profiles to assess the probiotic-induced changes in the microbial metabolism of the gut. Jointly, the analyses indicated that a probiotic formulation modulated a range of host metabolic pathways associated to fermentation of carbohydrates by different bacterial strains in the gut. Similar methods were used to describe the host–metabolite–microbiome associations in obese rodents, enabling the authors to conclude that both lean and obese animals could have specific metabolic phenotypes linked to their individual microbiomes (Waldram et al., 2009). Multivariate statistical modeling of the spectra from a series of metabolomic experiments revealed the pivotal contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice (Dumas et al., 2006), whereas principal components analysis (PCA) and partial least squares discriminant analysis (PLS-DA) profiling of urinary metabolites, which are influenced by gut bacteria, enabled the discrimination of cohorts of Crohn's disease (CD), ulcerative colitis (UC), and control patients, a distinction that is of pivotal importance for both management and prognostic reasons (Williams et al., 2009. Similar methods have been employed for computational meta-analysis of a collection of independent studies, including over 1000 phylogenetic array datasets, as a means to characterize the variability of human intestinal microbiota (Nikkila and de Vos, 2010). This type of analysis is an important step toward the ability to describe the extent and type of variability of the microbiota in the human gut, and establish relations between microbial taxa and their interaction with the host, intestinal location, or genotype.

Embedding Experimental Design and Data Decomposition

Multivariate statistics is generally used to determine the relationship between given types of data (e.g., omics data and a phenotype) obtained from the study subjects. In these studies, the aim of regression is not to predict the value of the phenotype but to derive reliable and validated relationships that can be studied further to select and interpret those genes, proteins, or metabolites that are most important with respect to the phenotype (Thissen et al., 2009). By combining these types of analyses with a study design to decompose the total data into data blocks that are associated with specific effects (such as ANOVA), the quality and interpretability of statistical regression models can be greatly improved by explicitly using the data structure. This was illustrated by Thissen et al. (2009) in a nutritional intervention study using Apoliprotein E3-Leiden transgenic mice to assess the relation between liver lipidomics and a plasma inflammation marker, Serum Amyloid A, where they have shown that ANOVA-PLS leads to a better statistical model that is more reliable and better interpretable compared to standard PLS analysis.

Network Analysis Methods

Network-based approaches have been used in a series of studies to characterize the microbiota abundance and temporal dynamics in diverse intestinal samples and correlate potential interactions between the identified species with respect to the healthy

and diseased states. Through topological network analysis, i.e., by computing several local and global network statistics, Naqvi et al. (2010) identified recurring patterns and motifs in gut microbiota datasets and subsequently fitted the network models to a family of well-studied graph models. This enabled to discriminate gut microbiota in alcoholic subjects and healthy subjects. In a different line, Trosvik et al. (2010) through a combination of nonlinear data modeling (regression) and simulations of the early phylum-level colonization process of the early infant gut colonization, addressed the ecological dynamics of the gut microbiota in infants. PCA of infant microbiota 16S rRNA gene microarray data showed that the main directions of variation were defined by three phylum-specific probes targeting *Bacteroides*, *Proteobacteria*, and *Firmicutes*. Nonlinear regression analysis identified several dynamic interactions between these three phyla. Simulations of the early phylum-level colonization process showed that, in general, varying the initial composition of phyla in the simulations had little bearing on the final equilibrium. Through a combination of spectroscopic, microbiomic, and multivariate statistical tools to analyze focal and urinary samples, Li et al. (2008) established structural differences in gut microbiomes of Chinese and American populations, which is consistent with population microbial and metabolic differences reported in epidemiological studies. In a thorough deep sequencing study supported by extensive bioinformatics analysis, Turnbaugh et al. (2006) have determined the organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins.

These bioinformatics and network analysis methods have been pivotal as well to characterize the (meta-) transcriptome of the human focal microbiota and that of the human and rodent host under a series of conditions. In a recent study by Booijink et al. (2010), metatranscriptomic analysis revealed subject-specific expression profiles with genes encoding proteins involved in carbohydrate metabolism being dominantly expressed. On the host side, studies such as those by Cavalieri et al. (2009), De Vogel-van den Bosch et al. (2008), and Bunger et al. (2007) generated significant new knowledge on the regulation and extension of the PPARα signaling network. A genome-wide mRNA expression analysis by Radonjic et al. (2009) of hepatic adaptation to high-fat diets revealed a switch from an inflammatory to steatotic transcriptional program.

Host–Microbe Interaction Models

Ascertaining the interactions between gut microbes and the host is crucial for the understanding of the GI-function. In a specially relevant study, van Baarlen et al. (2009), found that expression profiles of human mucosa displayed striking differences in modulation of NFκB-dependent pathways, notably after consumption of the probiotic bacterium *L. plantarum* in different growth phases. The study identified mucosal gene expression patterns and cellular pathways that correlated with the establishment of immune tolerance in healthy adults. Similar studies have been performed in pigs (Gross et al., 2008). A follow-up, comparative study with 3 different *Lactobacillus* species, revealed that the host gene expression is determined by the specific strain that is applied. Rizzetto and Cavalieri (2010) unveiled key interactions between yeast and the immune system by analyzing the transcriptional response of dendritic cells (DC) stimulated by the harmless *Saccharomyces cerevisiae* and of this

phagocytosed fungus. Pathway analyses provided valuable insights into the interactions and responses elicited by of DCs in the fungus.

Cross-Linking Top-Down and Bottom-Up Approaches

Different biological systems and scales require the application of appropriate modeling approaches. In general, top-down methods as those described above, are used to map the interactions and general structure of the systems understudy, and to pinpoint areas within the underlying networks that need to be subjected to more detailed examination to answer a particular biological question. This is then subsequently done through approaches involving more detailed models of smaller subsets. Ideally, these models generate novel or refined hypotheses that enable a redesign of the experimental setting and subsequent analysis of the outcomes. This procedure is highly iterative and conductive of an ever-more refinement model of the particular (sub)system understudy. This Systems Biology process is schematically exemplified in Figure 5.7 for the mammalian gut system.

Kinetic Modeling

Research as that described in the studies described above is crucial for the elucidation of the host–microbiota–nutrient interactions. The top-down approaches used are essential in helping to organize, structure, and interpret the wealth of information generated. Bottom-up approaches, in which the cellular and interaction networks are built from the genomic and other postgenomic information are particularly important to complement those top-down analyses in that they enable to generate specific, testable hypotheses and to make predictions of, for example, the effects of perturbations (e.g., mutations, stresses, or changed component composition) or the effect of different host genotypes on a given network. One such approach is that followed by de Graaf et al. (2009, 2010) who used stable-isotope metabolic flux analysis (MFA) to selective profile the human intestinal microbial metabolic products of carbohydrate food components and to measure the kinetics of their formation pathways, in a single experiment. Subsequent modeling of the ^{12}C contents and ^{13}C labeling kinetics allowed to determine the metabolic fluxes in the gut microbial pathways for the synthesis of lactate, formate, acetate, and butyrate. By manipulating substrate and microbiota composition in a purposeful manner, this approach enables the study of the modulation of human intestinal function by single nutrients, providing a new rational basis for achieving control of the specific gut metabolites (such as the SCFAs profile in this study). Linking such MFA and derived models for the quantitative analysis of material fluxes within the single cell as well as between different cell populations and organs, up to the whole-body level, with those of population dynamics and with the omics technologies and the analysis thereof, as described above, will provide a powerful way to explore gut function in an integrated manner and across the various levels of organization and time scales (de Graaf et al., 2009).

Kinetic models of given pathways, such as those of SCFA and lipid metabolism, or as described above, give valuable mechanistic insights into the underlying molecular interactions. These models are usually built from differential equations and solved through numerical simulation and other computational analyses such as metabolic

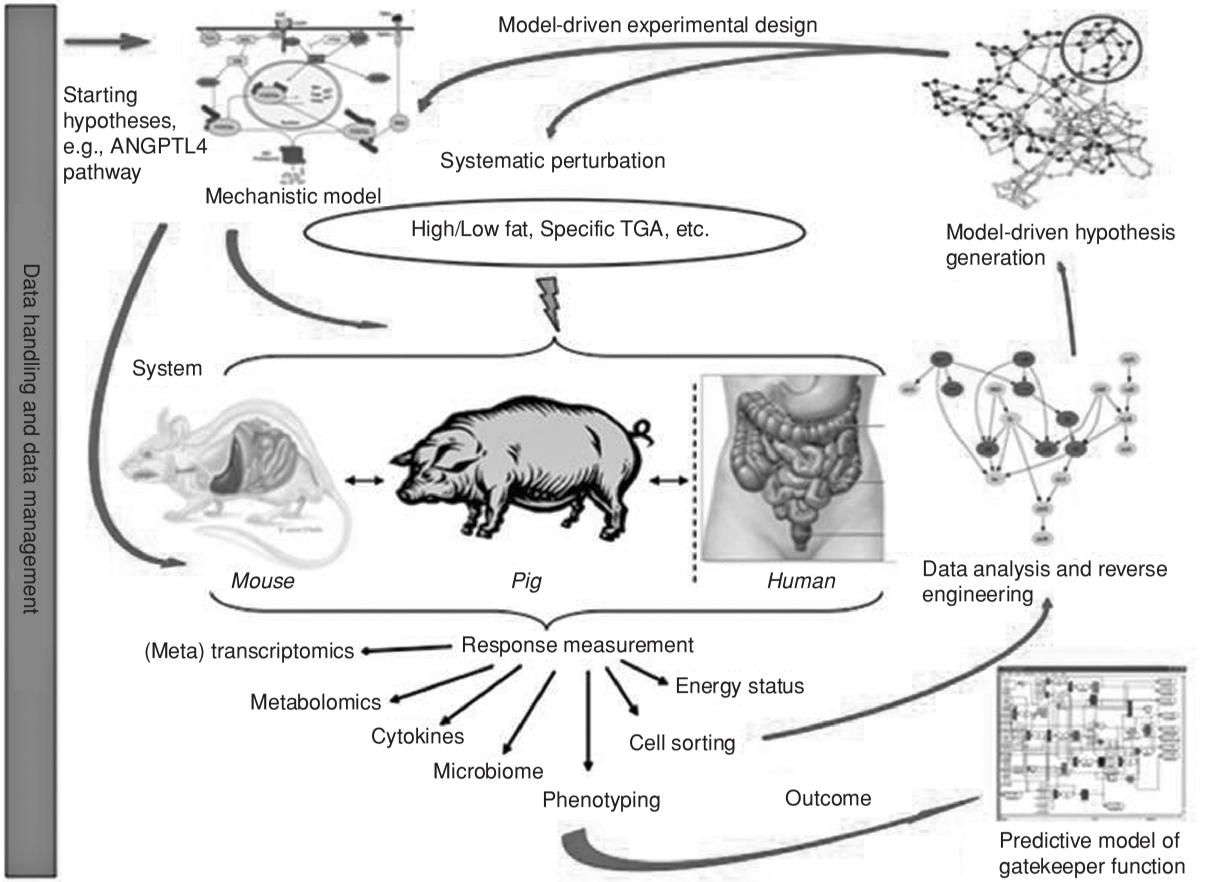


Figure 5.7 Schematic representation of an example of a general research workflow and the iterative "dry–wet" cycle centered around the cross-linking of bottom-up, model-driven hypothesis generation (roughly, the lower left half of the figure), and top-down, unbiased analysis and inference of genetic circuitry through reverse engineering from high-throughput data (lower part of the figure). The common underlying principle is to perturb the system (mice, rat, pigs, humans, or other mammalians) and analyze how it responds to the various perturbations relevant to the functioning of the intestine (and its associated microflora) as a gatekeeper. Mathematical models (graph, Boolean, or dynamic, depending on the level of knowledge available and the specific subsystem/pathway being handled) of initial hypotheses (such as the modulating function of ANGPLT4 pathway in obesity-induced high-saturated fat-linked chronic inflammation or the role of endogenous and ingested microbiota to host signaling via the NF-κB pathway) to direct experimental design of the perturbations.

control analysis, which determines how the control of flux is distributed in metabolic networks (de Graaf et al., 2009; Röling et al., 2010). A relevant, recent example is the modeling of carbohydrate degradation by human colonic microbiota, which was based on mass-balance (differential) equations to account for physiology of the intestine, metabolic reactions, and transport phenomena (Muñoz-Tamayo et al., 2010). The model was used to study various nutritional scenarios and to assess the role of the mucus on the system behavior, having provided reasonable, qualitative representation of the human colon. However, a major bottleneck of these models is that they also require detailed knowledge of the underlying molecular and physical mechanisms and of the respective model parameter values. These are in practice often difficult to obtain *in vivo* even after careful model reduction and parameterization.

Network Topology Models

Coarse-grain network topological models, in contrast, describe the interactions (edges) between molecular components (nodes) with less molecular detail than kinetic mechanistic models, but often include hundreds to thousands of components, up to the full genome scale (Puchałka et al., 2008; de Graaf 2009; Oberhardt et al., 2009; Ruppin et al., 2010). Genome-scale, stoichiometric constraint-based models, which describe the entire set of metabolic and transport reactions in an organism under the assumption of steady-state and by balancing mass and charge, have been developed for human, mice, and a series of gut-relevant bacteria, including some that are potential probiotic (Pastink et al., 2009; Teusink et al., 2009). These models, while lacking the detail of kinetic representations, provide a valuable framework with which to explore the metabolic space and capabilities of the organisms involved, to generate testable hypotheses of the relationship between genotype and phenotype, and to test the effect of external and internal perturbations (such as nutrients or other microbial species). Whereas relevant insights have been generated for single organisms, little work has been done so far in tackling the interactions of gut organisms with each other and with the mammalian host. Constraint-based modeling provides a valuable scaffold to account for interacting species and to ascertain the emergent properties arising from these interactions (Klitgord and Segrè, 2010; Röling et al., 2010). Albeit there are still considerable conceptual, technological, and computational challenges to overcome (such as the precise knowledge of the interactions, exchange and transport properties, heterogeneities, etc.), these approaches can be applied not only to a small number of interacting organisms, but, potentially, to the full gut microbiome. Owing to the various undergoing efforts worldwide in this area, much is expected in the next few years in this regard.

Boolean and Agent-Based Models

Other types of bottom-up, network models are those that attempt to map and capture regulatory interactions in a system by assuming that individual regulatory interactions are either on or off. Such so-called, Boolean approaches generate a kind of discrete network dynamics much the same way as in logical electronic circuits (de Graaf et al., 2009). Since these models require less experimental detail and knowledge of parameters than true dynamic models—as available, for example, for the epidermal growth

factor receptor (EGFR) signaling pathway (Samaga et al., 2009), or for canonical NFκB pathways (Calzone et al., 2010), discrete logic modeling has emerged as a means to link protein signaling networks with functional analysis of mammalian signal transduction complex systems such as those of the Toll-receptor signaling (Oda and Kitano, 2006), T-cell receptor (Saez-Rodriguez et al., 2009), and other signaling networks (Morris et al., 2010). These types of models enable thus also to describe, unto a reasonable extent, cross talk of different signaling pathways and are likely to become increasingly important in the future to handle immune responses in the gut to diet and gut microbiota, and to tackle the multiple interactions of this complex system.

Agent-based modeling (ABM) is a modeling technique based on the rules and interactions between the components of a system, simulating them in a "virtual world" to create an *in silico* experimental model. ABM is an approach that has been used in a number of relevant fields (e.g., to describe biofilm dynamics (Xavier et al., 2007), inflammatory cell trafficking, and epidemiological features (Ajelli et al., 2010)). Owing to its characteristics (object-oriented, rule-based, discrete-event, and discrete-time), it is in principle suited for the goal of dynamic knowledge representation and conceptual model verification. Its structure facilitates the development of aggregated modular multiscale models. Therefore, it has been proposed to use ABM as a unifying translational architecture for dynamic knowledge representation. An (2008) has presented a series of linked ABMs representing multiple levels of biological organization in the context of inflammation. An epithelial ABM derived from an *in vitro* model of gut epithelial permeability was concatenated with the endothelial/inflammatory cell ABM to produce an organ model of the gut. This model was validated against *in vivo* models of the inflammatory response of the gut to ischemia. Finally, the gut ABM was linked to a similarly constructed pulmonary ABM to simulate the gut–pulmonary axis in the pathogenesis of multiple organ failure. The behavior of this model was validated against *in vivo* and clinical observations on the cross talk between these two organ systems. Thus, albeit not mechanistic as detailed as true dynamic models, such ABM models are nevertheless useful as a navigation tool across the various levels of biological organization and hold the potential of enabling the coupling of scales.

Challenges Ahead

Some of the challenges ahead in these various approaches relate to the integration of the different omics datasets and the models derived of the analysis thereof. Recent reviews on the visualization (Gehlenborg et al., 2010) and integration of data and models (Zhang et al., 2008; Han et al., 2010; Kint et al., 2010) have addressed the major bottlenecks and possible solutions ahead in these top-down approaches. In their reviews, Raes and Bork (2008) and Röling et al. (2010) focus in particular on the necessary data types and methods that are required to unite molecular microbiology and ecology to develop an understanding of community function within their environments—including the gut—and discuss the potential shortcomings of these approaches. Despite the importance of both qualitative and quantitative models to unravel host–microbe interactions, the complexity and heterogeneity of the systems involved have so far hindered further developments. Ultimately, the goal is to develop interacting, detailed dynamic models of the gut function accounting for immune

responses to diet and microbiota, as well as nutrient and energy homeostasis. This will require both the cross-linking of top-down and bottom-up approaches and the integration of models covering a range of scales of biological organization. A way forward would be the development of multiscale models that tackle key interacting subsets of the host–microbiota–food system, as it has been done successfully for a number of systems, including type 1 diabetes through PhysioLab platform by Entelos (Shoda et al., 2010), the host–pathogen–therapy system in tuberculosis (Kirschner et al., 2010) and host–HIV–cocktail systems (Perelson, 2002). These approaches would then need to be complemented with thorough experimentally validated models of the microbiota involved such as those based on constraint-based, kinetic, and logic models above described.

Conclusions

The interplay of food, microbiota, and host is a complex mesh of intestinal-related functions that can be only understood from a systems perspective. The major challenges ahead lie in the integration of heterogeneous data and on the modeling of the many functions at various levels of biological organization. The long-term ambition in the field is to generate a solid knowledge-base and predictive mathematical frameworks on the functionality of the intestinal tract along its various spatial, temporal, and environmental dimensions. Linked to epidemiological, clinical, and comparative (post-) genomic background data, this will enable a thorough understanding of how specific host factors, nutrients, diets, and environmental conditions influence cell and organ function and how they thereby impact on health and disease. This systems knowledge will be pivotal for the development of rational intervention strategies for the prevention of human diseases such as diabetes, metabolic syndrome, obesity, and inflammatory bowel diseases. It will also be essential to fully exploit the intrinsic biological potential of host–feed–microbe interactions in livestock species, in order to optimize and "customize" animal feeds, to improve health and efficiency traits of the gut (by genetic selection), and for the development of prevention and/or intervention schemes. Moreover, it will enable generate *in silico* intestinal tract models that will advance our basic understanding and allow predicting the health effects of current and novel foods and feeds.

References

Ajelli, M., Goncalves, B., Balcan, D., et al. (2010) Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infectious Diseases* **10**, 190.

An, G. (2008) Introduction of an agent-based multi-scale modular architecture for dynamic knowledge representation of acute inflammation. *Theoretical Biology and Medical Modelling* **5**, 11.

Bäckhed, F., Leym, R.E., Sonnenburg, J.L., et al. (2005) Host-bacterial mutualism in the human intestine. *Science* **307**, 1915–1920.

Baregamian, N., Song, J., Bailey, C.E., et al. (2009) Tumor necrosis factor-α and apoptosis signal-regulating kinase 1 control reactive oxygen species release, mitochondrial autophagy,

and c-Jun N-terminal kinase/p38 phosphorylation during necrotizing enterocolitis. *Oxidative Medicine & Cellular Longevity* **2**, 297–306.

Bensona, A.K., Kellyb, S.A., Leggea, R., et al. (2010) Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proceedings of the National Academy of Sciences of the U S A* **107**, 18933–18938.

Biagi, E., Nylund, L., Candela, M., et al. (2010) Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PLoS One* **5**, e10667.

Booijink, C.C.G.M., El-Aidy, S., Rajilić-Stojanović, M., et al. (2010) High temporal and inter-individual variation detected in the human ileal microbiota. *Environmental Microbiology* **12**, 3213–3227.

Bunger, M., Van Den Bosch, H.M., Van Der Meijde, J., et al. (2007) Genome-wide analysis of PPAR{alpha} activation in murine small intestine. *Physiological Genomics* **30**, 192–204.

Calzone, L., Tournier, L., Fourquet, S., et al. (2010) Mathematical modelling of cell-fate decision in response to death receptor engagement. *PLoS Computational Biology* **6**, e1000702.

Camp, J.G., Kanther, M., Semova, I., et al. (2009) Patterns and scales in gastrointestinal microbial ecology. *Gastroenterology* **136**, 1989–2002.

Cani, P.D., Delzenne, N.M., Amar, J., et al. (2008) Role of gut microflora in the development of obesity and insulin resistance following high-fat diet feeding. *Pathologie Biologie* **56**, 305–309.

Cavalieri, D., Calura, E., Romualdi, C., et al. (2009) Filling gaps in PPARa signaling through comparative nutrigenomics analysis. *BMC Genomics* **10**, 596.

Claesson, M.J., O'Sullivan, O., Wang, Q., et al. (2009) Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS One* **4**, e6669.

de Graaf, A.A., Freidig, A.P., De Roos, B., et al. (2009) Nutritional systems biology modeling: from molecular mechanisms to physiology. *PLoS Computational Biology* **5**, e1000554.

de Graaf, A.A., Maathuis, A., de Waard, P., et al. (2010) Profiling human gut bacterial metabolism and its kinetics using [U-^{13}C]glucose and NMR. *NMR in Biomedicine* **23**, 2–12.

Derrien, M., van Passel, M.W.J., van de Bovenkamp, J.H.B., et al. (2010) Mucin-bacterial interactions in the human oral cavity and digestive tract. *Gut Microbes* **1**: 1–15.

Dethlefsen, L., Huse, S., Sogin, M.L., et al. (2008) The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biology* **6**, e280.

De Vogel-van den Bosch, H.M., Bünger, M., de Groot, P.J., et al. (2008) PPARalpha-mediated effects of dietary lipids on intestinal barrier gene expression. *BMC Genomics* **9**, 231.

De Wit, N.J., van den Bosch-Vermeulen, H., de Groot, P.J., et al. (2008) The role of the small intestine in the development of dietary fat-induced obesity and insulin resistance in C57BL/6J mice. *BMC Medical Genomics* **1**, 14.

Dumas, M-E., Barton, R.H., Toye, A., et al. (2006) Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proceedings of the National Academy of Sciences of the USA* **103**, 12511–12516.

Duncan, S.H., Belenguer, A., Holtrop, G., et al. (2007) Reduced Dietary Intake of Carbohydrates by Obese Subjects Results in Decreased Concentrations of Butyrate and Butyrate-Producing Bacteria in Feces. *Applied and Environmental Microbiology* **73**, 1073–1078.

Gehlenborg, N., O'Donoghue, S.I., Baliga, N.S., et al. (2010) Visualization of omics data for systems biology. *Nature Methods* **7**: S56–S68.

Gill, S.R., Pop, M., Deboy, R.T., et al. (2006) Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355–1359.

Gross, G., Van Der Meulen, J., Snel, J., et al. (2008) Mannose-specific interaction of *Lactobacillus plantarum* with porcine jejunal epithelium. *FEMS Immunology Medical Microbiology* **54**, 215–223.

Han, J., Antunes, L.C., Finlay, B.B., et al. (2010) Metabolomics: towards understanding host-microbe interactions. *Future Microbiology* **5**, 153–161.

Hartman, A.L., Lough, D.M., Barupal, D.K., et al. (2009) Human gut microbiome adopts an alternative state following smallbowel transplantation. *Proceedings of the National Academy of Sciences of the USA* **106**, 17187–17192.

Hildebrandt, M.A., Hoffmann, C., Sherrill-Mix, S.A., et al. (2009) High Fat Diet Determines the Composition of the Murine Gut Microbiome Independently of Obesity. *Gastroenterology* **137**, 1716–1724.

Hirasawa, A., Hara, T., Katsuma, S., et al. (2008) Free fatty acid receptors and drug discovery. *Biological & Pharmaceutical Bulletin* **31**, 1847–1851.

Hooper, L.V., Wong, M.H., Thelin, A., et al. (2010) Molecular analysis of commensal host-microbial relationships in the intestine. *Science* **291**, 881–884.

Ichimura, A., Hirasawa, A., Hara, T., et al. (2009) Free fatty acid receptors act as nutrient sensors to regulate energy homeostasis. *Prostaglandins & Other Lipid Mediators* **89**, 82–88.

Kint, G., Fierro, C., Marchal, K., et al. (2010) Integration of 'omics' data: does it lead to new insights into host-microbe interactions? *Future Microbiology* **5**, 313–328.

Kirschner, D.E., Young, D., & Flynn, J.L. (2010) Tuberculosis: global approaches to a global disease. *Current Opinions in Biotechnology* **21**, 524–531.

Klitgord, N. & Segrè, D. (2010) The importance of compartmentalization in metabolic flux models: yeast as an ecosystem of organelles. *Genome Information* **22**, 41–55.

Kurokawa, K., Itoh, T., Kuwahara, T., et al. (2007) Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Research* **14**, 169–181.

Lee, J.E., Lee, S., Sung, J., et al. (2010) Analysis of human and animal fecal microbiota for microbial source tracking. *Multidisciplinary Journal of Microbial Ecology* **5**, 362–365.

Ley, R.E. (2010) Obesity and the human microbiome. *Current Opinions in Gastroenterology* **26**, 5–11.

Ley, R.E., Backhed, F., Turnbaugh, P., et al. (2005) Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the USA* **102**, 11070–11075.

Ley, R.E., Turnbaugh, P.J., Klein, S., et al. (2006) Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023.

Li, M., Wang, B., Zhang, M., et al. (2008) Symbiotic gut microbes modulate human metabolic phenotypes. *Proceedings of the National Academy of Sciences of the USA*, **105**, 2117–2122.

Mahowald, M.A., Rey, F.E., Seedorf, H., et al. (2009) Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proceedings of the National Academy of Sciences of the USA* **106**, 5859–5864.

Mair, C., Plitzner, C., Pfaffl, M.W., et al. (2010) Inulin and probiotics in newly weaned piglets: effects on intestinal morphology, mRNA expression levels of inflammatory marker genes and haematology. *Archives of Animal Nutrition* **64**, 304–321.

Marco, M.L., de Vries, M.C., Wels, M., et al. (2010) Convergence in probiotic *Lactobacillus* gut-adaptive responses in humans and mice. Multidisciplinary *Journal of Microbial Ecology* **4**, 1481–1484.

Martin, F.P., Wang, Y., Sprenger, N., et al. (2008) Top-down systems biology integration of conditional prebiotic modulated transgenomic interactions in a humanized microbiome mouse model. *Molecular Systems Biology* **4**, 205.

Maslowski, K.M., Vieira, A.T., Ng, A., et al. (2009) Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* **461**, 1282–1286.

Membrez, M., Blancher, F., Jaquet, M., et al. (2008) Gut microbiota modulation with norfloxacin and ampicillin enhances glucose tolerance in mice. *FASEB Journal* **22**, 2416–2426.

Morris, M.K., Saez-Rodriguez, J., Sorger, P.K., et al. (2010) Logic-based models for the analysis of cell signaling networks. *Biochemistry* **49**, 3216–3224.

Muñoz-Tamayo, R., Laroche, B., Walter, E., et al. (2010) Mathematical modelling of carbohydrate degradation by human colonic microbiota. *Journal of Theoretical Biology* **7**, 189–201.

Naqvi, A., Rangwala, H., Keshavarzian, A., et al. (2010) Network-based modeling of the human gut microbiome. *Chemistry & Biodiversity* **7**, 1040–1050.

Nazli, A., Yang, P.C., Jury, J., et al. (2004) Epithelia under metabolic stress perceive commensal bacteria as a threat. *American Journal of Pathology* **164**, 947–957.

Nelson, K.E., Weinstock, G.M., & the Human Microbiome Jumpstart Reference Strains Consortium. (2010) A catalog of reference genomes from the human microbiome. *Science* **328** (5981), 994–999.

Nikkila, J. & de Vos, W.M. (2010) Advanced approaches to characterize the human intestinal microbiota by computational meta-analysis. *Journal of Clinical Gastroenterology* **44**, S2–S5.

Niot, I., Poirier, H., Tran, T.T., et al. (2009) Intestinal absorption of long-chain fatty acids: Evidence and uncertainties. *Progress in Lipid Research* **48**, 101–115.

Oberhardt, M.A., Palsson, B.Ø., & Papin, J.A. (2009) Applications of genome-scale metabolic reconstructions. *Molecular Systems Biology* **5**, 320.

Oda, K. & Kitano, H. (2006) A comprehensive map of the toll-like receptor signaling network. *Molecular Systems Biology* **2**, 2006.0015.

O'Hara, A.M. & Shanahan, H. (2006) The gut flora as a forgotten organ. *EMBO Reports* **7**, 688–693.

Palmer, C., Bik, E.M., DiGiulio, D.B., et al. (2007) Development of the human infant intestinal microbiota. *PLoS Biology* **5**: e177.

Pastink, M.I., Teusink, B., Hols, P., et al. (2009) Genome-scale model of Streptococcus thermophilus LMG18311 for metabolic comparison of lactic acid bacteria. *Applied and Environmental Microbiology* **75**, 3627–3633.

Perelson, A.S. (2002) Modelling viral and immune system dynamics. *Nature Reviews Immunology* **2**, 28–36.

Puchałka, J., Oberhardt, M.A., Godinho, M., et al. (2008) Genome-scale reconstruction and analysis of the *Pseudomonas putida* KT2440 metabolic network facilitates applications in biotechnology. *PLoS Computational Biology* **4**: e1000210.

Qin, J., Raes, J., MetaHit Consortium et al. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464** (7285), 59–65.

Radonjic, M., de Haan, J.R., van Erk, M.J., et al. (2009) Genome-wide mRNA expression analysis of hepatic adaptation to high-fat diets reveals switch from an inflammatory to steatotic transcriptional program. *PLoS One* **4**: e6646.

Raes, J. & Bork, P. (2008) Molecular eco-systems biology: towards an understanding of community function. *Nature Reviews Microbiology* **6**, 693–699.

Rajilić-Stojanović M., Heilig, H.G., Molenaar, D., et al. (2009) Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environmental Microbiology* **11**, 1736–1743.

Rajilić-Stojanović, M., Smidt, H., & de Vos, W.M. (2007) Diversity of the human gastrointestinal tract microbiota revisited. *Environmental Microbiology* **9**, 2125–2136.

Rey, F.E., Faith, J.J., Bain, J., et al. (2010) Dissecting the in vivo metabolic potential of two human gut acetogens. *Journal of Biological Chemistry* **285**, 22082–22090.

Rizzetto, L. & Cavalieri, D. (2010) A systems biology approach to the mutual interaction between yeast and the immune system. *Immunobiology* **215** (9–10), 762–769.

Rodenburg, W., Bovee-Oudenhoven, I.M., Kramer, E., et al. (2007b) Gene expression response of the rat small intestine following oral Salmonella infection. *Physiological Genomics* **30**, 123–133.

Rodenburg, W., Keijer, J., Kramer, E., et al. (2007a) *Salmonella* induces prominent gene expression in the rat colon. *BMC Microbiology* **7**, 84.

Rodenburg, W., Heidema, A.G., Boer, J.M., et al. (2008b) A framework to identify physiological responses in microarray-based gene expression studies: selection and interpretation of biologically relevant genes. *Physiological Genomics* **33**, 78–90.

Rodenburg, W., Keijer, J., Kramer, E., et al. (2008a) Impaired barrier function by dietary fructo-oligosaccharides (FOS) in rats is accompanied by increased colonic mitochondrial gene expression. *BMC Genomics.* **9**, 144.

Rolfe, D.F.S. & Brown, G.C. (1997) Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiological reviews* **77**, 731–758.

Röling, W.F., Ferrer, M., & Golyshin, P.N. (2010) Systems approaches to microbial communities and their functioning. *Current Opinions in Biotechnology* **21**(4), 532–538.

Ruppin, E., Papin, J., de Figueiredo, L., et al. (2010) Metabolic reconstruction, constraint-based analysis and game theory to probe genome-scale metabolic networks. *Current Opinions in Biotechnology* **21**, 502–510.

Saez-Rodriguez, J., Alexopoulos, L.G., Epperlein, J., et al. (2009) Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Molecular Systems Biology* **5**, 331.

Samaga, R., Saez-Rodriguez, J., Alexopoulos, L.G., et al. (2009) The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. *PLoS Computational Biology* **5**, e1000438.

Samuel, B.S., Shaito, A., Motoike, T., et al. (2008) Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. *Proceedings of the National Academy of Sciences of the U S A* **105**, 16767–16772.

Schwiertz, A., Taras, D., Schäfer, K., et al. (2010) Microbiota and SCFA in lean and overweight healthy subjects. *Obesity* **18**, 190–195.

Shoda, L., Kreuwel, H., Gadkar, K., et al. (2010) The Type 1 Diabetes PhysioLab Platform: a validated physiologically based mathematical model of pathogenesis in the non-obese diabetic mouse. *Clinical & Experimental Immunology* **161**, 250–267.

Somasundaram, S., Rafi, S., Hayllar, J., et al. (1997) Mitochondrial damage: a possible mechanism of the "topical" phase of NSAID induced injury to the rat intestine. *Gut* **41**, 344–353.

Somasundaram, S., Sigthorsson, G., Simpson, R.J., et al. (2000) Uncoupling of intestinal mitochondrial oxidative phosphorylation and inhibition of cyclooxygenase are required for the development of NSAID-enteropathy in the rat. *Alimentary Pharmacology & Therapeutics* **14**, 639–650.

Tap, J., Mondot, S., Levenez, F., et al. (2009) Towards the human intestinal microbiota phylogenetic core. *Environmental Microbiology* **11**, 2574–2582.

Teusink, B., Wiersma, A., Jacobs, L., et al. (2009) Understanding the adaptive growth strategy of Lactobacillus plantarum by in silico optimisation. *PLoS Computational Biology* **5**, e1000410.

Thissen, U., Wopereis, S., Van Den Berg, S.A., et al. (2009) Improving the analysis of designed-studies by combining statistical modelling with study design information. *BMC Bioinformatics* **7**(10), 52.

Tilg, H., Moschen, A.R., & Kaser, A. (2009) Obesity and the Microbiota. *Gastroenterology* **136**, 1476–1483.

Troost, F.J., van Baarlen, P., Lindsey, P., et al. (2008) Identification of the transcriptional response of human intestinal mucosa to *Lactobacillus plantarum* WCFS1 in vivo. *BMC Genomics* **9**, 374.

Trosvik, P., Stenseth, N.C., & Rudi, K. (2010) Convergent temporal dynamics of the human infant gut microbiota. *Multidisciplinary Journal of Microbial Ecology* **4**, 151–158.

Turnbaugh, P.J., Hamady, Y., Yatsunenko, T., et al. (2009) A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., et al. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031.

van Baarlen P., Troost, F.J., van der Meer, C., et al. (2010) Human mucosal in vivo transcriptome responses to three lactobacilli indicate probiotic modulatory capacities and therapeutic potential. *Proceedings of the National Academy of Sciences of the U S A* **108**, 4562–4569.

van Baarlen P., Troost, F.J., van Hemert, S., et al. (2009) Differential NF-kappaB pathways induction by *Lactobacillus plantarum* in the duodenum of healthy humans correlating with immune tolerance. *Proceedings of the National Academy of Sciences of the U S A* **106**, 2371–2376.

Van Den Bosch, H.M., De Wit, N.J., Hooiveld, G.J., et al. (2008) A cholesterol free, high-fat diet suppresses gene expression of cholesterol transporters in murine small intestine. *American Journal of Physiological Gastrointestinal Liver Physiology* **294**, G1171–G1180.

Vijay-Kumar, M., Aitken, J.D., Carvalho, F.A., et al. (2010) Metabolic Syndrome and Altered Gut Microbiota in Mice Lacking Toll-Like Receptor 5. *Science* **328**, 228–231.

Vrieze, A., Holleman, F., Zoetendal, E.G., et al. (2010) The environment within: how gut microbiota may influence metabolism and body composition. *Diabetologia* **53**, 606–613.

Waldram, A., Holmes, E., Wang, Y., et al. (2009) Top-down systems biology modeling of host metabotype-microbiome associations in obese rodents. *Journal of Proteome Research* **8**, 2361–2375.

Wall, R., Ross, R.P., Shanahan, F., et al. (2009) Metabolic activity of the enteric microbiota influences the fatty acid composition of murine and porcine liver and adipose tissues. *American Journal of Clinical Nutrition* **89**, 1393–1401.

Wen, L., Ley, R.E., Volchkov, P.Y., et al. (2008) Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature* **455**, 1109–1113.

Williams, H.R., Cox, I.J., Walker, D.G., et al. (2009) Characterization of inflammatory bowel disease with urinary metabolic profiling. *American Journal of Gastroenterology* **104**, 1435–1444.

Xavier, J.B., De Kreuk, M.K., Picioreanu, C., et al. (2007) Multi-scale individual-based model of microbial and bioconversion dynamics in aerobic granular sludge. *Environmental Science & Technology* **41**, 6410–6417.

Xie, L., Li, Y., & Wei, C. (2009) More than 9,000,000 unique genes in human gut bacterial community: estimating gene numbers inside a human body. *PLoS One* **4**, e6074.

Zengler, K., Toledo, G., Rappe, M., et al. (2002) Cultivating the uncultured. *Proceedings of the National Academy of Sciences of the U S A* **99**, 15681–15686.

Zoetendal, E.G., Rajilić-Stojanović, M., & de Vos, W.M. (2008) High throughput diversity and functionality analysis of the gastrointestinal tract microbiota. *Gut* **57**, 1605–1615.

Chapter 6
From Visual Biological Models Toward Mathematical Models of the Biology of Complex Traits

Marinus F.W. te Pas, Arend J.W. Hoekman, and Ina Hulsegge

Introduction

Complex Traits

Genetic diseases can be caused by a single gene defect or by multiple gene defects (Thomson, 1994). Similarly, traits can be regulated by a single gene or a single gene can be responsible for most of the genetic variation in that trait (called a major gene (Guo and Thompson, 1992; Janss et al., 1995, 1997; Thaller et al., 1996)). However, most traits in livestock science, such as traits related to production, reproduction, or health, have a complex genetic regulation based upon the combined expressions and functions of many genes (http://www.biology.duke.edu/nijhout/) (Nijhout, 2002; Nijhout et al., 2007). The expression of many genes is the net result of genetic composition and environmental effects. For most genes, the variation in the expression has only small effects on variation of the traits. This also makes that observed variation in complex traits can only be partly explained by variation in the expression of a single gene. These facts indicate the urge to a detailed measurement of the phenotype of the complex trait. Having as good as possible phenotypic data is vital for the association of genomic and biological data and the development of a biological model (Freimer and Sabbati, 2003). A biological model can be defined as (part of) the biological knowledge displayed as a model. In this chapter, we will often use physiological pathways as can be found in several databases (see below). However, morphological models, or models of complex molecules or organelles are also biological models.

Some of the complexity of the regulation of traits can already be seen in the number of quantitative trait loci (QTL) for a specific trait identified on the genetic maps available for individual species (http://www.thearkdb.org/arkdb/) (Hu et al., 2001). However, a comprehensive view of the regulation of the traits is lacking for most traits.

Systems Biology and Livestock Science, First Edition. Edited by Marinus F.W. te Pas, Henri Woelders, and André Bannink.

Research in the Post Genomic Era

Following the sequencing of the human genome an increasing number of animal genomes are being published including livestock species (chicken: Wallis et al., 2004; cattle: Bovine Genome Sequencing and Analysis Consortium, Elsik et al., 2009; pig: http://www.ncbi.nlm.nih.gov/genome/guide/pig/). Knowledge of the genome of a species includes the knowledge of the existence of (most of) the genes in a genome, but does not necessarily include insight into the regulation of the expression of the genome and how this regulates complex traits of the animal, or even how it is associated to them. Using sequence information, functional genomics investigates the expression of the genome at the RNA level (called transcriptomics), the protein level (proteomics), or the activity level (metabolomics). By getting an improved understanding of the functioning of the genome, the research aims to get an improved understanding of the biological processes that underlie the genetic regulation of complex traits.

Functional genomics intends to measure the entire genomic expression equivalent of a cell, tissue, or organ at a certain condition or physiological state. By comparing the expression profiles obtained under different physiological or environmental situations, insight is acquired into differences in genomic expression. In this way, functional genomics generated an ever-increasing amount of data during the last decades. The raw data of many experiments are available through specific databases (e.g., the Gene Expression Omnibus or GEO: http://www.ncbi.nlm.nih.gov/geo/) opening the opportunity to compare experiments and species. However, the large data content urges for thorough analyses of the data at several levels of aggregation of the biological system under study. Going through each of these levels is necessary toward building a biological model.

Recent research indicates that the expression of the genome is regulated in a complex way on all levels. The mechanism of transcription factors is well known. These mechanisms include the effect of enhancing or repressing proteins that regulate mRNA expression of genes by binding to specific regulatory sequences close-by or more distant to the location of the gene on the chromosome. More recently, it was discovered that these mechanisms are more complex as a result of small noncoding RNAs, called miRNAs that regulate the stability of mRNAs or the translation possibilities of the mRNA by binding to the mRNA (Kim, 2005). At the protein level, carbohydrate, lipid, or phosphate groups may be attached to the protein that regulates it half-life time or its activity (Burnett and Kennedy, 1954; Mann and Jensen, 2003). For example, attached groups may regulate the activity of an enzyme and thereby affect the composition of the metabolome of a cell. What exactly triggers a cell to decide which modification to use and what mechanisms are involved is often still unknown.

The transcriptomic and proteomic expression of the genome is the result of the interaction of the genome with the environment. Differences in the expression of the genome may be induced by food intake of the animal (Yoshizawa et al., 1997; Müller and Kersten, 2003) (see also Chapters 8 and 9), by environmental temperature (Goldspink, 1995) or daylight (Guido et al., 1999), or by pathogens including the host–pathogen interaction (Diehn and Relman, 2001) (see also Chapter 10). Other animals from the same or different species may also affect gene expression levels. Differences in the expression of the genome between individuals, which may be related to mutations in the genome sequence, are related to differences among

individual animals in observed traits (Andersson and Georges, 2004; Chesler et al., 2005). Knowledge of the regulatory mechanisms of important traits hence opens the ability to modulate the traits by management or breeding.

Functional genomics started the investigation of the expression of the genomes of livestock species (te Pas et al., 2005, 2008, 2009; Cagnazzo et al., 2006; Kadarmideen and Reverter, 2007; Wimmers et al., 2007). The picture that arises from this research is often complex and hampers a full understanding, which is necessary, however, to be able to modulate traits through the modulation of genome expression via animal management of breeding. Analyses with bioinformatics tools help in this stage of the research. Such analyses build on known biological models, which are often based upon the sparse knowledge of the physiological function associated with a limited number of genes (te Pas et al., 2008). These biological models may be used for mathematical modeling and a Systems Biology approach. When information of biological models is combined with mathematical modeling, a useful tool may be obtained that helps to understand, explain, and, preferably, even predict the regulation of these complex traits. Such an increased understanding is needed to provide insight into how the trait can be optimized for the purpose of livestock production. Also, the interactions between various traits can be studied in this manner. It can, for example, clarify the relation between productive traits and other traits such as animal health and animal welfare.

Objective

The objective of this chapter is to review methods used to develop biological models that represent regulation and causal relationships for complex traits in animals' use in livestock production. The methods used to generate these models will be discussed and it will be shown how their development is an essential step toward the development of mathematical models called Systems Biology. We define Systems Biology as the science generating mathematical (predictive) models, which are based on biological knowledge. These mathematical models may serve as a basis for the monitoring, the control, and the modulation of complex traits in an integrative and balanced way in farm animals in the future. Such achievement would add to sustainable change of livestock production system in the future.

Complex Traits and Biological Models

Biological models generally are based on basic knowledge of physiological processes. Although these physiological processes are related to the complex traits, at present, functional genomic regulatory mechanisms, gene functions, and their associated physiological function are only partly known. Therefore, biological models may only partly explain a complex trait. Because many genes have redundant functions the contributions of genes may also be hidden in the variations of other genes. Even in knock out mice, which completely turns off the expression of a gene, redundant gene functionality may hide the function of a gene (Heber et al., 2000; te Pas and Soumillion, 2001). Therefore, it should be kept in mind that genes may remain unnoticed when studying

or describing regulatory mechanisms. If not recognized by other methods, this may leave the explanation by the biological model of the complex trait incomplete. Thus, methods need to be used that are able to identify the currently unknown physiological function of certain genes influencing a complex trait.

Identify Genes That Influence Complex Traits

Several methods can identify genes as part of the regulatory mechanism of a complex trait. First of all, the expression profile of genes may indicate association with the trait. Statistical methods can show the relatedness of expression profile and changes in the complex trait phenotype.

Other methods use available physiological information of genes. Genes do not act as stand, alone units but are actively integrated in pathways or networks. Pathways show how genes work together based upon physiological information. Several databases that are accessible via the Internet show pathways in diverse graphical formats (Table 6.1). Networks show how genes (proteins) interact based upon knowledge about physical binding of proteins or proteins and DNA (Vazquez et al., 2003; Merico et al., 2009). Alternatively, networks may be created by combining pathways (te Pas et al., 2007, 2008). It should be noted that pathways and networks are graphical, not quantitative representations of physiology. Such biological models can be visually interpreted (Merico et al., 2009) and lead to increased understanding of complex traits. This is in line with biology as a descriptive science. However, future use of the models to modify

Table 6.1 Example of the different pathway databases.

Database	URL	Type of data
KEGG	http://www.genome.jp/kegg/	Metabolic and nonmetabolic pathways—focus on sequence of reactions
BioCarta	http://www.biocarta.com/	Metabolic and nonmetabolic pathways—focus on protein interactions and position in the cell
Reactome	http://www.reactome.org/	Metabolic and nonmetabolic pathways
Interactome	http://interactome.dfci.harvard.edu/	Human protein–protein interaction map
Database of interacting proteins	http://dip.doe-mbi.ucla.edu/dip/Main.cgi	Experimentally determined protein–protein interactions
BioGrid	http://www.thebiogrid.org/	Curated set of physical and genetic interactions
MetaCyc	http://metacyc.org/	Experimentally elucidated metabolic pathways from over 900 organisms

the traits requires a quantitative model that can be used to predict. To transform these into quantitative items combination with experimental data is necessary. This is a vital step to be taken before mathematical modeling can start.

Finally, genes lacking a physiological annotation may be added to the model using cluster analysis (Eisen et al., 1998). Cluster analysis can join together genes with similar expression profiles and suggests that such genes have common functionality. Thus, genes with unknown function can be ascribed the same function as genes with known function in the same cluster. Although a bit speculative, it may help to make the biological model more complete and improve the representation of the relationship with the complex trait. However, there is no guarantee that this will automatically also improve the insight in the regulation of the trait itself.

Generation of a Biological Model

Building a biological model that represents (aspects of) the functionality of an organ, tissue, or cell type always starts with collecting biological data at several biological levels. Often, functional genomics generates expression profiles at mRNA and/or protein levels (transcriptomics, proteomics). In addition, physiological information shows how the cells or tissues work at the biochemical level. It can include cell biology information about functioning, anatomy, histology, or histochemistry of the cells. Next to functional genomics, metabolomics provides data and insights on the metabolic flux through enzymes and pathways and thereby of the functioning of the proteome (i.e., the total protein expression of a cell, tissue, or organ). Combining the data generated by transcriptomics (i.e., the total genomic mRNA expression of a cell, tissue, or organ), proteomics, and metabolomics describing the cells or tissue for a range of physiological statuses is needed to generate knowledge about the functionality and cause-effects relationships of molecule fluxes through a system. Knowledge about the fluxes can make a biological model more quantitative as a first step toward a mathematical model. Here, we will describe the steps that can be taken to generate a biological model.

Analysis of Functional Genomics Research

A typical transcriptomics experiment generates data of all (or most of) the genes expressed in a tissue. The expression levels are either recorded directly in each animal and each status of the complex trait or as a ratio to another situation. The latter may, for example, be a pool of samples to generate a mean expression level related to the mean status of the trait. This type of research can generate two lists of genes: (1) a list of expressed genes (as opposite to a list of genes not expressed in the cells or tissue under investigation), and (2) a list of genes with differential expression when two situations are compared. Although these lists are important in themselves, they do not directly generate knowledge about the cell type or regulation of the difference between the two physiological states. Also, they do not deliver a quantitative understanding in terms of concentrations, massive ion fluxes. However, they can be used to derive essential elements to be included in a biological model.

Adding Biological Data to Lists of Genes

As a first step, the lists of genes can be analyzed with the data in the Gene Ontology (GO) database (http://www.geneontology.org/)(Barrell et al., 2008). This database provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data, and tools to access and process the provided data. The gene annotation includes known biological functions and cellular locations of the protein encoded by the gene. The database is organized top-down and consists of three major parts: (1) cellular component, (2) biological process, and (3) molecular function. Each of these is split into smaller parts describing further details. A gene is annotated first in one or more of the top categories (biological function, cellular localization, molecular function) and in more detailed (sub)functions. Although no direct information about livestock (production) traits are given, the interpretation of these biological data delivers the user background information about the regulation of the trait.

Another important feature of the GO database is the synonym function. Many genes have more than one name, and not all databases contain all names. Using the synonym function of the GO database opens the possibility to extract maximum information from other databases (Ashburner et al., 2000; Gene Ontology Consortium et al., 2004).

Genes Cooperate in Pathways

In a next step of the analysis toward building the biological model, it should be realized that genes usually do not act alone. Proteins may physically interact or may interact through a series of substrate–product combinations in which the product of one gene (protein) act as a substrate for the next one, for example, in a series of enzyme reactions. These functionalities may form networks of genes (proteins) or they may form pathways leading to specific functionalities of cells such as cell division or other cell features or fates of the cell (Kanehisa and Goto, 2000; Kanehisa et al., 2006, 2008). These pathways describe how genes interact (see Table 6.1). Pathways may only be partly used in a tissue (for the concept of subpathways, see te Pas et al., 2007). Similar to the gene ontologies in the GO database different types of pathways exist; biochemical pathways indicating lines of genes converting a substrate via several steps into an end product, morphological pathways indicating how cellular structures are built from physical interaction of proteins, and physiological pathways indicating how biological processes are regulated. Figure 6.1 shows an example of different forms of pathways as generated by the KEGG (Kyoto Encyclopedia of Genes and Genomes, http://www.genome.jp/kegg/) database (Kanehisa and Goto, 2000; Kanehisa et al., 2006, 2008). The figure shows an example of a biochemical pathway (Figure 6.1A, the well-known Citric acid cycle or Krebs cycle generating ATP energy via a sequence of enzymatic reactions), a morphological pathway (Figure 6.1B, the proteasome, a complex molecule build by the physical interaction of many peptides encoded by different genes), and a physiological pathway (Figure 6.1C, the sequence of events leading to prion disease (e.g., mad cow disease, scrapie) via resistance against protein degradation). The type of information gained from each type of model is different. A combination of pathways may be more accurately describing the complex trait phenotype than individual, separate models. To get a better understanding of

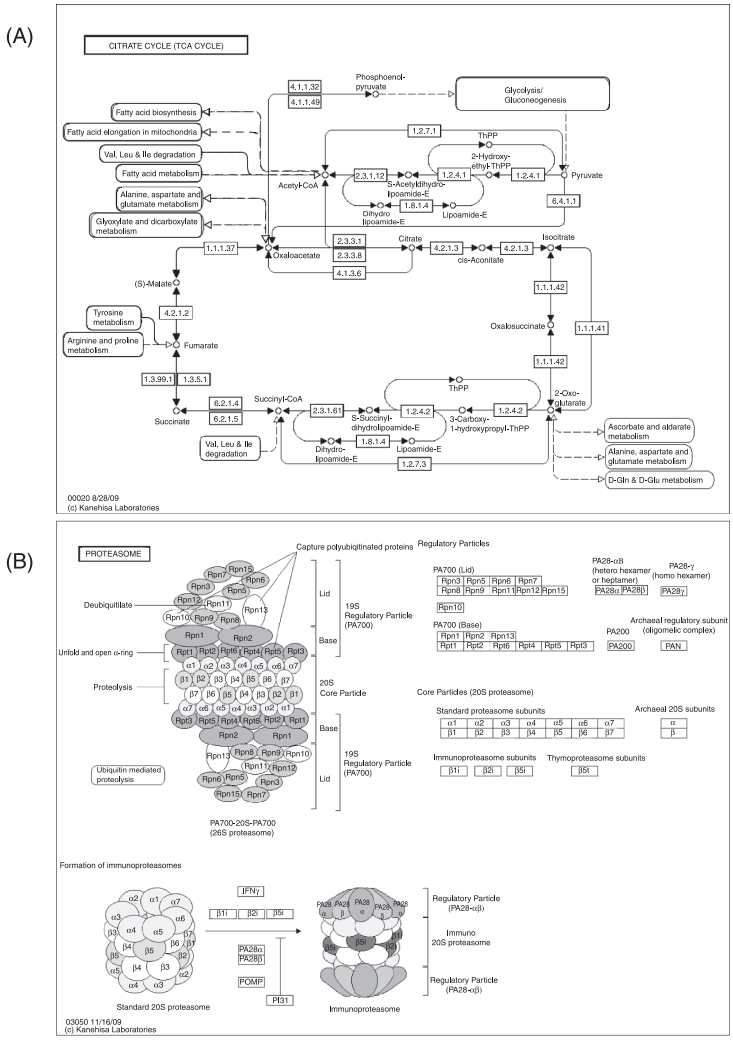


Figure 6.1 Three examples of biological pathways taken from the KEGG database (http://www.genome.jp/kegg/) (Kanehisa and Goto, 2000; Kanehisa et al., 2006, 2008). (A) The citrate cycle (also called TCA cycle or Krebs cycle), a typical biochemical pathway showing arrays of enzymes. Several input and output pathways are indicated making suggestions for networks of pathways. (B) The structure of the proteasome consisting of many peptides encoded by different genes—an example of a pathway showing the morphology of a cellular structure. (*Continued*)

pathways a Web site with a growing number of molecular animations is available (http://www.johnkyrk.com/).

Different databases may represent the same pathways, but in a different way. Each visual representation provides specific biological insights, thus, combining the data would give the most comprehensive representation of the biological model for the complex trait. Figure 6.2 shows an example of the WNT signaling pathway. The WNT

(C)

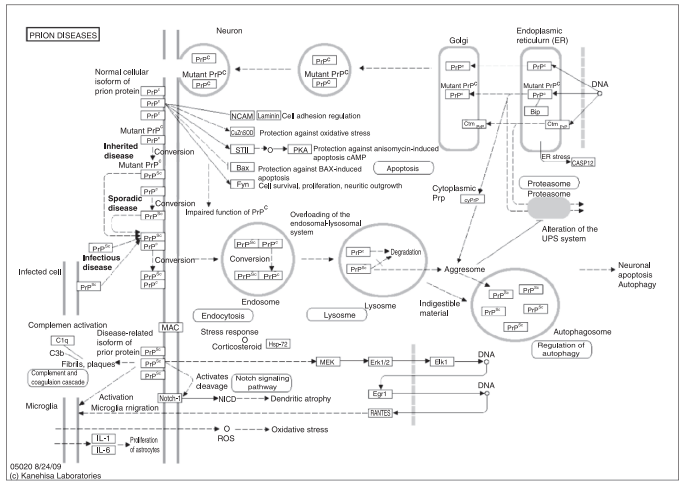


Figure 6.1 (*Continued*) (C) An overview of the biochemical events leading to prion disease, well known as scrapie or mad cow disease—a complex trait in which a single protein is converted in such a 3D structure that it resists degradation by the proteasome.

signaling pathway describes a network of proteins most well known for their roles in embryogenesis and cancer, but also involved in normal physiological processes in adult animals such as adult hippocampal neurogenesis (Lie et al., 2005). Figure 6.2A shows the pathway as composed by the KEGG database focusing on the biochemical sequence of reactions; Figure 6.2B shows the same pathway from the BioCarta database (http://www.biocarta.com/) focusing more on physical interactions of proteins and the cellular localization of the reaction process; Figure 6.2C and D show two representations from the same pathway in the Reactome database (http://www.reactome.org/); C was generated with the Cytoscape software (http://www.cytoscape.org/) which ONLY uses data of physical interactions from two proteins, and D shows how the database splits the pathway into small subpathways taking the other parts of the pathway as black boxes. The Cytoscape version indicates that the pathway divides into two parts with limited connection. This was also seen in the presentation of the KEGG database, but less clear. Thus, different representations by different databases of the same biological model can generate different insights for the explanation of the regulatory mechanisms involved with complex traits.

It is important to visualize the available experimental data related to investigate a complex trait in the pathway models in order to gain understanding of that trait. Different software tools exist to find the relevant pathways. One example is the often used Internet tool called DAVID (The Database for Annotation, Visualization, and Integrated Discovery) (http://david.abcc.ncifcrf.gov/home.jsp) (Dennis et al., 2003; Huang et al., 2009). Following statistical testing, the significant pathways from the KEGG and the BioCarta databases are shown. Unfortunately, only the relevant genes are indicated by an asterisk without indication of the other genes under investigation. Figure 6.3 shows the different types of output of a pathway analysis tool developed by te Pas et al. (2007, 2008). All the genes under investigation are encircled in the pathway (Figure 6.3A). In this specific investigation, the expression profiles of the genes were

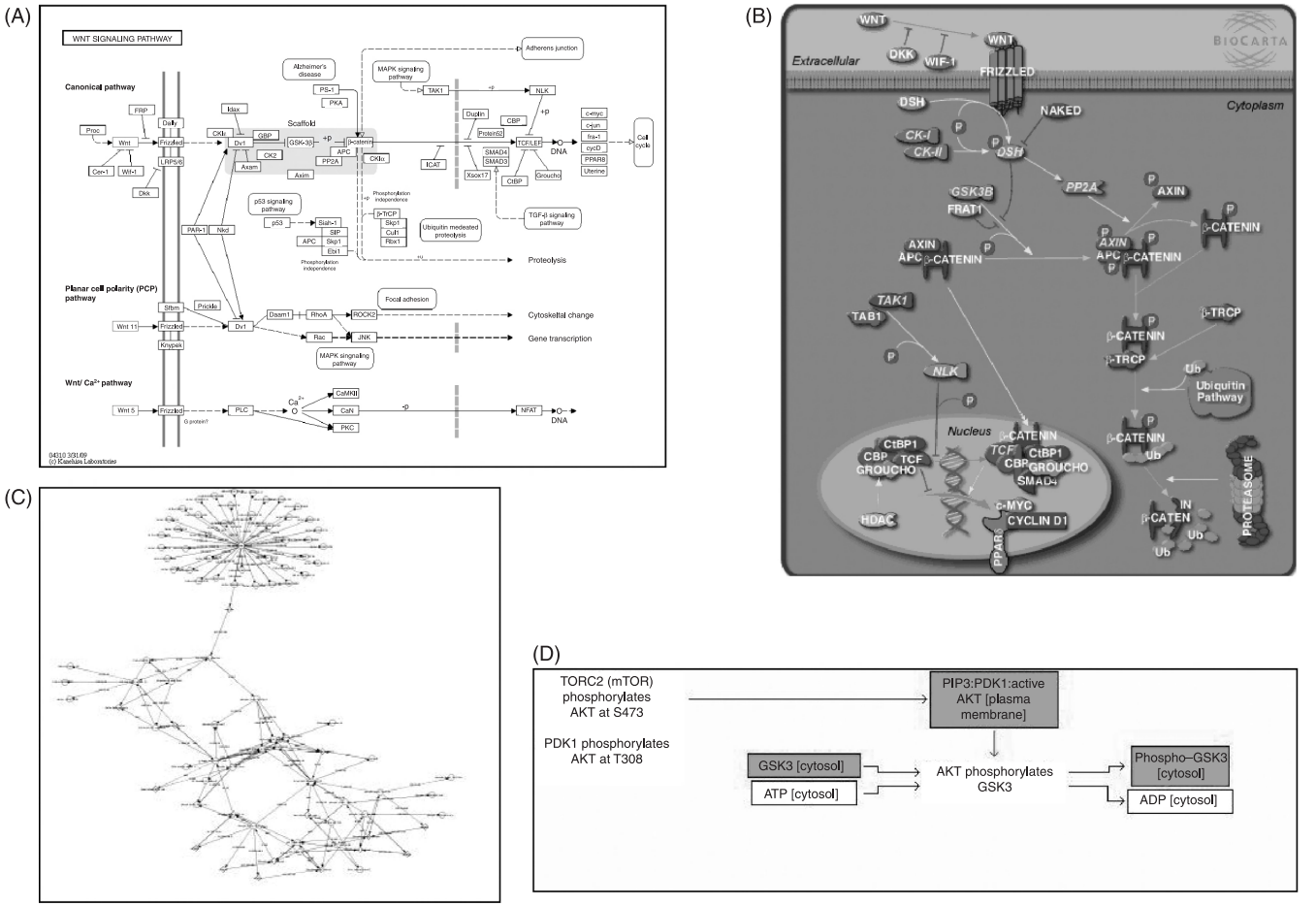


Figure 6.2 Different ways to present the same biochemical pathway. The WNT pathway was taken as an example: (A) from the KEGG database; (B) from the BioCarta database; (C) from the Reactome database exported via Cytoscape software. (D) The pathway as a whole is not constructed in the Reactome database but presented as units (black boxes) together in a main stream picture.

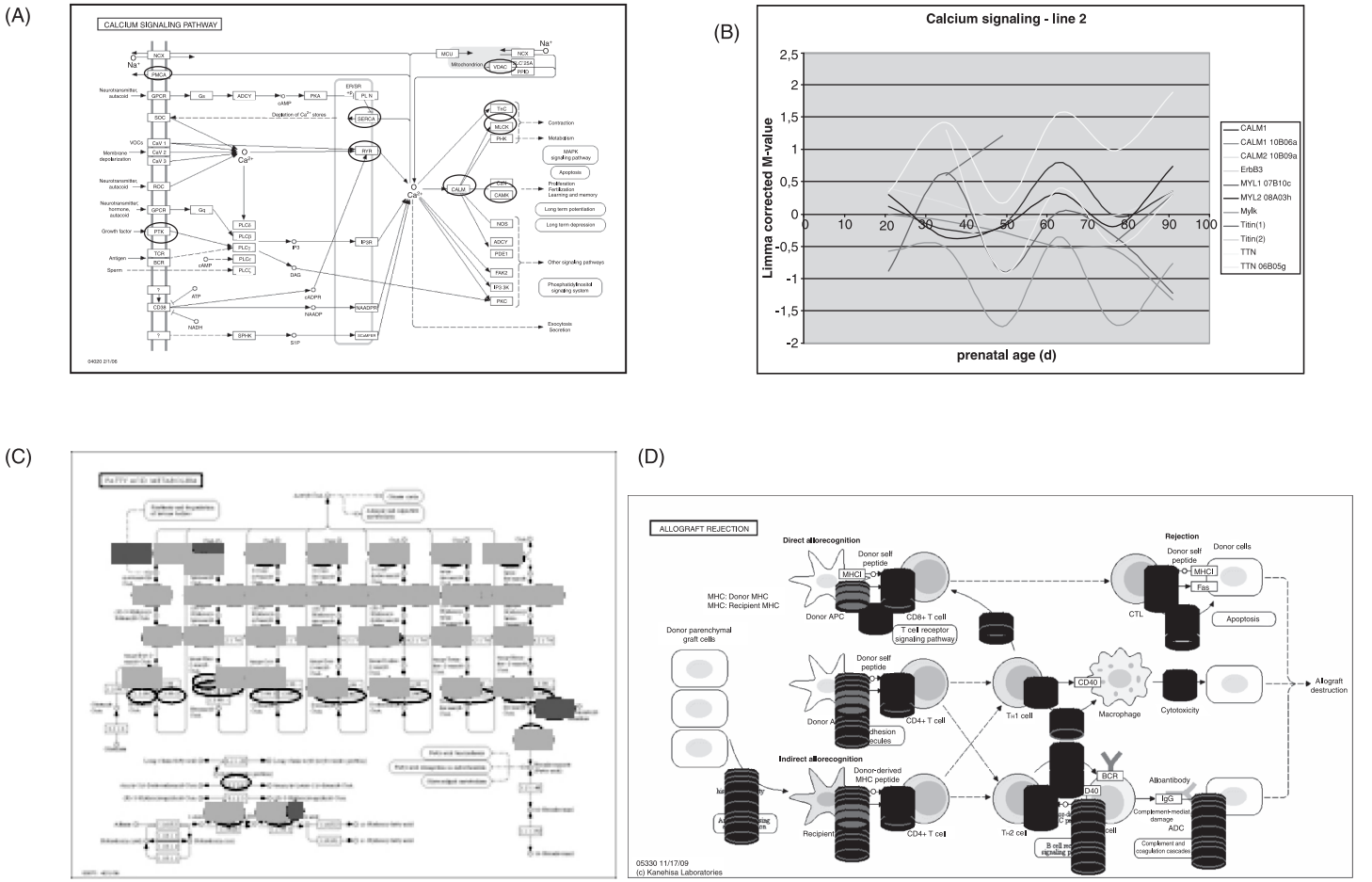


Figure 6.3 Examples visualizing experimental results investigating complex traits on biological pathways. (A, B). Using a pathway downloaded from the KEGG database first the genes under investigation are encircled (A) and the expression pattern of each gene in the time is indicated (B). Most genes of the pathway have similar regulated expression patterns in the time. (C) Alternatively, the genes in the pathway can be color coded: green color denotes up regulation, red color denoted down regulation of the gene expression in the present state of the complex trait as compared to another state. (D) Similar as (C), but each investigated animal is shown as a circle around the gene and each animal underneath the circle of the previous animal. (*See insert for color representation of this figure.*)

determined at seven moments during pig prenatal muscle tissue formation (te Pas et al., 2005, 2006; Cagnazzo et al., 2006). The profiles are shown in B indicating that all the genes in this pathway have similar expression profiles. Furthermore, these profiles were directly related to the muscle tissue formation trait (te Pas et al., 2005, 2007). It was concluded that this pathway must have an important place in the regulation of the trait. Indeed, calcium metabolism is known to be important for muscle fiber development as it may be involved in precursor cell differentiation (Przybylski et al., 1994; Baudier et al., 1995). In Figures 6.3C and D, a color coding indicates regulation of the expression: green circle indicates upregulated gene expression in this state of the complex trait as compared to a any other status of the complex trait, red indicates down regulation, and black indicates similarity of the expression level of the genes. In C, the means of the results per gene are shown; in D, the results for each animal are represented by a circle around each gene, with each next animal making a circle below the circle of the previous animal. This directly indicates the variation between individual animals of the regulation of the gene expression in relation to the phenotype of the trait under investigation (te Pas et al., 2007 and unpublished results).

From Pathways to Network of Pathways

A single pathway may end up in regulation of a specific function (i.e., a biological process such as cell division), but most pathways only describe a physiological process or a morphological feature in a cell. Therefore, a complex process may be executed by connecting and integrating pathways. The output of a pathway may be the input of a next pathway. In this manner, networks of pathways may be generated together regulating a biological process (te Pas et al., 2007, 2008). Such a network of pathways indicates the interactions between pathways including the direction (stimulation or inhibition) of the interaction. The individual pathways may be either used as black boxes or with the underlying biochemical information indicated. Alternative to the interaction of pathways also the genes may act in more than one pathway. Thus, a network of pathways may also be generated by linking together via common genes (te Pas et al., 2007, 2008).

A network of pathways can serve as a first biological model of a complex trait. Figure 6.4 shows a biological model highlighting interactions between pathways directed to the regulation of muscle formation and contraction—part of a model for the complex trait "skeletal muscle formation and physiological functioning." Using the pathway tool as described above, the results of the investigation of the skeletal muscle tissue formation in prenatal pigs were integrated into KEGG pathways. Next, regulated (sub) pathways were followed in diverse pathways. Thus, the individual pathways were linked together in a complex network of pathways, together regulating proliferation (i.e., growth) and function (i.e., contraction) of muscle tissue. Figure 6.4 shows that a biological model has been generated for two complex traits that are closely associated in physiological function of muscle tissue (te Pas et al., 2007). Similarly, in jejunum tissue, biological models based upon networks of pathways were generated for the genomic reaction of 1-day-old chicken to Salmonella enteritidis infection (Van Hemert et al., 2006a, 2006b). These pathways were associated to cellular reorganizations and cellular energy metabolism. The model comprises known and unknown

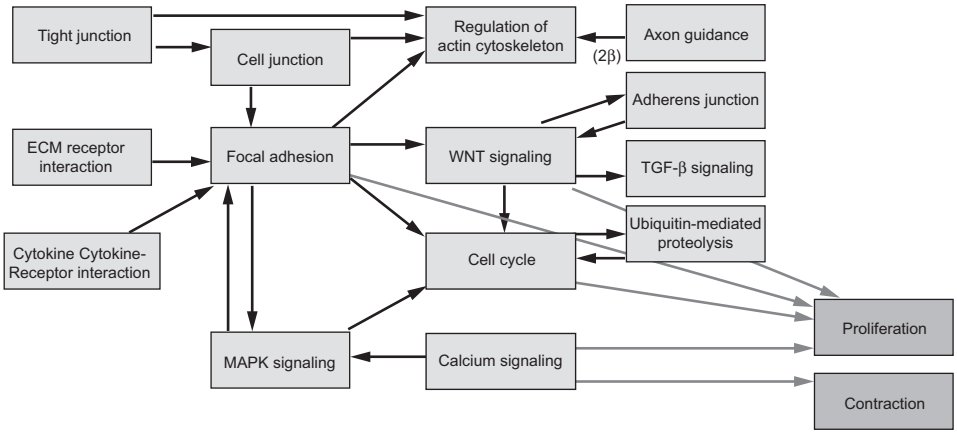


Figure 6.4 Biological model constructed using a network of pathways. The individual pathways were down loaded from the KEGG database and analyzed as shown in Figure 6.3C. As indicated in Figure 6.1A, the pathways were combined into a model where the analysis shown relevance to the trait, i.e., muscle tissue formation and functionality. (Adapted from te Pas et al., 2007.)

aspects of the host response to *Salmonella* infection, thus, increasing understanding of the biology of infection and resistance to *Salmonella* (te Pas et al., 2008).

Genes Cooperate in Networks

Generating pathways requires detailed information about reaction sequences or positional information of peptides in complex proteins or cellular structures. Such high-level information can explain many reaction constants and thus the reaction kinetics. If such high-level information is not available, networks can be generated using lower level information like (potential) protein–protein or protein–DNA (physical) interaction. Such relational networks can be built using specific softwares like Cytoscape (http://www.cytoscape.org/) (Shannon et al., 2003; Cline et al., 2007) or Osprey (http://biodata.mshri.on.ca/osprey/servlet/Index) (Breitkreutz et al., 2003). Cytoscape and Osprey are open source bioinformatics software platforms for visualizing complex molecular interaction networks and integrating these interactions with gene expression profiles and other data (e.g., associated physiological processes if association information is available).

The genes are the nodes and interactions are edges (i.e., lines between two genes) in visualizations of these networks. Quantitative levels of gene expression or relationships can be indicated in the network using color or shape of the nodes and edges. The number of relations with other genes (i.e., the number of edges) is often interpreted as the importance of a gene (node).

Such a network model can be regarded to be a different visual form of a biological model. The network model focuses on (physical) interactions and is unable to describe a sequence of (biochemical) reactions like the pathways model can. Furthermore, the network model describes relationships between genes or proteins, but due to the level of input knowledge with less certainty than the biological model that is built on

pathways information. Furthermore, the constructed networks tend to become too confusing to understand if the network describes larger datasets, and partial networks need to be constructed to gain insight from the network. Nevertheless, the biological model can provide insight in processes underlying the studied traits. Figure 6.2C presents an example of a biological network model.

Conclusions

Physiological knowledge can be used to generate biological models that provide insight in the regulation of biological processes involved with complex livestock traits such as production, reproduction, and health. Different methodologies and analysis tools produce different information models, each with their own characteristics, advantages, and disadvantages, which may even lead to different insights. The models may become quantitative to a limited extent in describing the expression levels of genes in a cell or tissue type. The model may describe the difference between two situations more quantitatively, depending upon the methodology used to measure the expression levels.

These models can be used to predict the importance of individual genes for regulating a biological process. However, this model application is hampered by the lack of knowledge about the genes in a model, for example, which genes are the important rate determining steps in pathways. This can be solved by combining the model with the results of an association study taking the genes with association to the trait as the important nodes in the model (Dixon et al., 2007; te Pas et al., unpublished results). Furthermore, these models cannot be used to predict the changes in a complex trait that is induced by environmental influences such as nutrition and housing management, or by genetic changes induced by breeding or historic effects on genotype. For this purpose, it is necessary to have quantitative understanding of the relationship between the conversion of the complex trait and changes in transcriptome or proteome expression profiles and to changing morphological/histological conditions of the cell or tissue. Such understanding can be obtained by combining the biological model with mathematical models (Laursen, 2009). Combining the models is possible since both models represent the physiology of a trait in a different form. The next two sections will describe these two additions to the biological models.

Association Studies Relating the Expression Levels of Genes or Proteins to Quantitative Traits

Until now we have a biological model of a trait based upon functional genomic and physiological knowledge. The model comprises a biochemical reaction path or complex associations of proteins and/or DNA. Furthermore, it may comprise other biological information such as morphology, etc. We assume that variation in the expression of genes in the biological model causes variation in the trait. To characterize the important genes within a biological model, the relationship between the expression profiles of the genes and the variation in the complex trait can be calculated. Several statistical methods can be used depending upon the importance of relations within

the statistical model (Lawless, 2003). Often, linear models are used, but for optimum traits quadratic associations may be more relevant. These models correct for all kinds of factors that may lead to false-positive associations such as variation in date of the measurements, animal feeding, and or housing, etc.

The result of such a study is a list of genes with association between the expression levels of the genes and the trait and the corresponding P-values for the linear or higher orders of associations. In a next step, this list is compared with the pathways and networks found. Some genes may be included in multiple pathways and networks. It is fair to assume that the genes with good association and a central position in the pathways or networks may be most important for the regulation of the trait.

However, presently, we have the problem that physiological information for many genes is still lacking, so not all genes will be found in the biological model. It is important to keep in mind that genes lacking physiological information may be even more important than the genes included in the pathways and networks. Therefore, the ultimate biological model explaining every aspect of a trait may be still far away. It could be interesting to include physiological information in the creation of the mathematical model to investigate the relative importance of the physiological unknown part of the biological model.

Classical association studies relating livestock production or reproduction trait phenotypes and genetic/genomic variation in genes or QTL are well known in the literature (Cardon and Bell, 2001; Lohmueller et al., 2003). In a next step, the association is determined between the complex trait phenotypes and continuous variables like transcriptome or proteome profiles (Devaux et al., 2001; Dixon et al., 2007). It is new to integrate the results of such an association study in a biological pathway (te Pas et al., 2010). Table 6.2 gives an example of part of such a study relating meat quality pH traits (measured at several hours postmortem) and transcriptomic analysis. The relations among genes, KEGG pathways, and complex trait are not directly clear in this example. This may be due to the trait. However, further integration with other complex meat quality traits and other association studies provides more insight into regulation of the meat quality traits (te Pas et al., 2010).

Table 6.2 Genes with association to the complex trait pH of meat quality at several time points postmortem and potential biochemical pathways (KEGG database).

Gene name	Meat quality trait pH (hours postmortem)	KEGG pathways
Fatty acid-binding protein, heart (H-FABP) (Muscle fatty acid-binding protein) (Fabp3)	pH6	PPAR signaling pathway
Heat shock 70 kDa protein 1A	pH24	MAPK signaling pathway; Antigen processing and presentation
GTP-binding protein RAD (RAD1) (aph-1)	pH24	Notch signaling pathway; Cell cycle yeast

Bioinformatics Toward Systems Biology: Biological Models Toward Mathematical Models

When having taken all these steps, where are we now? The aim of the research is to monitor, control, and change complex traits in livestock in a predetermined direction. As a first step, the underlying biological mechanism was investigated at as much as possible biological levels. Using bioinformatics enables to integrate all levels into a biological model comprising biochemical (physiological) pathways, histological and morphological pathways, and networks of pathways. Furthermore, networks based upon protein–protein and protein–DNA interactions can be integrated. Where possible, clustering analysis will be used to add genes with unknown biological function to the model. Furthermore, association studies indicated the most important genes of the model. In an optimal situation, this should provide a comprehensive picture of the biology that underlies the complex trait. In this manner, understanding and insight is obtained in the regulation of the complex traits of livestock, and possibly enabling some monitoring of the trait by measuring the expression levels of a few genes of the model instead of measuring the trait. However, the picture may have been hampered by the fact that genes with very small effects and genes with unknown function are missing in the model. Nevertheless, the biological model provides a wealth of information about the regulation mechanisms of complex traits at various levels of organization of the biological system.

Our aim was to monitor, control, and change the trait in a predetermined direction. Especially the latter two aims require that the reaction of the biological model to an action from outside can be predicted toward these aims. The biological model may do so qualitatively—i.e., showing the direction of the reaction or indicating a difference between two statuses, but not quantitatively—i.e., predicting the size and rate of changes in the biological model and the extent of the effect on the complex trait.

Adding mathematical modeling adds a further dimension to the model. Mathematical modeling describes the biological model as a set of mathematical formulas. This new type of model might be called a Systems Biology model.

Building the Systems Biology Model Using the Biological Model

Systems Biology can make a mathematical model from the biological model. Usually, the mathematical model is written in a special language called Systems Biology Markup Language (SBML) (Hucka et al., 2003; Strömbäck and Lambrix, 2005). The model will consist of a set of equations. The set of equations may describe each detail of the model or may take the details as a black box and models only the input and output, and how these are related. The latter is in itself enough for an empirical representation in the model that predicts the effect of changes in input on a complex trait. However, it excludes the incorporation of future-detailed measurements.

The biological model includes biochemical (physiological) pathways. A closer look at these pathways often indicates that they consist of arrays of enzymatic reactions. The first enzyme takes a substrate and converts it into a product. The reaction can be characterized by the reaction constant (Michaelis–Menten constant (Dowd and Riggs, 1965; Cornish-Bowden, 1995)). This consists of the association and dissociation

characteristics of the enzyme for the substrate and the product. It may include information about the reaction speed. It is important to remember that all enzymatic reactions are in principle equilibrium reactions, so there is always the possibility that the product is used to form the substrate back. The expression level, activity, and properties of the enzymes also influence this equilibrium. However, usually the reaction is directed to form the product and to continue to do so by the driving forces of the association constant of the substrate and the dissociation constant of the product. Often, the concentrations of substrate and product also force the reaction in that direction. If the values for the substrate concentration are known, and the association constant of the substrate and the dissociation constant of the product are known, the flux through the enzyme can be calculated. Thus, the model is a quantitative model with predictive capacity for the reaction if the substrate concentration changes, or what happens if a mutation changes the characteristics of the molecule to substrate or product.

In a biochemical pathway, a second enzyme follows after the first enzyme. The second enzyme uses the product of the first enzyme as substrate. The same mathematical modeling can be used for the second enzyme. This argues for making premade mathematical modules of such general reaction types as enzyme reactions. Perhaps the use of such premade modules will allow not-specialized researchers to generate mathematical models using biological models as a template. One should just keep in mind that the reaction constants and the values for the substrate concentration differ for each enzyme. Furthermore, since each next enzyme uses the product of the previous enzyme as a substrate the concentrations of substrate and product of each enzyme will change constantly. This may be a complicating factor in the model since changes in the reaction conditions or changes in the enzyme molecule characteristics may affect these concentrations. It would be best (if possible) if these changes themselves could be quantified.

Another complicating factor may be that some genes/enzymes are nodes in the pathway that are the starting point of several branches. Thus, the pathway can continue in several directions, even with the possibility to affect different aspects of a trait (e.g., apoptosis trait (pathway): apoptosis or cell death), or even different traits. Especially nodal enzymes in the networks of pathways and other networks may have complex modeling constants. Here, competition for the product of the enzyme takes place. This should be modeled keeping in mind that the flux through each of the subpathways can vary differently with changing environmental or tissue conditions.

Another important factor that needs to be modeled is the substrate of the first enzyme in a pathway. The substrate may come from other reactions/pathways in the cell or from substances imported in the cell such as food products. The first option is to relate the pathway/network to other pathways/networks not involved in the biological model of the complex trait under investigation. The second option is to relate the pathway/network to couple the metabolism to a metabolic state, or some drive of that, such as food intake, digestion, or exchange with the blood system. The second approach allows an integration of modeling of the trait with the functional response of other body organs or some general characteristics of animal performance and physiological state (e.g., see also Chapters 8 and 9). In both cases, scientific knowledge at different levels is generated.

Alternatively, if a (sub)pathway consists of a single array of enzymes, the whole pathway can be modeled as a black box with first substrate input and last product

output. This only means that apparent reaction characteristics at the level of the whole box have been modeled. The specific aim of this is to determine which details have to be modeled and which details can possibly be excluded. A disadvantage of this type of modeling is the lack of detail and less specific explanatory power or relationships to the detailed information at the molecular level. On the other hand, the model becomes simpler, which adds to the comprehension of the model and testing of hypothesis on how to affect a trait or animal response. Although simplified, such a model may explain more of the regulatory mechanisms of a complex trait due to its easy understandability (Boer et al., 2010). Furthermore, the role of genes without known physiological function added to the model, identified to be relevant by cluster analysis, may be placed inside a black box by an empirical representation instead of a mechanistic one.

The mathematical model needs to focus on genes whose variabilities were shown to be of higher importance in the biological model. If the mathematical model can support the role of this gene or provide an explanation for its role in the complex trait investigated, this will strengthen the central position of the gene in the biological model.

Mathematical models written in SBML or BIOPAX (Strömbäck and Lambrix, 2005) are available for the pathways in several databases. As an example of the type of available models, Figure 6.5 shows a part of the glycolysis pathway of the KEGG database (written in BIOPAX) and represent an enzymatic reaction, and a part of the WNT pathway from the Reactome database (written in SBML) representing protein complex formation.

From the available downloads it is not always clear which variables can be changed and how predictive the model actually is. However, these mathematical models can certainly be used as starting template to integrate into each other and integrate with other mathematical models and aid in the building of a new mathematical model of a complex trait. Furthermore, the mathematical models of networks of genes or proteins created with softwares like Cytoscape and Osprey should be made similarly using the specific network as a template. However, it should be noted that also for these networks mathematical modules are under development and visualized networks will be converted to mathematical models within the software in the near future.

The Mathematical Model: From Pathways and Networks to Model of the Complex Trait

If all the pathways and networks of reactions, and morphological structures are included in a biological model, a mathematical model can be constructed that predicts the reaction of the system to disturbances, e.g., environmental changes, or genetic mutations that change characteristics of specific molecules and structures. A mathematical model of the complex trait should integrate all the mathematical descriptions of individual pathways derived from biological models (e.g., Figure 6.4). The model can be fed with changing concentrations of substrates or changing molecular characteristics affecting biochemical reactions or morphological structures leading to a prediction of the changes to the regulatory mechanism. This will lead to a changed phenotype of the complex trait.

(A)

```
DB
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">KEGG</bp:
DB>
</bp:unificationXref>
p:conversion rdf:ID="600" />
p:biochemicalReaction rdf:ID="603">
p:XREF rdf:resource="#604" />
p:LEFT rdf:resource="#606" />
p:RIGHT rdf:resource="#607" />
</bp:biochemicalReaction>
p:unificationXref rdf:ID="604">
p:ID
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">rn:R00710
</bp:ID>
p:DB
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">KEGG</bp:
DB>
</bp:unificationXref>
```

(B)

```
pecies name="Skp1" compartment="c70101" id="s174150">
nnotation>
CT:physicalEntityClass>SEQUENCE</RCT:physicalEntityClass>
CT:content>
CT:speciesID>s174150</RCT:speciesID>
CT:sequenceStart>1</RCT:sequenceStart>
CT:sequenceEnd>-1</RCT:sequenceEnd>
</RCT:content>
</annotation>
</species>
pecies name="SCF-beta-TrCP1 complex" compartment="c70101"
id="s174155">
nnotation>
CT:physicalEntityClass>COMPOSITE COMPLEX</RCT:physicalEntityClass>
CT:content>
CT:speciesID>s174230</RCT:speciesID>
</RCT:content>
CT:content>
CT:speciesID>s174237</RCT:speciesID>
</RCT:content>
CT:content>
CT:speciesID>s174150</RCT:speciesID>
</RCT:content>
CT:PubmedID>10023660</RCT:PubmedID>
</annotation>
</species>
```

Figure 6.5 Mathematical models of an enzymatic reaction (A) and of a protein complex formation (B). (A): Part of the mathematical model of the Glycolysis/Gluconeogenesis pathway of KEGG; (B): Part of the WNT signaling pathway of Reactome.

The model predictions can be evaluated by experimental data. The result of such evaluations studies identifies imperfections in the model and guides in the decision of what must improve. This cycle can be repeated as long as necessary, until a mathematical model is obtained that predicts satisfactorily the effects of interest.

The Mathematical Model, the Biological Model, and the Complex Trait

In the end, if the mathematical model is optimal but the predictions of the changes in the complex trait remains suboptimal, this is an indication that the biological model is not complete. Repeating the whole procedure can improve the biological model as well as the mathematical model and systems biological model. The outcome of such a cycle of experimentation and mathematical modeling first of all delivers scientific knowledge about the biology underlying a complex trait, but also involves the integration of scientific disciplines and application of results. Perhaps, more importantly, the knowledge generates the possibility to change a complex trait in a predetermined direction because the effect of intended actions to change the complex trait can be calculated before starting the experiment.

Future Expectations

In theory, the constructed mathematical model describes the complex trait now as a kind of Mealy machine (Mealy, 1955). In computation theory, a Mealy machine is a finite state transducer using an adaptable input (e.g., substrate concentration) and a current state (e.g., the system with the alleles as it is) to produce an output (e.g., the biological regulatory mechanism leading to a complex trait phenotype). For each change in input and current state, a different output will be predicted. Thus, the complex trait phenotype can be monitored, controlled, and changed into a predetermined direction now. This will enable real balanced breeding to improve animal welfare and productivity simultaneously taking into account the dynamics and variation in the mechanisms involved with the effects exerted by physiological, genetic, and environmental factors (nutrition, management), including the representation of historic effects. This also improves sustainability of livestock production systems. This was a major aim of Systems Biology for livestock science.

Acknowledgments

The authors like to thank their colleagues in the Animal Breeding and Genomics Centre for helpful discussions. The work on pathways and Systems Biology was mainly financed by the grants KB-04–004-012 and KB-04–004-049 from the Dutch Ministry of Agriculture, Nature Conservation, and Food Quality under the Basic Knowledge program “Kennisbasis.”

References

Andersson, L. & Georges, M. (2004) Domestic-animal genomics: deciphering the genetics of complex traits. *Nature Reviews Genetics* **5**, 202–212.

Ashburner, M., Ball, C.A., Blake, J.A., et al. (2000) The Gene Ontology Consortium Gene Ontology: tool for the unification of biology. *Nature genetics* **25**, 25–29.

Barrell, D., Dimmer, E., Huntley R.P., et al. (2008) The GOA database in 2009-an integrated Gene Ontology Annotation resource. *Nucleic Acids Research* **37**, D396–D403.

Baudier, J., Bergeret, E., & Bertacchi, N. (1995) Interactions of Myogenic bHLH Transcription Factors with Calcium-Binding Calmodulin and S100a (.alpha..alpha.) Proteins. *Biochemistry* **34**, 7834–7846.

Boer, H.M.T., Veerkamp, R.F., Beerda, B., et al. (2010) Estrous behavior in dairy cows: identification of underlying mechanisms and gene functions. *Animal* **4**, 446–453.

Bovine Genome Sequencing and Analysis Consortium, Elsik C.G., Tellam R.L., Worley K.C., et al. (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**, 522–528.

Breitkreutz, B-J., Stark, C., & Tyers, M. (2003) Osprey: a network visualization system. *Genome Biology* **4**, R22 http://genomebiology.com/2003/4/3/R22

Burnett, G. & Kennedy, E.P. (1954). The enzymatic phosphorylation of proteins. *Journal of Biological Chemistry* **211**, 969–980.

Cagnazzo, M., te Pas, M.F.W., Priem, J., et al. (2006) Comparison of prenatal muscle tissue expression profiles of two pig breeds differing in muscle characteristics. *Journal of Animal Science* **84**, 1–10.

Cardon, L.R. & Bell, J.R. (2001) Association study designs for complex diseases. *Nature Reviews Genetics* **2**, 91–99.

Chesler, E.J., Lu, L., Shou, S., et al. (2005) Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nature Genetics* **37**, 233–242.

Cline, M.S., Smoot, M., Cerami, E. et al. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nature Protocols* **2**, 2366–2382.

Cornish-Bowden, A. (1995) *Fundamentals of Enzyme Kinetics* 3rd edition. Portland Press, London.

Dennis, G., Sherman, B.T., Hosack, D.A., et al. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology* **4**, P3. Available on: http://genomebiology.com/2003/4/9/R60.

Devaux, F., Marc, P., & Jacq, C. (2001) Transcriptomes, transcription activators and microarrays. *FEBS Letters* **498**, 140–144.

Diehn, M. & Relman, D.A. (2001) Comparing functional genomic datasets: lessons from DNA microarray analyses of host–pathogen interactions. *Current Opinion in Microbiology* **4**, 95–101.

Dixon, A.L., Liang L., Moffatt, M.F., et al. (2007) A genome-wide association study of global gene expression. *Nature Genetics* **39**, 1202–1207.

Dowd, J.E. & Riggs, D. (1965) A comparison of estimates of Michaelis-Menten kinetic constants from various linear transformations. *Journal of Biological Chemistry* **240**, 863–870.

Eisen, M.B., Spellman, P.T., Brown, P.O., et al. (1998) Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the USA* **95**, 14863–14868.

Freimer, N. & Sabbati, C. (2003) The human phenome project. *Nature Genetics* **34**, 15–21.

Gene Ontology Consortium, Harris, M.A., Clark, J., Ireland, A., et al. (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research* **32** (Database issue), D258–D261.

Goldspink, G. (1995) Adaptation of fish to different environmental temperature by qualitative and quantitative changes in gene expression. *Journal of Thermal Biology* **20**, 167–174.

Guido, M.E., de Guido, L.B., Goguen, D., et al. (1999) Daily Rhythm of Spontaneous Immediate-Early Gene Expression in the Rat Suprachiasmatic Nucleus. *Journal of Biological Rhythms* **14**, 275–280.

Guo, S.W. & Thompson, E.A. (1992) Monte Carlo method for combined segregation and linkage analysis. *American Journal of Human Genetics* **51**, 1111–1126.

Heber, S., Herms, J., Gajic, V., et al. (2000) Mice with Combined Gene Knock-Outs Reveal Essential and Partially Redundant Functions of Amyloid Precursor Protein Family Members. *The Journal of Neuroscience* **20**, 7951–7963.

Hu, J., Mungall, C., Law, A., et al. (2001) The ARKdb: genome databases for farmed and other animals. *Nucleic Acids Research* **29**, 106–110.

Huang, D.W., Sherman, B.T., & Lempicki, R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**, 1–13.

Hucka, M., Finney, A., Sauro, H.M., et al. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531.

Janss, L.L.G., Thompson, R., & van Arendonk, J.A.M. (1995) Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations. *Theoretical and Applied Genetics* **91**, 1137–1147.

Janss, L.L.G., van Arendonk, J.A.M., & Brascamp, E.W. (1997) Bayesian statistical analyses for presence of single major genes affecting meat quality traits in crossed pig population. *Genetics* **145**, 395–408.

Kadarmideen, H.N. & Reverter, A. (2007) Combined genetic, genomic and transcriptomic methods in the analysis of animal traits. *Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources* **2**, 042, 16 p. doi:10.1079/PAVSNNR20072042. Available on: http://www.cababstractsplus.org/cabreviews.

Kanehisa, M. & Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27–30.

Kanehisa, M., Araki, M., & Goto, S. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Research* **36**, D480–D484.

Kanehisa, M., Goto, S., Hattori, M., et al. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research* **34**, D354–D357.

Kim, V.N. (2005) MicroRNA biogenesis: Coordinated cropping and dicing. *Nature Reviews Molecular Cell Biology* **6**, 376–385.

Laursen, L. (2009) Biological logic. *Nature* **462**, 408–410.

Lawless, J.F. (2003) *Statistical Models and Methods for Lifetime Data*, 2nd Edn. Wiley-Interscience 630 p. (Wiley series in probability and statistics) New York: Wiley Hoboken, N.J.

Lie, D.C., Colamarino, S.A., Song, H.J., et al. (2005) Wnt signalling regulates adult hippocampal neurogenesis. *Nature* **437**, 1370–1375.

Lohmueller, K.E., Pearce, C.L., Pike, M., et al. (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genetics* **33**, 177–182.

Mann, M. & Jensen, O.N. (2003) Proteomic analysis of post-translational modifications. *Nature biotechnology* **21**, 255–261.

Mealy, G.H. (1955) A method to synthesizing sequential circuits. *Bell Systems Technical Journal* **34**, 1045–1079.

Merico, D., Gfeller, D., & Bader, G.D. (2009) How to visually interpret biological data using networks. *Nature Biotechnology* **27**, 921–924.

Müller, M. & Kersten, S. (2003) Nutrigenomics: Goals and Perspectives. *Nature Reviews Genetics* **4**, 315–322.

Nijhout, H.F. (2002) Genetic regulatory networks. In: *Encyclopedia of Evolution*. Oxford University Press, Oxford.

Nijhout, H.F., Reed, M.C., & Ulrich, C.M. (2007) A day in the life of cell metabolism. *Biological Theory* **2**, 124–127.

Przybylski, R.J., Szigeti, V., Davidheiser, S., et al. (1994) Calcium regulation of skeletal myogenesis. II. Extracellular and cell surface effects. *Cell Calcium* **15**, 132–142.

Shannon P., Markiel A., Ozier O., et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**, 2498–2504.

Strömbäck, L. & Lambrix, P. (2005) Representations of molecular pathways: an evaluation of SBML, PSI MI and BioPAX. Representations of molecular pathways: an evaluation of SBML, PSI MI and BioPAX. *Bioinformatics* **21**, 4401–4407.

Thaller, G., Dempfle, L., & Hoeschele, I. (1996) Maximumlikelihood analysis of rare binary traits under different modes of inheritance. *Genetics* **143**, 1819–1829.

Thomson, G. (1994) Identifying complex disease genes: progress and paradigms. *Nature Genetics* **8**, 108–110.

te Pas, M.F.W. & Soumillion, A. (2001) The use of physiologic and functional genomic information of the regulation of the determination of skeletal muscle mass in livestock breeding strategies to enhance meat production. *Current Genomics* **2**, 285–304.

te Pas, M.F.W., de Wit, A., Priem, J., et al. (2005) Transcriptome expression profiles in prenatal pigs in relation to myogenesis. *Journal of Muscle Research and Cell Motility* **26**, 157–165.

te Pas, M.F.W., Hulsegge, I., Pool, M.H., et al. (2007) Biochemical pathways analysis of microarray results: Regulation of myogenesis. *BMC Developmental Biology* **7**, 66. doi:10.1186/1471-213X-7-66.

te Pas, M.F.W., Jansen, J., Broekman, K.C.J.A., et al. (2009) Post mortem proteome degradation profiles of longissimus muscle in Yorkshire and Duroc pigs and its relationship with pork quality traits. *Meat Science* **83**, 744–751.

te Pas, M.F.W., Keuning, E., Hulsegge, B., et al. (2010) Longissimus muscle transcriptome profiles related to carcass and meat quality traits in fresh meat Pietrain carcasses. *Journal of Animal Science* **88**, 4044–4055.

te Pas, M.F.W., Pool, M.H., Hulsegge, I., et al. (2006) Analysis of the differential transcriptome expression profiles during prenatal muscle tissue development. *Archives of Animal Breeding Dummerstorf*, special issue **49**, 110–115.

te Pas, M.F.W., van Hemert, S., Hulsegge, I., et al. (2008) A pathways analysis tool for analyzing microarray data of species with low physiological information. *Advances in Bioinformatics*. ID 719468; doi:10.1155/2008/719468.

Van Hemert, S., Hoekman, A.J.W., Smits, M.A., et al. (2006a) Gene expression responses to a Salmonella infection in the chicken intestine differ between lines. *Veterinary Immunology and Immunopathology* **114**, 247–258.

Van Hemert, S., Hoekman, A.J.W., Smits, M.A., et al. (2006b) Early host gene expression responses to a Salmonella infection in the intestine of chickens with different genetic background examined with cDNA and oligonucleotide microarrays. *Comparative Biochemistry and Physiology* **1** (Part D), 292–299.

Vazquez, A., Flammini, A., Maritan, A., et al. (2003) Global protein function prediction in protein-protein interaction networks. *Nature Biotechnology* **21**, 697–700.

Wallis, J.W., Aerts, J., Groenen, M.A.M., et al. (2004) A physical map of the chicken Genome. *Nature* **432**, 761–764.

Wimmers, K., Murani, E., te Pas, M.F.W., et al. (2007) Associations of functional candidate genes derived from gene expression profiles of prenatal porcine muscle tissue with meat quality and carcass traits. *Animal Genetics* **38**, 474–484.

Yoshizawa F., Nagasawa T., Nishizawa N., et al. (1997) Protein synthesis and degradation change rapidly in response to food intake in muscle of food-deprived mice. *Journal of Nutrition* **127**, 1156–1159.

Chapter 7
Molecular Networks as Sensors and Drivers of Uterine Receptivity in Livestock

Stefan Bauersachs and Eckhard Wolf

Introduction

Successful reproduction depends on a cascade of biological processes, including maturation and selection of gametes, fertilization, pre- and postimplantation embryonic development, fetal growth, and birth. Importantly, biological filters involved in the various steps of the reproduction cascade may be overcome by assisted reproduction techniques (ART) that are increasingly used in humans and animals. In dairy cattle, a reduction in fertilization and embryonic survival rates has been suggested as the most important component for decreasing reproductive efficiency (Santos et al., 2004). A prerequisite for a successful pregnancy is a functional embryo–maternal communication to facilitate establishment, recognition, and maintenance of pregnancy. Via paracrine signals, the conceptus prepares its environment, the uterine endometrium, for attachment and implantation. Thus, analyzing the dynamic responses of the endometrium to a conceptus is a powerful approach to (i) identify biological processes that are stimulated or suppressed in the endometrium by the conceptus, and (ii) evaluate the quality of embryos regarding their ability to elicit physiological responses of the surrounding endometrium. Interestingly, the mechanisms of pregnancy recognition show in many aspects differences between mammalian species (Bazer et al., 2009). Therefore, the comparative analysis of common and species-specific mechanisms involved in preparation of the endometrium for implantation of the conceptus provides a unique opportunity for dissecting phylogenetically conserved and distinct pathways involved in this pivotal step of reproductive biology. Holistic and sensitive Omics-technologies characterizing the transcriptome, proteome, metabolome, etc., of cells or tissues facilitate the comprehensive description of molecular patterns of gametes, embryos, and their maternal environment. Importantly, dynamic changes of these patterns during development may point to genes or pathways that have an effect on reproductive success. Thus, molecular patterns identified by Omics-technologies can be viewed as "intermediate phenotypes" (Schadt, 2009), whose comprehensive description, interpretation, and modeling may help to understand the genetic basis of cellular functions that are important for fertility (Figure 7.1).

Systems Biology and Livestock Science, First Edition. Edited by Marinus F.W. te Pas, Henri Woelders, and André Bannink.

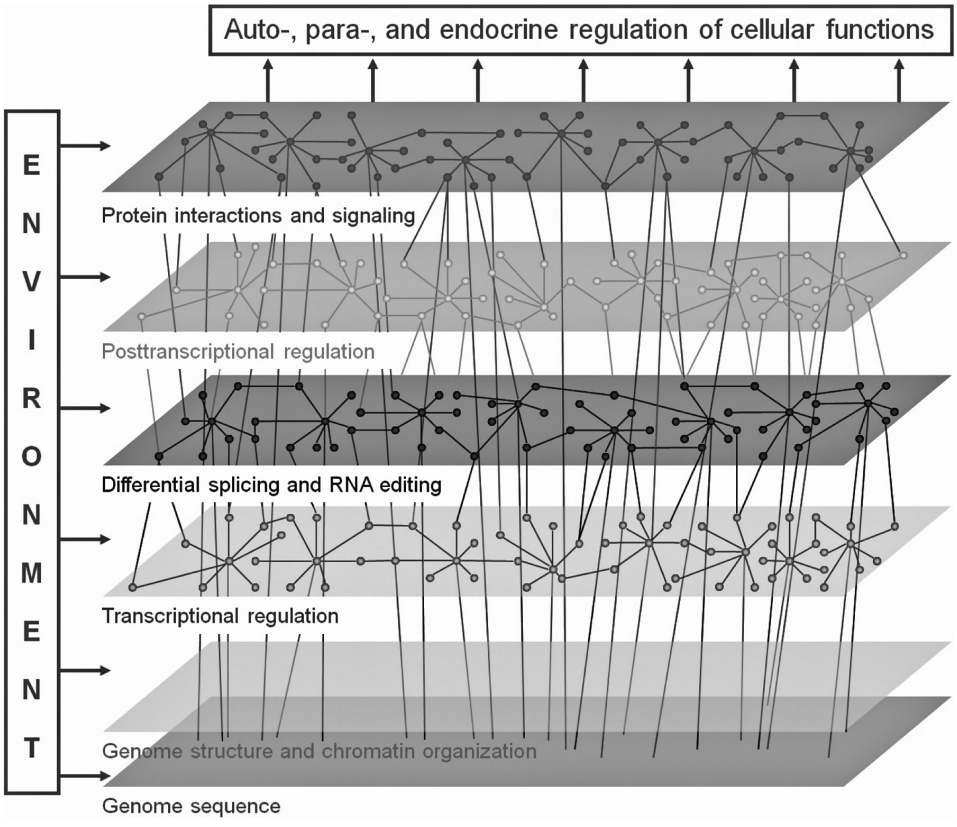


Figure 7.1 Flow of genetic information via different classes of molecules producing molecular patterns and networks, which affect cellular and organ functions. Integrating large-scale, high-dimensional molecular and physiological data holds promise for defining the molecular networks that respond to genetic and environmental perturbations of the physiological functions. The different layers of information provide a hierarchy of intermediate phenotypes, RNA being the most proximal non-DNA species of all molecular entities in the cell. Complex epigenomic mechanisms and interactions between the different classes of molecules modulate the flow of genetic information into biological functions.

Transcriptome Analysis as a Holistic Approach for the Study of Cellular Changes at the Molecular Level

Mammalian genomes contain approximately 20,000–22,000 protein-coding genes. The number of individual transcripts encoded by mammalian genomes is significantly higher due to transcript variants arising from the same gene and the fast growing world of noncoding RNAs that have structural and regulatory functions (Gustincich et al., 2006; Licatalosi and Darnell, 2010; Lindberg and Lundeberg, 2010; Orom et al., 2010). Furthermore, all these RNAs existing in a given tissue occur in very different abundances (Gustincich et al., 2006; Carninci et al., 2008). Since most physiological processes are accompanied by complex changes in the RNA profile, transcriptome analyses are a powerful approach for a system-wide description of cellular changes

at the molecular level. However, changes at the mRNA levels do not automatically predict similar changes at the protein level. To address the proteins themselves, as the functional players in the cell, would be the ideal way. However, this is limited due to the considerably larger differences in abundances and the extremely diverse chemical properties of individual proteins making them only partially accessible for current proteome analysis techniques (mass spectrometry and two-dimensional gel electrophoresis). In contrast, various analytical approaches have been developed to comprehensively profile mammalian transcriptomes (Stanton, 2001; Hoheisel, 2006; Bauersachs et al., 2008; Wang et al., 2009). Currently, the most powerful technologies are hybridization-based (DNA microarrays) or sequencing-based (RNA-Seq), both are able to generate comprehensive genome-wide expression profiles. The most widespread approach for the analysis of transcriptome changes is still the microarray technology. However, the RNA-Seq technology provides much more information on absolute transcript levels, transcript variants, and currently not annotated transcribed regions and is used more and more in different biological applications (Wang et al., 2009; Marguerat and Bahler, 2010). Due to their high performance, RNA-Seq technologies are particularly suited for the analysis of mammalian transcriptomes and will enable the detection of rare transcripts in complex tissues, such as the endometrium. For domestic animals, the next-generation sequencing technologies will also be helpful to improve the current gene annotation, to define the entire transcriptome, and finally provide sequence information for the design of comprehensive genome-wide microarrays. A comparison of the results derived from an RNA-Seq study and an Affymetrix microarray study of bovine endometrium at day 18 of pregnancy is shown in Table 7.1 (our unpublished results). The same RNA samples were used for both studies. The comparison revealed a consistent overlap between the results but many more differentially expressed genes (DEGs) for the RNA-Seq data.

In humans, the microarray technology has been used for the analysis of the endometrium transcriptome during the window of implantation and has provided remarkable insight into endometrial maturation and implantation (Giudice, 2004). Also, in domestic species such as sheep, cattle, swine, and the horse, several microarray studies have been performed on different reproductive tissues (Bauersachs et al., 2008; Evans et al., 2008; Spencer et al., 2008; Satterfield et al., 2009; Merkl et al., 2010;

Table 7.1 Comparison of Illumina RNA-Seq and Affymetrix GeneChip data derived from the analysis of bovine endometrium at day 18 of pregnancy.

	Number of genes	%
RNA-Seq DEGs (FDR 1% FC $\geq$ 2)	664	
Not represented on Affymetrix array	218	32.8
Not detectable with Affymetrix	37	5.6
Differential with Affymetrix	320	78.2
Affymetrix DEGs (FDR 1% FC $\geq$ 2)	336	
Contained in annotation used for RNA-Seq	297	88.4
Found as differential with RNA-Seq	278	93.6

DEGs (differentially expressed genes)
FDR (false discovery rate)
FC (fold change)

Ostrup et al., 2010) to characterize regulatory processes underlying the establishment and maintenance of pregnancy.

Resources for Functional Gene Annotation and Gene/Protein Interactions and Corresponding Analysis Tools

With the beginning of the systematic analysis of genomes, molecular functions were assigned to the newly identified genes. To provide a defined functional description and classification of genes, the Gene Ontology (GO) project has developed three structured, controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components, and molecular functions in a species-independent manner (Ashburner et al., 2000). Genes are assigned based on data from the literature, on belonging to a known protein family, but also merely based on the presence of conserved protein domains. There are numerous tools for the analysis of GO terms associated with a list of differentially expressed transcripts or proteins (see Table 7.2). Many of these tools provide quantitatively enriched GO terms associated with a gene list, i.e., GO terms for which significantly more associated genes were found than expected by chance. Although most of the annotated genes are assigned to GO categories, gene functions for large animals are mostly inferred from functions, which were only experimentally validated for the classical model organisms (Table 7.3). Therefore, the results of GO annotations for large animals need careful interpretation.

The processing of the results of such analyses can be very laborious due to the redundant structure of the GO categories. The "functional annotation clustering" tool of the database for annotation, visualization, and integrated discovery (DAVID) (Dennis et al., 2003; Huang et al., 2009) avoids this problem by clustering enriched functional categories that have overlapping gene contents. Furthermore, the DAVID "functional annotation clustering" tool uses information from a variety of databases, including GO data, KEGG pathways, BioCarta pathways, but also data from completely different sources, e.g., protein databases (SP PIR keywords), disease databases (OMIM Disease), protein domain (Interpro) and protein interaction databases (BIND), and others.

Another strategy for obtaining information on gene functions and interactions is text mining of MEDLINE abstracts (Clegg and Shepherd, 2008). The analysis tool CoPub (Frijters et al., 2008) combines text mining with the analysis of lists of DEGs or proteins to find gene–gene cocitation and cocitation of genes with keywords. This analysis identifies biological or disease-related keywords overrepresented within the DEGs and can be used as complement to GO analyses or as a basis for the construction of gene interaction networks.

Microarray datasets can also be characterized by comparison with gene sets derived from other gene expression studies or from defined functional categories by the use of the "gene set enrichment analysis" (GSEA) tool (Subramanian et al., 2005). GSEA compares a gene expression dataset with different collections of gene sets: positional gene sets, curated gene sets, motif gene sets, computational gene sets, and GO gene sets (for detailed explanation, see www.broadinstitute.org/gsea/msigdb/index.jsp). The genes of an expression dataset are ranked according to differential expression with the most significantly upregulated genes at the top and the most significantly

Table 7.2 Selected tools for the bioinformatic analysis of Omics-data.

Tool	Function	Web site/source
Gene Ontology (GO)	Provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data	geneontology.org
GO tools for analysis of gene expression/microarray datasets	Overview of tools for GO analysis of lists of differentially expressed genes/mRNAs or proteins	geneontology.org/GO.tools.microarray.shtml#goarray
Database for annotation, visualization, and integrated discovery (DAVID)	Functional annotation, functional classification, gene ID conversion, gene name batch viewer	david.abcc.ncifcrf.gov
CoPub text mining tool	Text mining tool for detection of co-occurring biomedical concepts in abstracts from the Medline literature database significantly linked to a differential gene set	services.nbic.nl/cgi-bin/copub3/CoPub.pl
Gene set enrichment analysis (GSEA)	Computational method that determines whether an *a priori* defined set of genes shows statistically significant, concordant differences between two biological states (e.g., phenotypes)	www.broadinstitute.org/gsea
KEGG pathway database	Collection of manually drawn pathway maps	www.genome.jp/kegg/pathway.html
BioCarta pathways	Open source for gene interaction networks/molecular pathways	http://www.biocarta.com/genes/index.asp
NCI/Nature pathway interaction database	Biomolecular interactions and cellular processes assembled into authoritative human signaling pathways	pid.nci.nih.gov/index.shtml

(*Continued*)

Table 7.2 (*Continued*)

Tool	Function	Web site/source
oPOSSUM	Detection of overrepresented transcription factor binding sites (TFBS) in the promoters of sets of genes	www.cisreg.ca/oPOSSUM
BioGRID (Biological general repository for interaction datasets)	Online interaction repository with data compiled through comprehensive curation efforts	http://thebiogrid.org
HPRD (Human protein reference database)	Centralized platform to visually depict and integrate information pertaining to domain architecture, posttranslational modifications, interaction networks, and disease association for each protein in the human proteome	www.hprd.org
BIND/BOND	Biomolecular interaction network database/ Biomolecular object network databank	http://bond.unleashedinformatics.com
Reactome	Online, open-source, curated pathway database encompassing many areas of human biology	www.reactome.org
STRING	Database and analysis tool for known and predicted protein–protein interactions	http://string.embl.de
Cytoscape	Open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data	www.cytoscape.org

Table 7.3 Percentage of annotations not inferred from electronic annotation (IEA) for selected model organisms.[a]

Species	Database	Gene products annotated	GO annotations	Non-IEA	(%)
Bos taurus	GOA@EBI	22,440	110,068	7,270	6.6
Caenorhabditis elegans	WormBase	17,688	114,213	59,122	51.8
Danio rerio	ZFIN	15,566	110,114	24,313	22.1
Drosophila melanogaster	FlyBase	12,747	76,970	61,333	79.7
Gallus gallus	GOA@EBI	18,185	79,614	3,316	4.2
Homo sapiens	GOA@EBI	18,410	222,229	124,324	55.9
Mus musculus	MGI	33,908	280,811	184,682	65.8
Rattus norwegicus	RGD	28,808	250,004	128,233	51.3

[a] Status: 12/2010.

downregulated genes at the bottom. On the basis of the positions of the genes of the gene sets in the ranked gene expression dataset, enrichment toward one end of the ranked list is calculated that indicates concordance of the gene set with the gene expression dataset. User-provided gene sets can also be used for comparison with the expression dataset. GSEA results can be helpful, for example, for drawing conclusions from regulatory mechanisms that are known for a given gene set or from gene sets that belong to defined functional categories or cellular pathways.

The tools described above reveal groups of genes associated with a certain common biological function. This does not automatically mean that these genes/proteins have direct interactions or are regulated by the same regulatory factor. With regard to the identification of a common gene regulation underlying observed gene expression changes, several tools (GSEA, DAVID, oPOSSUM (Ho Sui et al., 2007)) look for transcription factor binding sites in the promoter regions of the DEGs to find transcription factors whose binding sites are overrepresented. Unfortunately, these tools are based on known human, mouse, and rat promoter regions, i.e., the analysis in other species assumes conserved regulatory elements in the corresponding promoter regions. Nevertheless, a comparative study of human and bovine transcription factor binding sites (TFBS) encourages the use of human promoter databases for the inference of bovine gene regulation (Zadissa et al., 2007).

In order to better understand observed gene expression changes in the context of complex cellular processes, potential interactions between the identified genes or proteins and with other genes or proteins are analyzed and visualized. These analyses are based on data from different interaction databases/resources, e.g., BIND, HPRD, BioGRID (Table 7.1). Also, protein interactions and regulations can be obtained with the help of natural language preprocessing tools (Fundel et al., 2007). There are a number of open source tools available for interaction and pathway analysis, e.g., Reactome (Matthews et al., 2009), Cytoscape (Cline et al., 2007), and STRING (Jensen et al., 2009). The latter tool is a searchable database for known and predicted protein–protein interactions and also provides network visualization for a provided set of genes.

Identification of Biological Themes Related to Endometrial Remodeling and Receptivity in a Microarray Study of Bovine Endometrium During the Estrous Cycle

Endometrial gene expression is mainly regulated by the complex interplay of the ovarian steroid hormones such as estradiol and progesterone (Goff, 2004; Spencer et al., 2004). They act via the classical nuclear steroid hormone receptors, but also via nonclassical receptors such as progesterone receptor membrane component 1 and the novel family of membrane progestin receptors (Gellersen et al., 2009). Although the basic principles of hormonal regulations in the endometrium during the estrous cycle are known, the highly complex molecular responses in the endometrium to the ovarian hormones are not completely understood. Basically, progesterone is the key hormone for preparation of the endometrium for embryo implantation and maintenance of pregnancy (Bazer et al., 2008) and genes with increased expression levels in the luteal phase are probably regulated by progesterone, directly or indirectly. The supportive role of progesterone has been confirmed in a recent study, where a positive influence of progesterone on conceptus growth and development was found (Clemente et al., 2009). To identify genes playing a role in the context of fertility, two microarray studies of bovine intercaruncular endometrium during the estrous cycle were performed and different data analysis tools were used to characterize the obtained gene sets (Bauersachs et al., 2005, 2008; Mitko et al., 2008). These studies revealed several hundred genes differentially expressed between different stages of the estrous cycle. Two major groups of genes according to their expression profiles were observed that showed either highest mRNA levels during the estrus phase or highest levels during the luteal phase, respectively, corresponding to the steroid hormone profiles during the cycle. A minor group of genes exhibited highest mRNA levels on day 3.5. GO analysis, pathway analysis, and functional classification using the DAVID "functional annotation clustering" tool were used to infer regulated biological processes from gene expression data and to identify overrepresented biological themes from the list of differentially expressed genes. In addition, interaction networks were built for genes upregulated at estrus and genes upregulated at diestrus, respectively, to visualize potential interactions between the identified genes (Mitko et al., 2008). These networks were drawn on the basis of different types of interactions like binding, regulation, expression, transport, and assignment to a protein family or a biological process. Results from both studies characterized the estrus phase by overrepresentation of genes related to the functional terms "focal adhesion formation", "cell motility", "cytoskeleton", "extracellular matrix" (ECM), "ECM remodeling", and "cell growth". The interaction network for genes upregulated at estrus showed that the ECM proteins together with cytoskeletal proteins are linked to the process "focal adhesion" (cell-matrix adhesion). Coordinated regulation of genes involved in ECM remodeling during the estrous cycle is reflected by the expression patterns of genes encoding matrix metallopeptidases and their inhibitors in addition to genes coding for ECM constituents and genes involved in regulation of ECM proteins. In addition to genes related to ECM remodeling, a number of genes with lower mRNA levels at diestrus were identified that have been described in the context of positive regulation of invasive processes. Thus, decreased levels of these mRNAs during the luteal phase

may be characteristic for the noninvasive implantation process in cattle. This is also supported by the fact that the majority of genes from the functional category "invasive growth" with higher mRNA levels during the luteal phase have been described in the context of negative regulation of invasive growth.

The processes of angiogenesis and regulation of blood flow also play an important role for endometrial remodeling during the estrous cycle and the specific functions of the endometrium. Several genes related to these processes were identified as differentially expressed during the estrous cycle, including members of the angiopoietin family, transcription factors controlling the expression of vascular endothelial growth factors (VEGF) and their receptors, and other genes involved, e.g., in endothelial differentiation and regulation of blood flow. Furthermore, elevated concentrations of mRNAs coding for a variety of proteins involved in metabolic and transport processes were found during the luteal phase. Elevated transport and metabolism during the luteal phase may indicate increased secretion of nutrients (histotroph) necessary for the development of the embryo (Allison Gray et al., 2000).

Correlation of Gene Expression Data and Data from Genome-Wide Association Studies (GWAS) Links Differential Gene Expression with Phenotypes Related to Fertility

The studies described in the previous paragraphs aimed at the identification of DEGs during the estrous cycle and inferred from this a functional role in regulation of endometrial receptivity. To further substantiate a functional role of genes differentially expressed in endometrium during the estrous cycle in relation to fertility, the results of the transcriptome studies were linked to results of genome-wide association studies (GWAS). Whereas transcriptome and also proteome studies are restricted to experimental sets of animals with limited size, GWAS are performed on the whole population, allowing correlation with phenotypic traits. Linking findings from GWAS with transcriptome and other Omics-data derived from experimental settings can identify genetic loci that are associated with particular traits and thereby may help to unravel causal genes associated with the trait of interest. A possible scenario is outlined in Figure 7.2. Useful inks between these two types of datasets can be obtained in at least two different ways. First, candidate genes derived by microarray studies can be investigated for the presence of SNPs, and their influence on particular traits can be investigated in the breeding population. Second, genome-wide screening with increasingly dense sets of genetic markers will probably yield multiple associations, and it would be favorable to identify the most interesting ones. Again, a comparison with findings from transcriptome, proteome, and other Omics-studies might help to add functional validation to the positional information derived from the genome-wide association studies but also *vice versa*.

A first attempt to combine findings from microarray studies of endometrial tissue samples (genes with higher mRNA levels at diestrus compared to estrus (Bauersachs et al., 2005; Mitko et al., 2008) and selected candidate genes) with genome-wide association studies for fertility and production traits in dairy cattle was performed in a recent study (Pimentel et al., 2011). A total of 12 traits related to fertility and

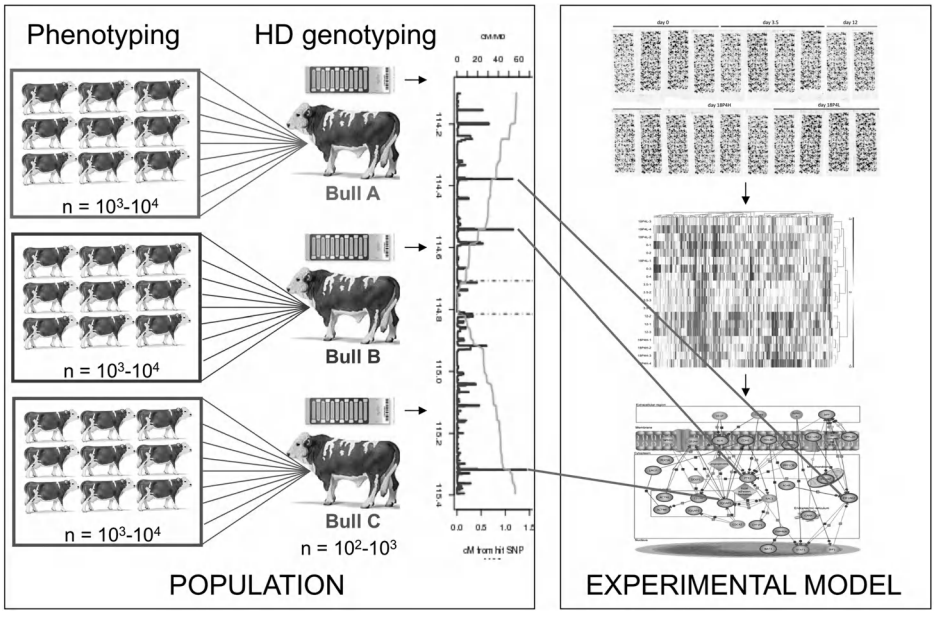


Figure 7.2 Linking data from gene expression and genome-wide association studies. (*See insert for color representation of this figure.*)

production that are typically included in national dairy cattle genetic evaluations were considered in this study. From the six fertility traits, two were heifer traits: nonreturn rate to 56 d (NRh) and interval from first to successful insemination (FLh); and four were cow traits: interval from calving to first insemination (CFc), nonreturn rate to 56 d (NRc), interval from first to successful insemination (FLc) and days open (DOc). Production traits were as following: milk yield (Mkg), fat yield (Fkg), protein yield (Pkg), fat percentage (Fpr), and protein percentage (Ppr). Another functional trait considered was somatic cell score (SCS). A set of 2294 Holstein–Friesian bulls genotyped for 39,557 single nucleotide polymorphisms (SNPs) was used, and a total of 111 SNPs were located on chromosomal segments harboring a candidate gene. Many of the SNP effects provided evidence for the antagonistic relationship between production and fertility at the molecular level. Interestingly, with respect to cattle breeding, five SNP alleles had favorable effects on yield and percentage traits, as well as on at least one fertility trait. For example, for the identified SNP within the tumor necrosis factor (ligand) superfamily, member 10 gene (*TNFSF10, TRAIL*) one allele had a positive effect on Fkg, Pkg, Fpr, and FLc. The TNFSF10 mRNA has been shown to be upregulated in human endometrium during the window of implantation (Riesewijk et al., 2003), in equine endometrium at day 12 of pregnancy (Merkl et al., 2010), and in bovine endometrium at day 18 of pregnancy (Bauersachs et al., 2006). Furthermore, a role of TNFSF10 in the modulation of the cytokine milieu at the implantation site has been suggested on the basis of the differential regulation of cytokines and chemokines in human endometrial stromal cells by TNFSF10 (Fluhr et al., 2009). For the gene interferon induced with helicase C domain 1 (*IFIH1*), specific allele increased Mkg and also the estimated breeding value (EBV) for FLc,

i.e., resulted in a reduction in the interval from first to successful insemination. *IFIH1* codes for an RNA helicase known to be involved in cellular recognition of RNA viruses (Wilkins and Gale, 2010). Song et al. (2007) reported that *IFIH1* is involved in the establishment of uterine receptivity to the conceptus during implantation in sheep. Effects in the same direction on Mkg, NRh, and FLh were found for a SNP located within the gene for insulin-like growth factor binding protein 7 (IGFBP7). Gene expression studies in other species underline the importance of this gene during early pregnancy. In equine endometrium, higher mRNA levels of IGFBP7 were found in samples derived from day 12 of pregnancy compared to nonpregnant samples (Merkl et al., 2010). Furthermore, abundant expression of IGFBP7 has been found in human glandular epithelial cells during the secretory phase and an *in vitro* knockdown experiment revealed a role of IGFBP7 protein in differentiation of these cells (Kutsukake et al., 2010). In a study of human endometrium during the menstrual cycle, an increase in expression during the receptive phase compared with the prereceptive phase, followed by a sharp increase in the late luteal phase, was found, suggesting an implication of IGFBP7 in endometrial physiology and receptivity (Dominguez et al., 2003). *APBA* APP-binding family A (*APBA*) is the gene for the X11α member of the X11 multidomain protein that is primarily expressed in neurons (reviewed in Rogelj et al., 2006). Knockout mice lacking functional *Apba1* expression show normal fertility, but a slight reduction in body weight gain (Mori et al., 2002; Ho et al. 2003), indicating a potential link of *Apba1* to production traits in livestock.

In general, the coupling of GWAS results to transcriptomic data can be a first step to a better understanding of the biology and mechanisms associated with SNPs (or genome fragments), which contribute to the improvement of specific traits. With regard to genomic selection, insights into the biological mechanisms underlying particular traits could help to search more directly towards preventing trade-offs. Gene expression data could be seen as novel phenotypes making genomic selection possible also for other than the currently accessible traits. The other way around, the link to GWAS data can help to confirm the functional role of genes in context of fertility in addition to their differential expression in the endometrium during the estrous cycle or early pregnancy.

Identification of Genes Involved in Preparation of the Bovine Endometrium for Embryo Implantation

In contrast to primates and rodents, in ruminants the time of implantation is late (in cattle after day 18 of gestation), when the trophoblast layer of the conceptus is elongated and fills out the entire pregnant uterine horn, and an epitheliochorial placenta is formed through a relatively noninvasive placentation process. Interferon tau (IFNT) has been identified as the embryonic pregnancy recognition signal in ruminants, which prevents the induction of luteolysis, thus enabling the establishment and maintenance of pregnancy (Bazer et al., 1997). This is mediated by the suppression of the genes for estrogen receptor-alpha (*ESR1*) and oxytocin receptor (*OXTR*), which prevents the pulsatile secretion of luteolytic prostaglandin F2 alpha ($PGF_{2\alpha}$) resulting in the maintenance of the ovarian corpus luteum and progesterone production (Spencer and

Bazer, 1996). In cattle, maximum secretion of IFNT was observed on day 17 (Bazer et al., 1997) in parallel to the time of maternal recognition of pregnancy. To get more detailed insights into the gene expression changes during the preimplantation period in bovine endometrium in response to the presence of a conceptus, transcriptome analyses were performed comparing samples recovered from day 18 pregnant animals and corresponding nonpregnant controls. In the first study, monozygotic twin cows were used, where one twin received two *in vitro*-produced embryos and the corresponding twin a sham transfer (Klein et al., 2006). Endometrial tissue samples were recovered on day 18 of pregnancy and the estrous cycle, respectively. Eighty-seven different genes were identified as upregulated in pregnant animals by the use of a combination of subtracted cDNA libraries and cDNA microarrays. GO analysis revealed that almost half of the obtained genes have been described as classical type I interferon-induced genes and could be directly assigned to the effects of IFNT, the embryonic pregnancy recognition signal in ruminants, on the endometrium. In the second experimental model, pregnancy was obtained by artificial insemination of heifers and control animals received a sham insemination (Bauersachs et al., 2006). Endometrial tissue samples were recovered on day 18 of pregnancy and the estrous cycle, respectively. In contrast to the first model, the control animals showed low serum progesterone levels due to a shortened estrous cycle. Thus, the differential expression of the identified genes was a mixed effect of embryonic signals and different steroid hormone levels. In this study, 179 DEGs were found, 109 with higher and 70 with lower mRNA abundance in pregnant animals. Similar to the first study, many mRNAs with higher abundance in pregnant animals were found, which have already been described earlier as being induced by interferons.

In addition to the typical interferon-induced genes, a number of upregulated genes were found with potentially important roles in establishment and maintenance of pregnancy. Among these genes was the transcription factor nuclear receptor subfamily 2, group F, member 2 (*NR2F2*, alias COUP-TFII), a nuclear orphan receptor, which is essential for progesterone control of implantation and a mediator of uterine epithelial-stromal cross talk in the mouse (Kurihara et al., 2007; Petit et al., 2007). During the peri-implantation period, NR2F2 regulates embryo attachment and decidualization through controlling ESR1 activity but is also required in the postimplantation period to facilitate placentation (Lee et al., 2010). Heterozygous *Nr2f2*-mutant mice showed decreased fecundity (Takamoto et al., 2005). Upregulation of *NR2F2* was also found in equine endometrium at day 12 of pregnancy compared to nonpregnant controls, suggesting a conserved role of *NR2F2* during establishment of pregnancy in mammalian species. A molecular interaction network for NR2F2 is shown in Figure 7.3. Most interestingly, *NR2F2* is a target gene of Indian hedgehog (IHH), which is induced by progesterone via the progesterone receptor (PGR) (Simon et al., 2009). In addition to the negative regulation of *ESR1* expression, NR2F2 also downregulates expression of the gene for oxytocin receptor (*OXTR*) in a complex with NR2F6 (Chu and Zingg, 1997). Furthermore, analysis of the microarray data for bovine endometrium assigned many of the DEGs to a variety of processes that are important with respect to the preparation of the endometrium for embryo attachment and implantation, such as endometrial remodeling (ECM remodeling, vascular remodeling), cell adhesion, and immunomodulation. Genes involved in immunomodulation are potentially important for tolerance of the fetal allograft by the maternal immune system. Most of them, if not all, are probably regulated by the embryonic IFNT.

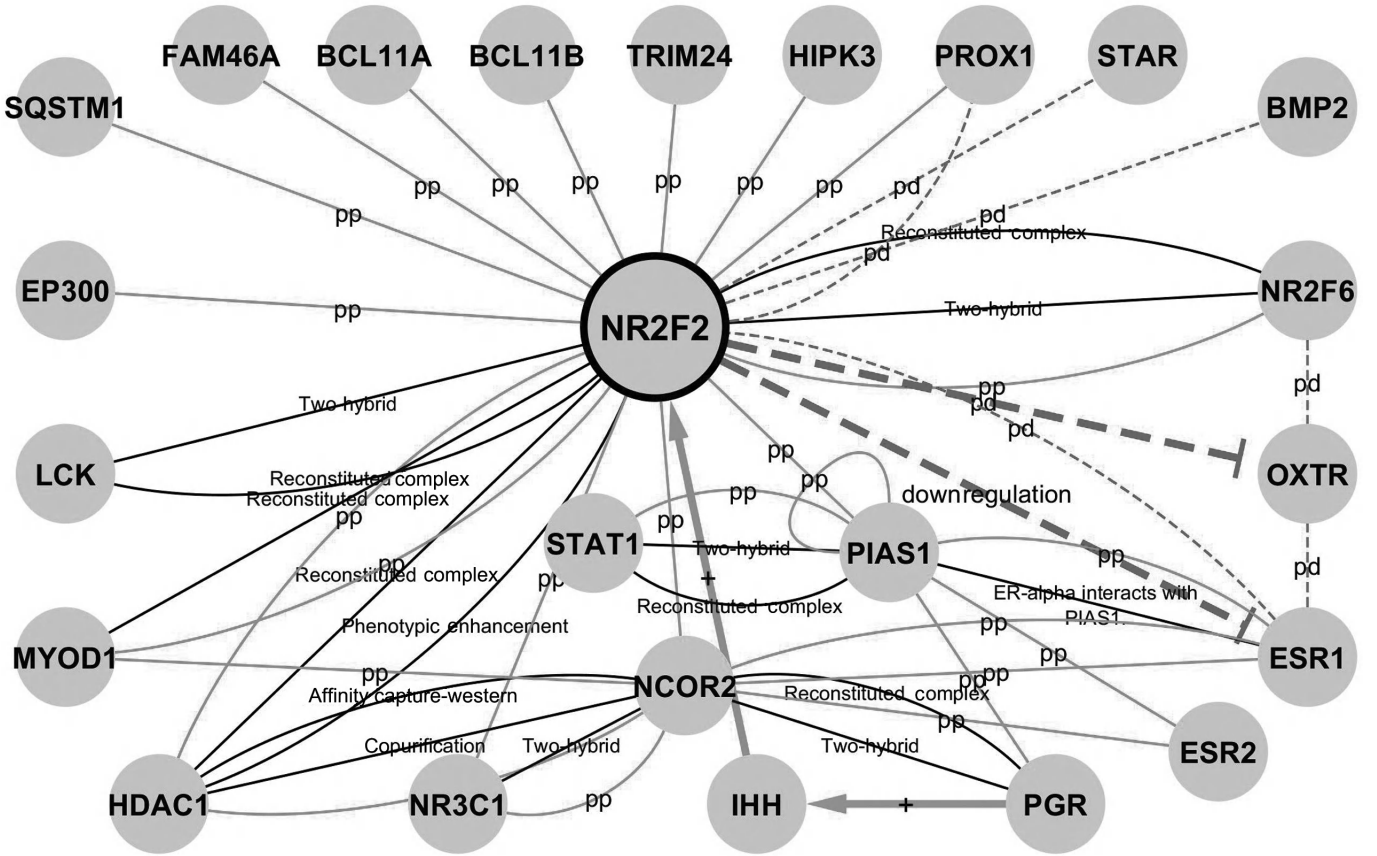


Figure 7.3 Interaction network for the essential transcription factor nuclear receptor subfamily 2, group F, member 2 (NR2F2). Interactors are shown with their official gene symbol. pp, protein–protein interaction; pd, protein–DNA interaction (promoter-binding). (*See insert for color representation of this figure.*)

Analysis of Gene Expression in Endometrium During the Preimplantation Phase in Porcine Endometrium

In swine, the implantation takes place after elongation of the conceptuses and placentation is noninvasive (epitheliochorial) similar to ruminants (Carter and Enders, 2004). However, the embryonic signal for maternal recognition of pregnancy in pigs is estrogen produced by the conceptus in increasing amounts from day 10 (Geisert et al., 1990) resulting in changes in the prostaglandin metabolism of the endometrium to prevent regression of the corpora lutea by prostaglandin F2alpha ($PGF_{2\alpha}$) (Ziecik, 2002). Recent research indicates that the estrogen signal from the conceptuses stimulates endometrial prostaglandin E2 (PGE_2) synthesis. Combined with a positive PGE_2 feedback loop in the endometrium, this synthesis leads to an increase in the PGE_2:$PGF_{2\alpha}$ ratio, which helps to overcome the luteolytic effect of $PGF_{2\alpha}$ (Waclawik et al., 2009). The porcine embryos remain free-floating in the uterine lumen until days 13–14 of gestation, when they appose and subsequently attach to the uterine luminal epithelium (Dantzer, 1985). The apposition and the onset of implantation are accompanied by extensive tissue remodeling of the endometrium (Cencic et al., 2003), where a pronounced vascularization is evident already from day 13 of gestation (Keys et al., 1986).

To characterize the processes involved in the initiation of placentation at the gene expression level, a microarray study of porcine endometrium at day 14 of pregnancy in comparison to corresponding nonpregnant controls was performed (Østrup et al., 2010). This study identified 263 DEGs between pregnant and nonpregnant sows. The identification of overrepresented GO terms for these genes revealed that most of the significantly enriched GO terms had allocated more upregulated than downregulated genes. These GO terms included: developmental process, transporter activity, calcium ion binding, apoptosis, cell motility, enzyme linked receptor protein signaling pathway, positive regulation of cell proliferation, ion homeostasis, and hormone activity. Only three terms had an overrepresentation of downregulated genes, namely, oxidoreductase activity, lipid metabolic process, and organic acid metabolic process. A number of the genes assigned to these terms are known to be involved in steroid hormone and prostaglandin metabolism. In the next step, an interaction network was built based on the genes assigned to the functional term developmental process (Østrup et al., 2010). This interaction network and known gene functions found in the literature identified the genes interleukin 6 receptor (*IL6R*), leukemia inhibitory factor receptor alpha (*LIFR*), interleukin 11 receptor, alpha (*IL11RA*), mucin 4 (*MUC4*), v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian) (*ERBB3*), fibroblast growth factor 9 (glia-activating factor) (*FGF9*), and fibroblast growth factor receptor 3 (*FGFR3*) likely to be involved in the process of placentation. *IL6R*, *LIFR*, and *IL11RA* are related to the interleukin gene families and involved in cytokine cell signaling. Among the growth factor-related genes were *FGF9*, *FGFR3*, and *ERBB3*. The differential gene expression and localization of IL11RA protein expression together with the described role of IL11 signaling for decidualization and regulation of trophoblast invasion in the mouse (Robb et al., 1998) indicated a role for IL11RA and IL11 signaling in porcine endometrium by inhibiting trophoblast invasion. Interestingly, the microarray study of porcine endometrium at day 14 of pregnancy identified differential expression of the cytokine receptor genes *IL6R*, *LIFR*, and *IL11RA* but the corresponding

cytokines were not found as differentially expressed. This suggests that regulation of signaling in this cytokine family in porcine endometrium is to great extent controlled by the expression of the specific receptors. A second interesting finding was the upregulation of the mRNA coding for the growth factor FGF9, which functions as an endometrial growth factor in humans (Tsai et al., 2002), in pregnant animals and the concomitant downregulation of *FGFR3* mRNA, coding for an FGF9 receptor. *FGFR3* downregulation and localization of FGF9 protein in the apical domain of the glandular epithelial cells in porcine endometrium of day 14 pregnant animals suggests that secreted FGF9 functions as an embryonic growth factor in the pig.

Analysis of Gene Expression in Endometrium During the Preimplantation Phase in Equine Endometrium

The nature of embryo–maternal communication and maternal recognition of pregnancy in equids is still not completely understood. A number of features of equine pregnancy are unique to the genus *Equus* and differ from other mammals. The equine embryo is completely enveloped by a tough glycoprotein capsule between days 7 and 21, which prevents the trophoblast from elongating and provides its typical spherical shape. The conceptus shows constant, self-induced mobility throughout the uterine lumen between days 6 and 17 after ovulation. Around day 17, as a result of increased conceptus diameter, increased uterine tone, and because of changes in the embryo's capsule and uterine environment, the conceptus becomes immobilized ("fixed") at the base of one of the uterine horns. At days 35–37, an "injection" of specialized, gonadotropin-secreting trophoblast cells into the maternal endometrium takes place followed by the establishment of a stable, microvillous contact of trophoblast cells with the luminal epithelium of the endometrium around days 40–42 (Allen, 2001; Allen and Wilsher, 2009).

In contrast to ruminants and swine, the nature of the embryonic pregnancy recognition signal to prevent luteolysis still remains unknown. However, the presence of a conceptus somehow uncouples the oxytocin-induced release of $PGF_{2\alpha}$ thereby preventing luteolysis (Goff, 1987; Sharp, 2000). The equine conceptus produces a number of different secretory products during early pregnancy, including steroids, prostaglandins, different proteins and peptides (Betteridge, 2000), such as interferon delta, a member of the type I interferon family (Cochet et al., 2009). Interestingly, the application of intrauterine devices has been demonstrated to prolong the luteal phase in the mare, indicating that a form of mechanotransduction by the migrating conceptus may also play a role in preventing production and release of $PGF_{2\alpha}$ (Rivera Del Alamo et al., 2008).

Very recently, two microarray studies of equine endometrium during early pregnancy were published. These studies analyzed days 8 and 12 (Merkl et al., 2010), and day 13.5 (Klein et al., 2010) of pregnancy and identified several hundred DEGs at days 12 and 13.5 of pregnancy. Gene set enrichment analysis, DAVID functional annotation clustering, and cocitation (CoPub) analysis were performed to identify overrepresented functional terms and biological pathways for the genes differentially expressed in day 12 pregnant endometrium. GSEA was performed to characterize the DEGs by

comparison with gene sets derived from other Omics studies. The greatest overlap with the genes upregulated at day 12 of pregnancy was found for a set of genes differentially expressed between two types of CD45 (PTPRC)$^-$ CD34$^+$ CD105 (endoglin)$^+$ stromal stem cells distinguished by the expression of CD31 (PECAM1) (Boquest et al., 2005). The CD31$^+$ cells are closely related to microvascular endothelial cells based on their upregulated transcripts. This agreed well with the results of DAVID and CoPub analysis, where terms related to angiogenesis/vascular remodeling were found to be quantitatively enriched. Furthermore, the mRNA coding for CD31 (PECAM1), a marker of endothelial cells that has also been described in context of angiogenesis (Woodfin et al., 2007), was found as 1.6-fold upregulated in the samples of day 12 pregnant endometrium. A substantial overlap was also found for the CD31$^+$ downregulated gene set that contains transcripts associated with ECM, transcripts that have been shown as expressed in early osteoblast differentiation, osteoclast-related transcripts, and transcripts typical of neuronal tissue (Boquest et al., 2005). Related terms were also found with DAVID and CoPub, such as extracellular region, tissue remodeling, bone remodeling, neurogenesis, and inflammation. Thus, GSEA revealed a biological characterization of many of the differentially expressed genes. Overall, GSEA identified biologically very different gene sets that could reflect (i) differential gene expression in different compartments of the endometrium and (ii) a response to different embryonic signals. This corresponds to the fact that the equine conceptus produces different molecules (Betteridge, 2000), such as progesterone, estradiol, and prostaglandins.

GSEA revealed many estrogen-induced genes and genes involved in regulation of estrogen signaling, but also genes known to be regulated by progesterone and prostaglandin E2. Likewise, at day 13.5, many genes with known or inferred functions are probably upregulated by embryonic estrogen. In addition, DAVID analysis revealed for the genes with elevated transcript levels overrepresentation of genes involved in cell–cell signaling, heat shock response, and genes coding for secretory proteins. Among the genes showing lower expression in pregnant mares on day 13.5, estrogen receptor 1 (*ESR1*) was of particular interest because of its potential involvement in the initiation of luteolysis in cyclic mares (Klein et al., 2010).

Based on the genes identified as upregulated at day 12 of pregnancy, putative interaction networks for genes related to the process of angiogenesis/vascular remodeling and genes described in context of steroid hormone and prostaglandin signaling were generated on the basis of a literature search, CoPub results and interactions from the Pathway Architect database (Stratagene) and other public protein interaction databases (Merkl et al., 2010). For the process of angiogenesis/vascular remodeling genes representing different regulatory levels were found, such as members of the angiopoietin family, of the VEGF system, hypoxia-induced genes, and genes regulating endothelial cell fate. Most of these genes have been described in context of positive regulation of angiogenesis, however, no difference in the proportion of blood vessels between pregnant and control samples was observed by quantitative stereology. A possible explanation could be a delay of changes in the expression of the corresponding proteins and the resulting biological consequences. In studies of vascular perfusion during early pregnancy, where transient changes in endometrial vascular perfusion accompanying the migrating embryonic vesicle have been shown (Silva et al., 2005), the microarray results indicate vascular remodeling at day 12 of pregnancy in response

to the migrating conceptus that likely could play a role in maternal support of conceptus growth and in preparing the uterus for the prospective pregnancy. Furthermore, deregulation of angiogenesis in the endometrium during early pregnancy has been found in pregnancy failure (Tayade et al., 2007).

The second interaction network related to steroid hormone and prostaglandin signaling was clearly dominated by estradiol with many estrogen-regulated genes. The identification of many estrogen-induced genes fits well with the finding that the equine embryo begins to secrete significant amounts of estrogens as early as day 10 after fertilization (Zavy et al., 1984; Choi et al., 1997). The interaction network contained also a number of negative regulators of *ESR1*, but also genes involved in regulation of growth and differentiation, and in estrogen metabolism. One important mediator of estrogen signaling in equine endometrium could be FGF9, which has been described as an autocrine endometrial stromal growth factor induced by E2 in human endometrial stroma (Tsai et al., 2002). The upregulation of the *FGF9* mRNA, the findings from other species and the concomitant upregulation of a putative *FGF9* antisense transcript make this gene an especially interesting candidate. Furthermore, a number of negative regulators of estrogen signaling, e.g., Kruppel-like factor 5 (*KLF5*), ERBB receptor feedback inhibitor 1 (*ERRFI1*), and heat shock 27kDa protein 2 (*HSPB2*), were found as upregulated, which could be an indication for either a negative feedback regulation in response to the estrogen signal or the result of progesterone action on the endometrium. The interaction network also contained a number of genes that function in prostaglandin signaling and metabolism. Similar to findings in the pig, where the PGE2 receptor EP2 (*PTGER2*) transcript level is increased in early pregnancy (Waclawik et al., 2009), mRNAs of PGE2 receptors EP3 (*PTGER3*) and EP4 (*PTGER4*) were upregulated. However, in contrast to studies in porcine endometrium, mRNA levels of prostaglandin E synthases did not differ between pregnant and nonpregnant equine endometrium. There was also no difference in mRNA levels for the known PGF synthases. Unlike in ruminants, where upregulation of mRNA for oxytocin receptor (*OXTR*) is prevented by the signaling of IFNT (Wolf et al., 2003), *OXTR* mRNA was slightly upregulated in equine endometrium at day 12 of pregnancy.

Although the results of this study revealed a response to different signaling molecules, a mechanical signaling induced by the migrating conceptus is not excluded. In a recent study, a small intrauterine device (water-filled plastic ball) was shown to induce prolonged luteal function (Rivera Del Alamo et al., 2008), supporting the concept of pregnancy recognition via mechanosensation. Some of the identified changes in mRNA expression levels at day 12 of pregnancy could in part reflect mechanosensation. Some of the upregulated genes were already described in the context of mechanotransduction. First, a direct response to mechanical force has been shown for *PECAM1* (Fujiwara, 2006). Second, upregulation of insulin-like growth factor binding protein 1 (IGFBP1) secretion in response to mechanical stretch was found in decidualized endometrial stromal cells (Harada et al., 2006). Third, Rho activation (members of the Rho GTPase family (*RND1*, *RND3*) (key regulators of cytoskeletal signaling) and a Rho GTPase activating protein (*ARHGAP29*) were upregulated at day 12 of pregnancy, which has been described in context of mechanotransduction-associated alveolar epithelial cell differentiation (Foster et al., 2010).

Altogether, these two microarray studies of equine endometrium during early pregnancy revealed potential target genes and pathways of conceptus-derived estrogens, progesterone, and prostaglandin E2 in the equine endometrium probably involved in the early events of establishment and maintenance of pregnancy in the mare.

Comparison of Gene Expression Datasets from Different Mammalian Species

The comparison of endometrial gene expression datasets between species of different mammalian groups can be used as an approach for the identification of genes related to common and species-specific mechanisms involved in establishment and maintenance of pregnancy. Such a comparison was done between the datasets for gene expression during the estrous cycle and on day 18 of pregnancy in bovine endometrium, and results of a number of similar microarray studies in human, mouse, and Rhesus monkey (Bauersachs et al., 2008). This analysis revealed an overlap of 70 genes that were differentially expressed during the estrous cycle in bovine endometrium and in at least one of the other studies. For 38 of these genes, the changes of mRNA levels were comparable to the changes found in the other studies. For 29 genes, expression changes were in opposite directions, for example, genes were downregulated in bovine endometrium during the luteal phase but upregulated in human endometrium during the window of implantation. For the remaining three genes, expression changes were contrary within the compared studies regarding the expression profile during the sexual cycle or the regulation by estrogen, respectively. This finding reflects (1) the differences between ruminant species and primates and rodents regarding histological changes in the endometrium during the cycle and the type of implantation of the embryo and (2) some common regulatory mechanisms between mammalian species.

For example, similar gene expression was found for claudin 4 (*CLDN4*), a cell adhesion molecule in tight junctions involved in intercellular sealing in simple and stratified epithelia (Tsukita and Furuse, 2002). CLDN4 has been found to selectively decrease Na^+ permeability in tight junctions. Likewise, dickkopf homolog 1 (*DKK1*) mRNA, coding for an inhibitor of WNT signaling (Glinka et al., 1998), has been found as upregulated in four human studies at the window of implantation time and in bovine endometrium during the luteal phase and at day 18 of pregnancy (Bauersachs et al., 2008). Correlated expression differences were also found in two human studies for nuclear protein 1 (*NUPR1*, candidate of metastasis 1, P8), *SLC1A1* and decidual protein induced by progesterone (*C10orf10*). Further, there seem to be some common regulatory mechanisms of the maternal immune system as indicated by the similar regulation of *C1R*, *SERPING1*, and *TAP1*.

In a recent study, GSEA was used to compare a microarray dataset from equine endometrium at day 12 of pregnancy to similar gene sets derived from other mammals (Merkl et al., 2010). Interestingly, the best enrichment toward the day 12 upregulated genes was found for genes upregulated in human endometrium during the window of implantation, indicating similarities in gene expression changes in equine and human endometrium during early pregnancy. Significant enrichment was also obtained for genes induced at day 14 of early pregnancy in porcine endometrium and at day 18 of

early pregnancy in bovine endometrium, but the number of genes overlapping with the genes significantly upregulated at day 12 of pregnancy in equine endometrium was relatively low. Higher numbers of overlapping genes were seen for genes regulated during the estrous cycle in bovine endometrium and estrogen-induced genes in general. Furthermore, the comparison of the overlapping genes of the gene sets from human, bovine, and porcine endometrium revealed a number of genes that probably have conserved functions across species such as crystallin, alpha B (*CRYAB*), *ERRFI1*, *FGF9*, insulin-like growth factor binding protein 2 (*IGFBP2*), *NR2F2*, stanniocalcin 1 (*STC1*), and *TNFSF10*.

The data analysis of the microarray studies of bovine, porcine, and equine endometrium during early pregnancy each revealed some major themes and pathways probably related to establishment and maintenance of pregnancy, in which some of these themes and pathways are likely to be common for the three species, whereas others are species-specific. A schematic overview of selected processes and functional groups of genes and their possible relationship to pregnancy recognition and endometrial receptivity is shown in Figure 7.4.

Identification of Fertility-Related Genes by the Analysis of Pathological Conditions

Transcriptome studies of the endometrium during the sexual cycle and early pregnancy revealed numerous DEGs reflecting the complex changes in the endometrium during the cycle and in the course of establishment and maintenance of pregnancy. This makes the identification of genes with essential and/or regulatory roles difficult. One approach to find the most crucial fertility-related genes is the analysis of pathological conditions, for example, those that lead to abnormal placenta formation and embryo development.

Analysis of Endometrial Responses to Clone Pregnancies in Comparison to IVF Pregnancies in Cattle

Somatic cell nuclear transfer (SCNT) cloning has been successfully performed in a number of species, but is particularly critical with respect to epigenetic abnormalities of the resulting embryos, fetuses, and offspring (reviewed in Shi et al., 2003). DNA hypermethylation was observed in some tissues of cloned bovine fetuses, but—to a lesser extent—also in fetuses derived from *in vitro*-produced embryos (Hiendleder et al., 2004b, 2006). However, it is largely unclear whether and how epigenetic changes cause developmental abnormalities and abortions of cloned embryos or fetuses. A number of studies suggested placental abnormalities as primary cause of pregnancy loss and abnormal fetal growth after transfer of SCNT embryos (reviewed in Palmieri et al., 2008). Observed placental changes in bovine SCNT pregnancies included, for example, a reduced number, but increased size of placentomes (Constant et al., 2006). Furthermore, transplacental leakage of maternal cells into the circulation of fetuses derived by SCNT, but not in IVF-derived fetuses was observed (Hiendleder et al., 2004a). These findings raised the question, how and when abnormal placental

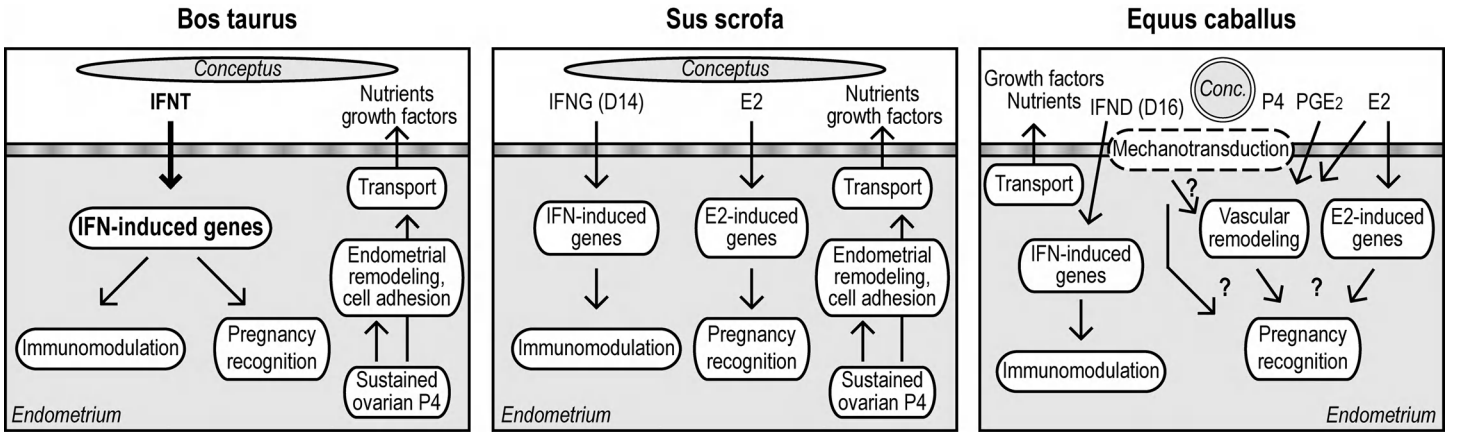


Figure 7.4 Overview of processes and signaling pathways inferred from transcriptome studies and existing knowledge related to pregnancy recognition and endometrial receptivity in cattle, swine, and horse. IFNT, interferon tau; IFNG, interferon gamma; IFND, interferon delta; E2, estrogen; P4, progesterone and its derivatives; PGE2, prostaglandin E2.

development is induced. To clarify if placental abnormalities may have their origin in abnormal embryo–maternal communication already in the preimplantation period, a microarray study of the response of the endometrium on day 18 of pregnancy to SCNT embryos versus embryos derived by *in vitro* fertilization (IVF) was performed (Bauersachs et al., 2009). To exclude specific effects of a particular embryonic genotype and to have a similar genetic variation in the SCNT and the IVF groups, several different nuclear donor cell lines were used for SCNT. Cluster analysis of the microarray data revealed a greater variation of mRNA profiles in the SCNT group than in the IVF group. Despite this variation in the SCNT group, 58 transcripts were differentially abundant comparing endometria from SCNT versus IVF pregnancies. For many of these genes an important role in implantation and/or placentation has already been shown or suggested. For the gene serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 14 (SERPINA14, also known as uterine milk protein or UTMP) upregulation of transcript levels in bovine endometrium during the ovulatory phase and during early pregnancy was shown (Ulbrich et al., 2009). In endometrium from SCNT pregnancies, SERPINA14 mRNA levels were lower compared to IVF pregnancies. Studies of SERPINA14 function indicate a role in mediating immunosuppressive effects of progesterone on the endometrium (Arck et al., 2007). The most interesting of the identified transcription factor genes was *NR2F2* for which an essential function in placentation in the mouse has been shown (see above) (Petit et al., 2007). Another interesting gene in the context of placenta formation was gap junction protein, alpha 1, 43 kDa (*GJA1*, also known as connexin 43) with lower transcript levels in SCNT pregnancies. A conditional deletion of the *Gja1* gene in the stromal cells of the murine endometrium, leading to impaired production of key angiogenic factors needed for the development of new blood vessels within the stromal compartment, resulted in the arrest of embryo growth and early pregnancy loss (Laws et al., 2008).

These findings indicate that abnormal placentation in bovine clone pregnancies may originate from a disturbed embryo–maternal communication starting already during the peri-implantation period.

At the same time, a second study was published that analyzed endometrial samples derived from SCNT pregnancies in comparison to IVF pregnancies and pregnancies after artificial insemination (AI) at day 20 of gestation (Mansouri-Attia et al., 2009). The authors provided evidence that the endometrium can be seen as a biological sensor able to fine-tune its physiology in response to embryos of different sources and developmental stages. Compared with AI, many biological functions and canonical pathways related to metabolism and immune function were found to be significantly altered in the endometrium of SCNT pregnancies at implantation. Gene expression differences between endometrium samples from IVF and AI pregnancies were less pronounced. In comparison to our study, a different experimental setup was used. SCNT embryos were derived from a single cell line, four embryos were transferred in case of SCNT and IVF, caruncular samples were analyzed in addition to samples from intercaruncular regions, and SCNT samples were compared to AI samples. Furthermore, different array platforms were used: a 13,000 elements oligonucleotide microarray in Mansouri-Attia et al. (2009) and a custom cDNA array representing 950 genes identified in the endometrium and the oviduct during the estrous cycle and early pregnancy (Bauersachs et al., 2007, 2009). Due to these differences in the experimental

design and the different array platforms, a direct comparison of the results is not possible. However, the principal findings of these two studies are similar. Both studies revealed many DEGs between SCNT pregnancies and pregnancies initiated with fertilized embryos (after AI or IVF). In general, these studies showed that there is an endometrial plasticity at the onset of implantation and deregulation of the maternal environment greatly influences the development of an embryo and the success of pregnancy.

Comparison of Porcine Endometrium from Day 30 of Clone Pregnancies with Normal Pregnancies Identified Genes Involved in Placenta Formation

In the pig, disturbed development of SCNT embryos to full-term fetuses is mainly caused by failures in extraembryonic tissue formation (Lee et al., 2007). Losses of clone pregnancies in pigs are mainly due to placental insufficiency. To understand how the maternal uterine environment responds to porcine SCNT embryos during early pregnancy, Ka et al. (2008) compared gene expression profiles in the endometrium from uteri containing SCNT embryos with endometrium from uteri containing embryos produced by natural mating on day 30 of pregnancy. Morphological analysis showed that extraembryonic tissues and fetuses derived from SCNT embryos were smaller than those derived from normal embryos. In addition, the uterine endometrium with SCNT embryos and also fetal membranes derived from SCNT embryos were less vascularized compared to normal pregnancy. Accordingly, a relatively large number of genes were found differentially expressed, most of them with decreased levels in endometrium from SCNT pregnancies. The DEGs included genes for enzymes involved in steroidogenesis and ECM remodeling and uterine secretory proteins. In line with the retarded development of the SCNT embryos, expression of mRNAs coding for steroidogenic enzymes was decreased, which could probably result in insufficient steroid supply to the placenta needed for maintenance of pregnancy. The expression of cathepsins B, D, H, L, and Y was also found to be decreased in endometrium from SCNT pregnancies. Different cathepsins were also found as differentially expressed in bovine and equine endometrium during the estrous cycle and early pregnancy (Bauersachs et al., 2006; Klein et al., 2006; Mitko et al., 2008; Merkl et al., 2010; Ostrup et al., 2010). Cathepsins belong to lysosomal cysteine, serine, and aspartic proteases, and have a variety of functions such as degradation of ECM molecules, activation of intracellular proteins and prohormones, and also regulation of the immune system and apoptosis (Conus and Simon, 2008; Mason, 2008; Obermajer et al., 2008). With regard to endometrial functions, cathepsins are implicated as regulators for implantation, placentation, and trophoblast invasion. The decreased expression of cathepsins in the endometrium with SCNT embryos could result in inappropriate endometrial tissue remodeling and/or a lack of other protein activation, which is important for the maintenance of pregnancy. From this study, conclusions can be drawn on the molecular differences between endometria of clone pregnancies and normal pregnancies. In addition to these conclusions, the results also provide genes that play an important role in the process of placenta formation, since deregulation due to insufficient embryonic signaling led to retarded placenta development.

Strategies and Approaches to Obtain Deeper Insights into and Better Understanding of the Processes Related to Establishment and Maintenance of Pregnancy

Most of the studies conducted so far were performed on the analysis of complete endometrial tissue samples, which have a complex composition containing luminal and glandular epithelial cells, stromal cells, blood vessels, and a variety of immune cells. This can result, for example, in sensitivity issues, since some genes are only expressed in a proportion of the endometrial cells, and can also lead to wrong conclusions with respect to putative interaction networks when two "interacting" genes are not expressed in the same cell type. To overcome this problem, a separate analysis of the most important endometrial cell types by the use of laser microdissection could be performed. However, the separate analysis of several parts of the endometrium would considerably increase the complexity of the transcriptome analyses.

Another approach to find genes or pathways associated with fertility is the analysis of endometrial transcriptomes across different mammalian species. So far, results of different microarray studies were already compared, but this has a number of limitations. First of all, microarray data cannot be directly compared due to different hybridization kinetics, nucleotide composition, and location of the probe within the target transcript. Furthermore, different array platforms were used in the respective studies and most of the used microarrays did not contain all known genes of the corresponding species. This problem could be overcome by the use of RNA-Seq, analyzing in principle all poly(A) RNAs present in a cell or tissue. Since the whole transcript is sequenced, the expression data is more quantitative than data from microarrays and can also be compared between different species for orthologous genes.

A third example for a strategy to get more insights into gene expression regulation in the endometrium is the analysis of microRNAs (miRNA). MicroRNAs have a central role in regulation of translation and stability of mRNAs and are themselves subject of complex regulation (Krol et al., 2010). An important role of miRNAs in endometrial gene expression can be postulated. For example, in human endometrium, differential expression of miRNAs during the menstrual cycle has been shown, suggesting a function in mediating responses to the ovarian hormones (Kuokkanen et al., 2010).

A formal proof for the relevance of a gene or mechanism would be the targeted alteration of its function followed by a comprehensive analysis of the phenotypic consequences. This reverse genetics approach is routinely used in mouse models, where a plethora of strategies for tailored genetic modification is available. The findings from genetically modified mouse models may also be relevant for other species. For instance, the reproductive deficits in *Nr2f2* mutant mice (Takamoto et al., 2005) pointed to an important function of this gene for implantation also in other species. In comparison to mouse, the spectrum of techniques for genetic engineering of livestock species is much more limited, although efficient techniques, such as lentiviral transgenesis (Hofmann et al., 2003, 2004) or nuclear transfer cloning using genetically modified cells (reviewed in Aigner et al., 2010) facilitate the generation of tailored animal models. However, such approaches are still labor, time, and cost intensive. A more feasible approach is the modification of embryos and the subsequent observation of their development including the cross talk with the maternal

environment. One example is the knockdown of specific gene products by using RNA interference. The first pilot experiments for this approach have already been performed in livestock species (Tesfaye et al., 2007, 2010).

Another possibility is the use of embryos carrying fluorescent marker genes which may help to localize early embryos in the uterus or—in later stages of pregnancy—to clearly distinguish maternal and embryonic/fetal portions of the placenta (Reichenbach et al., 2010). Reporter genes may be also used to label important steps in development (Wuensch et al., 2007), which may help to classify embryos according to their developmental potential and to characterize the differential responses of the endometrium, The endometrium may also be directly targeted, for example, by using morpholino antisense oligonucleotides to block specific gene products. This approach has been successfully used in the mouse model to dissect mechanisms of implantation (Luu et al., 2004; Nie et al., 2005) and may also be feasible in livestock models.

In summary, holistic and sensitive Omics-technologies characterizing the transcriptome, proteome, metabolome, and other molecular characteristics of cells or tissues facilitate the comprehensive description of molecular patterns of tissues that are associated with particular physiological or pathophysiological conditions. Importantly, dynamic changes of these patterns during development or disease may point to genes or pathways that have an effect on the trait under investigation. In addition to the development and implementation of Omics-phenotypes, a refinement of physiological readouts is urgently required. Those can be obtained, for example, by the development of noninvasive longitudinal techniques such as remote/indirect sensing or imaging. Integrating large-scale, high-dimensional molecular and physiological data holds promise for defining the molecular networks that respond to genetic and environmental perturbations of physiological functions, including reproduction.

References

Aigner, B., Renner, S., Kessler, B., et al. (2010) Transgenic pigs as models for translational biomedical research. *Journal of Molecular Medicine*, **88**, 653–664.

Allen, W.R. (2001) Fetomaternal interactions and influences during equine pregnancy. *Reproduction* **121**, 513–527.

Allen, W.R. & Wilsher, S. (2009) A review of implantation and early placentation in the mare. *Placenta* **30**, 1005–1015.

Allison Gray, C., Bartol, F.F., Taylor, K.M., et al. (2000) Ovine uterine gland knock-out model: effects of gland ablation on the estrous cycle. *Biology of Reproduction* **62**, 448–456.

Arck, P., Hansen, P.J., Mulac Jericevic, B., et al. (2007) Progesterone during pregnancy: endocrine-immune cross talk in mammalian species and the role of stress. *American Journal of Reproductive Immunology* **58**, 268–279.

Ashburner, M., Ball, C.A., Blake, J.A., et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* **25**, 25–29.

Bauersachs, S., Mitko, K., Blum, H., et al. (2007) Technical note: Bovine oviduct and endometrium array version 1: a tailored tool for studying bovine endometrium biology and pathophysiology. *Journal of Dairy Science* **90**, 4420–4423.

Bauersachs, S., Mitko, K., Ulbrich, S.E., et al. (2008) Transcriptome studies of bovine endometrium reveal molecular profiles characteristic for specific stages of estrous cycle and early pregnancy. *Experimental and Clinical Endocrinology & Diabetes* **116**, 371–384.

Bauersachs, S., Ulbrich, S.E., Gross, K., et al. (2005) Gene expression profiling of bovine endometrium during the oestrous cycle: detection of molecular pathways involved in functional changes. *Journal of Molecular Endocrinology* **34**, 889–908.

Bauersachs, S., Ulbrich, S.E., Gross, K., et al. (2006) Embryo-induced transcriptome changes in bovine endometrium reveal species-specific and common molecular markers of uterine receptivity. *Reproduction* **132**, 319–331.

Bauersachs, S., Ulbrich, S.E., Zakhartchenko, V., et al. (2009) The endometrium responds differently to cloned versus fertilized embryos. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 5681–5686.

Bazer, F.W., Burghardt, R.C., Johnson, G.A., et al. (2008) Interferons and progesterone for establishment and maintenance of pregnancy: interactions among novel cell signaling pathways. *Reproductive Biology* **8**, 179–211.

Bazer, F.W., Spencer, T.E., Johnson, G.A., et al. (2009) Comparative aspects of implantation. *Reproduction* **138**, 195–209.

Bazer, F.W., Spencer, T.E., & Ott, T.L. (1997) Interferon tau: a novel pregnancy recognition signal. *American Journal of Reproductive Immunology* **37**, 412–420.

Betteridge, K.J. (2000) Comparative aspects of equine embryonic development. *Animal Reproduction Science* **60–61**, 691–702.

Boquest, A.C., Shahdadfar, A., Fronsdal, K., et al. (2005) Isolation and transcription profiling of purified uncultured human stromal stem cells: alteration of gene expression after in vitro cell culture. *Molecular Biology of the Cell* **16**, 1131–1141.

Carninci, P., Yasuda, J., & Hayashizaki, Y. (2008) Multifaceted mammalian transcriptome. *Current Opinion in Cell Biology* **20**, 274–280.

Carter, A.M. & Enders, A.C. (2004) Comparative aspects of trophoblast development and placentation. *Reproductive Biology and Endocrinology* **2**, 46.

Cencic, A., Guillomot, M., Koren, S., et al. (2003) Trophoblastic interferons: do they modulate uterine cellular markers at the time of conceptus attachment in the pig? *Placenta* **24**, 862–869.

Choi, S.J., Anderson, G.B., & Roser, J.F. (1997) Production of free estrogens and estrogen conjugates by the preimplantation equine embryo. *Theriogenology* **47**, 457–466.

Chu, K. & Zingg, H.H. (1997) The nuclear orphan receptors COUP-TFII and Ear-2 act as silencers of the human oxytocin gene promoter. *Journal of Molecular Endocrinology* **19**, 163–172.

Clegg, A.B. & Shepherd, A.J. (2008) Text mining. *Methods in Molecular Biology* **453**, 471–491.

Clemente, M., De La Fuente, J., Fair, T., et al. (2009) Progesterone and conceptus elongation in cattle: a direct effect on the embryo or an indirect effect via the endometrium? *Reproduction* **138**, 507–517.

Cline, M.S., Smoot, M., Cerami, E., et al. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nature Protocols* **2**, 2366–2382.

Cochet, M., Vaiman, D., & Lefevre, F. (2009) Novel interferon delta genes in mammals: cloning of one gene from the sheep, two genes expressed by the horse conceptus and discovery of related sequences in several taxa by genomic database screening. *Gene* **433**, 88–99.

Constant, F., Guillomot, M., Heyman, Y., et al. (2006) Large offspring or large placenta syndrome? Morphometric analysis of late gestation bovine placentomes from somatic nuclear transfer pregnancies complicated by hydrallantois. *Biology of Reproduction* **75**, 122–130.

Conus, S. & Simon, H.U. (2008) Cathepsins: key modulators of cell death and inflammatory responses. *Biochemical Pharmacology* **76**, 1374–1382.

Dantzer, V. (1985) Electron microscopy of the initial stages of placentation in the pig. *Anatomy and Embryology (Berl)* **172**, 281–293.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., et al. (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biology* **4**, P3.

Dominguez, F., Avila, S., Cervero, A., et al. (2003) A combined approach for gene discovery identifies insulin-like growth factor-binding protein-related protein 1 as a new gene implicated in human endometrial receptivity. *The Journal of Clinical Endocrinology and Metabolism* **88**, 1849–1857.

Evans, A.C., Forde, N., O'Gorman, G.M., et al. (2008) Use of microarray technology to profile gene expression patterns important for reproduction in cattle. *Reproduction in Domestic Animals*, **43**(Suppl 2), 359–367.

Fluhr, H., Sauter, G., Steinmuller, F., et al. (2009) Nonapoptotic effects of tumor necrosis factor-related apoptosis-inducing ligand on interleukin-6, leukemia inhibitory factor, interleukin-8, and monocyte chemoattractant protein 1 vary between undifferentiated and decidualized human endometrial stromal cells. *Fertility and Sterility* **92**, 1420–1423.

Foster, C.D., Varghese, L.S., Gonzales, L.W., et al. (2010) The Rho pathway mediates transition to an alveolar type I cell phenotype during static stretch of alveolar type II cells. *Pediatric Research* **67**, 585–590.

Frijters, R., Heupers, B., Van Beek, P., et al. (2008) CoPub: a literature-based keyword enrichment tool for microarray data analysis. *Nucleic Acids Research* **36**, W406–W410.

Fujiwara, K. (2006) Platelet endothelial cell adhesion molecule-1 and mechanotransduction in vascular endothelial cells. *Journal of International Medicine* **259**, 373–380.

Fundel, K., Kuffner, R., & Zimmer, R. (2007) RelEx–relation extraction using dependency parse trees. *Bioinformatics* **23**, 365–371.

Geisert, R.D., Zavy, M.T., Moffatt, R.J., et al. (1990) Embryonic steroids and the establishment of pregnancy in pigs. *Journal of Reproduction and Fertility. Supplement* **40**, 293–305.

Gellersen B., Fernandes, M.S., & Brosens, J.J. (2009) Non-genomic progesterone actions in female reproduction. *Human Reproduction Update* **15**, 119–138.

Giudice, L.C. (2004) Microarray expression profiling reveals candidate genes for human uterine receptivity. *American Journal of Pharmacogenomics* **4**, 299–312.

Glinka, A., Wu, W., Delius, H., et al. (1998) Dickkopf-1 is a member of a new family of secreted proteins and functions in head induction. *Nature* **391**, 357–362.

Goff, A.K. (1987) Oxytocin stimulation of plasma 15-keto-13,14-dihydro prostaglandin F-2a during the oestrus cycle and early pregnancy in the mare. *Journal of Reproduction and Fertility. Supplement* **35**, 253–260.

Goff, A.K. (2004) Steroid hormone modulation of prostaglandin secretion in the ruminant endometrium during the estrous cycle. *Biology of Reproduction* **71**, 11–16.

Gustincich, S., Sandelin, A., Plessy, C., et al. (2006) The complexity of the mammalian transcriptome. *The Journal of Physiology* **575**, 321–332.

Harada, M., Osuga, Y., Takemura, Y., et al. (2006) Mechanical stretch upregulates IGFBP-1 secretion from decidualized endometrial stromal cells. *American Journal of Physiology. Endocrinology and Metabolism* **290**, E268–E272.

Hiendleder, S., Bebbere, D., Zakhartchenko, V., et al. (2004a) Maternal-fetal transplacental leakage of mitochondrial DNA in bovine nuclear transfer pregnancies: potential implications for offspring and recipients. *Cloning and Stem Cells* **6**, 150–156.

Hiendleder, S., Mund, C., Reichenbach, H.D., et al. (2004b) Tissue-specific elevated genomic cytosine methylation levels are associated with an overgrowth phenotype of bovine fetuses derived by in vitro techniques. *Biology of Reproduction* **71**, 217–223.

Hiendleder, S., Wirtz, M., Mund, C., et al. (2006). Tissue-specific effects of in vitro fertilization procedures on genomic cytosine methylation levels in overgrown and normal sized bovine fetuses. *Biology of Reproduction*, **75**, 17–23.

Ho, A., Morishita, W., Hammer, R.E., et al. (2003) A role for Mints in transmitter release: Mint 1 knockout mice exhibit impaired GABAergic synaptic transmission. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 1409–1414.

Hofmann, A., Kessler, B., Ewerling, S., et al. (2003) Efficient transgenesis in farm animals by lentiviral vectors. *EMBO Reports* **4**, 1054–1060.

Hofmann, A., Zakhartchenko, V., Weppert, M., et al. (2004) Generation of transgenic cattle by lentiviral gene transfer into oocytes. *Biology of Reproduction* **71**, 405–409.

Hoheisel, J.D. (2006) Microarray technology: beyond transcript profiling and genotype analysis. *Nature Reviews Genetics* **7**, 200–210.

Ho Sui, S.J., Fulton, D.L., Arenillas, D.J., et al. (2007) oPOSSUM: integrated tools for analysis of regulatory motif over-representation. *Nucleic Acids Research* **35**(Web Server issue), W245–W252.

Huang, D.W., Sherman, B.T., & Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57.

Jensen, L.J., Kuhn, M., Stark, M., et al. (2009) STRING 8 - a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Research* **37**, D412–D416.

Ka, H., Seo, H., Kim, M., et al. (2008) Gene expression profiling of the uterus with embryos cloned by somatic cell nuclear transfer on day 30 of pregnancy. *Animal Reproduction Science* **108**, 79–91.

Keys, J.L., King, G.J., & Kennedy, T.G. (1986) Increased uterine vascular permeability at the time of embryonic attachment in the pig. *Biology of Reproduction* **34**, 405–411.

Klein, C., Bauersachs, S., Ulbrich, S.E., et al. (2006) Monozygotic twin model reveals novel embryo-induced transcriptome changes of bovine endometrium in the preattachment period. *Biology of Reproduction* **74**, 253–264.

Klein, C., Scoggin, K.E., Ealy, A.D., et al. (2010) Transcriptional profiling of equine endometrium during the time of maternal recognition of pregnancy. *Biology of Reproduction* **83**, 102–113.

Krol, J., Loedige, I., & Filipowicz, W. (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nature Reviews Genetics* **11**, 597–610.

Kuokkanen, S., Chen, B., Ojalvo, L., et al. (2010) Genomic profiling of microRNAs and messenger RNAs reveals hormonal regulation in microRNA expression in human endometrium. *Biology of Reproduction* **82**, 791–801.

Kurihara, I., Lee, D.K., Petit, F.G., et al. (2007) COUP-TFII mediates progesterone regulation of uterine implantation by controlling ER activity. *PLoS Genetics* **3**, e102.

Kutsukake, M., Tamura, K., Yoshie, M., et al. (2010) Knockdown of IGF-binding protein 7 inhibits transformation of the endometrial gland in an in vitro model. *Molecular Reproduction and Development* **77**, 265–272.

Laws, M.J., Taylor, R.N., Sidell, N., et al. (2008) Gap junction communication between uterine stromal cells plays a critical role in pregnancy-associated neovascularization and embryo survival. *Development* **135**, 2659–2668.

Lee, D.K., Kurihara, I., Jeong, J.W., et al. (2010) Suppression of ERalpha activity by COUP-TFII is essential for successful implantation and decidualization. *Molecular Endocrinology* **24**, 930–940.

Lee, S.Y., Park, J.Y., Choi, Y.J., et al. (2007) Comparative proteomic analysis associated with term placental insufficiency in cloned pig. *Proteomics* **7**, 1303–1315.

Licatalosi, D.D. & Darnell, R.B. (2010) RNA processing and its regulation: global insights into biological networks. *Nature Reviews Genetics* **11**, 75–87.

Lindberg, J. & Lundeberg, J. (2010) The plasticity of the mammalian transcriptome. *Genomics* **95**, 1–6.

Luu, K.C., Nie, G.Y., & Salamonsen, L.A. (2004) Endometrial calbindins are critical for embryo implantation: evidence from in vivo use of morpholino antisense oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 8028–8033.

Mansouri-Attia, N., Sandra, O., Aubert, J., et al. (2009) Endometrium as an early sensor of in vitro embryo manipulation technologies. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 5687–5692.

Marguerat, S. & Bahler, J. (2010) RNA-seq: from technology to biology. *Cellular and Molecular Life Sciences* **67**, 569–579.

Mason, R.W. (2008) Emerging functions of placental cathepsins. *Placenta*, **29**, 385–390.

Matthews, L., Gopinath, G., Gillespie, M., et al. (2009) Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Research*, **37**, D619–D622.

Merkl, M., Ulbrich, S.E., Otzdorff, C., et al. (2010) Microarray analysis of equine endometrium at days 8 and 12 of pregnancy. *Biology of Reproduction* **83**, 874–886.

Mitko, K., Ulbrich, S.E., Wenigerkind, H., et al. (2008) Dynamic changes in messenger RNA profiles of bovine endometrium during the oestrous cycle: Focus on Mammalian Embryogenomics. *Reproduction* **135**, 225–240.

Mori, A., Okuyama, K., Horie, M., et al. (2002) Alteration of methamphetamine-induced striatal dopamine release in mint-1 knockout mice. *Neuroscience Research* **43**, 251–257.

Nie, G., Li, Y., Wang, M., et al. (2005) Inhibiting uterine PC6 blocks embryo implantation: an obligatory role for a proprotein convertase in fertility. *Biology of Reproduction* **72**, 1029–1036.

Obermajer, N., Jevnikar, Z., Doljak, B., et al. (2008) Role of cysteine cathepsins in matrix degradation and cell signalling. *Connective Tissue Research* **49**, 193–196.

Orom, U.A., Derrien, T., Beringer, M., et al. (2010) Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46–58.

Østrup, E., Bauersachs, S., Blum, H., et al. (2010) Differential endometrial gene expression in pregnant and nonpregnant sows. *Biology of Reproduction* **83**, 277–285.

Palmieri, C., Loi, P., Ptak, G., et al. (2008) Review paper: a review of the pathology of abnormal placentae of somatic cell nuclear transfer clone pregnancies in cattle, sheep, and mice. *Veterinary Pathology* **45**, 865–880.

Petit, F.G., Jamin, S.P., Kurihara, I., et al. (2007) Deletion of the orphan nuclear receptor COUP-TFII in uterus leads to placental deficiency. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 6293–6298.

Pimentel, E.C., Bauersachs, S., Tietze, M., et al. (2011) Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling. *Animal Genetics* **42**(3), 251–262.

Reichenbach, M., Lim, T., Reichenbach, H.D., et al. (2010) Germ-line transmission of lentiviral PGK-EGFP integrants in transgenic cattle: new perspectives for experimental embryology. *Transgenic Research* **19**, 549–556.

Riesewijk, A., Martin, J., Van Os, R., et al. (2003) Gene expression profiling of human endometrial receptivity on days LH+2 versus LH+7 by microarray technology. *Molecular Human Reproduction* **9**, 253–264.

Rivera Del Alamo, M.M., Reilas, T., Kindahl, H., et al. (2008) Mechanisms behind intrauterine device-induced luteal persistence in mares. *Animal Reproduction Science* **107**, 94–106.

Robb, L., Li, R., Hartley, L., Nandurkar, H.H., et al. (1998) Infertility in female mice lacking the receptor for interleukin 11 is due to a defective uterine response to implantation. *Nature Medicine* **4**, 303–308.

Rogelj, B., Mitchell, J.C., Miller, C.C., et al. (2006) The X11/Mint family of adaptor proteins. *Brain Research Reviews* **52**, 305–315.

Santos, J.E.P., Thatcher, W.W., Chebel, R.C., et al. (2004) The effect of embryonic death rates in cattle on the efficacy of estrus synchronization programs. *Animal Reproduction Science* **82–83**, 513–535.

Satterfield, M.C., Song, G., Kochan, K.J., et al. (2009) Discovery of candidate genes and pathways in the endometrium regulating ovine blastocyst growth and conceptus elongation. *Physiological Genomics* **39**, 85–99.

Schadt, E.E. (2009) Molecular networks as sensors and drivers of common human diseases. *Nature* **461**, 218–223.

Sharp, D.C. (2000) The early fetal life of the equine conceptus. *Animal Reproduction Science* **60–61**, 679–689.

Shi, W., Zakhartchenko, V., & Wolf, E. (2003) Epigenetic reprogramming in mammalian nuclear transfer. *Differentiation* **71**, 91–113.

Silva, L.A., Gastal, E.L., Beg, M.A., et al. (2005) Changes in vascular perfusion of the endometrium in association with changes in location of the embryonic vesicle in mares. *Biology of Reproduction* **72**, 755–761.

Simon, L., Spiewak, K.A., Ekman, G.C., et al. (2009) Stromal progesterone receptors mediate induction of Indian Hedgehog (IHH) in uterine epithelium and its downstream targets in uterine stroma. *Endocrinology* **150**, 3871–3876.

Song, G., Bazer, F.W., & Spencer, T.E. (2007) Pregnancy and interferon tau regulate RSAD2 and IFIH1 expression in the ovine uterus. *Reproduction* **133**, 285–295.

Spencer, T.E. & Bazer, F.W. (1996) Ovine interferon tau suppresses transcription of the estrogen receptor and oxytocin receptor genes in the ovine endometrium. *Endocrinology* **137**, 1144–1147.

Spencer, T.E., Johnson, G.A., Burghardt, R.C., et al. (2004) Progesterone and placental hormone actions on the uterus: insights from domestic animals. *Biology of Reproduction* **71**, 2–10.

Spencer, T.E., Sandra, O., & Wolf, E. (2008) Genes involved in conceptus-endometrial interactions in ruminants: insights from reductionism and thoughts on holistic approaches. *Reproduction* **135**, 165–179.

Stanton, L.W. (2001) Methods to profile gene expression. *Trends in Cardiovascular Medicine* **11**, 49–54.

Subramanian, A., Tamayo, P., Mootha, V.K., et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545–15550.

Takamoto, N., Kurihara, I., Lee, K., et al. (2005) Haploinsufficiency of chicken ovalbumin upstream promoter transcription factor II in female reproduction. *Molecular Endocrinology* **19**, 2299–2308.

Tayade, C., Fang, Y., Hilchie, D., et al. (2007) Lymphocyte contributions to altered endometrial angiogenesis during early and midgestation fetal loss. *Journal of Leukocyte Biology* **82**, 877–886.

Tesfaye, D., Lonergan, P., Hoelker, M., et al. (2007) Suppression of connexin 43 and E-cadherin transcripts in in vitro derived bovine embryos following culture in vitro or in vivo in the homologous bovine oviduct. *Molecular Reproduction and Development* **74**, 978–988.

Tesfaye, D., Regassa, A., Rings, F., et al. (2010) Suppression of the transcription factor MSX1 gene delays bovine preimplantation embryo development in vitro. *Reproduction* **139**, 857–870.

Tsai, S.J., Wu, M.H., Chen, H.M., et al. (2002) Fibroblast growth factor-9 is an endometrial stromal growth factor. *Endocrinology* **143**, 2715–2721.

Tsukita, S. & Furuse, M. (2002) Claudin-based barrier in simple and stratified cellular sheets. *Current Opinion in Cell Biology* **14**, 531–536.

Ulbrich, S.E., Frohlich, T., Schulke, K., et al. (2009) Evidence for estrogen-dependent uterine serpin (SERPINA14) expression during estrus in the bovine endometrial glandular epithelium and lumen. *Biology of Reproduction* **81**, 795–805.

Waclawik, A., Jabbour, H.N., Blitek, A., et al. (2009) Estradiol-17beta, prostaglandin E2 (PGE2), and the PGE2 receptor are involved in PGE2 positive feedback loop in the porcine endometrium. *Endocrinology* **150**, 3823–3832.

Wang, Z., Gerstein, M., & Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57–63.

Wilkins, C. & Gale, M., Jr. (2010) Recognition of viruses by cytoplasmic sensors. *Current Opinion in Immunology* **22**(1), 41–47.

Wolf, E., Arnold, G.J., Bauersachs, S., et al. (2003) Embryo-maternal communication in bovine - strategies for deciphering a complex cross-talk. *Reproduction in Domestic Animals* **38**, 276–289.

Woodfin, A., Voisin, M.B., & Nourshargh, S. (2007) PECAM-1: a multi-functional molecule in inflammation and vascular biology. *Arteriosclerosis, Thrombosis, and Vascular Biology* **27**, 2514–2523.

Wuensch, A., Habermann, F.A., Kurosaka, S., et al. (2007) Quantitative monitoring of pluripotency gene activation after somatic cloning in cattle. *Biology of Reproduction* **76**, 983–991.

Zadissa, A., McEwan, J.C., & Brown, C.M. (2007) Inference of transcriptional regulation using gene expression data from the bovine and human genomes. *BMC Genomics* **8**, 265.

Zavy, M.T., Vernon, M.W., Sharp, D.C. 3rd, et al. (1984) Endocrine aspects of early pregnancy in pony mares: a comparison of uterine luminal and peripheral plasma levels of steroids during the estrous cycle and early pregnancy. *Endocrinology* **115**, 214–219.

Ziecik, A.J. (2002) Old, new and the newest concepts of inhibition of luteolysis during early pregnancy in pig. *Domestic Animal Endocrinology* **23**, 265–275.

Chapter 8
Modeling Approaches to Link Various Levels of Organization in Animal Physiology

André Bannink, James France, and Jan Dijkstra

Introduction

With the rise of high-throughput technologies, vast amounts of so-called "omics" data are generated nowadays (Chapters 6 and 9). Also, comprehensive software tools or databases that can be used as Web applications on the Internet are available to scientists to handle and interpret these vast amounts of data (Chapters 6 and 9). These developments offer a whole new range of information, which means that new types of data and new physiological concepts come within reach for modeling the regulatory mechanisms underlying observed animal physiology. The -omics data essentially involve measurements at the molecular level of animal physiology in sampled tissues or fluids. An attempt to relate such -omics data to animal physiology requires that a precise and simultaneous record is made of the physiological state of the animal. The present capacity to generate huge amounts of new -omics data seems to have outstripped the capacity of the scientists to make sense of it all. Emphasis on the latter is essential to obtain a richer understanding of the links between -omics data and animal physiology. For example, Baldwin et al. (1980) already addressed the issues of identification for genetic improvement versus understanding of physiological response. Baldwin et al. (1980) noticed considerable biological variation between animals, with efficiencies of nutrient utilization below theoretical optima. He raised two questions: (1) Is it possible to identify animals that have high efficiencies and by genetic selection improve efficiency? (2) If the unfortunate metabolic decisions that less-efficient animals make were known, could the metabolism of those animals be manipulated such that their efficiencies would improve? A large amount of -omics data is available to address the first question Baldwin et al. (1980) raised, whereas the further integration of the -omics data with the metabolism is much less clear.

Various approaches may be taken to identify the relationships between -omics data and animal physiology. An empirical approach may be followed, identifying which observed physiological characteristics appear to be associated with changes in -omics observations, which have been collected simultaneously. This approach can reveal fairly quick answers to questions such as what molecular aspects, metabolic pathways,

Systems Biology and Livestock Science, First Edition. Edited by Marinus F.W. te Pas, Henri Woelders, and André Bannink.

or regulatory mechanisms of gene expression are associated with the observed physiological characteristics or phenotypic traits. More complex approaches attempt to use less of a black-box kind of approach, and try to develop a vision about how animal physiology is organized and by what mechanisms this organization is regulated. Although being more complex and detailed, the latter approach likely results in a model with mechanistic representation and broader applicability, both necessary for integration of information from various levels of organization and various physiological aspects (behavior, digestion, metabolism, production, health, gene expression). Both the more empirical and the more mechanistic approaches are referred to in literature by the same term "Systems Biology" (Woelders et al., 2011). Furthermore, both make use of similar terminology to discriminate between various applications.

In the following sections of this chapter, the link between the objectives of a modeling effort and type of representation of physiological functions will be discussed. Choices have to be made when representing various levels of biological organization. Further, modeling approaches and methods that allow integration of various levels of organization in order to represent physiological function, including the mechanisms governing its regulation and control, will be discussed. Finally, examples will be given to illustrate how a specific modeling objective dictates the modeling approach adopted and type of information required for building the model. This chapter does not intend to give an exhaustive review of modeling techniques for -omics data analysis; for that, the reader is referred to the other chapters (Chapters 6 and 9).

Levels of Organization

Before discussing some frequently used modeling techniques to capture the essence of physiological function in numbers and mathematical formulae, the importance of the concept of "level of organization" needs to be discussed. Although level of organization appears to be a rather abstract phrase that does not add any clarification at first sight, it actually is a useful term in discussions of animal physiology (France and Kebreab, 2008). It helps in finding common ground and pinpointing the differences in approach when various scientific disciplines work on the same physiological problem. The term can be used to draw attention to how and where there are differences in the type of information used and ways this information is interpreted. The term can be used to distinguish between (1) the various levels of detail involved with (or absent in) observational data, (2) the various entities and levels of detail needed to explain observed events and changes in physiological state or function, (3) the various levels of regulation of a physiological function. With respect to the observations made in livestock science, the level of organization ranges from the macromolecular level (the level of -omics data, the lowest level of organization) to the level of the whole animal or even herd or farm (the highest level of organization) (Figure 8.1).

However, the term certainly does not only address the level of detail of the observations made. The term also addresses other aspects, for example, the way observations on different levels of detail are interconnected in terms of the metabolic and physiological mechanisms involved, and the way physiological information and regulatory functions are organized and conducted. Each level has its own concepts and ideas, problems to address, and hypotheses and theories, which are described using its own

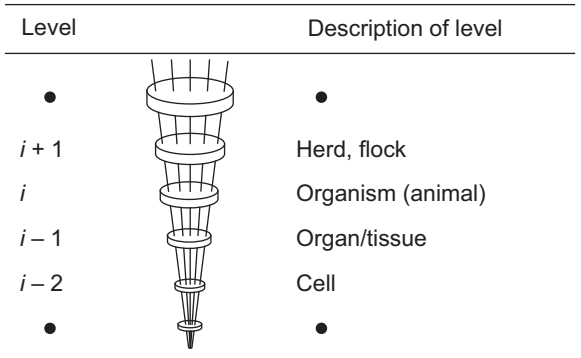


Figure 8.1 Levels of organization (at a higher or a lower level than the level *i* indicating a whole animal) of animal physiology (Dijkstra et al., 2005). Most Systems Biology approaches will focus on representation of levels of organization from *i* (organism) downwards to less than *i − 2* (molecular mechanisms).

language (Thornley and France, 2007). This certainly holds for the vast amounts of -omics information that can be gathered and analyzed nowadays, and will become available in livestock science in the near future (Chapters 6 and 9). In this regard, the difference in description or approach taken to study a physiological aspect in the -omics literature as opposed to the physiological literature is sometimes striking, although the same aspect is addressed. For example, from a (nutri)genomics viewpoint, gut health may be investigated by studying gene expression profiles in samples of gut tissue, and by deriving the most likely functionalities by making use of databases to analyze such data (Chapters 6 and 10). These functionalities may include a whole range of tissue functions (metabolism, transport, communication, immune response, repair, proliferation, and the networks/mechanisms regulating the expression of genes coding for these functions). From a physiological point of view, similar samples of gut tissue are collected. However, maintaining its integrity and functionality is of major concern when recording a physiological response by a tissue or organ; an aspect easily excluded when gathering -omics data. It is well known from trials testing the effect of nutrition on gut function and metabolic fluxes that there are major implications of state and filling of the gut lumen (and its consequences) for mesenteric blood flow, nutrient and metabolite fluxes, and functioning of gut tissues (Johnson et al., 1990; Baldwin, 1995; Bannink et al., 2006a). Ignoring these aspects has unforeseen consequences for the conclusions that can be drawn from -omics data.

The example indicates it is important to include information from higher levels of organization when interpreting observations at a lower level, and *vice versa*. It also illustrates that, depending on the focus of the research, a whole array of intervening levels of organization may unintentionally be excluded from the discussion. In doing so, often general and undefined terms are introduced in texts, such as "animal," "random," or "environmental" effects. In reality, the interconnection between these intervening levels of organization may prove to be extremely important in explaining variation in observed animal physiology that is of particular interest to farmers. With the aim of developing predictive tools that can be applied in livestock science and farming practice, a vision about which levels of organization (including their

interconnections) are to be represented in a mathematical model is of utmost importance (Cornish-Bowden and Cárdenas, 2005, 2007). Such a vision guides the development of mathematical models. However, the vision is based on a whole array of studies and observations, including -omics data. In this regard, the modeling and experimental work are clearly intertwined instead of being separate research efforts. This holds for the well-established physiological modeling efforts, which already addressed higher levels of organization than current -omics efforts. It will also hold for Systems Biology, however, when it aims to predict at higher levels of organization by making use of -omics data. Ideally, Systems Biology and physiological modeling should converge and be combined to realize current modeling objectives in livestock science (Woelders et al., 2011).

Representing Animal Physiology

No standard methodologies are available to identify the entities and relationships that underlie a physiological function. For example, a mechanistic, dynamic growth model developed for veal calves (Gerrits et al., 1997) also may serve as a conceptual framework for pig growth. However, to make it reliable for pigs, the whole model needs to be revisited and reparameterized because of the species-specific characteristics (Halas et al., 2004). Hence, there is no such thing as simply copying a framework to resolve another problem, although the general aspects of physiological function may appear similar, i.e., fat and protein accretion in a growing animal. As another example, the same holds for the application of models to represent digestive function in the gastrointestinal tract. Common concepts and a model framework might be developed for the functioning of the gastrointestinal tract (Bannink et al., 2008; Figure 8.2). However, every animal species, and probably the various genotypes within a species, have their own digestive characteristics requiring different variants of the common framework and separate parameterization. Furthermore, the digestive characteristics will depend on the type and amount of feed ingested as well. This means that nutritional factors need to be included when representing species or genotype differences in digestion and gut health. -Omics data are needed to identify the differences between species and genotypes, but they are preferably gathered in combination with variation in nutritional factors. For example, significant genotype $\times$ diet interactions for milk production have been demonstrated (Beerda et al., 2007). Only with sufficient variation in such observations can relationships be derived that will serve as descriptive elements in quantitative, predictive models.

Standard tools to analyze -omics data are not always able to cover the range of concepts and data needed to represent physiological function. Some physiological functions may require that -omics data are combined with other concepts proven to be relevant (Conner et al., 2010). Examples of physiological concepts often lacking in basic -omics datasets are representation of the physical–chemical compartmentation in cells (cell organelles, membranes, associated protein functions), tissues (various cell types, organization of cells), organs (various tissue types, combination of physiological functions and regulation), and in the whole body (functions of various organs, blood circulation, body reserves, distribution pools). Also, data on the physiological history of the animal are usually missing. A record of historical events or a track record of

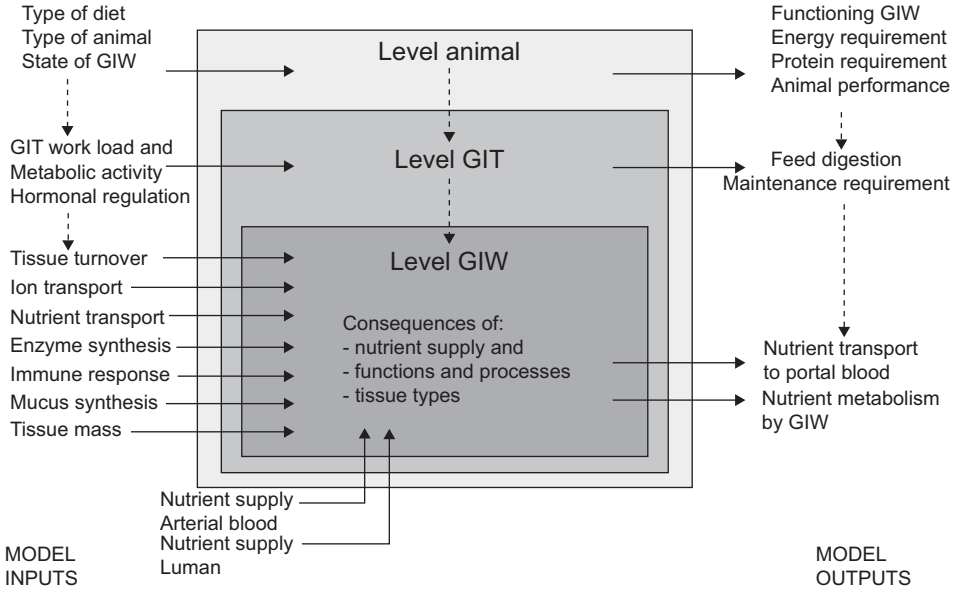


Figure 8.2 Schematic representation of the modeling approach at different levels of organization: the animal, the gastrointestinal tract (GIT), and tissue of the gastrointestinal wall (GIW). A distinction is made between model inputs (nutrient inputs, parameters for the physiological state of the GIT, productive functions), model representation (intracellular biochemical pathways of nutrients utilization), and model outputs (nutrient supply to portal blood, apparent nutrient utilization by the GIW).

the animal's history from a physiological perspective is often lacking. These historical aspects are important because they can interfere with regulation of gene expression and, hence, with observed physiological response, and be a cause of interanimal variation. Although it is probably less difficult to organize such historical observations in livestock science compared to the medical sciences, conducting such experiments is certainly hampered by time and costs. The review of Gluckman et al. (2009) is an illustrative example of the existence of such historic effects and discusses the developmental plasticity during development in order to let the organism match its environment. The influence of the environment can lead to stable changes in the epigenome of human individuals, which may make the individual more susceptible to the development of chronic cardiovascular and metabolic disease. The type of health problems and the life span of livestock obviously are not comparable to that of the humans. However, this does not exclude historic and epigenetic effects as important determinants of observed physiological function in farm animals.

It is concluded that besides the ability to handle and analyze -omics data to reveal the information they hold, additional modeling work is needed to quantify their physiological relevance. A physiological trait that demonstrates itself at different levels of organization, or at different moments in time, has to be represented as such in a model that aims to predict and explain under a wide range of management conditions.

Choice of Levels of Organization to Represent

With a different modeling objective, a different selection is made as to what levels of organization need to be represented in the model. For example, to predict the effect of SCD and DGAT polymorphisms in dairy cows on yield and composition of milk fat (Schennink et al., 2008; Heck et al., 2009), a model might be developed that includes a rather straightforward relationship between increased gene expression and activity of the stearoyl-CoA desaturase (SCD) and diacyl glycerol acyltransferase (DGAT) enzymes. This means that observations at the lowest level of organization (genotype) are associated directly with the highest level of organization (cow performance; milk fat yield). Such a relationship between DGAT gene expression and DGAT enzyme activity does not represent other key elements including translational and posttranslational control. Moreover, to predict the response of a cow with a certain genotype under various feeding regimes and farming conditions, other levels of organization also need to be taken into account. Nutrient supply and hormonal control of mammary function (as well as whole-cow function) will strongly affect animal performance in terms of production and composition of milk fat. Therefore, other aspects need to be included to become truly predictive (Figure 8.3). These aspects include nutrient supply (nutrition, composition of dietary crude fat, rumen fat metabolism; Dijkstra et al., 2000); intestinal fat digestion (Doreau and Chilliard, 1997); liver fat metabolism (Drackley et al., 2001); metabolism in adipose tissues (Baldwin, 1995); transport and udder uptake of nutrients and fatty acids from arterial blood (Volpe et al., 2010); and hormonal control and other regulatory processes of intracellular metabolism and secretory function of milk-secreting udder cells (Bionaz and Loor, 2007; Rius et al., 2010).

Adopted Model Representation

The representation adopted in the model varies with modeling objectives. A more mechanistic approach may be taken in which concepts and theories are represented explicitly to describe the mechanisms of how functioning of the biological system is organized. Alternatively, more empirical representations that describes the biological system directly from observations at the highest level observed and without presumptions and theories on the underlying mechanism, may be thought to suit a similar purpose. A comparison between dynamic, mechanistic growth models (France et al., 1987; Gerrits et al., 1997; Halas et al., 2004) and more empirical approaches (Oltjen et al., 1986; De Greef, 1992) shows that different assumptions and concepts were used (individual nutrients versus metabolizable energy, dynamic versus static, continuous versus discrete).

Another example is the distinct functionality of a more mechanistic approach to describe rumen fermentation in comparison to less mechanistic ones that chose not to represent concentration dependency of the enzyme- or protein-driven processes of feed degradation and microbial activity (Russell et al., 1992; Pitt et al., 1996).

There is no general rule as to whether a more mechanistic or a more empirical approach should be used. Growth performance and rumen fermentation can be modeled using both approaches, each having its own advantages and disadvantages (Thornley and France, 2007). It fully depends on the objectives of the exercise as to which approach suits purposes best. For practical application and extension, an

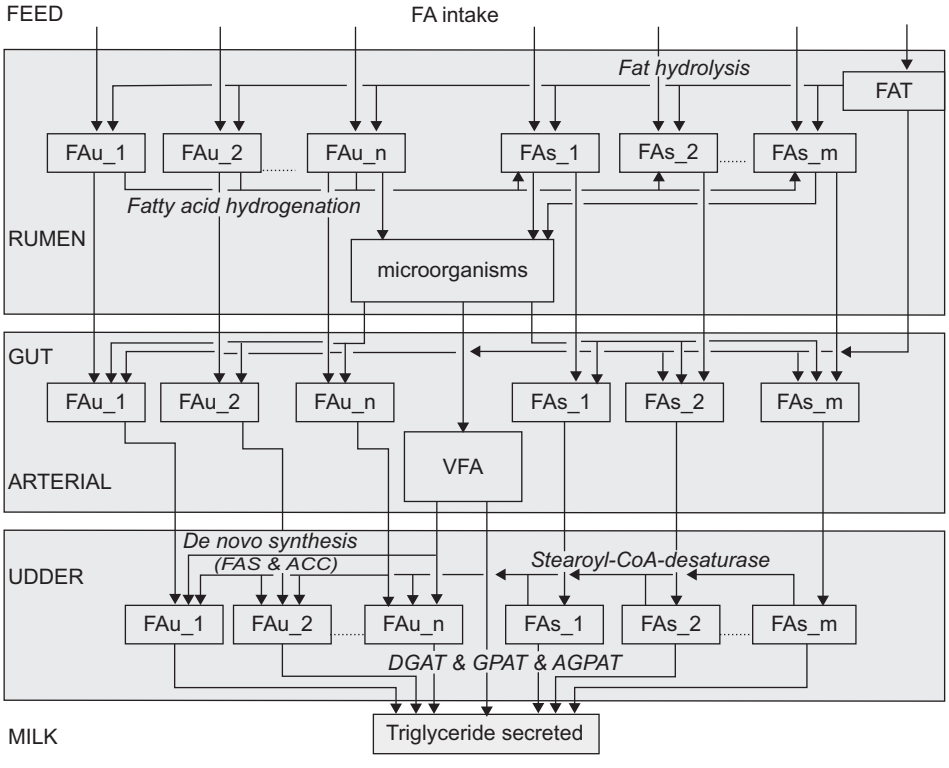


Figure 8.3 Simplified representation of the model elements needed to predict the variation in content of unsaturated fatty acids (FAu_i; i = 1. . .n) and saturated fatty acids (FAs_j; j = 1. . .m) in secreted milk fat. Elements that require representation are feed intake, dietary composition of FAu and FAs, microbial activity in the rumen responsible for lipolysis and biohydrogenation of unsaturated FA, arterial supply of VFA, FAu, and FAs to the secreting cells in the mammary gland. Arterial FAu and FAs represented as such, although the udder actually takes up plasma fatty acids in various forms (nonesterified FA, triglycerides circulating in chylomicrons, or very low-density lipoprotein). Not yet represented in the scheme, but highly relevant depending on stage of lactation and nutrition, are the utilization and supply of FAu and FAs by adipose tissue, and liver metabolism of VFA, FAu, and FAs (Chapter 9). Abbreviations used: ACC, acetyl-CoA carboxylase; FAS, fatty acid synthase; AGPAT, acyl glycerol phosphate acyltransferase; DGAT, diacyl glycerol acyltransferase; GPAT, glycerol 3-phosphate acyltransferase.

empirical model is often *a priori* thought more suitable because the use of practical indices of animal performance is relatively easy. However, with some modeling goals a more mechanistic approach may be required. For example, when the aim is to obtain an understanding at a certain level of organization, on the basis of a representation that makes use of elements from a lower level of organization and that addresses the concentration dependency of processes driven by enzymes or functional proteins. Also, with extrapolation to situations outside the area of data used to develop the empirical model, a mechanistic approach may be more advantageous. Thus, the specific aims of the modeling exercise dictates the approach adopted. The aim of an empirical

modeling approach will never be identical to that of a mechanistic approach, which makes them incomparable in a strict sense.

Representation of -Omics Data

With respect to modeling the meaning of -omics data in livestock science, the same decisions need to be made as with modeling physiological functions. The results of analyses of -omics data can be related directly to some observed physiological functions or even to whole-animal performance. This would be a rather empirical treatment of animal physiology aiming to associate -omics observations directly to observed variation in animal performance. On the other hand, an alternative approach can be chosen to represent the various physiological processes and underlying mechanisms, in an attempt to get an understanding of how physiological function is organized within the animal. This organization ranges from the lowest molecular level (chemical: nutrients, metabolites, gene regulatory factors, enzymes, transport proteins; physical: membranes, cell organelles, tissues, histology, morphology) to the highest whole-animal level of organization. Besides continuous relationships, models may also contain noncontinuous elements such as representation of discrete effects or logical operations. In particular with the onset of gene function, or when processes depend more on the presence or absence of hormones, for example, or when simply defining certain physiological or regulatory states in an empirical manner, use of such logical operations must be made.

Data gathered with -omics techniques are generally at the lowest level. Although perhaps valid from a chemical viewpoint, this approach does not necessarily hold from a physical and temporal (historical) point of view. Various levels of compartmentation exist, which means that a chemical compound can have different functions in different physical compartments. An example is the need to distinguish acetate in the cytosol of mammary gland cells and acetate in the mitochondrion (Volpe et al., 2010). Such compartmentation also exists from an anatomical point of view. For example, the location of the gastrointestinal tract sampled will have a decisive influence on experimental results. Also from a physicochemical point of view of compartmentation, different forms or states of the same chemical compound may well have to be represented (e.g., the distinction between different distribution pools). For example, for an understanding of absorption of volatile fatty acids (VFA) in an acid rumen environment, a distinction needs to be made between absorption rate of the dissociated and undissociated forms of these acids (Dijkstra et al., 1993). Furthermore, a distinction is necessary between the various states encountered for rumen papillae and their epithelia (Bannink et al., 2008; Penner et al., 2009).

Notwithstanding all these levels of organization that may be involved when interpreting -omics data, the aims of the modeling exercise determine which aspects should be treated in a rather empirical manner and which aspects should not be.

Modeling Approaches

Because the conceptual approach always reflects the aims of a modeling exercise, it is not determined as much by the wealth of data available, but rather depends on a vision

of how the system actually functions. This vision has to be available at the start of the modeling process, but certainly must develop further during the recurrent process of mathematical modeling, experimental testing and data gathering, and evaluation against knowledge presented in literature.

The intention of mathematical modeling in animal physiology should not be to duplicate life and to include as many details as possible, or with as much -omics detail as possible. Instead, it should attempt to distinguish the essential from the superfluous (France and Kebreab, 2008) and in this way assist in identifying the most relevant entities, concepts, and factors involved. Also, it should attempt to explain the animal physiological phenomena observed. Moreover, it has to be realized that these observations may also be the result of historical effects.

Systems Biology

Although "System Biology" obviously involves modeling at various levels of organization of animal physiology, we will not try to give a definition. The term Systems Biology includes a wide range of scientific research areas. Hood (2003) referred to Systems Biology as the integration of genome sequence and regulatory networks specifying gene behavior in order to find the logic of life of an organism. Similarly, Ideker (2004) defined Systems Biology as the integration of knowledge from diverse biological components and data into models of the system as a whole. Such definitions originate from situations in which modeling is applied to data gathered using high-throughput technologies. We would like to emphasize that for solving physiologically oriented problems in livestock science (related to feed production, animal production, animal metabolism, animal disease and behavior, and environmental and climate effects) the variability and dynamics of the physiological processes themselves are just as important as the regulatory mechanisms, which let that physiology be expressed. Preventing the uncoupling of cause and effect derived from data sources of different origin is crucial when developing a vision about the logic of animal life (Cornish-Bowden and Cárdenas, 2007). It means that variation in environmental, genetic, and physiological factors, and variation in gene expression may all be causally related, be of equal relevance, and become manifest in observational data in a fully confounded manner (Chapter 9).

The concept of Systems Biology is not new. Already in the sixties, Mihaljo Mesarovic (mathematician and engineer) commented that in spite of considerable interest and efforts at that time, the application of systems theory in biology has not quite lived up to expectation (Cornish-Bowden, 2005). We believe that Systems Biology will not be more than just a new name for already existing research (though practiced on ever-larger scales and equipment), if Systems Biology does not succeed in integrating knowledge from various disciplines and levels. Thus, if Systems Biology is to lead to expected benefits, a need for a shift in focus away from molecular characterization toward understanding functional activity is required. Indeed, integration of knowledge from a lower level may well give rise to new knowledge at the level of interest itself; the whole is more than the sum of the parts, as illustrated in various examples in Thornley and France (2007). Although the whole is more than the sum of the parts, it is explainable in terms of the parts and how they interact. Thus, in using -omics data

for modeling, focus should be on the understanding of variation in functional activity that arises from various molecular characterizations. This includes, for example, representing the causal factors (molecular mechanisms) involved with variation in the way gene expression is regulated under various physiological conditions. It also includes the various metabolic interactions between different molecules given a nonvarying gene expression.

Static and Empirical Approaches

Many kinds of mathematical equations can be used for constructing a model describing a physiological function. The choice of static equations means that dynamics and the time course of events cannot be handled by the model itself but have to be inputs. The choice of equations that do not contain entities defined at a lower level of organization means that the model is descriptive and unable to explain how changes in those entities are related to the physiological function observed. There are many examples of rather empirical approaches to address physiological problems in livestock production (feed intake, digestion, production, excretion, health, and welfare). Most models used in current practice probably have to be considered empirical and static. These may range from rules of the thumb (e.g., limits and advice for nutrition and management, avoiding risks of developing (subclinical) rumen acidosis and metabolic disorders) to more complex mathematical models predicting digestion, animal performance, and animal requirements (e.g., feed evaluation systems). Although static, empirical models may attempt to represent the underlying mechanisms at a lower level of organization (e.g., De Greef, 1992; Russell et al., 1992), dynamic, mechanistic models are without doubt more effective candidates to explore and explain in cases where physiological events have to be described as concentration-dependent processes (e.g., Dijkstra et al., 1992; Gerrits et al, 1997).

Mathematical Representation in Dynamic Models

With the development of any model, it is important to maintain internal consistency and a correct use of units (Baldwin, 1995). Only in this way a model fulfills physical, chemical, and physiological theories commonly accepted (e.g., balance of charge, energy, mass, and fluxes). Various text books are available on modeling quantitative aspects of the digestion and metabolism in farm animals, with various applications that range from modeling productive traits, modeling the impact of animal production on the environment, to modeling whole-farm systems (Dijkstra et al., 2005; Kebreab and France, 2008).

Because most physiological processes are concentration dependent and essentially nonlinear, mathematical representation of the underlying mechanisms mostly involves a dynamic system of ordinary nonlinear differential equations with a separate differential equation $dQ_{i,j}/dt$ to describe the rate of change of each state variable $Q_{i,j}$ (i.e., the quantity of i present in compartment j) at any time point t in response to the state of the system (rate:state formalism; Dijkstra et al., 2005). The $Q_{i,j}$ may be any entity represented in the model. The $dQ_{i,j}/dt$ can represent any physiological aspect, such as substrate degradation or absorption, nutrient utilization in specific tissues,

transport functions, and receptor binding. Many examples of this type of modeling can be given for physiologically oriented studies, ranging from growth models (Gerrits et al., 1997), microbial fermentation models (Dijkstra et al., 1992; Dijkstra, 1994), models predicting end-products of enzymatic digestion and microbial fermentation (Bannink et al., 2006b, 2009; Kebreab et al., 2009), models of mineral metabolism (Hill et al., 2008), models of organ function (Freetly et al., 1993; Hanigan and Baldwin, 1994), and infection models (Perelson and Weisbuch, 1997; Thornley and France, 2008, 2009). But, $dQ_{i,j}/dt$ can just as well represent changes in gene expression, enzyme activity, the effect of transcription factors, or any other associated regulatory and signaling mechanism investigated using -omics technologies. Similar to the abovementioned physiological examples, there are examples of the representation of gene expression and other -omics measurements (e.g., Zak et al., 2003). Stochasticity and modeling techniques to perform network analysis appears more relevant for the analysis of -omics data than for representing animal physiology. The obvious reason is the more straightforward relationship studied between influencing effect and physiological response. Description of such a relationship often considers a far more limited number of entities and a far more limited number of (types of) conditions to be analyzed by the model. The various modeling approaches used to represent gene regulatory networks are reviewed by De Jong (2002).

The remainder of this section gives an outline of the types of mathematical equations used to describe the relationships underlying observed physiological phenomena in deterministic, dynamic models to link information from various levels of organization in animal physiology. Parameters and units have been chosen arbitrarily and only serve to demonstrate the way a physiological process can be represented. Empirical, static approaches are often applied to represent physiological phenomena, but the mathematical equations are easy to comprehend and require no explanation here.

Conversion or Translocation Driven by Enzymes or Protein Functions

Most physiological processes are driven by protein activity. These may involve transport proteins binding to receptors or metabolic enzymes that catalyze the conversion of substrate. All these processes are essentially reversible and saturate with increasing rate and can be represented by equations similar to the basic Michaelis–Menten equation for substrate conversion:

$$v_1 = v_{\max 1}/(1 + M_1/[S_1]) \tag{8.1}$$

with

v_1 = actual conversion rate of substrate S_1 (mol S_1/d)
$v_{\max 1}$ = maximum conversion rate (mol S_1/d)
M_1 = affinity of enzyme for substrate S_1 (mol S_1/L)
$[S_1]$ = concentration of substrate S_1 (mol S_1/L)

Equation (8.1) describes the reversible binding of S_1 to enzyme, and $v_{\max 1}$ represents the effect of available enzyme concentration and M_1 the affinity of the enzyme for S_1. Both aspects may vary as a result of regulatory factors at the transcription level or the level of gene expression when considered proportional to transcription

rate. The parameters of this type of equation can be derived from either *in vivo* observations, knowledge of animal metabolism, or -omics studies and databases used to analyze -omics data. The problem and aim of the modeling exercise dictates to what extent general estimates from databases or literature apply. There are a wide variety of applications of this type of equation in the literature. For example, it has been applied to represent the interdependency between available hexose, ammonia, and soluble protein on rumen microbial activity and rumen digestive function of these microorganism (Dijkstra et al., 1992); the utilization of fatty acids in oxidative metabolism versus deposition of acetyl-CoA equivalents in adipose tissue in calves (Gerrits et al., 1997); and the interactions between bacterial infection, macrophages, cytokines, and lymphocytes (Gammack et al., 2005).

Multiple Effects

Equation (8.1) can be extended to represent the effect of another substrate S_2 alongside the converted substrate S_1. The effect of S_2 can be either stimulatory or inhibitory, and its effect on S_1 conversion can be represented in either an independent or dependent fashion. The following equations show an independent (Equation (8.2)) and a dependent (Equation (8.3)) stimulatory effect of S_2 on the conversion of S_1:

$$V_1 = v_{\max 1}/((1 + M_1/[S_1])(1 + M_2/[S_2])) \tag{8.2}$$

$$V_1 = v_{\max 1}/(1 + M_1/[S_1] + M_2/[S_2]) \tag{8.3}$$

with

M_2 = affinity of enzyme for substrate S_1 (mol S_2/L)
$[S_2]$ = concentration of substrate S_2 (mol S_2/L)

There are many examples and different expressions may be used. The choice of expression partly depends on the method used to estimate the parameters.

Competitive Inhibition

A special case of multiple effects is that of competitive inhibition between different substrates, with S_2 inhibiting the conversion of S_1 (Equation (8.4)) and S_1 inhibiting conversion of S_2 (Equation (8.5)):

$$v_1 = v_{\max 1}/((1 + M_1/[S_1])(1 + [S_2]/J_2)) \tag{8.4}$$

$$v_2 = v_{\max 2}/((1 + M_2/[S_2])(1 + [S_1]/J_1)) \tag{8.5}$$

with

v_1, v_2 = actual conversion rate of substrate S_1 and S_2, respectively (mol S_1 or S_2/d)
$v_{\max 1}, v_{\max 2}$ = maximum conversion rate for S_1 and S_2 respectively (mol S_1 or S_2/d)
M_1, M_2 = affinity of enzyme for substrate S_1 and S_2, respectively (mol S_1 or S_2/L)
J_1, J_2 = inhibitory constant for S_1 and S_2, respectively (mol S_1 or S_2/L)
$[S_1]$, $[S_2]$ = concentration of substrate S_1 and S_2, respectively (mol S_1 or S_2/L)

This form of expression was used by Bannink et al. (2008) to represent competitive inhibition between acetate, propionate, and butyrate for their activation by acyl-CoA-synthetases in rumen epithelium. Similar expressions also apply when representing competitive inhibition for multiple amino acids for mutual sodium-dependent transport sites in the mammary gland (Maas et al., 1998), and representing the effect of a specific inhibitor on enzyme activity such as the inhibiting effect of glutathione against glutamate on activity of glutamyl-cysteine synthetase (Reed et al., 2008).

Allosteric Effects, Threshold Values, or (Noncontinuous) Switch Functions

In reality, enzymatic reactions often follow a different pattern to that described by the basic Michaelis–Menten Equation (8.1). The binding between substrate and enzyme may be stimulated with increasing substrate concentration, leading to an acceleration of enzyme activity and substrate conversion. Such effects are referred to by the term allosteric effects, and result in a different equation form, known as Hill's equation, which allows for a steeper increase in enzyme activity with an increase in substrate concentration:

$$v_1 = v_{\max 1}/(1 + (M_1/[S_1])^{P_1}) \tag{8.6}$$

with P_1 = a value higher than 1 in case of a stimulatory effect, and a value between 0 and 1 in case of an inhibitory effect (no units).

Such allosteric effects occur in metabolic pathways in cells (Volpe et al., 2010), with the expression or transcription of genes (Matiatis and Reed, 2002), with translation of mRNA as discussed by Xu et al. (2001) for the effect of leucine on the mTOR-signaling pathway in pancreatic β-cells, and with effect of nutrient supply to the bovine mammary gland on milk protein synthesis (Rius et al., 2010).

Depending on the size of P_1, the curve for enzymatic conversion of S_1 becomes sigmoidal in shape (France et al., 2000). With high values of P_1, the sigmoidal shape becomes extreme and resembles a switch mechanism between no conversion ($[S_1] < M_1$) and maximum conversion ($[S_1] > P_1$). The concentration of S_1 at which such a switch will operate is determined by M_1, which acts as a threshold concentration for the onset of S_1 conversion.

Regulatory and Hormonal Effects

Hormonal effects can also be represented using Hill's equation described in the previous section. Various forms have been used, varying from a modulation of M and P through scaling of simulated hormone concentrations by a standard concentration (Gill et al., 1989), to inclusion of an independent term $1/(1+(M_h/[\text{hormone}])^{Ph})$ or $1/(1+([\text{hormone}]/J_h)^{Ph})$ to represent the stimulatory or inhibitory effect of hormones on a particular process with parameters $M_h, J_h,$ and P_h. The units of M_h and J_h depend on whether actual plasma concentrations or a relative scale compared to a reference value is used.

Conditional Statements and Logical Expressions

Representation of physiological processes using kinetic equations for enzymes and functional proteins can be a less suitable representation for some physiological aspects

or events. In particular, in cases where the quantity of molecules involved is very small or where single events or signals may evoke a physiological response, this form of representation is less suitable. In such cases, logical or Boolean expressions can be used, for example, to indicate the presence of certain cytokines that are produced in very small numbers (Thakar et al., 2007). Combinations of various modeling techniques have been applied to represent immune response as reviewed by Perelson and Weisbuch (1997). Further, logical expressions can be used to indicate whether a cow is in a certain physiological state or not, such as lactation, or presence of an immune challenge.

Anatomical, histological, and morphological aspects can be seen as aspects to be modeled conditionally, indicating a certain state. When of interest for representing the dynamics of enzyme activity or a physiological process, however, different compartments may be identified in the model, each with its own entities and with an exchange between compartments represented.

In addition to logical expressions, various conditional statements may be required to allow model description to correspond with generally accepted rules and theories for physical, chemical, and physiological processes. For example, a model has to satisfy the principle of mass conservation, thermodynamic principles, neutral electrical charge (balance of anions and cations), neutral metabolic charge (ATP production equals ATP utilization), acid–base chemistry, or other physicochemical aspects.

Representation of genotype in models of physiological aspects in animals is a special case of a conditional statement or a Boolean expression. The effect of genotype can be represented by modulation of the parameters values discussed in Chapter 6. Examples are the representation of the effect of DGAT polymorphisms (Schennink et al., 2008) in dynamic models that aim to predict the effects of nutrition and management as well as genotype on composition and production of milk fat (Shorten et al., 2004). The effect of polymorphisms on gene expression and translation rate affects the quantity of enzyme produced, hence, affecting parameter v_{max}. Other effects of polymorphisms that modulate the characteristics of an enzyme produced by the altered affinity for substrate(s) affect parameter *M.* Altered enzyme characteristics may also involve a changed susceptibility to regulatory factors or signals present or conducted to the same compartment accommodating the enzyme. When the model also attempts to represent the dynamics of the regulatory factor as well (i.e., the factor actually is a state variable in the model), the effect of a regulatory factor has to be represented by an additional term in the mathematical equation for enzyme activity (e.g., Equations (8.2)–(8.5) versus Equation (8.1)). Otherwise, the regulatory effect is represented by an altered parameterization of the equation of enzyme activity (e.g., Equation (8.6) versus Equation (8.1)). Regulatory effects may be a stimulatory, inhibitory, or any other modulatory effect on enzyme activity and can be parameterized by M, J, or P parameters. Which terms have to be included and what parameterizations need to be chosen, has to be deduced from insights gained from experimental data.

Historic and Memory Effects

The cumulative effect of a condition, challenge, or nutrient load in time can be represented in a model using a historic or memory function. The principle has been described by Neal and Thornley (1983) in a model representing mammary gland

tissue development and lactation in time. The same concept was used to represent the adaptation of rumen epithelia in a dynamic model of production, epithelial transport, epithelial metabolism, and portal appearance of the VFA (acetate, propionate, butyrate) in the rumen of cows (Bannink et al., 2008). The cumulative effect of the load of VFA was used as an initiator of morphological changes in rumen epithelia. It was postulated that this makes sense because the load of acid to be transported is a relevant factor for the amount of epithelial tissue needed to manage this acid (Dirksen et al., 1984). Another example is representation of the effect of variation in genotype and the effect of compensatory growth in a model that predicts the performance of veal calves with varying nutritional strategies (Gerrits et al., 2000). Related to this is the need to identify the effect of management history on ontogenic aspects (Conner et al., 2010). As a final example, modeling of historic effects is important when the aim is to become predictive instead of diagnostic with respect to the effect of management on animal health. Also, for identifying the relationship between lactation performance and cow health, indicators of metabolic imbalance and immune competence are needed. Such indicators may guide the nutrition of high-yielding cows to account for their genotype as well as management history. Studying such historic effects using -omics technology can help trace the effects of nutritional management from a historic perspective and elucidate how in time the interaction between genotype and nutrition lead to a compromised health status of modern dairy cows.

The aim of including historic or tissue memory effects, to be traced from analysis of -omics data, into a model can be achieved by a representation of a memory function. An additional entity or state variable may be introduced describing the cumulative or historic effect, which can be formulated to have its effect on a change in parameters values discussed in Sections "Conversion or translocation driven by enzymes or protein functions" through "Conditional statements and logical expressions."

Some Examples of Physiological Aspects

Interaction between Nutrition and Digestive Physiology

The extent and site of feed digestion in the gastrointestinal tract of farm animals determines the amount and profile of nutrients available for productive functions (e.g., growth, milk, offspring) and maintenance aspects (e.g., metabolic costs of physiological functions such as transport, tissue turn-over, secretions, immune response, basal cell metabolism). However, digestion also affects the tissues directly responsible for it (Bannink et al., 2006a). The tissues in the rumen wall in dairy cows are of interest because of the large amount of end products of rumen microbial fermentation that are absorbed, transported, and metabolized by them. The case of cattle is of interest because of the intensive feeding regimes they are exposed to. Tissues of the rumen wall adapt to these feeding regimes, but literature suggests that certain nutritional management or cow susceptibility may lead to subclinical rumen acidosis, which is mentioned as a major cause of many health problems in cattle through lactation (Enemark et al., 2002). The concept of subclinical acidosis is also of practical significance because a high level of rumen acidity can have a detrimental effect on fiber digestion (Dijkstra et al., 1992) and hence decrease the efficiency of utilization of roughages.

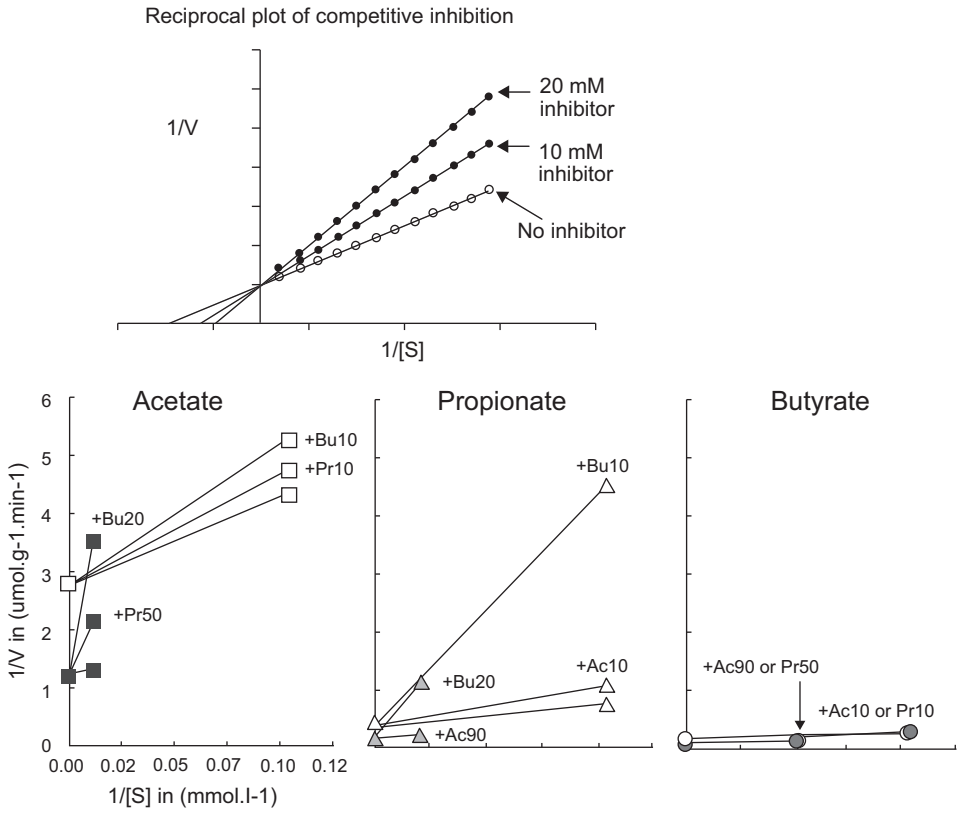


Figure 8.4 Comparison of enzyme assays in rumen epithelium by Harmon et al. (1991; closed symbols) and Ash and Baird (1973; open symbols). The effect of inhibiting VFA on the activity of VFA activation is demonstrated by a double-reciprocal plot of cosynthetase activity (V in μmol/g tissue/min) and VFA-concentration of the activated VFA type (S in mmol/L); (A) acetate, □ and Ac; (B) propionate, △ and Pr; (C) butyrate, ○ and Bu. Codes and numbers that guide the symbols indicate the type and concentration (in mmol/L) of inhibiting VFA (absence of a guiding code indicates absence of inhibiting VFA). (The graphs were reproduced with permission from Bannink et al., 2008.)

Prediction of the effect of nutrition on rumen wall function first of all requires prediction of microbial activity and the amount and type of VFA produced (Bannink et al., 2006b, 2008). Next, the effect of absorbed VFA on rumen wall tissues needs to be understood. Response of these tissues to VFA exposure may be alteration of the enzymatic activity per unit tissue. Also, tissue morphology may be altered affecting the amount of tissue, and hence the amount of enzyme, present in whole tissue. However, early enzyme assay studies indicated a relatively small effect of nutrition on enzymatic activity in tissue samples in cattle (Figure 8.4; Bannink et al., 2008). The amount of enzyme available would have to be represented by the parameter v_{max} discussed in section "Modeling approaches," affinity of the enzyme for a substrate by parameter M, and regulatory effects on enzyme activity by parameters J and P.

Also, more recent work on the expression of genes encoding for acylCoA-synthetases of acetate, propionate, and butyrate give either rather small effects or results, which remain difficult to interpret (Baldwin et al., 2007). Conner et al. (2010) reviewed the effect of nutrition on gene expression in digestive tissues of ruminants but focused on data available for developmental changes in tissues of juvenile ruminants. They concluded that a thorough understanding of the mechanism involved requires more detailed knowledge of gene expression during various physiological states to well-designed, more-complex functional investigation and hypothesis-driven experiments. Although this may be true, it cannot be discounted at the moment that gene expression of metabolic pathways and transport functions might not change dramatically with nutritional strategy. It may be that the adaptation response mainly consists of a morphological change of the rumen wall (Dirksen et al., 1984; Bannink et al., 2008, 2010a) without apparent changes in gene expression per unit of tissue as established in early enzyme assays, or even a lower gene expression (Penner et al., 2009), which may also have been affected by morphological changes.

Discussion in the current literature on rumen function makes it clear that simply collecting -omics data on gene expression in rumen wall tissue, without giving proper consideration to the quantification of rumen physiological aspects (microbial fermentation, rumen wall morphology, and function), is unlikely to generate the insight needed to be able to predict the risk of acidosis and abnormal rumen function (Penner et al., 2009; Conner et al., 2010). Current results remain too inconclusive, whereas the problem of acidosis and related health problems in cattle remain prevalent in practice (Enemark et al., 2002). A breakthrough in the current impasse might be forced with a combined approach of using new -omics technologies and simultaneously making efforts to improve physiological-oriented predictive models.

Interaction between Nutrition, Methanogenesis, Fat Metabolism, and Milk Fat Quality

Currently, there is much interest in both modulation of milk fatty acid composition and mitigation of methane emission in dairy cows. Both may, in principle, be achieved by nutritional measures as well as animal breeding. A genetic background to milk fat composition has been demonstrated (Schennink et al., 2008), although nutrition can be considered to have an equal or larger impact (Bionaz and Loor, 2007). A genetic background to methanogenesis, other than the common effect of feed intake, feed digestibility, and cow productivity, remains to be established, however. There is much interest in profiling microorganisms in the rumen and finding a genetic background of differences among individuals for this profile and accompanying methane production. However, there are limitations to what simply profiling the microbiota might explain. Recently, Hook et al. (2009) investigated the quantity and diversity of methanogens in the rumen of dairy cows subjected to long-tern feeding of monensin to reduce methanogenesis. In contrast to a reduction of methanogenesis with monensin by 6% for more than 6 months (Odongo et al., 2007), Hook et al. (2007) were not able to establish any effect on the quantity or diversity of methanogens in the rumen. This result confirms that there may be a serious risk just demonstrating the presence and profile of microbiota in the rumen will not give an indication of their activity and of

the consequences for rumen function. As with the previous example, it is probably necessary to combine the use of -omics techniques with functional measurements of methanogenesis or with modeling rumen physiology to become predictive with respect to methanogenesis. Dijkstra et al. (1992, 1994) used three microbial entities (cellulolytic bacteria, amylolytic bacteria, protozoa) and their mutual interaction (predation by protozoa) to represent the most relevant processes for rumen fermentation. It is likely that this representation is also required to address details on activity of methanogens that may become available in the near future.

Again, the potential appears high for using -omics technologies to gain further insight into the effect of nutritional strategies on the dynamics of methanogens and their activity in the rumen environment. With this insight it can probably be answered with more confidence whether it is feasible to breed for low-methane individuals (apart from the ongoing breeding for higher milk yield and hence less methane per unit of milk produced) without unforeseen detrimental effects.

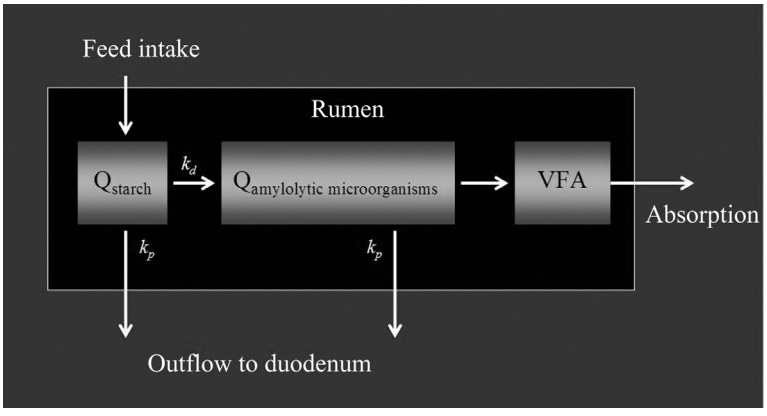
The other current interest, changing milk fat composition, may also be related to methanogenesis. Vlaeminck et al. (2006) showed the pattern of milk fatty acids to be correlated with the profile of VFA produced in the rumen (next to feed intake and amount of organic matter fermented, the most important determinant of methane; Figure 8.5).

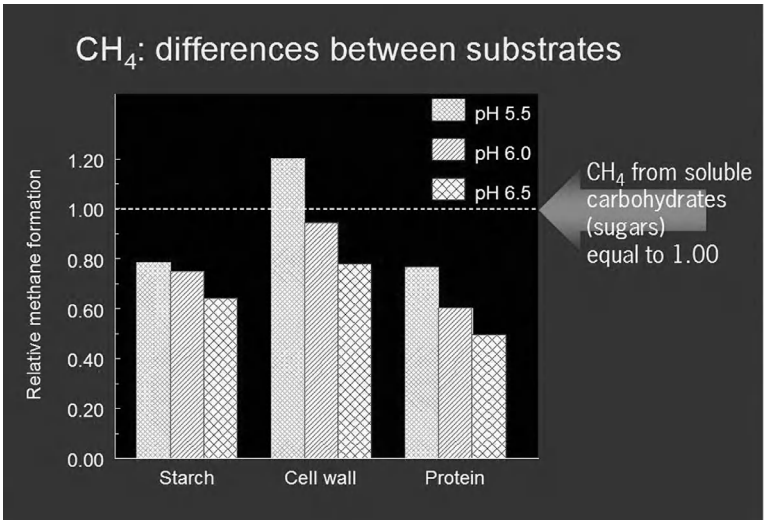
This suggests that milk fat composition may be an indicator of methanogenesis. However, milk fat composition is not only dependent on the rumen profile of the microbiota and related production of VFA, hydrogen, and methane (Figure 8.5). It is also affected by the fatty acid profile of dietary fat and its consequences for rumen microbial activity (Dijkstra et al., 2000), for the profile of fatty acids transported to the mammary gland, and for fatty acid metabolism and *de novo* synthesis of fatty acid in mammary secretory cells (Bionaz and Loor, 2007; Figure 8.3). These processes again may be influenced by genetic background of the cow (Schennink et al., 2008). Furthermore, mobilization of fat from adipose tissue during early lactation and the strong effects of nutrition and stage of lactation on the amount of milk fat produced (e.g., Odongo et al., 2007, established a 9% reduction in milk fat in combination with a 6% reduction of methane), complicates things even further and means that the milk fatty acid profile is probably confounded by a large number of factors besides VFA in the rumen.

Although proving very complex, there are good arguments to explore further the regulatory mechanism of fatty acid supply to the rumen environment, to the cow's intermediary metabolism and to the mammary secretory cells. For application and accurate prediction in practice, predictive models need to be developed that allow combined evaluation of all relevant factors with respect to milk fat composition, but also with respect to rumen fermentation, digestion, and cow productivity.

Interaction between Nutrition, Preabsorptive and Postabsorptive Processes

The profile of nutrients available for uptake by the mammary secretory cells of the cow (VFA, various forms of long-chain fatty acids, glucose, amino acids) depends on the extent and the gut compartment in which digestion takes place (preabsorptive aspects), and the effects on mobilization of body reserves, nutrient partitioning between





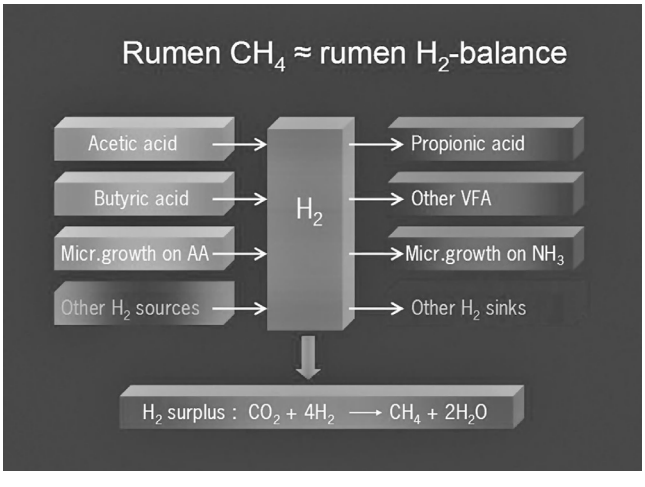


Figure 8.5 (A) Diagram of dynamic model for substrate and microorganisms, VFA, and methane; (B) effect of fermented substrates on methane; (C) diagram of rumen H_2-balance as outcome of VFA production and microbial growth (Bannink et al., 2010b).

organs and nutrient utilization by mammary secretory cells. At present, new insights are being gained into the regulation of milk protein synthesis with regard to dependency of nutrient supply and hormones by use of -omics technologies to elucidate the molecular mechanisms (Rius et al., 2010), including mTOR signaling proteins, protein kinase, and ribosomal protein S6. To allow postabsorptive models to provide accurate predictions, crucial aspects need to be included as represented in models of mammary gland metabolism (Hanigan and Baldwin, 1994; Shorten et al., 2004; Volpe et al., 2010) and of adipose tissue metabolism (Baldwin, 1995). Moreover, they need to be linked to models of digestive or preabsorptive processes (e.g., Dijkstra et al., 1992; Bannink et al., 2006a, 2006b).

Many aspects have not been accommodated in such models, for example, the recent finding of Van Knegsel et al. (2007) that extent of negative energy balance and mobilization of body reserves is affected by amount of glucogenic nutrients. A postabsorptive model needs to accommodate the preabsorptive effects of nutrition, management, and cow genotype as well to become predictive in practice. Both aspects are of high relevance and without doubt, further development of pre- as well as postabsorptive models to predict cow performance and milk synthesis benefits from the use of modern -omics technologies to elucidate the mechanisms involved and test hypothesis *in vivo*. The challenge lies ahead to demonstrate where the largest achievements can be made.

Implications and Perspectives

Regulation of physiological processes in farm animals is complex and covers several levels of organization. Furthermore, the regulatory mechanisms are of a highly dynamic nature and the result of changing environmental influences (nutrition, management, disease), physiological factors (phenotypic traits), and regulatory mechanisms (regulation of gene expression). Sufficient comprehension of these dynamics is needed to address most problems in livestock science, instead of considering animal characteristics and responses fixed. Such comprehension requires that different levels of biological organization need to be included in modeling efforts that aim to become predictive. As much use as possible has to be made of both the outcomes of data analyses at the level at which -omics data are gathered, and observation of physiological aspects at higher levels of organization. The specific aim of the modeling effort determines the boundaries set upon model representation, and the assumptions and concepts adopted. Which levels of organization require representation, and which entities need to be defined, depends on the modeling objectives and the vision one has of how observed physiological function is organized. This vision is created by both -omics data and physiological insights, instead of either one. Modeling functional aspects of animal physiology is always needed when the aim is to address physiological problems. Otherwise, the risk of becoming not predictive must be considered real. Current developments in Systems Biology appear very promising. When combined with modeling of physiological aspects in livestock sciences, there appears to be great potential for actually becoming more predictive, thereby assisting industry and farmers in solving problems. Linking data from various levels of organization and from various disciplines can lead to the development of models that are able to separate causes, and

which may be an improvement upon current insights in practice. Improved predictive capacity is the challenge that needs to be addressed. Instead of investing in large-scale monitoring studies, research activities are better directed at finding causal relationships and testing hypothesis formulated with the help of mathematical modeling. The latter should not be limited to modeling large volumes of -omics data alone.

Acknowledgments

The authors like to thank their colleagues in the Centre of Animal Nutrition and the Animal Breeding and Genomics Centre for helpful discussions. Part of the work was financed by the Dutch Ministry of Agriculture, Nature Conservation, and Food Quality under the Basic Knowledge program "Systems Biology." Funding through the Canada Research Chairs Program is also acknowledged.

References

Ash, R., & Baird, G.D. (1973) Activation of volatile fatty acids in bovine liver and rumen epithelium. *Biochemical Journal* **136**, 311–319.

Baldwin, R.L. (1995) *Modeling Digestion and Ruminant Metabolism*. Chapman & Hall, London.

Baldwin, R.L., El-Kadi, S.W., McLeod, K.R., et al. (2007) Intestinal and ruminal epithelial and hepatic regulatory gene expression as affected by forage to concentrate ratio in bulls. In: *Proceedings 2nd International Symposium on Energy and Protein Metabolism and Nutrition*, Vichy, France (Eds. I. Ortigues_Marty, N. Miraux, & W. Brand-Williams), pp. 293–294. EAAP publication 124, Wageningen Academic Publishers, Wageningen.

Baldwin, R.L., Smith, N.E., Taylor, J., et al. (1980) Manipulating metabolic parameters to improve growth rate and milk secretion. *Journal of Animal Science* **51**, 1416–1428.

Bannink, A., Dijkstra, J., & France, J. (2010a) Representing tissue mass and morphology in mechanistic models of digestive functions in ruminants. In: *Modelling Nutrient Utilization in Farm Animals* (Eds. D. Sauvant et al.), pp. 171–177. Wageningen Academic Publishers, Wageningen.

Bannink, A., Dijkstra, J., Koopmans, S.-J., et al. (2006a) Physiology, regulation and multifunctional activity of the gut wall: a rationale for multicompartmental modelling. *Nutrition Research Reviews* **19**, 227–253.

Bannink, A., Ellis, J.L., France, J., et al. (2009) Prediction of starch digestion in the small intestine of lactating cows. In: *Proceedings of the XIth International Symposium on Ruminant Physiology. Digestion, metabolism, and effects of nutrition on reproduction and welfare*, Clermont-Ferrand, France (Eds. Y. Chilliard et al.), pp. 68–69. Wageningen Academic Publishers, Wageningen.

Bannink, A., France, J., Lopez, S., et al. (2008) Modelling the implications of feeding strategy on rumen fermentation and functioning of the rumen wall. *Animal Feed Science and Technology* **143**, 3–26.

Bannink, A., Kogut, J., Dijkstra, J., et al. (2006b) Estimation of the stoichiometry of volatile fatty acid production in the rumen of lactating cows. *Journal of Theoretical Biology* **238**, 36–51.

Bannink, A., Smits, M.C.J., Kebreab, E., et al. (2010b) Simulating the effects of grassland management and grass ensiling on methane emission from lactating cows. *Journal of Agricultural Science, Cambridge* **148**, 55–72.

Beerda, B., Ouweltjes, W., Šebek, L.B.J., et al. (2007) Effects of genotype by environment interactions on milk yield, energy balance, and protein balance. *Journal of Dairy Science* **90**, 219–228.

Bionaz, M. & Loor, J.J. (2007) Identification of reference genes for quantitative real-time PCR in the bovine mammary gland during the lactation cycle *Physiological Genomics* **29**, 312–319.

Conner, E.E., Li, R.W., Baldwin, R.L., et al. (2010) Gene expression in the digestive tissues of ruminants and theor relationships with feeding and digestive processes. *Animal* **4**, 993–1007.

Cornish-Bowden, A. (2005) Making systems biology work in the 21st century. *Genome Biology* **6**, 317.

Cornish-Bowden, A. & Cárdenas, M.L. (2005) Systems biology may work when we learn to understand the parts in terms of the whole. *Biochemical Society Transactions* **33**, 516–519.

Cornish-Bowden, A. & Cárdenas, M.L. (2007) Vertical integration from 'omics' to the whole organism. In: *Proceedings 2nd International Symposium on Energy and Protein Metabolism and Nutrition*, Vichy, France (Eds. I. Ortigues_Marty, N. Miraux, & W. Brand-Williams), pp. 463–472. EAAP publication 124, Wageningen Academic Publishers, Wageningen.

De Greef, K.H. (1992) Prediction of production. Nutrition induced tissue partitioning in growing pigs. Ph.D. Thesis, *Wageningen Agricultural University*, Wageningen, NL.

De Jong, H. (2002) Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. *Journal of Computational Biology* **9**, 67–103.

Dijkstra, J. (1994) Simulation of the dynamics of protozoa in the rumen. *British Journal of Nutrition* **72**, 679–699.

Dijkstra, J., Boer, H., Van Bruchem, J., et al. (1993) Absorption of volatile fatty acids from the rumen of lactating dairy cows as influenced by volatile fatty acid concentration, pH and rumen fluid volume. *British Journal of Nutrition* **69**, 395–396.

Dijkstra, J., Forbes, J.M., & France, J. (2005) *Quantitative Aspects of Ruminant Digestion and Metabolism*. CAB International, Wallingford, UK.

Dijkstra, J., Gerrits, W.J.J., Bannink, A., et al. (2000) Modelling lipid metabolism in the rumen. In: *Modelling Nutrient Utilization in Farm Animals* (Eds. J.P. McNamara, J. France, & D.E. Beever), pp. 25–36. CAB International, Wallingford, UK.

Dijkstra, J., Neal, H.D. St. C., Beever, D.E., et al. (1992) Simulation of nutrient digestion, absorption and outflow in the rumen: model description. *Journal of Nutrition* **122**, 2239–2256.

Dirksen, G., Liebich, H.G., Brosi, G., et al. (1984) Morphologie der pansenschleimhaut und fettsaureresorption beim Rind- bedeutende faktoren fur gesundheit und Leistung. *Zentralblatt fur Veterinarmedizin* **A 31**, 414–430.

Doreau, M. & Chilliard, Y. (1997) Digestion and metabolism of dietary fat in farm animals. *British Journal of Nutrition* **78**(Suppl. 1), S15–S35.

Drackley, J.K., Overton, T.R., & Douglas, G.N. (2001) Adaptations of Glucose and Long-Chain Fatty Acid Metabolism in Liver of Dairy Cows during the Periparturient Period. *Journal of Dairy Science* **84**, E100–E112.

Ellis, J.L., Dijkstra, J., Kebreab, E., et al. (2008) Aspects of rumen microbiology central to mechanistic modelling of methane production in cattle. *Journal of Agricultural Science, Cambridge* **146**, 213–233.

Enemark, J.M.D., Jørgensen, R.J., & Enemark, P.S. (2002) Rumen acidosis with special emphasis on diagnostic aspects of subclinical rumen acidosis: a review, *Veterinarija ir Zootechnika* **20**, 16–29.

France, J. & Kebreab, E. (2008) *Mathematical Modelling in Animal Nutrition*, 574 pp. CAB International, Wallingford, UK.

France, J., Dijkstra, J., Dhanoa, M.S., et al. (2000) Estimating the extent of degradation of ruminant feeds from a description of their gas production profiles observed in vitro: derivation of models and other mathematical considerations. *British Journal of Nutrition* **83**, 143–150.

France, J., Gill, M., Thornley, J.H.M., et al. (1987) A model of nutrient utilization and body composition in beef cattle. *Animal Production* **44**, 371–385.

Freetly, H.C., Knapp, J.R., Calvert, C.C., et al. (1993) Development of a mechanistic model of liver metabolism in the lactating cow. *Agricultural Systems* **41**, 157–195.

Gammack, D., Ganguli, S., Marino, S., et al. (2005) Understanding the immune response in tuberculosis using different mathematical models and biological scales. *Multiscale Modelling & Simulation* **3**, 312–345.

Gerrits, W.J.J., Dijkstra, J., & France, J. (1997) Description of a model integrating protein and energy metabolism in preruminant calves. *Journal of Nutrition* **127**, 1229–1242.

Gerrits, W.J.J., Van Der Togt, P.L., Dijkstra, J., et al. (2000) Evaluation of a growth model of preruminant calves and modifications to simulate short-term responses to changes in protein intake. In: *Modellling Nutrient Utilization in Farm Animals* (Eds. J.P. McNamara, J. France, & D.E. Beever), pp. 163–174. CAB International, Wallingford, UK.

Gill, M., Beever, D.E., & France, J. (1989) Biochemical bases needed for the mathematical representation of whole animal metabolism. *Nutrition Research Reviews* **2**, 181–200.

Gluckman, P.D., Hanson, M.A., Buklijas, T., et al. (2009) Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. *Nature Reviews Endocrinology* **5**, 401–408.

Halas, V., Dijkstra, J., Babinszky, L., et al. (2004) Modelling of nutrient partitioning in growing pigs to predict their anatomical body composition. 1. Model description. *British Journal of Nutrition* **92**, 707–723.

Hanigan, M.D. & Baldwin, R.L. (1994) A mechanistic model of mammary gland metabolism in the lactating cow. *Agricultural Systems* **45**, 369–419.

Harmon, D.L., Gross, K.L., Krehbiel, C.R., et al. (1991) Influence of dietary forage and energy intake on metabolism and acyl-CoA synthetase activity in bovine ruminal epithelial tissue. *Journal of Animal Science* **69**, 4117–4127.

Heck, J.M.L., van Valenberg, H.J.F., Dijkstra, J., et al. (2009) Seasonal variation in the Dutch bovine raw milk composition. *Journal of Dairy Science* **92**, 4745–4755.

Hill, S.R., Knowlton, K.F., Kebreab, E., et al. (2008) A model of phosphorus digestion and metabolism in the lactating dairy cow. *Journal of Dairy Science* **91**, 2021–2032.

Hood, L. (2003) Systems biology: integrating technology, biology, and computation. *Mechanisms of Ageing and Development* **124**, 9–16.

Hook, S.E., Northwood, K.S., Wright, A.-D.G., et al. (2009) Long-term monensin supplementation does not significantly affect the quantity or diversity of methanogens in the rumen of the lactating dairy cow. *Applied and Environmental Microbiology* **75**, 374–380.

Ideker, T. (2004) Systems biology—what you need to know. *Nature Biotechnology* **22**, 473–475.

Johnson, D.E., Johnson, K.A., & Baldwin R.L. (1990) Changes in liver and gastrointestinal tract energy demands in response to physiological workload in ruminants. *Journal of Nutrition* **120**, 649–655.

Kebreab, E. & France, J. (2008) *Mathematical Modelling in Animal Nutrition*. CAB International, Wallingford, UK.

Kebreab, E., Dijkstra, J., Bannink, A., et al. (2009) Recent advances in modeling nutrient utilization. *Journal of Animal Science* **87**, E111–E122.

Maas, J.A., France, J., Dijkstra, J., et al. (1998) Application of a mechanistic model to study competitive inhibition of amino acid uptake by the lactating bovine mammary gland. *Journal of Dairy Science* **81**, 1724–1734.

Matiatis, T. & Reed, R. (2002) An extensive network of coupling between gene expression machines. *Nature* **416**, 499–506.

Neal, H.D.S.C. & Thornley, J.H.M. (1983) The lactation curve in cattle: a mathematical model of the mammary gland. *Journal of Agricultural Science, Cambridge* **101**, 89.

Odongo, N.E., Bagg, R., Vessie, G., et al. (2007) Long-term effects of feeding monensin on methane production in lactating dairy cows. *Journal of Dairy Science* **90**, 1781–1788.

Oltjen, J.W., Bywater, A.C., Baldwin, R.L., et al. (1986) Development of a dynamic model of beef cattle growth and composition. *Journal of Animal Science* **62**, 86–97.

Penner, G.B., Taniguchi, M., Guan, L.L., et al. (2009) Effect of dietary forage to concentrate ratio on volatile fatty acid absorption and the expression of genes related to volatile fatty acid absorption and metabolism in ruminal tissue. *Journal of Dairy Science* **92**, 2767–2781.

Perelson, A.S. & Weisbuch, G. (1997) Immunology for physicists. *Reviews of Modern Physics* **69**, 1219–1267.

Pitt, R.E., Van Kessel, J.S., Fox, D.G., et al. (1996) Prediction of ruminal volatile fatty acids and pH within the net carbohydrate and protein system. *Journal of Animal Science* **74**, 226–244.

Reed, M.C., Thomas, R.L., Pavisic, J., et al. (2008) A mathematical model of glutathione metabolism. *Theoretical Biology and Medical Modelling* **5** (doi:10.1186/1742-4682-5-8).

Rius, A.G., Appuhamy, J.A.D.R.N., Cyriac, J., et al. (2010) Regulation of protein synthesis in mammary glands of lactating dairy cows by starch and amino acids. *Journal of Dairy Science* **93**, 3114–3127.

Russell, J.B., O'Connor, J.D., Fox, D.G., et al. (1992) A net-carbohydrate and protein system for evaluating cattle diets. I. Ruminal fermentation. *Journal of Animal Science* **70**, 3551–3561.

Schennink, A., Heck, J.M.L., Bovenhuis, H., et al. (2008) Milk fatty acid unsaturation: genetic parameters and effects of Stearoyl-CoA Desaturase (SCD1) and Acyl CoA:Diacylglycerol Acyltransferase 1 (DGAT1). *Journal of Dairy Science* **91**, 2135–2143.

Shorten, P.R., Pleasants, T.B., & Upreti, G.C. (2004) A mathematical model for mammary fatty acid synthesis and triglyceride assembly: the role of stearoyl CoA desaturase (SCD). *Journal of Dairy Research* **71**, 385–397.

Thakar, J. Pilione, M., Kirimanjeswara, G., et al. (2007) Modeling systems-Level regulation of host immune responses. *PLoS Computational Biology* **3**, e109. doi:10.1371/journal .pcbi.0030109.

Thornley, J.H.M. & France, J. (2007) *Mathematical Models in Agriculture*, revised 2nd Edition, 906 pp. CAB International, Wallingford, UK.

Thornley, J.H.M. & France, J. (2008) Modelling bovine spongiform encephalopathy. *Journal of Agricultural Science, Cambridge* **146**, 183–194.

Thornley, J.H.M. & France, J. (2009) Modelling foot and mouth disease. *Preventative Veterinary Medicine* **89**, 139–154.

Van Knegsel, A.T.M., Van den Brand, H., Dijkstra, J., et al. (2007) Dietary energy source in dairy cows in early lactation: energy partitioning and milk composition. *Journal of Dairy Science* **90**, 1467–1476.

Vlaeminck, B., Fievez, V., Tamminga, S., et al. (2006) Milk odd- and branched-chain fatty acids in relation to the rumen fermentation pattern. *Journal of Dairy Science* **89**, 3954–3964.

Volpe, V., Cant, J.P., Boston, R.C., et al. (2010) Development of a dynamic mathematical model for investigating mammary gland metabolism in lactating cows. *Journal of Agricultural Science, Cambridge* **148**, 31–54.

Woelders, H., te Pas, M.F.W., Bannink, A., et al. (2011) Systems biology in animal sciences. *Animal* (in press). doi:10.1017/S1751731111000036.

Xu, G., Kwo, G., Cruz, W.S., et al. (2001) Metabolic regulation by leucine of translation initiation through the mTOR signaling pathway by pancreatic β-cells. *Diabetes* **50**, 353–360.

Zak, D.E., Gonye, G.E., & Schwaber, J.S. (2003) Importance of input perturbations and stochastic gene expression in the reverse engineering of genetic regulatory networks: insights from an identifiability analysis of an in silico network. *Genome Research* **13**, 2396–2405.

Chapter 9
Systems Biology and Animal Nutrition: Insights from the Dairy Cow During Growth and the Lactation Cycle

Juan J. Loor, Massimo Bionaz, and Guido Invernizzi

Introduction

The biological complexity of agricultural animals unavoidably requires a Systems Biology approach, i.e., a way to systematically study the complex interactions in the animal using a method of integration instead of reduction (Loor and Cohick, 2009). Important goals of Systems Biology are to uncover the underlying links (pathways, regulatory networks, and structural organization) within and between tissues (e.g., adipose and liver, mammary fat pad, and mammary parenchyma), and also to discover new emergent properties that may arise from examining the interactions between all components of a system. This integrative approach provides the means to arrive at a holistic view of how the organism functions (Bruggeman and Westerhoff, 2007). Work in model organisms during the past 15 years has demonstrated the applicability of high-throughput technologies to discern biological networks (Lin and Qian, 2007; Feist and Palsson, 2008).

The advent of high-throughput sequencing, transcriptomics, and proteomics technologies has dramatically accelerated the rate at which biological and genetic information can be collected from agricultural animals (Lippolis and Reinhardt, 2008; Loor, 2010). In the context of animal nutrition and animal metabolism, a substantial body of work linking the transcriptome and the genome has been conducted in chickens (e.g., Cogburn et al., 2007). Fewer studies have addressed the role of nutrition on transcriptional adaptations in other agricultural species, let alone attempted a more holistic approach using "omics" and bioinformatics tools. Despite progress in the area of transcriptomics and bioinformatics in livestock research, application of a Systems Biology approach is still in its infancy.

Dairy cows represent both an economically important livestock species and a unique biological model of mammalian adaptations partly because of an unrivaled requirement for dietary energy (Allen et al., 2005). Extensive microbial fermentation in the rumen (forestomach) results in a constant reliance on gluconeogenesis, a process that

Systems Biology and Livestock Science, First Edition. Edited by Marinus F.W. te Pas, Henri Woelders, and André Bannink.

increases dramatically immediately after parturition. The transition from pregnancy into lactation in mammals, and particularly dairy cattle, is characterized by metabolic adaptations in major organs (e.g., mammary, liver, adipose) that allow the animal to adjust to the need of synthesizing milk for the neonate. Different physiological stages during growth or the lactation cycle provide powerful models to study how long-term signals interact to affect long-term tissue and whole-animal function.

This chapter focuses on the use of transcriptomics and bioinformatics as tools to study the complex interactions that occur in liver, adipose, and mammary as dairy cows adapt to changes in physiological state and nutritional management during postnatal growth, lactation, or the transition from pregnancy to lactation. The goal is to provide specific examples of how these combined approaches could advance our understanding of tissue function beyond the classical biochemical pathways.

The Peripartal Dairy Cow as a Model for Whole-Animal Systems Biology

Achieving homeostasis during the transition from late-pregnancy to lactation represents a monumental task in modern high-producing dairy cows. In early postpartal dairy cows, a series of biological mechanisms bring about the prioritization for milk production at the cost of body reserves (Leroy et al., 2008). Drastically reduced insulin concentrations and alterations in adipose cellular response to this hormone cause a reduction in lipogenesis to extremely low rates. Adipose tissue of high-yielding dairy cows has an increased sensitivity to lipolytic stimuli (low insulin, high catecholamines, and high glucocorticoid concentrations; McNamara, 1989). As a result, lipolytic activity postpartum increases and lipid is mobilized among other things to supply nonesterified fatty acids (NEFA) to the udder.

These NEFA are primarily supplied to the liver where they are fully oxidized to CO_2, converted to ketone bodies, or esterified into triacylglycerol (TAG) either for delivery into blood as very-low density lipoproteins (VLDL) or for storage as cytosolic lipid droplets (Drackley, 1999). Because ruminants have inherently low rates of VLDL synthesis and secretion, accumulation of TAG in liver cells as well as extensive output of ketone bodies such as β-hydroxybutyrate into the circulation likely afflict all postpartal dairy cows. The metabolic load placed on the peripartal cow liver is exacerbated by the decrease in feed intake and the strong increase in milk production rate, causing a negative energy balance (NEB) that can occur as early as 10 days prior to parturition (Drackley, 1999). This NEB also seems to be exacerbated by feeding higher energy diets, i.e., allowing cows to overconsume energy, during the dry period (Dann et al., 2006; Bertoni et al., 2009).

The nature of the physiologic and metabolic adaptations during the transition from pregnancy to lactation is multifaceted and involves key tissues and cell types, further underscoring the need for a Systems Biology approach to identify control points, which go beyond the classical metabolic pathways. We have recently proposed (Loor, 2010) that in order to address the complex metabolic phenotypes of the peripartal period, there is a need to identify at the very least transcript variations in liver, mammary, and adipose that might underlie variation in metabolism and health. The approach could initially encompass a transcriptomic characterization of each tissue as

well as immune cells (e.g., neutrophils) over a wider range of nutritional treatments of practical relevance, and also across cows of different genetic merit (Loor, 2010). Within individual experiments, the data generated would allow for the identification of underlying gene networks and pathways that can be linked to a particular metabolic or health phenotype.

In the next sections we first introduce a bioinformatics approach that makes use of commercial software (Ingenuity Pathways Analysis®, IPA) as well as freely available online resources (DAVID, KEGG pathways) to mine transcriptomics data in order to assign biological relevance to gene transcription. Specific examples of the transcriptomics and bioinformatics approach are then presented using primarily data generated from our lab with dairy cattle. We do not attempt a full discussion of the findings, but want to highlight novel aspects of the system. These aspects have been uncovered recently and serve as an example how a Systems Biology approach may help to understand the underlying links between various tissues at various physiological states and in response to nutrition.

The Mammary Gland as a Unique System

The importance of the mammary gland is not only related to nutrition of the offspring, but milk and milk-related products are of high importance to human nutrition in many societies. Worldwide human consumption of bovine milk and milk-derived products has increased markedly in the last decades, especially in countries where this food was not part of the traditional diet (Wiley, 2007). The bovine mammary gland is an extraordinary organ able to produce more than 30,000 kg of milk in a complete lactation cycle (~305 d), with an estimated average of ~6% of body weight per day in high-producing Holstein cows. For example, the top-producing US Holstein in 1997 produced ca. 100 kg milk/d (Holstein Association USA ((http://www.holsteinusa.com/index.jsp), July 2, 2009).

The anabolic capacity of this organ in modern dairy cows, but also in other species, is so remarkable that it has been suggested, probably exaggerating, that from a metabolic perspective the cow can be considered an appendage of the mammary gland (Bauman et al., 2006). The milk yield potential, which particularly for dairy cows has been substantially improved through artificial selection programs, can be optimized through appropriate overall management with diet playing a pivotal role. Proper nutritional management can satisfy requirements for energy, protein, carbohydrate, lipid, minerals, and vitamins as well as prevent digestive and metabolic disorders. The quality of milk can also be manipulated by nutrition, particularly by specific components of the diet such as fatty acids.

In the past several decades, there have been extraordinary advances on the knowledge of the physiology of the lactating mammary gland (reviewed by Bauman et al., 2006). Despite those efforts, the physiological and cellular adaptations required for the synthesis and secretion of milk remain largely unknown. In our view, it is safe to conclude that most advances to date for improving bovine milk yield and composition through specific dietary components (e.g., fatty acids) have been made through an empirical and reductionist approach. The lack of availability and the costs of alternative techniques for use in livestock species have certainly limited the development of

a holistic and integral view, at the least at the molecular level, of the physio-cellular adaptations in mammary tissue. The recent development of microarray platforms for livestock species in combination with bioinformatics provides a valid alternative to overcome the previous limitations.

Overall findings from a recent Systems Biology analysis by Lemay et al. (2007) provided unique insights into the molecular basis that underlies the physiology of murine mammary gland during the course of lactation. However, their bioinformatics approach had several limitations, some highlighted by the authors themselves. In our view, one of the most important drawbacks was the lack of consideration of all the information that was available in the comparison of the 10 time points examined. This limited a more holistic view of the underlying biological adaptations.

Transcriptomics was used recently to study the adaptations of the bovine mammary gland between day 5 ($\pm$ 5) before parturition (day -5) and day 10 ($\pm$ 5) after parturition (Finucane et al., 2008). Besides the excessive variation in the sampling times (large variation in sampling days between animals), the study used bioinformatics approaches to a limited extent (Gene Ontology (GO) analysis only) without the application of appropriate statistics in the functional analysis. The main findings of the study were that twice as many genes were significantly downregulated than upregulated at day 10 postpartum, and that many of the upregulated genes were related to transport activity, lipid and carbohydrate metabolism, and cell signaling pathways. Downregulated genes were related to cell cycle and cell proliferation, DNA replication and chromosome organization, microtubule-based processes, and protein and RNA degradation.

The overall data generated by the above study in bovine mammary underscored the tremendous anabolic capacity of the mammary gland, particularly for the synthesis of lipid, which also is a characteristic in mouse mammary (Rudolph et al., 2007). Those findings confirmed what has been known for a long time about ruminant mammary gland. A novel piece of information from that study was the decrease in chromosome reorganization after parturition, which also has been found in mouse mammary (Rudolph et al., 2007) and suggests that the mammary gland reduces chromosome reorganization in order to allow for a "consistent" transcriptome. This finding points to epigenetics as an important event during the control of milk synthesis.

Systems Analysis of the Cow Mammary Transcriptome

We have built a Systems Biology pipeline using a large transcriptomics experiment in mammary tissue considering 9 time points beginning a month prior to parturition through 300 days into the subsequent lactation. Functional and gene network analyses of the more than 6000 differentially expressed genes (DEG) from microarray data, filtered by quality and statistical significance (false discovery rate or FDR ≤ 0.001), were mined using several bioinformatics tools (DAVID at http://david.abcc.ncifcrf.gov/home.jsp; Huang da et al., 2009b, 2009c; IPA at http://ingenuity.com/, Kyoto Encyclopedia of Genes and Genomes (KEGG) at http://www.genome.jp/kegg/) implemented by a novel approach developed in our lab and termed dynamic impact approach (DIA). All time points after parturition were compared with the month prior to parturition ($+1$ versus -30, $+15$ versus -30, and so on) as well as with each

preceding time point (+15 versus +1, +30 versus +15, and so on). Those comparisons allowed for a complete visualization of the mammary gland's adaptations to the onset of lactation as well as each discrete lactating stage relative to the late-pregnant state, all of which exhibit distinct features (Grossman and Koops, 2003).

The data generated by our microarray experiment uncovered that bovine mammary experiences a large degree of transcriptomics adaptations (i.e., large number of DEG with overall time effect at FDR = 0.001 and p-value $\leq$0.001 between comparisons) both at the onset of copious milk secretion (1 day) and when milk yield declines (at 240 days postpartum) suggesting a large degree of transcriptional control of milk secretion. In addition, from day 30 prior to parturition through 60 days postpartum we observed a consistent increase in the number of DEG, reaching more than 3000 at 30 days postpartum. These results reveal a different pattern than observed in mouse mammary (Lemay et al., 2007), or in a previous small-scale bovine microarray study (Finucane et al., 2008).

Limitations and Potential Solutions for the Proposed Analytical Approach

A preliminary analysis of our mammary microarray data was conducted using the gene-enrichment approach (also known as overrepresented approach or ORA) (Huang da et al., 2009a). This statistical approach is the most widely used in high-throughput biological function analysis. It is based on the assumption that in order for pathways, functions, or terms in a given gene list to be biologically relevant, they have to be overrepresented, i.e., must be present in the gene list at a greater proportion compared to the background (that is the proportion of the genes in that pathway, function, or term on the microarray platform used for transcriptomics). In other words, the genes associated with a specific pathway, function, or term have to be present in the list of DEG in a significantly greater proportion compared to the proportion of genes of that pathway, function, or term in the microarray that are picked randomly.

The preliminary analysis clearly indicated that the enrichment analysis was not adequate to address the objective of our experiment. In fact, an orthodox application of the statistics, i.e., the use of Benjamini–Hochberg FDR adjustment to account for multiple comparisons (Benjamini and Hochberg, 1995), to the functional analysis provided almost no overrepresented functions or pathways when data were mined in DAVID or IPA. On the one hand, those results suggested that there were transcriptomics adaptations encompassing most of the known functions of the mammary gland during lactation. However, more importantly, those results highlighted several limitations of the ORA approach, which have been underscored previously (Draghici et al., 2007; Huang da et al., 2009a). A serious limitation in our longitudinal study of the mammary transcriptome was the well-known impossibility of comparing the functional analysis between gene lists (Huang da et al., 2009a). This precluded a full investigation of the dynamism of the functional adaptations during a time-course experiment. Another important limitation was that the larger the gene list, the lower the likelihood of finding significantly enriched functions/pathways and *vice versa* (Huang da et al., 2009a). For those reasons we have undertaken a different approach, which is not based on ORA but on the calculation of the impact of DEG on functions and pathways; it is referred as DIA.

The foundation of the DIA approach lies on the fact that pathways (or functions) are proportionally affected or impacted by several factors: the proportion of DEG (i.e., their encoded proteins) among all genes in the microarray that belong to a pathway, function, or term, the average significance (*p*-value) of the change in expression of those DEG, and the average fold change of the DEG composing the pathway, function, or term.

The DIA approach to investigate affected pathways is based on the combination between percentage of up- or downregulated DEG compared with all genes on the microarray that belong to the pathway/function, the mean –log *p*-value of the up- and downregulated DEG, and the mean of log (base 2) fold change of the DEG. An overall impact of DEG on the pathway/function is calculated for each comparison. The approach has several caveats, such as the fact that no specific gene/s is/are considered "rate-limiting" (which is not in itself a limitation considering that many, if not most, of the genes on the pathway need to be affected to have an impact on the metabolic flux or function (Morandini, 2009)). Also, up- or downregulation does not always coincide with an increase or decrease in metabolic flux or function, because proteins with regulatory roles often exist within pathways or functions.

The DIA also estimates the potential direction of the flux (increased or decreased). The DIA was implemented in Microsoft Excel and used in combination with DAVID and KEGG. The mammary dataset with significant DEG and relative Entrez Gene IDs was uploaded in DAVID and results without any thresholds were downloaded. For KEGG pathways analysis, the whole KEGG bovine pathway database was downloaded from KEGG Web site as a platform for analysis of impact and direction of flux within the DEG. For the KEGG analysis, the DIA also clusters the pathways at a higher level of organization and calculates the average impact and flux to provide a readable overview of the pathways (Figure 9.1).

The Proposed Analytical Approach Assessed

The validity of the DIA, which was explained in section "Limitations and potential solutions for the proposed analytical approach," was confirmed by the high association between the calculated impact of pathways and functions with functional measurements (milk yield, fat, protein, and lactose content in milk) and known biology of the lactating mammary gland. For example, the approach uncovered BTA6, known to be the chromosome with the highest association to milk yield, milk protein, and fat percentage (Sheehy et al., 2009), as the most impacted chromosome; the DIA revealed milk, mammary epithelial, mammary gland, and lactating mammary gland among the most impacted "tissues"; and lactose synthesis was the most impacted biological process. All those terms, except for milk at 60 versus −30 days in milk (DIM), were not significantly enriched (Benjamini–Hochberg FDR ≤ 0.05) using the ORA approach both in IPA and DAVID.

Application of DIA to the KEGG pathways using our mammary time-course microarray data suggested that metabolic pathways were the most impacted and induced during lactation, with lipid and carbohydrate metabolism being the most important (Figure 9.1). Among carbohydrate metabolism, galactose metabolism was the most impacted and induced followed by the TCA cycle, whereas ascorbate and aldarate

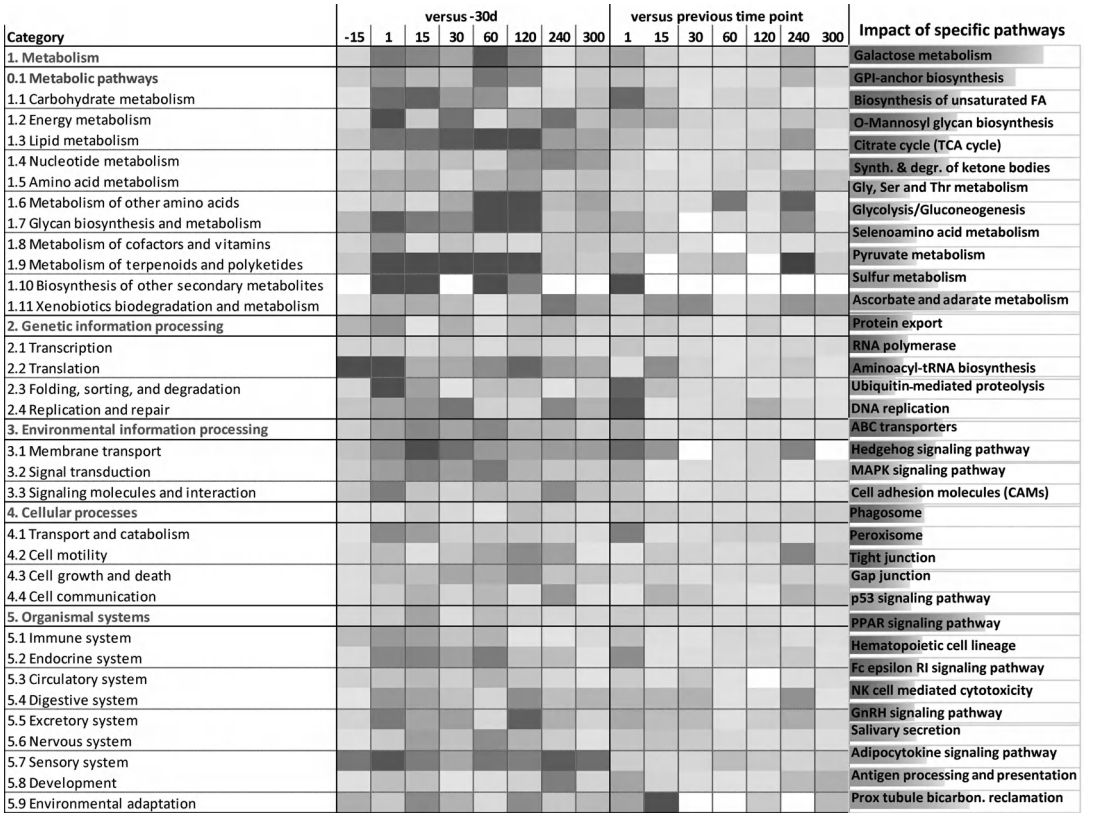


Figure 9.1 Overall calculated dynamic impact (DIA) among KEGG pathways using differentially expressed genes (DEG) in bovine mammary tissue during the lactation cycle. Shown are the main pathway classification category (e.g., metabolism, genetic information processing; left column) and corresponding subcategories (e.g., carbohydrate metabolism in metabolism category, transcription in genetic information processing). The heat map denotes estimated increase (red shade) or decrease (green shade) of metabolic flux or signaling through the pathway at each time point (−15, 1, 15, 30, 60, 120, 240, and 300 days from parturition) relative to −30 days or relative to a consecutive time points. The estimates are calculated as ((% of DEG/genes for the particular pathway in the microarray) × average $\log_2$ fold change of DEG × average –log p-value of DEG). The most impacted (denoted by the length of the bar) pathways with an estimated increase (red bar) or decrease (green bar) in flux in each general category during lactation are shown in the right column. (*See insert for color representation of this figure.*)

metabolism, pyruvate metabolism, and glycolysis/gluconeogenesis were among the most inhibited during lactation. Among lipid metabolism, biosynthesis of unsaturated fatty acids and synthesis and degradation of ketone bodies were among the most induced (Figure 9.1). Overall, the functional analysis using DIA suggested that from a metabolic standpoint the mammary gland increases lipid and carbohydrates metabolism, i.e., increases synthesis of carbohydrates and lipid molecules and decreases utilization of glucose for energy production with an apparent increase in the utilization of ketone bodies for energy production. Concurrently, it was revealed that there is a considerable inhibition of amino acid metabolism, protein synthesis, and RNA metabolism during the course of lactation (Figure 9.1). The cell cycle together with DNA metabolism, cell assembly and organization, and gene expression appeared greatly inhibited during the whole lactation, whereas signaling and interactions between cells were generally enhanced. Cell signaling together with development of the immune system appeared to be substantially impacted and induced, whereas immune cell trafficking was inhibited. One of the most impacted and inhibited subfunctions during the whole lactation was the major histocompatibility complex, particularly the class I.

Interestingly, most of the pathways induced during lactation were inhibited between 120 and 240 days postpartum, which coincided with a substantial drop in milk yield. In addition, among signaling pathways the one involving peroxisome proliferator activated receptor (PPAR) appeared to be the most induced during lactation. On the basis of those results, we have tested the role of PPAR, and particularly PPARγ among subtypes of PPAR, in regulating expression of genes involved in milk fat synthesis. Data from such investigation provided evidence of a likely role for this nuclear receptor (Kadegowda et al., 2009).

Overall, the use of DIA to investigate the temporal transcriptomics of the mammary tissue suggested the following. (1) From a functional point of view, that the mammary gland of the modern high-producing dairy cow concentrates all its efforts in producing milk. Hence, this effort leads to a shutdown of nonessential or milk synthesis- and -secretion-competitive functions. The data allowed us to infer a mechanism where the genes coding for milk-related proteins (milk components and protein essential for milk synthesis) have a substantial increase in expression, thus, greatly enhancing the competitive advantage in the use of the protein synthesis machinery whose availability decreases during lactation. This appears to be a feature in common between bovine and mouse mammary (Lemay et al., 2007). (2) The organization of the tissue for milk production has to be set before the onset of lactation. During lactation, the tissue attempts to maintain a stable organization and transcriptional program. This also appears to be similar to mouse mammary (Lemay et al., 2007). (3) The mammary gland uses a substantial amount of resources to enhance its immune system capacity, but it prevents an overly sensitive activation of the immune response (i.e., inhibition of major histocompatibility complex (MHC)).

Transcriptional Networks in Systems Analysis

The analysis of pathways, functions, and terms is only a part of a System Biology approach. In fact, the viability of cells as well tissues, organs, or the entire

organism is dependent upon interactive networks among molecular components. Therefore, central to the definition of System Biology is the analysis of interactive networks.

Among global interactive networks those formed by transcription factors (TF) and their putative target genes (i.e., TF networks) are an important element of a Systems Biology approach. This analysis allows transcriptional networks to be uncovered that may play important regulatory functions and, being TF able to regulate expression of many genes, provides avenues for potential manipulation (e.g., via nutrients). The TF networks are generated by investigating the potential impact of one or more TF on the expression of DEG. In case of our longitudinal study of bovine mammary tissue we relied on the IPA Knowledge Base. The analysis allowed the identification of more than 100 TF, which may play a pivotal role in orchestrating the adaptations of the mammary gland to lactation. A large TF regulatory network (i.e., a larger number of known relationships of TF with DEG) was uncovered for v-myc myelocytomatosis viral oncogene (*MYC*) and tumor protein p53 (*TP53*), both of which were downregulated during lactation, and for lysine (K)-specific demethylase 5B (*KDM5B*) and YY1 transcription factor (*YY1*), both of which were upregulated.

Several TF had a tremendous increase in expression through lactation (e.g., E74-like factor 2 and ankyrin repeat and SOCS box-containing 11), but were not able to produce known transcriptional networks with other DEG. The lack of underlying networks is consequence of the limited availability of information about those TF. In-depth molecular studies will probably reveal downstream genes and overall functional impacts of the TF uncovered. It will be important to conduct additional molecular studies to assign a functional meaning to those TF networks.

The TF network analysis within a System Biology approach revealed previously unrecognized networks of genes with major TF at play. This is of interest because manipulation of those TF networks, both directly (e.g., inhibiting or activating one or several TF in the network) or indirectly (e.g., through selection or genetic engineering), might lead to improvements in milk synthesis capacity, efficiency of nutrient utilization (e.g., optimal dietary N and P), or regulation of mammary metabolism in order to prevent/treat metabolic disorders.

Clustering Approaches as a Tool in Systems Biology

We have conducted clustering analysis to further evaluate the coexpression of functions/pathways and key TF that control the expression of DEG during lactation in bovine mammary tissue. Using Genesis (Sturn et al., 2002) we conducted k-means clustering analysis that partitions the observations (in our case, the DEG) into k clusters (k denotes a number of clusters decided *a prior*) in which each observation belongs to a cluster with the nearest mean. This is the most appropriate approach for clustering time-course data. The k-mean clustering analysis identified 16 clusters that best describe the temporal pattern of gene expression among the more than 6000 DEG affected by time at an FDR = 0.001. This number of clusters was deemed most appropriate using the minimum gain of power of the figure of merit as criterion. The figure of merit is a system used to characterize the performance of a device, system, or method, relative to its alternatives; the analysis in the clustering-decision process

reaches the best number of clusters among alternatives when the gain of power of prediction of an additional cluster reaches a minimum (Yeung et al., 2001)

We used DAVID and IPA to allow uncovering functions/pathways significantly enriched in each cluster. The IPA was used to investigate the functional networks among genes and those potentially key TF-controlling expression of the cluster of genes. In addition to providing an in-depth analysis of potential TF controlling the cluster, we examined overrepresented DNA binding motifs for TF in promoter or coding regions of genes within the clusters. For this we used previously developed bioinformatics tools (a tool developed by Tabach et al., 2007 and cREMaG from Piechota et al., 2010).

Application of the above-mentioned approaches uncovered several functional clusters. Five clusters grouped genes with evident upregulation during lactation. Those clusters were significantly enriched with milk-related products and functions, lipid biosynthetic processes, tight junctions, endomembrane system and transport (particularly endoplasmic reticulum and Golgi membranes), lactation, mammary gland development, signal, extracellular region, and secretion. Four clusters grouped genes downregulated during lactation. Those clusters contain significantly enriched functions/components such as those involved in protein synthesis, immune response, chromatin remodeling, and cell cycle.

The analysis also uncovered several clusters of genes that were enriched in particular chromosomes. For example, Chr 6 was enriched by the milk protein gene cluster (i.e., the cluster composed of genes with the greatest upregulation during lactation), Chr 19 and 23 were enriched by the cluster of genes significantly enriched with cell cycle and chromosome modification. Interestingly, the cluster of genes most upregulated during lactation encompassed, besides genes coding for milk protein, those coding the major fatty acid binding protein (FABP), isoform (*FABP3*), and DnaJ [Hsp40] homolog subfamily C member 12. The latter is involved in protein assembly and export.

The network analysis uncovered a large degree of interactions between gene products of the same upregulated clusters. The unsupervised TF analysis using IPA highlighted that many TF are potentially involved in controlling expression of genes within the clusters. Among TF, the hepatocyte nuclear factor 4 alpha (*HNF4A*) appeared to be common among nearly all the upregulated clusters, which may indicate that this TF is a plausible candidate for future molecular studies.

The search for significantly enriched motifs in the clusters of genes using the method developed by Tabach et al. (2007) uncovered AREB6.03 (Atp1a1 regulatory element binding factor 6) as being the most significantly enriched in one of the clusters of the most upregulated genes during lactation. The biological significance of this finding for the mammary gland is unknown. Progress in understanding the function of AREB6 has been made (e.g., Ikeda et al., 1998), but specific molecular studies in bovine will be required to understand the function of this TF in the context of the mammary gland biology. The reader also has to consider that our TF motif analysis was heavily limited by the low numbers of DEG with a suitable promoter region (ca. 2000 out of more than 6000 DEG). Future progress in bovine genome sequencing and annotation will hopefully bridge the gap and deliver a more complete comprehension of these data.

The analysis of significantly enriched motifs using cREMaG provided further information about the genes found in the k-mean clusters. The motifs for LIM homeobox 3,

myocyte enhancer factor 2A, and glucocorticoid receptor were significantly enriched within the cluster of genes with the largest upregulation during lactation; whereas, the motif for nuclear factor NF-kappa-B p65 subunit was significantly enriched in genes within the cluster with the second most upregulated pattern during lactation. The interferon regulatory factor 2 motif was uniquely enriched within a cluster of down-regulated genes during lactation. The significance of these findings, with exception of glucocorticoid receptor (Doppler et al., 1989), is not apparent; however, this opens up possibilities for future investigations.

Applying the Systems Biology approach to microarray data of mammary tissue during pregnancy and lactation allowed us to confirm several previously established responses but also provided new information on other salient pathways and functions underlying the mammary gland's adaptations to copious milk synthesis. However, the most novel outcome of this approach was the prospect of viewing the mammary gland as a complex interactive system where activation or inactivation of pathways and induction or inhibition of functions is occurring simultaneously; more importantly, the gene products encompassing these pathways/functions specifically interact in order to properly carry out their assigned function(s).

As stated in the introduction to this chapter, application of the Systems Biology approach to livestock is still in its infancy and several limitations still exist, both of a technical and mechanistic nature (i.e., knowledge about specific molecules and their interactions). The mammary gland has, for the most part, been studied in isolation disregarding the fact that (as an organ) it is composed of additional cells beside epithelial. An orthodox Systems Biology approach will require consideration of the interactions of one tissue with others composing the organism as well as the intimate relationship between neighboring cells. In section "Intertissue cross talk during prepubertal bovine mammary development," we provide an example of a Systems Biology approach to uncover the interactive network in the main compartments of the mammary gland during preweaning development.

Intertissue Cross Talk During Prepubertal Bovine Mammary Development

The development of the mammary gland in preweaned dairy calves is crucial for future milk production. It is characterized by an important cross talk between the two major developing tissues: the parenchyma (PAR) and the mammary fat pad (MFP). In a recent study (Piantoni et al., 2010), we investigated the effects of milk replacers (fed for ~60 days), with different levels of fat and protein, or different rates of intake, on transcriptomics in PAR and MFP of preweaned calves. The two tissues were carefully dissected and RNA extracted for microarray analysis. Despite the large morphological changes due to type of milk replacer (Daniels et al., 2009a, 2009b), analysis of tissue collected at slaughter indicated that diet had only a modest effect on the transcriptome (Piantoni et al., 2007). Besides the evaluation of treatment effects, the availability of dissected samples and tools for high-throughput analysis allowed for the first time a direct transcriptomics comparison between PAR and MFP. Furthermore, it provided the means of using a Systems Biology approach to investigate potential cross talk between the developing PAR and MFP.

The transcriptomics differences between the two tissues were remarkable, with more than 9000 DEG (FDR-corrected p-value ≤ 0.05). A cutoff of 1.5-fold differences in gene expression between tissues was applied in order to identify more pronounced DEG. This resulted in more than 1400 DEG. A functional analysis using IPA and DAVID was conducted to identify the most enriched functions in genes highly expressed in one tissue versus the other tissue. Not surprisingly, the top significantly enriched functions among DEG that are more expressed in PAR compared to MFP were related to proliferation and morphogenesis of epithelial tissue, whereas, the top significantly enriched functions among DEG that are more expressed in MFP versus PAR were related to adipose tissue (Piantoni et al., 2010).

To evaluate the TF with main roles in controlling highly expressed DEG in one tissue versus the other, we conducted a transcriptional network analysis. The TF in each list were extrapolated and all known effects on transcription of DEG were analyzed. This allowed uncovering several TF (e.g., peroxisome proliferator-activated receptor γ (*PPARG*) and *HNF4A* among DEG that were more expressed in MFP versus PAR, and *MYC* and *TP53* in DEG that were more expressed in PAR versus MFP) and relative networks as chief hubs in controlling the specific transcriptomics differences between PAR and MFP. More importantly, the data revealed that these TF are not only central during lactation (see section "transcriptional networks in Systems Biology") but also during prepubertal mammary development. With the aim to investigate the potential cross talk between the two tissues, we identified all possible DEG in PAR versus MFP that encode for secreted signaling molecules (i.e., cytokines and growth factors), and we analyzed possible interactions with DEG that were more expressed in MFP versus PAR that encode for receptors, and *vice versa*. The analyses suggested a large degree of interaction between the two tissues and allowed us to envisage a reciprocal influence during development.

Several cytokines and growth factors, such as interleukin 1 beta (*IL1B*), osteopontin (*SPP1*), chemokine (C-X-C motif) ligand 10 (*CXCL10*), platelet-derived growth factor alpha polypeptide (*PDGFA*), dickkopf homolog 1 (*DFF1*), and neuregulin 1 (*NRG1*) were among the signaling molecules with likely higher production (i.e., higher gene expression) and secretion in the PAR compared to MFP. The network analysis indicated that those signaling molecules upon reaching the MFP could have increased lipid accumulation (e.g., through increased expression of *PPARG* and several lipogenic enzymes) and have reduced its proliferation rate (Piantoni et al., 2010).

Cytokines and growth factors released preferentially by MFP such as adiponectin (*ADIPOQ*), leptin (*LEP*), fibroblast growth factor 2 (*FGF2*), interleukin 13 (*IL13*), interleukin 7 (*IL7*), and jagged 1 (*JAG1*) could have a potential role in the organization and proliferation of PAR. A model of cross talk between PAR and MFP on the basis of the overall functional analysis was proposed. This model provides the targets for future hypotheses-driven studies.

The discovery of networks of genes with seemingly important roles in the reciprocal influence between the two adjacent tissues in the developing mammary gland was an important achievement. We are now poised to design studies to uncover, among other things, factors (including nutrients) that may influence the dynamics of the TF networks. In the future, the discovery of agonists or antagonists of these interactive networks would be powerful tools in the hands of dairy farmers that may desire to enhance mammary gland development.

Effect of Prepartal Nutrient and Energy Intake on Dairy Cow Adipose Tissue

The biological active role of liver and mammary in coordination of animal physiology, particularly during the transition from pregnancy into lactation in dairy cows is well known. Until a decade ago, adipose tissue was considered as a mere passive energy storage organ in the body, with some additional corollary functions such as providing cushion and thermoregulation. Due to the large and rapid rise of obesity-related diseases in Western countries, research on adipose tissue biology has exploded in recent years (Hausman et al., 2009). It is now evident that adipose tissue is far from being a passive tissue, but that it has an active role in regulating whole-body metabolism, appetite, and the immune system. Adipose tissue appears to have an important degree of cross talk with other tissues through the release of endocrine molecules (Hausman et al., 2009; Lee et al., 2009). The tissue appears to be very sensitive to energy status of the organism (Lee et al., 2009).

Several studies carried out in model organisms or humans have shown that energy status has a quick and strong impact on adipose tissue transcriptomics (Higami et al., 2006; Swindell, 2008; Palou et al., 2010). This impact involves the regulation of lipid metabolism in particular, but also immune-related functions. Nutrigenomic studies in animal models using high-throughput technologies have mainly focused on short- and long-term transcriptomics and metabolomics in adipose tissue as it pertains to high-fat feeding (Kim and Park, 2008; Shearer et al., 2008; Jobgen et al., 2009), calcium, dairy products (Bruckbauer et al., 2009), leucine supplementation (Jobgen et al., 2009), coffee (Fukushima et al., 2009), soy proteins (Takamatsu et al., 2004), cocoa (Matsui et al., 2005), and specific long-chain fatty acids such as conjugated linoleic acid (LaRosa et al., 2006).

In livestock species, adipose has been studied predominantly in meat animals such as pig, chickens, and beef with a strong emphasis on subcutaneous and intramuscular fat (Hausman et al., 2009). Relatively few nutrigenomics studies have been carried out in livestock using high-throughput technologies and most have made limited use of bioinformatics. Therefore, the outcomes of those efforts have been primarily lists of genes or proteins for future investigations (e.g., Wang et al., 2009) with few systematic insights into the tissue's adaptation to nutrition.

We recently studied the transcriptomics adaptations of bovine subcutaneous adipose tissue from the beginning of pregnancy through early lactation in cows fed diets designed to meet (~100% of net energy requirements; 1.21 Mcal/kg diet dry matter) or exceed (~150%; 1.63 Mcal/kg diet dry matter, i.e., energy-overfed) energy requirements during the entire dry period (~65 days; Janovick and Drackley, 2010). The higher energy diet led to greater accumulation of body fat, as measured by body condition score (Janovick and Drackley, 2010), and robust transcriptional adaptations with more than 3000 DEG affected (FDR = 0.05) by the interaction of time × diet (Janovick et al., 2009). In response to the higher energy diet prepartum and using a cutoff of $P \leq 0.01$ for the comparison between diets at each time point plus the FDR-corrected p-value ≤ 0.05 for the interaction effect, we uncovered more than 1500 DEG at 2 weeks prepartum, less than 100 at 1 day after calving, and ~200 DEG at 2 weeks postpartum compared to control (i.e., diet to meet 100% of energy requirements). These longitudinal adaptations suggested that the transcriptome responded

quickly to prepartal energy overfeeding, but there was little carryover effect after parturition.

Using a combination of ORA tools and the DIA described above for the mammary gland, it was evident that the higher energy diet had a strong impact on metabolism and other cellular functions compared to the requirement diet (Figure 9.2 shows results from DIA and ORA for the diet comparison at −14 DIM). In cows overfed energy compared to controls, the DIA analysis of KEGG pathways and the enrichment analysis of pathways in DAVID and IPA indicated a large activation of energy metabolism and lipid synthesis, including *de novo* fatty acid synthesis. The analysis also indicated a pivotal role of PPAR signaling and a larger induction of protein synthesis in the adipose tissue of overfed cows. Interestingly, almost all the KEGG pathways were induced in adipose tissue of energy-overfed cows at two weeks prior to parturition, but those pathways were strongly inhibited in the same group at parturition (i.e., +1 versus −14 DIM), suggesting that they are tightly controlled by homeorhetic adaptations required for energy repartitioning at the onset of lactation (Drackley, 1999).

Network analysis uncovered large interactions among DEG in the comparison of the higher energy to the requirement diet at −14 days from parturition (Figure 9.2). Among TF in the network CCAAT/enhancer binding protein alpha and beta (*CEBPA* and *CEBPB*), both upregulated by overfeeding energy at −14 days, produced the transcriptional networks with the largest number of DEG suggesting that these TF are central in orchestrating prepartal adipose transcriptional adaptations to high-energy diet. On the basis of the high-impact of PPAR pathways, we evaluated the relevance of PPARγ in controlling expression of DEG in response to overfeeding energy at −14 days. The PPARγ appeared to be central as well (Figure 9.2) with a large control among most of the lipogenic genes that were affected (see Loor, 2010, for a description of them). In addition, among the functions that were significantly enriched (FDR = 0.15) we found growth (inhibited) and activity (induced) of neurons, inhibition of growth of the immune system, and induction of protein synthesis. From previous experiments with pregnant, nonlactating dairy cows, it is well established that blood insulin increases dramatically and chronically soon after animals are fed higher energy diets, either during the entire dry period or during the last 3 weeks prepartum (Dann et al., 2006; Loor et al., 2006). Through network analysis we evaluated the potential effect of elevated blood insulin on gene networks. Just as it would be expected in nonruminants this hormone appears to control most of the networks, and it is affected significantly by overfeeding metabolizable energy (Figure 9.2).

The Systems Biology approach allowed us to uncover that the adipose transcriptome responds rapidly to overfeeding dietary energy through marked induction of genes involved in lipid accumulation and protein synthesis. Central to these adaptations were PPAR signaling together with the CEBP family of TF, which are part of a large interactive network underlying these coordinated adaptations. From a physiological standpoint, the overall molecular adaptations observed might have been driven by the large insulin surge caused by prepartal energy overfeeding.

The adipose tissue plays a pivotal role during transition from pregnancy to lactation due to the release of NEFA. As previously shown (Dann et al., 2006), overfeeding dietary energy prepartum increases NEFA concentration in the immediate postpartal period, often with negative effects on cow health (Drackley, 1999; Bertoni et al., 2009).

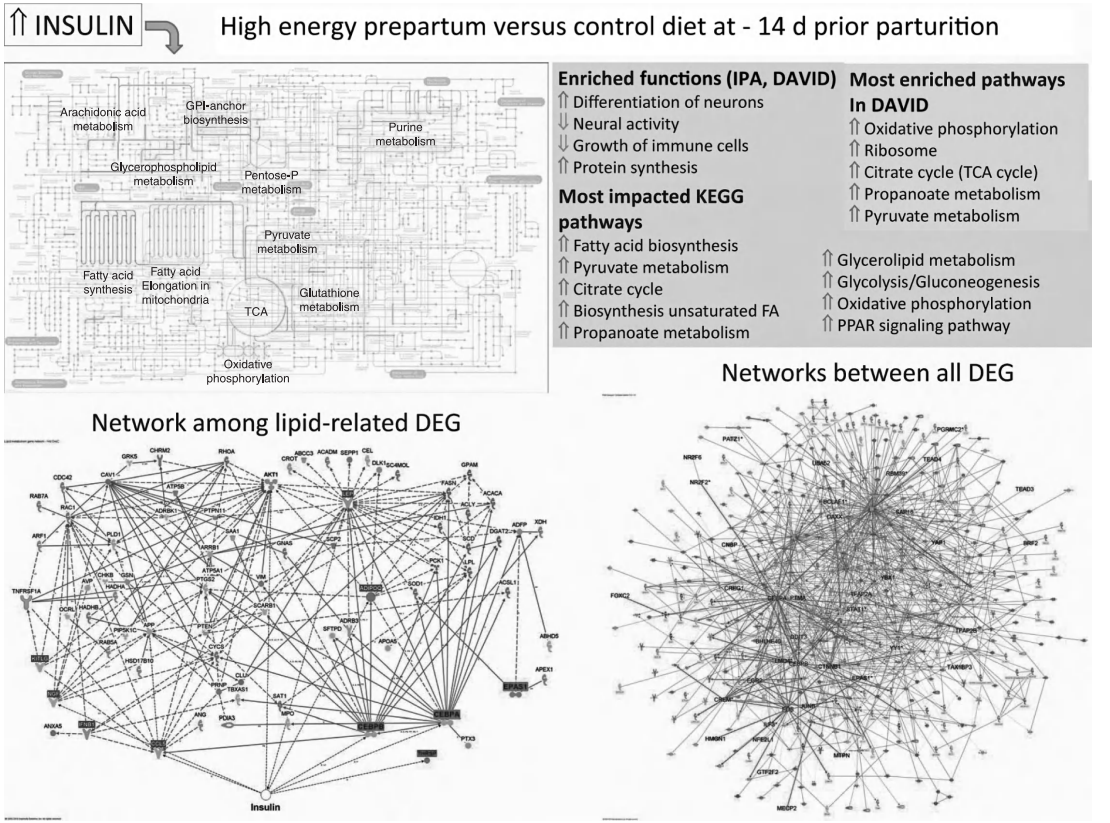


Figure 9.2 Summary of the effects of overfeeding energy during the dry period on bovine subcutaneous adipose tissue metabolic flux (KEGG pathways), main biological functions (DAVID and Ingenuity Pathways Analysis® (IPA)), and network analysis to uncover putative links between blood insulin and DEG in energy-overfed cows versus controls. Arrows denote upregulation (↑) or downregulation (↓) in energy overfed cows versus controls. In the KEGG pathway orange-to-red denotes an increase in flux, while green denotes a decrease in flux through the specific pathway. Among lipid-related DEG highlighted in red in the insulin gene network are transcription factors (TF), and in blue are cytokines and growth factors. Within this network and the network of all DEG, red denotes upregulation and green downregulation of expression of the particular gene in cows overfed energy versus controls. In addition, in the network encompassing all DEG the TF are in large bold font. Microarray data are from the study of Janovick et al. (2009). (*See insert for color representation of this figure.*)

Excessive concentrations of NEFA in blood can be deleterious to liver (hence, the cow) because of this organ's involvement in metabolizing the NEFA. Besides metabolizing NEFA, liver has additional crucial roles during the transition from pregnancy to lactation, e.g., it is the major site of gluconeogenesis and actively participates in the immune response (Drackley et al., 2006; Bertoni et al., 2009). Thus, the importance of studying the liver's adaptations to prepartal dietary energy is evident. As in mammary tissue, the combination of microarray technology with bioinformatics is a unique mean for applying a Systems Biology approach to liver.

Effect of Prepartal Nutrient and Energy Intake on Dairy Cow Liver

In nonruminants, the differentiation of hepatocytes and the function of the adult liver are controlled through the coordinated expression of a large number of genes (Columbano and Ledda-Columbano, 2003). Environmental (including nutrients), autocrine, endocrine, or paracrine signals all contribute to changes in hepatic gene expression (Columbano and Ledda-Columbano, 2003). Therefore, transcriptomics analyzed thorough bioinformatics tools are ideal to help identify regulatory mechanisms in the bovine liver that are sensitive to nutrient balance during the transition from pregnancy into lactation.

With current practice of feeding dairy cows, the energy density of prepartal diets is increased during the last 2–3 weeks before parturition in the hope to maximize feed intake before parturition and to provide an adjustment period for the rumen to adapt to the higher energy diets typically fed at the onset of lactation (Grummer et al., 2004). However, our data suggest that cows that are moderately overfed (more than 140% of net energy requirements) during this phase of the dry period, even without becoming obese or obtaining a too high body condition score, may be placed at greater risk for peripartal health problems (Dann et al., 2006). A consistent finding in our studies and others (Rabelo et al., 2003; Grummer et al., 2004) is that cows allowed *ad libitum* access to higher than required energy diets (net energy of lactation greater than 1.50 Mcal/kg dry matter) at the end of the dry period have a stronger decrease in feed intake before parturition and a lower increase in feed intake postpartum. Therefore, a more controlled energy content of the diet (i.e., diets that meet but do not greatly exceed energy requirements prepartum) may be beneficial by promoting a more consistent feed intake around parturition (Drackley et al., 2006). The amount of energy intake prepartum is known to induce changes in blood and liver tissue metabolic indicators after parturition (Drackley, 1999; Dann et al., 2006).

In two studies, we used a cDNA microarray consisting of 7872 annotated cattle genes to evaluate liver transcriptional adaptations to the prepartal level of energy intake and the change in physiological state (Loor et al., 2005, 2006). Diets were fed *ad libitum* to exceed net energy requirements (~140% of requirements), were restricted to provide less than estimated requirements (~80%), or were fed to meet (~100%) the calculated energy requirements during the dry period. When the individual datasets (Loor et al., 2005; 2006) were combined and reanalyzed statistically (Bionaz et al., 2007b), we found more than 4790 DEG (FDR $\leq$ 0.05) due to the interaction of treatment $\times$ time. An initial approach used for mining the data consisted of clustering analysis using the same criteria as for the mammary time-course experiment

(reported above). This analysis resulted in 13 clusters of genes with correlated expression profiles (Figure 9.3). In order to uncover coordinated molecular adaptations to prepartum dietary energy level in liver, we performed for each cluster an enrichment analysis of GO biological process, GO cellular component, and KEGG pathways using DAVID. In addition, we also have analyzed functions and most enriched pathways using Ingenuity Pathway Analysis.

The cluster analysis allowed uncovering many genes with a coordinated adaptation to prepartal level of dietary energy. Interestingly, among the 13 clusters generated by our data, 5 clusters (ca. 25% of DEG) had a similar pattern in both restricted and energy-overfed cows versus controls (clusters 2, 5, 8, and 13, and to some degree also cluster 12; Figure 9.3). The other 8 clusters revealed several differences between groups (Figure 9.3). For example, cluster 1, cluster 9, cluster 10 and, to some extent, also cluster 12, contained genes with greater temporal upregulation in liver of energy-restricted cows than energy-overfed or control cows; whereas, cluster 4 and 6 contained genes with greater downregulation in energy-restricted cows than the other groups (Figure 9.3).

The functional analysis of the clusters identified few highly significant (FDR < 0.05) enriched terms (biological process, cellular component, functions, or pathway) in the clusters (Figure 9.3). However, most of the clusters presented overrepresented functions. The genes in cluster 2, characterized by a strong downregulation by both energy-overfed and restricted versus controls, were overrepresented with terms related to induction of inflammation, suggesting that those dietary management approaches similarly down-regulate genes involved in inflammatory response, in particular the complement pathway (Figure 9.3). The IPA analysis confirmed the enrichment of inflammatory-related functions, in particular acute phase response, with high enrichment of several signaling pathways involving nuclear receptors that control cholesterol synthesis (e.g., farnesoid X-activated receptor and liver X receptor beta). IPA also identified among the top enriched functions in this cluster tissue morphology (mainly involved in increasing number of cells), lipid metabolism (mainly increasing oxidation of fatty acids), and cell-to-cell interaction (general activation of cells).

Those findings and the coordinated downregulation of inflammatory response-associated genes by either overfeeding or restricting dietary energy prepartum were novel. The complement system components are synthesized (ca. 90%) by liver and participate in the activation of the immune system (Qin and Gao, 2006). The biological consequence of down-regulation of the complement system would be a reduction in inflammatory-like responses after parturition compared to control cows. The inflammatory-like conditions typical in peripartal cows have detrimental influence on performance (Bionaz et al., 2007a; Bertoni et al., 2009); thus, the reduction of the complement system should have prevented or reduced the postpartal inflammatory-like response.

The synthesis of complement system components appears to behave as positive acute-phase proteins (Qin and Gao, 2006); thus, a decrease in expression of these molecules is indicative of a decreased acute-phase reaction in both energy-overfed and energy-restricted cows after parturition. However, the evaluation of temporal expression of genes coding for acute-phase proteins (e.g., serum amyloid A 1 and ceruloplasmin) indicated a stronger acute-phase reaction postpartum in energy-restricted cows compared to the other groups. In addition, a strong down-regulation of the

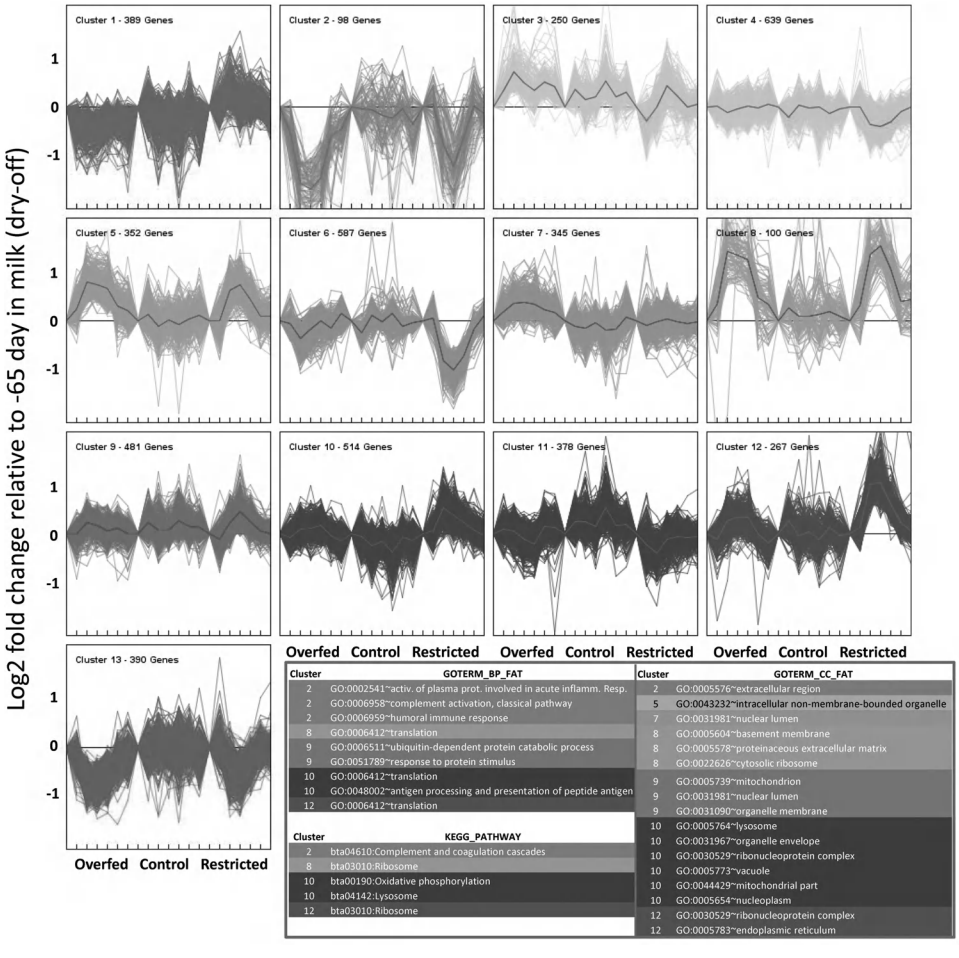


Figure 9.3 *k*-means clustering analysis using Genesis software (Sturn et al., 2002) of ~4790 DEG (false discovery rate ≤ 0.05) due to treatment × time in liver of cows underfed energy (ca. 80% of energy requirements), overfed energy (ca. 150% of energy requirements), or fed energy to meet requirements (control) prepartum (from dry-off to parturition). The X-axis depicts the dietary treatments imposed by time point (−30, −14, 1, 14, 28, and 49 days relative to parturition). The Y-axis depicts fold change in expression compared with −65 days relative to parturition (i.e., the first sample before cows were assigned to treatments). The average trend in expression pattern for all the genes composing each cluster is shown in pink. Genes composing each cluster have correlated expression profiles and may take part in the same or similar biological processes. Reported also are the GO biological process, GO cellular component, and KEGG pathways in DAVID that were enriched significantly (Benjamini–Hochberg multiple comparison correction <0.05). To help in data interpretation, the color of the table matches the color of the cluster. Microarray data are from a reanalysis of Loor et al. (2005, 2006). (*See insert for color representation of this figure.*)

complement system can increase susceptibility of the cow to bacteria and virus infections because it will reduce the capacity of the immune system to respond to such challenges (Qin and Gao, 2006).

Another common feature of the two diets versus controls that was uncovered by the cluster analysis is the similar and strong up-regulation of genes involved in protein synthesis (clusters 8 and 12 in Figure 9.3). The cluster 5, which grouped genes having a similar pattern as those in cluster 8 and 12, was significantly enriched in the DAVID analysis by genes involved in cytoskeletal organization. The IPA analysis confirmed the significant enrichment of protein synthesis in cluster 8. No additional functions or pathways were enriched at an FDR < 0.05. The high-enrichment and coordinated upregulation of protein synthesis-related genes by over- and under-feeding energy in the diet is a novel finding but an explanation for this response is not apparent when considering that ruminant liver protein synthesis is rather unaffected by food intake (Connell et al., 1997). However, it could be possible that the coordinated cytoskeletal modification might require large increases in protein synthesis.

The functional analysis using DAVID of the clusters with a greater temporal increase in expression in liver of energy-restricted versus energy-overfed or control cows (clusters 9, 10, and 12) highlighted a significant enrichment of terms related to catabolic activity of mitochondria (e.g., oxidative phosphorylation) and protein synthesis (no pathways or functions were enriched significantly in IPA). Those results indicated a coordinated up-regulation of catabolic activity by restricted energy feeding prepartum. The consideration of all the terms enriched with an EASE score of 0.10 (a correction of the *p*-value implemented in DAVID) uncovered that also cluster 1 was enriched by the same terms, and that the increase in catabolic activity of the mitochondria was due to an increase of glucose catabolism (probably thorough increase of TCA cycle). In IPA, cluster 1 was significantly-enriched by the fatty acid metabolism pathway, indicating a greater degree of oxidation of fatty acids in the mitochondria of energy-restricted versus energy-overfed or control cows. In particular, this cluster included genes such as carnitine palmitoyltransferase 1A, acyl-CoA dehydrogenase very long chain, acetyl-CoA acyltransferase 1, and cytochrome P450 family 3 subfamily A polypeptide 4. In addition, antigen processing and presentation was enriched in cluster 10. The coordinated increase in expression of genes involved in antigen processing and presentation indicates a significant degree of responsiveness of liver from energy-restricted cows to the presence of antigens, followed by an immune response. This conclusion appears to be supported by the larger acute-phase reaction in those cows as indicated by the greater expression of the *SAA1* and *CP*.

The clusters 3 and 7, which grouped genes with greater increase in expression (especially prepartum) in energy-overfed than energy-restricted or control cows, were not enriched with any terms at an FDR-corrected *p*-value <0.05 in DAVID or IPA. However, using as cutoff the EASE score <0.10 in DAVID we observed in both clusters an enrichment of terms related to negative regulation of RNA metabolism and processing, including transcription. The latter appears to be in contradiction with the greater number of up-regulated genes between -65 to $+14$ DIM observed in energy-overfed versus energy-restricted cows (Bionaz et al., 2007b).

The clusters 4 and 6 grouped genes with expression patterns characterized by a greater temporal decrease in energy-restricted versus energy-overfed or control cows. Cluster 4 was significantly enriched by the WNT signaling pathway, which is involved

in liver development, regeneration, metabolism, and maintenance of normal function (Behari, 2010). No functions or pathways were significantly enriched in IPA. At an EASE score <0.10 as cutoff, we found that cluster 4 was enriched in DAVID by genes involved in apoptosis and cellular response to stress; in addition, both clusters 4 and 6 were enriched by terms related to proteolysis and DNA damage response. Enrichment analysis of clusters 4 and 6 suggests that liver of energy-restricted animals had either a decreased ability to respond to damage and stress, or that liver experienced less stress. The second suggestion appears to be the most likely explanation because liver from obese rats had a decreased ability to respond to a damaging insult induced by reactive metabolites (Corcoran et al., 1989). In our experiment, energy-overfed cows had greater TAG accumulation in liver (Loor et al., 2006) and steatosis has been associated with increased liver damage (Adinolfi et al., 2001) and greater mitochondrial ROS production coupled with oxidative damage of the DNA (Gao et al., 2004). However, as suggested by clusters with a coordinated upregulation in expression in energy-restricted cows versus the other groups, energy-restricted cows appeared to have experienced an increase capacity for postpartal oxidation of fatty acids, which may have increased the production of ROS.

The more than three-fold increase in prepartal serum insulin in energy-overfed versus restricted cows could have primed the liver to accumulate more triglycerides postpartum (Loor et al., 2006); however, none of the clusters was significantly enriched by insulin signaling pathways or related terms, not even considering the noncorrected *p*-value. To further evaluate potential effects of insulin on liver, we performed a large network analysis considering all the interactions of the DEG in each cluster with insulin.

Results from the analysis suggested a minor effect of insulin on the clusters of genes, with a proportionally greater effect on cluster 10, followed by cluster 8 and 9 (the expression of 2.3, 2.0, and 1.7%, respectively, of genes in those clusters could be affected by insulin). Our analysis suggests a lack of direct role of insulin in driving coordinated transcriptome adaptations in liver of peripartal dairy cows. The higher concentration of plasma insulin in energy-overfed cows could have had an indirect effect on liver by increasing accumulation of triglycerides in adipose tissue resulting in a more sustained postpartal release of NEFA into blood (Loor et al., 2006).

To further characterize the coordinated transcriptomics adaptations to energy prepartum, we performed network analysis using IPA. The clusters that could form the largest networks of genes encompassing all possible relationships (e.g., protein/protein, DNA/protein, etc) were cluster 10 and 12, with more than 36% of the genes able to form networks, followed by cluster 4 (with 35% of the genes in cluster able to form networks). Cluster 3 was the most highly enriched by TF (ca. 46% of genes were TF).

Interestingly, except for cluster 8 where the top function among the networks of genes was protein synthesis, all the other networks in clusters were highly enriched by genes involved in cell cycle, cell death, and cell growth and proliferation. In fact, the largest transcriptional networks were enriched by TF such as MYC, TP53, and FOS, which are involved in cell cycle/death. In this last analysis it has to be taken into account that those TF are among the most studied; thus, it is not surprising that these TF have the greatest number of interactions in the original IPA knowledge base. Few clusters contained TF other than the above among the most important in terms of

numbers of genes within the network controlled. For example, in cluster 1, the largest transcriptional networks were produced by PPARA, XBP1, and STAT, while the TF with the largest connections in the networks of cluster 2 was NFE2L2 (nuclear factor erythroid-derived 2-like 2) and PPARA. In addition, HIF1A (hypoxia inducible factor 1 alpha subunit) was among the TF with the largest network in cluster 5.

The overall analysis surprisingly uncovered very few functions/pathways with a coordinated transcriptional regulation. Among those, protein synthesis appears to have been coordinately induced by either overfeeding or restricting dietary energy prepartum, with an overall larger increase in restricted cows. It also appears that there was a coordinated downregulation of genes involved in liver response to inflammation with both dietary treatments. The restriction of dietary energy prepartum appears to have coordinately increased expression of genes involved in fatty acid oxidation and energy production (with a likely role of PPARA and STAT) mostly involving mitochondrial components. Because of the well-known response of PPAR to long-chain fatty acids, our results point at potential practical uses of dietary lipid supplementation during the peripartal period. However, as we argued before (Loor, 2010), more in-depth studies with various doses and combinations of long-chain fatty acids need to be conducted with peripartal cows.

From a Systems Biology perspective, it would be important to conduct functional studies, which address the relevance of the pathways/functions that have been found to be altered by nutritional management and physiological state. A starting point in such efforts could entail the identification and testing of TF networks and their targets as we have proposed recently (Loor et al., 2007; Loor, 2010). Several transcription regulators, their target genes, and the molecular functions that they affect have been uncovered among the DEG (for additional detailed examples, refer to Loor, 2010). Those regulators could be potential markers that can be targeted via management or nutritional measures in the future.

Dietary Lipid Supplementation, Ruminal Metabolism, and the Mammary Gland Transcriptome

In the past two decades, there has been substantial interest in the possibility to modulate bovine, caprine, and ovine milk fatty acid profiles through nutrition with the goal to improve the nutritional properties of milk fat with respect to human health (Bauman et al., 2006). It is now well established that dietary factors such as level of forage and grain, source and amount of supplemental lipid (e.g., marine polyunsaturated FA, saturated FA), and their combinations can have marked effects on ruminal microbial lipid metabolism (Lourenco et al., 2010), which largely dictates the type and amount of LCFA (trans-18:1, conjugated 18:2, and 18:3 isomers versus saturated LCFA) that is available to tissues for metabolism (Shingfield et al., 2010).

Previous studies have resulted in a better understanding of the productive response of the animals, the underlying differences between species, and have increased our knowledge of the metabolism of dietary and ruminally derived LCFA in the mammary gland (Shingfield et al., 2010). Although there have been several studies of transcriptional adaptations of bovine mammary tissue to dietary lipids, they have focused mostly on lipogenic target genes (e.g., Bauman et al., 2008). Large-scale transcriptomics

adaptations of bovine mammary tissue in response to dietary lipids remain largely unknown.

The use of microarrays in rodent mammary gland in response to supplemental lipid has allowed for a more holistic view of the systematic adaptations in mammary tissue. An initial study in mice underscored the importance of gene transcription in the regulation of lipid synthesis (Rudolph et al., 2007). Although with technical, but unavoidable, limitations (caprine RNA on a bovine microarray), a study of the caprine mammary transcriptome in response to fat supplementation (Ollier et al., 2009) provided evidence of the suitability of large-scale transcriptomics to study responses to nutrients by the ruminant mammary gland.

Ruminal unsaturated lipid metabolism gives rise to a wide variety of 18:1 and 18:2 fatty acids containing trans double bonds, some which have been clearly shown (e.g., trans10–18:1, trans10, cis12–18:2) to alter bovine mammary lipid metabolism (Shingfield et al., 2010). Previous work from our lab (Bionaz and Loor, 2008) suggested that the complex dynamic process of regulation of milk fat synthesis in the bovine mammary gland may be influenced by LCFA through the PPARγ transcriptional network. Using immortalized bovine mammary epithelial cells, we recently provided demonstration that saturated LCFA affect transcription of a large array of genes involved in milk fat synthesis (Kadegowda et al., 2009).

In order to better understand the longer term systemic adaptations in biological processes and gene networks in the mammary gland in response to dietary fat supplementation, we conducted (Invernizzi et al., 2010) a transcriptomics analysis of the mammary tissue in mid-lactating cows fed a control diet (CTR), a saturated fat diet (EB), or a diet supplemented with blend of fish oil and soybean oil (FSO) for a period of 3 weeks. The temporal adaptations of 29 genes associated with milk fat synthesis have been reported previously (Invernizzi et al., 2010). At the level of production, results clearly indicated that the FSO diet decreased milk fat synthesis markedly, while feeding EB had no effect (Invernizzi et al., 2010). In addition, the composition of LCFA in milk fat was substantially affected by the treatments, with FSO-fed cows yielding greater amounts (moles/day) of trans-LCFA and lower amounts of *de novo*-synthesized LCFA in milk (Invernizzi et al., 2010). Those results confirmed previous work with similar types of diets (Shingfield et al., 2006). Despite the large effect at the level of milk fat synthesis, which was most dramatic at ~1 week postinitiation of feeding, only 2 of 29 genes were affected significantly at day 21 of feeding. Quite unexpectedly, most transcriptomics adaptations to supplemental fat were observed at 7 days postfeeding. Those results led us to conclude that control mechanisms regulating fat synthesis in response to milk fat-depressing diets occur quite rapidly, i.e., the mammary gland adapts/adjusts its metabolism via transcriptional mechanisms (Invernizzi et al., 2010)

In this section we present results from the same samples analyzed by a microarray but only at day 21 of treatments. The statistical analysis at day 21 was covariate-adjusted to day 0 (prior to feeding treatment diets) to account for potential animal variation. We found at 3 weeks of treatment a total of 1432 DEG (FDR $\leq$ 0.05 for the overall effect and a p-value $\leq$ 0.01 for each comparison) in mammary tissue due to feeding EB versus CTR, 847 DEG between FSO and CTR, and 1137 between the two lipid-supplemented treatments. Those data indicated a larger effect of saturated LCFA (i.e., EB) on the bovine mammary tissue transcriptome. The greater sensitivity

of bovine cells to saturated versus unsaturated LCFA seems to be quite consistent across different bovine cell types, as shown by previous findings from our lab using immortalized mammary or kidney epithelial cells (Invernizzi et al., 2009; Kadegowda et al., 2009; Thering et al., 2009).

The bioinformatics analysis of DEG was performed with Ingenuity Pathway Analysis (IPA relies on the ORA approach) and the DIA approach described in the first sections of this chapter. Analysis of DEG between EB and FSO using IPA showed that the top enriched functions (FDR ≤ 0.05) after 21 days of supplemental lipid feeding were associated with lipid metabolism, molecular transport, small molecule biochemistry and carbohydrate metabolism. Interpretation of the functional analysis suggests that mammary tissue of cows fed EB versus FSO decreased lipid synthesis processes (LCFA, triglyceride, and cholesterol), catabolism of fatty acids, and metabolism of carbohydrate. In addition, the same gene list contained as the most enriched pathways acute phase response, oxidative phosphorylation, and TCA cycle.

The results of pathway analysis reinforced the finding of a greater degree of energy utilization by mammary cells of cows fed for 21 days with FSO than EB; furthermore, supplemental FSO appeared to have induced a more pronounced immune response by mammary cells compared to EB. The DIA analysis results supported the above conclusions (Figure 9.4). On the basis of known biochemistry of the ruminant mammary gland (Bauman et al., 1970), a reduction in energy production in mammary cells can be partly explained by the inhibition of cytosolic isocitrate dehydrogenase and the pentose phosphate shunt (more induced by FSO than EB), both of which are the major sources of NADPH for mammary cells. The functional analysis suggest that this coordinated mechanism could partly have reduced *de novo* fatty acids synthesis in favor of greater uptake of preformed LCFA in cows fed EB versus FSO. However, the calculated *de novo* synthesis of LCFA was greater due to EB than FSO at day 21 (Invernizzi et al., 2010).

A reduction of NADPH also might involve an inhibition of glutathione metabolism (Invernizzi et al., 2009), which is dependent on NADPH. Glutathione metabolism was significantly enriched and was downregulated in the comparison of DEG between EB and FSO. The inhibition of glutathione metabolism would imply an alteration in the antioxidant status and/or the level of reactive oxygen species in mammary cells when saturated fat is fed. Opposite to our results with marine LCFA (rich in 20:5n-3 and 22:5n-3), soybean oil supplementation (with high concentration of 18:2n-6 LCFA) for 5 days decreased murine mammary mRNA expression of enzymes of the pentose phosphate shunt, mitochondrial citrate transporter, and enzymes of fatty acid synthesis (Rudolph et al., 2007). Thus, there may be species differences in sensitivity to type of dietary unsaturated LCFA.

The functional adaptations of the cows fed EB versus FSO was opposite to what would be expected from evaluating milk fat percentage and yield, i.e., saturated LCFA feeding does not result in lower milk fat percentage or yield and should not affect mammary lipogenic pathways (Shingfield et al., 2010). However, the data also indicated a large inhibition of lipid and glucose catabolism due to EB, suggesting that the mammary gland was likely using saturated LCFA as energy source or for esterification to TAG to produce milk fat. Under such scenario, the reduction in expression of genes involved in lipid metabolism pathways may be a way to control flux through these pathways. Because most changes in lipogenic gene expression assessed by qPCR

Category	FLUX and IMPACT			Most impacted patwhays in each major category	
	EB versus CTR	FSO versus CTR	EB versus FSO	EB versus CTR	FSO versus CTR
1. Metabolism				Phenylalanine metabolism	FA elongation in mitochondria
0.1 Metabolic pathways				FA elongation in mitochondria	Histidine metabolism
1.1 Carbohydrate metabolism				Histidine metabolism	Phenylalanine metabolism
1.2 Energy metabolism				Primary bile acid biosynthesis	Folate biosynthesis
1.3 Lipid metabolism				Cyanoamino acid metabolism	Tyrosine metabolism
1.4 Nucleotide metabolism				Taurine & hypotaurine metabolism	Tryptophan metabolism
1.5 Amino acid metabolism				Linoleic acid metabolism	Primary bile acid biosynthesis
1.6 Metabolism of other amino acids				Tyrosine metabolism	Glycosphingolip biosynt (globo series)
1.7 Glycan biosynthesis and metabolism				Retinol metabolism	One carbon pool by folate
1.8 Metabolism of cofactors and vitamins				Drug metab - cytochrome P450	Inositol phosphate metabolism
1.9 Metabolism of terpenoids and polyketides				Metab of xenobiotics by cyt P450	O-Mannosyl glycan biosynthesis
1.10 Biosynthesis of other secondary metabolism				Arachidonic acid metabolism	Starch & sucrose metabolism
1.11 Xenobiotics biodegradation and metabolism				Nitrogen metabolism	Sphingolipid metabolism
2. Genetic information processing				Homologous recombination	RNA polymerase
2.1 Transcription				RNA polymerase	Aminoacyl-tRNA biosynthesis
2.2 Translation				Protein export	Protein export
2.3 Folding, sorting, and degradation				RNA degradation	Mismatch repair
2.4 Replication and repair				Mismatch repair	Spliceosome
3. Environmental information processing				mTOR sign pathway	Calcium sign pathway
3.1 Membrane transport				Calcium sign pathway	Phosphatidylinositol sign system
3.2 Signal transduction				Phosphatidylinositol sign. system	ABC transporters
3.3 Signaling molecules and interaction				ECM-receptor interaction	mTOR sign pathway
4. Cellular processes				Regulation of autophagy	Lysosome
4.1 Transport and catabolism				Peroxisome	Regulation of autophagy
4.2 Cell motility				p53 sign pathway	p53 sign pathway
4.3 Cell growth and death				Gap junction	Adherens junction
4.4 Cell communication				Oocyte meiosis	Tight junction
5. Organismal systems				Aldosterone-regulated Na reabsorpt	Complement & coagulation cascades
5.1 Immune system				Cytosolic DNA-sensing pathway	Hematopoietic cell lineage
5.2 Endocrine system				PPAR sign pathway	Pancreatic secretion
5.3 Circulatory system				Complement & coagulation cascades	Cytosolic DNA-sensing pathway
5.4 Digestive system				Proximal tubule bicarb reclamation	Toll-like receptor sign pathway
5.5 Excretory system				Cardiac muscle contraction	Aldosterone-regulated Na reabsorpt
5.6 Nervous system				Long-term potentiation	RIG-I-like receptor sign pathway
5.7 Sensory system				NOD-like receptor sign pathway	Circadian rhythm - mammal
5.8 Development				Gastric acid secretion	NOD-like receptor sign pathway
5.9 Environmental adaptation				Pancreatic secretion	Insulin sign pathway

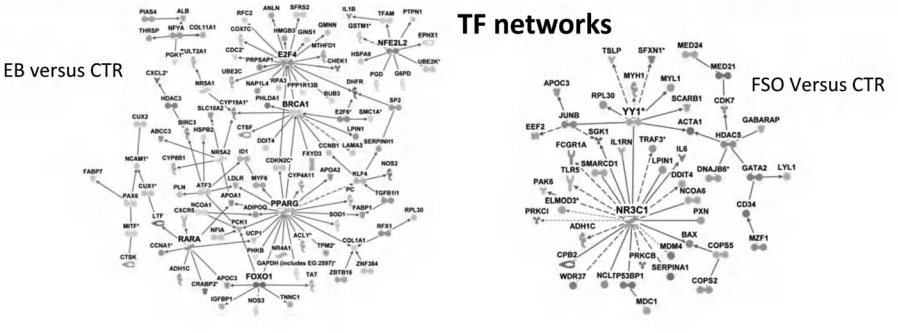


Figure 9.4 Overall calculated dynamic impact (DIA) in KEGG pathways of DEG in bovine mammary tissue from midlactation cows fed a control diet (CTR), the control diet supplemented with saturated lipid (EB), or the control diet supplemented with a blend of fish oil and soybean oil (FSO) for 3 weeks. Shown are the main pathway classification groups (left column) and corresponding subgroups. The heat map denotes potential increase (red shade) or decrease (green shade) of metabolic flux or signaling through the pathway for each treatment comparison. The overall impact is denoted by the size of the blue bar (the larger the bar the greater the impact of DEG on the category of pathways). The most impacted pathways with the overall flux (red shade denotes increases and green shade denotes decreases) for the comparison EB versus CTR and FSO versus CTR are shown in the right column. The transcription factor (TF) networks produced by DEG in EB versus CTR and FSO versus CTR are reported in the bottom panel. The TF in each network are highlighted by larger font. Details of the animal experiment and portions of the microarray analysis have been reported previously (Invernizzi et al., 2009, 2010). (*See insert for color representation of this figure.*)

were observed at 7 than 21 days post treatment (Invernizzi et al., 2010), it is likely that the "bigger-picture" response to long-term supplemental lipid represented a type of adaptation despite the sustained reduction in milk fat synthesis through day 21 in cows fed FSO. The results point to an interpretation of the large-scale transcriptomics data in which the changes observed due to FSO are a consequence of the mammary tissue attempting to counteract the milk fat depression effect, whereas, in the case of EB, mammary tissue is controlling a milk fat synthesis response due to influx of saturated LCFA.

The analysis of KEGG pathways using DIA (Figure 9.4) suggested an activation of amino acid metabolism in FSO versus CTR. In general, feeding FSO or EB versus CTR resulted in an overall activation of phenylalanine, histidine, and tryptophan metabolism (Figure 9.4). In addition, primary bile acid biosynthesis was a highly impacted pathway by both EB and FSO versus CTR but in an opposite fashion, i.e., flux was upregulated by FSO and downregulated by EB (Figure 9.4). Although it is unlikely that mammary cells synthesize bile acids per se, the significance of this metabolic pathway likely is due to an increase in the availability of cholesterol in blood as a response to lipid supplementation (e.g., Loor et al., 2005). Such a response is supported by the fact that negative regulation of cholesterol import and negative regulation of sterol and cholesterol transport (among GO biological processes) were the most impacted in the comparison of EB versus CTR; it can be envisioned that such responses at the level of mammary tissue favored an increase in cholesterol import into the mammary gland, thus, decreasing the need to use LCFA for cholesterol synthesis within the tissue. Among the pathways related to genetic information processing, protein export was activated by FSO versus CTR but inhibited with EB versus CTR. This finding suggests a potential decrease in protein secretion due to EB and increase in protein secretion due to FSO. However, milk protein percentage and yield did not differ due to diet (Invernizzi et al., 2010).

The PPAR signaling pathway was the most-inhibited pathway under the organismal system category in the comparison EB versus CTR. Studies performed in our lab support the idea that PPARγ exerts a crucial role in mammary gland lipid metabolism in response to saturated long-chain fatty acids or very long-chain polyunsaturated fatty acids, with an activation of the nuclear receptor network by saturated LCFA (Kadegowda et al., 2009). Because PPARγ is a nuclear receptor controlling expression of several genes involved in milk fat synthesis and its activity can be modulated by LCFA, the practical implications of our findings are exciting. By increasing or decreasing supplementation of specific LCFA in the diet, it might be possible to modulate milk fat synthesis for a desire purpose, for example, reduce energy utilization by mammary gland during the peripartal period as a means to lessen NEB.

The large degree of inhibition of the PPAR signaling pathway due to EB appears to contradict our previous *in vitro* data. However, besides the *in vitro* nature of our previous work, we evaluated the short-term effect (12 hours post treatment) while in the present analysis we are clearly dealing with long-term consequences of dietary lipid supplementation in addition to the "confounding" of ruminal and inter-organ metabolism. As mentioned above, the fact that *in vivo* and *in vitro* data contrast might be explained by the need of the mammary gland to control milk fat synthesis due to influx of saturated LCFA. The data from the present analysis clearly indicated that feeding LCFA, both saturated and unsaturated, has a strong and long-term

transcriptomics effect on the mammary gland. In the case of FSO, despite the pronounced negative effect on milk fat synthesis (Invernizzi et al., 2010) driven partly by short-term transcriptional adaptations in key lipogenic genes (Invernizzi et al., 2010), the mammary gland attempted to counteract the milk fat depressing effect by further transcriptional adaptations beyond the classical lipogenic pathway. This hypothesis will need to be tested further. However, if it held true, this new knowledge might be exploited in practical terms to help in preventing milk fat depression.

Network analysis of TF and their targets within DEG was conducted to aid in the identification of putative mechanisms of differential regulation of mammary tissue in response to dietary EB or FSO (Figure 9.4). Feeding EB up-regulated expression of E2F transcription factor 4, p107/p130-binding (*E2F4*), nuclear factor (erythroid-derived 2)-like 2 (*NFE2L2*), retinoic acid receptor, alpha (*RARA*), forkhead box O1 (*FOXO1*), and nuclear transcription factor Y, alpha (*NFYA*); whereas, it down-regulated expression of breast cancer 1 early onset (*BRCA1*) and nuclear receptor subfamily 5 group A member 2 (NR5A2). In addition, there were a large number of PPARγ target genes affected (Figure 9.4) and most were downregulated. The majority of those genes are involved in transport and glyceroneogenesis (i.e., pyruvate carboxylase (*PC*); phosphoenolpyruvate carboxykinase 1 (soluble), *PCK1*). Among TF affected by feeding FSO, YY1 transcription factor (*YY1*, downregulated), histone deacetylase 5 (*HDAC5*, upregulated), and nuclear receptor subfamily 3 group C member 1 (glucocorticoid receptor) (*NR3C1*, upregulated) formed the largest networks with other target DEG (Figure 9.4). Those DEG are involved in response to stimulus, development, transcription, and programmed cell death.

The above TF uncovered by our study are potential candidates for future molecular studies in order to further characterize networks regulating milk fat synthesis beyond the classical lipogenic genes. Once their role in mammary tissue has been demonstrated, an additional step would be to determine whether they bind specific nutrients including LCFA and what effects that may bring about. These approaches will lead to identification of specific agonists and antagonists that could be used to manipulate milk fat synthesis and/or other functional aspects of mammary function (e.g., milk protein synthesis).

A detailed analysis of the mammary response to supplemental lipid was beyond the scope of this manuscript. This brief discussion, in conjunction with qPCR data reported previously (Invernizzi et al., 2010), allowed us to propose that mammary gland is capable of adapting to the specific cellular concentration of LCFA, which is largely dictated by diet and ruminal metabolism. In the milk fat-depression scenario, the mammary gland of cows fed for more than 20 days with FSO counteracted the antilipogenic effect by increasing expression of genes involved in lipid synthesis, particularly the synthesis of cholesterol and TAG. In the milk fat-enhancing scenario, the large influx of saturated LCFA the mammary gland controlled lipid metabolic fluxes by decreasing the expression of genes coding for enzymes involved in those biological processes.

The analysis, interpretation, and conclusion of this nutrigenomic study only included mammary tissue, but the mammary is highly dependent on other tissues such as liver and adipose which also are sensitive to LCFA. There is evidence that those tissues also are altered by dietary fat (Harvatine et al., 2009; Thering et al., 2009), thus, the ensuing biological events can potentially influence mammary gland adjustments

to LCFA. This represents another important consideration in the context of Systems Biology, i.e., mammary function is one important element of an integrated system. Future studies of lipid supplementation would benefit from simultaneous analyses of molecular and flux data across key organs or cell types.

Perspectives

Although still far from "complete" Systems Biology as exemplified by work in model organisms (e.g., Ishii et al., 2007), use of transcriptomics coupled with bioinformatics analysis and blood and tissue-level data, have allowed for a more holistic study of the multifaceted adaptations of livestock tissues. This adaptation not only refers to nutrition strategies but also to changes in physiological state. This information provides the basis for more detailed functional studies that could encompass evaluations of gene function (e.g., via gene silencing) and subsequent effects on transcriptional networks. There are ongoing efforts to develop algorithms for identifying TF and their regulatory networks in livestock tissues (e.g., Hudson et al., 2009). These adaptations undoubtedly will add to the currently available tools for bioinformatics analysis.

Current limitations for a more complete Systems Biology approach in livestock include the lack of other "omics" data (e.g., metabolomics, proteomics, microRNAomics), incomplete genome annotation, lack of measurements of biochemical fluxes and tissue nutrient flows, and the lack of additional bioinformatics tools to analyze with a greater confidence the time course of the regulation of tissue function, or to interpret multiple treatment experimental designs. Despite those limitations, the Systems Biology approach appears extremely promising. Besides leading to the discovery of regulatory targets, the Systems Biology approach might help to address a broader spectrum of basic and practical applications including interpretation of phenotypic data, metabolic engineering, or interpretation of lactation phenotypes.

Acknowledgments

The authors express their sincere appreciation to J. K. Drackley, S. L. Rodriguez-Zas, H. A. Lewin, R. M. Akers, and R. E. Everts (University of Illinois) for their support and collaboration during conduct of the studies presented in this chapter.

References

Adinolfi, L.E., Gambardella, M., Andreana, A., et al. (2001) Steatosis accelerates the progression of liver damage of chronic hepatitis C patients and correlates with specific HCV genotype and visceral obesity. *Hepatology* **33**, 1358–1364.

Allen, M.S., Bradford, B.J., & Harvatine, K.J. (2005) The cow as a model to study food intake regulation. *Annual Reviews of Nutrition* **25**, 523–547.

Bauman, D.E., Brown, R.E., & Davis, C.L. (1970) Pathways of fatty acid synthesis and reducing equivalent generation in mammary gland of rat, sow, and cow. *Archives of Biochemistry and Biophysics* **140**, 237–244.

Bauman, D.E., Mather, I.H., Wall, R.J., et al. (2006) Major advances associated with the biosynthesis of milk. *Journal of Dairy Science* **89**, 1235–1243.

Bauman, D.E., Perfield 2nd, J.W., Harvatine, K.J., et al. (2008) Regulation of fat synthesis by conjugated linoleic acid: lactation and the ruminant model. *Journal of Nutrition* **138**, 403–409.

Behari, J. (2010) The Wnt/β-catenin signaling pathway in liver biology and disease. *Expert Review of Gastroenterology & Hepatology* **4**, 745–756.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289–300.

Bertoni, G., Trevisi, E., & Lombardelli, R. (2009) Some new aspects of nutrition, health conditions and fertility of intensively reared dairy cows. *Italian Journal of Animal Science* **8**, 491–518.

Bionaz, M. & Loor, J.J. (2008) Gene networks driving bovine milk fat synthesis during the lactation cycle. *BMC Genomics* **9**, 366.

Bionaz, M., Drackley, J.K., Rodriguez-Zas, S.L., et al. (2007b) Uncovering adaptive hepatic gene networks due to prepartum plane of dietary energy and physiological state in periparturient Holstein cows. *Journal of Dairy Science* **90**(E-Suppl. 1), 971.

Bionaz, M., Trevisi, E., Calamari, L., et al. (2007a) Plasma paraoxonase, health, inflammatory conditions, and liver function in transition dairy cows. *Journal of Dairy Science* **90**, 1740–1750.

Bruckbauer, A., Gouffon, J., Rekapalli, B., et al. (2009) The effects of dairy components on energy partitioning and metabolic risk in mice: A microarray study. *Journal of Nutrigenetics and Nutrigenomics* **2**, 64–77.

Bruggeman, F.J. & Westerhoff, H.V. (2007) The nature of systems biology. *Trends in Microbiology* **15**, 45–50.

Cogburn, L.A., Porter, T.E., Duclos, M.J., et al. (2007) Functional genomics of the chicken – a model organism. *Poultry Science* **86**, 2059–2094.

Columbano, A. & Ledda-Columbano, G.M. (2003) Mitogenesis by ligands of nuclear receptors: An attractive model for the study of the molecular mechanisms implicated in liver growth. *Cell Death and Differentiation* **10**(Suppl. 1), S19–S21.

Connell, A., Calder, A.G., Anderson, S.E., et al. (1997) Hepatic protein synthesis in the sheep: effect of intake as monitored by use of stable-isotope-labelled glycine, leucine and phenylalanine. *British Journal of Nutrition* **77**, 255–271.

Corcoran, G.B., Salazar, D.E., & Chan, H.H. (1989) Obesity as a risk factor in drug-induced organ injury. III. Increased liver and kidney injury by furosemide in the obese overfed rat. *Toxicology and Applied Pharmacology* **98**, 12–24.

Daniels, K.M., Capuco, A.V., McGilliard, M.L., et al. (2009a) Effects of milk replacer formulation on measures of mammary growth and composition in Holstein heifers. *Journal of Dairy Science* **92**, 5937–5950.

Daniels, K.M., McGilliard, M.L., Meyer, M.J., et al. (2009b) Effects of body weight and nutrition on histological mammary development in Holstein heifers. *Journal of Dairy Science* **92**, 499–505.

Dann, H.M., Litherland, N.B., Underwood, J.P., et al. (2006) Diets during far-off and close-up dry periods affect periparturient metabolism and lactation in multiparous cows. *Journal of Dairy Science* **89**, 3563–3577.

Doppler W., Groner, B., & Ball, R.K. (1989) Prolactin and glucocorticoid hormones synergistically induce expression of transfected rat beta-casein gene promoter constructs in a mammary epithelial cell line. *Proceedings of the National Academy of Sciences of the U S A* **86**, 104–108.

Drackley, J.K. (1999) Biology of dairy cows during the transition period: The final frontier. *Journal of Dairy Science* **82**, 2259–2273.

Drackley, J.K., Donkin, S.S., & Reynolds, C.K. (2006) Major advances in fundamental dairy cattle nutrition. *Journal of Dairy Science* **89**, 1324–1336.

Draghici, S., Khatri, P., Tarca, A.L., et al. (2007) A systems biology approach for pathway level analysis. *Genome Research* **17**, 1537–1545.

Feist, A.M. & Palsson, B.O. (2008) The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. *Nature Biotechnology* **26**, 659–667.

Finucane, K.A., McFadden, T.B., Bond, J.P., et al. (2008) Onset of lactation in the bovine mammary gland: Gene expression profiling indicates a strong inhibition of gene expression in cell proliferation. *Functional and Integrative Genomics* **8**, 251–264.

Fukushima, Y., Kasuga, M., & Nakao, K. (2009) Effects of coffee on inflammatory cytokine gene expression in mice fed high-fat diets. *Journal of Agricultural and Food Chemistry* **57**, 11100–11105.

Gao, D., Wei, C., Chen, L., et al. (2004) Oxidative DNA damage and DNA repair enzyme expression are inversely related in murine models of fatty liver disease. *American Journal of Physiology Gastrointestinal and Liver Physiology* **287**, G1070–G1077.

Grossman, M. & Koops, W.J. (2003) Modeling extended lactation curves of dairy cattle: A biological basis for the multiphasic approach. *Journal of Dairy Science* **86**, 988–998.

Grummer, R.R., Mashek, D.G., & Hayirli, A. (2004) Dry matter intake and energy balance in the transition period. *Veterinary Clinics of North America Food Animal Practice* **20**, 447–470.

Harvatine, K.J., Perfield 2nd, J.W., & Bauman, D.E. (2009) Expression of enzymes and key regulators of lipid synthesis is upregulated in adipose tissue during CLA-induced milk fat depression in dairy cows. *Journal of Nutrition* **139**, 849–854.

Hausman, G.J., Dodson, M.V., Ajuwon, K., et al. (2009) Board-invited review: The biology and regulation of preadipocytes and adipocytes in meat animals. *Journal of Animal Science* **87**, 1218–1246.

Higami, Y., Barger, J.L., Page, G.P., et al. (2006) Energy restriction lowers the expression of genes linked to inflammation, the cytoskeleton, the extracellular matrix, and angiogenesis in mouse adipose tissue. *Journal of Nutrition* **136**, 343–352.

Huang da, W., Sherman, B.T., & Lempicki, R.A. (2009a) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**, 1–13.

Huang da, W., Sherman, B.T., & Lempicki, R.A. (2009b) Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nature Protocols* **4**, 44–57.

Huang da, W., Sherman, B.T., Zheng, X., et al. (2009c) Extracting biological meaning from large gene lists with David. *Current Protocols in Bioinformatics* Chapter **13**: Unit 13. 11.

Hudson, N.J., Reverter, A., Wang, Y., et al. (2009) Inferring the transcriptional landscape of bovine skeletal muscle by integrating co-expression networks. *PLoS One* **4**, e7249.

Ikeda, K., Halle, J.P., Stelzer, G., et al. (1998) Involvement of negative cofactor nc2 in active repression by zinc finger-homeodomain transcription factor areb6. *Molecular and Cellular Biology* **18**, 10–18.

Invernizzi, G., Thering, B.J., Bionaz, M., et al. (2009) New insights on mammary tissue responses to dietary lipids using transcriptomics. In: *Proceedings XIth International Symposium on Ruminant Physiology*, Clermont-Ferrand, France, Page 540. Wageningen Academic Publishers, Wageningen.

Invernizzi, G., Thering, B.J., McGuire, M.A., et al. (2010) Sustained upregulation of stearoyl-CoA desaturase in bovine mammary tissue with contrasting changes in milk fat synthesis and lipogenic gene networks caused by lipid supplements. *Functional and Integrative Genomics* **10**, 561–575.

Ishii, N., Nakahigashi, K., Baba, T., et al. (2007) Multiple high-throughput analyses monitor the response of E. coli to perturbations. *Science* **316**, 593–597.

Janovick, N.A. & Drackley, J.K. (2010) Prepartum dietary management of energy intake affects postpartum intake and lactation performance by primiparous and multiparous Holstein cows. *Journal of Dairy Science* **93**, 3086–3102.

Janovick, N.A., Loor, J.J., Ji, P., et al. (2009) Overfeeding energy prepartum dramatically affects peripartal expression of mRNA transcripts in subcutaneous adipose tissue compared with controlling energy intake prepartum. *Journal of Dairy Science* **92**(E-Suppl. 1), 709.

Jobgen, W., Fu, W.J., & Gao, H. (2009) High fat feeding and dietary L-arginine supplementation differentially regulate gene expression in rat white adipose tissue. *Amino Acids* **37**, 187–198.

Kadegowda, A.K., Bionaz, M., Piperova, L.S., et al. (2009) Peroxisome proliferator-activated receptor-gamma activation and long-chain fatty acids alter lipogenic gene networks in bovine mammary epithelial cells to various extents. *Journal of Dairy Science* **92**, 4276–4289.

Kim, Y.J. & Park, T. (2008) Genes are differentially expressed in the epididymal fat of rats rendered obese by a high-fat diet. *Nutrition Research* **28**, 414–422.

LaRosa, P.C., Miner, J., Xia, Y., et al. (2006) Trans-10, cis-12 conjugated linoleic acid causes inflammation and delipidation of white adipose tissue in mice: A microarray and histological analysis. *Physiological Genomics* **27**, 282–294.

Lee, D.E., Kehlenbrink, S., Lee, H., et al. (2009) Getting the message across: Mechanisms of physiological cross talk by adipose tissue. *American Journal of Physiology Endocrinology and Metabolism* **296**, E1210–E1229.

Lemay, D.G., Neville, M.C., Rudolph, M.C., et al. (2007) Gene regulatory networks in lactation: Identification of global principles using bioinformatics. *BMC Systems Biology* **1**, 56.

Leroy, J.L., Vanholder, T., Van Knegsel, A.T., et al. (2008) Nutrient prioritization in dairy cows early postpartum: mismatch between metabolism and fertility. *Reproduction of Domestic Animals* **43**(Suppl. 2), 96–103.

Lin, J. & Qian, J. (2007) Systems biology approach to integrative comparative genomics. *Expert Reviews in Proteomics* **4**, 107–119.

Lippolis, J.D. & Reinhardt, T.A. (2008) Centennial paper: proteomics in animal science. *Journal of Animal Science* **86**, 2430–2441.

Loor, J.J. (2010) Genomics of metabolic adaptations in the peripartal cow. *Animal* **4**, 1110–1139.

Loor, J.J. & Cohick, W.S. (2009) ASAS centennial paper: Lactation biology for the twenty-first century. *Journal of Animal Science* **87**, 813–824.

Loor, J.J., Dann, H.M., Everts, R.E., et al. (2005) Temporal gene expression profiling of liver from periparturient dairy cows reveals complex adaptive mechanisms in hepatic function. *Physiological Genomics* **23**, 217–226.

Loor, J.J., Dann, H.M., Janovick-Guretzky, N.A., et al. (2006) Plane of nutrition prepartum alters hepatic gene expression and function in dairy cows as assessed by longitudinal transcript and metabolic profiling. *Physiological Genomics* **27**, 29–41.

Loor, J.J., Everts, R.E., Bionaz, M., et al. (2007) Nutrition-induced ketosis alters metabolic and signaling gene networks in liver of periparturient dairy cows. *Physiological Genomics* **32**, 105–116.

Lourenco, M., Ramos-Morales, E., & Wallace, R.J. (2010) The role of microbes in rumen lipolysis and biohydrogenation and their manipulation. *Animal* **4**, 1008–1023.

Matsui, N., Ito, R., Nishimura, E., et al. (2005) Ingested cocoa can prevent high-fat diet-induced obesity by regulating the expression of genes for fatty acid metabolism. *Nutrition* **21**, 594–601.

McNamara, J.P. (1989) Regulation of bovine adipose tissue metabolism during lactation 5. Relationships of lipid synthesis and lipolysis with energy intake and utilization. *Journal of Dairy Science* **72**, 407–418.

Morandini, P. (2009) Rethinking metabolic control. *Plant Science* **176**, 441–451.

Ollier, S., Leroux, C., de la Foye, A., et al. (2009) Whole intact rapeseeds or sunflower oil in high-forage or high-concentrate diets affects milk yield, milk composition, and mammary gene expression profile in goats. *Journal of Dairy Science* **92**, 5544–5560.

Palou, M., Sánchez, J., Priego, T., et al. (2010) Regional differences in the expression of genes involved in lipid metabolism in adipose tissue in response to short- and medium-term fasting and refeeding. *Journal of Nutritional Biochemistry* **21**, 23–33.

Piantoni, P., Bionaz, M., Graugnard, D.E., et al. (2010) Functional and gene network analyses of transcriptional signatures characterizing pre-weaned bovine mammary parenchyma or fat pad uncovered novel inter-tissue signaling networks during development. *BMC Genomics* **11**, 331.

Piantoni, P., Graugnard, D., Daniels, K.M., et al. (2007) Prepubertal nutrition effects on bovine mammary parenchyma and fat pad gene expression profiles. *Journal of Dairy Science*, **90**(Suppl. 1), 269.

Piechota, M., Korostynski, M., & Przewlocki, R. (2010) Identification of cis-regulatory elements in the mammalian genome: the cREMaG database. *PLoS One* **5**, e12465.

Qin, X. & Gao, B. (2006) The complement system in liver diseases. *Cellular and Molecular Immunology* **3**, 333–340.

Rabelo, E., Rezende, R.L., Bertics, S.J., et al. (2003) Effects of transition diets varying in dietary energy density on lactation performance and ruminal parameters of dairy cows. *Journal of Dairy Science* **86**, 916–925.

Rudolph, M.C., McManaman, J.L., Phang, T., et al. (2007) Metabolic regulation in the lactating mammary gland: A lipid synthesizing machine. *Physiological Genomics* **28**, 323–336.

Shearer, J., Duggan, G., Weljie, A., et al. (2008) Metabolomic profiling of dietary-induced insulin resistance in the high fat-fed c57bl/6j mouse. *Diabetes Obesity and Metabolism* **10**, 950–958.

Sheehy, P.A., Riley, L.G., Raadsma, H.W., et al. (2009) A functional genomics approach to evaluate candidate genes located in a QTL interval for milk production traits on BTA6. *Animal Genetics* **40**, 492–498.

Shingfield, K.J., Bernard, L., Leroux, C., et al. (2010) Role of trans fatty acids in the nutritional regulation of mammary lipogenesis in ruminants. *Animal* **4**, 1140–1166.

Shingfield, K.J., Reynolds, C.K., Hervás, G., et al. (2006) Examination of the persistency of milk fatty acid composition responses to fish oil and sunflower oil in the diet of dairy cows. *Journal of Dairy Science* **89**, 714–732.

Sturn, A., Quackenbush, J., & Trajanoski, Z. (2002) Genesis: Cluster analysis of microarray data. *Bioinformatics* **18**, 207–208.

Swindell, W.R. (2008) Comparative analysis of microarray data identifies common responses to caloric restriction among mouse tissues. *Mechanisms of Ageing and Development* **129**, 138–153.

Tabach, Y., Brosh, R., Buganim, Y., et al. (2007) Wide-scale analysis of human functional transcription factor binding reveals a strong bias towards the transcription start site. *PLoS One* **2**, e807.

Takamatsu, K., Tachibana, N., Matsumoto, I., et al. (2004) Soy protein functionality and nutrigenomic analysis. *Biofactors* **21**, 49–53.

Thering, B.J., Graugnard, D.E., Piantoni, P., et al. (2009) Adipose tissue lipogenic gene networks due to lipid feeding and milk fat depression in lactating cows. *Journal of Dairy Science* **92**, 4290–4300.

Wang, Y.H., Bower, N.I., Reverter, A., et al. (2009) Gene expression patterns during intramuscular fat development in cattle. *Journal of Animal Science* **87**, 119–130.

Wiley, A.S. (2007) The globalization of cow's milk production and consumption: biocultural perspectives. *Ecology of Food and Nutrition* **46**, 281–312.

Yeung K.Y., Haynor, D.R., & Ruzzo, W.L. (2001) Validating clustering for gene expression data. *Bioinformatics* **17**, 309–318.

Chapter 10
Host–Pathogen Interactions

Mari A. Smits and Dirkjan Schokker

Introduction

Despite huge research efforts to understand the interactions between hosts and pathogens during an infection, it is still very difficult to translate the generated fundamental knowledge into strategies to prevent and fight infection diseases effectively, not only in humans but also in livestock animals. This inability is most probably due to the focus of researchers on particular components of the "host–pathogen interaction" system. Nevertheless, these focused approaches have been very successful in the identification of major virulence and host defense mechanisms. However, the research field is not able yet to explain the course and severity of infections due to the fact that several hundreds of host- and pathogen-encoded components interact with each other in a complex manner, each contributing to the infection process, either quantitatively or qualitatively. Therefore, appropriate methodologies are required to integrate and analyze the vast amount of existing and future interaction data in order to get a more complete picture of host–pathogen interaction networks. The field of Systems Biology provides novel concepts and methodologies that allow such analyses on the behavior of biological systems. This chapter describes the first steps in the application of Systems Biology to study the behavior of two highly interacting biological systems, namely that of the invading pathogen and that of the defending host.

In a general sense, the research field of host–pathogen interactions is well developed due to its impact and importance of infectious diseases that threaten human and animal health. The field is closely linked to our growing fundamental understanding of the host's innate and adaptive immune system and to our current understanding of the wide variety of mechanisms that pathogenic microorganisms use to invade, reside, and replicate within hosts. Quantitative and qualitative aspects of host–pathogen interactions determine whether pathogens are able to invade hosts, survive and replicate, spread throughout the body, and transmitted to other hosts. These aspects also determine the pathology and severity of the disease resulting from the interplay between host and pathogen: either elimination or colonization of the microbe without causing clinical signs of disease, or colonization of the microbe causing diseas, ranging from mild to deadly. Both microbial virulence as well as host disease susceptibilities are frequently seen as traits of pathogens and hosts, respectively. However, this is not

Systems Biology and Livestock Science, First Edition. Edited by Marinus F.W. te Pas, Henri Woelders, and André Bannink.

correct since both depend on dynamic interactions between the two (Casadevall and Pirofski, 2000).

Recent studies focusing on host–pathogen interactions increasingly use high-throughput -omics technologies that generate ten to several hundred of thousands of data points. They provide genome-wide global views of the molecular structures and molecular compositions of biological samples. Such studies have shed light on several virulence strategies used by microbes, on a number of defense strategies used by hosts, and on several mechanisms by which the interactions between hosts and pathogens are influenced by external factors, such as nutrition and stress. However, biological functions do not simply manifest themselves from the addition of the properties of system components, but rather arise from the dynamic interactions of these components. In addition, individual studies usually focus on one specific biological level, i.e., genes, cytokines, or macrophage activity, on particular cells or tissues, i.e., dendritic cells (DCs) or spleen, a particular time frame, and either host-response or pathogen inference. However, to understand the genetics and physiology of host–pathogen interactions, it is required to get data information from different time frames and different scales, i.e., genes, molecules, networks, pathways, cells (host as well as pathogens), tissues, organs, organisms.

Unfortunately, it is still difficult to analyze such multiscale systems as a whole. As described elsewhere in this book, the framework of obtaining a better view of the behavior of complex biological systems is now beginning to emerge through the application of Systems Biology. The goal of the application of Systems Biology in host–pathogen research is the development of models that describe the biology of host–pathogen interactions. These models will provide a framework to predict (aspects of) the outcome of "host–pathogen interaction" system in response to changes in the environment, host, and pathogens. The availability of such models may provide a sound innovative basis for improving the prevention and intervention of infectious diseases in farmed animals. Several examples, approaches, and perspectives are given in the rest of this chapter.

Data Explosion and the Rationale for Systems Biology Approaches

Data Explosion

Livestock research experiences an enormous data explosion. Complete genome sequences of major livestock species are available for several years now (see, e.g., http://www.ebi.ac.uk/ and http://www.ncbi.nlm.nih.gov/) and projects to sequence individual livestock genomes are underway. In addition, data on the genetic variability of livestock genomes, especially with regard to single nucleotide polymorphisms (SNP), is rapidly expanding. The field of pathogen genomics is even more mature compared to that of livestock genomics. Because of their relatively small sizes, pathogen genomes were the first to be completely sequenced. These efforts have provided researchers fundamental insights into the biology of pathogens, evolutionary relationships, and the determinants of virulence. The availability of a new-generation DNA sequencing machines using massively parallel approaches increases the amount of DNA- and RNA-based information as well as information on genetic variation (Mardis, 2008a, 2008b). With this type of information, the genetic potential of organisms can be

documented and, theoretically, their complete genetic potential for traits and characteristics can be assessed. Advanced statistical approaches are currently successfully applied to investigate direct correlations between genome sequence variants and phenotypic characteristics (Mardis, 2008b).

High-throughput functional genomic approaches are also increasingly used in this area for the identification of system components, such as genes, transcripts and proteins, and processes, such as ligand–receptor interactions, cell-to-cell communication (hormones, cytokines), cell proliferation, cell differentiation, and cell motility. Such studies use a wide variety of *in vitro*, *ex vivo*, and *in vivo* infection systems and focus on either individual cells, tissues, and/or organs. They compare infected hosts with uninfected hosts, hosts that differ in their susceptibility to (particular) pathogens, as well as the effect of pathogens (serotypes, isolates, mutants) that differ in virulence characteristics. These studies shed light on the composition and dynamics of the biological systems and help to identify the molecular and cellular mechanisms involved in the phenotypic characteristics displayed by the system. Especially mRNA expression profiling is a powerful tool to study the behavior of host–pathogen interaction systems (http://www.ncbi.nlm.nih.gov/sites/entrez?db=geo). Technical improvements in the area of functional genomic approaches will further enhance our insight into the spatiotemporal dynamics of gene expression (Yashiro et al., 2009). In a similar way, technical advances in the field of mass spectrometry will also contribute to a data explosion in proteomic and metabolomic research (Lippolis and Reinhardt, 2008). Furthermore, high-throughput cell-based assays are widespread and provide massive data on system parameters that go beyond the level of cellular components (Wunder et al., 2008).

Rationale for Systems Biology Approaches

A number of specific host–pathogen interactions have been addressed by various different experimental and theoretical approaches. For example, the effect of pathogen W on the gene expression of *in vitro*-cultured cell type X, or the effect of a host serum factor Y on the expression level of virulence factor Z. As indicated in the previous paragraph and elsewhere in this book, the behavior of biological systems cannot be derived from the simple addition of the properties of (sub)system components, but emerge from the dynamic interactions between all system components. Fortunately, the framework of obtaining a better view of the properties of biological systems is now beginning to emerge through the application of Systems Biology. Unlike traditional research, Systems Biology approaches take into account the information derived from multiple biological scales and multiple time frames. The goal of host–pathogen Systems Biology is to develop models representing the biological mechanisms underlying host–pathogen interaction at the various scales and time frames, both from the perspective of the host as well as that of the pathogen. Undoubtedly, such models will consist of different modules, each describing a specific aspect of the interaction between hosts and pathogens. Each module will include key biological building blocks as nodes and quantitative parameters representing the flow of material and information within the system. An important issue hereby is to identify the external variable factors affecting model outputs. They include host and pathogen genetic factors, historical factors, and factors of the production environment, such as housing, management, nutrition, humidity, pathogen load, stress, etc. The development of such models in

which the dynamic (kinetic) relationships between system components and the effect of external factors are represented by mathematical equations, will allow the computational simulations of system responses. However, it should be noted here that at this stage the development of models that integrate various, different aspects, such as metabolic aspects, immunological aspects, and health aspects, as a function of two independent genomes and multiple environmental factors is a major challenge. It is particularly difficult because two different systems (host and pathogen) as well as their interactions need to be understood. As indicated later on in this chapter, the novel concepts and methods that are or will be developed within the Systems Biology arena, may help to meet these challenges.

The expectation is that not only the models but even the individual modules of such models will provide a robust framework to predict aspects of the outcome of host–pathogen interactions in response to external changes and/or changes in host and microbe genetics. The expectation is also that the availability of such models will contribute to the improvement of diagnosis, development of improved drugs, discovery of new methods for therapeutic intervention, and prevention of infectious diseases. Like in the human biomedical field, Systems Biology may be used to develop veterinary practices that are more preventive and predictive than they are now. Specific benefits would be: the assessment of the probability of farm animals to develop (specific) infectious diseases, the selection of animals adapted to specific health management programs, the selection of animals with lower disease susceptibility to assist in developing early warning systems for disease, and the identification of new targets for diagnosis, prevention, and intervention.

Infection Biology

Infection biology is a research area that combines the fields of immunology and microbiology (virology, bacteriology, parasitology, and epidemiology) in order to get more insight in the development of infectious diseases. In this paragraph, we briefly discuss the immune system, which is classified as a two-component system consisting of innate and adaptive immunity. The innate system is the most dominant one and is activated when microbes are recognized by components that are conserved among broad groups of microorganisms or when host cells send out alarm signals. Innate immune defenses are not antigen-specific, they operate in a generic way, and do not lead to immunological memory. The adaptive system provides an antigen-dependent lagged response, it requires the specific recognition of "nonself" antigens, and leads to immunological memory. This memory is used to quickly eliminate pathogens that infect the host more than once. The most important host cells involved in innate and adaptive immunity are indicated in Tables 10.1 and 10.2. At the end of this paragraph, we also describe briefly some major virulence factors of pathogens, as well as some examples of mechanism that pathogens use to evade host immunity.

Innate Immunity

Before innate immune mechanisms are activated, pathogens encounter several barriers that protect host organisms from infection. Mechanical factors such as the skin

and epithelial layers act as a first line of defense. The shedding of the outermost layers of barrier tissues removes adhered microorganisms. The flushing action of saliva and tears protect mouth and eyes. Moreover, movement of cilia keeps the respiratory tract free of microorganisms, and the mucus layer of the gastrointestinal tract traps microorganisms and prevents direct contact with epithelial cells. Chemical factors, such as lysozymes and phospholipases are present in saliva, nasal secretions, and tears, destabilize bacterial membranes and break down bacterial cell walls. Gastric secretions and sweat have a low pH to prevent bacterial growth. Furthermore, in the gastrointestinal and respiratory tract, defensin proteins are present, which contain antimicrobial activity. In the lungs, surfactant proteins can bind to pathogens (opsonize), which makes them more susceptible to the action of phagocytes. Biological factors, such as the microbiota in the gut and skin, also contribute to host defense. They secrete toxins or compete for nutrients, thereby preventing adhesion, colonization, and growth of invading pathogens. Important aspects of the innate immune system are: complement activation, recognition of pathogenic invaders, inflammation and recruitment of leukocytes to infection sites, and activation of the adaptive immune system.

Complement Activation

The complement system consists of approximately 25 small inactive proteins that circulate in the blood. After activation, protein cleavage events occur, leading to the release of cytokines and to the initiation of further protein cleavage events. This results in a massive activation of membrane bound complexes that kill invading cells. The complement system is part of the innate immune system since it does not change in time. However, the complement system can be brought into action by products (antibodies) of the adaptive immune system via the so-called classical complement pathway. The alternative complement pathway and the mannose-binding lectin pathway are activated without the presence of antibodies (nonspecific immune response). Although the three pathways contain different receptors, in all three pathways the key protein C3 is proteolytic activated. C3 has a unique internal thioester bond, which is exposed to the molecular surface upon activation and which covalently binds to invading microorganisms (Law et al., 1980). Activation of the complement system results in opsonization, chemotaxis, and lysis. Through opsonization, phagocytosis of the invading pathogens is enhanced. Chemotaxis contributes to attracting different immunological cells such as macrophages and neutrophils. Direct lysis and bursting of the invading pathogen by disrupting the membrane is a major protective function of the complement system.

Recognition of Pathogenic Invaders

At the molecular level, various host-encoded receptors are in place to discriminate between self and nonself (foreign) components. These receptors are the primary sensors of the host to detect foreign (microbial) products. Practically all cells of the innate immune system, such as macrophages, DCs, neutrophils, natural killer cells, and mast cells, express such receptors. And also epithelial cells use them to constantly screen their environment for the presence of unwanted pathogens. A special branch

of these receptors is the pattern recognition receptors (PRRs), which are usually located on the surface of a number of different host cells. These PRRs can recognize pathogen-associated molecular patterns (PAMPs) (Janeway, 1989; Medzhitov, 2001). PAMPs are conserved molecular patterns and are uniquely produced by microbes. PAMPs are often invariant between microorganisms of a given class and are essential for their survival. Because of these typical features they are very suitable targets for innate immune recognition (Tapping, 2009). Examples of PRRs are the Toll-like receptors (TLRs), C-type lectin receptors (CLRs), and the intracellular-located nucleotide-binding oligomerization domain (NOD) family of proteins. TLRs recognize an array of PAMPs, e.g., lipopolysaccharide (LPS), peptidoglycan, or lipoeichoic acids, lipoproteins, CpG DNA, dsRNA, zymosan, and flagellin, which are found in bacteria, viruses, or fungi. The TLR family size differs in species, humans have approximately 12 TLRs, whereas sea urchin has up to 222 (Hibino et al., 2006). CLRs bind specific carbohydrate structures on both self-antigens and cell wall components of pathogens in a calcium-dependent manner (Cambi et al., 2005). After recognition of PAMPs by PRRs, intracellular signaling pathways are activated leading to activation of transcription factors such as those belonging to the NF-κB family which, in turn, are able to induce transcriptional programs of the innate immune system, for example, the synthesis and secretion of inflammatory cytokines (Kumar et al., 2009). Furthermore, CLR-ligand binding leads to antigen internalization and degradation, followed by antigen processing and presentation to DCs (Beg, 2002; Figdor et al., 2002; Dam and Brewer, 2010).

Inflammation and Leukocytes

Inflammation is initiated by cells, such as macrophages, DCs, and mast cells, which are present in nearly all tissues. After infection, these cells become activated and secrete eicosanoids and cytokines. These molecules cause an increased blood flow into infected areas and induce the migration of additional white blood cells (leukocytes) to these sites. The cytokines include chemokines that induce chemotaxis and interferons (IFNs) that induce antiviral effects, such as the inhibition of the protein synthesis machinery. The inflammatory mediators not only recruit immune cells to the site of infection, but also promote the healing of affected tissues. Therefore, leukocytes can move unrestricted through the body for surveillance and protection. Most innate leukocytes are not able to replicate, but are the progeny of multipotent hematopoietic stem cells. A schematic overview of the different leukocytes and their functions is presented in Table 10.1.

Activation of Adaptive Immunity

DCs are central in the induction and regulation of immune responses (Banchereau et al., 2003). In fact, DCs are thought to be the link between innate and adaptive immunity, because of their ability to induce adaptive immunity. Their main function is to process antigens and present it on the surface to other cells of the immune system. They are mainly present in tissues that are in close contact with the external environment. After activation of mature DCs, they migrate to lymphoid nodes where they interact with T cells and B cells of the adaptive immune system. There they

Table 10.1 Leukocytes and their functions

Class	Function	Reference
Mast cell	Releasing granules containing histamine and heparin (allergies)	(Marshall and Jawdat, 2004)
Monocyte	Phagocytosis and activation of immune system, antigen presentation	(Serbina et al., 2008)
Macrophage	Phagocytosis and activation of immune system, antigen presentation	(Serbina et al., 2008
Dendritic cell (DC)	Antigen uptake and presentation (stimulating of naïve T cells)	(Howard et al., 2004)
Neutrophil	Phagocytosis and activation of immune system	(Serbina et al., 2008)
Basophil	Killing parasites covered with antibody and allergies	(Schroeder, 2009)
Eosinophil	Killing parasites covered with antibody and allergies	(Cadman and Lawrence, 2010)
Natural killer cell	Tumor rejection and eliminate virus-infected cells	(Hamerman et al., 2005)
$\gamma\delta$ T cells	Immunoregulation and cytotoxicity	(Hayday, 2000)

orchestrate the development of an appropriate adaptive immune response (Finkelman et al., 1996; Heath and Carbone 2001). Immature DCs gather self-antigens from apoptotic cells and induce tolerance (Huang et al., 2000; Steinman et al., 2000). DCs are constantly in communication with other cells in their neighborhood, either by direct cell-to-cell contact or by cytokines. DCs produce a wide variety of cytokines, including interleukins (ILs), IFNs, growth factors, and chemokines. Different combinations of these cytokines can lead to particular differentiations of T cells. For example, high IL4 expression leads to a Th2 differentiation, whereas high IL12 and IFNG expression leads to Th1 differentiation (Mosmann and Sad, 1996). This results in the priming and activation of the adaptive immune system for attack against the presented antigens. There are significant differences in the cytokines produced by the various DCs. For example, the lymphoid DCs have also the ability to produce huge amounts of type-1 IFNs, which recruit more activated macrophage to allow phagocytosis.

Adaptive Immunity

The cells of the adaptive immune system are called lymphocytes, a specific type of leukocytes, and constitute for the major part of B- and T-cells. They circulate through the body within the lymphatic system and the blood. Approximately 20–40% of the white blood cells are lymphocytes. Adaptive immunity is based on the recognition and binding of specific determinants (epitopes) of nonself antigens by T-cell receptors (TCRs) and immunoglobulins (Igs). TCRs are expressed on the surface of T-cells, whereas Igs are produced by B-cells, either as cell membrane bound receptors or as secreted molecules (antibodies). At their N-termini these receptors have loops

Table 10.2 Lymphocytes and their functions.

Class	Function	Reference
Cytotoxic T cell	Eliminate virus-infected and tumor cells (allograft)	(Williams and Bevan, 2007)
B cell	Secretion of antibodies and generating memory cells	(McHeyzer-Williams and McHeyzer-Williams , 2005)
Th1	Maximizes the killing efficacy of the macrophages, proliferation of cytotoxic T cells and produces opsonizing antibodies (bacteria and some viruses)	(Romagnani, 2006)
Th2	Stimulate B-cell proliferation, induce class switching and increase neutralizing antibody production (helminthes)	(Romagnani, 2006)
Th17	Antimicrobial immunity	(Romagnani, 2006)
Regulatory T cells	Homeostasis of immune system and self-antigen tolerance	(Romagnani, 2006)

of peptides that fold into a pocket for epitope binding, the primary sequence of the peptide loops provide a specific binding structure to an epitope (Sadofsky, 2001). After the primary adaptive immune response, immunological memory is created leading to rapid response against pathogens that infect the host more than once. A schematic overview of the major cell types of the adaptive immune response and their functions is presented in Table 10.2.

Generation of Diversity

TCRs and Igs must be able to distinguish between many different potential antigens and therefore these molecules must be produced in a very large variety of configurations. Indeed, animals and humans are capable to produce more than 10^{14} different Igs or TCRs. To generate such variation, T- and B-cells use site-specific DNA recombination mechanism. This mechanism involves rearranging and assembling variable (V), diversity (D), and joining (J) gene segments and is called V(D)J recombination (Hozumi and Tonegawa, 1976; Brack et al., 1978). V(D)J recombination occurs between Ig and TCR gene segments that are flanked by conserved recombination signal (RS) sequences. A palindromic heptamer and an AT-rich nonamer make up RSs and these are adjacent to each of the coding segments. The heptamers and nonamers can be separated by either 12 or 23 base pair spacers (Lieber, 1991). V(D)J recombination starts when DNA double-strand breaks between the V, D, J segments and flanking RSs are introduced. The latter process is induced by recombination of activating gene (RAG)-1 and RAG-2, which are exclusively expressed in lymphocytic cells (Schatz et al., 1989; Oettinger et al., 1990). Next, RS ends are accurately joined, while coding ends are modified by potentially adding or removing nucleotides (Bassing et al., 2002). Therefore, this variety of antigen receptors is not germ line encoded, but somatically generated through a process known as clonal selection. Through these mechanisms

the adaptive immune system contains an immense diversity of TCRs and Igs. As a consequence, they are capable of initiating an immune response to an enormous diversity of foreign antigens.

Virulence Factors

Virulence factors are molecules expressed by bacteria, viruses, fungi, and protozoa that enable them to survive on or within a host. Frequently, these virulence factors cause disease in the host as they interfere with certain functions of (the) host (cells). Virulence factors are produced by pathogens to achieve specific properties for survival. These includes adherence to host components and colonization, inhibition and evasion of the host immune system, entry into host cells, and the acquisition of the necessary nutrients. Pathogens have developed a remarkable variety of strategies and molecular machinery to accomplish survival in hosts (Aldridge et al., 2005; Jenner and Young, 2005). In many cases, the pathogens interfere with the common transcriptional program of the innate immune system. In other cases, pathogens exploit vital cellular processes such as modulation of cell cycle progression, modulation of the actin cytoskeleton, or modulation of secretory pathways. An extensive description of these strategies goes beyond the scope of this review. Only some examples will be briefly described here, just to demonstrate the diversity in strategies. (i) After the infection of DCs by the HIV-1, the virus interferes with the host transcriptional machinery in such a way that only an interferon-stimulated-gene cluster is induced while other immune signaling gene clusters are inhibited (Izmailova et al., 2003). This results in the recruitment of T-cells and macrophages, the main target for HIV replication, while avoiding the activation of immune responses. (ii) *Streptococcus pyogenes* produces a surface component that recognizes adhesive matrix molecules of the host. This component mediates attachment to epithelial layers and commit to infection (Fisher et al., 2008). (iii) Gram-negative bacteria possess the so-called type III secretion system (TTSS), which acts as supramolecular syringes and inject bacterial proteins directly into the cytoplasm of host cells. Many different TTSS translocated proteins have been identified in various bacterial species, including *Escherichia coli, Salmonella,* and *Yersinia*. The injected bacterial proteins interfere with normal cellular functions, usually at the transcriptional level, in order to convert the hostile environment of the pathogen into a beneficial one. A multitude of cellular effects of TTSS-based injections have been described in literature, including disruption of adhesion complexes, disruption of phagocytosis, inhibition of cytokine production, and induction of apoptosis (Nomura and He, 2005; Ly and Casanova, 2007). (iv) Evading the immune system can also be a property of specific surface components of microorganisms, such as capsules. Such components can inhibit phagocytosis (Hyams et al., 2010) and protect the pathogen outside the host (Roberts, 1996). (v) A major group of virulence factors are bacterial toxins, consisting of two groups: endotoxins, which are an intrinsic part of the pathogens (such as capsule) and secreted exotoxins. Exotoxins can act as superantigens (Kotb, 1995), damage the host cell membrane (Bhakdi et al., 1985), or act in the cytoplasm of the host cell (Iglewski and Kabat, 1975). They cause a variety of effects, such as hemolysis and necrosis of host cells or inhibiting protein synthesis (Dinges et al., 2000). Endotoxins are primarily found on gram-negative bacteria. For example,

LPS is present in gram-negative bacteria and causes a range of detrimental effects in the host, including septic shock (Galanos and Freudenberg, 1993).

Transmission

Pathogens are transmitted from one source to another. Pathogens can be acquired by direct contact with body fluids, objects, aerosolized droplets, or by ingesting of contaminated food. Transmission of pathogens may also occur via a vector, which could be mechanical or biological. Biological vectors deliver pathogens to new hosts in an active manner, for example, by a byte, mechanical vectors transmit pathogens in a passive manner, for example, by adherence to the legs of a fly. Transmission parameters are influenced by a number of different factors, including environmental factors (temperature, humidity), pathogen load, host density, genetic background of populations, genetic factors of pathogens, host immune competence, vaccination, nutrition, etc. There is a complex relationship between transmission and virulence characteristics. This relationship affects the longer term coevolution of hosts and pathogens. In general, one sees more severe clinical signs and higher death rates during the first wave of new emerging disease.

Complexity and Scales

Space

The final goal of host–pathogen Systems Biology is to understand the physiology and infectious disease development from the level of molecules, to cellular networks, host cells, viruses, bacterial pathogens, tissues, organs, organisms, up to whole populations. Since each biological level is already complex in its own, the complexity of the whole system is enormous. Individual components are continually created, destroyed, and circulated throughout the body, which increases the temporal and spatial diversity even further. The whole systems spans about ten orders of magnitudes in space scale, ranging from the size of a molecule to the size of a population. The temporal scale spans from microseconds, for biochemical reactions, to the life span of organisms and population in years. For the sake of simplicity, here we define and describe only five biological levels from the perspective of the host, namely the molecular, cellular, tissue, organism, and population level.

Molecular

The complexity of the host on the molecular level is mostly explained by the quantitative and qualitative variability of receptor molecules, the molecules involved in signaling cascades, and the variation in the effector molecules. Mathematical models could shed light on, for example, the efficiency of PAMP binding to PRRs or describe the kinetics of the interactions between epitopes and TCRs.

Cellular

The complexity at the cellular level has been investigated by a plethora of -omics technologies. This provided insight into the complex signaling pathways that cells use to adapt their transcriptional program in response to signals, picked up at the cell surface or in the cytoplasm. At this time, only the global frameworks of a number of these signaling pathways have been established. Models could help to understand how these signaling pathways interact with each other in different cell types and how they influence cellular transcriptional regulation and affect the resulting properties of cells.

Tissue

The complexity at the tissue level is nicely demonstrated by the complexity and dynamics of the immune system. As described before, the immune system is highly complicated and appears to be precisely tuned to detecting and eliminating infections. It is made up of numerous different types of cells that communicate with each other and migrate through the body and have different jobs to do. Immune responses involve the collective and coordinated response of approximately 10^{12} cells. Mathematical modeling of (parts of) the immune system may predict how various interactions together result in particular immunological phenomena. Furthermore, the effect of potential drugs can be tested *in silico*.

Organism

The complexity at the level of an organism is illustrated by the ability of pathogens evading the immune system and migrating into different tissues. For example, *Mycobacterium tuberculosis* can hide themselves in cells of the immune system (Pieters and Gatfield, 2002) and subsequently induce systemic infection. They can reside in multiple tissues and multiple organs. Communication between these tissues and organs exist, as well as communication toward the immune system. Since each organ contains tissue specific cells and other abiotic characteristics, pathogens are associated to a certain biological environment (niche). Models describing host–pathogen interactions at the organism level are useful for the identification of the critical bacterial factors responsible for successful infection and the main components of the immune response for successful defense. The first Boolean-based model that covers multiple aspects of host–pathogen interactions have been published now and some examples will be described in section "Interaction models."

Population

The complexity of host–pathogen interaction at the population level is mostly captured by ecological and epidemiological models. These models are top-down approaches and may predict antigenic shifts, spatial-temporal effects (of vaccination), and transmission characteristics of infectious disease outbreaks. Moreover, genetics also contributes to the complexity at the population level, each individual has its genetic background and will, therefore, react differently to infection. The SIR model, for example, computes

the theoretical number of infected individuals with a disease in a closed population over time. The model involves equations relating to the number of ***S***usceptible animals, the number of ***I***nfected animals, and number of ***R***ecovered animals. The SIR-derived SEIR model includes exposed animals that do not have (yet) the disease (Li et al., 1999) and the SIS model takes into account that recovered animals can be infected again (Hethcote and Van Den Driessche, 1995). Also, combinations of susceptible, infected, recovered, exposed, carrier state animals and animals with maternal immunity are possible (Hethcote, 2000).

Time

The time course of an infection not only differs from pathogen to pathogen but also differs between hosts. The time elapsed between exposures to a pathogen and when clinical signs first appear may be as short as minutes to as long as several years. The same is true for other phases of the infection process. Adherence is one of the first events that connect a pathogen to structures of the body. After adherence, pathogens can migrate and spread to other parts of the body and start to multiply to cause an infection. After multiplication begins, pathogens can continue to multiply and resist the defense mechanisms of the host, or a state of balance is achieved that causes a chronic infection, or a carrier state is achieved without disease symptoms, or the body is able to destroy and eliminate the invading pathogen. Usually, innate immune responses will show-up in an attempt to eliminate the invading pathogens. However, immune evasion mechanisms allow several pathogens to resist innate defense mechanisms. In such cases, innate immunity alone is not sufficient and an adaptive immune response needs to be build-up. The adaptive immune response is time delayed because extended proliferation of B and/or T cells is necessary to produce an effective amount of specific immune cells and/or antibodies. Usually, this takes several days. At the same time, memory cells are generated that will be present for years. These memory cells allow a much quicker specific immune response after follow-up infections with the same pathogen or pathogens bearing homologous antigens.

Mathematical Models

Models can represent certain biological systems or phenomena and are by definition imperfect. However, models can help to understand (parts of) systems by deduction to only the most important components. In Systems Biology, mathematical modeling is required to structure and represent the relationships between components of the systems at various different levels. As indicated before, the information of the various components of biological systems is rapidly increasing by application of -omics technologies. With the help of mathematical models and computational tools, these data can be used to improve our understanding of the functioning of biological systems. By *in silico* simulations and experimental validations, hypotheses obtained with these models can be tested, accepted or rejected, and adjusted to new or improved hypotheses. This iterative cycle is a typical aspect of mathematical modeling in Systems Biology.

Mathematical Frameworks

Mathematical models are a collection of equations and variables, which describe a system. With these equations, simulations in time can be run and predictions can be made. The variables of models can have diverse sets of values, such as real numbers, integer numbers, Boolean values, or strings. Equations describe the relations between variables and thus represent properties of the system. Because of the large variety in variables and equations, models can be classified into different groups: static versus dynamic models, deterministic versus stochastic (probabilistic) models, and linear versus nonlinear models. Static models do not account for time scales, in contrast to dynamic models. In deterministic models, the outcome is precisely determined by known relations among states and events. Contrary, in stochastic models, all variables are described by probability distributions. In deterministic models, a specific input always produces an identical output, this is not the case in stochastic models due to randomness. Linearity or nonlinearity depends on the context, but in general when all operators exhibit linearity, the model is defined as linear. Since different models answer different questions, multiple mathematical models are in use and operate on different scales. Here, we describe briefly a selection of models used to represent aspects of host–pathogen interaction.

Boolean functions are used to describe gene regulatory networks (see also Chapter 1). Moreover, Boolean functions are either *true* or *false*, often depicted by numerical 0 or 1. By using logical operators like conjunction (AND), disjunction (OR), and complement or negation (NOT), different scenarios for gene regulation can be simulated (Table 10.3). However, more operators are known such as NOR, XOR, XNOR, NAND, TRUE or FALSE, each having a different output on the regulation of genes.

Although Boolean networks have a deterministic nature, Garg et al. (2009) have managed to model stochasticity and robustness in gene regulatory networks. Boolean functions are also used to model the dynamic interactions between pathogens with the host immune system, for example, with *Bordetella bronchiseptica or Bordetella pertussis* (Thakar et al., 2007). Both gram-negative pathogens are closely related but cause different diseases in their host (Mattoo et al., 2001; Parkhill et al., 2003). The model of Thakar et al. encompasses immunological cells, cytokines, antibodies, and antigens, represented in 18 nodes that are common to both pathogenic species (Thakar et al., 2007). After loading in BooleanNet for simulation studies (Albert et al., 2008), three different phases could be identified for *Bordetella* infections and the

Table 10.3 Different scenarios for gene regulation of genes A and B.

Input		Output			
A	B	A AND B	A OR B	NOT B	NOT A
0	0	0	0	1	1
0	1	0	1	0	1
1	0	0	1	1	0
1	1	1	1	0	0

predicted differences in the clearance between the two species could be confirmed by experimental data. The model was also used to simulate secondary infections. Again, this resulted in differences between the two species. As expected, primary Bordetella infections had a predicted effect on the progression of a secondary infection.

ODEs (ordinary differential equations) are often used to describe temporal dynamic events in systems. ODEs consist of one independent variable, here time, and one or more derivatives to the independent variable. Furthermore, these different variables can be expressed by one or more equations. A relatively simple model is the Lotka–Volterra model (reviewed by Wangersky, 1978), also known as predator–prey model, which models the interactions between two species. The following equations describe this relation between preys and predators:

$$\frac{dN}{dt} = aN - bNP$$

$$\frac{dP}{dt} = cNP - dP$$

where N is the number of preys and P the number of predators. Furthermore, dN/dt and dP/dt denote the growth of these populations in time t. Interactions between preys and predators are described with the following parameters, a, b, c, and d. Parameter a is the natural growth rate of preys, b is the predation rate coefficient, c is the efficiency of converting preys into predators, and d is natural death of predators. These ODE models and more enhanced models are already applied in the field of microbial communities competing for food (Kaunzinger and Morin, 1998) and in host–pathogen interactions (Hethcote, 2000; Fenton and Perkins, 2010). A time plot of simulation with preys and predators is given in Figure 10.1.

Cellular automata models consist of many identical uncomplicated cells, which together are capable of displaying complex behavior (Wolfram, 1984). These models comprise a grid of cells, and each cell has a limited number of states, like "on" and "off" or "0," "1," "2," and "3." The grid is also limited to a certain set of dimensions. A set of cells is called a neighborhood, each cell has its own neighborhood. Before simulations are run ($t = 0$), each cell has an assigned state. When simulation starts, t is incremented by 1 and the state of all cells are updated by fixed rules. Cellular automata models are applied in biology, e.g., insect population dynamics (Hassell et al., 1991), HIV dynamics (dos Santos and Coutinho, 2001), and cellular dynamics in the immune system (Celada and Seiden, 1992; Seiden and Celada, 1992). The "Game of Life" is an example of simple rules leading to complex behavior (Gardner, 1970). Three requirements have to be met: (i) the seeding pattern must not have simple proof that the population can grow limitless, (ii) seeding patterns should apparently grow limitless, and (iii) the seeding patterns change considerably for a period of time, before ending in three possible situations: disappear completely, going to a stable configuration, or entering an oscillating phase (repeating an endless cycle of two or more periods) (Gardner, 1970). It is an infinite grid of square cells, where each cell can be either alive ("on") or dead ("off"). All cells interact with their eight neighbors, which are the horizontal, vertical, and diagonal adjacent cells. Then, for the time-step the following transition occurs (simultaneously): (i) live cells with fewer than two live

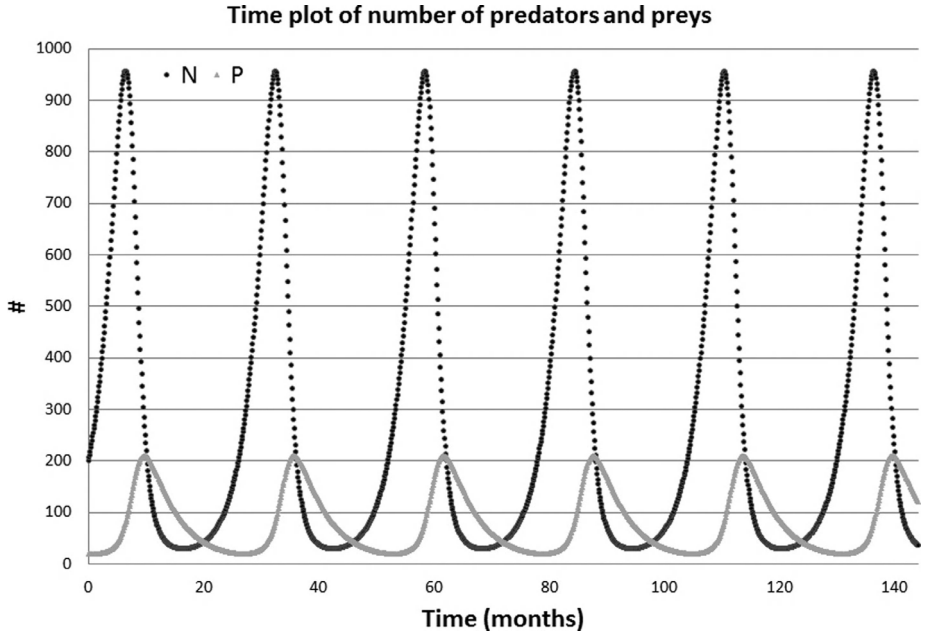


Figure 10.1 Time plot of simulation with preys and predators. The y-axis depicts the number of prey (black circles) and predator (grey triangles), whereas the x-axis depicts time in months. The following parameter values were used, a = 0.4, b = 0.005, c = 0.00075, d = 0.2, initial N (N_0) = 200, initial P (P_0) = 20, for the prey-predator model.

neighbors die, (ii) live cells with two or three live neighbors stay alive, (iii) live cells with more than three live neighbors die, and (iv) dead cells with exactly three live neighbors become alive. A few simple examples are depicted in Figure 10.2.

Partial differential equations (PDEs) are able to describe multiscale systems. When a system consist of different types of components, it is sometimes necessary to distinguish between these components by the use of PDEs. Almost all PDEs have

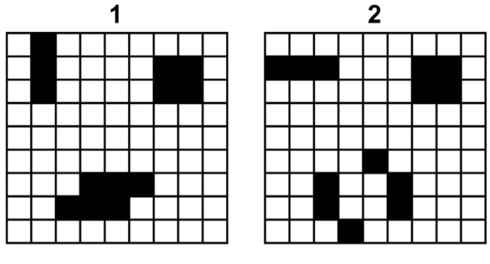


Figure 10.2 Game of Life. With only one simple rule "*A live cell with two live neighbors, or any cell with three live neighbors, is alive at the next step*" (Gardner, 1970), different patterns can be generated. The system is seeded state "1" (left) and the subsequent time point (right) the new state "2" will be achieved. The following time point the system will go back to its initial state, and thus a loop between state "1" and "2" is established (upper-left; blinker, upper-right; block, bottom; toad).

infinitely many solutions, thus, solving PDEs is often finding properties of the solutions. A relatively simple PDE to solve is $\mu_t + \mu_x = 0$. The general solutions is $\mu(x, t) = \varphi(x - t)$ for any function φ of one variable. However, most PDEs are very difficult or impossible to solve, like indefinite integrals. For example, PDEs describing different cell types in a tissue, such as macrophages and tumor cells (Owen and Sherratt, 1999), or morphogen gradient concentrations and axis development (Baker et al., 2008). In section "Interaction models," a *M. tuberculosis* infection model will be described. In this model, spatial movement and temporal development dynamics of cells are described with PDEs (Gammack et al., 2004). The model consisted of coupled reaction–diffusion–advection equations, describing dynamics of resting and infected macrophages, intra- and extracellular bacteria, as well as important chemokines affecting granuloma growth.

Computational Tools

Because of the plethora of mathematical models describing biological systems, also a variety of software and tools are (freely) available. For example, tools for creating and editing models, time-series simulation, and many other analyses, but some tools also interact with model repositories and are able converting models in different formats. These tools cover different mathematical frameworks, such as ODEs, PDEs, discrete stochastic simulation, discrete events, logical models, etc. Moreover, consensus languages for biological models were generated, such as Systems Biology graphical notation (SBGN) (Le Novere et al., 2009) and Systems Biology markup language (SBML) (Hucka et al., 2003). The SBGN project aims to standardize graphical notation, for example, for signaling pathways and other biochemical processes occurring in cells. These graphical notations can be expressed in mathematical formulas and written in SMBL, which is machine readable. These initiatives are necessary, because sharing and understanding of data and models will be easier. Table 10.4 provides several tools covering most of the available options and frameworks. They were extracted from a more comprehensive overview of available tools in the field of Systems Biology (http://sbml.org/SBML_Software_Guide/SBML_Software_Matrix).

Interaction Models

The outcome of an infection with pathogens is dependent on the interplay between pathogen-related, host-related, and environmental factors (Figure 10.3). Each of these individual players can be regarded to be build up of subsystems or modules, as extendedly described for the host in the previous paragraphs. As a consequence, modules interact with each other within and between hosts, pathogens, and environment, respectively. In addition, during the time course of an infection, these interactions may change. As indicated before, the development of a comprehensive model of a "host–pathogen system" is currently not feasible. However, several models have been described in literatures that describe the behavior of well-characterized subsystems or modules of host–pathogen interaction systems (Bumann, 2009). In this section,

Table 10.4 A selection of software/tools for modeling[a].

Software/ Tool	Capabilities					Frameworks							SBML[b]	
	Creating/ editing models	Simulation (time-series)	Analyses (e.g., flux balance)	Database (inventory of models)	Utility (converting models)	Ordinary differential equations	Differential-algebraic equations	Partial differential equations	Discrete stochastic simulation	Discrete events	Logical (e.g., Boolean)	Other	Import	Export
ByoDyn	•	•	•	•	•	•	•		•	•		•	•	
CellDesigner	•	•			•	•	•		•	•	•		•	•
COPASI	•	•	•		•	•			•				•	•
Cytoscape	•				•								•	
JSim	•	•	•		•	•	•	•		•	•	•	•	
PottersWheel	•	•	•		•	•	•		•	•			•	•
ProMoT	•		•	•		•	•			•	•		•	•
PySCeS	•	•	•		•	•				•			•	•
SBToolbox2	•	•	•			•			•				•	•
SBW	•	•	•		•	•	•		•				•	•
Vcell	•	•		•		•	•	•	•		•	•	•	•

[a] Extracted from SBML site http://sbml.org/SBML_Software_Guide/SBML_Software_Matrix.
[b] SBML, Systems Biology markup language.

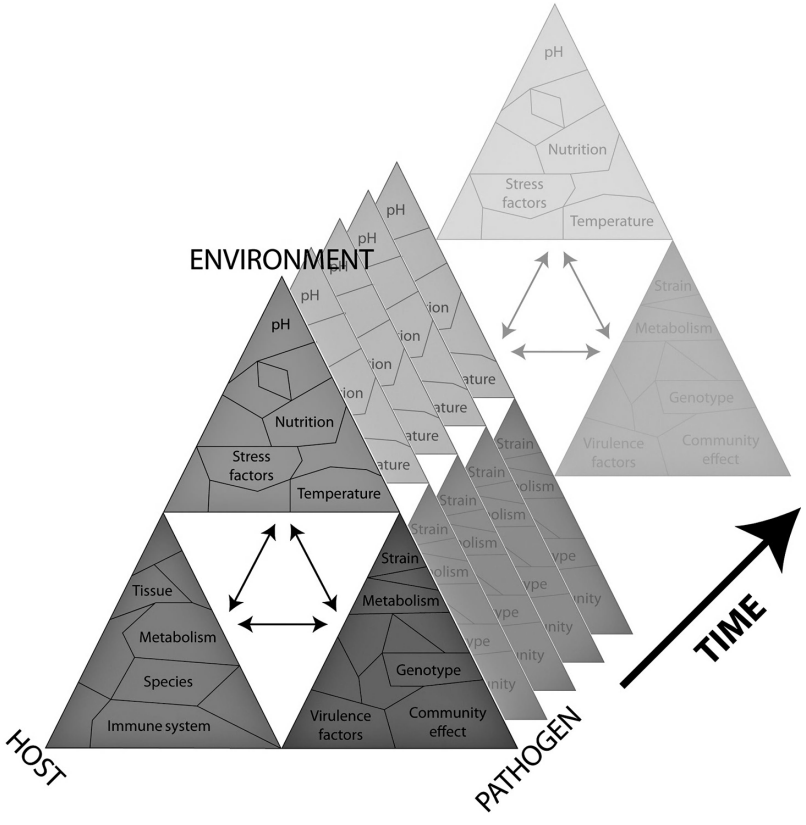


Figure 10.3 Host–pathogen interactions in conjunction with environment. The three key players, environment, host, and pathogen, are divided into various modules. These modules can be of different scales, i.e., genotype or community effect, and interaction can occur between different modules, however, modules of different key players can also interact. To make it even more complex, over time module characteristics can change and therefore possibly also their interaction.

several examples are provided about how System Biology models are generated and used to understand important aspects of host–pathogen interactions.

Pathogens

Metabolomic networks of prokaryotic microorganisms belong to the best-understood biological systems, since they have been studied for many years at the level of metabolic fluxes, enzymes, and reaction kinetics (Feist et al., 2007; Ishii et al., 2007). Therefore, pathogen metabolism is one of the most suitable subsystem for modeling. Indeed, *in silico* models have been generated that represent the quantitative interactions of more than thousand *E. coli* enzymes and metabolites (Feist et al., 2007). This model is able to predict the metabolic behavior of more than a thousand *E. coli* mutants

in vitro. Since the interaction of a pathogen with a particular host dramatically affects the metabolism of pathogens, sometimes with a direct link to virulence (Milenbachs et al., 1997; Exley et al., 2005; Bruggemann et al., 2006), similar approaches are used now to build high-resolution reconstructions of microbial metabolism of pathogens in infected cells and/or tissues. The idea behind these modeling efforts is that the *in vivo* metabolism of pathogens provides important information to understand their lifestyle within the host as well as to understand nutrient-based host defense mechanisms. *In silico* simulations with these models may generate new hypotheses about *in vivo* metabolism of pathogens and might be of value for the rational design of novel antibiotics or attenuated live vaccines.

Raghunathan et al. (2009), for example, recently developed a model describing the metabolic capacity of *Salmonella typhimurium* during host–pathogen interaction. The model was reconstructed by the integration of existing genomic, proteomic, and phenotypic data. The authors developed a genome-scale metabolic network using more than thousand genes, involved in many different metabolic and transport reactions. The model was used to study growth parameters under several *in vivo* and *in vitro* conditions using flux balance analysis and *in silico* gene essentiality analysis. Model predictions and experimental data showed an overlap of more than 80% for growth and virulence phenotypes. In an extended model, gene expression data, obtained from *Salmonella*-infected macrophages, were used to identify sets of metabolic pathways required for replication inside host cells. In this case, model development started with an annotated genome, and through several iterative steps, it moved to a genome-wide metabolic network that could be used to predict *in silico* the functional properties of the "*Salmonella*-into-host-cell" system. The model pointed toward essential nodes of the metabolic network and toward relationship between nutritional factors and microbial genetic factors and provided insight into their functional contribution toward systems behavior: growth versus killing of *Salmonella*. The model could also be used as a toolbox to visualize the functional properties as represented by large-scale -omics datasets. With this model, the authors expect to arrive at a conserved set of metabolic reactions that intertwine with those of the host cell. Such metabolic reactions might be novel targets for therapeutic intervention.

Receptor-Ligand Kinetics

The kinetics of receptor–ligand interactions codetermines the speed and efficiency of virulence and host defense mechanisms. Modeling this "subsystem" requires first the definition of the set of components involved, as well as their interactions, often on the basis of empiric observations. Subsequently, parameters have to be inferred that quantify the intensity of these interactions and cellular concentrations of the components. In a further step, a mathematical formalism or simulation method has to be selected (Goldstein et al., 2004). The simplest model describing the kinetics of receptor–ligand interaction is Equation 10.1 (Figure 10.4), where L stand for ligand, R for receptor, k+1 for binding association, k−1 for binding disassociation, and LR for ligand–receptor complex. Binding studies mostly measure specific binding, in other words, measuring the complex. After several steps

Equation 10.1

$$L + R \underset{k-1}{\overset{k+1}{\rightleftharpoons}} LR$$

Equation 10.2

$$LR = \frac{[R]_{tot} \cdot [L]}{K_d + [L]}$$

Figure 10.4 Equations of receptor–ligand interaction kinetics. Equation 10.1 describes the kinetics of a receptor (R) and ligand (L) interaction. Often, the complex (LR) is measured and therefore these kinetics can also be described as Equation 10.2 where $[R]_{tot}$ is the total number of nonbound receptors and Kd the equilibrium dissociation constant.

the equation can be rearranged with the complex on the left and similar to enzyme kinetics function on the right (Equation 10.2, Figure 10.4), where LR is the ligand–receptor complex, R_{tot} is the total number of nonbound receptors, L is the ligand concentration, and K_d is the equilibrium dissociation constant ($K_d = k+1/k-1$). This formula is similar to calculating enzyme activity, the well-known Michaelis–Menten kinetics.

The model sketched above is a good description of the core process, however, in nature multiple ligands could be present competing for the same receptor and sometimes ligands can trigger multiple responses by associating and dissociating from receptors. The latter is known as serial triggering and was proposed to explain how antigen presenting cells, having low densities of peptide–MHC complexes, could trigger a whole set of TCRs over a short period of time (Valitutti et al., 1995). Wofsy et al. developed a mathematical model to test whether this serial engagement of peptide-MHC really occurs (Wofsy et al., 2001). With this model, it was calculated that approximately 50–200 receptors could be triggered by only one peptide–MHC complex (Valitutti et al., 1995; Wofsy et al., 2001). Furthermore, each peptide–MHC complex is being subjected to serial engagement and is an increasing function of the dissociation rate constant.

Next to the serial engagement concept, another concept was put forward, namely that of kinetic proofreading (McKeithan, 1995). Kinetic proofreading is used by TCRs to discriminate between ligands during the ligand–receptor bond (dwell time). The TCR undergoes modifications during this dwell time, but when the ligand dissociates then the modifications are reversed (McKeithan, 1995). Thus, the duration of the dwell time determines whether activation signals are induced or not. Later, "kinetic discrimination" was proposed as a more realistic model that also accounted for TCR ligand sensitivity and the resulting biological response (positive or negative). Compared to kinetic proofreading where many intermediate steps are needed (Rabinowitz et al., 1996), here agonist and antagonist peptides are distinguished in one single step. These examples demonstrate that complex ligand–receptor kinetics can be described in relatively simple equations. The models are valuable to study the characteristics and dynamics receptor–ligand interactions.

Intracellular Signaling

Cellular host–pathogen interaction models are frequently based on large-scale -omics data. They comprise interaction networks that represent the cellular effects of the attachment, recognition, or invasion of by pathogens. In the case of viral infection, the cell response is usually initiated by a few viral components in an attempt to transform the cellular machinery into an environment suitable for virus replication. On the basis of extended viral–host interaction studies, usually under *in vitro* conditions, several mechanistic models have been developed describing specific aspects of the biological events in infected cell. A classical example is the recognition of viral components by TLRs and the induction of Toll-like signaling pathways that play a crucial role in the induction of antiviral and inflammatory immune responses (Wong et al., 2009) (Figure 10.5).

Other models describe the binding of viral proteins to host molecules and that result into a change of cellular metabolism, like the inhibition of host mRNA synthesis or the inhibition of cytokine synthesis. In fact, on the basis of long-lasting intensive research in the field of molecular virology, a prototype of the human virtual infected cell has been developed (Navratil et al., 2010). This model contains more than 100 viral "infectomes," including the viral pathogens HCV, HBV, HIV, HHV, HPV, and has resulted into several hypotheses for the development of novel antiviral strategies based on interventions of cellular functions. The first attempts to use such system approaches for the development of new vaccines show promising results, particularly in the identification of molecular signatures or biomarkers that are induced immediately after vaccination and that correlate with protection after experimental or natural infections (Pulendran, 2009). Furthermore, systems approaches are also applied now to understand the functional properties of vaccine adjuvants (Mosca et al., 2008).

A nice demonstration of the identification of key cellular molecules, which play decisive roles in the cellular functions that arise after interaction with pathogens, is given by the recent work of Amit et al., in mammalian DCs (Amit et al., 2009). These cells play a central role in the induction of the adaptive immune system through antigen presentation after identification of invading pathogens. First, the authors measured gene expression profiles of the DCs after interaction with five different pathogen (bacterial and viral) encoded ligands for a number of TLRs at nine different time points after stimulating. These profiles were used to identify 144 putative regulator genes involved in driving DC responses. Systematic perturbation of the regulator gene expression, using RNAi knock-down technology, followed by large-scale gene expression measurements was subsequently used to associate the putative regulators to their targets. From this, a picture emerged that showed the complexity of TLR-mediated sensing and signaling in DCs, with several key regulators being connected to different targets in feed-forward and feed-back loops (see Figure 1.7, Chapter 1). Feed-forward circuits respond to persistent rather than transient stimulation, protecting the system from responding to spurious signals. The authors identified 13 "known" as well as 11 "new" key regulatory factors involved in dendritic inflammatory or antiviral responses. Twelve of the key regulators were also found to be associated with autoimmune and related diseases in genome-wide SNP association studies. The identified regulator and fine-tuner molecules are probably the main driver molecules that modulate the functional properties of DCs. Thus, this information is of great help

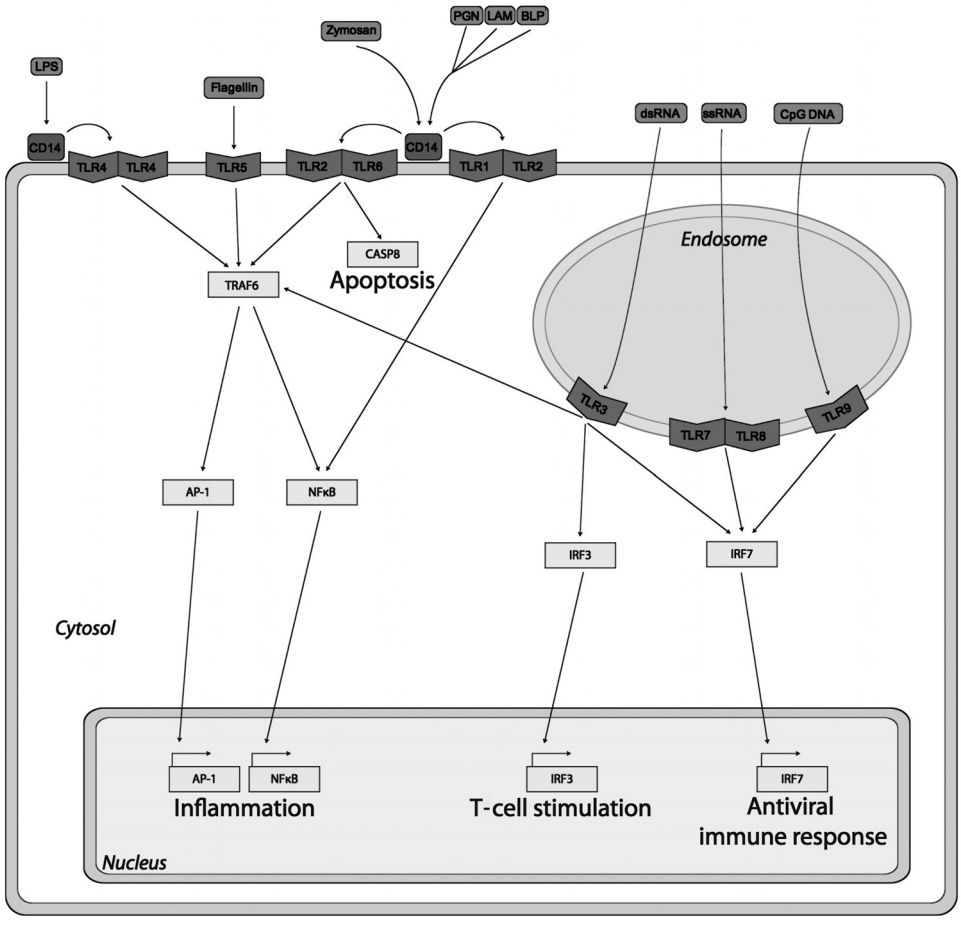


Figure 10.5 Simplified view of Toll-like receptors (TLRs), their ligands, and the downstream signaling cascade. TLR1, TLR2, TLR4, TLR5, and TLR6 are positioned on the cell membrane, whereas TLR3, TLR7, TLR8, and TLR9 are located inside an endosome. Different TLRs can bind different components of pathogens, after which various signaling cascades are triggered. This results in the onset of transcription of genes involved in inflammatory responses, T-cell stimulation or antiviral immune responses.

to explain and describe how pathogen-sensing pathways in DCs achieve specificity and sensitivity. Quantitative data of such key drivers molecules should become part of future (mathematical) models that capture and predict the behavior of DCs upon exposure to a particular pathogen.

Another example is published by Gilchrist et al. (2006) and further improved by Litvak et al. (2009) in which regulatory factors were identified that orchestrate the response of macrophages upon exposure to LPS. They used several different datasets, including time-dependent gene expression data, comparative sequence analysis, protein–protein interaction data, and quantitative protein localization

data to develop a model. The model was able to explain the behavior of the LPS/macrophage system: predictions produced by the model could be validated by experimental data.

Tissue/Organs

Several attempts have been described to model parts of the immune system. Folcik et al., generated the Basic Immune Simulator to identify potential targets for prevention of infectious diseases (Folcik et al., 2007). The model simulates the actions and interactions of autonomous "agents or entities" of the immune system and is used to study the relationship between cells of the innate and adaptive branch of the immune system and reproduces its complex behavior. The model is using basic molecular and cellular knowledge of the immune system. Also, Kalita et al. modeled and simulated the immune system computationally (Kalita et al., 2006). Macrophages, DCs, neutrophils, natural killer cells, B-cells, T-helper cells, complement proteins, and pathogenic bacteria are represented in the model. In fact, as early as 1995 the development of a C-language-based version of an IMMune system SIMulator (C-ImmSim) was started (Bernaschi and Castiglione, 2001; Castiglione et al., 2004). This computational program is based on known mechanisms that direct the humoral and cellular response of the immune system at the cellular level to specific antigens. The model includes several known aspects of the adaptive immunity, including the diversity of lymphocyte receptors due to genomic rearrangements in their encoding V-, J-, and C-regions, MHC restriction, clonal selection, development of immunocompetent T cells, antigen processing, antigen presentation, cell–cell cooperation, cell mobility, chemotaxis, hyper mutation of antibodies, maturation of the cellular and humoral response, and memory. Thus, the model represents the process of antigen processing and immune development that occurs locally in a small tissue part of organs like spleen, tonsil, and lymph nodes.

Recently, Rapin et al. provided an extended version of C-ImmSim that combined the power of genomic information with the abilities of C-ImmSim. The extended model takes into account, not only the genetic variation of lymphocyte receptors, but also the amino acid sequence variation in immunogenic proteins of pathogens. To this end, bioinformatic tools for T and B cell epitope predictions are used to mimic the recognition and binding of epitopes together with the activation and cooperation from T cells to stimulate host immune responses. This allowed the authors to perform *in silico* immunization experiments with real antigens. To assess the utility of the model, several different simulations were performed. In one instance, the authors performed a classical immunization experiment with the gag protein of the HIV-1 virus using a primary and secondary immunization. The model predicted a typical primary and secondary immune response, including the development memory since the secondary response appeared much faster than the primary response. In another simulation, the model predicted the phenomenon of affinity maturation of lymphocytes, i.e., higher proliferation levels and dominating outgrowth of lymphocyte clones with a high affinity to specific epitopes. Also, the advantage of MHC heterozygosity over homozygosity to clear viral infections was predicted in a simulation with influenza virus H1N1. Thus, this simulator

produces characteristics and dynamics that are consistent with basic immunological knowledge.

In our institute, we recently initiated efforts to identify key molecules that potentially drive the biological behavior of a developing intestine of young chicks. To this end, we generated gene correlation networks on the basis of time-series data (Schokker et al., 2009) to identify highly connected genes (or hub genes) in uninfected and *Salmonella*-infected animals. In healthy chickens, the identified hub genes were predominantly related to developmental processes, whereas in infected chicks the hub genes were more involved in processes related to host responses to pathogens (Schokker et al., 2011 accepted for publication). Furthermore, the data suggested that the major drive of the healthy system was focused on cellular development and cellular differentiation, whereas the infected system was focusing more on intercellular communication. The hypothesis is that the identified hubs are major drivers of system behavior and therefore putative candidates to modulate the system. We are now using this information to generate a mathematical model, using ODEs, in order to contribute to a better understanding of intestinal processes related to intestinal infections with gram-negative pathogens.

Organisms

Maybe the most challenging models are those that can predict the outcome of an infection on the organism level. An early example in this field has been published by Raman et al. (2010). They report an extensive model that covers multiple aspects of host–pathogen interaction of *M. tuberculosis* and integrates information from various biological levels. The host–pathogen interaction system is modeled as a Boolean network. The model accounts for several different steps in the infection process, including several mechanisms of pathogen invasion, defense of the host, and various defense mechanisms of the pathogen. The model consists of 56 host-related nodes (26 molecules, 11 cellular processes, 19 cell types or cell states), 18 pathogen-encoded components, and 12 quantitative parameters like bacterial load and growth, delayed onset of adaptive immunity, phagocytosis, and apoptosis. The nodes in the model represent molecules, processes, and cells, and their connections and interdependencies are represented by sequential processes that describe activities related to inhibition, activation, signaling, and proliferation or recruitment of cells. With this model, the authors try to identify the critical bacterial factors responsible for successful infection and the main components of the immune response for successful defense. At the end, the authors want to predict the influence of various factors and events, such as bacterial growth rate and delay in adaptive immune onset, on disease outcome, which is either bacterial clearance (no disease), bacterial persistence (carrier state), or bacterial growth (active disease).

Conclusions

Genomic technologies and tools offer major new opportunities to understand the genetic and nongenetic components involved in host–pathogen interactions.

Currently, large-scale genetic information is already used in the livestock industries to introduce the predictive principles (models) of "genomic selection" (Meuwissen and Goddard, 2001; Calus et al., 2008). Although genomic selection approaches are also applied in the field of host–pathogen interaction, this has not resulted yet in the selection of animals with improved disease resistance traits. A critical goal of host-pathogen research in livestock species is the identification of (new) targets for prevention and intervention and the development of (genomic-based) assays to select for disease resistance/susceptibility traits. This is a complex and very challenging task. As a first step, it is necessary to characterize infected animals in as comprehensive a manner as possible, thus studying the host and the pathogen at different space and time scales and at different biological levels, i.e., genomic, metabolic, immunologic, nutritional, health trait. The next step is most challenging: merging the data and providing a comprehensive and systems level perspective of host and pathogen responses. With the application of Systems Biology approaches, researchers will be able to better integrate and analyze data to improve our understanding of the spatial- and temporal-events leading to traits. In the future, the application of System Biology approaches will allow the animal sciences to predict the effect of both genetic as well as environmental changes on systems outcome through computational simulation. Compared to "genomics selection" this brings the animal sciences a next step forward in the area of predictive biology. Such predictive approaches will allow the monitoring and further improvements of traits by optimizing genotype–environment interactions.

As can be concluded from this chapter, the use of Systems Biology-based models that describe host–pathogen interaction and that integrate signaling, metabolic, immunologic, and health aspects as a function of two independent genomes and several environmental factors is still in its infancy. Disease is the outcome of the complex interplay between various pathogenic factors, as well as the host immune systems. Models that describe aspects of this interplay, some of which have been described in this chapter, are just a first step toward making sense of this complex interplay. Nevertheless, the available models already provide valuable insights into the importance of specific components for controlling infectious disease development. For example, they already provide insight into the role of specific pathogenic factors for cytokine expression and regulation. In addition, currently available models facilitate the integration and evaluation of new hypotheses. Of course, the current models still have major limitations, which originate for a great part from our limited understanding of the mechanistic details and quantitative aspects of virulence and host immune mechanisms. Therefore, most current models represent approximations of processes involved in host–pathogen interactions, as shown by the Boolean models in which nodes can have only two states (on or off; active or inactive). However, such a cellular state represents the outcome of a number of molecular events. Despite such limitations, several current models already provided new insights into the complex interplay between hosts and pathogens. The example described in 10.6.5 nicely demonstrates that systems-level modeling is an important step toward a holistic understanding of complex biological systems that can be used to predict the outcome of an infection, such as active disease, persistence, or clearance. Since we have only just arrived at the threshold of the area of Systems Biology, the expectation is that in the future it would be possible to develop a "virtual animal" model representing comprehensive

quantitative information on host–pathogen interactions, that takes into account both environmental as well as host- and pathogen-related genotypic variation.

References

Albert, I., Thakar, J., Li, S., et al. (2008) Boolean network simulations for life scientists. *Source Code for Biology and Medicine* **3**, 16.

Aldridge, P.D., Gray, M.A., Hirst, B.H., et al. (2005) Who's talking to whom? Epithelial-bacterial pathogen interactions. *Molecular Microbiology* **55**, 655–663.

Amit, I., Garber, M., Chevrier, N., et al. (2009) Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326**(5950): 257–63.

Baker, R.E., Gaffney, E.A., & Maini, P.K. (2008) Partial differential equations for self-organization in cellular and developmental biology. *Nonlinearity* **21**, R251–R290.

Banchereau, J., Paczesny, S., Blanco, P., et al. (2003) Dendritic cells: controllers of the immune system and a new promise for immunotherapy. *Annals of the New York Academy of Sciences* **987**, 180–187.

Bassing, C.H., Swat, W., & Alt, F.W. (2002) The mechanism and regulation of chromosomal V(D)J recombination. *Cell* **109 Suppl**, S45–S55.

Beg, A.A. (2002) Endogenous ligands of Toll-like receptors: implications for regulating inflammatory and immune responses. *Trends in Immunology* **23**, 509–512.

Bernaschi, M. & Castiglione, F. (2001) Design and implementation of an immune system simulator. *Computers in Biology and Medicine* **31**, 303–331.

Bhakdi, S., Tranum-Jensen, J., & Sziegoleit, A. (1985) Mechanism of membrane damage by streptolysin-O. *Infection and Immunity* **47**, 52–60.

Brack, C., Hirama, M., Lenhard-Schuller, R., et al. (1978) A complete immunoglobulin gene is created by somatic recombination. *Cell* **15**, 1–14.

Bruggemann, H., Hagman, A., Jules, M., et al. (2006) Virulence strategies for infecting phagocytes deduced from the in vivo transcriptional program of Legionella pneumophila. *Cellular Microbiology* **8**, 1228–1240.

Bumann, D. (2009) System-level analysis of Salmonella metabolism during infection. *Current Opinions in Microbiology* **12**, 559–567.

Cadman, E.T. & Lawrence, R.A. (2010) Granulocytes: effector cells or immunomodulators in the immune response to helminth infection? *Parasite Immunology* **32**, 1–19.

Calus, M.P., Meuwissen, T.H., de Roos, A.P., et al. (2008) Accuracy of genomic selection using different methods to define haplotypes. *Genetics* **178**, 553–561.

Cambi, A., Koopman, M., & Figdor, C.G. (2005) How C-type lectins detect pathogens. *Cellular Microbiology* **7**, 481–488.

Casadevall, A. & Pirofski, L.A. (2000) Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease. *Infection and Immunity* **68**, 6511–6518.

Castiglione, F., Poccia, F., D'Offizi, G., et al. (2004) Mutation, fitness, viral diversity, and predictive markers of disease progression in a computational model of HIV type 1 infection. *AIDS Research and Human Retroviruses* **20**, 1314–1323.

Celada, F. & Seiden, P.E. (1992) A Computer-Model of Cellular Interactions in the Immune-System. *Immunology Today* **13**, 56–62.

Dam, T.K. & Brewer, C.F. (2010) Lectins as pattern recognition molecules: the effects of epitope density in innate immunity. *Glycobiology* **20**, 270–279.

Dinges, M.M., Orwin, P.M., & Schlievert, P.M. (2000) Exotoxins of Staphylococcus aureus. *Clinical Microbiology Reviews* **13**, 16–34, table of contents.

dos Santos, R.M.Z. & Coutinho, S. (2001) Dynamics of HIV infection: A cellular automata approach. *Physical Review Letters* **8716**, 4.

Exley, R.M., Shaw, J., Mowe, E., et al. (2005) Available carbon source influences the resistance of Neisseria meningitidis against complement. *The Journal of Experimental Medicine* **201**, 1637–1645.

Feist, A.M., Henry, C.S., Reed, J.L., et al. (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology* **3**, 121.

Fenton, A. & Perkins, S.E. (2010) Applying predator-prey theory to modelling immune-mediated, within-host interspecific parasite interactions. *Parasitology* **137**, 1027–1038.

Figdor, C.G., van Kooyk, Y., & Adema, G.J. (2002) C-type lectin receptors on dendritic cells and Langerhans cells. *Nature Reviews Immunology* **2**, 77–84.

Finkelman, F.D., Lees, A., Birnbaum, R., et al. (1996) Dendritic cells can present antigen in vivo in a tolerogenic or immunogenic fashion. *The Journal of Immunology* **157**, 1406–1414.

Fisher, M., Huang, Y. S., Li, X., et al. (2008) Shr is a broad-spectrum surface receptor that contributes to adherence and virulence in group A streptococcus. *Infection and Immunology* **76**, 5006–5015.

Folcik, V.A., An, G.C., & Orosz, C.G. (2007) The Basic Immune Simulator: an agent-based model to study the interactions between innate and adaptive immunity. *Theoretical Biology and Medical Modelling* **4**, 39.

Galanos, C. & Freudenberg, M.A. (1993) Mechanisms of endotoxin shock and endotoxin hypersensitivity. *Immunobiology* **187**, 346–356.

Gammack, D., Doering, C.R., & Kirschner, D.E. (2004) Macrophage response to Mycobacterium tuberculosis infection. *Journal of Mathematical Biology* **48**, 218–242.

Gardner, M. (1970) Fantastic Combinations of John Conways New Solitaire Game Life. *Scientific American* **223**, 120.

Garg, A., Mohanram, K., Di Cara, A., et al. (2009) Modeling stochasticity and robustness in gene regulatory networks. *Bioinformatics* **25**, i101–i109.

Gilchrist, M., Thorsson, V., Li, B., et al. (2006) Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4. *Nature* **441**(7090), 173–178.

Goldstein, B., Faeder, J.R., & Hlavacek, W.S. (2004) Mathematical and computational models of immune-receptor signalling. *Nature Reviews Immunology* **4**, 445–456.

Hamerman, J.A., Ogasawara, K., & Lanier, L.L. (2005) NK cells in innate immunity. *Current Opinion in Immunology* **17**, 29–35.

Hassell, M.P., Comins, H.N., & May, R.M. (1991) Spatial Structure and Chaos in Insect Population-Dynamics. *Nature* **353**(6341), 255–258.

Hayday, A.C. (2000) [gamma][delta] cells: a right time and a right place for a conserved third way of protection. *Annual Reviews Immunology* **18**, 975–1026.

Heath, W.R. & Carbone, F.R. (2001) Cross-presentation, dendritic cells, tolerance and immunity. *Annual Review of Immunology* **19**, 47–64.

Hethcote, H.W. (2000) The mathematics of infectious diseases. *Siam Review* **42**, 599–653.

Hethcote, H.W. & Van Den Driessche, P. (1995) An SIS epidemic model with variable population size and a delay. *Journal of Mathematical Biology* **34**, 177–194.

Hibino, T., Loza-Coll, M., Messier, C., et al. (2006) The immune gene repertoire encoded in the purple sea urchin genome. *Developmental Biology* **300**, 349–365.

Howard, C.J., Charleston, B., Stephens, S.A., et al. (2004) The role of dendritic cells in shaping the immune response. *Animal Health Research Reviews* **5**, 191–195.

Hozumi, N. & Tonegawa, S. (1976) Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proceedings of the National Academy of Sciences of the U S A* **73**, 3628–3632.

Huang, F.P., Platt, N., Wykes, M., et al. (2000) A discrete subpopulation of dendritic cells transports apoptotic intestinal epithelial cells to T cell areas of mesenteric lymph nodes. *Journal of Experimental Medicine* **191**, 435–444.

Hucka, M., Finney, A., Sauro, H.M., et al. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531.

Hyams, C., Camberlein, E., Cohen, J.M., et al. (2010) The Streptococcus pneumoniae capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. *Infection and Immunity* **78**, 704–715.

Iglewski, B.H. & Kabat, D. (1975) NAD-dependent inhibition of protein synthesis by Pseudomonas aeruginosa toxin. *Proceedings of the National Academy of Sciences of the U S A* **72**, 2284–2288.

Ishii, N., Nakahigashi, K., Baba, T., et al. (2007) Multiple high-throughput analyses monitor the response of E. coli to perturbations. *Science* **316**(5824), 593–597.

Izmailova, E., Bertley, F.M.N., Huang, Q., et al. (2003) HIV-1 Tat reprograms immature dendritic cells to express chemoattractants for activated T cells and macrophages. *Nature Medicine* **9**, 191–197.

Janeway, C.A., Jr. (1989) Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harbor Symposium in Quantative Biology* **54** Pt 1, 1–13.

Jenner, R.G. & Young, R.A. (2005) Insights into host responses against pathogens from transcriptional profiling. *Nature Reviews Microbiology* **3**, 281–294.

Kalita, J.K., Chandrashekar, K., Hans, R., et al. (2006) Computational modelling and simulation of the immune system. *International Journal of Bioinformatics Research and Applications* **2**, 63–88.

Kaunzinger, C.M.K. & Morin, P.J. (1998) Productivity controls food-chain properties in microbial communities. *Nature* **395**(6701), 495–497.

Kotb, M. (1995) Bacterial pyrogenic exotoxins as superantigens. *Clinical Microbiology Reviews* **8**, 411–426.

Kumar, H., Kawai, T., & Akira, S. (2009) Toll-like receptors and innate immunity. *Biochemical Biophysical Research Communications* **388**, 621–625.

Law, S.K., Lichtenberg, N.A., & Levine, R.P. (1980) Covalent binding and hemolytic activity of complement proteins. *Proceedings of the National Academy of Sciences of the U S A* **77**, 7194–7198.

Le Novere, N., Hucka, M., Mi, H., et al. (2009) The Systems Biology Graphical Notation. *Nature Biotechnology* **27**, 735–741.

Li, M.Y., Graef, J.R., Wang, L. et al. (1999) Global dynamics of a SEIR model with varying total population size. *Mathematical Biosciences* **160**, 191–213.

Lieber, M.R. (1991) Site-specific recombination in the immune system. *FASEB Journal* **5**, 2934–2944.

Lippolis, J.D. & Reinhardt, T.A. (2008) Centennial paper: Proteomics in animal science. *Journal of Animal Science* **86**, 2430–2441.

Litvak, V., Ramsey, S.A., Rust, A.G., et al. (2009) Function of C/EBPdelta in a regulatory circuit that discriminates between transient and persistent TLR4-induced signals. *Nature Immunology* **10**, 437–443.

Ly, K.T. & Casanova, J.E. (2007) Mechanisms of Salmonella entry into host cells. *Cellular Microbiology* **9**, 2103–2111.

Mardis, E.R. (2008a) The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**, 133–141.

Mardis, E.R. (2008b) Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics* **9**, 387–402.

Marshall, J.S. & Jawdat, D.M. (2004) Mast cells in innate immunity. *Journal of Allergy and Clinical Immunology* **114**, 21–27.

Mattoo, S., Foreman-Wykert, A.K., Cotter, P.A., et al. (2001) Mechanisms of Bordetella pathogenesis. *Frontiers in Bioscience* **6**, E168–E186.

McHeyzer-Williams, L.J. & McHeyzer-Williams, M.G. (2005) Antigen-specific memory B cell development. *Annual Review of Immunology* **23**, 487–513.
McKeithan, T.W. (1995) Kinetic proofreading in T-cell receptor signal transduction. *Proceedings of the National Academy of Sciences of the U S A* **92**, 5042–5046.
Medzhitov, R. (2001) Toll-like receptors and innate immunity. *Nature Reviews Immunology* **1**, 135–145.
Meuwissen, T.H. & Goddard, M.E. (2001) Prediction of identity by descent probabilities from marker-haplotypes. *Genetics Selection Evolution* **33**, 605–634.
Milenbachs, A.A., Brown, D.P., Moors, M., et al. (1997) Carbon-source regulation of virulence gene expression in Listeria monocytogenes. *Molecular Microbiology* **23**, 1075–1085.
Mosca, F., Tritto, E., Muzzi, A., et al. (2008) Molecular and cellular signatures of human vaccine adjuvants. *Proceedings of the National Academy of Sciences of the U S A* **105**, 10501–10506.
Mosmann, T.R. & Sad, S. (1996) The expanding universe of T-cell subsets: Th1, Th2 and more. *Immunology Today* **17**, 138–146.
Navratil, V., Lotteau, V., & Rabourdin-Combe, C. (2010) The virtual infected cell: a systems biology rational for antiviral drug discovery. *Medical Science (Paris)* **26**, 603–609.
Nomura, K. & He, S.Y. (2005) Powerful screens for bacterial virulence proteins. *Proceedings of the National Academy of Sciences of the United States of the USA* **102**, 3527–3528.
Oettinger, M.A., Schatz, D.G., Gorka, C., et al. (1990) RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248**(4962), 1517–1523.
Owen, M.R. & Sherratt, J.A. (1999) Mathematical modelling of macrophage dynamics in tumours. *Mathematical Models & Methods in Applied Sciences* **9**, 513–539.
Parkhill, J., Sebaihia, M., Preston, A., et al. (2003) Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. *Nature Genetics* **35**, 32–40.
Pieters, J. & Gatfield, J. (2002) Hijacking the host: survival of pathogenic mycobacteria inside macrophages. *Trends in Microbiology* **10**, 142–146.
Pulendran, B. (2009) Learning immunology from the yellow fever vaccine: innate immunity to systems vaccinology. *Nature Reviews Immunology* **9**, 741–747.
Rabinowitz, J.D., Beeson, C., Lyons, D.S., et al. (1996) Kinetic discrimination in T-cell activation. *Proceedings of the National Academy of Sciences of the U S A* **93**, 1401–1405.
Raghunathan, A., Reed, J., Shin, S., et al. (2009) Constraint-based analysis of metabolic capacity of Salmonella typhimurium during host-pathogen interaction. *BMC Systems Biology* **3**, 38.
Raman, K., Bhat, A.G., & Chandra, N. (2010) A systems perspective of host-pathogen interactions: predicting disease outcome in tuberculosis. *Molecular Biosystems* **6**, 516–530.
Rapin, N., Lund, O., Bernaschi, M., et al. (2010) Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One* **5**, e9862.
Roberts, I.S. (1996) The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annual Review of Microbiology* **50**, 285–315.
Romagnani, S. (2006) Regulation of the T cell response. *Clinical & Experimental Allergy* **36**, 1357–1366.
Sadofsky, M.J. (2001) The RAG proteins in V(D)J recombination: more than just a nuclease. *Nucleic Acids Research* **29**, 1399–1409.
Schatz, D.G., Oettinger, M.A., & Baltimore, D. (1989) The V(D)J recombination activating gene, RAG-1. *Cell* **59**, 1035–1048.
Schokker, D., Hoekman, A.J., Smits, M.A., et al. (2009) Gene expression patterns associated with chicken jejunal development. *Developmental & Comparative Immunology* **33**, 1156–1164.

Schokker, D., Smits, M.A., Hoekman, A.J., et al. (2010) Effects of Salmonella on spatial-temporal processes of jejunal development in chickens. *Developmental & Comparative Immunology* **34**, 1090–1100.

Schroeder, J.T. (2009) Basophils beyond effector cells of allergic inflammation. *Advances in Immunology* **101**, 123–161.

Seiden, P.E. & Celada, F. (1992) A Model for Simulating Cognate Recognition and Response in the Immune-System. *Journal of Theoretical Biology* **158**, 329–357.

Serbina, N.V., Jia, T., Hohl, T.M., et al. (2008) Monocyte-mediated defense against microbial pathogens. *Annual Review of Immunology* **26**, 421–452.

Steinman, R.M., Turley, S., Mellman, I., et al. (2000) The induction of tolerance by dendritic cells that have captured apoptotic cells. *Journal of Experimental Medicine* **191**, 411–416.

Tapping, R.I. (2009) Innate immune sensing and activation of cell surface Toll-like receptors. *Seminars in Immunology* **21**, 175–184.

Thakar, J., Pilione, M., Kirimanjeswara, G., et al. (2007) Modeling systems-level regulation of host immune responses. *PLoS Computational Biology* **3**, e109.

Valitutti, S., Muller, S., Cella, M., et al. (1995) Serial triggering of many T-cell receptors by a few peptide-MHC complexes. *Nature* **375**(6527), 148–151.

Wangersky, P.J. (1978) Lotka-Volterra Population Models. *Annual Review of Ecology and Systematics* **9**, 189–218.

Williams, M.A. & Bevan, M.J. (2007) Effector and memory CTL differentiation. *Annual Review of Immunology* **25**, 171–192.

Wofsy, C., Coombs, D., & Goldstein, B. (2001) Calculations show substantial serial engagement of T cell receptors. *Biophysical Journal* **80**, 606–612.

Wolfram, S. (1984) Cellular Automata as Models of Complexity. *Nature* **311**(5985), 419–424.

Wong, J.P., Christopher, M.E., Viswanathan, S., et al. (2009) Activation of toll-like receptor signaling pathway for protection against influenza virus infection. *Vaccine* **27**, 3481–3483.

Wunder, F., Kalthof, B., Muller, T., et al. (2008) Functional cell-based assays in microliter volumes for ultra-high throughput screening. *Combinatorial Chemistry and High Throughput Screening* **11**, 495–504.

Yashiro, Y., Bannai, H., Minowa, T., et al. (2009) Transcriptional profiling of hematopoietic stem cells by high-throughput sequencing. *International Journal of Hematology* **89**, 24–33.

Chapter 11
Systems Biology in Livestock Science and Commercial Livestock Business

Marinus F.W. te Pas, Henri Woelders, and André Bannink

Introduction

This book demonstrates that Systems Biology is an emerging interdisciplinary science combining biology at all its levels of organization. This requires information from biochemistry, biophysics, bioinformatics, and physiology and it includes the interactions between them and mathematics to develop quantitative, predictive models. The aim of Systems Biology is to provide insight in the complexity of biological processes, which underlie the characteristics of living organisms.

Many production traits or production-related traits in livestock are the result of complex biological processes. Livestock traits have a genetic basis with interactions with the environment being of a variable importance. Production traits—or the products they relate to—are either produced in tissues and excreted (e.g., milk), are the tissues themselves (e.g., muscle mass—meat yield), or are the product of the functioning of combinations of tissues (e.g., reproduction, health). Genetic selection to optimize production is thereby accompanied by a changed cellular and tissue metabolism and/or composition. Animals nutrition can be directed at an optimal performance with respect to production, health, and environmental impact. Farm management may allow a certain animal behavior or induce stress to the animals (interaction with the environment) affecting the response to be expected from genetic selection and nutrition. Breeding, nutrition, and management all may induce unwanted side effects with optimization on a narrow window of selection of traits, and may affect animal health or welfare, and hence consumer's concern.

Thus, livestock management is complex. Because Systems Biology aims to unravel the complex relations of life, it is expected that commercial livestock industry will benefit from it. However, it is important that not only science has a say in this but also the livestock industry gives an opinion on the benefits that may be expected from Systems Biology. Is the concept of Systems Biology known in the livestock industry? Does the livestock industry have experience with Systems Biology research? What are the expectations of the livestock industry from Systems Biology? This chapter reflects

Systems Biology and Livestock Science, First Edition. Edited by Marinus F.W. te Pas, Henri Woelders, and André Bannink.

Box 11.1 Experts from the Livestock Industries

Prof. Dr. Leo den Hartog, Professor Farm Management in Animal Production at Wageningen University and Director R&D and Quality Affairs at Nutreco. Nutreco is a global leader in animal nutrition and fish feed.

Dr. Theo van Kempen and Dr. John Newbold, Provimi Research and Innovation. Newbold and van Kempen are Science Director for Ruminant and Swine, respectively. Van Kempen is also Adjunct Professor at the North Carolina State University, USA. Provimi is one of the leaders in the international animal feed industry.

Dr. Pramod Mathur, Senior Research Geneticist at IPG. IPG is an independent enterprise, strategic partner of Topigs—one of the three largest pig breeders in the world.

Dr. Alfred de Vries, Manager Genetic Products of CRV. CRV is an international enterprise in the field of cattle improvement.

Dr. Gerard Albers and Pierre Cherel, DVM, of Hendrix Genetics. Albers is Director Research and Technology of Hendrix Genetics. Cherel, DVM, is a Research Geneticist located in France and involved in several more fundamental projects. Hendrix Genetics is a leading multispecies breeding company (poultry, layers, turkeys, pigs, aquaculture) serving at least half of the world production of laying hens as well as turkeys and is the second largest pig breeder in the world.

the opinion of a number of leading scientists and managers from the livestock industry itself (see Box 11.1) on the potential of Systems Biology.

Expectations from Systems Biology for Livestock Science and Industrial Innovations

Animal Feed Industry

What Is Systems Biology?

The definition of what Systems Biology is was discussed in most interviews. For instance, Newbold and Van Kempen, science directors at Provimi, stated that the definition of Systems Biology remains still rather vague. What does it exactly stand for, how is it defined, and how does it differ from previous modeling work? What knowledge will it add? According to Den Hartog, R&D and Quality Affairs Director at Nutreco, the lack of a straightforward definition of Systems Biology makes it difficult to define the opportunities that lie ahead. Different types of system definitions can be distinguished when seeking solutions for the practice of livestock production. Systems may be the whole production chain, a specific farming system or a specific multi-site production system, and intra-animal physiological and regulatory systems. Another requirement is that Systems Biology needs to deliver the mode of action of an opportunity in order to become applicable for animal feed industry. In this respect, Van Kempen, Newbold, and Den Hartog all indicated that a black-box approach must

be considered insufficient for introduction of a new nutritional or farm management concept in the market. To have Systems Biology applied by industry, statements are needed on the factors that play a role, on how they interact and on what mechanisms are involved. Often, science demonstrates associations between factors and animal responses, but for industry, it should be clarified how they can be used and how they actually have to be interpreted.

Furthermore, Den Hartog indicated that answers need to be generated within the system of interest to the industry as a starting point, which may go beyond that of an individual animal. This means that Systems Biology needs to start from a well-defined problem for the industry or some specific consumer's demands.

Integration

Independently of each other, the research directors of the two animal feed companies value the integration of research results from the diverse biological levels. With all the scientific achievements and new technologies becoming available, Den Hartog explained that an integrated approach from a farm perspective remains most important. Various disciplines and different viewpoints on a problem have to come together when on-farm problems are to be solved. In the animal feed industry, integration of known aspects of the feed and interactions with the animals of different genotype and phenotype is already applied. Functioning of the gastrointestinal tract is a key issue to address at this moment for both Provimi and Nutreco. The focus is on optimizing the relation between the nutritional requirements of the animal and the offered nutrients, and on the interaction with the environment (e.g., immune responses to pathogens). The role of feed composition in relation to these traits is of major importance. Additionally, there is a strong focus on transition phases in the life of farm animals (i.e., birth, weaning, lactation, animal housing) related to animal health and well-being.

Systems Biology aims to integrate "omics" data and mechanisms at the molecular level, to mechanisms at the level of the animal, including the response of the animal to its environment. Indeed, the interesting factors for animal feed companies such as Nutreco and Provimi go beyond the level of the animal, including groups of animals, the animal's environment, farm management, and expectations from consumers/society. When Systems Biology is to deliver applicable answers for the animal feed industry, these factors need to be taken into account as well. For example, Den Hartog indicated that a huge part (about 30%) of the theoretical genetic potential is estimated not to be utilized because of suboptimal farm circumstances or suboptimal feeding. Taking the large variation among animals into account may contribute to the possibility to make more and better use (with less problems and side effects) of the genetic potential of individual animals. It needs to be defined and understood how diet, nutritional strategy, and other environmental factors interact with genotype. Models that are developed with a Systems Biology approach must obtain a truly predictive capacity. This means that they should become able to quantitatively predict the interactions between nutrition, genotype, and environment. These relationships are quite diverse and may refer to, for example, effects across generations, effects on the carbon footprint of livestock production, or the need to eradicate *Salmonella* from the food chain. Although every livestock species has its own specific problems to be solved, to become helpful, Systems Biology should deliver a better understanding of (1) the

variation expressed among animals and (2) the variations in the interactions among animals and the environmental conditions, such as the relationship between nutrition, genotype, and environment. Systems Biology should do better than traditional monodisciplinary approaches can deliver. Customized feeding (precision feeding) of individual animals or farm-specific strategies are needed to exploit the full potential and to prevent common problems in livestock production.

Also, applicability of omics techniques themselves was discussed with Van Kempen and Newbold. They questioned how to make use of the possibilities offered by the new omics technologies using the Systems Biology approach. They stated that results collected with the new omics technologies are challenging, but puzzling as well. Such omics methods result in large datasets, which require interpretation. Also, at least at present, use of omics analysis still brings high costs and the need for specialized labs. Furthermore, there is a need to go beyond the stage of identifying associations between omics data and the animals' traits or responses of interest. Besides the ability to measure many things with new high-throughput technologies, the ability to integrate all these data into a view on cause and effect in the animal is considered a very important condition for Systems Biology to be useful for industry.

A Systems Biology approach may help industry if some important conditions are met. Firstly, Systems Biology has to be able to quantify effects at different biological levels to indicate positive effects on animal performance as an end result. Secondly, it has to add up information up to the level, which accommodates the interest of industry, which is often metabolites and physiological regulatory mechanisms. Physiological mechanisms have, of course, also been studied in the past, but without the possibilities of omics techniques that are currently available. For example, metabolomics is considered a useful new technology because it can deliver a more complete picture of metabolites and factors involved with the regulation of animal physiology and animal response under various conditions, which remained unnoticed or unidentified before. Van Kempen stated that Provimi has already trials in which the range of metabolites is measured in search for clues what is happening in the animal, for example, in the case of a new disease. In this manner, these omics technologies may deliver a more complete picture and contribute to a formulation of new (types of) hypotheses or ideas to be explored further and applied by industry.

For example, Newbold mentioned an area of research, which seems rather unexplored so far, but which is considered to be highly important in bringing together nutrition, genetics, and genomics. Newbold expects that the interactions between these expertises will bring new knowledge that can explain why animals react diversely to feed substances, feeding strategies, and farm management. It may also lead to prediction of the animal's response for future use. The question remains how this research actually can be organized, because combining these disciplines will require very extensive and costly trials, including many animals as well as costly techniques to be used for every single animal. Another area deserving more attention is a more detailed description and understanding of the components in feed and their pre- as well as postabsorptive effects on the target animal (with respect to species, type of animal, physiological state).

However, there are also concerns. Van Kempen and Newbold expressed their concern that software currently available to analyze omics results may not be conclusive enough (too much black-box and too general) to pinpoint the mode of action and

cause and effect relationships in a specific well-defined target animal. Van Kempen indicated that a major problem for industry is the lack of reasonable explanations for the inconsistent results from different trials with the same nutritional measure or problem investigated, but performed at different locations or by different research groups or in different lab. Moreover, often, no effects are seen in large-scale experiments performed by industry when positive results had been obtained in smaller scale, detailed experiments conducted at research institutes and universities. It would be a major step forward if a Systems Biology approach helps out by demonstrating why certain observations are not reproduced in some of the trials. This could lead to a much faster development of successful new products by the feed industry as solutions for specific problems in the farming practice.

Animal Breeding Industry

Expectations: Genomics as a Starting Point

Mathur, senior geneticist as IPG, stated that livestock industry is moving away from traditional farming to a technology- and innovation-based industry. The global competition requires a faster rate of progress in delivering the quality as desired by consumers. Whether it is milk, meat, or eggs, the traditional emphasis on increasing the quantity of production are becoming obsolete. There is a higher demand for quality to meet the consumer and societal needs. The consumers now are looking for healthy, nutritive products, produced with proper care of environment and animal welfare. Many of these characteristics are hard to improve through conventional means. The sequencing of genomes of some of the livestock species (e.g., cattle, pigs, and poultry) has provided some tools for identification of SNPs, genomic region, and some genes related to the traits of interest. But these are bits and pieces of the puzzle. An integrated approach through Systems Biology will be of interest to understand the systems structure and dynamics as well as the control and design mechanisms for the biological processes. This will enhance our understanding of the new phenotypes for new breeding goals in livestock production. It is otherwise very difficult and expensive to measure some of these characteristics and attempt to make genetic progress in strictly the conventional ways.

Currently, a major part of the livestock industry looks upon Systems Biology as an interesting emerging discipline that should be able to contribute to a better understanding of important biological functions. However, Systems Biology is still in its infancy, especially for livestock. There are hardly any results that the industry can directly apply. The science has to grow and provide tools that can be applied in the livestock industry in a cost-effective manner.

De Vries of cattle breeding organization CRV indicated that genomic selection has become very important for CRV. With genomic selection, one predicts the breeding value of individual animals in the absence of direct phenotypic measurements. Predictions are based on the relationship between large numbers of consecutive genome fragments and phenotypical performance as established in a reference population. The current genomic selection methods use an additive model, i.e., all effects of genes (or rather of pieces of the genome) are simply added, without allowing for interactions

between genes. Moreover, as genomic selection only relies on associations established in a reference population, the results may not be fully translatable to other populations or breeds, for example, in other countries. This may be due to genetic differences as well as interactions with environmental factors. In his opinion, Systems Biology could help predicting these interactions in order to solve this problem. A related issue is that CRV uses a multitude of (small) herds, which have variation in management and environmental factors, to test bulls or to find associations for genomic selection, whereas some other breeding organizations use just a few defined large testing herds. The advantages of the CRV approach are the wide involvement of the farmers and the validity of the results over the tested range of environments. But, on the other hand, this approach makes it difficult to actually describe and quantify how genotypes and environmental factors interact.

Similarly, Hendrix Genetics' research director Albers indicated the research aims with regard to breeding value estimations. Presently, the use of (structural) genomics using several types of SNP-CHIPs or sequencing in order to describe genetic variation and relate it to phenotypic variation is generally applied. Preparations to sequence the entire genomes of a number of chickens are ongoing, aiming to know the complete sequence of breeding lines. In France, a candidate gene project showed several interesting candidate genes. In the Netherlands, also some functional genomics (transcriptomics) results were obtained, but Dr. Albers indicates that he is less interested in gene expression profiling. Breeding (breeding business) is about selling genotypes that have a predictable potential for producing certain phenotypes. The process between the two is in itself of no direct interest to him unless the knowledge of this process would yield useful biomarkers that can be used for more efficient genetic selection programs.

Cherel stated that Systems Biology in itself is an interesting and logical option as a follow-up of genomics. The potential of predicting traits is always challenging for the livestock industry. And forecasting the output of animal productions is more than ever important in a risk-averse economic environment. But, when it comes to expectations for the livestock industry, he thinks that although in livestock research and especially nutrition modeling is not new, today, application of Systems Biology to livestock species might be a premature attempt. The outcomes of successful attempts to model livestock production through a Systems Biology approach would be extremely valuable to the industry. But, Systems Biology approaches need to demonstrate step by step their capabilities to integrate complex datasets, and to model results of a growing number of genetic and environment effects. Livestock may just not be the easiest test case for such a demonstration. This is noteworthy in a contemporary context where functional genomics recently delivered more data than applications, thus motivating the demand of proofs of concept first.

Mathur added that Systems Biology offers a huge potential, but that it is important to be cautious in the optimism. The science should not overestimate the potential, or should not make false promises that will be hard to meet. In biological science, the past experience is that the more we discover the more complex it gets. The approaches in genomics, transcriptomics, proteomics, and several others "omics" will generate a huge volume of data. A lot of development in bioinformatics will be required to manage and process the data. Further, developments will be required also in understanding and making use of the information.

Another important aspect is development of methods for accurately measuring functions of minute cellular particles and molecules. Also, for these aspects, a high-throughput and ability of processing large volumes of information will be required.

It will be very important to select the right question and hypothesis relevant to the industry. Gathering molecular and biological data with respect to a single type of disease, or a single type of protein can take several years. In that time frame either the pathogens can mutate and become ineffective or completely wipe off the industry. Therefore, choosing the right question and delivering the expected result in a timely manner will be an important aspect of how relevant Systems Biology can be for industry. In most cases, the solutions do not need to be complex, or even be complete. A lot of progress can be made with less than perfect but simple, targeted and timely information.

Expectations: Application of Systems Biology for Specific Problems and Questions

Albers indicated that Systems Biology might be interesting for scientific aims. For example, it may add to the knowledge on interactions between host and parasites. But apart from pure academic use, Albers is skeptical about Systems Biology producing usable results for the breeding industry. However, Systems Biology may have some use for gaining knowledge about processes taking place at a whole-animal level that may not be directly measurable. In the process from DNA to phenotype of an animal, Albers is skeptical for a role of Systems Biology. The whole animal is important. And not only the whole animal but also its interactions with other animals, its nutrition, and other environmental factors. This is the complete system. And it is quite unlikely that Systems Biology would be able to encompass this entire system, completely and with sufficient detail and accuracy, to be of use to help in predicting this complete system adequately and correctly. One could try and limit the system by modeling only part of the system, for example, an organ or tissue, but then, is that Systems Biology or reductionism? The work that is currently done in simulating (parts of) the functioning of an animal are interesting and probably useful to understand physiological processes or diseases. But these models may yet not be complete enough to explain the phenotype of the entire animal from its genotype. And that is what breeding is about. The development of ascites in broilers is mentioned as an example of what would seem a relatively simple process, but even in this case a systems analysis is not likely to produce a model with good predictive value.

In contrast, De Vries indicated several specific problems where he thinks that Systems Biology may be helpful to find solutions. As indicated above, in relation to breeding value estimation, it is important to address the interactions between genes or between genes and environment. But, additionally one would like to be able to take into account that relationships may be nonlinear, which adds to the difficulty of predicting a phenotype or a breeding value. One other reason that a Systems Biology approach may be helpful is that selection in most cases is not toward a single trait but rather toward a combination of traits. These traits may influence each other as they rely on the same (energy/nutrient) resources or make use of the same intermediary metabolism or regulation mechanisms. (As an example, interactions

between milk yield, energy metabolism, body composition, and fertility is mentioned). Systems Biology could help understanding such interactions between trait systems, and define how traits can be optimally combined in the genetic models. Similarly, understanding of the relations between genes plus environment and phenotype could make it easier to "translate" results from one population or even one breed to another.

One other aspect is the fact that in some cases we may not know which parameters could be used best to estimate a phenotype. Milk measurements like the NIR spectrometry for determining the fatty acid composition are adding to the repertoire of possibilities. For some traits, we may not know (yet) which parameters could be best suited to estimate a phenotypic trait and can also be obtained in large numbers easily and cheaply. Systems Biology could help in defining which (deep) phenotypic parameters could be used.

Another knowledge gap where Systems Biology could importantly add to understanding is the rumen. Due to complex processes by the bacteria in the rumen, cattle can live on grass—a low nutrient food. However, this mechanism results in emissions of high levels of methane, which is a potent greenhouse gas. So, one goal would be to reduce methane emission. On the other hand, the effective functioning of the rumen and the rest of the digestive tract is important for the energy requiring trait systems like milk production in the udder, fertility, and disease resistance. Systems Biology could help in understanding the interactions between the animal and the rumen flora in relation to nutrient use and methane emission.

Mathur believes that Systems Biology should prove to be most useful in areas that are relatively new and difficult to handle compared to the traits of quantity and efficiency of livestock production. New subjects of major interest are health, welfare, and product quality, production efficiency including robustness, and emissions and environmental issues. An example is animal welfare. Consider the concern about castration of little piglets. This intervention is becoming a huge animal welfare issue and a ban on castration is expected to be imposed on pig producers soon. Also, producers do not benefit from castration because production of castrated males is less efficient. However, if boars are not castrated, the meat of some entire males may have a penetrating "animal," "sweat," "fecal," or "urine" like odor called boar taint. One can measure and quantify, and select against boar taint in conventional ways. Genomic analysis with the 60K SNP chip has allowed identifying some SNPs and genomic regions associated with boar taint. But, the biological background remains hard to understand. Therefore, selection using these SNPs may have unwanted side effects on fertility of females and aggression of males is less than optimal.

Another example is animal health, which is a very complex trait, and difficult and expensive to measure. Usually, challenge experiments are required that provide only few phenotypes. A Systems Biology approach will enhance our understanding in host pathogen interactions and allow the development of diagnostic tools and bioassays for the measurement of immune responses on a larger scale and at a lower price.

Mathur elaborated on how Systems Biology could work to fulfill these promises. Let us take an example of animal health; one of the approaches related to Systems Biology

is Proteomics. Proteomics contributes to the study of Systems Biology by revealing protein–protein interactions as well as protein expression levels. For example, it can be used to suggest pathways involved in cellular function or in a pathogenic infection. The use of two-dimensional electrophoresis can create proteome maps of cells and tissues and provides the relative abundance of proteins present in a sample. Further, mass spectrometry can be used to determine masses as well as the sequences of peptides from a given protein. Proteomic techniques can be used to identify possible pathways that are involved in normal cellular function and in pathogen-altered functions, and thus can serve as one of the initial steps in the study of Systems Biology.

Cherel suggests an indirect route for Systems Biology toward livestock industrial use: His advice would be to concentrate first on model organisms, as yeast or mice, of which well-defined and more extensive genetic models are available and which may be more suited for building up causative and quantitative links from genotype or environment to phenotype. Livestock industry is obviously concentrating on the most productive genotypes only, and is unlikely to explore genome variation at large and all of its consequences. Variation in environment conditions may also hinder the productive use of data from livestock species. Production environments tend to be standardized overall, from a management perspective. But, practically speaking, and from an organism side, the chronic and multiple health infectious challenges, the continuous optimization of feed, interactions among production groups members, and to a lesser extent climate variation all contribute to sources of noises in observations of the phenotype. It may prove difficult to use most of these types of variation in modeling approaches just because some factors, such as infectious pressure as an example, are not documented.

In subsequent stages, outbred selected lines from model organisms may be used as a more realistic animal model for quantitative genetic variation relevant to animal breeders, but in a fully controlled environment. Furthermore, although other traits may be more important, Cherel's advice would be to start with well defined but complex traits such as growth (rate), for which model species can offer realistic proxies.

A major determinant of growth (rate) is nutrition (which is, for example, approximately 60% of the costs of pig production). Historically, application of complex nutrition models have always been limited by the ability to realistically measure animal genotypes capabilities with regard to fat or lean tissue deposition rates in response to variation in nutrient input. Revisiting those response curves may be a good starting point for Systems Biology, although this represents a major undertaking as it covers numerous physiological functions: feed digestion, nutrients absorption, storage, tissue build up, and energy control. Economic interest of such modeling is large and known for decades but may get a chance to go a few steps forward using additional sources of data in a Systems Biology context. This would be a boon if it can provide additional hints on physiological traits relevant to growth phenotype, which could be further explored for genetic variation.

One may argue that chicken shows numerous attributes of a model species while being a major contributor of animal products. Still, an extensive range of genetic models is probably needed to set up a System Biology approach, controlling environment is possible but not standard in production, and livestock industry is unlikely to sustain the needed genetic variability in their normal operations!

Conclusions

Scientists from animal feed industry conclude that Systems Biology has potential for them, provided that certain conditions are met. The scientists from the different breeding industries have more varying views on the relevance of Systems Biology for the industry.

Den Hartog thinks Systems Biology may become a promising approach within livestock sciences. A Systems Biology approach needs to include aspects of livestock production systems taking into account the interaction of the environment with the animal. To become applicable for the animal feed industry, it has to also highlight the biological mechanism underlying suggested solutions for problems. Van Kempen and Newbold conclude that industry might benefit from Systems Biology, but also indicate that perhaps industry does not realize or recognize this yet. The term Systems Biology remains too vague and unknown, as long as there are no good examples available of successful applications of Systems Biology. Systems Biology certainly seems helpful in developing new ways of conducting research and in improving the process of hypothesis formulation and testing. Nevertheless, industry must often focus on the rather short term, which contrasts with the rather long-term research agenda probably needed for successful Systems Biology projects. Systems Biology should deliver convincing results with respect to the physiological mechanisms involved, including the side effects to be expected, to make it applicable for industry and for the development of successful feeding strategies and products for the animal feed market.

Mathur and De Vries think there may also be useful applications of Systems Biology regarding the breeding industry. De Vries expects that Systems Biology will provide better predictability of results, especially when interactions between animal and environmental factors are involved. Mathur warns that livestock industry has to function in a very cost-effective way to retain global competitiveness. The Systems Biology needs to provide practical knowledge that can be applied directly by the livestock industry in a cost-effective way. Contrary to this, Cherel thinks that Systems Biology is a very interesting long-term ambition that may be important for defining better-characterized phenotypes and will be helpful in interpretation of omics results. Albers believes that Systems Biology may be very interesting for academic purposes and for fundamental research, but at present he sees a limited use for the breeding industry.

Considering these views from experts in the field, we can state that Systems Biology is still in its infancy, especially for livestock science. However, if Systems Biology proves itself by demonstrating that it can deliver applicable results it can have a golden future in livestock science and livestock industry.

Index

Note: Italicized page locators indicate figures/photos; tables are noted with *t*.

Systems Biology and Livestock Science, First Edition. Edited by Marinus F.W. te Pas, Henri Woelders, and André Bannink.