Animal, page 1 of 12 © The Animal Consortium 2011
doi:10.1017/S1751731111000036

animal

Systems biology in animal sciences

H. Woelders[1†], M. F. W. Te Pas[1], A. Bannink[2], R. F. Veerkamp[1] and M. A. Smits[1]

[1]*Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, PO Box 65, 8200 AB Lelystad, The Netherlands;* [2]*Animal Nutrition, Wageningen UR Livestock Research, PO Box 65, 8200 AB Lelystad, The Netherlands*

(Received 16 April 2009; Accepted 21 July 2010)

Systems biology is a rapidly expanding field of research and is applied in a number of biological disciplines. In animal sciences, omics approaches are increasingly used, yielding vast amounts of data, but systems biology approaches to extract understanding from these data of biological processes and animal traits are not yet frequently used. This paper aims to explain what systems biology is and which areas of animal sciences could benefit from systems biology approaches. Systems biology aims to understand whole biological systems working as a unit, rather than investigating their individual components. Therefore, systems biology can be considered a holistic approach, as opposed to reductionism. The recently developed 'omics' technologies enable biological sciences to characterize the molecular components of life with ever increasing speed, yielding vast amounts of data. However, biological functions do not follow from the simple addition of the properties of system components, but rather arise from the dynamic interactions of these components. Systems biology combines statistics, bioinformatics and mathematical modeling to integrate and analyze large amounts of data in order to extract a better understanding of the biology from these huge data sets and to predict the behavior of biological systems. A 'system' approach and mathematical modeling in biological sciences are not new in itself, as they were used in biochemistry, physiology and genetics long before the name systems biology was coined. However, the present combination of mass biological data and of computational and modeling tools is unprecedented and truly represents a major paradigm shift in biology. Significant advances have been made using systems biology approaches, especially in the field of bacterial and eukaryotic cells and in human medicine. Similarly, progress is being made with 'system approaches' in animal sciences, providing exciting opportunities to predict and modulate animal traits.

Keywords: systems biology, mathematical modeling, animal traits

Implications

The recent progress in high-throughput omics technologies allows the rapid identification of all the cellular components relevant to animal traits. This paper explains how systems biology can be used to analyze and integrate these data in order to find and understand the biological mechanisms that arise from the dynamic interactions of these components. Some important concepts of systems biology are explained and potential fields of application in animal science and a number of examples thereof are given. This could contribute to stimulation of the application of systems biology approaches in animal science.

Introduction

Systems biology is an integrative approach, looking at the whole system as working together, rather than to its individual components. Although some disciplines in biology have historically used such an integrative approach, two developments have recently given a strong impetus to the integrative study of biological systems. These two developments are the strong increase in our abilities to gather data on expression of genes and on the presence and interactions of gene products, and the exponential increase of the computational power with which we can analyze and try to understand these data. Systems biology is a rapidly expanding field of research and is applied in a number of biological disciplines. In animal sciences, omics approaches are increasingly used, yielding vast amounts of data, but systems biology approaches to extract understanding from these data of biological processes and animal traits are not yet frequently used. This paper aims to explain what systems biology is and which areas of animal sciences could benefit from systems biology approaches.

What is systems biology?

Various approaches in systems biology

Researchers from various scientific backgrounds and expertise value various aspects of systems biology differently.

† E-mail: henri.woelders@wur.nl

Some biologists emphasize that systems biology is the use of statistics, bioinformatics and computational modeling to analyze, filter, combine, and integrate mass 'omics' data in order to extract knowledge from the huge data sets obtained from laboratory experiments. Others emphasize that it includes the building of mathematical models in which the dynamic (kinetic) relationships between the system components are represented by mathematical equations, allowing *in silico* simulations of the system response. Again others emphasize that systems biology is characterized by an iterative cycle of hypothesis formulation – laboratory experiments – and renewed hypothesis formulation.

The common notion is that, even if we would be able to know all components of a system but not describe their interactions, we would not be able to understand how biological systems work, or how we can predict the behavior of biological systems. Only by knowing and describing the interactions between system components, the function and performance of (sub) systems can be understood. Thus, systems biology can be described as the study of the emergence of functional properties that are present in a biological system but not in its individual components.

Top-down and bottom-up approaches

Many papers on systems biology distinguish so called 'top-down' and 'bottom-up' approaches (e.g. Kitano, 2000; Palsson, 2002; Bruggeman and Westerhoff, 2007). The distinction is illustrated in Figure 1. Bottom-up refers to constructing a mathematical model of the system from (a selected number of) its components of which the properties are known, similar to engineering. The direction of working is from known components upward toward the system. In a top-down approach, one may not *a priori* know which components of the system are specifically important to look at. Therefore, one starts at the top of the system (e.g. an animal) and by introducing a change or perturbation to the system (e.g. application of an infectious agent), one takes an as broad as possible top-down view of the response of the system (e.g. by measuring changes in transcriptome or proteome profiles). The direction is downward, from the system toward components. Top-down approaches are often referred to as 'reverse engineering' (e.g. Palsson, 2002). Bottom-up approaches are an example of deductive reasoning: from known or assumed properties of the components one deduces system functions. Top-down approaches are an example of inductive reasoning: from multiple observations on how the system reacts to perturbations one infers which components have a critical effect and how the system may function.

Modeling

Modeling of a biological system may be done with increasing levels of complexity and completeness of the model (Figure 2). As a first step, the components of a system can be identified merely as being present or absent, for example, on the basis of the gene expression profile of a particular cell type. Then, correlations between specific expression profiles resulting from various perturbations of the system can be used in principle component analyses, whereas Bayesian models can be used to identify regulatory or priority relationships. The regulation of the responses of a biological system to various perturbations can be described in terms of gene networks and metabolic networks, for example, in qualitative modeling of regulated metabolic pathways (e.g. Simao *et al.*, 2005). Predictive whole-cell metabolic models

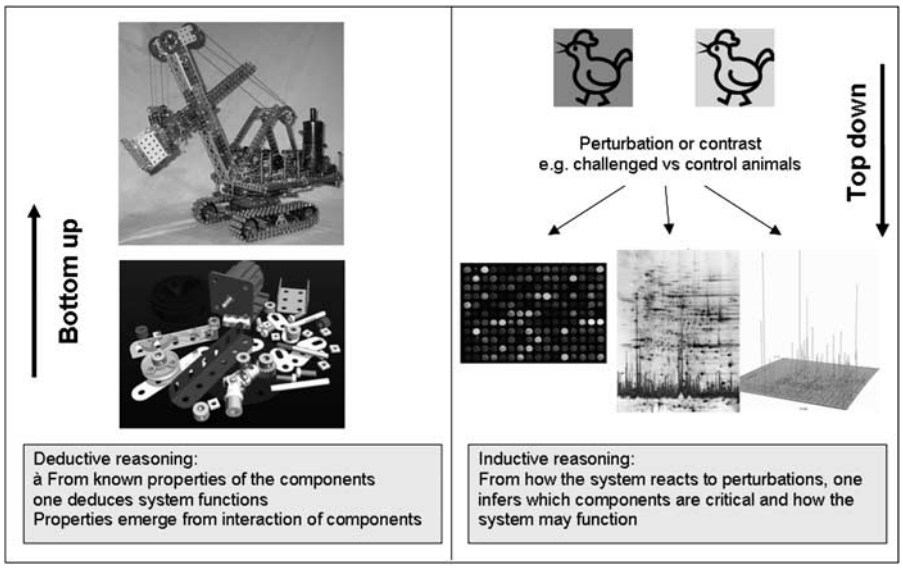

Figure 1 Bottom-up and top-down approaches in systems biology. Left panel: bottom-up construction of a model from (a selected number of) its components of which the properties are known, similar to engineering. Bottom-up approaches are an example of deductive reasoning: from known or assumed properties of the components one deduces system functions. Properties emerge from interaction of components. Right panel: one may not *a priori* know, which components of the system are specifically important to look at. Therefore, in top-down approaches, an as broad as possible (genome wide, proteome wide, etc.) top-down view is taken of the system. From the response of the system to perturbations one infers, which components have a critical effect and how the system may function.

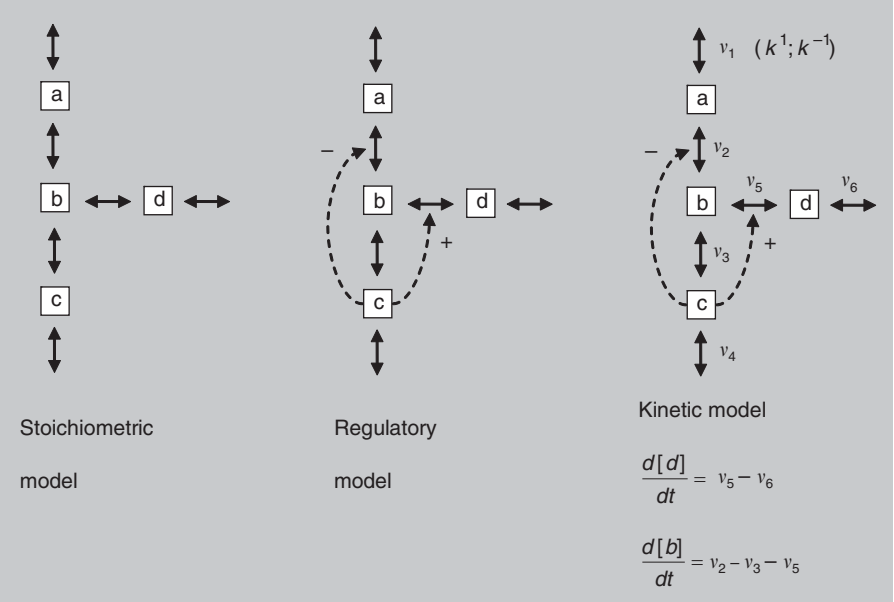

Figure 2 Modeling of a biological system may be done with increasing levels of complexity and completeness of the model. For clarity, a very simple system is shown, consisting of only four components (a to d). Solid arrows denote production of one component from another. In the case of metabolites, for example, the solid lines could denote enzymatic reactions. As a first step, the components of a system can be identified merely as being present or absent, for example, on the basis of the gene expression profile of a particular cell type (stoichiometric model). By including positive or negative regulatory influences (allosteric or regulation of gene expression; dashed arrows) one can arrive at a qualitative regulatory model. Finally, kinetic models may be constructed to simulate system behavior quantitatively. This requires that the (kinetic) properties of the relevant components of the system are known in sufficient detail.

can be developed using constraints-based modeling that uses the successive imposition of governing constraints (such as mass conservation, thermodynamics, capacity and nutritional environment) to eliminate network functions that fall outside the range of constraints, which govern the system (see section 'Some examples of modeling approaches in single cells'). Finally, kinetic models may be constructed to simulate system behavior. This requires that the (kinetic) properties of a large enough number of components of the system are known in sufficient detail. Components could be genes, mRNAs, molecular regulators, hormones, enzymes, receptors, metabolites, biochemical or signaling pathways, membrane pumps, cell organelles and possibly cells, tissues, organs, etc., depending on the organizational levels to be represented in the model. A major problem is that the (kinetic) properties of molecular components may be influenced by their immediate microenvironment. Enzyme activity data, for example, may be derived from studies with purified enzymes (e.g. in an artificial *in vivo*-like medium, van Eunen *et al.*, 2010), but may also be inferred from measurements of cell lysates, complete cells or organelles. Present high resolution imaging techniques (Megason and Fraser, 2007) may help to determine some of the kinetic relationships inside living cells (Shav-Tal *et al.*, 2004).

Part of mechanistic models may be black boxes. For instance, cell organelles, cells or even organs may be the principal building blocks of a model system, as long as the kinetic properties of that building block are known or can be assumed. In this way, the model may contain 'lumped' rate constants or 'lumped' fluxes through a pathway, a cell organelle or an organ (e.g. Snoep, 2005).

Iterative and integrative

Frequently, systems biology is characterized by iterative cycles of hypothesis formulation – laboratory experiments – data analysis – model building – *in silico* validation – renewed formulation of hypotheses – re-modeling – and experimental testing of systems behavior. Iterative cycles may include top-down approaches combined with mechanistic (bottom-up) modeling. For example, as indications are obtained from top-down approaches as to which system components could be relevant for proper functioning, these components may be included in a subsequent mechanistic (bottom-up) model. Both approaches are therefore complementary.

One of the features of systems biology is that it strives for integration of information from many different sources and possibly from different levels of biological organization. Integration of genomic, transcriptomic, proteomic and metabolomic data from laboratory experiments as well as information from various databases, can bring the description of cellular regulation beyond that of gene networks (e.g. Ideker *et al.*, 2001). The inclusion of higher levels of biological organization up to the organ or organism level is, however, a mighty challenge. For example, a systems biology approach of infectious diseases could aim to span levels of organization from the molecular level (pathogen sensing receptors) up to the population level, and span a time scale from days (innate immunity) to decades (lifelong protection, Young *et al.*, 2008). As explained in section 'Modeling', at the higher levels of organization, the building blocks of a model do not necessarily include all the components of the subcellular or molecular level. One example is the elaborate mathematical modeling of the human menstrual cycle by

(Reinecke and Deuflhard, 2007) in which the model components may be organs or tissues with defined overall ('lumped') kinetic properties, whereas in other parts of the model a more detailed description up to the level of hormone receptors and intracellular signaling is used. Examples of mechanistic modeling on the level of organs can also be seen in the modeling of human organs in the International Union of Physiological Sciences (IUPS) Physiome Project (www.physiome.org.nz/), for instance, the mechanical and anatomical/physiological modeling of the human heart (Noble, 2008). Although the ambition of the Physiome project is to span and integrate levels of organization ranging from molecular components to organs and beyond, it is acknowledged that no one model can encompass this wide range of organization levels in all detail (Hunter and Nielsen, 2005). Instead, the ambition is to establish a framework for handling a hierarchy of computational models in which the parameters of a particular model in the hierarchy can be understood in terms of the physics or chemistry of the appropriate model(s) at a lower hierarchical level.

Databases and computational platforms

A strong impetus for the current interest in systems biology is the exponential growth of data. During the last two decades, a steadily increasing amount of genomic information has become available (http://www.ncbi.nlm.nih.gov/ and http://www.ebi.ac.uk/). The sequencing of whole genomes and the development of high-throughput methods to screen vast numbers of gene transcripts, proteins or metabolites isolated from tissues delivers great opportunities to identify the majority of systems components and their spatial and temporal dynamics. The steady increase of computational power and miniaturization of analysis equipment further nourishes the drive to accumulate even more data. The technological-driven developments and the gathering of ever increasing amounts of data into ever more detail, created the need to develop methods to explain the meaning of all these data and to analyze the organization of the biological system under consideration. Indeed, much emphasis is currently put on both high-throughput data acquisition as well as mathematical modeling (www.nature.com/focus/systemsbiologyuserguide).

Without the aim to be exhaustive, we briefly identify a few general tools required for systems biology approaches. First of all, software tools are required for the statistical analysis of raw omics data and the systematic storage of data. Other tools are necessary that allow access and the use of knowledge present in a variety of databases like Gene Ontology, Kyoto Encyclopedia of Genes and Genomes, BioCarta, Pathway Commons, etc. Again other software tools are required to build and analyze regulatory networks and to connect them into larger networks (e.g. Osprey, Cytoscape, Ingenuity Pathway Analysis, Acuity, MetaCore, etc.). Various tools are available for mechanistic modeling of cells. Examples are Virtual Cell (www.nrcam.uchc.edu), E-Cell (http://www.e-cell.org/) and Silicon Cell (http://www.siliconcell.net). The first two are modeling environments for cell biology that are used to calculate what happens in cells. Silicon Cell is not a real package of software for simulations, but rather a database of kinetic models for specific pathways that can be interrogated over the Internet (Snoep, 2005). With regard to modeling at the organ level, the website of the Physiome consortium (www.physiome.org.nz/) provides links to various relevant databases and computational tools.

The systems biology markup language (SBML, Finney *et al*., 2006) was developed to allow a number of existing simulation software packages to communicate with each other, and to enable the exchange of (parts of) software developed in one tool to software developed in another tool. The SBML webpage http://sbml.org/index.psp lists over 100 software systems and databases of biological models.

Applications in systems biology

Some examples of modeling approaches in single cells

Prokaryotes, single-celled eukaryotes and well-defined cultures of mammalian cells are attractive for systems biology approaches, as they lack the extra layers of complexity that multi-cellular organisms have. Various bacteria and fungi are economically important organisms, as they are used for the production of a variety of biochemicals. Genomes from various production micro-organisms, a variety of bacterial pathogens, and from a number of commensal, symbiotic and environmental micro-organisms have been sequenced now (Pallen and Wren, 2007, http://xbase.bham.ac.uk/taxon.pl). Systems biology approaches are used to understand the metabolism of production organisms in order to improve product output or to identify potential drug targets in pathogenic micro-organisms. Similarly, systems biology approaches are being used with cell cultures of mammalian cells to study important processes, like signaling, relevant to understanding cellular function and dysfunction, for example, in immunology and cancer research. A few pertinent examples of such modeling studies in single cells are presented in the following three sections.

Metabolic networks in bacteria; constraint-based modeling

One approach to integrate genomic and other omics data to predict system function in micro-organisms is to derive constraint-based genome-scale metabolic models (Joyce and Palsson, 2006; Schuetz *et al*., 2007; Raghunathan *et al*., 2009). This is an example of a 'top-down' approach, as the possible ways that a cell could function are inferred from genome-wide data sets.

Constraint-based modeling starts with assembling as much as possible knowledge on the genes and other cellular components. From the available annotated genome and other omics data, gene–protein–reaction relationships are described. The equations for most metabolic reactions and their stoichiometries are available, as well as information on the function and location of the reactions. This allows the construction of metabolic models that represent almost entire microbial genomes. With up to 1000 biochemical reactions, these genome-scale models allow to predict network function, for example, by using flux balance analysis (Fell and Small, 1986).

The constraint-based metabolic model of *Salmonella enterica* serovar typhimurium reported by Raghunathan *et al.* (2009) contains 1083 genes, 973 proteins, 744 metabolites and 1087 metabolic reactions, of which 1018 are gene associated. Estimates for the metabolite concentrations under the condition of steady state were obtained from literature or from experimental measurements. A network was built using gene to protein and protein to reaction associations, and flux balance analysis was performed using the biomass production reaction as 'objective function'. The model finds various alternate optimal solutions of flux distributions, all achieving maximal biomass yield and flux variability analysis was additionally used to identify the ranges individual fluxes can take within this constraint. The study allowed to simulate and predict the utilization of various carbon and nitrogen sources and to predict the effect on various mutations (gene knockouts) on the growth and virulence phenotype of the bacterium under various conditions, including the conditions inside host cells. These predictions were found to be in line with experimental data obtained in the same study or taken from literature. Similar models have been constructed for other bacteria, for example, *Escherichia coli* (Schuetz *et al.*, 2007; Feist *et al.*, 2010).

Signaling in dendritic cells (DCs); gene regulatory networks
Gene regulatory network modeling is another way to organize and analyze complicated cell systems on the basis of transcriptomic data. The aim is to identify the key genes that orchestrate specific physiological responses of cells and to analyze how they are connected to each other, and to effector genes that generate specific cell responses. A recent study of immune responses of mammalian (mouse) DCs (Amit *et al.*, 2009) first identified putative regulator–target relations on the basis of correlated expression and then studied the effect of systematic perturbation of the regulator genes. Gene expression profiles were made of DCs at nine time points after stimulating them with five pathogen derived 'bacterial' and 'viral' ligands for a number of Toll-like receptors (TLRs), and specific and shared genes that respond to each stimulus were identified. These profiles were used to identify 144 candidate regulator genes, as well as a signature of 118 marker genes that captured the complexity of the response. Then, a systematic perturbation study was done by knocking down candidate regulator genes in DCs by RNA interference. The cells were then stimulated and the expression of the 118 marker genes (and 10 control genes) was profiled. The changes in signature gene expression resulting from knocking down specific candidate regulator genes were used to associate regulators to their targets. The emerging picture shows the complexity of TLR-mediated sensing and signaling in DCs, with effectors (activators or suppressors) being connected to many different targets in feed-forward and feedback loops. Feed-forward circuits respond to persistent rather than transient stimulation, protecting the system from responding to spurious signals. A total of 13 'known' as well as 11 'new' key factors (hub genes) of inflammatory or antiviral responses were identified. Twelve of the identified regulators are in linkage disequilibrium with single nucleotide polymorphisms (SNPs) associated with autoimmune and related diseases in genome-wide association studies.

Kinetic modeling of cellular metabolism
As biological systems most often comprise many coupled (often non-linear) relations, mathematical modeling could be necessary to correctly predict its behavior. Erroneous outcomes could be obtained if one attempts to predict system behavior from intuitive reasoning or qualitative modeling. The outcome of simulations done with a kinetic mathematical model are quantitative and 'correct', that is, these outcomes are true for the model as it is defined.

Thus, kinetic modeling can predict system behavior and functional properties that cannot be recognized from the individual properties of the system components. A nice example is seen in the studies on glycolytic pathways in *Trypanosoma brucei* (Haanstra *et al.*, 2008). In these studies, an 'exact' kinetic model of a part of intracellular energy metabolism was used, comprising a large number of coupled differential equations that contain (Michaelis–Menten) kinetic descriptions of metabolic reactions. The model unexpectedly predicted that the glucose transporter is the best drug target candidate, rather than glyceraldehyde-3-phosphate dehydrogenase, which is the drug target that is worked on most. The model explained that *T. brucei* needs its glycosome, an organelle that contains the glycolytic enzymes, to prevent a lethal increase ('explosion') of phosphorylated intermediates of glycolysis. This prediction was confirmed experimentally.

Another example of kinetic modeling of cellular metabolism is a study on signaling pathways in cell cultures of human cells (reviewed in Kholodenko, 2006). The mechanistic model shows how interactions of components lead to oscillations and other forms of spatial and temporal dynamics of signaling molecules. These dynamics confer the specificity in the signaling message in transduction pathways in which different receptors are connected to different cellular responses through the same signaling intermediates. These studies indicated the relevance of receptor tyrosine kinase signaling for major human diseases ranging from developmental defects to cancer to chronic inflammatory syndromes and diabetes.

Human health-related systems biology, on a higher than cellular level
Many pharmaceutical companies apply systems biology approaches for drug development and drug testing purposes (Aksenov *et al.*, 2005). They combine various omics data, obtained from biopsies of organs in preclinical (animal) models, with physiological and pathobiological data and use these in Bayesian networks-based inference frameworks to characterize the molecular pathways affected by a compound of clinical interest. In addition, in cancer research, there is much interest in using systems biology approaches, whereby top-down inference methods and mechanistic modeling are combined, using both *in vitro* and *in vivo* data.

This methodology has great potential for unraveling cancer disease mechanisms and for devising effective therapeutics (Khalil and Hill, 2005).

In order to address the influence of the genetic make-up of individual patients in the susceptibility to multi-factorial diseases, it is necessary to investigate how genes interact to produce phenotypes. As there is little statistical power to detect interactions between genes in human population studies, interactions of mutations and their effect on phenotypes are usually determined in model organism like *Caenorhabditis elegans*. Such studies aim to understand how specific mutations affect phenotypes, and to construct systematic genetic interaction networks that provide insights into the susceptibility of humans to diseases (Lehner, 2007). For example, Schadt *et al*. (2005) described a multi-step statistical procedure, integrating DNA variation and gene expression data with complex data on traits, that can predict whether genetic variations are independent, causative or reactive relative to a certain trait. With this method they identified three new obesity-related genes.

Systems biology in human medicine may involve modeling on various levels of biological organization, up to the population level (e.g. Young *et al*., 2008). For instance, models that can be used to understand a disease on the level of an organ or entire organism, and to predict the efficacy of a drug or treatment ('the virtual patient') can be of great value during the development of new treatments or drugs (Klauschen *et al*., 2007). In this context, physiological models of cancer growth and therapy have been used to suggest optimal chemotherapeutic regimens in breast cancer (Arakelyan *et al*., 2002). Similarly, a model of the heart was developed to characterize the pathophysiology underlying electrocardiographic dysfunction and predict the actions of drugs (Noble, 2008). This is part of the IUPS Physiome project, mentioned earlier.

Systems biology in animal sciences

Data explosion in animal sciences

The animal sciences are in the midst of a data explosion. Complete genome sequences of economically important livestock species have been determined now (http://www.ebi.ac.uk/, http://www.ncbi.nlm.nih.gov/, http://www.ark-genomics.org/ and http://www.animalgenome.org) and projects are underway to sequence genomes of individual animals. Furthermore, information on the genetic variability of livestock genomes at the level of SNPs and copy number variation (CNV) is rapidly expanding, as methods for the detection of several ten- to hundred-thousands of SNPs, CNVs and epigenetic variations in hundreds of individual animal genomes are continuously improving. For the major livestock species 50 to 60 K SNP chips are commercially available now and the number of features per array will certainly increase soon. The availability of new ultra-high-throughput genome sequencing machines using massively parallel sequencing approaches increase the speed and capacity of sequencing and the discovery and measurement of genetic variation enormously (Fox *et al*., 2009). Application of these tools results in enhanced data throughputs at decreasing costs per animal. With these new technologies, the genetic potential of individual animals can be documented and, theoretically, the complete genetic potential for traits can be assessed.

A similar increase in throughput capability is seen in the area of functional genomics applications in livestock species, which allow the identification of cellular components resulting from transcription, translation, protein interactions and metabolic pathways. In addition, studies on supracellular levels provide data on system parameters like protein secretion, ligand-receptor interactions, cell communication, proliferation and differentiation.

DNA microarrays that contain ten thousands of probes, representing the genes encoded by the whole genome, have become commercially available for all major livestock species since a couple of years. These microarrays allow the simultaneous measurement of the transcriptional activity of all genes. Consequently, mRNA expression data are rapidly becoming available for many different tissues of livestock species and for a multitude of environmental conditions (see Gene Expression Omnibus (GEO) repository, http://www.ncbi.nlm.nih.gov/sites/entrez?db=geo). Gene expression profiling by transcriptome analysis has proven to be a very powerful tool in studying the biology of traits. Such studies have identified (new) genes, molecular (signaling) pathways and regulatory networks involved in a variety of biological processes that are associated with animal- or breed-specific traits. Current developments in massively parallel sequencing approaches and in the array-based platform technologies will further enhance our abilities to discover and profile mRNA and regulatory RNA species (Yashiro *et al*., 2009) with ever increasing throughput and at reducing costs. These developments will further increase our insights into the spatio-temporal dynamics of livestock gene expression.

Technological advances in chromatography and mass spectrometry have also led to an increased throughput in large-scale protein analysis (proteomics) and metabolite analysis (metabolomics). These technologies are increasingly applied in livestock research (reviewed by Lippolis and Reinhardt, 2008). Similarly, high-throughput functional assays can provide massive data on system parameters that go beyond the level of cellular components (Kittler *et al*., 2008; Wunder *et al*., 2008). Furthermore, advanced statistical and integrative approaches are emerging to investigate correlations between the expanding omics information and livestock traits.

Rationale for the use of systems biology in animal sciences

High-throughput technologies allow the animal sciences to identify and characterize the molecular components of traits, and the variation therein, with ever increasing speed. However, traits and quantitative aspects of traits do not simply arise from the sum of the properties of individual components of the 'trait system' under investigation but depend on dynamic interactions between these components at various biological levels. As indicated in sections 'What is systems

biology?' and 'Applications in systems biology', the prospects of obtaining a better view of the behavior of biological systems is now beginning to emerge through the application of systems biology. The development of predictive mathematical models representing the dynamic interaction between (molecular) components of trait system will allow *in silico* simulations of trait-system responses. An important factor hereby is to know how various parameters of the production environment (housing, nutrition, climate, pathogen load and stress) and the animal's genome affect these interactions and the resulting behavior of the trait system. The rationale of the application of systems biology in animal sciences is the development of predictive models of animal (sub)traits that help to understand the biology of traits and that can be applied for the prediction, modulation and improvement of traits (Quackenbush, 2007).

Mathematical models are already being applied in the practice of animal husbandry for some time. However, these models do not address the full complexity of trait systems; they usually predict the effect of only a single type of variable, for example, diet or quantitative trait loci (e.g. Bannink *et al.*, 2006). Integration of existing models with omics-based knowledge may lead to significant improvements of their predictability and applicability, and we expect that such improved models can cope with a much broader spectrum of environmental and genetic variables. We anticipate that omics-based models will contribute to innovation in livestock husbandry and will lead to improvement of animal traits within a timeframe of approximately 3 to 5 years.

Differences in the objectives of systems biology

Systems biology approaches in animal sciences differ from those in humans and model organisms. Model organisms are used to unravel universal fundamental biological processes of cells. In most cases, model organisms are more attractive for studying the effects of multiple variants (mutants and strains), multiple generations and multiple environmental conditions. Systems biology approaches in humans mainly aim to understand the perturbation of a normal functioning system, that is, the development of disease. Disease involves an impairment of one or multiple system components resulting into a change or loss of a particular functional property. Human systems biology contributes in developing tools and methods that help to transform the current 'reactive' practice in medicine, into a more predictive, preventive and personalized approach.

Systems biology in animal sciences is widely oriented, compared with human health research, in that it addresses improvement, rather than disturbances of existing trait systems. Traits of interest are not only related to health, but also to production and quality parameters, welfare, robustness, environmental footprints, etc. In animal sciences, therefore, systems biology is applied to develop predictive models that aid in the improvement of health-, product- and sustainability-based (sub) traits. Such models require an understanding at a rather high biological level of organization, far from the level at which omics data are gathered. As in the human field, systems biology can contribute to more predictive veterinary practices. Specific benefits would be in assessment of the probability of animals to develop disease, selection of animals adapted to specific health management programs, early warning for disease, and identification of new targets for diagnosis and treatment. It should be noted here that herd approaches rather than individual treatments have to be adopted. The major challenge of the application of system biology in animal sciences is to go beyond prevention and prediction and to arrive at a stage where we can further improve animal traits like disease resistance, product quality and fertility.

Advantages of using livestock

Systems biology approaches focusing on animal traits can take full advantage of the availability of a variety of divergently selected lines that differ quantitatively in specific traits. Such animal populations have often been well-characterized in terms of measurable traits, like milk yield or disease incidence, and may provide a valuable resource for discovery and validation research. For example, variation in genetically determined resistance to infectious diseases is found in all major livestock species (Gibson and Bishop, 2005). Moreover, for all major livestock species, divergently selected lines exist that have been selected specifically for disease susceptibility traits. Variability in disease resistance is frequently reflected by differences in gene expression patterns (e.g. van Hemert *et al.*, 2006) and in the activity of specific (signaling) pathways (e.g. Te Pas *et al.*, 2008). Integration of such knowledge with large-scale genetic (SNP), physiological, immunological and/or metabolic data will help to understand both the genetically as well as the environmentally induced mechanisms of disease resistance.

Another advantage of the use of farm animal species is the existence of extended animal recording programs, archiving phenotypic performance data for management, genetic, reproductive, health and economic purposes. Such recordings have, for example, allowed the development of sophisticated genetic improvement programs that use advanced statistical methods to predict breeding values of individual animals (see section 'Perspectives of systems biology in animal sciences'). Systems biology approaches will further benefit from the availability of large numbers of phenotyped animals, although for a comprehensive 'trait-system' analysis, additional phenotypic data at deeper physiological levels as well as a standardized ontology of animal phenotypes will be required.

Finally, compared with humans, it is much easier to obtain biological material, taken at specific time points and time intervals and from specific parts of the body, from animals kept under well-controlled and well-monitored conditions, and from animals in which specific traits are deliberately perturbed in a well controlled manner. Furthermore, large collections of biological material such as blood, eggs and milk, from a large number of animals with well-documented management and performance recordings are readily accessible. Nevertheless, the rising demands for access to specific

biological materials may require the set up of biobanks (Asslaber and Zatloukal, 2007).

Contrasting these advantages for systems biology in animal sciences, some specific drawbacks are the limited number of commercially available biological tools and research kits, the almost complete absence of congenic knockout and knockdown mutants, and the relatively high costs of animal experimentation.

Perspectives of systems biology in animal sciences

In the current practice of animal husbandry, performance traits are usually monitored by end-point measurements (e.g. milk yield and weight-gain), indicating that current practices are 'reactive' in nature. However, through application of high-throughput technologies as described in section 'Data explosion in animal sciences', animal husbandry can get access to sophisticated tools and methods for multi-target and multi-parameter measurements to identify and characterize the molecular components of trait systems. The knowledge derived from such measurements can be used to generate accurate and comprehensive predictions of performance characteristics of animals kept under normal or specified conditions. One of the first and very successful applications of this type of 'predictive biology' in animal sciences is 'genomic selection'. 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 ($>$50 000) of consecutive genome fragments and phenotypical performance as established in a reference population (Meuwissen *et al.*, 2001; Calus *et al.*, 2008). Several breeding industries already apply genomic selection procedures as it accelerates the rate of genetic gain considerably. The next step in this area is to use the increasing knowledge on genotype–phenotype relationships for the development of precision mating systems to maximize the use of non-additive variation of traits and to minimize cumulative effects of inbreeding. Another application is the sorting of animals, which are best equipped for optimal performance under defined environmental and management conditions.

With the genetic approaches described above, researchers are able to use correlations between genotypes and phenotypes without the need to understand the underlying biological mechanisms. Using the high-throughput data of functional genomic studies, researchers can go one step further and try to understand the spatial- and temporal-dependent biological mechanisms leading to traits. As indicated in sections 'What is systems biology?' and 'Applications in systems biology', the most promising approach to translate the molecular composition of trait systems into meaningful biological information, is through the application of the methods developed in the systems biology arena. Knowledge on the relationships between molecular composition, biological mechanisms and the behavior of trait systems allows the identification of 'molecular signatures' that can be used to monitor trait development. This will provide farmers with information that can be used during the production phase to optimize management to support and maintain normal homeostasis. Knowledge on the relationships between molecular composition, biological mechanisms and the behavior of trait systems can also be used to design knowledge-based treatments or intervention strategies to prevent impaired health, to improve the quality of livestock end products, to decrease environmental impact, etc. A practical example, which is the subject of an increasing number of research projects worldwide, is the development of nutritional strategies to modulate the activity of the intestinal innate immune system to enhance disease resistance. Furthermore, improved understanding of the biology of traits may also aid in the identification of genomic variation causally related to specific traits. Finally, systems biology may also provide a framework to address the interaction between different biological mechanisms in a particular tissue or between tissues. A practical example in this context is the general observation that the regulation of the synthesis of fatty acids is related to the regulation of innate immune responses (Sordillo *et al.*, 2009), suggesting a relationship between milk fatty acid composition in milk and udder health. It is currently unknown whether and how these two biological systems interact with each other in mammary epithelial cells of cattle.

Examples of projects that benefit from a systems biology approach

Some specific aspects of carefully chosen animal traits have arrived at a research stage at which it is attractive to proceed with a systems biology approach. Disease susceptibility, for example, requires understanding of the interactions between host and pathogen at the cellular level. In this area, a wealth of relevant gene expression data is available on the basis of which regulatory gene networks can be built. Such networks identify the key driver molecules that determine the behavior of the host–pathogen interaction system under investigation. For example, as described in section 'Some examples of modeling approaches in single cells', Amit *et al.* (2009) generated transcriptional networks to identify core regulator and fine-tuner molecules of dendritic mammalian cells that interact with a variety of pathogens. This was of great help in explaining how pathogen-sensing pathways achieve specificity and sensitivity and to identify the basic driver molecules of crucial cellular functions of DCs. Quantitative data of these key driver molecules are important parameters to be included into mathematical models that capture and predict the behavior of DCs upon exposure to a specific pathogen. As the main function of DCs is to process antigen material and present it on the surface to other cells of the immune system and act as messengers between the innate and adaptive immunity, such models might be of critical importance for the development of 'knowledge-based' vaccines to prevent infectious disease.

In our own institute, we identified several areas in our research programs that could benefit from a follow-up by a systems biology approach. In one area, we aim to mathematically model a specific aspect of intestinal health, namely

the intestinal efficiency of chicken to eliminate *Salmonella*. To get a genome-wide view of the changes that take place in the intestine after a challenge with *Salmonella*, we measured gene expression in a time series (over 3 million data points; Schokker *et al*., 2010). Using GeneNet (R software package), we inferred gene regulatory networks for the normal and *Salmonella* infected condition (Figure 3). This allowed the identification of so-called hub genes, the central elements in gene regulatory networks (Schokker *et al*., 2010). As these hub genes have the potential to orchestrate an array of intra and intercellular processes in the intestine, their activities are important parameters to be included in the equations of the mathematical model we are working on. Other necessary components of this model will be derived from other biological levels of the intestinal trait system and include the influx, number, and activity of various different immune cells and the rate of synthesis of various cytokines. Quantitative measurements of these components have been taken and will allow critical parameter estimations.

In another area, we aim to construct a mathematical model that predicts a product quality trait, that is, the secretion of unsaturated fatty acids from bovine mammary epithelial cells. To this end, we are adopting an existing dynamic model (Shorten *et al*., 2004) and are adding missing elements that describe and explain in more detail the regulation of milk fat composition. These regulatory elements were discovered using transcriptomic measurements in bovine mammary tissues from biopsies taken at various stages of lactation (Bionaz and Loor, 2008) or from cows differing in their milk fatty acids composition (Mach N. *et al*., manuscript in preparation). These elements include key transcription regulators and their ligands involved in controlling fatty acid metabolism, specifically the transcription factors sterol regulatory element binding protein-1 and -2, and the peroxisome proliferator-activated receptor-γ. In addition, we are also exploring how to represent genetic factors that are known to affect milk fatty acid composition, like diacylglycerol acyl transferase and stearoyl-coenzyme A desaturase 1, as an additional variable to the model. In the future, the functionality of this model will be evaluated against independent experimental observations on nutritional, physiological and genetic control of milk fat composition. A validated model should allow the simulation of the effects of system inputs (fatty acids and regulatory molecules absorbed from blood) on system outputs (fatty acid composition in milk) and the interdependencies between its key components (fatty acid metabolites, ligands, gene variants, gene activities, etc.). To allow predictions of milk fatty acid composition based on nutritional inputs, such an udder epithelium model may be combined with models representing aspects of liver and adipose tissue, and with the already available 'rumen' model (Bannink *et al*., 2006). The latter is a mechanistic model predicting the rates of production of volatile fatty acids in the rumen, the hydrolysis of fat and the hydrogenation and transfer of fatty acid isomers. This model describes the interdependencies between types of feed substrate, fatty acids and micro-organisms present in the rumen of dairy cattle.

A third line of research is in the area of female fertility in dairy cattle. Fertility is a very important trait in animal production generally, and in dairy cattle in particular. Concurrent with an increase of milk yield, a decrease in dairy cow fertility has been observed during the last decades (Veerkamp *et al*., 2003; Pryce *et al*., 2004). This decline in fertility is manifested in alterations in hormone patterns

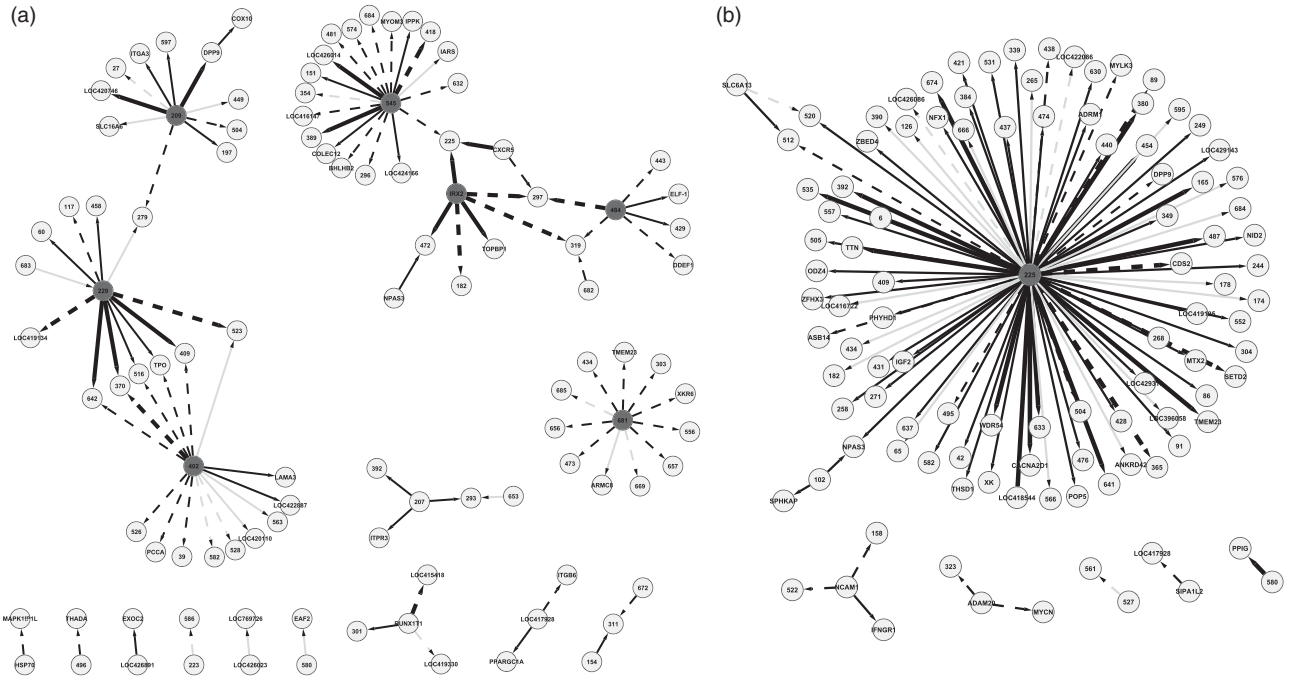

Figure 3 Gene regulatory networks of genes of intestinal tissue that reside in the top 100 most significant edges for (a) control and (b) *Salmonella* infected chickens.

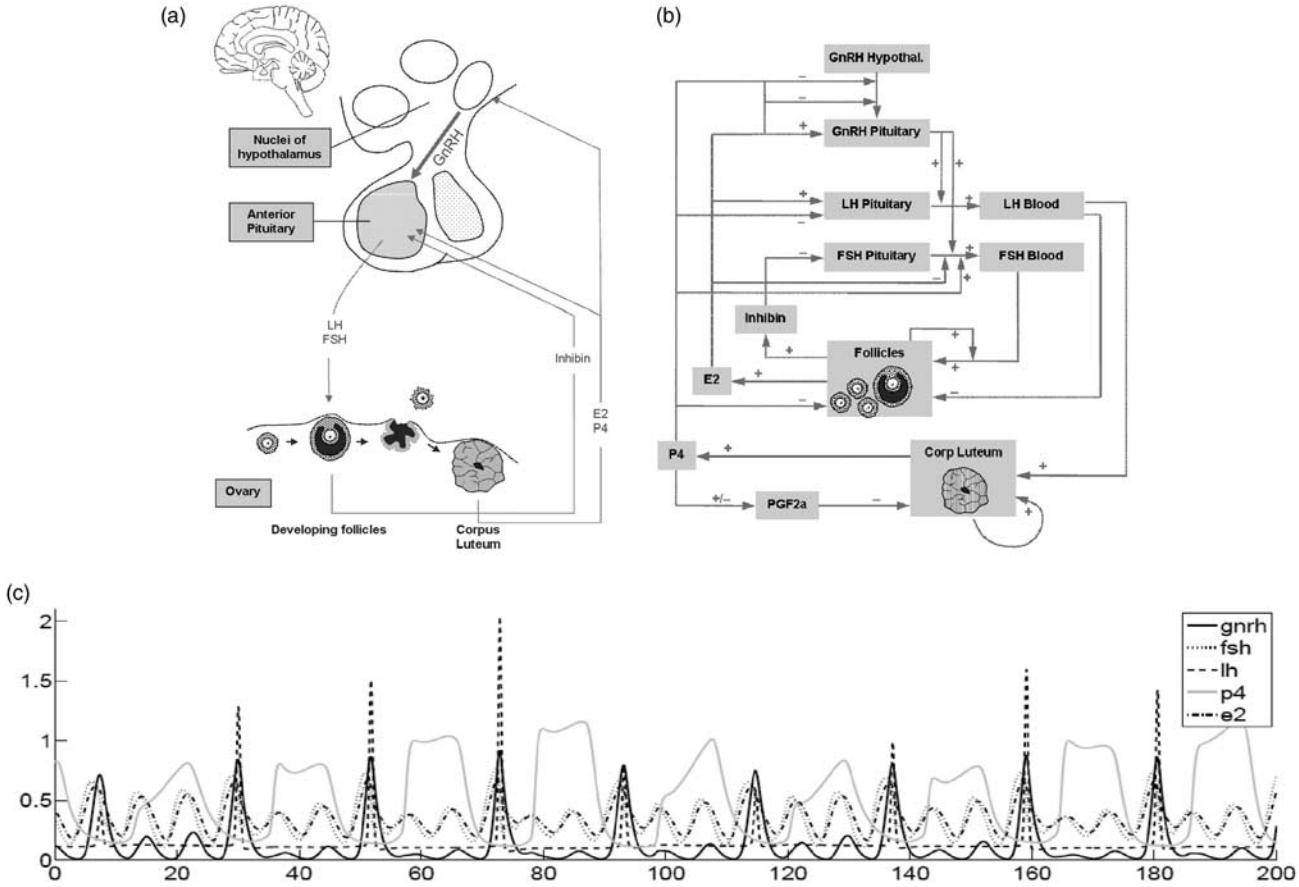

Figure 4 Modeling of bovine estrous. (a) The regulation of estrous is controlled by the complex interplay of various organs and hormones. (b) Owing to the various positive and negative feedback loops, it is very difficult to predict the behavior of this system from the known qualitative relations between components. A mathematical model was therefore constructed in which the relations between system components are expressed quantitatively by a set of equations (Boer *et al.*, 2010). (c) The model allows quantitative simulation and prediction of the behavior of the system. Simulations shown here are only those for gonadotropin releasing hormone (gnrh), follicle stimulating hormone (fsh), luteinizing hormone (lh), progesterone (p4) and estradiol (e2; from Boer *et al.*, 2010).

during the estrous cycle, reduced expression of estrous behavior and lower conception rates (Wiltbank *et al.*, 2006).

In this research area, we combine functional genomics data and systems biology approaches in order to learn more about the regulation of estrous behavior in dairy cattle. Currently, we apply various analyses, bioinformatics and modeling approaches to identify associations between whole genome gene expression profiles in different areas of the brain and the cow's heat score (Kommadath *et al.*, 2010).

In the same research area, we use a bottom-up mathematical modeling of the estrous cycle of cows. The regulation of estrous is controlled by the interplay of various organs and hormones. As illustrated in Figure 4, even with a limited number of components, the system may become too complex to predict its behavior from the known qualitative relations between components. We have now developed a mathematic model of the bovine estrous (Boer *et al.*, 2010), containing 12 ordinary differential equations and 54 parameters. With the current parameterization, the model generates estrous cycles of 21 days with three peaks of follicle stimulating hormone and three corresponding waves of follicle growth per cycle. The output of the model is surprisingly well in line with empirical data. In future study, we want to use this model to determine the level of control exerted by various system components on the functioning of the system. Examples of such model applications are to explore the mechanisms that influence the pattern of follicular waves or to study hormone patterns associated with subfertility. The model can serve as a basis for more elaborate models and simulations, with the ability to study effects of external manipulations and genetic differences. Possible extensions of the model could be in the field of energy metabolism, stress, disease and factors affecting the expression of estrous behavior.

Conclusions

Systems biology is an emerging interdisciplinary research field combining biology, mathematics and computational science, which aims at building models for dynamic interactions of system components. Such models enable to predict the outcome of a biological system in response to a given external factor. Animal sciences have arrived at the threshold of a genomics data explosion. It is now in a position to make most effective use of the improved knowledge on the

structure, variation and expression of animal genomes. The application of systems biology approaches using this genomic information will provide better insight into the biology of animal traits. Consequently, it will provide opportunities to monitor, modulate, and improve animal traits. Systems biology approaches require a close collaboration between many different disciplinary scientific communities that share resources, technologies and knowledge, and that are willing to integrate their data sets. With the development of systems biology approaches, we are entering the era of a predictive theoretical biology for farm animal traits.

Acknowledgment

This study was financially supported by the Ministry of Agriculture, Nature and Food Quality (project code KB-04-004-049). The paper represents the authors' views and does not necessarily represent a position of the Ministry.

References

Aksenov SV, Church B, Dhiman A, Georgieva A, Sarangapani R, Helmlinger G and Khalil IG 2005. An integrated approach for inference and mechanistic modeling for advancing drug development. FEBS Letters 579, 1878–1883.

Amit I, Garber M, Chevrier N, Leite AP, Donner Y, Eisenhaure T, Guttman M, Grenier JK, Li W, Zuk O, Schubert LA, Birditt B, Shay T, Goren A, Zhang X, Smith Z, Deering R, McDonald RC, Cabili M, Bernstein BE, Rinn JL, Meissner A, Root DE, Hacohen N and Regev A 2009. Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. Science 326, 257–263.

Arakelyan L, Vainstein V and Agur Z 2002. A computer algorithm describing the process of vessel formation and maturation, and its use for predicting the effects of anti-angiogenic and anti-maturation therapy on vascular tumor growth. Angiogenesis 5, 203–214.

Asslaber M and Zatloukal K 2007. Biobanks: transnational, European and global networks. Briefings in Functional Genomics & Proteomics 6, 193–201.

Bannink A, Kogut J, Dijkstra J, France J, Kebreab E, Van Vuuren AM and Tamminga S 2006. Estimation of the stoichiometry of volatile fatty acid production in the rumen of lactating cows. Journal of Theoretical Biology 238, 36–51.

Bionaz M and Loor JJ 2008. Gene networks driving bovine milk fat synthesis during the lactation cycle. BMC Genomics 9, 366.

Boer HMT, Stötzel C, Röblitz S, Deuflhard P, Veerkamp RF and Woelders H 2010. Mathematical model of the bovine oestrous cycle. In Food, feed, energy and fibre from land, a vision for 2020. Annual Conference of the British Society of Animal Science, Agricultural Research Forum and the World Poultry Science Association, Belfast, UK, p. 163.

Bruggeman FJ and Westerhoff HV 2007. The nature of systems biology. Trends in Microbiology 15, 45–50.

Calus MP, Meuwissen TH, de Roos AP and Veerkamp RF 2008. Accuracy of genomic selection using different methods to define haplotypes. Genetics 178, 553–561.

Feist AM, Zielinski DC, Orth JD, Schellenberger J, Herrgard MJ and Palsson BØ 2010. Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. Metabolic Engineering 12, 173–186.

Fell DA and Small JR 1986. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. Biochemical Journal 238, 781–786.

Finney A, Hucka M, Bornstein BJ, Keating SM, Shapiro BE, Matthews J, Kovitz BL, Schilstra MJ, Funahashi A, Doyle JC and Kitano H 2006. Software infrastructure for effective communication and reuse of computational models. In System modeling in cellular biology: from concepts to nuts and bolts (ed. Z Szallasi, J Stelling and P Periwal), pp. 355–378. MIT Press, Cambridge, MA, USA.

Fox S, Filichkin S and Mockler TC 2009. Applications of ultra-high-throughput sequencing. Methods in Molecular Biology 553, 79–108.

Gibson JP and Bishop SC 2005. Use of molecular markers to enhance resistance of livestock to disease: a global approach. Revue scientifique et technique (International Office of Epizootics) 24, 343–353.

Haanstra JR, van Tuijl A, Kessler P, Reijnders W, Michels PA, Westerhoff HV, Parsons M and Bakker BM 2008. Compartmentation prevents a lethal turbo-explosion of glycolysis in trypanosomes. Proceedings of the National Academy of Sciences of the United States of America 105, 17718–17723.

Hunter P and Nielsen P 2005. A strategy for integrative computational physiology. Physiology (Bethesda) 20, 316–325.

Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R and Hood L 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292, 929–934.

Joyce AR and Palsson BO 2006. The model organism as a system: integrating 'omics' data sets. Nature Reviews Molecular Cell Biology 7, 198–210.

Khalil IG and Hill C 2005. Systems biology for cancer. Current Opinion in Oncology 17, 44–48.

Kholodenko BN 2006. Cell-signalling dynamics in time and space. Nature Reviews Molecular Cell Biology 7, 165–176.

Kitano H 2000. Perspectives on systems biology. New Generation Computing 18, 199–216.

Kittler R, Pelletier L and Buchholz F 2008. Systems biology of mammalian cell division. Cell Cycle 7, 2123–2128.

Klauschen F, Angermann BR and Meier-Schellersheim M 2007. Understanding diseases by mouse click: the promise and potential of computational approaches in Systems Biology. Clinical and Experimental Immunology 149, 424–429.

Kommadath A, Mulder HA, de Wit AAC, Woelders H, Smits MA, Beerda B, Veerkamp RF, Frijters ACJ and te Pas MFW 2010. Gene expression patterns in anterior pituitary associated with quantitative measure of oestrous behaviour in dairy cows. Animal 4, 1297–1307.

Lehner B 2007. Modelling genotype-phenotype relationships and human disease with genetic interaction networks. Journal of Experimental Biology 210, 1559–1566.

Lippolis JD and Reinhardt TA 2008. Centennial paper: proteomics in animal science. Journal of Animal Science 86, 2430–2441.

Megason SG and Fraser SE 2007. Imaging in systems biology. Cell 130, 784–795.

Meuwissen TH, Hayes BJ and Goddard ME 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–1829.

Noble D 2008. Computational models of the heart and their use in assessing the actions of drugs. Journal of Pharmacological Sciences 107, 107–117.

Pallen MJ and Wren BW 2007. Bacterial pathogenomics. Nature 449, 835–842.

Palsson B 2002. In silico biology through "omics". Nature Biotechnology 20, 649–650.

Pryce JE, Royal MD, Garnsworthy PC and Mao IL 2004. Fertility in the high-producing dairy cow. Livestock Production Science 86, 125–135.

Quackenbush J 2007. Extracting biology from high-dimensional biological data. Journal of Experimental Biology 210, 1507–1517.

Raghunathan A, Reed J, Shin S, Palsson B and Daefler S 2009. Constraint-based analysis of metabolic capacity of Salmonella typhimurium during host-pathogen interaction. BMC Systems Biology 3, 38.

Reinecke I and Deuflhard P 2007. A complex mathematical model of the human menstrual cycle. Journal of Theoretical Biology 247, 303–330.

Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, Sieberts SK, Monks S, Reitman M, Zhang C, Lum PY, Leonardson A, Thieringer R, Metzger JM, Yang L, Castle J, Zhu H, Kash SF, Drake TA, Sachs A and Lusis AJ 2005. An integrative genomics approach to infer causal associations between gene expression and disease. Nature Genetics 37, 710–717.

Schokker D, Smits MA, Hoekman AJ, Parmentier HK and Rebel JM 2010. Effects of Salmonella on spatial-temporal processes of jejunal development in chickens. Developmental & Comparative Immunology 34, 1090–1100.

Schuetz R, Kuepfer L and Sauer U 2007. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Molecular Systems Biology 3, 119.

Shav-Tal Y, Singer RH and Darzacq X 2004. Imaging gene expression in single living cells. Nature Reviews Molecular Cell Biology 5, 855–861.

Shorten PR, Pleasants TB and Upreti GC 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.

Simao E, Remy E, Thieffry D and Chaouiya C 2005. Qualitative modelling of regulated metabolic pathways: application to the tryptophan biosynthesis in *E. coli*. Bioinformatics 21 (Suppl. 2), ii190–ii196.

Snoep JL 2005. The silicon cell initiative: working towards a detailed kinetic description at the cellular level. Current Opinion in Biotechnology 16, 336–343.

Sordillo LM, Contreras GA and Aitken SL 2009. Metabolic factors affecting the inflammatory response of periparturient dairy cows. Animal Health Research Reviews/Conference of Research Workers in Animal Diseases 10, 53–63.

Te Pas MFW, van Hemert S, Hulsegge B, Hoekman AJW, Pool MH, Rebel JMJ and Smits MA 2008. A pathway analysis tool for analyzing microarray data of species with low physiological information. Advances in Bioinformatics 2008 Article ID 719468, 1–7.

van Eunen K, Bouwman J, Daran-Lapujade P, Postmus J, Canelas AB, Mensonides FI, Orij R, Tuzun I, van den Brink J, Smits GJ, van Gulik WM, Brul S, Heijnen JJ, de Winde JH, Teixeira de Mattos MJ, Kettner C, Nielsen J, Westerhoff HV and Bakker BM 2010. Measuring enzyme activities under standardized in vivo-like conditions for systems biology. The FEBS Journal 277, 749–760.

van Hemert S, Hoekman AJ, Smits MA and Rebel JM 2006. Gene expression responses to a Salmonella infection in the chicken intestine differ between lines. Veterinary Immunology and Immunopathology 114, 247–258.

Veerkamp RF, Beerda B and van der Lende T 2003. Effects of genetic selection for milk yield on energy balance, levels of hormones, and metabolites in lactating cattle, and possible links to reduced fertility. Livestock Production Science 83, 257–275.

Wiltbank M, Lopez H, Sartori R, Sangsritavong S and Gumen A 2006. Changes in reproductive physiology of lactating dairy cows due to elevated steroid metabolism. Theriogenology 65, 17–29.

Wunder F, Kalthof B, Muller T and Huser J 2008. Functional cell-based assays in microliter volumes for ultra-high throughput screening. Combinatorial Chemistry & High Throughput Screening 11, 495–504.

Yashiro Y, Bannai H, Minowa T, Yabiku T, Miyano S, Osawa M, Iwama A and Nakauchi H 2009. Transcriptional profiling of hematopoietic stem cells by high-throughput sequencing. International Journal of Hematology 89, 24–33.

Young D, Stark J and Kirschner D 2008. Systems biology of persistent infection: tuberculosis as a case study. Nature reviews. Microbiology 6, 520–528.