
Systems Biology in Animal Production and Health, Vol. 2

Haja N. Kadarmideen
Editor

Systems Biology in Animal Production and Health, Vol. 2

 Springer

Editor

Haja N. Kadarmideen
Faculty of Health and Medical Sciences
University of Copenhagen
Frederiksberg C, Denmark

ISBN 978-3-319-43330-1 ISBN 978-3-319-43332-5 (eBook)
DOI 10.1007/978-3-319-43332-5

Library of Congress Control Number: 2016956674

© Springer International Publishing Switzerland 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland
The registered company address is Gewerbstrasse 11, 6330 Cham, Switzerland

Foreword

The increased prominence of “systems biology” in biological research over the past two decades is arguably a reaction to the reductionist approach exemplified by the genome sequencing phase of the Human Genome Project. A simplistic view of the genome projects was that the genome sequence of a species, whether humans, model organisms, plants or farmed animals, represents a blueprint for the organism of interest, and thus characterising the sequence would reveal the relevant instructions. Subsequent targets for the reductionist or cataloguing approach were complete lists of transcripts (transcriptomes) and proteins (proteomes) for the organism of interest. The ‘omics approach to the comprehensive characterisation of an organism, tissue or cell has also been extended to metabolites and hence metabolomes. A catalogue of parts, however, is insufficient to understand how an organism functions. Thus, a holistic approach that recognises the interactions between components of the system was required. Given the size and complexity of the data and the possible interactions, it was necessary to use advanced mathematical and computational methods to attempt to make sense of the data. Thus, “systems biology” in the ‘omics era is widely considered to concern the use of mathematical modelling and analysis together with ‘omics data (genome sequence, transcriptomes, proteomes, metabolomes) to understand complex biological systems. The predictive aspect of these models is viewed as particularly important. Moreover, it is desirable that the models’ predictions can be tested experimentally. Systems biology, therefore, contributes in part to converting large ‘omics data sets from data-driven biology experiments into testable hypotheses.

Systems approaches and the use of predictive mathematical models in biological systems long pre-date the post genome project (re-)emergence of systems biology. Population biologists/geneticists, epidemiologists, agricultural scientists, quantitative geneticists and plant and animal breeders have been developing and successfully exploiting predictive mathematical models and systems approaches for decades.

Quantitative geneticists and animal breeders, for example, have been remarkably successful at developing statistical animal models that are effective predictors of future performance. For decades, these successes were achieved without any knowledge of the underlying molecular components. The accuracy of these models has been increased by using high-density molecular (single nucleotide polymorphism, SNP)

genotypes in so-called genomic selection. However, whilst the sequences and genome locations of the SNP markers are known little is known about the functional impact or relevance of the individual SNP loci. Further improvements could be achieved through the use of genome sequence data and by adding knowledge of the likely effects of the sequence variants whether coding or regulatory. Thus, there is a growing commonality between the systems approaches of quantitative geneticists and animal breeders and the ‘omics version of systems biology.

Animals are not only complex biological systems but also function within wider complex systems. The recognition that an animal’s phenotype is determined by a combination of its genotype and environmental factors simply restates the latter. The environmental factors include, amongst others, feed, pathogens and the microbiomes present in the gastrointestinal tract and other locations. The ‘omics technologies allow not only the characterisation of the components of the animal of interest, but also those of its commensal microbes and the microbes, including pathogens present in its environment.

As noted earlier, it is desirable that the mathematical models developed in systems biology are predictive and that the associated hypotheses are testable. Genome editing technologies which have been demonstrated in farmed animal species facilitate hypothesis testing at the level of modifying the genome sequence that determines components of the system of interest.

This volume of *Systems Biology in Animal Production and Health*, edited by professor Haja Kadarmideen, explores some aspects of both quantitative genetics and ‘omics led approaches to applying systems approaches to tackling the challenges of improving animal productivity and reducing the burden of disease. The book contains some chapters with R codes and other computer programs, workflow/pipeline for processing and analysing multi-omic datasets from lab all the way to interpretation of results. Hence, this book would be useful particularly for students, teachers and practitioners of integrative genomics, bioinformatics and systems biology in animal and veterinary sciences.

Villa-Vialaneix et al. (chapter “[Depicting Gene Co-expression Networks Underlying eQTL](#)”) address the challenge of identifying the gene networks that capture the interaction between genes from eQTL data. The application of systems approaches to specific traits of interest in agriculture and biology are reviewed by *Schroyen et al.* (chapter “[Applications of Systems Biology to Improve Pig Health](#)”), *Fukumasu et al.* (chapter “[Systems Biology Application in Feed Efficiency in Beef Cattle](#)”), and *Vailati-Riboni et al.* (chapter “[Nutritional Systems Biology to Elucidate Adaptations in Lactation Physiology of Dairy Cows](#)”). The analysis of transcriptomic data and specifically RNA-Seq data are described in greater detail by *Mazzoni and Kadarmideen* (chapter “[Computational Methods for Quality Check, Preprocessing and Normalization of RNA-Seq Data for Systems Biology and Analysis](#)”).

Finally, farmed animal species are not only important for agriculture but are also used for basic biological research and as models in biomedical research. *Mashayekhi et al.* (chapter “[Systems Biology and Stem Cell Pluripotency: Revisiting the Discovery of Pluripotent Stem Cell](#)”) describe a systems perspective on pluripotency.

Professor Alan L. Archibald FRSE
Deputy Director, Head of Genetics and Genomics
The Roslin Institute and Royal (Dick) School of Veterinary Studies
University of Edinburgh
Easter Bush, Midlothian EH25 9RG, UK

Preface

Systems biology is a research discipline at the crossroad of statistical, computational, quantitative and molecular biology methods. It involves joint modeling, combined analysis and interpretation of high-throughput omics (HTO) data collected at many “levels or layers” of the biological systems within and across individuals in the population. The systems biology approach is often aimed at studying associations and interactions between different “layers or levels”, but not necessarily one layer or level in isolation. For instance, it involves study of multidimensional associations or interaction among DNA polymorphisms, gene expression levels, proteins or metabolite abundances. With modern HTO biotechnologies and their decreasing costs, hugely comprehensive multi-omic data at all “levels or layers” of the biological system are now available. This “big data” at lower costs, along with development of genome scale models, network approaches and computational power, have spearheaded the progress of the systems biology era, including applications in human biology and medicine. Systems biology is an established independent discipline in humans and increasingly so in animals, plants and microbial research. However, joint modeling and analyses of multilayer HTO data, in large volumes on a scale that has never been seen before, has enormous challenges from both computational and statistical points of view. Systems biology tackles such joint modeling and analyses of multiple HTO datasets using a combination of statistical, computational, quantitative and molecular biology methods and bioinformatics tools. As I wrote in my review article (*Livestock Science* 2014, 166:232–248), systems biology is not only about multilayer HTO data collection from populations of individuals and subsequent analyses and interpretations; it is also about a philosophy and a hypothesis-driven predictive modeling approach that feeds into new experimental designs, analyses and interpretations. In fact, systems biology revolves and iterates between these “wet” and “dry” approaches to converge on coherent understanding of the whole biological system behind a disease or phenotype and provide a complete blueprint of functions that leads to a phenotype or a complex disease.

It is equally important to introduce, alongside systems biology, the sub-discipline of *systems genetics* as a branch of systems biology. It is akin to considering “genetics” as a sub-discipline of “biology”. It is well known that quantitative genetics/genomics links genome-wide genetic variation with variation in disease risks or a performance (phenotype or trait) that we can easily measure or observe in a

population of individuals. However, systems genetics or systems genomics not only performs such genome-wide association studies (GWAS), but also performs linking genetic variations (e.g. SNPs, CNVs, QTLs etc.) at the DNA sequence level with variation in molecular profiles or traits (e.g. gene expression or metabolomic or proteomic levels etc. in tissues and biological fluids) that we can measure using high-throughput next- and third-generation biotechnologies. The systems genetics approach is still “genetics”, because we are looking at those genetic variants that exert their effects from DNA to phenotypic expression or disease manifestations through a number of intermediate molecular profiles. Hence, systems genetics derives its name, as originally proposed in my earlier article (*Mammalian Genome*, 2006, 17:548–564), by being able to integrate analyses of all underlying genetic factors acting at different biological levels, namely, QTL, eQTL, mQTL, pQTL and so on. I have provided a complete up-to-date review and illustration of systems genetics or systems genomics and multi-omic data integration and analyses in our review paper published in *Genetics Selection Evolution* (2016), 48:38. Overall, systems genetics/genomics leads us to provide a holistic view on complex trait heredity at different biological layers or levels.

Whether it is systems biology or systems genetics, the gene ontology annotation is one of the most important and valuable means of assigning functional information using standardized vocabulary. This would include annotation of genetic variants falling into functional groups such as trait QTL, eQTL, mQTL, pQTL. Molecular pathway profiling, signal transduction and gene set enrichment analyses along with various types of annotations form the “icing on cake”. For this purpose, several bioinformatics tools are frequently used. Most chapters in this book and its associated volume cover these aspects.

I would like to point out that systems biology approaches have been proven to be very powerful and shown to produce accurate and replicable discoveries of genes, proteins and metabolites and their networks that are involved in complex diseases or traits. In very practical terms, it delivers biomarkers, drug targets, vaccine targets, target transcripts or metabolites, genetic markers, pathway targets etc. to diagnose and treat diseases better or improve traits or characteristics in animals, plants and humans. In the world of genomic prediction and genomic selection, there have been an increasing number of studies that have shown high accuracy and predictive power when models include functional QTLs such as eQTL, mQTL, pQTL which, in fact, are results from systems genetics methods.

This book and its associated volume cover the above-mentioned principles, theory and application of systems biology and systems genetics in livestock and animal models and provides a comprehensive overview of open source and commercially available software tools, computer programming codes and other reading materials to learn, use and successfully apply systems biology and systems genetics in animals.

Overall, I believe this book is an extremely valuable source for students interested in learning the basics and could form as a textbook in higher educational institutes and universities around the world. Equally, the book chapters are very relevant and useful for scientists interested in learning and applying advanced HTO studies, integrative HTO data analyses (e.g. eQTLs and mQTLs) and computational

systems biology techniques to animal production, health and welfare. One of the chapters focuses on stem cell research in animal models elucidating systems biology of pluripotency with translational applications for human neurological and brain diseases. The two volumes of this book is a result of contributions from highly reputed scientists and practitioners who originate from renowned universities and multinational companies in the UK, Denmark, France, Italy, Australia, USA, Brazil and India. I would like to thank the publisher Springer for inviting me to edit two volumes on this subject, publishing in an excellent form and promoting the book across the globe. I am grateful to all contributing authors and co-authors of this book. I also wish to thank Ms. Gilda Kischinovsky from my research group for proofreading and the staff at Springer involved in production of this book. Last but not least, I wish to thank my wife and children who have given me moral support and strength while I reviewed and edited this book.

Copenhagen, Denmark
September 2016

Haja N. Kadarmideen

Contents

Depicting Gene Co-expression Networks Underlying eQTLs.	1
Nathalie Villa-Vialaneix, Laurence Liaubet, and Magali SanCristobal	
Applications of Systems Biology to Improve Pig Health	33
Martine Schroyen, Haibo Liu, and Christopher K. Tuggle	
Computational Methods for Quality Check, Preprocessing and Normalization of RNA-Seq Data for Systems Biology and Analysis	61
Gianluca Mazzoni and Haja N. Kadarmideen	
Systems Biology Application in Feed Efficiency in Beef Cattle	79
Heidge Fukumasu, Miguel Henrique Santana, Pamela Almeida Alexandre, and José Bento Sterman Ferraz	
Nutritional Systems Biology to Elucidate Adaptations in Lactation Physiology of Dairy Cows	97
Mario Vailati-Riboni, Ahmed Elolimy, and Juan J. Loor	
Systems Biology and Stem Cell Pluripotency: Revisiting the Discovery of Induced Pluripotent Stem Cell	127
Kaveh Mashayekhi, Vanessa Hall, Kristine Freude, Miya K Hoeffding, Luminita Labusca, and Poul Hyttel	