Bioinformatics for Immunomics

Immunomics Reviews
An Official Publication of the International Immunomics Society

---

This peer-reviewed book series offers insight on immunology for 21st century. The technological revolution has borne advances in high-throughput instrumentation and information technology, initiating a renaissance for biomathematics, and biostatistics. Cross-fertilization between genomics and immunology has led to a new field called immunomics, transforming the way in which theoretical, clinical and applied immunology are practiced. Immunomics Reviews will cover integrative approaches and applications to the theory and practice of immunology and explore synergistic effects resulting from a combination of technological advances and the latest analytical tools with the traditional fields of basic and clinical immunology.

For other titles published in this series, go to
www.springer.com/series/7528

Darren R. Flower · Matthew N. Davies
Shoba Ranganathan
Editors

Bioinformatics for Immunomics

Springer

Editors
Darren R. Flower
University of Oxford
RG20 7NN
United Kingdom
darren.flower@jenner.ac.uk

Matthew N. Davies
University of London
London WC1E 7HX
United Kingdom
m.davies@mail.cryst.bbk.ac.uk

Shoba Ranganathan
Macquarie University
Sydney
Australia
shoba.ranganathan@mq.edu.au

ISBN 978-1-4419-0539-0 e-ISBN 978-1-4419-0540-6
DOI 10.1007/978-1-4419-0540-6
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2009927086

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Contributors

P. Antal
Department of Measurement and Information Systems, Budapest University of Technology and Economics, Magyar tudosok korutja 2., 1117, Rm IE 423, Budapest, Hungary

M.N. Davies
The Jenner Institute, University of Oxford, High Street, Compton, Berkshire RG20 7NN, UK
m.davies@mail.cryst.bbk.ac.uk

Carmen M. Díez-Rivero
Facultad de Medicina, Departamento de Immunología (Microbiología I), Universidad Complutense de Madrid, Pabellón 5º, planta 4ª, 28040 Madrid, Spain
cmdiezri@med.ucm.es

A. Falus
Department of Genetics, Cell- and Immunobiology, Semmelweis University, Nagyvárad tér 4, Budapest 1089, Hungary
Faland@dgci.sote.hu

D.R. Flower
The Jenner Institute, University of Oxford, High Street, Compton, Berkshire RG20 7NN, UK
darren.flower@jenner.ac.uk

Jason A. Greenbaum, Ph.D.
La Jolla Institute for Allergy & Immunology, 9420 Athena Circle, La Jolla, CA 92037, USA
jgbaum@liai.org

G. Hajós
Department of Measurement and Information Systems, Budapest University of Technology and Economics, Magyar tudosok korutja 2., 1117 Budapest, Hungary

G. Hullám
Department of Measurement and Information Systems, Budapest University of Technology and Economics, Magyar tudosok korutja 2., 1117 Budapest, Hungary
hullam.gabor@mit.bme.hu

Javed Mohammed Khan
Department of Chemistry and Biomolecular Sciences & ARC Centre of Excellence in Bioinformatics, Macquarie University, Sydney, NSW 2109, Australia

Sneh Lata
Institute of Microbial Technology, Sector39A, Chandigarh, India
sneh@imtech.res.in

Steven G.E. Marsh
Department of Haematology, Royal Free Hospital, Pond Street, Hampstead, London NW3 2QG, UK
marsh@ebi.ac.uk

A. Millinghoffe
Department of Measurement and Information Systems, Budapest University of Technology and Economics, Magyar tudosok korutja 2., 1117 Budapest, Hungary
milli@mit.bme.hu

Bjoern Peters, Ph.D.
La Jolla Institute for Allergy & Immunology, 9420 Athena Circle, La Jolla, CA 92037, USA
bpeters@liai.org

G.P.S. Raghava
Institute of Microbial Technology, Sector39A, Chandigarh, India
raghava@imtech.res.in

Shoba Ranganathan
Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore 117597

Pedro Reche
Facultad de Medicina, Departamento de Immunología (Microbiología I), Universidad Complutense de Madrid, Pabellón 5º, planta 4ª, 28040 Madrid, Spain,
pareche@med.ucm.es

James Robinson
Anthony Nolan Research Institute, Royal Free Hospital, Pond Street, Hampstead, London NW3 2QG, UK
jrobinso@ebi.ac.uk

Alessandro Sette, Ph.D.
La Jolla Institute for Allergy & Immunology, 9420 Athena Circle, La Jolla, CA 92037, USA
alex@liai.org

Cs. Szalai
Inflammation Biology and Immunogenomics Research Group, Hungarian Academy of Sciences, Nagyvárad tér 4, Budapest 1089, Hungary

Jon Timmis, B.Sc. (Wales), Ph.D. (Wales)
Department of Computer Science, University of York, York, UK

Joo Chuan Tong
Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613

Randi Vita, M.D.
La Jolla Institute for Allergy & Immunology, 9420 Athena Circle, La Jolla CA 92037, USA
rvita@liai.org

Laura M. Zarebski, Ph.D.
La Jolla Institute for Allergy & Immunology, 9420 Athena Circle, La Jolla, CA 92037, USA
laura@liai.org

Introduction

Like many words, the term "immunomics" equates to different ideas contingent on context. For a brief span, immunomics meant the study of the Immunome, of which there were, in turn, several different definitions. A now largely defunct meaning rendered the Immunome as the set of antigenic peptides or immunogenic proteins within a single microorganism – be that virus, bacteria, fungus, or parasite – or microbial population, or antigenic or allergenic proteins and peptides derived from the environment as a whole, containing also proteins from eukaryotic sources. However, times have changed and the meaning of immunomics has also changed. Other newer definitions of the Immunome have come to focus on the plethora of immunological receptors and accessory molecules that comprise the host immune arsenal. Today, Immunomics or immunogenomics is now most often used as a synonym for high-throughput genome-based immunology. This is the study of aspects of the immune system using high-throughput techniques within a conceptual landscape borne of both clinical and biophysical thinking.

Within an immunogenomic or immunomic framework, chapter "A Bioinformatic Platform for a Bayesian, Multiphased, Multilevel Analysis in Immunogenomics" describes a bioinformatic platform for undertaking Bayesian analysis at multilevels. How the phenotypic behaviour of the immune system emerges from the interaction of its genome-encoded components should be of paramount interest to all involved in its investigation. Saying this is one thing; but actually achieving it is quite another. Bayesian statistics can provide an insightful route to manifesting data which is both rigorous and of true utility.

Clearly, the genome, the epinome, the proteome, the glycome, the metabolome, and all the rest of the – omes and – omics that have come to dominate our current perceptions are of direct relevance to burgeoning understanding of immunology and immunological processes. Genes, proteins, carbohydrates, and lipids, as well glycoproteins and lipoproteins, together with the peptides and small molecules too, all take part in an extraordinary range of interactions that manifest themselves as our immune response to pathogen challenge. It is clear, if still largely unacknowledged, that a pivotal turning point has been reached; several key technologies have achieved long-awaited maturity, most notably predictive immunoinformatic methods and post-genomic strategies.

In another popular definition, immunomics can stand as a synonym for system biology techniques applied to the study of Immunology. For many scientists, immunology is the pre-eminent example of systems behaviour in biology. Of course, the whole of biology – indeed the whole of the physical universe – behaves as a system, and exhibits characteristic systems behaviour. Since the immune system is innately hierarchical and exhibits confounding complexity at each tier of this cascading or branching hierarchy, as a system it can be said to exhibit emergent behaviour at all levels. Yet at the heart of the immune system are arrays of straightforward if not uncomplicated molecular recognition events, each of which is essentially indistinguishable from other biomolecular interactions. It is only when our frail and constrained mortal minds study something do we see the components in isolation. This is the glory – and the limitation – of reductionist approaches to scientific insight and discovery.

Since the discovery of antibodies and MHC restriction, humoural and cellular immunologists have sought to understand the nature of these biomacromolecular interactions, seeking to analyse them in the most fundamental way. Systems biology seeks to analyse higher levels of the immune system with the same degree of rigour, by both analysing the system as it exhibits itself at these individual levels and by integrating detailed, low-level, small-scale molecular or mesoscopic information and more overtly macroscopic measurements with more intrinsically qualitative anatomical, functional, and phenotypic data. Thus, Systems Biology – or, in this context, Systems Immunomics – can be said to function at various length scales from the atomic to the macroscopic.

Biological systems, of which immunological systems are an example, are seldom binary entities on the whole organism scale, any more than their cascading sub-systems – be they organ, tissue, or cellular – are binary entities at subsidiary levels. They operate stochastically, subject to random fluctuations and exhibit clear non-linear behaviour. Immunology only truly manifests itself at the level of the whole organism, but at every intermediate level down to that of the molecule, significant and often unexpected emergent behaviour within experimental systems is observed.

Many tools exist within systems biology. Some tools are based on capitalising on the latent power of simulation, be that simulations of abstract theoretical or mathematical models or molecular simulations of precise descriptions of molecular system. Other tools are analytical tools that can be used together to effect the synthesis of competing thesis and antithesis through the integration of measured data.

The simplest types of systems model include network maps, which reticulate pathway components producing complex cellular representations akin to circuit diagrams, and so-called logical models, which describe immunological process in terms of sets of relatively simple rules, framed as Boolean logic, such as if X AND Y but NOT Z then A = 1. There are many other more complex and mathematically demanding models available; these include correlation models and kinetic modelling.

Correlation models are familiar to anyone who has ever tried statistical approaches to the prediction problem. Multiple linear regression or Partial Least Squares or neural networks, or, indeed, any of a hundred other data mining techniques, can be used to identify commonalities of exchange or cooperation within or between the measured

outputs of different signalling or regulatory pathways. Kinetic models, on the other hand, try to picture the spatio-temporal behaviour of each and every individual component within the system. They are the zenith and apotheosis of complexity with the currently available approaches within systems biology.

It is also possible to combine these different kinds of model. One can use a hybrid of a kinetic model and one based on Boolean logic. This is particularly useful when one wishes to fuse data of different granularity. It is possible, for example, to build a detailed kinetic model for part of a pathway and then to fill in the lacuna within the available data by modelling the rest using much simpler Boolean models.

The word "bioinformatics" has formed part of the scientific lingua franca since the early 1990s; yet a simple and straightforward, and comprehensive and inclusive definition remains strangely elusive. A particularly succinct epitome of the discipline is: "Bioinformatics is the application of informatics methods to biological macromolecules."

There are many reasons for this failure of orismology, partly arising because bioinformatics is in constant flux; undergoing relentless change, growth, and differentiation: you cannot easily name something that is never still enough to describe.

Bioinformatics has greatly expanded over the years, allowing for both new sub-disciplines to emerge within it and for bioinformatics to merge with other disciplines producing new and exciting hydrids. Sub-disciplines have tended to focus on areas of applications, such as neuroinformatics, transcriptomics, or proteomics, while hybrids have included text mining or statistical genetics. Immunoinformatics is another important sub-discipline. It deals specifically with the unique problems of the immune system. Like Bioinformatics, immunoinformatics complements, but never replaces, practical experimentation. It helps, and in a systematic way, researchers to answer the key questions in the still highly empirical world of immunology.

The scope and focus of bioinformatics is constantly developing and expanding to encompass more and more new areas of application. However, it is clear that Bioinformatics concerns itself with medical, genomic, and biological information and supports both basic and clinical research. Bioinformatics develops computer databases and algorithms for accelerating, simplifying, and thus enhancing, research in bioscience. Within this, however, the nature and variety of different bioinformatic activities are hard to quantify. Bioinformatics is as much a melting pot of interdisciplinary techniques as it is a branch of information science: it operates at the level of protein and nucleic acid sequences, their structures, and their functions, using data from microarray experiments, traditional biochemistry, as well as theoretical biophysics.

Databases are a key component of research in bioinformatics and immunoinformatics. They have been so and will remain vital for the foreseeable future. They are, or should be, as much tools as the algorithms used to search, analyse, and interrogate them. Bioinformatics is largely concerned with data handling, mainly through the annotation of macromolecular sequence and structure databases. A number of chapters in this book describe the application and development of databases within the immunoinformatic domain. Chapters "IPD – The Immuno Polymorphism Database" and "The IMGT/HLA Database", by Professor Marsh and co-workers,

describe two world-leading resources: IPD and IMGT/HLA. Chapter "Ontology Development for the Immune Epitope Database" by Bjorn Peters and colleagues neatly summarises on-going development of the IEDB database. Chapter "Databases and Web-Based Tools for Innate Immunity" extends and completes this strand by describing a variety of databases aimed at the archiving of data relating to the innate immune system.

The growth of Bioinformatics is a clear success story of the informatic applications in bioscience. The services of bioinformaticians remain much in demand by forward-thinking biologists of many kinds. As new genomes are, for example, sequenced, biologists and immunologists wish to know many things: where and what post-translational modification there are; the location of protein within a cell; which proteins will substrates for proteases or kinases or other enzymes; even down to the pKa of a particular residue within a certain enzyme active site residue. The list seems, and is, endless, or nearly so. Attempting to address all of these possibilities in a systematic and effective manner using experiment only would be prohibitive to the point of intractability, in terms of time, resource, and that most precious quantity of all: human labour. The only practical and practicable solution is the deployment of bioinformatics.

Bioinformatics focuses on analysing molecular sequence and structure data, molecular phylogenies, and the analysis of post-genomic data generated by genomics, transcriptomics, and proteomics. Bioinformatics seeks to find solutions to two key challenges. First, the prediction of Function from Sequence, which can be performed using global homology searches, motif databases searches, and the formation of multiple sequence alignments. Chapter "Discovery of Conserved Epitopes Through Sequence Variability Analyses" indicates the wisdom of this assertion; it addresses the prediction of conserved epitopes within an immunomics and immunoinformatic context. Chapter "Defining the Elusive Molecular Self" picks up on this with its analysis of the molecular nature of the immune self.

Secondly, the prediction of Structure from Sequence, which may be attempted using secondary structure prediction, threading, and comparative, or so-called homology, modelling. Chapter "Structural Immunoinformatics: Understanding MHC-Peptide-TR Binding" provides a lucent and definitive description of the use of 3-dimensional structural data, as derived from experiment and computation, within the province of immunoinformatic investigation. As yet, the full power of 3-dimensional data has not been realised. Structure-based computation based on dynamic simulation and hypothesis-guided modelling has so much to reveal, but as yet the potential is not matched by available computing resources. The next few years should see this approach beginning to bear fruit as more and more studies are undertaken.

It is also an implicit assumption that knowledge of a structure facilitates prediction of function. In reality, all predictions of function rely on identifying similarity between sequences or between structures. When this similarity is very high, and thus is intrinsically reliable, then useful inferences may be drawn, but as similarity falls away any conclusions that are inferred become increasingly uncertain and potentially misleading. Thus, provenance is everything; and provenance and

annotation. Bioinformatics still concerns handling and analysing data, often basing the classification of sequences or structures into coherent groups on the rigorous annotation of macromolecular sequence and structure databases. The Tepidas system described in chapter "TEPIDAS: A DAS Server for Integrating T-Cell Epitope Annotations" addresses the integration of data sources for the rigorous and reliable annotation of T cell epitopes.

Vaccines were for so long a moribund market, yet they have recently re-emerged as the most hopeful growth area for the Pharmaceutical Industry. Public health requirements safeguard vaccine supply of vaccines and in the absence of competition – Influenza apart, only two to three manufacturers target each vaccine-preventable disease – this has led to a recent increase in unit price for specialty vaccines. The launch of pioneering products – Wyeths Prevnar or Merck's Gardasil, for example – together with a much more favourable regulatory framework has made vaccines a key focus of the biotechnology and pharmaceutical industries. Paediatric vaccines currently hold sway over the global market for vaccines, yet adult vaccines will help drive future growth. The cancer vaccine market, led by vaccines targeting cervical cancer, is the most lucrative area of vaccine development: by 2012, cancer vaccines will account for around 30% of all vaccine revenues. As discussed in chapter "Computational Vaccinology", Immunomics and Systems Immunomics, at least in their informatic and computational guise, have much to offer vaccine design and discovery and the still emergent science of Vaccinology.

Returning to our first theme, the term vaccinology is said by many to have been coined by Jonas Salk to distinguish the systematic scientific study of vaccines – and thus how to develop and discover them – from the practice of vaccination as a medical art. In recent times, another term, immunovaccinology has been adopted by some to further differentiate the study of vaccine discovery and development based on a sound understanding of immunology, if such a thing exists, from what many might consider the highly empirical, microbiology-based science of vaccinology, as practiced in year gone by. Davies and Flower give a concise examination of how immunoinformatics has and can impact upon the pursuance of a rational yet systematic approach to vaccine discovery.

The next stage in the evolution of Immunomics and Systems Immunomics will come as closer collaborative links are forged between immunoinformaticians and experimentalists searching for new and deeper understanding of immunology, within and between both academic and commercial organisations. Immunomics must be both client and provider acting as a consumer of existing techniques and as an inspiration for other techniques. In this regard, the prime acolyte is that branch of computer science known as artifical immune systems or AIS research. The power and potential of AIS is amply demonstrated by chapter "Tunable Detectors for Artificial Immune Systems: From Model to Algorithm". This looks at how the immune system is able to provide a metaphor in the development of tunable detectors.

Despite the need for more accurate prediction algorithms, able to cover ever more MHC alleles in ever more species, the lack of persuasive evaluations of known methods continues to hamper and stymie uptake of this technology. In order that Immunoinformatic approaches might one day become universally used

by experimental immunologists, methods should be tested over a wide range of alleles, species, and sequence-distinct peptides, with their accuracy reaching a high statistical significance. This will be greatly facilitated by adoption of a cyclically and progressive process of using and refining models and experiments.

The effective implementation of immunoinformatic strategies within Immunomics and Systems Immunomics will deliver an unprecedented dividend of great if unquantifiable magnitude. Methods that accurately predict individual components of the immune response or allow us to model the behaviour of the whole system or part thereof will be the most vital of tools for tomorrow's immunologists and vaccinologist. Immunoinformatic prediction, within the broader system immunomics context, remains a grand scientific problem, being both challenging, and thus exciting, and of true practical value. Moreover, the proper realisation of Systems Immunomics requires not only a deep appreciation of immunological mechanisms but also requires one to integrate many other disciplines, both experimental and theoretical. To enable this requires more than improved methods and software, it necessitates building immunoinformatics into the basic strategy of immunological investigation and it needs the confidence of experimentalists to commit laboratory work on this basis. Within the context of immunomics and systems immunomics, the synergy of experimental and informatics-based disciplines will enhance significantly our ability to understand and manipulate immunology process, leading to the augmented discovery of new laboratory reagents and diagnostics, in addition to new biomarkers and candidate vaccines.

M.N. Davies
S. Ranganathan
D.R. Flower