Review

Cell PRESS

Feature Review

Enabling a systems biology approach to immunology: focus on innate immunity

Jennifer L. Gardy[1,*], David J. Lynn[2,*], Fiona S.L. Brinkman[2] and Robert E.W. Hancock[1]

[1] Centre for Microbial Diseases and Immunity Research, 232 - 2259 Lower Mall, University of British Columbia, Vancouver, British Columbia, Canada, V6T 1Z4
[2] Department of Molecular Biology and Biochemistry, 8888 University Drive, Simon Fraser University, Burnaby, British Columbia, Canada, V5A 1S6

Immunity is not simply the product of a series of discrete linear signalling pathways; rather it is comprised of a complex set of integrated responses arising from a dynamic network of thousands of molecules subject to multiple influences. Its behaviour often cannot be explained or predicted solely by examining its components. Here, we review recently developed resources for the systems-level investigation of immunity. Although innate immunity is emphasized here, its considerable overlap with adaptive immunity makes many of these resources relevant to both arms of the immune response. We discuss recent studies implementing these approaches and illustrate the potential of systems biology to generate novel insights into the complexities of innate immunity.

Understanding innate immunity requires systems-level analysis

Innate immunity: recent developments reveal its complexity

Interest in the innate immune response [1], often described as our first line of defence against invading pathogens, has exploded in recent years, leading to an increased understanding of its importance in protection against and susceptibility to a range of infectious agents, in addition to the discovery of the complex communication between the innate and adaptive immune systems. This heightened focus has enabled not only an increasingly detailed dissection of many of the key signalling pathways involved [2–4] but also the realization that the innate immune response is much more complex than previously imagined [5].

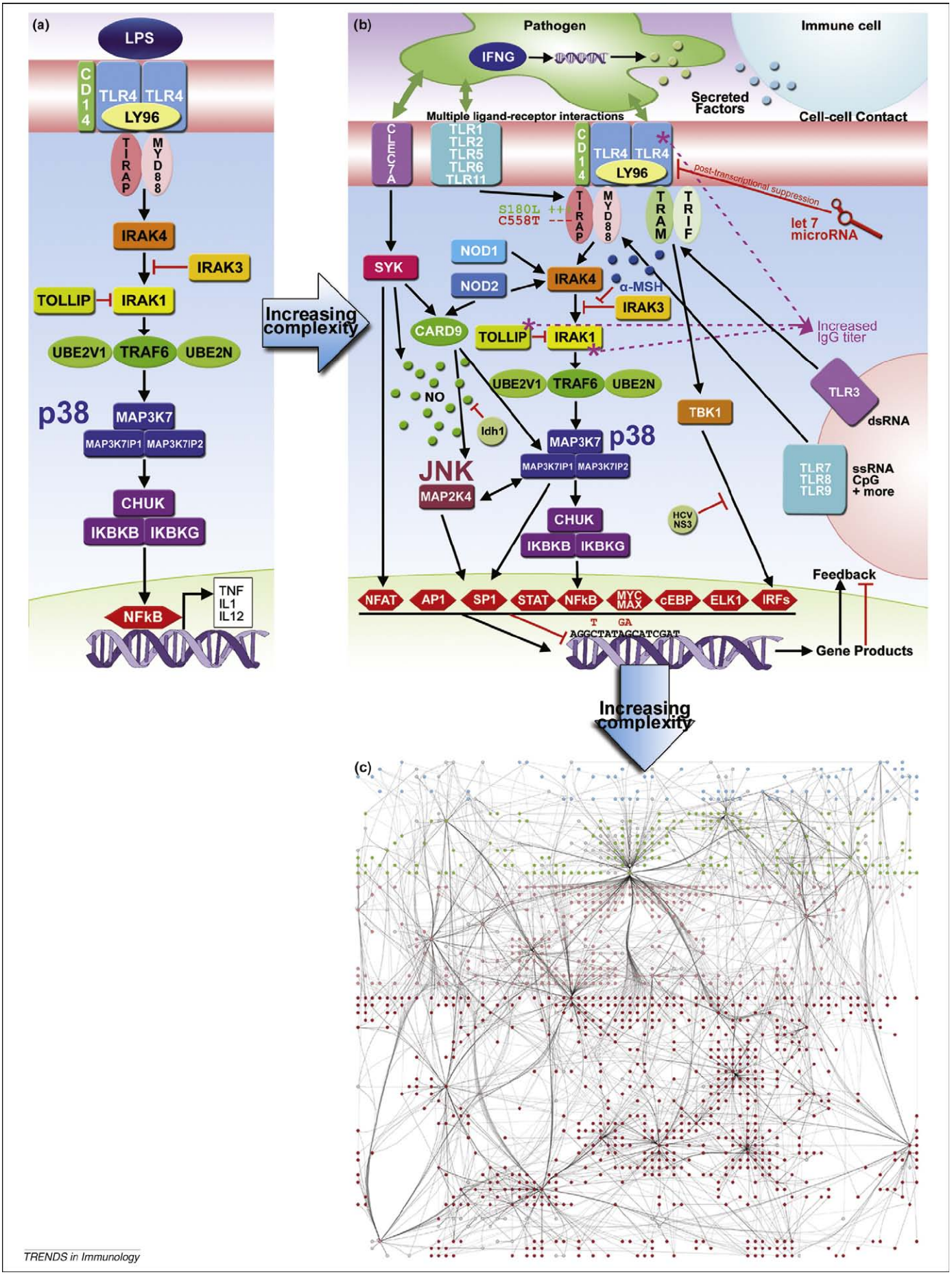
Innate immunity can no longer be simply thought of as a set of discrete signalling pathways activated by a pathogen binding to a receptor (Figure 1a); instead, the innate immune response is a complex network of interconnected pathways with activities dependent on many factors (Figure 1b, 1c). It is now clear that single ligands can trigger multiple signal transduction pathways, although it is rare that a single stimulus would be involved in an infection. More commonly, multiple ligands simultaneously stimulate a range of receptors, which in turn activate several signalling pathways [6]. These pathways often exhibit cross-talk, complex feedback or feed-forward loops [7] and diverse mechanisms of regulation. The importance of post-transcriptional regulation of signalling, including regulation mediated by microRNAs (miRNAs) [8] and post-translational modifications (e.g. proteolysis, ubiquitination, phosphorylation or acetylation [9]) is becoming increasingly apparent. Downstream of the pathways, specific subsets of transcription factors (TFs) are activated to initiate the appropriate gene expression response to a given stimulus [10]. For example, recently, the amplification and attenuation kinetics of a gene-regulatory network involving the TFs CCAAT/enhancer binding protein delta (Cebpd), activating transcription factor 3 (Atf3) and nuclear factor kappa B (NF-κB), was shown to distinguish between transient and persistent signalling triggered through Toll-like receptor 4 (Tlr4) [11]. The environment surrounding the cell can also have a profound effect on its response to a pathogen. Contact with other immune cell-types [12], for example, or exposure to compounds including cytokines, hormones or extrinsic small molecules, can also impact upon the immune response. At the population level, single nucleotide polymorphisms (SNPs) can alter protein–protein interactions or transcriptional regulation, and add another layer of complexity [13].

The immune response is also not simply a function of the host; the pathogen itself and regulated variations in its virulence along with complex interplay with the rest of the microbiome can all lead to variation in the innate immune response. For example, there are pronounced differences in the macrophage responses to viable organisms, such as *Mycobacterium tuberculosis,* compared with the response to inactivated bacteria [14], highlighting the fact that viable pathogens can actively manipulate the host response. Similarly, host factors can influence the pathogen; interferon-γ (IFN-γ), for example, stimulates the transcription of *Pseudomonas aeruginosa* virulence factors [15].

Yet another layer of complexity is introduced when the cell-type and species specificity of the innate immune response are considered. Certain receptors and signalling

Corresponding author: Hancock, R.E.W. (bob@cmdr.ubc.ca)
* Joint first authors. These authors contributed equally to this work..

1471-4906/$ – see front matter doi:10.1016/j.it.2009.03.009 Available online 8 May 2009


(a)
LPS
CD14
TLR4
TLR4
LY96
TIRAP
MYD88
IRAK4
IRAK3
TOLLIP
IRAK1
UBE2V1
TRAF6
UBE2N
p38
MAP3K7
MAP3K7IP1
MAP3K7IP2
CHUK
IKBKB
IKBKG
NFkB
TNF
IL1
IL12
Increasing complexity
(b)
Pathogen
Immune cell
IFNG
Secreted Factors
Multiple ligand-receptor interactions
Cell-cell Contact
CLEC7A
TLR1
TLR2
TLR5
TLR6
TLR11
CD14
TLR4
TLR4
LY96
post-transcriptional suppression
let 7 microRNA
S180L +++
C558T ---
TIRAP
MYD88
TRAM
TRIF
SYK
NOD1
NOD2
IRAK4
α-MSH
IRAK3
CARD9
TOLLIP
IRAK1
Increased IgG titer
UBE2V1
TRAF6
UBE2N
TLR3
dsRNA
NO
Idh1
TBK1
MAP3K7
p38
MAP3K7IP1
MAP3K7IP2
JNK
MAP2K4
TLR7
TLR8
TLR9
ssRNA
CpG
+ more
HCV NS3
CHUK
IKBKB
IKBKG
Feedback
NFAT
AP1
SP1
STAT
NFkB
MYC MAX
cEBP
ELK1
IRFs
T
GA
AGGCTATAGCATCGAT
Gene Products
Increasing complexity
(c)
TRENDS in Immunology

Box 1. What is systems biology?

Systems biology represents a powerful and comprehensive paradigm for biology that stands in contrast to the reductionist approach that has tended to dominate. In the reductionist approach, researchers attempted to understand a complex entity (e.g. a cell, an organ or a disease) by breaking it down into smaller, more tractable units for study, such as genes, complexes or pathways. The reductionist approach was characterised by the concept that a system is the sum of its parts and simply identifying and characterizing these parts would be sufficient to generate predictions about the system's behaviour. This is clearly not the case, however, because the extensive biomolecular cataloguing of the 'omics' era has given rise to many more questions than it has answered.

The systems biology approach regards a system as more than the sum of its parts – its behaviour arises not from simply the presence of its building blocks but through the complex relationships among them. Indeed, systems display what are termed emergent properties, which are behaviours made possible only through the interaction of a system's components, and which cannot be predicted by looking at a single component alone. An ant colony or a flock of birds flying in unison represent two such emergent properties because they are both traits that could not have been predicted easily from observing a single ant or bird.

Understanding of a biological system, its components and its emergent properties benefits from harnessing as many data types as possible, including catalogues of genes, proteins, RNAs, small molecules and cells; their interactions with each other and measurements of the concentration, regulation and behaviour of these entities under a range of conditions. A variety of high-throughput technologies are typically used to collect these measurements, with bioinformatics methods used to warehouse and integrate the data and mathematical modelling approaches used to generate predictions regarding the system's behaviour. These predictions are then tested experimentally – sometimes they are validated and, in other cases, the resulting data are used to further refine the prediction. This iterative cycle of model building and testing continues until a conclusion has been reached.

pathways are only observed in specific cell-types, and there can be marked differences depending on the species. TLR2, for example, is expressed at very low levels in human endothelial cells (ECs) and is sequestered intracellularly, whereas it is expressed at high levels in murine ECs and translocated to the cell surface [16]. Mouse models are often used as surrogates for investigating human disease and infection, despite the fact that the mouse and human innate and adaptive immune systems both differ at several crucial points (reviewed in Ref. [17]). Finally, it is probably not entirely appropriate to consider the innate and adaptive immune systems as separate entities because recent work has blurred the boundaries between the two and revealed many molecules and cell-types that act as bridges between innate and adaptive immunity or as effectors in both systems [18].

Several additional examples of the complexity of the innate response have recently been demonstrated in studies of patients deficient in key innate immunity signalling molecules. Given its role as a key adaptor in TLR signalling, one would expect that patients deficient in myeloid differentiation primary response gene 88 (MyD88) would exhibit markedly hampered immune responses; indeed this prediction is true in the murine model. In humans, however, MyD88-deficient patients are susceptible to pyogenic bacterial infections, but otherwise show normal resistance to most pathogens [19]. Furthermore, the infectious phenotype resolves itself after childhood, as does a similar deficiency in patients with a mutation in the downstream interleukin-1 receptor-associated kinase 4 (IRAK4) gene [20]. These observations point to a level of complexity that goes far beyond our current models and indicate that if we are to truly understand the nature of the immune response, the mode of action of immunomodulatory compounds and the pathogenesis of immune system disorders, we require an analytical approach that takes into account this complexity.

Why employ systems biology approaches?

A system as intricate as innate immunity necessitates the detailed level of investigation provided by systems biology approaches (Box 1). This approach to biological experimentation views a system of interest as not just a set of discrete components, but rather as a complex product of the interactions between these components and their relationship with the surrounding environment. This new experimental paradigm is driven by high-throughput methodologies and would not be possible without the extensive catalogues of genes, proteins and other biomolecules, and their interactions, which have come out of the 'omics' era ('omics' being a catch-all phrase describing high-throughput approaches to analysing biological systems, including genomics, transcriptomics, proteomics, metabolomics and pharmacogenomics). These resources have permitted the development of new approaches, such as the use of genome-wide association studies to identify new susceptibility genes or the use of interaction data to identify more accurate biomarkers for disease progression or outcome.

Although immunologists have utilized the high-throughput experimental techniques used by systems biology for many years, it is only recently that immunology-focused resources and workflows have been developed for downstream analysis of the resulting datasets. These resources are enabling new insights into the complex nature of the innate immune response and are painting a much more detailed picture of immunity than ever before, with recent studies revealing new components of the system, new regulatory mechanisms and new biomarkers of immunological disorders.

Figure 1. Traditional views of innate immune signalling fail to capture its full complexity. **(a)** Illustration of a canonical view of mammalian TLR4 signalling via NFκB, in which the process is represented as a simple linear pathway. **(b)** Demonstration that this simple pathway is under multiple levels of control. Selected examples shown here include: (i) signalling through multiple receptors including TLRs, nucleotide-binding oligomerisation domains (NODs) and dectin-1 [84]; (ii) inhibition of IRF3 phosphorylation by hepatitis C virus protein NS3–4A [85]; (iii) inhibition of nitric oxide by *Staphylococcus aureus* L-lactate dehydrogenase (ldh1) [86]; (iv) IFNγ induction of *P. aeruginosa* virulence factor transcription [15]); (v) inhibition of TLR4 signalling through sequestration of IRAK3 by the neuropeptide α-melanocyte-stimulating hormone (αMSH) [87]; (vi) let7 microRNA-mediated post-transcriptional suppression of TLR4 levels [88]; (vii) SNPs with divergent effects (e.g. S180L and C558T in Toll-interleukin 1 receptor (TIR) domain containing adaptor protein (TIRAP) - S180L confers protection against several infections [89], whereas C558T is associated with an increased risk of meningal tuberculosis infection [90]); (viii) interacting SNPs (e.g. a SNP in each of TLR4, toll interacting protein (TOLLIP) and IRAK1 that interact to increase the immunoglobulin G (IgG) titre in response to pertussis vaccination [91]); and (ix) SNPs in immune gene promoters that lead to disease [92]. **(c)** Illustrates a more accurate depiction of the complexity of the system. InnateDB [39] was used to retrieve interactions for each of the 47 proteins shown in (b) and the resulting network was visualized using the Cerebral [93] plugin in Cytoscape [94]. It contains 1346 genes or proteins and 2531 interactions.

Here, we first highlight some of the emerging resources and workflows available to the innate immunology community, many of which can be implemented by researchers without extensive bioinformatics experience. We also describe recent novel insights into innate immunity that have been generated using these new approaches. We present leading-edge systems studies from other fields that might serve as a source of inspiration for future investigation of the innate immune response, and discuss some of the challenges facing this rapidly growing field. Although the focus of this review is on innate immunity, it is worth mentioning that much of our discussion pertains also to adaptive immunity, as the two systems share receptors, signal transduction pathways, regulatory systems and effector mechanisms, and adaptive immunity often transitions directly from innate immunity. Indeed, virtually all of the resources and workflows presented here are equally applicable to the analysis of adaptive immunity.

The immune response is a complex entity with many possible inputs, influences and outcomes, and as this review demonstrates, systems biology holds the promise of allowing us to both better understand its nature, and generate predictions and hypotheses about its behaviour under particular conditions.

Emerging experimental approaches and computational resources

Transcriptomics: meta-analysis for the reliable measurement of gene expression

Transcriptomics remains one of the most popular methods for the investigation of the immune response on a genome-wide scale, enabling researchers to uncover processes and pathways that are differentially regulated in a condition of interest and/or generate hypotheses about co-regulation of a specific set of genes. Microarrays have led to several new insights into innate immunity both in the past (reviewed in Ref. [21]) and more recently (see later), whereas new technologies, including next-generation sequencing (reviewed in Ref. [22]) and exon and microRNA arrays, represent the future of transcriptomics. Regardless of the technology used, the long lists of differentially expressed genes generated by transcriptomics experiments can be noisy, with many false positives and false negatives. To fully realize the benefits of this data source, therefore, systems biologists have recognized that it is often necessary to integrate multiple datasets from diverse sources for the purposes of meta-analysis, in which datasets are combined and only genes behaving similarly across several independent experiments are considered as true positives.

Meta-analysis requires easy access to relevant data. In addition to resources such as Array Express (Table 1) [23] and the Gene Expression Omnibus (GEO) [24], both of which are generalised repositories for transcriptomics data from all species, conditions and platforms, several groups have created publicly available databases of innate (and adaptive) immunology-relevant transcriptomics datasets. These include the Reference Database of Immune Cells (RefDIC) [25] and the Immune Response *In Silico* database (IRIS) [26]. RefDIC stores Affymetrix GeneChip profiles of unstimulated human and mouse immune cells, whereas IRIS contains Affymetrix data from activated and differentiated human immune cells. A third repository at the Institute for Systems Biology (ISB) contains Affymetrix data derived from TLR ligand-stimulated mouse macrophages (www.innateimmunity-systemsbiology.org) [27].

The aforementioned resources, although valuable, can be difficult to harness for meta-analysis. The larger general repositories contain data generated by many discrete array platforms that can be difficult to integrate, whereas the immune-specific compendia focus on only a small number of cell types or conditions. The recent Immunological Genome Project initiative (Table 1) [28] aims to overcome these obstacles and is the first transcriptomics project to take a truly systems-level approach to the analysis of immune cell populations. A network of laboratories is generating rigorously standardised genome-wide gene expression datasets, profiling over 200 different mouse immune cell populations under a variety of conditions, including genetic polymorphisms, gene knockdowns or knockouts and drug treatments. These data are being made freely available through the project website. Although the project is presently focused on mouse cells and contains more data from lymphoid cell lineages than myeloid cells, the consortium might expand its efforts to humans in the future, and they are actively soliciting suggestions for other cell types for analysis from the immunology research community.

Meta-analysis has already led to new insights regarding a conserved host innate immune response to a range of pathogens. Combining data from 32 *in vitro* human studies enabled the identification of a cluster of 511 genes representing a common host response to bacteria, viruses and selected other pathogens, which included many poorly characterised genes and genes not previously known to be involved in innate immunity [29]. A similar approach was recently implemented to examine mechanisms of lung inflammation *in vivo* using rodent and primate models [30]. A set of core upregulated and downregulated genes was identified as contributing to lung inflammation, regardless of whether the cause was an infection, asthma or an airborne pollutant; many of theses genes overlapped with those found in the earlier study [29].

Transcriptomics will undoubtedly continue to provide novel insights into the innate immune response, although as discussed later, the real power of these data will be realized through its analysis in the context of the interactome and other regulatory networks.

Transcriptional regulation: building genetic regulatory networks

Because TFs bind to short, degenerate sequence motifs that occur by chance throughout the genome sequence, the computational identification of TFs responsible for the regulation of a specific gene is usually confounded by a high number of false positive TF binding sites (TFBSs) [31]. Fortunately, high-throughput approaches for the experimental verification of protein–DNA interactions, including chromatin-immunoprecipitation (ChIP) array platforms (ChIP-chip) and the more recent ChIP-seq technology, in which the array step is replaced by massively parallel sequencing, are emerging as powerful methods for identifying genetic regulatory networks on a genome-wide scale [32].

Table 1. Selected bioinformatics resources for systems-level analysis of innate immunity.

Bioinformatics resource	URL	Details
Immunology-specific resources		
Immune Response *In Silico* (IRIS)	http://share.gene.com/clark.iris.2004/iris/iris.html	Immune specific genes identified from multiple microarray expression datasets.
Innate Immunity Database	http://db.systemsbiology.net/IIDB	Repository of genomic annotations and experimental data for over 2000 mouse immune-related genes derived from over 100 microarray experiments.
Innate Immunity in Heart, Lung and Blood Disease	http://www.innateimmunity.net/	Project to discover and model the associations between nucleotide sequence variations (SNPs and Indels) in the genes of the innate immunity pathway in humans.
InnateDB	http://www.innatedb.ca	Database and analysis platform facilitating systems-level analyses of the innate immune system and beyond.
Macrophages.com	http://www.macrophages.com	Community website for researchers with an interest in macrophage biology.
Reference Database of Immune Cells (RefDIC)	http://refdic.rcai.riken.jp	Database of quantitative mRNA and protein profiles, specifically for immune cells and tissues.
The Immunological Genome Project	http://www.immgen.org	Project generating rigorously standardised genome-wide gene expression datasets in 200 different mouse immune cell populations.
The Immunology Database and Analysis Portal (ImmPort)	http://www.immport.org/	The ImmPort system provides IT support in the production, analysis, archiving and exchange of scientific data for researchers supported by NIAID's DAIT.
The Immunome Database	http://bioinf.uta.fi/Immunome/	Database that contains information about human immunity related proteins, their domain structure and the related ontology terms.
Pathway databases		
Integrating Network Objects with Hierarchies (INOH)	http://www.inoh.org/	Pathway database of model organisms, including human, mouse, rat and others. Currently, 62 pathways.
Kyoto Encyclopedia of Genes and Genomes (KEGG)	http://www.genome.ad.jp/kegg/	Database of biological systems, consisting of genes and proteins, endogenous and exogenous chemicals and ligands and molecular wiring diagrams of interaction and reaction networks.
NCI-Nature Pathway Interaction Database (PID)	http://pid.nci.nih.gov	Biomolecular interactions and cellular processes assembled into authoritative human signalling pathways. Currently, 83 Human Pathways.
NetPath	http://www.netpath.org	Curated resource of signal transduction pathways in humans. 10 immune and 10 cancer signalling-pathways are available.
Pathguide	http://www.pathguide.org/	Comprehensive listing of web-accessible network and pathway resources.
Reactome	http://www.reactome.org/	Curated knowledgebase of biological pathways. Inferred orthologous events in 22 non-human species.
Interaction databases		
Biomolecular Interaction Network Database (BIND)	http://bond.unleashedinformatics.com	Manually curated molecular interaction database now integrated into the commercial BOND database.
Database of Interacting Proteins (DIP)	http://dip.doe-mbi.ucla.edu/	Database of experimentally determined interactions between proteins in 270+ species. 57 000+ interactions.
IntAct	http://www.ebi.ac.uk/intact/	Database system and analysis tools for literature-derived or user-submitted protein interaction data. 180 000+ interactions from a broad range of species.
Molecular Interaction Database (MINT)	http://mint.bio.uniroma2.it/mint/	Database of experimentally verified protein–protein interactions mined from the scientific literature by expert curators. 110 000+ interactions from a broad range of species.
Pathogen Interaction Gateway (PIG)	http://molvis.vbi.vt.edu/pig/	Database dedicated to the study of host-pathogen protein–protein interactions.
The Biological General Repository for Interaction Datasets (BioGRID)	http://www.thebiogrid.org/	Database of protein and genetic interactions from selected model organism species. Over 198 000 interactions from six different species.
VirusMINT	http://mint.bio.uniroma2.it/virusmint/	Database annotating in a structured format interactions between human and viral proteins and integrating this information in the human protein-interaction network.
Network analysis and visualization		
CEll REgion-Based Rendering And Layout (CEREBRAL)	http://www.pathogenomics.ca/cerebral/	Java plugin enhancing Cytoscape's functionality by generating more pathway-like representations of a network and enabling the visualization of gene expression data from multiple conditions.
Cytoscape	http://www.cytoscape.org/	Software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data.
HUB oBjects Anlyzer (HUBBA)	http://hub.iis.sinica.edu.tw/Hubba/	A web-based service designed to explore the essential nodes in a network.

Table 1 (*Continued*)

Bioinformatics resource	URL	Details
Network Analysis Tools (NEAT)	http://rsat.ulb.ac.be/rsat/index_neat.html	Web-based access to a collection of modular tools for the analysis of networks (graphs) and clusters (e.g. microarray clusters, functional classes).
Other useful sites		
Array Express	http://www.ebi.ac.uk/microarray-as/ae/	A public repository for transcriptomics data, which is aimed at storing MIAME- and MINSEQE-compliant gene expression data.
Gene Expression Omnibus	http://www.ncbi.nlm.nih.gov/geo/	A gene expression or molecular abundance repository supporting MIAME compliant data submissions.
International HapMap Project	http://www.hapmap.org/	The goal of the International HapMap Project is to develop a haplotype map of the human genome and to describe the common patterns of genetic variation in humans.
miRBase	http://microrna.sanger.ac.uk	Central online repository for microRNA nomenclature, sequence data, annotation and target prediction. The current release (10.0) contains 5071 miRNA loci from 58 species.

Although these methods, to date, have been implemented to locate binding sites for a small number of TFs at a time, they have already been successfully applied to the study of innate immune regulation. ChIP-chip has been used to validate predicted associations between nuclear factor-κB-1 (NFKB1), interferon response factor 1 (IRF1) and the promoters of co-expressed genes in TLR-stimulated murine macrophages [33]. More recently, this technology was instrumental in the discovery of a cluster of innate immune response genes that are regulated by signal transducer and activator of transcription 3 (STAT3) downstream of the cytokine leukaemia inhibitory factor (LIF) [34]. ChIP-seq has also been employed in the discovery of STAT1 binding sites in interferon-responsive human genes [35].

A combination of comprehensive computational and high-throughput experimental approaches to TFBS identification might represent the most promising approach to the analysis of gene regulatory networks. The ISB maintains the Innate Immune Database (Table 1), which stores both predicted TFBSs, representing consensus predictions from a suite of bioinformatics techniques and ChIP-chip verified TFBSs [27]. The data currently includes 2000 mouse genes known to have a role in the macrophage response to lipopolysaccharide (LPS), but the approach employed is easily extendable to a larger gene set and to human data.

Proteomics: generating the innate immunity interactome

In the same way that transcriptomics is used for the quantification of gene expression at the RNA level and the consequent inference of gene-regulatory networks, proteomics has opened up explorations at the protein level, both quantitatively (reviewed in Ref. [36]) and in the generation of large-scale protein–protein interaction data (reviewed in Ref. [37]). In addition to genome-scale proteomics studies that have attempted to capture a global picture of an host interactome (defined as the network of interacting proteins and other biomolecules) [37], several key studies have begun to collate the vast amount of information specifically regarding the immune interactome, representing the interactions involving the genes and gene products known to participate in the immune response. Until recently, much of this immunity-specific data was present in the biomedical literature, but had not found its way into the most popular publicly available interaction databases (see Table 1 and individual database URLs therein), an endeavour that was crucial to enable systems-level analyses. The first release of innate immunity-specific interaction data occurred with the publication of a manually constructed map of the mammalian TLR signalling network [38]. More recently, our own efforts on the InnateDB project (Table 1) have collated, reviewed, annotated and made available more than 7000 innate immunity-relevant interactions involving 2000 human and mouse genes [39].

As new proteomics data are generated, two emerging areas are likely to expand our knowledge of the innate immunity interactome. First, investigators are now looking beyond the host interactome to that of the pathogen and the host–pathogen interface because recent studies have revealed that many pathogen-encoded proteins directly interact with molecules of the innate immune response. An increasing number of viral proteins, for example, are involved in viral evasion of the innate immune system through interactions with key proteins in innate immunity pathways (reviewed in Ref. [40]). A recent study has provided a proteome-wide map of the interactions between hepatitis C virus-encoded proteins and human proteins using a yeast two-hybrid approach and literature mining [41], whereas the VirusMINT (Table 1) [42] and Pathogen Interaction Gateway (PIG) [43] databases provide online resources that describe known host-pathogen interactions.

Second, more specialized avenues of investigation are providing increasing detail concerning how post-translational modifications and protein dynamics influence the innate immune response. For example, the emerging field of 'phosphoproteomics' will undoubtedly assist in unravelling the dynamics of signalling cascades integral to the innate immune response. Such studies provide a global picture of protein phosphorylation, an important mechanism of cellular signal transduction, with the ultimate goal being the identification of the complete 'kinome' of a cell. The effectiveness of this approach was recently demonstrated by an analysis of Jun terminal kinase (JNK) signalling in *Drosophila melanogaster* [44] using phosphoproteomics and RNAi (see later), which yielded a comprehensive map of this important signalling cascade and several novel activators and repressors.

Genome-wide RNAi screens: high-throughput insight into phenotype

Large-scale RNAi screens, in which genes are progressively knocked down using appropriate inhibitory RNA molecules, have enabled the functional annotation of hundreds or thousands of genes in a single experimental series and have the potential to rapidly identify crucial innate immune signalling genes. A recent study [45] used this technology to screen *Caenorhabditis elegans* chromosome 1 for innate immunity regulators. Murine homologs of the 32 regulators identified were then further analysed, again using RNAi, and 11 key regulators were identified, including some that had not been previously implicated in innate immunity. Similarly, another study used RNAi in human DLD1 colon adenocarcinoma cells to identify regulators of the Wnt/β-catenin signalling pathway [46], a pathway that has been recently implicated as having a key role in the innate immune response [47]. As the authors noted [46], RNAi screens, by themselves, are powerful in identifying genes that contribute in some way to a phenotype but provide few mechanistic insights. To overcome this, both of the aforementioned studies integrated data from the RNAi screens with interaction networks (see discussion later), thereby enabling the identification of an evolutionarily conserved protein interaction network with a role in innate immunity [45] and the discovery of a novel regulator of Wnt/β-catenin signalling, the angiogenic factor with G patch and Forkhead-associated (FHA) domains 1 (AGGF1) [46].

RNAi is also emerging as an important tool to shed light on host-pathogen interactions. This technology was applied recently to identify the host proteins involved in HIV [48] and West Nile virus [49] infections. More than 250 host factors were identified as important in the HIV viral life cycle; many of these were implicated as being involved in innate immune signalling pathways and were found to be highly expressed in immune cells. In the West Nile virus investigation, nearly 300 novel host factors were identified. Again, several innate genes were identified, including β-defensins and IRF3.

miRNAs and innate immunity

miRNAs were discovered about ten years ago and are now recognized as crucial post-transcriptional regulators of gene expression, working through a variety of mechanisms, including mRNA degradation and regulation of translation, for up to 30% of all transcripts. Despite their novelty, there has been tremendous progress in identifying miRNAs and characterizing their functions. miRBase (Table 1) [50], the most comprehensive database of miRNA nomenclature, sequence, annotation and target prediction, currently contains around 6000 miRNA sequences from 58 species and represents an excellent resource for researchers wishing to determine which miRNAs might potentially target their gene(s) of interest. InnateDB [39] is also currently updating its interaction information to include cases where a specific miRNA is known to regulate an innate immunity-relevant gene, providing a more focused view of how this type of regulation can impact immunity than is offered by the more general miRBase database [51].

Microarray-based platforms that enable the profiling of miRNA expression on a large or even global scale are also now available [52] and their impact upon systems biology is just becoming apparent. A recent study investigated the leukocyte miRNA response to LPS stimulation [53]. Five miRNAs were identified as consistently responsive to LPS stimulation, whereas bioinformatics approaches identified more than 30 candidate target genes influenced by these miRNAs that are central to the innate immune response; these included IRAK1 and 2, several members of the mitogen-activated protein (MAP) kinase signalling pathway, multiple interleukins and several key proteins involved in signalling, apoptosis and transcriptional activation.

Genetic polymorphisms in innate immunity genes and the promise of genome-wide association studies

Another layer of complexity that requires consideration in systems-level analyses of innate immunity is the influence of genetic polymorphisms and their effects on gene and protein expression, host-pathogen interactions and molecular signalling (reviewed in Refs. [54–56]). The Innate Immunity in Heart, Lung and Blood Disease Project aims to discover and model associations among nucleotide sequence variations in human innate immunity genes, eventually relating these variations to airway diseases. Over 80 genes have been re-sequenced to date, including several cytokines, chemokines, interleukins and TLRs, with the data and related analysis tools available on their website (Table 1).

Structural variations such as gene copy number variation (CNV) are also of interest, and technologies such as array comparative genomic hybridisation (CGH) are enabling assays of CNV on a genome-wide scale (reviewed in Ref. [57]). For example, β-defensin CNV is strongly associated with the inflammatory skin disease psoriasis [58]. However, the influence of such variations on the innate immune response to pathogens remains largely unexplored.

Genome-wide association (GWA) studies, made possible through the International HapMap Project [59], and the availability of platforms such as the Affymetrix GeneChips and Illumina BeadChips that can genotype up to 1 million SNPs in parallel, are providing a powerful new approach for identifying genes that contribute to disease susceptibility, including autoimmune and chronic inflammatory disorders (reviewed in Refs [60,61]). Although these studies have been successful in identifying several new genes associated with disease, including several innate immune pathway genes [61], in most cases, only a small number of markers are found to be statistically significant, accounting for just a fraction of the estimated genetic components of these diseases. By identifying those pathways or interaction sub-networks that are statistically enriched in these datasets, one might be able to uncover more true-positive associations, including markers that fall below the traditional significance thresholds. The inclusion of network and pathway data also has the potential to provide more profound mechanistic insights into the complex diseases targeted by GWA studies.

Bringing it all together: computational tools and workflows for integrating and analysing disparate data

Examining data in a network context: databases and visualization tools

Simply collecting measurements on a large scale is rarely sufficient for generating novel systems-level insights. Instead, as demonstrated by several of the studies described earlier, measurements must be placed in their proper biological context, a step that frequently involves integrating quantitative data with a biomolecular interaction network.

Statistical analysis of the gene categories and pathways enriched in a gene list is a popular approach to investigating large datasets [39] and can reveal much about the biological processes underlying a phenomenon of interest. Pathway analysis, however, relies on associating genes of interest to known biological pathways, which are limited in scope and are often annotated as basic linear cascades (Figure 1a). By contrast, network analysis can expand this simple perspective on signalling to a more complete picture of the relationships among genes, proteins, RNAs and/or other molecules (Figure 2). By investigating a larger set of interactions, network-based analyses, which might include visualization of interacting networks, topological characterisation and more, have the potential to reveal as-yet unknown signalling cascades or pathways, functionally relevant sub-networks and the central molecules (often called 'hubs' or 'bottlenecks') of these networks. For example, investigation of the mammalian TLR signalling network confirmed MyD88 as a bottleneck protein through which signals from many different TLRs are funnelled, and also demonstrated the existence of multiple positive and negative feedback and feed-forward loops that can be used to characterize and predict the system's dynamics [38]. Similarly, analysis of a network of TFs and the genes they regulate in the response of mouse macrophages to LPS identified several hubs (highly connected molecules or nodes) that represent key TFs, including nuclear factor erythroid derived 2, like 2 (Nrf2), Atf3, E26 avian leukemia oncogene 1 (Ets1) and Irf1 [62].

Advances in the mapping of interactomes (reviewed in Ref. [63]) have led to an explosion in the volume of network and pathway datasets. Pathguide (Table 1) [64], a comprehensive listing of web-accessible network and pathway resources, increased in content by 43% within three years of its initial publication in 2006. With almost 300 unique databases available, selecting a suitable one might seem an intractable problem for researchers. Certain databases, however, have emerged as leading players in the field of interactomics (Table 1). Among the most comprehensive and popular non-commercial interaction databases are IntAct [65], MINT [66] and BioGRID [67], each of which stores more than 100 000 interactions across a range of species and incorporates information in the form of the experimental types and publications associated with each interaction.

Of particular interest to immunologists is InnateDB [39], a recently developed, freely available database and analysis platform. InnateDB integrates data from many of the most popular interaction and pathway databases, including all of those mentioned earlier, into a single resource that consequently represents one of the most comprehensive repositories of human and mouse molecular interactions. Further manual curation has led to the inclusion of many immune-relevant interactions that are not present in other databases. Unlike other websites, InnateDB is also a complete analysis platform, offering comprehensive annotation and database cross-references for each gene and protein, in addition to seamlessly integrated, user-friendly bioinformatics tools. These include a method for identifying pathways (of which InnateDB stores over 3,000) and ontological terms (describing molecular function, biological process and cellular compartment) that are statistically over-represented in a user-specified list of genes of interest. Associated quantitative data, such as gene expression data, can be included from up to ten conditions at one time and can be overlaid on pathways and networks of interest. Users also have the ability to construct orthologous interaction networks in other species and to explore and visualize their data in a network and/or pathway context. Further development is underway, including incorporation of a tool for detecting over-represented transcription factor interactions.

Applying the tools: systems-level studies provide novel insights into innate immunity

Beyond network visualization and topological analysis, more complex workflows are necessary if one wishes to integrate quantitative measurements with interaction data to shed new light on the processes underlying innate immunity (Box 2). In many cases, the starting point for such an analysis is a time-course microarray experiment. With two or more time points and/or multiple conditions, many studies opt to pre-process the data to reduce its complexity, a step most often accomplished through clustering (grouping according to correlated expression patterns; reviewed in Ref. [68]). It is worth mentioning that the identity of individual clusters is highly dependent on the clustering methodology used. However, the advantage of this strategy is the decrease in complexity of subsequent analyses, such that each cluster of interest, rather than the complete dataset, can be used for further analysis. For example, one study [69] used hierarchical clustering to group differentially expressed genes with similar temporal expression profiles in a baboon lung model of *Escherichia coli*-induced sepsis. Selected clusters were used to generate networks that were then examined for functional enrichment and topological features, providing a broad overview of the functional timeline of sepsis. A similar approach was adopted in a study that identified a potential early feedback loop involving suppressor of cytokine signalling 3 (Socs3) in interferon-γ-stimulated murine macrophages [70].

Analysis of an LPS-stimulated murine macrophage model [71] illustrates a particularly insightful approach to investigating the innate immune response to LPS. After clustering by temporal-expression profile, a second set of clusters was created by grouping genes that contained promoter elements with binding sites for similar transcription factors. A transcriptional regulatory network was then constructed, interconnecting individual TFs to the genes in the clusters they were predicted to regulate. Expression

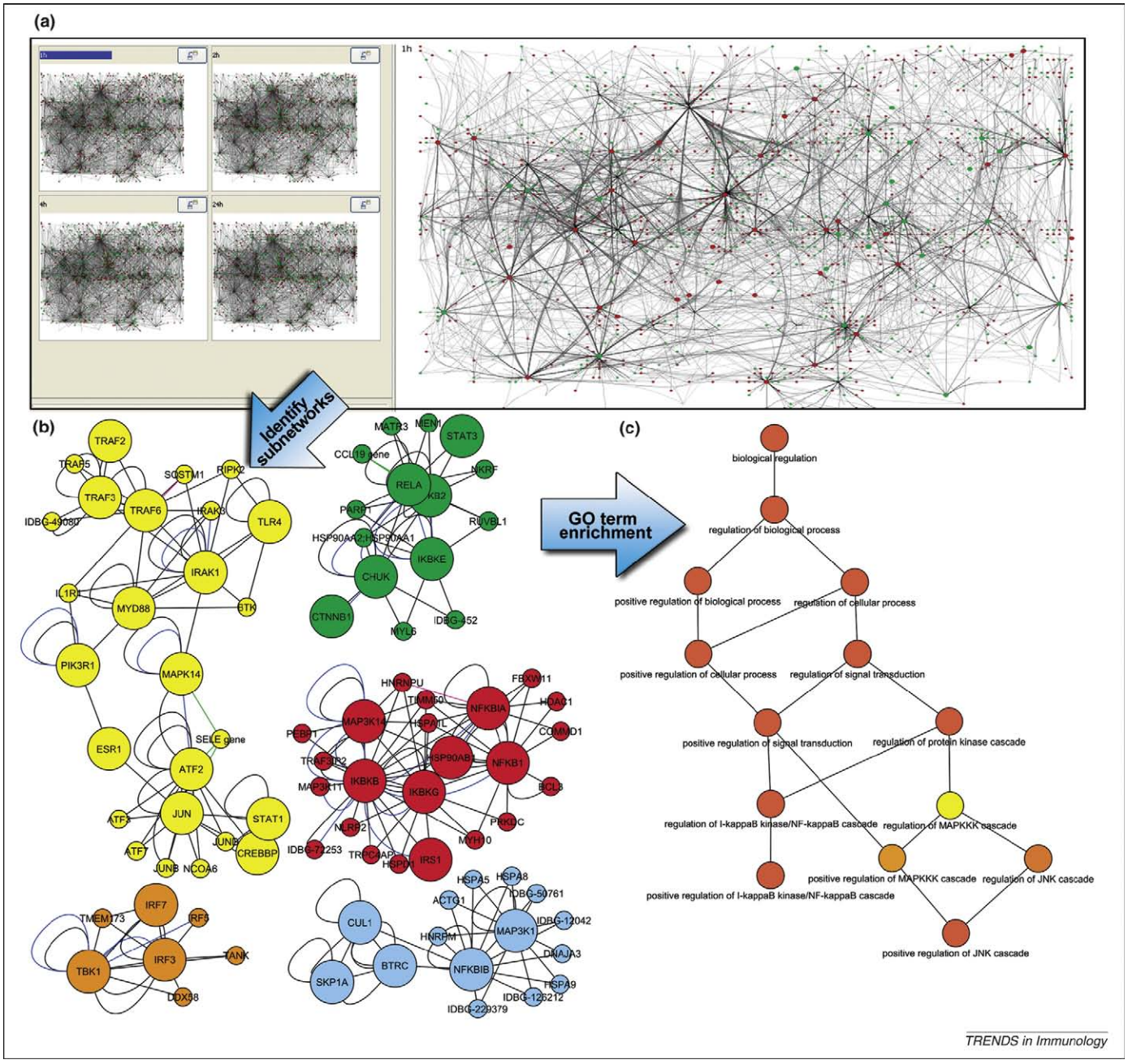


Figure 2. Network-based visualization and analysis. **(a)** Illustrates the interactions among the members of the human TLR signalling pathway retrieved from InnateDB and was visualized in one click directly from the InnateDB website using a webstart version of Cytoscape with the Cerebral plugin installed. The open-source Cytoscape interaction viewer and analysis environment [94] provides extensive capabilities for displaying and editing networks, whereas plugins developed by its user community add novel functionality. Cerebral (CEll REgion Based Rendering And Layout) provides a pathway-like layout for complex networks [93], a view that is perhaps more intuitive for biologists. Here, sample expression data have been included to illustrate Cerebral's ability to display multiple quantitative datasets simultaneously. The top 100 bottleneck nodes as identified by the HUB oBjects Anlyzer, (HUBBA, [95]) are shown as larger nodes in the network. HUBBA and other tools can be used to identify topologically interesting features of a network, such as hubs, bottlenecks and cliques, which are often of biological relevance; hubs might represent key regulatory molecules, bottlenecks typically indicate essential proteins and cliques might represent functional complexes. The MCODE Cytoscape plugin [96] was then used to identify the five most significant sub-networks from this larger network, which are shown in **(b)**. The largest of these clusters (yellow) was then passed to Cytoscape's BiNGO plugin for the analysis of statistically enriched GO (gene ontology) terms [97]. Several over-represented GO terms are significantly associated with this cluster and **(c)** shows a hierarchy of those terms pertaining to signal transduction, in which a deeper orange colour is reflective of a smaller p-value.

data were overlaid onto the transcriptional regulatory network and network cliques (a term for groups of highly connected components that often represent functional protein complexes) that were uniquely responsive at specific time points were further analysed. Three important regulators were predicted: Atf3, which had not been implicated previously in the LPS response; E26 avian leukemia oncogene 2 (Ets2), which activates several inflammatory mediators, and nuclear factor (erythroid-derived 2)-like 2 (Nfe2l2), which was predicted to form a regulatory circuit with Atf3 through kelch-like ECH-associated protein 1 (Keap1).

Several studies have taken an even more detailed approach to the systems-level investigation of innate immunity, beginning with a landmark paper from the ISB in Seattle, USA [72]. In a time-course study of the

Box 2. Incorporating bioinformatics resources into a systems workflow

As the studies reviewed here demonstrate, a successful systems biology workflow spans both the bench and the computer. After high-throughput collection of experimental data measuring transcript, protein or phosphorylation levels across multiple time points and/or conditions, the data must be analysed computationally, using all or some of the steps described below; to investigate the possible involvement of key groups of genes, pathways, transcription factors, biological processes etc, in particular immunological events. For each step, we have noted the computational resources that could be employed (See Table 1 for URLs).

- **Retrieval of existing gene expression data for meta-analysis:** Immunological Genome Project, IRIS, IIDB, RefDIC, ArrayExpress, GEO
- **Clustering and pre-processing of data to reduce complexity:** Cytoscape with Cerebral or clusterMaker plugin, numerous independent clustering tools available online.
- **Functional enrichment analysis based on functions (ontology) or features (e.g. involvement of receptors, pathways or transcription factors):** InnateDB (pathway and gene ontology enrichment), IIDB (TFBS enrichment), Cytoscape with BiNGO plugin (gene ontology enrichment).
- **Retrieval of a comprehensive interaction network surrounding the genes of interest:** InnateDB (interactions, pathways), IntAct (interactions), BioGRID (interactions), MINT (interactions), BIND (interactions), DIP (interactions), VirusMINT (host-pathogen interactions), PIG (host-pathogen interactions), INOH (pathways), KEGG (pathways), PID (pathways), NetPath (pathways), Reactome (pathways).
- **Network visualization and data overlay:** Cytoscape with Cerebral or VistaClara plugin.
- **Topological or module-based analysis of the network to identify key regulators, signalling networks and functional complexes:** Cytoscape with MCODE or jActiveModules plugins, HUBBA, NEAT.
- **Addition of information beyond the interactome:** miRBase (miRNAs and their target genes), Immunome (SNPs), HapMap (SNPs).

murine macrophage response to LPS, a cluster of early responsive TFs was identified and, from this, a target factor (Atf3) was selected for further analysis. Transcription factors known to interact with Atf3 were retrieved and the promoters of genes in a second early response cluster were scanned for potential binding sites for Atf3 and its interactors. Of these genes, those encoding interleukin 6 (*Il6*) and *Il12b* were selected for follow-up using mathematical modelling of Atf3 and reticuloendotheliosis oncogene (Rel) binding to the promoters of these genes. Predictions made by the model were then tested *in vivo*, confirming that Atf3 acts as a negative regulator of the *Il6* and *Il12b* genes and is part of a negative feedback loop in response to Tlr4 stimulation. The model was later expanded to include a second regulator, Cebpd [11].

A third time-course experiment from the ISB investigated the effect of multiple TLR agonists on the murine macrophage response [33]. Early response TFs were identified through clustering, and downstream target genes were identified based on the enrichment of these transcription factor binding sites in the promoters of genes in delayed response clusters, by utilizing a signal-processing measurement termed time-lagged correlation (TLC). A regulatory network was constructed and expanded through the inclusion of TF interaction data from publicly available sources. This study implicated TGFB-induced factor homeobox 1 (Tgif1) in macrophage activation.

Several other studies have implemented variations on these approaches. An analysis of publicly available asthmatic-mouse-lung expression datasets [73] used an unsupervised learning algorithm called Module Networks [74] to predict potential regulatory modules and consequently implicated serum amyloid A 3 (*Saa3*) and several other genes in lung inflammation. By retrieving interactors of module members and adding them to the network, insight was gained into the potential regulators of the Il13 pathway, which is important in both lung inflammation and the innate and adaptive immune responses [75]. Newly identified regulators of Il13 signalling included transforming growth factor β 1 (Tgfb1) and Jun-B oncogene (Junb).

New directions: inspiration from studies outside immunology

As immunologists move towards more complex systems-level analyses, it is worthwhile looking to leading-edge studies in other disciplines for inspiration. One recent study [76] has demonstrated how insights derived from a network of functional associations can be complemented with data from other disparate sources to identify novel disease-susceptibility genes. A network was constructed containing four known breast cancer susceptibility genes and 114 genes co-expressed with these genes in a large number of patients. Each gene was connected to others in the network on the basis of multiple types of functional associations derived from data across multiple species. The 114 genes were then ranked based on the strength of their functional associations with the four known risk genes and high-ranking genes were selected for downstream analysis. A yeast two-hybrid experiment was then performed to generate a physical interaction map centred on one of these, the hyaluronan-mediated motility receptor (HMMR). Several HMMR interactors were known to form complexes with the breast cancer 1 (BRCA1) protein, suggesting a physical association that was confirmed subsequently by further experimentation. Knockdown studies employing small interfering RNA (siRNA) revealed a genetic interaction between the two genes and pointed towards a role for HMMR overexpression in centrosome amplification and tumorigenesis. SNP analysis of the HMMR gene revealed three SNPs, all associated with an increased risk of breast cancer, leading the authors to conclude that HMMR is a novel breast-cancer susceptibility gene. This study thus provides a potential roadmap for researchers in innate immunity to identify important new regulators of pathways involved in the innate immune response. One might, for example, identify all the potential interactors of proteins involved in TLR signalling and prioritise those with multiple lines of evidence supporting the interaction and those that have no known role in the pathway. Methods (for example, siRNA) could then be used to knockdown the candidate genes and the effects on downstream components of TLR signalling, such as NF-κB activation, could be assayed.

Systems-level analyses are not simply useful for identifying novel regulators; they can also be used for classification. Unsupervised clustering was recently applied to identify distinct phenotypic classes of stem cells based on 153 microarrays [77]. Using the MATISSE algorithm [78], which combines expression data and interaction network analysis to identify enriched sub-networks, the authors constructed PluriNet, an interaction network containing genes and proteins that are crucial to pluripotency. Applying a similar approach to microarray data that are being generated through the Immunological Genome Project is likely to be able to successfully identify distinct immune cell phenotypes and the essential sub-networks that differentiate them from each other.

New imaging technologies are also enabling systems-level investigations of individual cells. Recently, proteomics and imaging technologies have been combined to visualize the behaviour of individual cancer cells responding to a drug, with respect to both space and time [79]. Using a library of over 1000 unique fluorescence-tagged gene products, individual cells were treated with camptothecin and observed for 48 h. By observing the changes in expression and localisation of the tagged proteins, the researchers were able to draw several interesting conclusions regarding the mode of action of camptothecin. This study marks one of the first investigations to use large-scale imaging successfully in a systems analysis, and a recent review of imaging techniques suggests that this field holds much potential for advancing systems biology in the coming years, particularly with respect to quantifying cellular dynamics [80].

At present, our view of the innate immune response is largely based on static measurements of what is, in fact, a very dynamic system. These measurements are typically captured through system-wide gene expression studies, which are limited to assessing transcript levels at selected time points. These time points are usually so far apart that the fine detail of the dynamics of the response is lost. Furthermore, simply capturing transcript levels does not reflect changes in the localization of proteins, which, as illustrated by transcription factors translocating between the nucleus and cytoplasm, is a crucial aspect of signalling. Furthermore, recent studies have highlighted the importance of regulation through rapidly changing interaction and feedback loop kinetics in determining the nature of the response to a given stimulus [11]. These new imaging technologies offer the possibility of monitoring many of the key proteins in innate immunity in real time and space and promise to provide a far more comprehensive picture of the dynamics of the response.

Towards new insights: the future of innate immunity systems biology

Although systems biology holds considerable promise for discovery and new insights into processes as complex as innate immunity, the road forward is not without obstacles. These challenges (discussed in some detail in Box 3) are often not specific to the study of innate immunity but represent challenges for systems biology at large. They include the need for greater coverage of network data; increased accuracy of available information and the implementation of appropriate data and experimental reporting standards; improvements in the integration of diverse data types and, perhaps most importantly, shifting the mindset of researchers away from reductionism to more holistic systems biology approaches.

Much of the information on the molecular interactions and signalling pathways involved in innate immunity lies buried in the biomedical literature where it is inaccessible for systems biology approaches. Researchers should be encouraged by the journals to submit annotation on new interactions to any of the widely used interaction databases.

Box 3. Outstanding Questions

Coverage

Only an estimated 15% of the human interactome is known at present [98]. Furthermore, many known biomolecular interactions are not annotated in any of the most popular databases. Curation efforts, such as in InnateDB [39], are part of the solution, but manual review of the literature is time-consuming. Until a more complete literature-supported interactome is available, a variety of approaches are being implemented to infer probable interactions from other data. These include interactions that are inferred, based on orthology [99], co-expression [100] and co-evolution-based predictions [101] and predictions based on the co-occurrence of multiple features, such as motifs, domains, common subcellular localisation and ontology [102] [103].

Accuracy

An interaction observed in a single large-scale study, such as a yeast two-hybrid screen, is less reliable than an interaction observed in several different studies. Although the leading interaction databases take care to note the amount of evidence for given interactions, the experimental approaches used to identify that interaction, and the associated publication(s), there is no physical interaction database as yet that integrates this information into a scoring scheme. Although experienced users often manually screen out interactions that they judge to be potential false-positives, novice users might not appreciate the distinction between high- and low-quality interaction data. Implementing an edge-weighting scheme, such as that employed by the functional interaction database STRING (www.string-db.org) [104], would facilitate the distinction between higher and lower quality interactions. Furthermore, interactions are context-dependent; they do not occur in all cell or tissue types and depend on specific conditions. Many interactions already in the databases do not have such contextual annotation available and, indeed, much of the detail required (MIMIx guidelines [105]) to accurately annotate an interaction is frequently missing in published articles [106].

Integration

To date, many studies have focused on either protein–protein networks or gene regulatory networks and these two types of networks are rarely combined into a single comprehensive view. With the availability of large-scale data on miRNA-target gene interactions and the phosphoproteomics-derived 'kinome', the ability to integrate these diverse datasets must be a priority. Data standards will be crucial to enable the sharing and integration of such disparate datasets (reviewed in Ref. [107]), as is the need for user-friendly platforms that enable biologists, who have the in-depth knowledge of their systems but not the computational background, to intuitively navigate and analyse complex datasets.

The data resulting from the above steps is used to generate hypotheses, which can then be tested experimentally and either confirmed or refined through further analysis.

As discussed earlier, there are notable differences in the innate immune response in humans and mice. Systems biology approaches must be cognisant of these differences, but it is crucial for innate immunity researchers to clearly specify the species under investigation; this lack of clarity in experimental papers is one of the most prevalent roadblocks in curation efforts undertaken within InnateDB.

Systems biology approaches to innate immunity also need to account for differences in the interactome and pathways in specific cell-types, and must recognize that cell population dynamics undoubtedly shape specific responses. Signalling in a macrophage, for example, might be very different to signalling in a dendritic cell. In the same light, the interactome is not a static entity; it is dynamic and changes according to the specific temporal expression of genes and proteins, and in response to particular influences.

As we have discussed in this review article, despite these obstacles, systems biology approaches to investigating innate immunity hold the promise of providing important new insights into this first line of defence against invading pathogens and how dysregulation of innate immunity leads to uncontrolled inflammation and disease. Systems biology not only has the potential to accelerate the discovery of new regulators of innate immunity but will also provide more comprehensive insights into the kinetics of regulation at the transcriptional, protein–protein interaction and post-transcriptional levels. Through an understanding of such kinetics, together with consideration of the innate response as a complex network of interactions, we will probably make tremendous strides in understanding many aspects of this response that are presently not well-characterized. We might, for example, begin to understand how innate immunity can distinguish between different pathogens and danger signals to mount an appropriate response, despite having a much smaller repertoire of receptors and diversity than is utilized in the adaptive response. Systems biology approaches will also enable further leaps forward in our understanding of how certain pathogens manipulate the signalling networks of the host response to their own benefit. Such approaches might also be key to revealing how polymorphisms in the host signalling and gene regulation leading to variation in susceptibility to disease.

One of the ultimate goals of systems biology is the *in silico* modelling of a system and prediction of the effects of a perturbation, such as a gene knockout or the presence of a therapeutic compound. Although complex, progress in modelling of innate immunity is being made [81], including recent mathematical models of IκB-NFκB signalling [82] and systemic inflammation [83]. With the accelerating pace of systems-level investigations of innate immunity, the possibility of an *in silico* immune system – a computational environment where one can predict how modulations of the system will alter the response – moves ever closer.

Acknowledgements

We wish to thank Mr Raymond Lo for his kind assistance with the manuscript. The authors' systems biology work has been funded by Genome Canada and Genome BC through the Pathogenomics of Innate Immunity (PI2) project and by the Foundation for the National Institutes of Health and the Canadian Institutes of Health Research under the Grand Challenges in Global Health Research Initiative (Grand Challenges ID: 419). D.J.L and J.L.G hold Postdoctoral Trainee Awards from the Michael Smith Foundation for Health Research (MSFHR) and J.L.G also holds a Sanofi Pasteur CIHR fellowship. F.S.L.B is a Canadian Institutes of Health Research (CIHR) New Investigator and a MSFHR Senior Scholar. R.E.W.H holds a Canada Research Chair (CRC).

References

1 Zanker, K.S. (2008) General introduction to innate immunity: Dr. Jekyl/M$_r$. Hyde quality of the innate immune system. *Contrib. Microbiol.* 15, 12–20
2 Akira, S. (2006) TLR signaling. *Curr. Top. Microbiol. Immunol.* 311, 1–16
3 Kanneganti, T.D. *et al.* (2007) Intracellular NOD-like receptors in host defense and disease. *Immunity* 27, 549–559
4 Thompson, A.J. and Locarnini, S.A. (2007) Toll-like receptors, RIG-I-like RNA helicases and the antiviral innate immune response. *Immunol. Cell Biol.* 85, 435–445
5 Smith, K.D. and Bolouri, H. (2005) Dissecting innate immune responses with the tools of systems biology. *Curr. Opin. Immunol.* 17, 49–54
6 Brikos, C. and O'Neill, L.A. (2008) Signalling of toll-like receptors. *Handb Exp Pharmacol* 21–50
7 Hu, X. *et al.* (2008) Regulation of interferon and Toll-like receptor signaling during macrophage activation by opposing feedforward and feedback inhibition mechanisms. *Immunol. Rev.* 226, 41–56
8 Bi, Y. *et al.* (2009) MicroRNAs: novel regulators during the immune response. *J. Cell. Physiol.* 218, 467–472
9 Sunnerhagen, P. (2007) Cytoplasmatic post-transcriptional regulation and intracellular signalling. *Mol. Genet. Genomics* 277, 341–355
10 Walhout, A.J. (2006) Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. *Genome Res.* 16, 1445–1454
11 Litvak, V. *et al.* (2009) Function of C/EBPdelta in a regulatory circuit that discriminates between transient and persistent TLR4-induced signals. *Nat Immunol* 10, 437–443
12 Reschner, A. *et al.* (2008) Innate lymphocyte and dendritic cell cross-talk: a key factor in the regulation of the immune response. *Clin. Exp. Immunol.* 152, 219–226
13 Kawaguchi, Y. *et al.* (2007) Contribution of single nucleotide polymorphisms of the IL1A gene to the cleavage of precursor IL-1alpha and its transcription activity. *Immunogenetics* 59, 441–448
14 Ehrt, S. *et al.* (2001) Reprogramming of the macrophage transcriptome in response to interferon-gamma and Mycobacterium tuberculosis: signaling roles of nitric oxide synthase-2 and phagocyte oxidase. *J. Exp. Med.* 194, 1123–1140
15 Wu, L. *et al.* (2005) Recognition of host immune activation by Pseudomonas aeruginosa. *Science* 309, 774–777
16 Shuang, Chen. *et al.* (2007) Differential expression of Toll-like receptor 2 (TLR2) and responses to TLR2 ligands between human and murine vascular endothelial cells. *J. Endotoxin Res.* 13, 281–296
17 Mestas, J. and Hughes, C.C. (2004) Of mice and not men: differences between mouse and human immunology. *J. Immunol.* 172, 2731–2738
18 Borghesi, L. and Milcarek, C. (2007) Innate versus adaptive immunity: a paradigm past its prime? *Cancer Res.* 67, 3989–3993
19 von Bernuth, H. *et al.* (2008) Pyogenic bacterial infections in humans with MyD88 deficiency. *Science* 321, 691–696
20 Casanova, J.L. *et al.* (2008) Revisiting human primary immunodeficiencies. *J. Intern. Med.* 264, 115–127
21 Ricciardi-Castagnoli, P. and Granucci, F. (2002) Opinion: Interpretation of the complexity of innate immune responses by functional genomics. *Nat. Rev. Immunol.* 2, 881–889
22 Wang, Z. *et al.* (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63
23 Parkinson, H. *et al.* (2009) ArrayExpress update–from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res.* 37, D868–D872
24 Barrett, T. *et al.* (2009) NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.* 37, D885–D890

25 Hijikata, A. *et al.* (2007) Construction of an open-access database that integrates cross-reference information from the transcriptome and proteome of immune cells. *Bioinformatics* 23, 2934–2941
26 Abbas, A.R. *et al.* (2005) Immune response *in silico* (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun.* 6, 319–331
27 Korb, M. *et al.* (2008) The Innate Immune Database (IIDB). *BMC Immunol.* 9, 7
28 Heng, T.S. and Painter, M.W. (2008) The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* 9, 1091–1094
29 Jenner, R.G. and Young, R.A. (2005) Insights into host responses against pathogens from transcriptional profiling. *Nat. Rev. Microbiol.* 3, 281–294
30 Pennings, J.L. *et al.* (2008) Identification of a common gene expression response in different lung inflammatory diseases in rodents and macaques. *PLoS One* 3, e2596
31 Kolchanov, N.A. *et al.* (2007) Combined experimental and computational approaches to study the regulatory elements in eukaryotic genes. *Brief. Bioinform.* 8, 266–274
32 Collas, P. and Dahl, J.A. (2008) Chop it, ChIP it, check it: the current status of chromatin immunoprecipitation. *Front. Biosci.* 13, 929–943
33 Ramsey, S.A. *et al.* (2008) Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics. *PLOS Comput. Biol.* 4, e1000021
34 Langlais, D. *et al.* (2008) Regulatory network analyses reveal genome-wide potentiation of LIF signaling by glucocorticoids and define an innate cell defense response. *PLoS Genet.* 4, e1000224
35 Robertson, G. *et al.* (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* 4, 651–657
36 Phizicky, E.M. and Grayhack, E.J. (2006) Proteome-scale analysis of biochemical activity. *Crit. Rev. Biochem. Mol. Biol.* 41, 315–327
37 Kuroda, K. *et al.* (2006) Systems for the detection and analysis of protein-protein interactions. *Appl. Microbiol. Biotechnol.* 71, 127–136
38 Oda, K. and Kitano, H. (2006) A comprehensive map of the toll-like receptor signaling network. *Mol Syst Biol* 2, 15
39 Lynn, D.J. *et al.* (2008) InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Mol. Syst. Biol.* 4, 218
40 Bowie, A.G. and Unterholzner, L. (2008) Viral evasion and subversion of pattern-recognition receptor signalling. *Nat. Rev. Immunol.* 8, 911–922
41 de Chassey, B. *et al.* (2008) Hepatitis C virus infection protein network. *Mol. Syst. Biol.* 4, 230
42 Chatr-aryamontri, A. *et al.* (2009) VirusMINT: a viral protein interaction database. *Nucleic Acids Res.* 37, D669–D673
43 Driscoll, T. *et al.* (2009) PIG–the pathogen interaction gateway. *Nucleic Acids Res.* 37, D647–D650
44 Bakal, C. *et al.* (2008) Phosphorylation networks regulating JNK activity in diverse genetic backgrounds. *Science* 322, 453–456
45 Alper, S. *et al.* (2008) Identification of innate immunity genes and pathways using a comparative genomics approach. *Proc. Natl. Acad. Sci. U. S. A.* 105, 7016–7021
46 Major, M.B. *et al.* (2008) New regulators of Wnt/beta-catenin signaling revealed by integrative molecular screening. *Sci. Signal.* 1, ra12
47 Irazoqui, J.E. *et al.* (2008) Role for beta-catenin and HOX transcription factors in *Caenorhabditis elegans* and mammalian host epithelial-pathogen interactions. *Proc. Natl. Acad. Sci. U. S. A.* 105, 17469–17474
48 Brass, A.L. *et al.* (2008) Identification of host proteins required for HIV infection through a functional genomic screen. *Science* 319, 921–926
49 Krishnan, M.N. *et al.* (2008) RNA interference screen for human genes associated with West Nile virus infection. *Nature* 455, 242–245
50 Griffiths-Jones, S. *et al.* (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.* 36, D154–D158
51 Pedersen, I. and David, M. (2008) MicroRNAs in the immune response. *Cytokine* 43, 391–394
52 Yin, J.Q. *et al.* (2008) Profiling microRNA expression with microarrays. *Trends Biotechnol.* 26, 70–76
53 Schmidt, W.M. *et al.* (2009) *In vivo* profile of the human leukocyte microRNA response to endotoxemia. *Biochem. Biophys. Res. Commun.* 380, 437–441
54 Misch, E.A. and Hawn, T.R. (2008) Toll-like receptor polymorphisms and susceptibility to human disease. *Clin. Sci. (Lond.)* 114, 347–360
55 Dickinson, A.M. and Holler, E. (2008) Polymorphisms of cytokine and innate immunity genes and GVHD. *Best Pract. Res. Clin. Haematol.* 21, 149–164
56 Carneiro, L.A. *et al.* (2008) Nod-like proteins in inflammation and disease. *J. Pathol.* 214, 136–148
57 Emanuel, B.S. and Saitta, S.C. (2007) From microscopes to microarrays: dissecting recurrent chromosomal rearrangements. *Nat. Rev. Genet.* 8, 869–883
58 Hollox, E.J. *et al.* (2008) Psoriasis is associated with increased beta-defensin genomic copy number. *Nat. Genet.* 40, 23–25
59 Frazer, K.A. *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851–861
60 Xavier, R.J. and Rioux, J.D. (2008) Genome-wide association studies: a new window into immune-mediated diseases. *Nat. Rev. Immunol.* 8, 631–643
61 Zhernakova, A. *et al.* (2009) Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat. Rev. Genet.* 10, 43–55
62 Tegner, J. *et al.* (2006) Systems biology of innate immunity. *Cell. Immunol.* 244, 105–109
63 Charbonnier, S. *et al.* (2008) The social network of a cell: recent advances in interactome mapping. *Biotechnol. Annu. Rev.* 14, 1–28
64 Bader, G.D. *et al.* (2006) Pathguide: a pathway resource list. *Nucleic Acids Res.* 34, D504–D506
65 Kerrien, S. *et al.* (2007) IntAct–open source resource for molecular interaction data. *Nucleic Acids Res.* 35, D561–D565
66 Chatr-aryamontri, A. *et al.* (2007) MINT: the Molecular INTeraction database. *Nucleic Acids Res.* 35, D572–D574
67 Breitkreutz, B.J. *et al.* (2007) The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res.*
68 Kerr, G. *et al.* (2008) Techniques for clustering gene expression data. *Comput. Biol. Med.* 38, 283–293
69 Zhu, H. *et al.* (2007) Temporal dynamics of gene expression in the lung in a baboon model of *E. coli* sepsis. *BMC Genomics* 8, 58
70 Raza, S. *et al.* (2008) A logic-based diagram of signalling pathways central to macrophage activation. *BMC Syst. Biol.* 2, 36
71 Nilsson, R. *et al.* (2006) Transcriptional network dynamics in macrophage activation. *Genomics* 88, 133–142
72 Gilchrist, M. *et al.* (2006) Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4. *Nature* 441, 173–178
73 Novershtern, N. *et al.* (2008) A functional and regulatory map of asthma. *Am. J. Respir. Cell Mol. Biol.* 38, 324–336
74 Segal, E. *et al.* (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.* 34, 166–176
75 Brubaker, J.O. and Montaner, L.J. (2001) Role of interleukin-13 in innate and adaptive immunity. *Cell Mol Biol (Noisy-le-grand)* 47, 637–651
76 Pujana, M.A. *et al.* (2007) Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat. Genet.* 39, 1338–1349
77 Muller, F.J. *et al.* (2008) Regulatory networks define phenotypic classes of human stem cell lines. *Nature* 455, 401–405
78 Ulitsky, I. and Shamir, R. (2007) Identification of functional modules using network topology and high-throughput data. *BMC Syst. Biol.* 1, 8
79 Cohen, A.A. *et al.* (2008) Dynamic proteomics of individual cancer cells in response to a drug. *Science* 322, 1511–1516
80 Kherlopian, A.R. *et al.* (2008) A review of imaging techniques for systems biology. *BMC Syst. Biol.* 2, 74
81 Klauschen, F. *et al.* (2007) Understanding diseases by mouse click: the promise and potential of computational approaches in Systems Biology. *Clin. Exp. Immunol.* 149, 424–429
82 Cheong, R. *et al.* (2008) Understanding NF-kappaB signaling via mathematical modeling. *Mol. Syst. Biol.* 4, 192
83 Foteinou, P.T. *et al.* (2009) Modeling endotoxin-induced systemic inflammation using an indirect response approach. *Math. Biosci.* 217, 27–42
84 Underhill, D.M. (2007) Collaboration between the innate immune receptors dectin-1, TLRs, and Nods. *Immunol. Rev.* 219, 75–87

85 Otsuka, M. *et al.* (2005) Interaction between the HCV NS3 protein and the host TBK1 protein leads to inhibition of cellular antiviral responses. *Hepatology* 41, 1004–1012
86 Richardson, A.R. *et al.* (2008) A nitric oxide-inducible lactate dehydrogenase enables Staphylococcus aureus to resist innate immunity. *Science* 319, 1672–1676
87 Taylor, A.W. (2005) The immunomodulating neuropeptide alpha-melanocyte-stimulating hormone (alpha-MSH) suppresses LPS-stimulated TLR4 with IRAK-M in macrophages. *J. Neuroimmunol.* 162, 43–50
88 Chen, X.M. *et al.* (2007) A cellular micro-RNA, let-7i, regulates Toll-like receptor 4 expression and contributes to cholangiocyte immune responses against Cryptosporidium parvum infection. *J. Biol. Chem.* 282, 28929–28938
89 De Jager, P.L. *et al.* (2007) The role of the Toll receptor pathway in susceptibility to inflammatory bowel diseases. *Genes Immun.* 8, 387–397
90 Hawn, T.R. *et al.* (2006) A polymorphism in Toll-interleukin 1 receptor domain containing adaptor protein is associated with susceptibility to meningeal tuberculosis. *J. Infect. Dis.* 194, 1127–1134
91 Kimman, T.G. *et al.* (2008) Association of interacting genes in the toll-like receptor signaling pathway and the antibody response to pertussis vaccination. *PLoS One* 3, e3665
92 Smith, A.J. and Humphries, S.E. (2008) Cytokine and cytokine receptor gene polymorphisms and their functionality. *Cytokine Growth Factor Rev*.
93 Barsky, A. *et al.* (2007) Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. *Bioinformatics* 23, 1040–1042
94 Shannon, P. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504
95 Lin, C.Y. *et al.* (2008) Hubba: hub objects analyzer–a framework of interactome hubs identification for network biology. *Nucleic Acids Res.* 36, W438–443
96 Bader, G.D. and Hogue, C.W. (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4, 2
97 Maere, S. *et al.* (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21, 3448–3449
98 Bader, S. *et al.* (2008) Interaction networks for systems biology. *FEBS Lett.* 582, 1220–1224
99 Ramani, A.K. *et al.* (2008) A map of human protein interactions derived from co-expression of human mRNAs and their orthologs. *Mol. Syst. Biol.* 4, 180
100 Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted gene co-expression network analysis. *BMC Bioinformatics* 9, 559
101 Juan, D. *et al.* (2008) High-confidence prediction of global interactomes based on genome-wide coevolutionary networks. *Proc. Natl. Acad. Sci. U. S. A.* 105, 934–939
102 Rhodes, D.R. *et al.* (2005) Probabilistic model of the human protein-protein interaction network. *Nat. Biotechnol.* 23, 951–959
103 Marcatili, P. *et al.* (2008) The MoVIN server for the analysis of protein interaction networks. *BMC Bioinformatics* 9 (Suppl 2), S11
104 Jensen, L.J. *et al.* (2009) STRING 8–a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* 37, D412–D416
105 Orchard, S. *et al.* (2007) The minimum information required for reporting a molecular interaction experiment (MIMIx). *Nat. Biotechnol.* 25, 894–898
106 Cusick, M.E. *et al.* (2009) Literature-curated protein interaction datasets. *Nat. Methods* 6, 39–46
107 Brazma, A. *et al.* (2006) Standards for systems biology. *Nat. Rev. Genet.* 7, 593–605