Network representations of immune system complexity

Naeha Subramanian,1,2* Parizad Torabi-Parizi,2,3 Rachel A. Gottschalk,2 Ronald N. Germain2 and Bhaskar Dutta2*

The mammalian immune system is a dynamic multiscale system composed of a hierarchically organized set of molecular, cellular, and organismal networks that act in concert to promote effective host defense. These networks range from those involving gene regulatory and protein–protein interactions underlying intracellular signaling pathways and single-cell responses to increasingly complex networks of in vivo cellular interaction, positioning, and migration that determine the overall immune response of an organism. Immunity is thus not the product of simple signaling events but rather nonlinear behaviors arising from dynamic, feedback-regulated interactions among many components. One of the major goals of systems immunology is to quantitatively measure these complex multiscale spatial and temporal interactions, permitting development of computational models that can be used to predict responses to perturbation. Recent technological advances permit collection of comprehensive datasets at multiple molecular and cellular levels, while advances in network biology support representation of the relationships of components at each level as physical or functional interaction networks. The latter facilitate effective visualization of patterns and recognition of emergent properties arising from the many interactions of genes, molecules, and cells of the immune system. We illustrate the power of integrating ‘omics’ and network modeling approaches for unbiased reconstruction of signaling and transcriptional networks with a focus on applications involving the innate immune system. We further discuss future possibilities for reconstruction of increasingly complex cellular- and organism-level networks and development of sophisticated computational tools for prediction of emergent immune behavior arising from the concerted action of these networks. Published 2015. This article is a U.S. Government work and is in the public domain in the USA.

How to cite this article:

INTRODUCTION

Decades of reductionist biological study have resulted in cataloging the enormous numbers of components (genes, gene products, metabolic intermediates, macromolecules, and cells) that make up a living creature, especially complex metazoans. In an effort to place these extensive lists of parts into a coherent framework that helps illuminate their functional and/or structural relationships and that can also help provide insight into how perturbations of the cellular or organismal environment lead to changes in cell or organism behavior, biologists have increasingly invested in building network models of these relationships. A network is an artificial construct that organizes complex multivariate relations in a structured format for systematic mathematical analysis and intuitive visual representation. Nodes and edges are the building blocks of any

*Correspondence to: bhaskar.dutta@nih.gov, naeha.subramanian@systemsbiology.org
1Institute for Systems Biology, Seattle, WA, USA
2Laboratory of Systems Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA
3Critical Care Medicine Department, Clinical Center, National Institutes of Health, Bethesda, MD, USA

Conflict of interest: The authors have declared no conflicts of interest for this article.
network, also called a graph, where nodes represent variables and edges represent relationships between the variables. As this is a fairly general concept, network analysis has been applied to several fields, including sociology, air-traffic, electric power grids, and counter-terrorism. In recent times, network analysis has also been applied extensively in biology, especially in computational and systems biology research. In most biological networks, nodes are macromolecules, such as genes, RNAs, proteins, metabolites, or cells, and a given network can be composed of more than one type of macromolecule or cell. The edges in these networks can depict connections between entities under general or specific conditions, and can represent anything from a physical interaction to regulatory or functional relationships.

Comprehensive reviews have appeared on molecular networks and their applications. In this article, we focus specifically on network concepts in the context of the immune system, which is composed of many dynamic, multiscalar processes. Immune cells sense infection and other environmental cues through a variety of extracellular and intracellular receptors. Ligation of these receptors leads to signaling cascades consisting of many dynamic processes including signal-induced protein binding, phosphorylation, degradation, and nuclear localization. These signaling events lead to changes in gene expression, and subsequently to the production of both effector proteins required to combat infection and proteins involved in regulation of the ensuing, potentially host-damaging, response. The number of molecular players or variables involved in any such activity can vary from hundreds to thousands, making immune responses immensely complex. This complexity is amplified by the multiscalar nature of the immune system, as these signaling and transcriptional responses occur in the context of diverse and dynamic cell–cell interactions. The different types of immune cells, from innate myeloid cells to lymphocytes, receive cues from different classes of receptors, produce distinct effector molecules, and depend on cues from each other as well as from the tissue microenvironment to respond appropriately. Pathogen invasion of host cells further exaggerates this complexity; while immune cells mount an inflammatory response to contain infection, pathogens attempt to evade immune recognition and modulate the host response by secreting a variety of virulence factors. The nature and magnitude of the host immune response is therefore dictated by extensive dynamic molecular and cellular interactions within and between host and pathogen, which can be represented by multiple types of networks.

Given the intricate interactions that underlie immune function, the use of networks as mathematical abstractions of complex relationships is not only suitable but also necessary for understanding immune responses. Here, we apply network concepts at different molecular and cellular scales to elucidate some of the unique aspects of the immune system that make it particularly suitable for network-based studies. In section Molecular Networks, we introduce different types of molecular networks and the resources through which they can be accessed. In section Intracellular Molecular Networks in Innate Immunity, we discuss how application of network analysis methods can provide insights into intracellular molecular networks underlying the innate immune response. In addition to intracellular interactions, crosstalk between immune cells as well as their spatial organization and migration within tissues in vivo are critical to the effective development of innate and adaptive immune responses. These intercellular interactions, which are functions of space and time, are represented as networks in section Intercellular and Organismic Networks. Another unique aspect of immunity, i.e., host–pathogen crosstalk, is represented as interaction between host and pathogen networks in section A Network Perspective of Host-Pathogen Interactions (HPIs). Each of these areas is large in and of itself and we cannot be comprehensive in this short review; our objective is to introduce general network biology concepts and elucidate how network analysis can be used to represent and interrogate different aspects of the immune system.

**MOLECULAR NETWORKS**

Network analysis has mostly been applied to biological systems to model macromolecules at the genome scale. In this section, we discuss the general concepts of these molecular networks, available resources, and their applications. Based on the types of macromolecules and the nature of their interactions, networks can be classified into different categories, such as metabolic, gene regulatory, protein–protein interaction (PPI), genetic interaction and signaling networks. These networks have been created using a combination of experimental and computational approaches. Technological breakthroughs and computer-aided automation have enabled high-throughput experiments and the efficient generation of large unbiased datasets that are particularly valuable for network construction. Molecular networks have also been created by manual curation of vast amounts of existing literature by domain experts. Although this process is painstakingly long and labor intensive, the resulting molecular networks...
are of good quality with low false-positive rates. An alternate approach is to apply text-mining algorithms to scientific articles and abstracts to predict molecular networks. As expected, this strategy may expedite the process of network construction at the cost of increased false-positive rates. A list and brief description of databases for gene regulatory, PPI, and signaling networks is provided in Table 1.

Gene regulatory networks (GRNs) represent control of gene expression or gene-specific RNA amounts by regulatory components such as transcription factors (TFs), activators, repressors, and microRNAs (miRNAs). Hence, each node in the GRN either corresponds to a target gene or a regulator of the gene or its RNA product. Several experimental and computational methods have been used for constructing GRNs. Two complementary experimental approaches are often used to generate data for these networks; Yeast one-hybrid assays can identify TFs that regulate a gene, while ChIP (chromatin immunoprecipitation)-chip and ChIP-seq experiments can predict potential target genes of a TF by mapping protein–DNA interactions. Several algorithms have been developed for computationally predicting GRNs, which are categorized and compared in two recent review articles. Recently, a collaborative and community-driven approach was used for predicting GRNs through the consortium DREAM (Dialogue on Reverse Engineering Assessment and Methods).

Many biological processes are dependent on physical interactions between individual proteins and spatial patterns of movement of numerous proteins within the cell. Thus, elucidating and interpreting these interactions is essential for understanding the resulting biological function. PPI networks attempt to capture large-scale interaction information, where each node represents a protein and edges correspond to physical interactions. High-throughput experimental techniques, such as yeast two-hybrid (Y2H) and affinity purification–mass spectrometry (AP-MS), are the most widely used methods for identifying PPIs. While Y2H can identify physical and binary interactions between a protein pair in vivo, AP-MS is suitable for identifying whole protein complexes with multiple interacting partners. Computational approaches use different types of information, such as evolutionary relationships, three-dimensional protein structure, and the presence of specific protein domains, to predict PPI networks. Network representation of protein–protein and gene regulatory interactions allows us to study the topological properties of these interactions. In Figure 1(a) and (b), we highlight the contrast between the architecture of PPI networks and GRNs.

One concern in the PPI network construction community is the lack of reproducibility among networks generated in different laboratories using different (and in some cases, nominally the same) experimental methods, implying a high rate of false-positive and/or false-negative interaction calls in the various studies. Some computational algorithms have attempted to address this issue by assigning confidence scores to each of the interactions, but others have focused on the lack of standard operating procedures among laboratories using a single method such as MS to study PPIs. Indeed, standardization of protocols between laboratories has markedly increased the agreement in PPI networks derived from such studies.

Signaling networks respond to inputs from extracellular and intracellular stimuli, process the inputs to distill signal from noise, and ultimately integrate the incoming information to ensure an appropriate functional response (Figure 1(c)). These networks show architectural similarity to electric circuits, as they have distinct components for sensing, processing/filtering, and transmission of signals. Unlike GRNs and PPI networks, the nature of the nodes and edges of signaling networks can be diverse. Nodes in a typical signaling network most often represent proteins and protein complexes, but can also represent genes, RNAs, chemical compounds, or any variable that can impact transmission of the signal from the starting point, often a receptor, to the end of the cascade, for example, a TF or gene. The edges of signaling networks can be physical binding, phosphorylation, ubiquitination, glycosylation, or transcriptional regulation, as in the case of signal-induced transcription of signaling components.

While global molecular networks represent all possible interactions and regulations that can be achieved by the components, overlaying high-throughput data obtained from experimental perturbations onto the network can identify regions of the network functionally active under a specific condition. By analogy, if a global network represents the comprehensive street map of a city, overlaying data can indicate regions of high traffic volume at a specific time of the day. Several algorithms that integrate high-throughput data with GRNs, PPI networks, and signaling networks have been developed. The performance of these algorithms depends on the choice of network resource, signal-to-noise levels in high-throughput data, sensitivity and specificity of the scoring metric, and statistical tests for estimating the significance.

Representing complex relationships using networks facilitates effective visualization, which in
<table>
<thead>
<tr>
<th>Name</th>
<th>Website</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transcription regulatory network</strong></td>
<td>TRANSFAC(^{12}) <a href="http://www.gene-regulation.com/pub/databases.html">http://www.gene-regulation.com/pub/databases.html</a></td>
<td>Most comprehensive database with public and professional (requires subscription) versions. The professional version contains information on (~20k) transcription factors (TFs), (~80k) target genes, (~40k) DNA sites, and (~300k) promoter sequences.</td>
</tr>
<tr>
<td>AnimalTFDB(^{13})</td>
<td><a href="http://www.bioguo.org/AnimalTFDB/">http://www.bioguo.org/AnimalTFDB/</a></td>
<td>A database of TFs classified into 72 families based on their DNA-binding domains predicted in 50 sequenced animal genomes. Also contains information on transcription cofactors and chromatin remodeling factors from these genomes.</td>
</tr>
<tr>
<td>TFdb(^{14})</td>
<td><a href="http://genome.gsc.riken.jp/TFdb/">http://genome.gsc.riken.jp/TFdb/</a></td>
<td>Enlists mouse TF genes, their related genes, and the associated gene ontology (GO) terms.</td>
</tr>
<tr>
<td>TFCat(^{15})</td>
<td><a href="http://www.tfcat.ca">http://www.tfcat.ca</a></td>
<td>Literature curated catalog of mouse and human TFs and sequence-based prediction of additional TFs.</td>
</tr>
<tr>
<td>ITFP(^{16})</td>
<td><a href="http://itfp.biosino.org/itfp/">http://itfp.biosino.org/itfp/</a></td>
<td>A platform for computational prediction and classification of TFs from protein sequence of human, mouse and rat. TF target genes are predicted from gene expression data using ARACNE algorithm.</td>
</tr>
<tr>
<td><strong>Interaction network</strong></td>
<td>STRING(^{18}) <a href="http://string-db.org">http://string-db.org</a></td>
<td>A database of known and predicted protein interactions obtained from diverse sources, such as experimental repositories, computational predictions, and public text collections. It contains around 2.5 million proteins from 630 organisms and includes physical and functional interactions.</td>
</tr>
<tr>
<td>BIND(^{19})</td>
<td><a href="http://bind.ca">http://bind.ca</a></td>
<td>Stores information on interacting partners of a protein, such as RNA, DNA, other proteins, molecular complexes, and small molecules. Molecular interactions were obtained from published literature.</td>
</tr>
<tr>
<td>HPRD(^{20})</td>
<td><a href="http://www.hprd.org">http://www.hprd.org</a></td>
<td>A database of manually curated high-confidence protein interactions in human. It also stores information on protein domain architecture, post-translational modification, and disease associations.</td>
</tr>
<tr>
<td>MIPS(^{21})</td>
<td><a href="http://mips.helmholtz-muenchen.de/proj/ppi/">http://mips.helmholtz-muenchen.de/proj/ppi/</a></td>
<td>A repository of manually curated protein–protein interactions (PPIs) in mammals collected from the scientific literature. Only reliable physical interactions are included in the database.</td>
</tr>
<tr>
<td>IntAct(^{22})</td>
<td><a href="http://www.ebi.ac.uk/intact/main.xhtml">http://www.ebi.ac.uk/intact/main.xhtml</a></td>
<td>Stores protein interaction data that are either curated from literature or from direct user submission. IntAct recently merged with MINT, which is another PPI database.</td>
</tr>
<tr>
<td><strong>Signaling pathway/network</strong></td>
<td>KEGG(^{23}) <a href="http://www.genome.jp/kegg/">http://www.genome.jp/kegg/</a></td>
<td>A large collection of signaling pathways for nearly 1500 species. The pathways can be browsed and data can be freely downloaded in KGML format.</td>
</tr>
<tr>
<td>Name</td>
<td>Website</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>REACTOME</td>
<td><a href="http://www.reactome.org">http://www.reactome.org</a></td>
<td>A manually curated and peer-reviewed database of biochemical reaction networks in human. For several species, such as mouse, fly, worm, plant, and bacteria, the reaction network is inferred through orthologous events. The network can be visualized through web browsers and downloaded in multiple formats.</td>
</tr>
<tr>
<td>WikiPathways</td>
<td><a href="http://wikipathways.org/index.php/WikiPathways">http://wikipathways.org/index.php/WikiPathways</a></td>
<td>A freely available resource for visualizing and downloading pathways for ~25 species. Unlike other databases, the WikiPathways resource relies on participation by the scientific community. Pathways can be created and edited by users.</td>
</tr>
<tr>
<td>InnateDB</td>
<td><a href="http://www.innatedb.com">http://www.innatedb.com</a></td>
<td>A well-curated and widely used resource of the genes and proteins involved in mammalian innate immunity and their interaction network. The web interface allows different systems biology analyses, such as enrichment of pathways and GO terms, network, and TF-binding site analyses.</td>
</tr>
<tr>
<td>PathogenPortal</td>
<td><a href="http://pathogenportal.org">http://pathogenportal.org</a></td>
<td>This website stores, updates, and integrates genomic, transcriptomic, metagenomic, and proteomic data for diverse hosts, pathogens, and vectors. The interactions between hosts, pathogens, and vectors can be visualized as networks from the user-friendly interface.</td>
</tr>
<tr>
<td>Systems Virology Center</td>
<td><a href="https://www.systemsvirology.org">https://www.systemsvirology.org</a></td>
<td>Provides analysis tools and multiple ‘omics’ data to obtain a better understanding of how viruses interact with and regulate host networks.</td>
</tr>
<tr>
<td>Systems Influenza</td>
<td><a href="http://www.systemsinfluenza.org">http://www.systemsinfluenza.org</a></td>
<td>A knowledgebase of transcriptomics, proteomics, and lipidomics data for studying interaction of influenza virus with the host immune system.</td>
</tr>
<tr>
<td>ImmGen</td>
<td><a href="http://www.immgen.org">http://www.immgen.org</a></td>
<td>A collaborative project that generated gene expression data under carefully standardized conditions surveying all cell types of lymphoid and myeloid lineage in mice. The data, metadata, genomic modules, and networks can be accessed through web browsers.</td>
</tr>
</tbody>
</table>
FIGURE 1 | Architecture of human gene regulatory and protein–protein interaction networks. (a) Human transcription factor regulatory network (TRN). The human TRN was downloaded from Ref 17. The network is displayed using the Cytoscape force-directed layout. Each node in the network is a transcription factor (TF), and edges represent transcriptional regulation. As transcriptional regulation is directed, network edges are directional from a TF to its target. The network has 3107 nodes and 6887 edges. The network clearly shows a hierarchical architecture as observed in Ref 17. Only a small subset of TFs regulate most of the other TFs, which is obvious from the modularity of the network architecture. The network can be divided into regions that are either autocratic or democratic. In the autocratic regions, a TF is usually regulated by a single TF whereas in the democratic regions a TF is regulated by multiple TFs. Circles highlight some of the autocratic regions. (b) Human protein–protein interaction (PPI) network. The network was downloaded from the Human Protein Reference Database (HPRD) and displayed using the Cytoscape force-directed layout. Each node in this network is a protein. The edges of the network represent protein–protein associations observed in the literature and manually curated in the database. As PPIs do not have directionality, the network edges are nondirectional. The PPI network has 9251 nodes and 38,869 edges, hence it is much bigger compared with the gene regulatory network (GRN). Unlike the GRN, the PPI network shows lack of hierarchy. (c) A schematic representation of a typical signaling network. Ligands are sensed by specific membrane-bound receptors (R1 and R2) followed by signal transduction, often mediated by adapter proteins that associate with the effector domains of receptors. The complex signaling circuitry further propagates and processes the signal through multiple steps, including signal integration, amplification (for instance by phosphorylation (P) or dephosphorylation of specific mediators), and noise reduction. Finally, the actuator, usually a TF, directs expression of appropriate target genes based on the processed signal. Different cellular components and their analogs in electric circuits are indicated on the left and right, respectively.

turn aids in easy recognition of patterns and emergent properties arising from these interactions. In network visualization, nodes can be laid out in different ways, e.g., random, circular, hierarchical, or spring embedded. Different properties of the nodes and edges can also be represented by their color, size, shape, and border. Table 2 enlists some of the visualization tools widely used for rendering molecular networks. Among these tools, Cytoscape is one of the oldest and the most widely used. Some of the reasons behind this popularity are its user-friendly interface and its open-source platform, which allow any developer to create plug-in software.61 Most of the plug-ins not only help in network rendition but also allow different network-based analyses within the interactive visualization framework. Similarly, several visualization and analysis tools are also available through R Bioconductor, which is the biggest repository of bioinformatics tools. The igraph package provides a large number of sophisticated analysis tools along with network visualization tools, and is used by diverse research communities.
TABLE 2 | Visualization of Biological Networks (Following Are Some of the Major Tools for Visualization of Networks and Pathways)

<table>
<thead>
<tr>
<th>Tool</th>
<th>Plugin/Packag (Total Downloads)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytoscape</td>
<td>ClueGO (13822) ^52</td>
<td>Identifies enriched gene ontology (GO) and pathway terms from a list of genes and represents interrelations of enriched terms as a network.</td>
</tr>
<tr>
<td></td>
<td>BiNGO (9975) ^53</td>
<td>Determines which GO categories are statistically overrepresented in a set of genes or a subgraph of a biological network.</td>
</tr>
<tr>
<td></td>
<td>CluePedia (6801) ^55</td>
<td>Extends the ClueGO plug-in by rendering a network of genes corresponding to each enriched GO and pathway term using data and existing knowledge.</td>
</tr>
<tr>
<td></td>
<td>MCODE (6425) ^56</td>
<td>Finds densely connected regions or clusters in a biological network. The biological interpretation of the clusters depends on the type of network.</td>
</tr>
<tr>
<td></td>
<td>jActiveModules (5728) ^48</td>
<td>Overlays gene expression data on biological networks to identify expression-activated subnetworks or network hotspots.</td>
</tr>
<tr>
<td>R Bioconductor</td>
<td>Rgraphviz (63259)</td>
<td>Renders R graph objects and provides multiple options for layout, node, and edge properties. It is used by several Bioconductor packages for network rendering.</td>
</tr>
<tr>
<td></td>
<td>Pathview (7726) ^57</td>
<td>Overlays gene expression data onto KEGG canonical pathway map.</td>
</tr>
<tr>
<td></td>
<td>BioNet (5309) ^18</td>
<td>Scores each node by differential expression and identifies significantly differentially expressed subgraphs from large biological networks. Subgraphs are rendered using 2D and 3D visualization.</td>
</tr>
<tr>
<td></td>
<td>RedeR (3851) ^59</td>
<td>A package that combines R-based computational analysis with Java-based visualization for dynamic network visualization and manipulation.</td>
</tr>
<tr>
<td></td>
<td>cisPath (2722)</td>
<td>A package for management, visualization, and editing of PPI networks. cisPath creates HTML files, which can be visualized using standard web browsers.</td>
</tr>
<tr>
<td></td>
<td>igraph</td>
<td>A general-purpose graph package in R and Python language. Provides hundreds of functions for creating, manipulating, optimizing, and rendering graphs.</td>
</tr>
<tr>
<td>Web-tools</td>
<td>STRING ^18</td>
<td>The web interface can render a network of query gene(s) and their direct neighbors from the underlying network database, which is created by combining different sources, e.g., co-expression, text mining, and co-occurrence.</td>
</tr>
<tr>
<td></td>
<td>Pathway Commons ^60</td>
<td>Allows the user to search, visualize, and download network neighbors of query genes from publicly available pathway databases.</td>
</tr>
</tbody>
</table>

Unlike Cytoscape, use of R Bioconductor or igraph tools requires programming experience. STRING ^18 and Pathway Commons ^60 are examples of web-based resources for visualizing smaller networks, typically interacting partners of queried gene(s).

Deriving conclusions from computational and visual analyses of any large-scale network is a formidable challenge. Some researchers have focused on studying topological properties of the nodes and finding their biological relevance. For example, Barabasi and Albert ^62 argued that genetic and signaling networks are scale-free (i.e., network degree distribution follows a power law) and the networks were developed through preferential attachment during evolution. In scale-free networks, "hub" nodes (i.e., nodes with large numbers of interacting partners) play the most important roles and the network will cease functioning if these nodes are deleted. In a complementary study, Yu et al. ^63 observed that it is not the hub nodes but the nodes with high betweenness centrality (i.e., network nodes that have many ‘shortest paths’ going through them) that are essential in a protein network. Specific experimentally testable hypotheses are often derived from a network by combining its topological properties with biological information. In the next section, we discuss specific examples where network analysis has led up to experimentally testable hypotheses.

INTRACELLULAR MOLECULAR NETWORKS IN INNATE IMMUNITY

The functions of the immune system, such as protective host defense under normal conditions or
autoimmune disease when dysregulated, are encoded in interactions between its constituents operating in underlying complex dynamic molecular and cellular networks. These include GRNs, PPI networks, and signaling and metabolic networks introduced in section Molecular Networks. In this section, we discuss these networks in the context of the innate immune system, which detects microbial infection or tissue damage by employing a repertoire of pattern recognition receptors (PRRs), which are present on the surface of the cell or in the cytoplasm. These include the Toll-like receptors (TLRs), nucleotide binding and oligomerization domain-like receptors (NLRs), C-type lectin receptors (CLRs), the retinoic acid-inducible gene I-like receptors (RLRs), as well as several putative sensors of intracellular DNA.

The processes that enable an effective innate response can be broadly grouped into four categories, namely transcriptional, translational, spatial, and functional networks (Figure 2), which are higher level processes encompassing the GRNs, PPI networks, and signaling networks discussed previously. Each node in these diverse networks is a complex biological entity (e.g., innate sensors such as TLRs, NLRs, RLRs, and CLRs) or a biological process (e.g., transcription, translation, subcellular location, signaling, and heterogeneity) that is itself comprised of underlying subnetworks (Figure 2). That is, each node in the functional cellular network can be recast as a more complex subgraph that explodes the higher level node into its constituent parts with new nodes and edges. For instance, the process of ‘transcription’ in turn comprises of the nodes ‘cofactor, TF, promoter/enhancer’ and edges depicting connections between these constituent nodes (Figure 2). The edges in the higher order network represent functional connections in the context of an innate insult such as pathogen invasion or exposure to endogenous signals arising from tissue damage.

To perform network analysis of a complex system such as the innate immune system, information must thus be captured and integrated from many hierarchical levels. These levels include genomic information (DNA sequences and associated polymorphisms, epigenetic regulation), transcriptional information (RNA sequences and abundances, alternate splice forms, noncoding RNAs such as Inc RNAs and miRNAs), translational information (protein abundances, RNA-binding regulatory proteins, post-translational modifications), spatial information (protein localization to subcellular structures such as cell membranes, cytosol mitochondria, endosomes, or peroxisomes), cellular signaling processes (protein–protein and protein–DNA interactions, organization of proteins, lipids, and other molecular entities into pathways), cellular heterogeneity (information on functionally distinct cell subtypes or stochastic differences between single cells of the same subtype), cell–cell interactions, and organization of cells into tissues, organs, organ systems, and ultimately the resulting phenotypes (Figure 2). Recent technological advances have enabled us to query these many levels through acquisition of large-scale ‘omics’ datasets that have facilitated network biology research in the context of innate immune molecular mechanisms. Genomics (sequencing of whole genomes) is complemented by transcriptomics (characterization and quantification of mRNA species), proteomics (characterization and quantification of proteins and associated modifications), and more recently with metabolomics and lipidomics (comprehensive characterization of abundance of lipids and metabolites). The wiring of networks representing the data gathered in this manner can in turn be expected to vary depending upon the nature of the immune cell [e.g., macrophage, dendritic cell (DC), neutrophil, T cell, or B cell] as well as between single cells of seemingly homogeneous immune cell populations. Below, we discuss several prominent examples in which network analysis methods have been applied to varying types of experimental datasets to provide new insights into the innate immune response.

Transcriptional/Gene Regulatory Networks

Most studies utilizing network biology approaches to elucidate innate immune responses have focused on deciphering transcriptional networks and GRNs controlled by TLRs, very recently extending to include other innate immune sensors. This is because many functional responses to TLR ligands can be readily recapitulated in vitro using bone marrow-derived macrophages and DCs. Early studies utilized microarray-based transcriptome analysis over multiple time points and clustering of genes with similar expression kinetics. In one such example, gene clusters from TLR4-stimulated macrophages were mined for both induced TFs and for cis-regulatory elements to identify a prominent group of genes regulated by activating TF 3 (ATF3). Network analysis and visualization predicted ATF3 as part of a transcriptional complex that also contains members of the nuclear factor (NF)-κB family of TFs. Probable ATF3-associated TFs were selected and potential transcriptional targets of the ATF3 regulatory complex were identified based on enrichment of ATF3, NF-κB, and associated TF-binding sites within close proximity. Using this method, 30 target genes
were identified, many of which were validated using a ChIP-to-chip approach whereby ATF3-bound DNA was assessed for enrichment of selected genes.

Scanning for TF-binding sequence motifs and gene expression dynamics have since been used to elucidate macrophage transcriptional networks on a larger scale. In one study, time-lagged expression correlation between TF genes and clusters of potential target genes was used to infer a network of associations in response to multiple TLR agonists. As an independent measure of association, the promoters of differentially expressed genes were scanned for the presence of TF-binding motifs. A combined P-value based on both the time-lag correlation and motif scanning was determined and TF–gene cluster pairs were ordered based on this combined P-value to construct...
a regulatory network that was further expanded through the inclusion of TF interaction data from publicly available sources. The inferred network of associations between TF genes and co-expressed gene clusters was validated with targeted ChIP-on-chip experiments, and yielded insights into the macrophage activation program. One of the major findings was the identification of a novel regulator, TGFβ-induced factor homeobox 1 (Tgif1), which was only detected using the integration of both expression data and promoter scanning evidence. In another example, a combination of gene expression analysis and TF-binding site motif scanning algorithms was used to infer a network of associations between TFs and target genes in macrophages activated with dsRNA. This network of associations predicted that a member of the forkhead family of TFs, FOXO3, acts as a negative regulator of the type 1 interferon (IFN) response, a linkage that was confirmed experimentally using FOXO3-null macrophages. Genome-wide ChIP-seq analysis identified the IRF7 gene as one of the critical targets of FOXO3; this observation was functionally validated by gene deletion studies that in turn identified FOXO3 as a negative regulator of IRF7 transcription. FOXO3, IRF7, and type 1 IFN were further demonstrated to form a coherent feed-forward regulatory circuit that was experimentally validated by ChIP analysis, in vitro functional assays, and in vivo during vesicular stomatitis virus lung infection of FOXO3 KO and IRF7 KO mice. Together, these findings established the FOXO3–IRF7 regulatory loop as a critical circuit that optimizes the antiviral response to achieve the appropriate balance between host defense and unchecked, potentially host-detrimental inflammation.

The study of transcriptional networks has also yielded insights into human responses to PRR ligands (pathogen-associated molecular patterns, PAMPs) in vivo. In one study, microarray data were generated from peripheral blood leukocytes isolated from human subjects at various times after injection of bacterial endotoxin. Using Ingenuity Systems Inc. knowledgebase and network analysis tool, the authors constructed a prototypical interaction network in blood leukocytes containing 292 genes and their direct interactions, and a global endotoxin-induced gene interaction network of 1556 genes. This approach was used to identify changes in functional modules at different times during an endotoxin-induced immune response. Using the same human microarray dataset, a more recent study used network component analysis to derive a dynamic regulatory network including 10 TFs, 99 target genes, and 149 regulatory interactions. By incorporating regulatory network information, this method facilitated prediction of TF activities over time and allowed for the identification of novel co-regulated gene clusters and target genes within these clusters.

The integration of genome-wide mRNA expression data with network perturbation using methods such as RNAi knockdown has further facilitated construction of biologically relevant innate immune transcriptional networks. One particularly extensive study employed a combination of gene expression and RNAi perturbation to reconstruct a transcriptional network controlling responses of DCs to distinct PAMPs. Specifically, gene expression profiles obtained at different time points poststimulation with PAMPs were used to identify candidate regulators that were then perturbed using lentiviral shRNA knockdown to determine how these regulators affect the target genes. This overall approach identified 2322 significant regulatory connections and revealed the functions of 125 TFs, chromatin modifiers, and RNA-binding proteins. These data enabled the construction of a network model consisting of 24 hub regulators and 76 fine-tuners that help to explain how pathogen-sensing pathways achieve specificity. Non-negative matrix factorization was used to define two major states induced by the PAMPs, namely ‘inflammatory’ versus ‘antiviral’ programs as well as distinct subnetworks for each state. Within each subnetwork, the authors found the predominance of multiple feed-forward circuits. The network reconstruction strategy utilized in this study has been reviewed by the authors. Recently, the same group has developed a high-throughput ChIP method to reconstruct the genome-wide dynamic binding network of 25 TFs and 4 chromatin marks at four time points following LPS stimulation of DCs. Analysis of over 180,000 TF–DNA interactions revealed a temporally multilayered architecture of TF networks. Coupling these two approaches, i.e., unbiased perturbation of TFs with genome-wide binding of the same factors, will be a robust approach for systematically elucidating the GRNs controlling the innate immune response.

In addition to construction of GRNs underlying the innate immune response, integrative systems approaches based on genome-wide RNAi screens have been used to identify useful interventions in infection. One especially exhaustive study employed genome-wide RNAi screening to identify host factors required for influenza virus replication as potential targets for therapeutic intervention. An arrayed siRNA library targeting more than 19,000 human genes was used to transfect human lung epithelial cells. Data from two independent screens analyzed using an integrative analysis approach, including redundant siRNA activity, interactome, and ontology-based
analysis, identified 295 host cell factors required for early-stage influenza virus replication. Protein interaction data derived from Y2H databases, including published datasets, were used to construct a host–pathogen interaction (HPI) map depicting associations between the identified host factors, viral-encoded proteins, and other cellular proteins. The integration of RNAi and interactome datasets produced a network containing 181 confirmed host cellular factors that mediate 4266 interactions between viral or cellular proteins, and act as rate-limiting ‘hubs’ in cellular pathways or processes required for influenza virus replication. A number of candidate genes involved in influenza virus replication were biologically validated by experiments showing that depletion of these genes resulted in reduced viral gene expression. Further functional assays were performed to identify factors specifically involved in virus entry or postentry viral replication. Finally, small-molecule inhibitors of several host factors including vATPase (vacuolar ATPase) and CAMK2B (a ubiquitously expressed calcium sensor that regulates diverse cellular functions) were shown to antagonize influenza virus replication providing new opportunities for the development of host factor-directed antiviral therapies.

Considerable interest has also focused on the analysis of GRNs that direct cellular diversity and cell fate decisions within the innate immune system. A combination of ChIP-on-chip data and analysis of TF-binding sites has been used to demonstrate a regulatory network involving the TFs PU.1 and Gfi1 that orchestrate innate (macrophage) and adaptive (B lymphocyte) immune cell fates in multipotential progenitors. This network was experimentally validated in vitro and in vivo by following B-cell development in Gfi1 knockout and PU.1 heterozygous animals. The same group has previously assembled and mathematically modeled a GRN that controls macrophage versus neutrophil cell fate choice and that accounts for the onset and resolution of mixed lineage patterns during cell fate determination. Another study has used a network perturbation approach to investigate causal influences or crosstalk between signaling modules involved in the cytoskeletal response of neutrophils to chemoattractant. More recently, network analyses have been employed to reveal a spectrum of macrophage activation states extending the current M1- versus M2-polarization model. The authors generated a dataset of 299 macrophage transcriptomes by stimulating macrophages with diverse activation signals including PRR ligands, cytokines, and metabolic cues. Network analyses revealed common transcriptional regulators associated with all types of macrophage activation states as well as regulators related to stimulus-specific programs. Finally, the Immunological Genome Project consortium database has been utilized to elucidate transcriptional regulatory networks that control lineage commitment and functional diversification of DCs (DC progenitor cells, lymphoid-tissue, nonlymphoid-tissue DCs, and migratory DCs) across the murine immune system (Table 1). Thus, network analysis of transcriptional and ChIP data has also aided in identification of GRNs underlying distinct immune cell phenotypes and cell states in homeostatic and activated conditions.

Genome-wide mRNA expression profiling followed by network analysis has also been applied to study cell fate determination within the adaptive immune system. One such study focused on elucidating the molecular mechanisms by which naive T cells develop into effector Th17 cells. Network analysis of transcription profiles from wild-type and IL-23R KO T cells following in vitro exposure to Th17 polarizing conditions (TGF-β1 and IL-6 with or without IL-23) identified serum glucocorticoid kinase 1 (SGK1), an inducible salt-sensing kinase, as an essential node downstream of IL-23 signaling. A curated database of PPIs was used to construct a network model that connects known proteins of the IL-23R signaling pathway to the TFs whose function is dysregulated in IL-23R KO cells. The network’s nodes were ranked based on a centrality measure, i.e., the fraction of IL-23R-affected TFs downstream of that node in the network. SGK1 emerged as the highest ranking node suggesting that it acts both as a transcriptional target of IL-23R signaling and as a kinase that may mediate the transcriptional effects of the pathway. These predictions were tested by examining the differentiation and maintenance of Th17 cells in SGK1 KO mice, as well as in mice lacking SGK1 specifically in IL-17-producing CD4+ T cells. The results showed that SGK1 was indispensable for IL-23R-dependent stability and maintenance of Th17 cells. Furthermore, mice fed a high salt diet (SGK1 is salt-inducible) showed a marked increase in the frequency of Th17 cells in the gut and developed a more severe form of experimental autoimmune encephalitis than mice fed a normal diet. This increased severity was dramatically reduced in SGK1-deficient mice, confirming the role of SGK1 in induction of pathogenic Th17 cells.

Another emerging area of focus is the elucidation of GRNs established during cellular reprogramming to derive clinically relevant cells for regenerative medicine. A network biology platform called CellNet...
The authors analyzed expression data from 56 published reports and using CellNet to compare the GRNs of in vitro engineered cells with those of their target cell types in vivo, and they found that cells derived via directed differentiation from embryonic stem cells more closely resemble their in vivo counterparts than products of direct conversion by ectopic expression of lineage-specific TFs. They further discovered that directly converted cells fail to adequately silence expression programs of the starting population and that the establishment of unintended GRNs is common to virtually every cellular engineering paradigm. CellNet was used to spot aberrant regulatory networks and predict candidate transcriptional regulators that could be targeted for improving cell derivations. Experimental testing of these predictions showed that as predicted, improved B-cell to macrophage conversion, both transcriptionally and functionally, was observed by knocking down the B-cell regulators Pou2af1 and Ebf1. In addition, CellNet revealed an unexpected intestinal program controlled by the master regulator Cdx2 during conversion of fibroblasts into induced hepatocytes (iHeps), and experimental validation of this prediction showed long-term functional engraftment of mouse colon by iHeps, thereby establishing their broader potential as endoderm progenitors and demonstrating direct conversion of fibroblasts into intestinal epithelium.

**Signaling Networks**

While the accessibility of gene expression datasets generated from PAMP-stimulated cells has contributed to the prevalence of large-scale transcriptional networks, elucidation of signaling networks demands incorporation of protein expression and PPI data into network analysis methods. This has been accomplished by advances in MS, which have allowed for acquisition of extensive proteomic datasets. In one such study, the authors systematically explored the signal transduction pathways responsible for regulating cellular antiviral defense and IFN production by using a comprehensive proteomic approach to assess the human innate immunity interactome controlling production of type 1 IFN. In a pull-down approach, 58 genes with known involvement in the regulation of type 1 IFN were used as baits for interacting proteins; a total of 264 complexes were purified and analyzed by MS. Using these data, the authors mapped a network of 260 interacting proteins and 401 unique interactions. Many of the interactions not present in existing BioGRID or STRING databases were independently validated by immunoprecipitation and Cytoscape was used to visualize the interconnectivity of the network and combine the interactions into a single map. Functional consequences of these interactions on antiviral responses were assessed using overexpression and/or RNAi analyses. Dynamic changes in this interaction network were observed upon stimulation with several distinct PAMPs, revealing nodes of potential receptor crosstalk.

Integrative systems approaches combining genomic, proteomic, and network perturbation methods have yielded further insights into complex signaling circuits. A study combining transcriptional profiling, lentiviral shRNA-mediated knockdown, small-molecule perturbations, and SILAC-based quantitative phosphoproteomics, uncovered 35 signaling regulators, including 16 known regulators, involved in the TLR4 response of mouse DCs. In particular, they identified CRKL, a tyrosine kinase adaptor, as an important modulator, and polo-like kinases 2 and 4 as critical activators of the antiviral response. More recently, the same group has used an integrative approach to identify genes involved in the cytosolic DNA-sensing network. They selected 809 candidates after transcriptomic and quantitative proteomic analyses (protein–protein and protein–DNA interactions) and perturbed the candidates using high-throughput RNAi and small-molecule inhibition to functionally validate the network. These networks, generated from integrated datasets, provide a major resource for immunologists to mine and identify novel players in signaling pathways and can aid in the construction of comprehensive computational models to generate predictions of the innate response.

**Spatial Networks**

Activation of signal transduction networks by extracellular or intracellular PRRs is encoded not only into transcriptional networks, GRNs, and PPI networks but also into spatial patterns of activation and relocation of numerous proteins. Many proteins are targeted with exquisite precision to specific locations in the cell, or undergo rapid directed changes in localization that influence the outcome of or are essential to signal generation. In many cases, innate sensors and/or adaptors must relocalize to specific cellular compartments or intracellular membranes for optimal assembly of signaling complexes and effector function. For instance, TLRs signal from the plasma membrane or endosomes; RIG-I, MDA-5, and NLRP3
relocalize mitochondria upon activation via association with MAVS\textsuperscript{91–94}; and NOD1 and NOD2 signal from endosomes.\textsuperscript{95–97} The nature of subcellular association can also influence the quality and duration of the response. This has been observed in the context of viral infection, where peroxisomal MAVS induces rapid IFN-independent expression of defense factors that provide short-term protection, whereas mitochondrial MAVS activates an IFN-dependent signaling pathway with delayed kinetics, which amplifies and stabilizes the antiviral response.\textsuperscript{98} Thus, there is increasing evidence that organelar structures and intracellular movement are integral to the development of an effective immune response. Integration of this additional dimension into network analysis and visualization methods through development of appropriate tools should allow for a more holistic representation of innate networks as well as prediction of emergent properties that arise from the concerted action of these networks.

**Functional Networks**

Most studies discussed thus far have assessed responses to single, purified PRR ligands; however, this is a scenario far from biological reality. During a natural infection, pathogens present a cocktail of ligands that activate diverse cell surface-associated and intracellular PRRs to initiate combinatorial signaling cascades\textsuperscript{99} (Figure 2). Crosstalk between these signaling cascades gives rise to multifaceted, pliable, and functional networks in which dynamic interactions between the underlying components dictate the quality, magnitude, and duration of the ensuing host response. In this context, each ‘signaling pathway’ linked to a PRR-innate sensor is in reality a much more extensive multicomponent network comprised of several nodes representing constituent signaling molecules and edges representing the spatial, temporal, and functional connections between these constituents. The existence of such a multifactorial network composed of multiple innate sensors with some degree of functional redundancy may enable macrophages, DCs, and other immune cells to detect pathogens or danger with greater precision resulting in robust synergistic responses. For instance, synergistic inflammatory responses have been observed upon simultaneous stimulation of myeloid cells with different TLR ligands,\textsuperscript{100} TLR and NLR ligands,\textsuperscript{101} and TLR and CLR ligands.\textsuperscript{102} Crosstalk can also occur between PRRs and other receptors expressed by immune cells. For instance, dual B-cell receptor and TLR and/or NLR engagement can fine-tune functional B-cell responses.\textsuperscript{103,104} Similarly, crosstalk between PRRs and phagocytic receptors is indicated by the observation that phagosomal TLR agonists enhance major histocompatibility complex (MHC) class II–mediated antigen presentation.\textsuperscript{105} Additionally, synergy between PRRs and cytokine receptors can be inferred from the observation that IFN-\(\gamma\) augments TLR-induced gene transcription.\textsuperscript{106} Finally, transcriptional induction of some PRRs by microbial signals allows for sequential triggering of PRRs that may permit greater accuracy in mounting a response commensurate to the nature and extent of the threat while preventing excessive inflammation. For instance, triggering of TLRs is critical for priming NLRP3 expression and subsequent inflammasome activation.\textsuperscript{107} Thus, functional innate responses result from crosstalk within and between PRR, cytokine, phagocytic, and other pathways upon recognition of multiple antigens simultaneously presented by a microbe. Assessment of responses to combined rather than single stimuli is therefore an essential parameter that should be incorporated in the study of these complex signaling networks engaged by pathogens.

**INTERCELLULAR AND ORGANISMAL NETWORKS**

The existence of several different types of immune cells and their diverse cellular states lends additional levels of complexity that make holistic analysis of the immune system a challenging task. New advances in single-cell technologies such as single-cell microfluidics,\textsuperscript{108,109} transcriptomics,\textsuperscript{110} quantitative proteomics,\textsuperscript{111} multiparameter flow cytometry, mass cytometry,\textsuperscript{112} and live cell imaging\textsuperscript{113} are allowing researchers to understand how immune populations may generate and leverage cellular heterogeneity to achieve the range and plasticity required during dynamic processes such as cellular differentiation or host defense.\textsuperscript{66,114} One study combined molecular profiling data and database mining approaches to elucidate intercellular communication networks that dictate blood stem cell fate decisions and showed that intracellular and intercellular networks, and their interactions, are dynamic as differentiation occurs over time.\textsuperscript{115} Therefore, the nominal definition of a cell fails to convey the dynamic nature of cellular states, as intercellular communications change intracellular networks that define cellular behavior.

Given the dynamic nature of cell–cell interactions, there has been a growing interest in representing and modeling complex intercellular interactions in multicellular systems as networks with the ultimate goal of understanding and predicting their global behavior in a systematic and scalable way.\textsuperscript{116} Several
large-scale analyses to describe the structure and dynamics of intercellular networks have been performed, in the context of the immune system. One such study probed the cytokine-mediated connections between cells of the immune system (immune cells), parenchymal and stromal cells within tissues (body cells), and between ‘immune’ and ‘body’ cells. This cytokine network was constructed using available databases and revealed a tiered connectivity between immune cells, whereas body cells lacked such a connectivity pattern. The structural motifs of the immune cytokine network resembled those present in the social and World Wide Web superfamilies owing to the presence of clique motifs. The apparently greater degree of interconnectivity among immune cells, relative to other nonimmunological physiological systems as evidenced by this particular study, highlights the possible particular importance of visualizing immune responses as networks at the cellular level. When intercellular interactions are investigated in vivo, in the context of biological systems, the three-dimensional positioning of cells within tissues must be also considered, just as intracellular physical positioning of network components is crucial to efficient signaling, as described above. In a recent study of intercellular metabolic networks in the human brain, consideration of the tissue microenvironment in the model generation process added an additional layer of information that rendered networks more representative of the in vivo experience.

As a result of extensive studies involving sophisticated static and live imaging techniques combined with more traditional immunological assays, the murine lymph node (LN) is an optimal example of a tissue environment in which the many dimensions of immunologically relevant cellular interactions have been studied in great detail. Here, we provide an illustrative example of an innate immune response network in the LN. The lymphatic system is a network of open-ended structures that collect interstitial fluid that even under normal conditions extravasates into tissues. Interspersed along lymphatic vessels are LNs. These complex structures are classically considered the site where adaptive immune responses are initiated. However, it has recently been shown that a multicellular innate immune network is at play in the LN protecting the host from systemic dissemination of lymph-borne organisms. In Figure 3 we graphically display this complex biological system in parallel with a network rendition of the intercellular connections. Under steady-state conditions (Figure 3(a)) a heterogeneous population of innate immune cells is located in close proximity to the macrophages that line the subcapsular sinus, where lymph drains. The physical proximity of the innate cells and macrophages allows for efficient intercellular communication via secreted cytokines following a perturbation (i.e., infection with an intracellular or extracellular bacterium, e.g., as seen in Figure 3(b) and (c)). In the network rendition, we emphasize the spatial positioning requirements for effective intercellular interactions by only highlighting those nodes that can display a potential connection (dotted line) with the central node and graying out the others. Figure 3(b) and (c) represents the system after infection with an intracellular or extracellular organism, respectively. The sequence and timing of events is indicated by the gradual color progression from light gray (earlier events) to black (later events) and the thickness of the edges indicates the relative contribution of a particular pathway. For example, following infection with an extracellular bacterium (depicted in Figure 3(c)), the main pathway that is activated involves IL-1β release from the subcapsular sinus macrophages, which in turn recruits neutrophils to the LN, culminating in containment of the infection.

Cellular migration between tissues is the next level in further defining dynamic networks as they apply to complex biological communities. Upon immune challenge in the periphery, antigen can drain freely or be actively transported into the draining LN via afferent lymphatics. Here, antigen processing and presentation by professional antigen-presenting cells (APCs) takes place. T lymphocytes constitutively circulate through lymphoid organs, entering LNs through specialized venules and then migrating within the LN in search for cognate antigen. This migratory pattern is directed by encounters with APCs and by the stromal network in the LN. Upon cognate antigen recognition, as during an immune challenge, T-cell activation and proliferation occur. T cells exit the LN via lymphatics and enter the systemic circulation. Effector T cells then have the ability to home to sites of infection. It has been suggested that T cells receive instructive signals that direct them preferentially to certain sites of inflammation during the initial phase of priming. Once in the tissue, activated T cells carry out their effector function, in part through the secretion of cytokines.

Visualization of a network that is dynamic in four dimensions poses several challenges. In Figure 4(b), we depict a complex network of temporal and spatial variations in organization representing the biological system in Figure 4(a). Each node in this complex environment is an active and dynamic multifaceted network on its own. For example, the naive T-cell node must contain information about the
FIGURE 3 | A multicellular innate immune network in lymph nodes. An innate immune response circuit in the lymph node (LN) shown as a biological system (left panels) and a network (right panels). Left panels: (a) Innate effector cells are prepositioned in proximity to subcapsular sinus macrophages under steady-state conditions. (b) Upon exposure to intracellular bacteria draining to the LN through the lymphatic system, macrophages are activated, release cytokines, and engage innate effector cells. Their activation, in turn, leads to cytokine production that enables the macrophages to contain the infection. (c) Exposure to extracellular organisms also leads to macrophage activation. Production of IL-1β leads indirectly to neutrophil recruitment from the circulation, which leads to containment of the infection. Right panels: Network rendition of the biological system. Dotted edges in (a) represent potential connections. The color progressions of edges from light gray to black in (b) and (c) indicate the temporal progression of events. Thickness of the edges indicates the relative contribution of a particular connection.

migratory behavior of the cells, which in itself is a set of intracellular molecular events that allow the cell to interact with the extracellular matrix.\textsuperscript{119,123} T-cell migratory behavior influences the likelihood of antigen encounter in the LN.\textsuperscript{125,126} Therefore, there must also be information on intercellular interactions between T cells and APCs in the LN, as this interaction allows for the activation of T cells in the presence of cognate antigen. Additionally, each edge in this network represents an intricate set of molecular events, such as a change in the state of a cell (edge between naïve T cell and activated T cell) or in migration of a cell to distant organs (edge connecting blood to skin). We use different colors to indicate behavior under steady-state conditions (red arrows in Figure 4(b)) and make use of the color progression to indicate temporally sequential events (light gray to black). With the help of this example, we delineate the response of the host immune system as an interaction network, representing dynamic, spatial, and contextual interactions between diverse immune cells. As we learn more about the link between structure and function and the temporal evolution of immune responses, one must develop the tools to integrate these additional dimensions of information into network renditions or predictions of complex biological systems. A final layer of complexity can be added when one seeks to represent the...
A NETWORK PERSPECTIVE OF HPIs

Despite significant progress in medicine, infectious diseases such as HIV/AIDS, tuberculosis, malaria, and hemorrhagic fever, continue to remain major killers worldwide. Several pathogens can modulate host defense mechanisms to their advantage and actively manipulate the host immune response by producing a variety of virulence factors, thereby influencing the outcome of infection.127–129 Similarly, host factors can influence the pathogen; for instance, host-derived lysophospholipids can enhance secretion of a key virulence factor flagellin from Salmonella, and IFN-γ stimulates the transcription of Pseudomonas aeruginosa virulence factors.131 A deeper understanding of the underlying pathophysiology and molecular interactions between the host and pathogen is therefore central to overcoming the challenges posed by infections. From a network biology perspective, HPIs can be conceptualized as crosstalk between two molecular networks: one corresponding to the host and the other to the pathogen. These molecular networks have specific points of interaction and quite often pathogen proteins mimic eukaryotic PPI domains to invade the host PPI network. For example, the nonstructural protein 1 (NS1) of influenza A virus contains domains that bind to SH2 and SH3 domains of the regulatory

interaction between the networks of a multiscalar host and an invading pathogen discussed in the next section.
subunits of host PI3K. Other studies have shown that interaction between host and influenza A PPI domains can determine the virulence of the virus. Similar examples of host molecular mimicry exist in bacteria; for instance, interaction of the *Escherichia coli* type III effector protein espF with human sorting nexin 9 protein (SNX9) induces host membrane remodeling. These studies suggest that interactions between host and pathogen molecular networks are not random but rather well orchestrated, directed events that can influence both host and pathogen behavior. A holistic view of HPIs therefore requires a systematic and comprehensive study of host and pathogen molecular networks separately and in combination.

The different databases available for host genome-scale PPIs are reviewed in detail in section **Molecular Networks**. As virus genomes are much smaller, genome-scale PPIs were first generated in *E. coli* bacteriophage T7; this PPI consisted of only 2.5 viral-protein interactions. Subsequently, genome-scale PPI maps have also been generated for other viruses, such as hepatitis C, vaccinia, herpes, and SARS-associated coronavirus. All these studies used Y2H experiments to identify the interactions. Most viruses, except herpesviruses, have fairly small PPIs. For two members of the herpesvirus family, i.e., Kaposi sarcoma-associated herpesvirus and varicella-zoster virus, a comparatively large number of PPIs, 123 and 173, respectively, have been identified. Topological analyses of PPI networks of several viruses have revealed that nodes in viral PPI networks tend to form a single and highly connected module with many hubs and few peripheral nodes. This observation is in striking contrast to the network topology of eukaryotic PPIs that have few hubs and several peripheral nodes.

Compared to viral PPI networks, bacterial PPI networks are much larger and more complex. PPI networks have been generated for several bacteria including *Helicobacter pylori*, *Campylobacter jejuni*, *Treponema pallidum*, *Mycobacterium tuberculosis*, and *Bacillus subtilis*. Several studies have identified PPI networks in *E. coli*, a well-characterized Gram-negative bacterium, using experimental and computational approaches. The *E. coli* PPI network is one of the largest bacterial networks currently available with more than 10,000 interactions. These experimental PPI networks have been most widely generated using two methods, namely Y2H systems and tandem AP-MS. Topological analyses of bacterial PPI networks have revealed that they follow power-law distributions, i.e., they have few hubs and many peripheral proteins. Similar to bacterial PPI networks, protozoan pathogen PPI networks as in the case of *Plasmodium falciparum* also follow scale-free distribution.

Collection of HPI data in a high-throughput and systematic way has become available only in recent times. Several web-based resources are currently available that compile HPI data from different sources, including high-throughput experiments, literature curation, and computational predictions. A list of some of the widely used resources and their descriptions are included in Table 3. All these resources have come up with their knowledge base, which stores data in a systematic format and provides tools to query the database and display results through their web interface. Some of these applications also provide sophisticated tools to visualize interactive HPI networks. As most of the databases mainly rely on similar sources, the results are expected to overlap. However, some discordance between the results is not completely unexpected. Some databases are tailored for host- and pathogen-specific information. For example, GPS-Prot stores only human–HIV interactions, while PATRIC is a repository of bacterial interactions with many host species. A few recent studies have created meta-databases, such as PHISTO, PHI-base, and HPIDB, that combine interactions from a large number of host and pathogen species. Topological analysis of proteins involved in interaction of host and pathogen networks indicates that both bacterial and viral pathogens in general target hub and bottleneck proteins in host networks. These target proteins are associated with processes that help propagate infection or evade host defense mechanisms, for instance, evasion of apoptosis, inhibition of innate immunity, and reorganization of the actin cytoskeleton.

Computational predictions of HPI are useful when experimental data are unavailable, incomplete, or noisy. Several types of computational approaches have been used for this purpose; for instance, machine learning-based approaches have been applied for automated mining of scientific literature to predict HPIs. Although these methods are cheap and scalable, their performance should be carefully measured by estimating sensitivity and specificity. Interactions between human proteins and several pathogen proteins have also been predicted based on interlogs, an approach that extends protein interactions based on sequence homology and orthology. Because proteins usually interact through specific domains, algorithms for predicting domain–domain interactions can help identify potential PPIs. For instance, computational prediction of short linear sequence motifs common to both HIV-1 and human
TABLE 3 | Publicly Available Host–Pathogen Interaction Databases and Tools

<table>
<thead>
<tr>
<th>Database</th>
<th>Website</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHISTO149</td>
<td><a href="http://www.phisto.org">http://www.phisto.org</a></td>
<td>This meta-database combines interaction data from different sources using computational tools and currently contains ~2400 interactions between human and 300 pathogen strains (247 viral, 45 bacterial, 3 fungal, and 5 protozoan). The web interface also allows for network analysis and visualization.</td>
</tr>
<tr>
<td>PHI-base150</td>
<td><a href="http://www.phi-base.org">http://www.phi-base.org</a></td>
<td>A comprehensive database containing experimentally verified pathogenicity, virulence, and effector genes from bacterial, fungal, and oomycete pathogens, which infect a broad range of hosts, such as plant, vertebrate, and fungal species. The data can be easily downloaded and visualized with network-rendering tools. The current version contains more than 4000 interactions between 166 pathogens and 110 hosts.</td>
</tr>
<tr>
<td>HPIDB151</td>
<td><a href="http://agbase.msstate.edu/hpi/main.html">http://agbase.msstate.edu/hpi/main.html</a></td>
<td>A host–pathogen interaction (HPI) meta-database that contains 22,841 unique interactions between 49 hosts and 319 pathogens obtained by combining interspecies interaction data from six publicly available databases, i.e., BIND, MINT, PIG, GENERIC, REACTOME, and IntAct. The website does not provide network visualization, but the interaction network information can be downloaded and visualized using tools such as Cytoscape.</td>
</tr>
<tr>
<td>Phinet152</td>
<td><a href="http://www.phidias.us/phinet/">http://www.phidias.us/phinet/</a></td>
<td>A HPI network visualization tool that uses PHIDIAS knowledgebase created by combining manually curated data with existing databases, such as KEGG and MiNet. PHIDIAS database currently includes 100 pathogens, including 58 bacteria, 37 viruses, 2 parasites, and 3 fungi. The multilayered and interactive network rendition also provides detailed information on cellular localization of host–pathogen proteins and complexes.</td>
</tr>
<tr>
<td>PATRIC153</td>
<td><a href="http://www.patricbrc.org">http://www.patricbrc.org</a></td>
<td>A database that provides diverse types of data and bioinformatic tools specifically for bacterial pathogens. Host-bacterial protein–protein interactions (PPIs) can be downloaded and viewed using interactive network diagrams.</td>
</tr>
<tr>
<td>VirhostNet154</td>
<td><a href="http://pbildb1.univ-lyon1.fr/virhostnet/login.php">http://pbildb1.univ-lyon1.fr/virhostnet/login.php</a></td>
<td>A manually curated database of high-confidence virus–virus and human–virus PPIs for 180 distinct viral species. It also provides additional information about the proteins, such as structural and functional annotation, gene ontology term, pathway membership, and INTERPRO domain.</td>
</tr>
<tr>
<td>CAPIH155</td>
<td><a href="http://bioinfo-dbb.nhri.org.tw/capih">http://bioinfo-dbb.nhri.org.tw/capih</a></td>
<td>A web-based tool for comparison of PPIs between HIV and four different hosts, i.e., human, chimpanzee, rhesus macaque, and mouse. CAPIH can highlight host-specific interactions due to genetic differences and displays the results, including HPI networks, through interactive Java Applets.</td>
</tr>
<tr>
<td>GPS-Prot156</td>
<td><a href="http://www.gpsprot.org">http://www.gpsprot.org</a></td>
<td>A meta-database that integrates PPI data between human and HIV from three publicly available databases, including MINT, BioGRID, and HPRD. It also provides Java Applet-based interactive tool for interactive visualization of HPI networks.</td>
</tr>
</tbody>
</table>

Proteins has provided the basis for understanding the interaction with and modification of the host PPI network by the virus.\textsuperscript{165,166} To improve computational predictions, pipelines can be created that include combinations of the above approaches and impose additional constraints.\textsuperscript{157} For example, interaction between a pair of host and pathogen proteins can be predicted based on availability of interacting domain pairs from relevant databases. Once all such pairwise interactions have been predicted, the results can be filtered to retain only hub genes in the host PPI network, as these genes are the ones most likely to be targeted by pathogen proteins to gain access to the host network(s).
CONCLUSIONS AND TRANSLATIONAL IMPLICATIONS

The immune system is particularly well suited for the application of network reconstruction methods. From dynamic molecular interactions underlying intracellular signaling pathways, controlled by GRNs and also regulated by epigenetic modifications, to intercellular crosstalk between the many distinct immune cell subtypes, organization of cells into tissues and organs, and finally, to the overall response of an organism, there are multiple hierarchical levels that are amenable to network-based modeling and visualization. Although we are still far from the eventual goal of applying network-based approaches to the construction of comprehensive predictive models of the immune response to diverse perturbations, the examples described in the preceding sections highlight how systems-level network analyses have accelerated our understanding of molecular interactions underlying the immune response. Large-scale ‘omics’ approaches need to be combined with recent advances in computational biology and mathematical tools to allow for the generation of comprehensive quantitative and predictive network models across multiple scales of organization in the immune system. Iterative refinements of such dynamic models through experimental validation should eventually allow us to simulate immune behavior and infer responses that cannot be predicted a priori.

In addition to deciphering networks that underlie the immune response, the field of immune network biology also holds promise for unraveling genetic and molecular networks that drive complex human diseases, and drug-target networks that can help power new drug discovery. Unbiased genetic approaches, such as genome-wide association studies (GWAS), have uncovered many genetic loci and processes that are dysregulated in human immune disorders. These causal genetic alleles and the products they encode are in turn rooted within complex molecular and cellular networks. Each disease-associated allele can thus result in dysregulation of multiple circuits, and disease states can be considered emergent properties of molecular networks rather than the result of single gene mutations. Integrating large-scale high-dimensional data such as genome-wide single-nucleotide polymorphism genotyping data, GWAS data, whole genome transcriptome, next-generation sequencing, and ChIP-seq data will provide a more comprehensive view of disease states and will be a key to better understanding of the molecular networks underlying physiological states associated with disease. A major challenge will be construction of predictive probabilistic networks that can infer causality between disease phenotypes and underlying molecular circuits. Development of sophisticated computational methods for statistically associating specific network behaviors with upstream gene alleles and downstream clinical outcomes will be critical to define causality, and to determine biomarkers of disease, and candidate targets for therapeutic intervention.

Finally, network-based approaches hold potential for the identification of novel actionable drug targets as well as the repositioning of established targets. This may be accomplished by the integration of drug-target networks with PPI and other biological networks. Recently, the recognition that disease-causing alleles are embedded in complex molecular and cellular networks has shifted the focus of drug discovery from the ‘one-target, one-drug’ model to a ‘multitarget, multidrug’ model aimed at systemically modulating multiple nodes in a network. In addition, studies have revealed that many diseases have common genetic origins and that drug targets are often involved in multiple diseases, indicating potential for application of known drugs to new diseases. The ultimate goal will be to integrate information from large-scale ‘omics’ approaches with causal disease pathways and drug-target analysis in order to build a comprehensive computational platform to drive multitarget multidrug discovery. In future clinical applications, we envision construction of individual-specific network models for personalized medicine that may empower physicians to treat patients according to previously unutilized and uncoordinated levels of biological information. Integrating this information from multiple biological levels remains a challenge that will lay the cornerstone for a functional systems approach to human immunology that can reveal the precise cellular and molecular networks that contribute to human diseases and how they can be targeted to ameliorate disease.

ACKNOWLEDGMENTS

This work was generously supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, and institutional support funds from the Institute for Systems Biology to NS.
REFERENCES


109. Kellogg RA, Gomez-Sjoberg R, Leyrat AA, Tay S. High-throughput microfluidic single-cell analysis...


