Chapter 22
Systems-Level Analyses of the Mammalian Innate Immune Response

David J. Lynn, Jennifer L. Gardy, Christopher D. Fjell, Robert E.W. Hancock, and Fiona S.L. Brinkman

Abstract The regulation of the innate immune response, our first line of defence against infectious disease, does not involve simple linear pathways but rather complex inter-connected networks of interactions, regulatory loops, and multifaceted transcriptional responses. Given this complexity, systems biology approaches to investigate the host innate immune response are essential. InnateDB (www.innatedb.com) is a publicly available database and integrated analysis platform specifically designed to facilitate systems-level analyses of the mammalian innate immune response and is one of the most comprehensive databases of all human and mouse molecular interactions (130,000+) and pathways (3,000+). Building upon this, more than 12,900 innate immunity-relevant molecular interactions have been contextually annotated through detailed review of the literature providing novel insight into the innate immunity interactome. Integrated bioinformatics solutions include the ability to investigate user-supplied quantitative data in a network and pathway context using pathway, ontology and transcription factor over-representation analyses, and network visualisation and analysis tools. In this chapter, we introduce innate immunity as a complex system and a provide a detailed step-by-step guide to using InnateDB and other bioinformatics tools to investigate the host response to infectious disease.

Keywords Systems immunology · Innate immunity · Database · InnateDB · Pathway analysis · Network analysis

D.J. Lynn (✉)
Department of Molecular Biology and Biochemistry, 8888 University Drive, Simon Fraser University, Burnaby, BC, Canada, V5A 1S6; Animal & Bioscience Research Department, Teagasc, Grange, Dunsany, Co. Meath, Ireland
e-mail: dlynn@sfu.ca; david.lynn@teagasc.ie

S. Choi (ed.), *Systems Biology for Signaling Networks*, Systems Biology 1,
DOI 10.1007/978-1-4419-5797-9_22,

22.1 Introduction to Innate Immunity

All organisms are faced with a constant bombardment of microbes which they must recognize and mount an appropriate immune response against. The immune response has traditionally been divided into two different branches, the adaptive immune response and the innate immune response. Adaptive immunity, which evolved in the first jawed vertebrates, is the immune response that is primarily associated with T and B lymphocytes (although other cell types, such as dendritic cells, are also important). The adaptive immune response, through the rearrangement of antigen receptor gene segments, can produce vast numbers of receptors capable of recognizing many millions of foreign antigens. Adaptive immunity discriminates between self and non-self and is also associated with immunological memory whereby the immune response can mount a faster and more efficient response to an antigen that it has previously encountered.

Innate immunity, on the other hand, represents a much more ancient system and is the first line of defense against infectious agents in species ranging from insects to plants to mammals, including humans. Until a little over a decade ago, innate immunity was considered to be of little interest in species with an adaptive immune system. However, since the discovery of innate immunity receptors and pathways that are evolutionarily conserved between *Drosophila* and humans (Medzhitov et al. 1997) there has been an explosion of interest in innate immunity. Innate immunity is now known to be critical for the response to pathogens not only in plants and invertebrates but also in humans and other vertebrates, effectively dealing with most short-term microbial insults without the need for an adaptive response. Innate immunity also responds very quickly to an infectious agent (within hours) in comparison to adaptive immunity, which often takes several days to mount a sufficient response. Innate immunity is now also recognized as essential to inducing an appropriate adaptive response and importantly, the effector mechanisms of innate immunity overlap considerably with those of the adaptive immune response (MacLeod and Wetzler 2007; Manicassamy and Pulendran 2009).

Innate immunity may be constitutive or inducible in response to pathogens. The constitutive components of the innate immune system include the barrier functions of the epithelial layers of the skin and mucous membranes, which prevent entry of pathogens. Physiological barriers such as temperature and pH can limit the survival conditions of particular microorganisms. Furthermore, certain enzymes, such as lysozyme, are constitutively expressed and under some circumstances can destroy invading pathogens by disrupting bacterial membrane integrity.

Most defense mechanisms are, however, inducible and require the specific recognition of infectious microorganisms. Epithelial cells act not only as a physical barrier to pathogens, but are also a primary site for the production of inducible chemokines and host defense peptides, such as defensins, that orchestrate innate immunity. Other cell types such as macrophages, neutrophils, and dendritic cells, can phagocytose or internalize invading pathogens and kill them. These cell types are also responsible in a large part for the production of pro-inflammatory cytokines such as interleukins and tumor necrosis factor (TNF). Inflammation is a component of innate immunity that plays a number of key roles in combating infections that are centered around

local effects such as the recruitment of additional immune response cells to the site of infection. If not tightly regulated, however, inflammation can be detrimental to the host and an overwhelming immune response can lead to what is sometimes called a cytokine storm. Sepsis, for example, is a dysregulated systemic inflammatory response that results in more than 200,000 deaths a year in the United States alone (Angus et al. 2001).

Cytokines also cause the induction of acute-phase proteins, which modulate immunity serving a broad variety of functions ranging from inhibition of microbes through to coagulation and regulation of inflammation. Other cytokines known as interferons are induced in response to microbial signals (particularly viral RNA) and essentially induce an antiviral state, although they also have roles in defense against other microbes ranging from intracellular bacteria to parasites. These proteins also activate natural killer (NK) cells and macrophages that can directly kill or engulf infected cells.

Despite the lack of true antigenic specificity evident in adaptive immunity, which is possible by the rearrangement of antibody and receptor genes, the germ-line encoded components of the innate immune system still permit quite a broad coverage in defense against pathogens. This is accomplished through the recognition of conserved motifs or signatures (often termed pathogen associated molecular patterns or PAMPs) on the surface or within the invading pathogens. These pathogen signature molecules are usually functionally important to the microorganism, are conceptually conserved, and are relatively resistant to mutation. A wide variety of pathogen signatures have been identified including lipopolysaccharide (LPS), peptidoglycan, lipoteichoic acid, lipopeptides, flagellin, bacterial CpG DNA, and viral nucleic acids. The receptors responsible for interactions with bacterial signatures are known as pattern recognition receptors (PRRs) (Medzhitov and Janeway 1997). PRRs are known to be expressed by a range of cell types, including leukocytes and epithelial cells. A wide array of structurally and functionally diverse PRRs has evolved.

The best-studied family of PRRs in humans are the Toll-like receptors (TLRs) (see for review Akira 2006). Ten human TLR genes have been identified and the encoded receptors have specificity for a range of PAMPs (Chuang and Ulevitch 2000, 2001; Medzhitov et al. 1997; Rock et al. 1998; Takeuchi et al. 1999). TLR2, for example, recognizes lipoteichoic acid, LPS is the ligand for TLR4, TLR3 recognizes double-stranded RNA, and TLR5 recognizes flagellin, a protein component of bacterial flagella. Activation of one of the TLRs by its relevant PAMP results in a signaling cascade that leads to the activation of nuclear factor kappa B (NF-κB) and other transcription factors. These transcription factors then regulate the expression of a very large array of effector genes.

22.2 Complexity of Innate Immunity – Why Systems Approaches are Necessary

Over the last ten years, considerable progress has been made in understanding the innate immune response including the detailed investigation of many of the critical

signaling pathways involved. It is becoming increasingly clear, however, that the innate immune response does not involve straightforward linear pathways but rather complex inter-connected networks of pathways and molecular interactions, feed-forward and feedback regulatory loops, multifaceted transcriptional responses, and regulation at the transcriptional, post-transcriptional, and post-translational levels (Gardy et al. 2009; Lee and Kim 2007; Tegner et al. 2006) (Fig. 22.1).

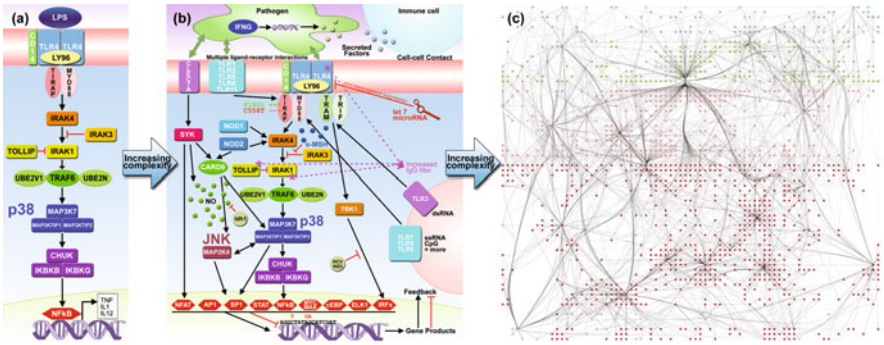


Fig. 22.1 Appreciating the complexity of the innate immune response. (**a**) Illustrates a canonical view of mammalian TLR4 signaling via NF-κB, in which the process is represented as a simple linear pathway. (**b**) Demonstrates that this simple pathway is in fact much more complex and is under multiple levels of control. (**c**) Illustrates a more accurate depiction of the complexity of the system. InnateDB was used to retrieve interactions for each of the 47 proteins shown in (**b**) and the resulting network was visualized using Cerebral/Cytoscape. It contains 1,346 genes or proteins and 2,531 interactions. Reprinted from Gardy et al. (2009), with permission from Elsevier

It has become evident that the TLR signaling pathways (discussed above) are just one of several PRR pathways. The importance of cytoplasmic PRRs in the intracellular recognition of pathogens has become increasingly clear. These cytoplasmic PRRs include the nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs) (Inohara and Nunez 2001; Kanneganti et al. 2007) and the retinoic acid-inducible gene 1 (RIG-1)-like receptors (RLRs) (Thompson and Locarnini 2007; Yoneyama et al. 2004). RIG-I, for example, signals via several proteins, including mitochondrial antiviral signaling (MAVS), to activate transcription factors such as NF-κB and interferon regulatory factor 3 (IRF3), which regulates the expression of type I interferons. A broad range of additional pathways are now implicated as having a role in innate immunity.

Adding to this complexity is that a single pathogen signature molecule can trigger multiple different pathways simultaneously – both RIG-I and TLR3 recognize dsRNA, for example (Honda and Taniguchi 2006), while LPS stimulates a variety of myeloid differentiation primary response gene 88 (MyD88)-dependent and (MyD88)-independent pathways. Furthermore, during infection it is likely that several different pathogen signatures are presented simultaneously. Simultaneous stimulation of TLRs by multiple ligands has been shown to result in complex synergistic responses in terms of cytokine production (Hsueh et al. 2009).

Many of the signaling pathways in the innate immune response also display additional layers of complexity including cross-talk between pathways, such as the recently described cross-talk between the TLR4 and cAMP pathways in macrophages (Wall et al. 2009). There is also apparent redundancy in signaling components leading to unexpected phenotypes, such as the case of patients deficient in MyD88 (von Bernuth et al. 2008). MyD88 is a central adaptor molecule in TLR signaling – yet MyD88-deficient patients do not show broad susceptibility to infection as would be expected. Feedback and feedforward loops provide an additional layer of signaling complexity. The suppressor of cytokine signaling (SOCS) proteins, for example, have been identified as a negative feedback loop in cytokine signaling (Dimitriou et al. 2008).

The complexity of the innate immune response is particularly apparent at the transcriptional and post-transcriptional levels. In terms of the transcriptional response, hundreds or even thousands of genes may be differentially expressed. Exposure of peripheral blood mononuclear cells (PBMCs) to LPS (a component of Gram-negative bacteria), for example, induces a potent innate immune response. This single stimulus induces changes in several hundred genes; some of which are specifically involved in immunity and others which form part of a general response to injury (Brownstein et al. 2006). Adding to the complexity is the fact that the response to a stimulus varies not only with the cell type being examined but also with the particular time point. Monocytes, B-cells, T-cells, and dendritic cells, for example, have all been shown to have different responses to LPS (Mookherjee et al. 2009), while significant changes are observed in the LPS response between 2, 7, and 24 h (Nilsson et al. 2006), with complex dynamics occurring depending on the duration of exposure.

Although important transcription factors such as NF-κB and activator protein 1 (AP-1) are often discussed as the downstream activators of the innate immune response, it is now evident that the transcriptional response is regulated by panels of transcriptional factors rather than one or two factors. These panels of transcriptional factors are differentially activated over the course of the innate immune response and have complex behavior depending on the type and strength of microbial stimulus. Several transcription factors, for example, have been shown to be activated in response to the immunomodulatory host defense peptide, LL-37 (Mookherjee et al. 2009). Additionally, a series of papers by the Institute of Systems Biology in Seattle, has examined which transcription factors are active over the time-course of the LPS response in macrophages, revealing that transcription factor binding sites in the promoter regions of genes at later time points were enriched for early response factors (Gilchrist et al. 2006; Ramsey et al. 2008). Recently, a regulatory network, involving the mouse transcription factors CCAAT/enhancer-binding protein delta (Cebpd), activating transcription factor 3 (Atf3) and NF-κB, was found to distinguish between transient and persistent signaling via TLR4 (Litvak et al. 2009).

Until recently, much of what was known about which transcription factors regulate which genes was computationally predicted based on binding site analysis of a gene's promoter. Computational prediction of gene regulation by transcription factors, however, has a high false positive rate due to the fact that the short

transcription factor binding site sequences can occur by chance throughout the genome (Kolchanov et al. 2007). Fortunately, transcription factor binding can now be verified on a genomic scale by the powerful experimental methods based on chromatin immunoprecipitation (ChIP) (Collas and Dahl 2008). In these techniques, DNA sequences that are bound by proteins are either identified by microarray "chips" (ChIP-chip) or by sequencing (ChIP-seq). Using ChIP-chip methods, the transcription factors NF-κB and interferon regulatory factor 1 (IRF1) were determined to bind a set of co-expressed genes in TLR-stimulated murine macrophages (Ramsey et al. 2008) and a gene cluster regulated by the transcription factor STAT3 was found to be downstream of leukemia inhibitory factor (LIF) (Langlais et al. 2008). Similarly, using ChIP-seq, STAT1 was found to bind a large number of interferon-responsive human genes (Robertson et al. 2007).

Aside from transcriptional regulation, other multiple layers of regulation all add to the complexity of the innate immune response. MicroRNAs, for example, are now being shown to be pivotal regulators of the innate immune response (Bi et al. 2009; Pedersen and David 2008). MicroRNAs are short (typically ~22 nucleotides), non-coding RNA molecules that bind specifically to mRNA and act as post-transcriptional regulators of gene expression altering mRNA stability and translation efficiency. More than 700 microRNAs have been found in mammalian cells and they play roles in a range of processes including development, homeostasis, and differentiation, as well as regulation of immune system responses in both adaptive and innate immunity. The microRNA let-7i, for example, has been shown to regulate TLR4 expression (Chen et al. 2007), while mir-146 regulates several signaling proteins involved in the innate immune response (Taganov et al. 2006).

Changes to the activity of proteins can also occur due to post-translational modifications which alter the biological activity of the protein without necessarily changing gene expression. Such modifications include degradation (proteolysis), as well as covalent modifications (for example ubiquitination, phosphorylation, and acetylation). The protein Eyes absent 4 (EYA4) has recently been shown to enhance the innate immune response to viruses. This activity is abolished through the mutation of its threonine-phosphatase domain, revealing that this protein regulates innate immunity via changes in the phosphorylation of target signaling proteins (Okabe et al. 2009). A series of other phosphorylation events are known to be essential to signaling in innate immunity. Additionally, ubiquitination, a reversible covalent modification of proteins involving the addition of ubiquitin, has also emerged as a key post-translational mechanism in regulating innate immunity (Bhoj and Chen 2009). For example, two E3 ubiquitin ligases, cellular inhibitor of apoptosis protein 1 (cIAP1) and cIAP2, have been shown to be required for NLR signaling (Bertrand et al. 2009).

To account for these many layers of regulation and complexity (and numerous more not discussed here) in investigations of the innate immune response, researchers are now adopting systems biology approaches. In the next section we will discuss some computational tools that are emerging to facilitate such efforts.

22.3 Computational Resources for Innate Immunity

As the complexity of the innate immune response has become appreciated and researchers begin to adopt systems biology approaches to investigate innate immunity (Andersen et al. 2008; Gilchrist et al. 2006; Oda and Kitano 2006; Tegner et al. 2006), a variety of bioinformatics resources with an innate immunity or broader immunology focus have been developed that facilitate these efforts. Several groups, for instance, offer immunology-relevant transcriptomics data sets, in addition to the large microarray repositories. These databases and data sets include the Reference Database of Immune Cells (RefDIC, refdic.rcai.riken.jp) (Hijikata et al. 2007), the Immune Response In Silico database (IRIS, share.gene.com/clark.iris.2004/iris/iris.html) (Abbas et al. 2005), and resources from the Institute for Systems Biology (ISB), including Affymetrix expression data from TLR ligand-stimulated mouse macrophages (www.innateimmunity-systemsbiology.org) (Korb et al. 2008). Additionally, the Immunological Genome Project (www.immgen.org) (Heng and Painter 2008) is a network of laboratories generating rigorously standardized genome-wide gene expression data sets of over 200 different mouse immune cell populations under a variety of conditions. Other valuable resources include the Innate Immune Database (db.systemsbiology.net/IIDB), which stores both predicted and ChIP-chip verified transcription factor binding sites (Korb et al. 2008). The IIDB currently includes data on the regulation of 2000 mouse genes that are differentially expressed in LPS-stimulated macrophages.

Of particular note is InnateDB (www.innatedb.com), the first database and integrated analysis platform specifically designed to facilitate systems-level analyses of the innate immune response (Lynn et al. 2008). Although InnateDB manual curation, as the name suggests, has focused on annotating molecular interactions and pathways involved in the innate immune response, InnateDB is, in fact, one of the most comprehensive databases of all human and mouse molecular interactions and pathways, consisting of more than 130,000 molecular interactions and 3,000+ pathways, integrated from the major public molecular interaction and pathway databases. To enrich our knowledge of innate immunity networks and pathways, the InnateDB curation team has contextually annotated more than 12,900 innate immunity-relevant molecular interactions through the review of 3,000 plus biomedical articles. Interactions are annotated using Open Biomedical Ontology (OBO) controlled vocabulary for terms such as cell type, tissue type, interaction detection method, etc., in compliance with the Proteomics Standards Initiative Molecular Interaction (PSI-MI) 2.5 XML format (Hermjakob et al. 2004), and in a manner that adheres to the recently proposed "minimum information required for reporting a molecular interaction experiment" (MIMIx) guidelines (Orchard et al. 2007; Smith et al. 2007).

InnateDB is also an analysis platform, offering seamlessly integrated, user-friendly bioinformatics tools, including pathway and ontology analysis, network

visualization and analysis, and the ability to upload and analyze user-supplied gene expression or other quantitative data in a network and/or pathway context. In this chapter, we present a guide to using InnateDB and selected other computational resources to explore quantitative -omics data in a more systems-oriented manner.

22.4 A Walk Through the Analysis of a Smallpox Gene Expression Data Set Using InnateDB – Pathways, Processes, and Interaction Networks

22.4.1 Introduction

One of the most powerful features of InnateDB is the ease with which one can quickly analyze a gene list with associated quantitative data – for example, differentially expressed genes identified through a microarray experiment – and generate hypotheses about the molecular functions, biological processes, signaling pathways, and molecular interaction networks or sub-networks that may be contributing to the phenomenon being investigated.

Previously, interpretation of such data sets required submitting one's data to multiple Web sites. Each Web site typically had its own preferred file format, often offered only a single type of analysis, and generated output in different formats. In contrast, the analysis environment provided by InnateDB brings many of the most popular types of analysis together "under one roof". With just a few clicks, an InnateDB user can submit a single file to multiple types of analyses and can download the results in a simple, intuitive spreadsheet format or interactively visualize their results in the context of a molecular interaction network.

In the following sections, we will demonstrate how InnateDB can be used to interpret the results of a microarray experiment. We will begin with examining the type of data that can be submitted to InnateDB, and will move on to how to retrieve gene, interaction, and pathway data as well as perform Gene Ontology (Ashburner et al. 2000) and pathway over-representation analyses. We encourage the reader to follow along with our analysis using the sample data file provided (InnateDB_Sample_Data.txt).

The sample data set is from an interesting paper published in the *Proceedings of the National Academy of Sciences* (Rubins et al. 2004) and is freely available to all readers at http://www.pnas.org/content/101/42/15190.full. In this study, the researchers examined the host response to smallpox. Smallpox is a highly contagious disease caused by variola virus that can result in several distinct clinical outcomes. While some smallpox infections lead only to minor disease, most infections lead to "ordinary smallpox" – a disfiguring disorder in which the body is covered in pustules which can, in severe cases, merge into a single large confluent pustule, essentially detaching the outermost layer of skin. Ordinary smallpox is fatal in approximately 30–75% of cases depending on confluence of the pustules, while

"hemorrhagic smallpox" – an especially severe form of the disease that can develop in certain patients – is nearly 100% fatal.

Through vaccination, smallpox has been eradicated. Since 1977, no natural cases of the disease have occurred, and the only remaining stocks of the virus exist in tightly controlled international repositories in Atlanta, Georgia and Novosibirsk, Russia. Concerns over use of the virus as an agent of bioterrorism remain, however, and thus research into the disease continues (Cohen and Enserink 2002).

The variola virus causing smallpox is a double-stranded DNA virus whose genome encodes 197 proteins and many of which are known to interfere with the host's innate immune response (Seet et al. 2003). In the Rubins et al. study (Rubins et al. 2004), the authors explored the host response to smallpox using microarrays to gain further insight into how the virus subverts the host's innate defenses. Working under Biosafety Level 4 and using the virus stock housed at Atlanta's Centers for Disease Control and Prevention, the authors infected macaques with smallpox virus, using these primates as a surrogate for the human system. Twenty-two macaques were exposed to smallpox through various routes of administration and peripheral blood samples were collected at several time points following infection. The study generated a significant amount of data; however, for the purposes of our demonstration, we will be examining only the data from two time points – three and four days after a combined IV/aerosol infection. The complete data set is available at http://microarray-pubs.stanford.edu/smallpox/raw_data.html. Our sample data file contains fold-change values (and associated *p*-values) for approximately 500 genes that were up- or down-regulated at least 1.5-fold at day 3 and/or day 4 in comparison to day 0 (uninfected). Note that although the experiment was performed in macaques, human microarrays were used to profile gene expression changes. InnateDB's analysis tools can only be used on human or mouse data at present, although orthology predictions may be used to map data from other species to human/mouse identifiers.

22.4.2 Preparing Data for Analysis in InnateDB

Following a microarray experiment or any other experimental approach in which a list of interesting genes is generated, a researcher must prepare his/her gene list for analysis. In the case of preparing microarray or other quantitative data for submission to InnateDB, we recommend the following steps:

1. Ensure each gene in the data set is identified with one of the cross-reference identifiers (accession numbers) InnateDB supports. Supported identifiers include Ensembl IDs (gene, protein, or transcript) (Hubbard et al. 2009), EntrezGene (Maglott et al. 2005), RefSeq (Pruitt et al. 2009), and UniProt (The UniProt Consortium 2008) (Table 22.1). If your data is in a different format than those supported, there are a number of freely available Web-based tools that can translate from one type of identifier to another, for example, IDconverter (Alibes et al. 2007).

Table 22.1 InnateDB supports several common identifiers. Here, the various identifiers for the human Toll-like receptor 4 gene are shown as an example

Identifier	Web site	Format
Ensembl Gene	http://www.ensembl.org	ENSG00000136869
Ensembl Transcript	http://www.ensembl.org	ENST00000394487 ENST00000355622
Ensembl Protein	http://www.ensembl.org	ENSP00000377997 ENSP00000363089
Entrez Gene	http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene	7099
RefSeq	http://www.ncbi.nlm.nih.gov/RefSeq/	NM_138554
UniProt	http://www.uniprot.org/	O00206

Table 22.2 Sample data for submission to InnateDB. Each gene is identified using one of the accepted identifiers listed in Table 22.1, and fold change and associated *p*-values are given for each gene at two time points

Ensembl Gene	Day 3 fold change	Day 3 *p*-value	Day 4 fold change	Day 4 *p*-value
ENSG00000128274	1.5	0.015	1.7	0.014
ENSG00000206410	1.8	0.020	1.7	0.007
ENSG00000143878	1.1	0.026	3.3	0.008
ENSG00000014257	2	0.049	1.7	0.011
ENSG00000172594	1.5	0.010	−0.3	0.029

2. Although InnateDB can be used in the analysis of a list of gene identifiers without any associated quantitative data, such analyses are more powerful if quantitative data is included with the gene list. In our sample data set, we have fold change in gene expression values from two time points – day 3 and day 4 – for each of our genes. This data is also associated with *p*-values that provide a measure of the biological variability associated with each fold-change value. These *p*-values can be included in your file for submission to InnateDB, although they are not required. InnateDB expects fold-change values to be absolute values (not log values) in the format where +2 represents a twofold increase in gene expression and –2 represents a twofold decrease.
3. In most microarray data sets, one encounters the issue of multiple probes mapping to a single gene. In these situations, InnateDB averages the multiple fold change and associated *p*-values such that a single average fold change and *p*-value is associated with each gene. One may prefer, however, to adopt an alternative approach to these cases – for example, keeping only the probe with the most significant *p*-value. If desired, such changes must be made to the data set prior to upload to InnateDB.

4. Before submitting data to any analytical tool, it is worthwhile considering the comparisons that one may ultimately want to make. In an experiment with a single condition, it is often sufficient to simply analyze the complete list of genes differentially expressed at that condition. In experiments with multiple conditions, however, such as drug or no drug, and/or with multiple time points, it is preferable to identify subsets of genes of interest for analysis as considering the entire set can lead to errors in interpretation. For example, in an analysis comparing drug-treated cells to placebo-treated cells, an over-representation analysis performed on the complete data set may not yield meaningful results – it is only when genes expressed in response to the drug but not the placebo are considered that the analysis generates meaningful results. Before uploading a complex data set, consider splitting it into multiple files, one for each subset of genes you wish to analyze. With our sample data set, we may wish to consider each time point independently, or consider them both together.

22.4.3 Uploading Data to InnateDB

Once a tab-delimited text or Excel spreadsheet (.xls only) containing your genes of interest and, if desired, associated quantitative measurements and *p*-values has been prepared, it can be uploaded to the InnateDB Web site. This is illustrated in Fig. 22.2.

Fig. 22.2 Uploading data to InnateDB. Genes of interest can be uploaded directly from any tab-delimited text or Excel spreadsheet format (.xls only) or, provided less than 1000 genes are submitted, can be pasted into the text box. One of four common identifiers must be used to identify the genes, and InnateDB must be told which type of identifier was used and which column of the data set contains the identifiers. Expression data can also be provided

1. From the *Data Analysis* tab, select a file to upload by clicking on the "*Upload File*" button. One can upload a *tab-delimited* or *Excel* file of protein/gene identifiers or accession numbers and obtain a list of all genes, proteins, pathways, interactors, or interactions that they are associated with. Alternatively, click on the "Web Form" button and paste your tab-delimited data in the text box (max. 1000 lines). Note that there should be only one accession number per row. Probes that map to multiple genes should be removed.
2. On the next page, click on the column headers to specify which column in your data file contains the identifiers/accession numbers for each gene (and which database they come from) (Fig. 22.3). This is called the "*Cross-reference ID*". You can only specify one cross-reference ID column. Accession numbers from the following databases are currently accepted: Ensembl, RefSeq, EntrezGene, and UniProt. You must also specify the *Cross-reference database*. This is the database where the identifiers in the cross-reference column come from. In the sample data set column 1 contains Ensembl IDs which should be used as the cross-reference IDs.
3. If you have included *gene expression data*, identify which columns contain the gene expression values and their associated *p*-values (Fig. 22.3). You may also identify the column containing the probe IDs if you have included them in your

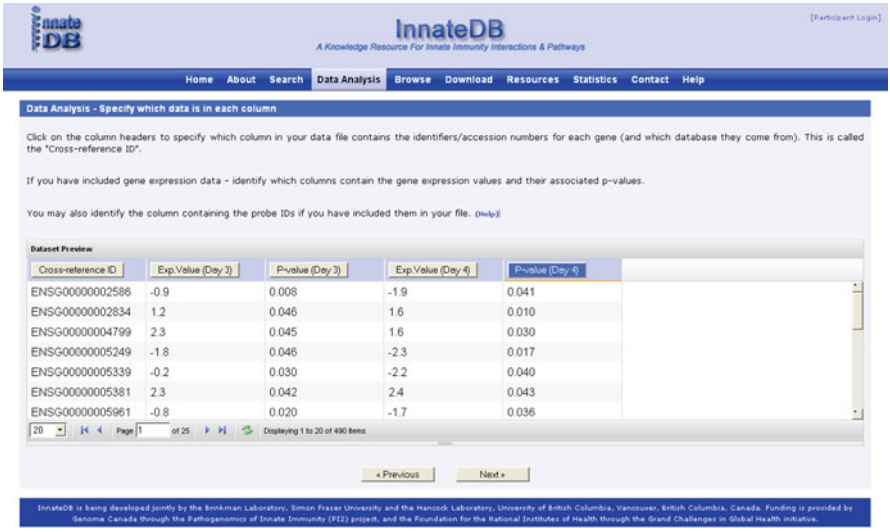


Fig. 22.3 Specify which columns in your uploaded file contain the relevant data. Click on the column headers to specify which column in your data file contains the identifiers/accession numbers for each gene (called the Cross-reference ID column) and which database they come from (called the Cross-reference database). If you have included gene expression data, identify which columns contain the gene expression values and their associated *p*-values. You may also identify the column containing the probe IDs

file (not included in the sample data set). Including quantitative data, for instance, gene expression values, is optional but recommended. It is a very useful way to investigate one's data in a pathway and interaction network context and to carry out subsequent analyses, such as pathway over-representation analysis. Remember, expression values must be in the format where a value of +2 represents a twofold increase in expression and a value of –2 a twofold decrease in expression. You can specify values from up to ten different conditions or time points. You can also specify a name for each condition. The sample data set contains gene expression data for day 3 and day 4 in column 2 and 4, respectively, with associated *p*-values for these conditions in column 3 and 5.

4. Choose whether you want to return interactions, interactors, genes, or pathways associated with your list of genes or proteins (Fig. 22.4). Returning a *list of interactions* allows one to identify all interactions in InnateDB involving the genes (or their encoded products) in the uploaded list and to construct a network of these interactions for visualization and further analysis. Detailed annotation and evidence are then available for each interaction. The resulting interaction network may then be downloaded in a variety of supported formats or interactively visualized. Returning a *list of interactors* allows one to identify all molecules in InnateDB which interact with the genes (or their encoded products) in the uploaded list. Returning a *list of genes* provides detailed annotation for each gene in the uploaded list and is a prerequisite to performing a Gene Ontology over-representation analysis. Returning a *list of pathways* provides pathway annotation for each gene in the uploaded list and is a prerequisite to performing a pathway over-representation analysis.

Fig. 22.4 Choose which data to return. Choose whether you want to return interactions, interactors, genes, or pathways associated with your list of genes or proteins

22.4.4 Performing a Gene Ontology Over-Representation Analysis

An over-representation analysis (ORA) examines a gene list for the occurrence of annotation terms, such as a Gene Ontology (GO) (Ashburner et al. 2000) term or pathway membership, which are more prevalent in the data set than expected by chance. The InnateDB over-representation analysis tools provide a number of statistical methods for pathway and GO analysis, including the Hypergeometric distribution (default), the Fisher exact test, and the Chi-square test. *P*-values are automatically corrected using the Benjamini and Hochberg correction for the false discovery rate (Benjamini and Hochberg 1995), although one can also opt to use the more conservative Bonferroni correction. In this way, annotations that occur more frequently than expected in a gene list can be identified, and may point toward a biological process or pathway that is being differentially regulated in the condition of interest.

1. Gene Ontology is a standardized method for representing gene product attributes such as their functions, the biological processes they participate in, and their cellular compartment, and thus is a useful way of finding out if a gene set has known roles in a wide variety of biological processes. To perform a GO ORA, first upload a gene list via the Data Analysis page as described above, and select the *Return a list of genes* option. You will be taken to the gene results page, providing detailed annotation for each gene you have uploaded, including the gene name, species, orthologs, chromosomal location and number of interactions in which it is a participant.
2. From the gene results page, click on the red *Ontology ORA* button at the top of the page.
3. You must now specify the type of data set you have provided, as it influences how InnateDB performs its statistical analysis (Fig. 22.5). If you have uploaded a complete microarray data set, where gene expression values for all probes on the array have been provided regardless of whether they are differentially expressed or not, choose the first option – *Complete microarray data set.* You must then specify a fold change and *p*-value cutoff to distinguish differentially expressed genes from genes that are unchanged in their expression. The default is a fold change in expression of 1.5 with a p-value < 0.05. Using this option the ORA algorithm considers the proportion of differentially expressed genes on the array in the calculation of statistical significance. This option, however, usually performs significantly more statistical tests than the second option discussed below as almost all GO terms are represented in the data set. This can mean that the correction for multiple testing is particularly conservative. The benefit of this option is that it can calculate the statistics for up to 10 conditions or time points simultaneously.

   If you are only providing a subset of genes from the microarray, as with the example data set, choose the second option – *Data set consisting of a subset of genes from the entire array*. In our example, although we have uploaded gene expression data for two conditions, the expression values will not be used in

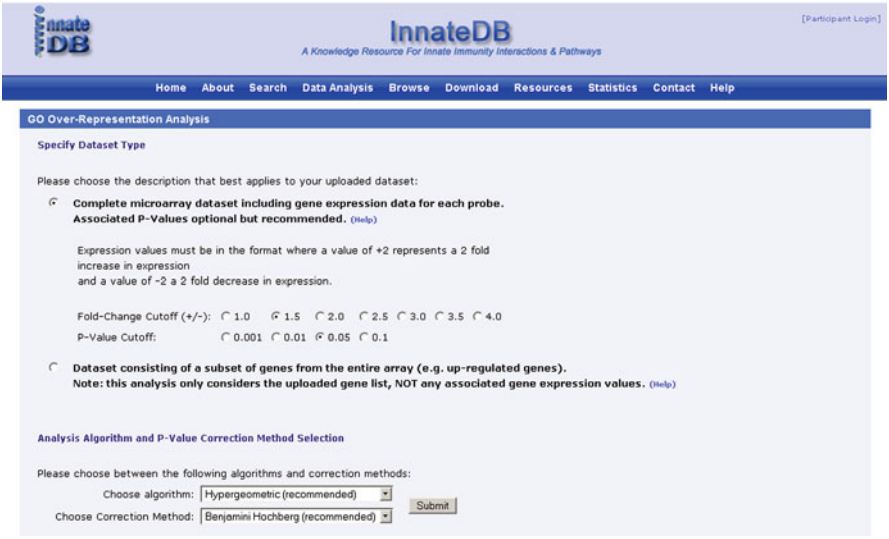
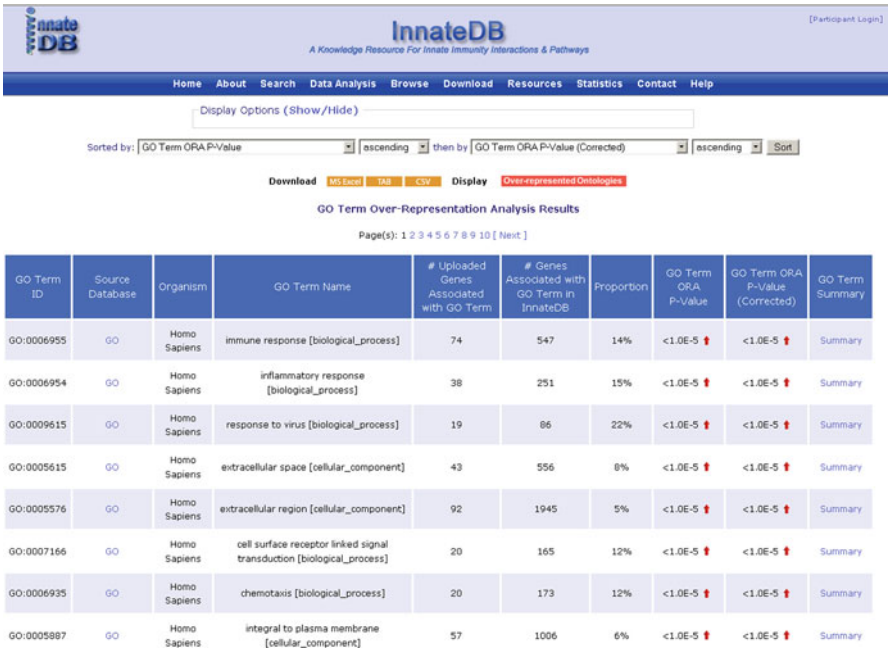


Fig. 22.5 Selecting the type of data for an over-representation analysis (ORA). To perform the appropriate statistical analysis, InnateDB must know whether a complete microarray data set has been uploaded or whether only a subset of genes from such an experiment was provided. For the purposes of our sample data set, the latter option – a subset of genes – should be selected. The analysis and correction methods can be changed using the drop down menus below, although we strongly recommend the use of the default parameters

the over-representation analysis. The analysis will be performed on the gene list provided regardless of any expression values and will calculate the statistics for the gene list and not each condition separately. If you wish to analyze each time point separately you will need to create separate files for each data set. The algorithm used in this option is slightly different to the first option as the proportion of differentially expressed genes on the array is unknown and the expected proportion is calculated based on the relative proportion of genes in InnateDB (for a given species). The benefit of this option is that it tends to be less conservative than the first option and it is often easier to interpret one data set at a time.
WARNING: If you try to analyze a subset of genes using the entire data set algorithm or vice versa your results will not be correct.

4. As discussed above there are several statistical methods and multiple testing correction methods that one may choose for the analysis. You may change these options using the dropdown menus at the bottom of the screen (Fig. 22.5), however, we recommend using the default selections for optimal performance. Click *Submit* to commence the analysis.
5. The results will be displayed as a table (Fig. 22.6) ordered by the GO over-representation *p*-value, with the most significant GO term associated with the data set at the top of the table and other GO terms beneath it. The *GO Term Name* column contains the GO term itself, followed by its classification in brackets – molecular function, biological process, or cellular compartment. #



GO Term ID	Source Database	Organism	GO Term Name	# Uploaded Genes Associated with GO Term	# Genes Associated with GO Term in InnateDB	Proportion	GO Term ORA P-Value	GO Term ORA P-Value (Corrected)	GO Term Summary
GO:0006955	GO	Homo Sapiens	immune response [biological_process]	74	547	14%	<1.0E-5	<1.0E-5	Summary
GO:0006954	GO	Homo Sapiens	inflammatory response [biological_process]	38	251	15%	<1.0E-5	<1.0E-5	Summary
GO:0009615	GO	Homo Sapiens	response to virus [biological_process]	19	86	22%	<1.0E-5	<1.0E-5	Summary
GO:0005615	GO	Homo Sapiens	extracellular space [cellular_component]	43	556	8%	<1.0E-5	<1.0E-5	Summary
GO:0005576	GO	Homo Sapiens	extracellular region [cellular_component]	92	1945	5%	<1.0E-5	<1.0E-5	Summary
GO:0007166	GO	Homo Sapiens	cell surface receptor linked signal transduction [biological_process]	20	165	12%	<1.0E-5	<1.0E-5	Summary
GO:0006935	GO	Homo Sapiens	chemotaxis [biological_process]	20	173	12%	<1.0E-5	<1.0E-5	Summary
GO:0005887	GO	Homo Sapiens	integral to plasma membrane [cellular_component]	57	1006	6%	<1.0E-5	<1.0E-5	Summary

Fig. 22.6 Gene Ontology ORA results. InnateDB results are displayed in tabular format. The Display Options panel allows the columns of table to be customized, while sorting of columns in either ascending or descending order is permitted. The results table can be downloaded in one of several formats using the download buttons. GO terms are initially sorted in ascending order of *p*-value, and *red arrows* indicate the statistically significant enrichment of a specific GO term

Uploaded Genes Associated with GO Term denotes how many genes in the submitted list are annotated with that GO term, while # *Genes Associated with GO Term in InnateDB* denotes how many genes in InnateDB (either human or mouse, depending on the data you have uploaded) are annotated with that term. *Proportion* is the former count divided by the latter. *GO Term ORA P-Value* reflects the likelihood of seeing that GO term enriched in a gene list of that size by chance alone, while *GO Term ORA P-Value (Corrected)* is this value after correction for multiple testing. *P*-values denoted by a red arrow provide a quick visual reference as to which GO terms are significantly more enriched in the uploaded data set than expected by chance. Clicking on *Summary* will provide a definition of the GO term as well as a list of all genes in the gene list associated with that term and all genes in InnateDB associated with that term.

6. Any results table in InnateDB can be redrawn or reordered according to a user's needs using the *Display Options* and *Sorted By* controls. Using the *Display Options* controls, set InnateDB to display 200 rows. Scrolling through the updated results table, you will notice that in our sample data set more than 200 GO terms are over-represented when uncorrected *p*-values are considered, while approximately half as many are significantly enriched when corrected

p-values are used instead. As with all such statistical analyses, the InnateDB ORA tools provide a guide as to which GO terms or pathway annotations are statistically over-represented – critical evaluation is needed to decide which results are biologically significant. Correction for multiple testing can be very conservative in some cases, with some terms potentially being of biological significance without being statistically significant. Conversely, some terms are found to be statistically significant but are unlikely to be of biological relevance since many genes/proteins are associated with several GO terms.

7. Complete results tables can be downloaded in a variety of formats. Select *MS Excel, TAB,* or *CSV* to open the data in one of these three formats suitable for a spreadsheet. You may wish to edit this file to remove terms with a *p*-value of > 0.05. In our case, deleting any corrected *p*-value greater than 0.05 leaves us with 72 over-represented GO terms (note that this number may change as InnateDB is continually updated).

Examining the results of our GO ORA reveals that the functions and processes associated with early smallpox response genes are heavily skewed toward the innate immune response, while cellular compartment annotations are primarily extracellular, consistent with a scenario in which chemokines and cytokines are being released in response to the virus and activating key innate immune signaling pathways. Interferon (IFN), interleukin-1 (IL-1), and immunoglobulin all appear to play important roles in the host response to smallpox, and we also observe the presence of terms indicating cellular activation, proliferation, and differentiation. Note that the purpose of this chapter is not to provide an in-depth re-analysis of this smallpox data set but to provide a guide as to how one might perform such analysis using InnateDB.

22.4.5 Performing a Pathway ORA

Pathways are the biochemical engines for transducing signals (often received by receptors) into output responses (e.g., activation of a transcription factor and downstream gene expression). The principles underlying a pathway ORA are identical to those of the GO ORA – pathway annotations associated with genes in the gene list are identified, and those that occur more often in the gene list than would be expected by chance are identified. The pathway ORA tends to provide a more focused picture of processes differentially regulated in a condition of interest than the GO ORA and is often the second step in analysis of microarray data. InnateDB automatically tests for over-representation of differentially expressed genes in the more than 3,000 pathways that are collectively annotated in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al. 2007), the NCI-Nature Pathway Interaction Database (PID) (http://pid.nci.nih.gov), the Integrating Network Objects with Hierarchies (INOH) pathway database

(http://www.inoh.org/), and the NetPath (http://www.netpath.org) and Reactome databases (Joshi-Tope et al. 2005).

For pathway over-representation analyses, it is worth including both up- and down-regulated genes in each analysis, as both can be informative in the context of a signaling pathway that may contain both positive and negative regulations. For example, up-regulation of the components of a pathway and down-regulation of the pathway's negative regulators are both equally informative. In our sample analysis, we will consider both time points together.

1. To perform a pathway ORA, select *Return a list of pathways* from the Data Analysis upload tool. You will be taken to the pathway results page, listing each of the pathways that the genes you have uploaded participate in. Certain genes may not be annotated as belonging to any pathways, while many genes belong to multiple pathways.
2. From the pathway results page, select the *Pathway ORA* button. As in the GO ORA, select whether the gene list represents a complete microarray or a subset of genes of interest.
3. The results are displayed in a similar format to those of the GO ORA (Fig. 22.7), with the pathway name, numbers of uploaded and total number of genes

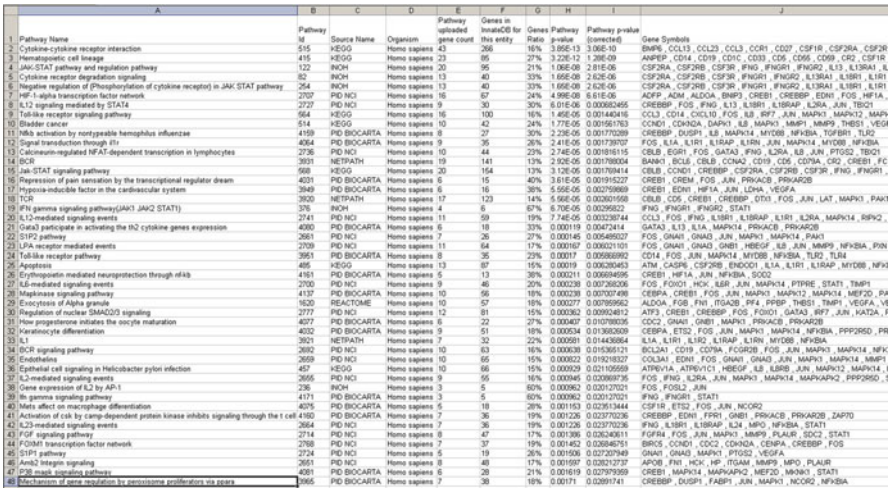
Pathway Over-Representation Analysis Results

Page(s): 1 2 3 4 5 6 7 8 9 10 [Next]

InnateDB Pathway ID	Source Database	Organism	Pathway Name	# Uploaded Genes Annotated in Pathway	# Genes Annotated in Pathway in InnateDB	Proportion	Pathway ORA P-Value	Pathway ORA P-Value (Corrected)	Pathway Summary	Cerebral
515	KEGG	Homo sapiens	Cytokine-cytokine receptor interaction	43	266	16%	<1.0E-5	<1.0E-5	Summary	Cerebral
415	KEGG	Homo sapiens	Hematopoietic cell lineage	23	85	27%	<1.0E-5	<1.0E-5	Summary	Cerebral
122	INOH	Homo sapiens	JAK-STAT pathway and regulation pathway	20	95	21%	<1.0E-5	<1.0E-5	Summary	Cerebral
82	INOH	Homo sapiens	Cytokine receptor degradation signaling	13	40	33%	<1.0E-5	<1.0E-5	Summary	Cerebral
254	INOH	Homo sapiens	Negative regulation of (Phosphorylation of cytokine receptor) in JAK STAT pathway	13	40	33%	<1.0E-5	<1.0E-5	Summary	Cerebral
2707	PID NCI	Homo sapiens	HIF-1-alpha transcription factor network	16	67	24%	<1.0E-5	<1.0E-5	Summary	Cerebral
2727	PID NCI	Homo sapiens	IL12 signaling mediated by STAT4	9	30	30%	<1.0E-5	0.00068	Summary	Cerebral
564	KEGG	Homo sapiens	Toll-like receptor signaling pathway	16	100	16%	0.00001	0.00144	Summary	Cerebral
514	KEGG	Homo sapiens	Bladder cancer	10	42	24%	0.00002	0.00156	Summary	Cerebral
4159	PID BIOCARTA	Homo sapiens	Nfkb activation by nontypeable hemophilus influenzae	8	27	30%	0.00002	0.00177	Summary	Cerebral
4064	PID BIOCARTA	Homo sapiens	Signal transduction through il1r	9	35	26%	0.00002	0.00174	Summary	Cerebral
2736	PID NCI	Homo sapiens	Calcineurin-regulated NFAT-dependent transcription in lymphocytes	10	44	23%	0.00003	0.00182	Summary	Cerebral

Fig. 22.7 Pathway ORA results. Similar to GO ORA results, pathway ORA results are also displayed in a table. In this case, however, each pathway contains a link to the Cerebral visualization tool, which launches a viewer that will draw the pathway in question and color it with any uploaded quantitative data

belonging to that pathway, and uncorrected and corrected *p*-values all shown. Clicking on *Summary* provides a description of the pathway and the database from which it was sourced, as well as an overview of the genes from your list belonging to that pathway versus all genes in InnateDB belonging to that pathway. Download your results in any of the three available formats and view the resulting spreadsheet (Fig. 22.8).



	A	B	C	D	E	F	G	H	I	J
1	Pathway Name	Pathway Id	Source Name	Organism	Pathway uploaded gene count	Genes in InnateDB for this entity	Genes Ratio	Pathway p-value	Pathway p-value (corrected)	Gene Symbols
2	Cytokine-cytokine receptor interaction	515	KEGG	Homo sapiens	43	266	16%	3.85E-13	3.06E-10	BMP6 , CCL13 , CCL23 , CCL3 , CCR1 , CD27 , CSF1R , CSF2RA , CSF2R
3	Hematopoietic cell lineage	415	KEGG	Homo sapiens	23	85	27%	3.22E-12	1.28E-09	ANPEP , CD14 , CD19 , CD1C , CD33 , CD5 , CD55 , CD59 , CR2 , CSF1R
4	JAK-STAT signaling and regulation pathway	122	INOH	Homo sapiens	20	95	21%	1.06E-09	2.81E-06	CSF2RA , CSF2RB , CSF3R , IFNG , IFNGR1 , IFNGR2 , IL13 , IL13RA1 , IL
5	Cytokine receptor degradation signaling	82	INOH	Homo sapiens	13	40	33%	1.65E-08	2.62E-06	CSF2RA , CSF2RB , CSF3R , IFNGR1 , IFNGR2 , IL13RA1 , IL15RA , IL1R1
6	Negative regulation of (Phosphorylation of cytokine receptor) in JAK STAT pathway	264	INOH	Homo sapiens	13	40	33%	1.65E-08	2.62E-06	CSF2RA , CSF2RB , CSF3R , IFNGR1 , IFNGR2 , IL13RA1 , IL15RA , IL1R1
7	HIF-1-alpha transcription factor network	2707	PID NCI	Homo sapiens	16	67	24%	4.99E-08	6.61E-06	ADFP , ADM , ALDOA , BNIP3 , CREB1 , CREBBP , EDN1 , FOS , HIF1A ,
8	IL12 signaling mediated by STAT4	2727	PID NCI	Homo sapiens	9	30	30%	6.01E-06	0.000682455	CREBBP , FOS , IFNG , IL13 , IL18R1 , IL18RAP , IL2RA , JUN , TBX21
9	Toll-like receptor signaling pathway	564	KEGG	Homo sapiens	16	100	16%	1.45E-05	0.001440416	CCL3 , CD14 , CXCL10 , FOS , IL8 , IRF7 , JUN , MAPK1 , MAPK12 , MAP
10	Bladder cancer	514	KEGG	Homo sapiens	10	42	24%	1.77E-05	0.001561763	CCND1 , CDKN2A , DAPK1 , IL8 , MAPK1 , MMP1 , MMP9 , THBS1 , VEGI
11	NfkB activation by nontypeable hemophilus influenzae	4159	PID BIOCARTA	Homo sapiens	8	27	30%	2.23E-05	0.001770289	CREBBP , DUSP1 , IL8 , MAPK14 , MYD88 , NFKBIA , TGFBR1 , TLR2
12	Signal transduction through il1r	4064	PID BIOCARTA	Homo sapiens	9	35	26%	2.41E-05	0.001739707	FOS , IL1A , IL1R1 , IL1RAP , IL1RN , JUN , MAPK14 , MYD88 , NFKBIA
13	Calcineurin-regulated NFAT-dependent transcription in lymphocytes	2736	PID NCI	Homo sapiens	10	44	23%	2.74E-05	0.001916116	CBLB , EGR1 , FOS , GATA3 , IFNG , IL2RA , IL8 , JUN , PTGS2 , TBX21
14	BCR	3931	NETPATH	Homo sapiens	19	141	13%	2.92E-05	0.001788004	BAMH1 , BCL6 , CBLB , CCNA2 , CD19 , CD5 , CD79A , CR2 , CREB1 , FC
15	Jak-STAT signaling pathway	568	KEGG	Homo sapiens	20	154	13%	3.12E-05	0.001709414	CBLB , CCND1 , CREBBP , CSF2RA , CSF2RB , CSF3R , IFNG , IFNGR1 ,
16	Repression of pain sensation by the transcriptional regulator dream	4031	PID BIOCARTA	Homo sapiens	6	15	40%	3.61E-05	0.001915227	CREB1 , CREM , FOS , JUN , PRKACB , PRKAR2B
17	Hypoxia-inducible factor in the cardiovascular system	3949	PID BIOCARTA	Homo sapiens	6	16	38%	5.56E-05	0.002598669	CREB1 , EDN1 , HIF1A , JUN , LDHA , VEGFA
18	TCR	3920	NETPATH	Homo sapiens	17	123	14%	5.56E-05	0.002601558	CBLB , CD5 , CREB1 , CREBBP , DTX1 , FOS , JUN , LAT , MAPK1 , PAK1
19	IFN gamma signaling pathway(JAK1 JAK2 STAT1)	376	INOH	Homo sapiens	4	6	67%	6.70E-05	0.00299822	IFNG , IFNGR1 , IFNGR2 , STAT1
20	IL12-mediated signaling events	2741	PID NCI	Homo sapiens	11	59	19%	7.74E-05	0.003238744	CCL3 , FOS , IFNG , IL18R1 , IL18RAP , IL1R1 , IL2RA , MAPK14 , RPK2 ,
21	Gata3 participate in activating the th2 cytokine genes expression	4080	PID BIOCARTA	Homo sapiens	6	18	33%	0.000119	0.00472414	GATA3 , IL13 , IL1A , MAPK14 , PRKACB , PRKAR2B
22	S1P2 pathway	2681	PID NCI	Homo sapiens	7	26	27%	0.000145	0.005485027	FOS , GNAI1 , GNAI2 , JUN , MAPK1 , MAPK14 , PAK1
23	LPA receptor mediated events	2709	PID NCI	Homo sapiens	11	64	17%	0.000167	0.006021101	FOS , GNAI1 , GNAI2 , GNB1 , HBEGF , IL8 , JUN , MMP9 , NFKBIA , PXN
24	Toll-like receptor pathway	3951	PID BIOCARTA	Homo sapiens	8	35	23%	0.00017	0.005866992	CD14 , FOS , JUN , MAPK14 , MYD88 , NFKBIA , TLR2 , TLR4
25	Apoptosis	486	KEGG	Homo sapiens	13	87	15%	0.00019	0.006280453	ATM , CASP6 , CSF2RB , ENDOD1 , IL1A , IL1R1 , IL1RAP , MYD88 , NFKI
26	Erythropoietin mediated neuroprotection through nf-kb	4161	PID BIOCARTA	Homo sapiens	6	13	36%	0.000211	0.006994095	CREB1 , HIF1A , JUN , NFKBIA , SOD2 , VEGFA
27	IL6-mediated signaling events	2700	PID NCI	Homo sapiens	9	46	20%	0.000238	0.007268206	FOS , FOXO1 , HCK , IL6R , JUN , MAPK14 , PTPRE , STAT1 , TIMP1
28	Mapkinase signaling pathway	4137	PID BIOCARTA	Homo sapiens	10	56	18%	0.000238	0.007007498	CEBPA , CREB1 , FOS , JUN , MAPK1 , MAPK12 , MAPK14 , MEF2D , PA
29	Exocytosis of Alpha granule	1620	REACTOME	Homo sapiens	10	57	18%	0.000277	0.007595952	ALDOA , FGB , FN1 , ITGA2B , PF4 , PPBP , THBS1 , TIMP1 , VEGFA , VI
30	Regulation of nuclear SMAD2/3 signaling	2777	PID NCI	Homo sapiens	12	81	15%	0.000362	0.009924812	ATF3 , CREB1 , CREBBP , FOS , FOXO1 , GATA3 , IRF7 , JUN , KAT2A ,
31	How progesterone initiates the oocyte maturation	4077	PID BIOCARTA	Homo sapiens	6	22	27%	0.000407	0.010788035	CDC2 , GNAI1 , GNB1 , MAPK1 , PRKACB , PRKAR2B
32	Keratinocyte differentiation	4032	PID BIOCARTA	Homo sapiens	9	51	18%	0.000534	0.013682609	CEBPA , ETS2 , FOS , JUN , MAPK1 , MAPK14 , NFKBIA , PPP2R5D , PR
33	IL1	3921	NETPATH	Homo sapiens	7	32	22%	0.000581	0.014436864	IL1A , IL1R1 , IL1R2 , IL1RAP , IL1RN , MYD88 , NFKBIA
34	BCR signaling pathway	2692	PID NCI	Homo sapiens	10	63	16%	0.000639	0.015365121	BCL2A1 , CD19 , CD79A , FCGR2B , FOS , JUN , MAPK3 , MAPK14 , NFK
35	Endothelins	2659	PID NCI	Homo sapiens	10	65	15%	0.000822	0.019218327	COL3A1 , EDN1 , FOS , GNAI1 , GNAI3 , JUN , MAPK1 , MAPK14 , MMP1
36	Epithelial cell signaling in Helicobacter pylori infection	457	KEGG	Homo sapiens	10	66	15%	0.000929	0.021105559	ATP6V1A , ATP6V1C1 , HBEGF , IL8 , ILBRB , JUN , MAPK12 , MAPK14 ,
37	IL2-mediated signaling events	2655	PID NCI	Homo sapiens	9	55	16%	0.000945	0.020899735	FOS , IFNG , IL2RA , JUN , MAPK3 , MAPK14 , MAPKAPK2 , PPP2R5D , S
38	Gene expression of IL2 by AP-1	236	INOH	Homo sapiens	3	5	60%	0.000962	0.020127021	FOS , FOSL2 , JUN
39	Ifn gamma signaling pathway	4171	PID BIOCARTA	Homo sapiens	3	5	60%	0.000962	0.020127021	IFNG , IFNGR1 , STAT1
40	Mets affect on macrophage differentiation	4075	PID BIOCARTA	Homo sapiens	5	18	28%	0.001153	0.023513444	CSF1R , ETS2 , FOS , JUN , NCOR2
41	Activation of csk by camp-dependent protein kinase inhibits signaling through the t cell	4160	PID BIOCARTA	Homo sapiens	7	36	19%	0.001226	0.023770236	CREBBP , EDN1 , FPR1 , GNB1 , PRKACB , PRKAR2B , ZAP70
42	IL23-mediated signaling events	2664	PID NCI	Homo sapiens	7	36	19%	0.001226	0.023770236	IFNG , IL18R1 , IL18RAP , IL24 , MPO , NFKBIA , STAT1
43	FGF signaling pathway	2714	PID NCI	Homo sapiens	8	47	17%	0.001386	0.026240611	FGFR4 , FOS , JUN , MAPK1 , MMP9 , PLAUR , SDC2 , STAT1
44	FOXM1 transcription factor network	2768	PID NCI	Homo sapiens	7	37	19%	0.001452	0.026946751	BIRC5 , CCND1 , CDC2 , CDKN2A , CENPA , CREBBP , FOS
45	S1P1 pathway	2724	PID NCI	Homo sapiens	5	19	26%	0.001506	0.027207949	GNAI1 , GNAI3 , MAPK1 , PTGS2 , VEGFA
46	Amb2 Integrin signaling	2651	PID NCI	Homo sapiens	8	48	17%	0.001697	0.028212737	APOB , FN1 , HCK , HP , ITGAM , MMP9 , MPO , PLAUR
47	P38 mapk signaling pathway	4081	PID BIOCARTA	Homo sapiens	6	28	21%	0.001619	0.027979369	CREB1 , MAPK14 , MAPKAPK2 , MEF2D , MKNK1 , STAT1
48	Mechanism of gene regulation by peroxisome proliferators via ppara	3965	PID BIOCARTA	Homo sapiens	7	38	18%	0.00171	0.02891741	CREBBP , DUSP1 , FABP1 , JUN , MAPK1 , NCOR2 , NFKBIA

Fig. 22.8 A spreadsheet downloaded from InnateDB containing pathway ORA results. Downloading the results of a pathway ORA is often useful, as it enables a user to assess which pathways have been duplicated in their results. Duplicate pathways can be removed from the results or be amalgamated into a larger pathway containing all of the known components of a given pathway

4. Because InnateDB sources its pathway data from multiple databases, each with its own interpretations of the components of a given pathway, you will observe some degree of duplication in the results; however, this is outweighed by the extra annotation that can be obtained from different data sources. In our sample results, for instance, the interleukin-1 pathway is seen four times: three records are from the INOH database and distinguish IL-1 signaling through three distinct pathways (p38, I-Kappa-B Kinase (IKK)-NFκB, and c-Jun N-terminal kinase (JNK)), while one result from the PID NCI database contains over twice as many pathway components as the INOH records. We recommend inspecting one's results and amalgamating the data from duplicate pathways into a single representative pathway containing as many members as possible.

   In the pathway ORA results for the smallpox data set, the predominance of innate immune processes was again noted; however one now has a more specific idea of pathways that are being affected. The involvement of the Janus kinase (JAK) – signal transducer and activator of transcription (STAT) and osteopontin pathways, for example, point toward a role for interferon-alpha signaling in the

response to smallpox, while interferon-gamma is also implicated. Both T cell and B-cell signaling processes are enriched, as are traditional innate immune signaling pathways, such as TLR signaling, the p38 mitogen-activated protein kinase (MAPK) pathway, and NFκB.

22.4.6 Visualizing Pathway Data with Cerebral

When a pathway of interest has been identified, it is often preferable to view that pathway in graphical format rather than as a table showing the individual components of the pathway. Previously, this would have required downloading the interactions between the pathway members, formatting them for input to interaction visualization software, and then manually adjusting the resulting diagram to create a more biologically relevant pathway view. With InnateDB, however, the above steps can be completed in a single click, even for a user unfamiliar with interaction visualization. Furthermore, InnateDB allows users to quickly "paint" a pathway diagram with quantitative data, such as the fold changes from our microarray experiment.

Network and pathway visualization in InnateDB is carried out using Cerebral (Cell Region-Based Rendering And Layout) (Barsky et al. 2007). Cerebral is a plug-in for the Cytoscape biomolecular interaction viewer (Shannon et al. 2003) that draws a biological network in a more pathway-like layout. The screen is divided into layers, one for each cellular compartment, and the nodes representing genes and proteins are confined to their associated layer. Cerebral is also designed to facilitate the simultaneous comparison of quantitative data from multiple conditions, such as microarray data from different time points. Furthermore, as a Cytoscape plug-in, Cerebral allows users to access the powerful functions associated with this program. Discussion of Cytoscape's features is beyond the scope of this chapter. We suggest the reader visit the Cytoscape Web site (http://www.cytoscape.org/) to familiarize themselves with this important tool and work through the many online tutorials that are available.

1. In the pathway ORA results, locate "IFN gamma signaling pathway (JAK1 JAK2 STAT1)" from INOH (Pathway ID 376). In this pathway, four of the six pathway components are differentially expressed in our gene list. Click the *Cerebral* button to launch a Java Web start instance of Cytoscape with the Cerebral plug-in installed. It is not necessary to have Cytoscape installed on your computer, as the Web start will download and install a temporary instance of the program. A recent version of Java must be installed on your computer.
2. Cerebral will load the requested pathway and lay it out according to the annotated localization layer for each gene/protein in the pathway (Fig. 22.9). Nodes are placed in a layer according to the localization inferred from Gene Ontology (Ashburner et al. 2000) or InnateDB curation of subcellular localization. For display purposes, nodes with multiple potential localizations will be assigned a single preferential localization; however, all annotated localizations will be displayed in the *Node Attribute Browser.*

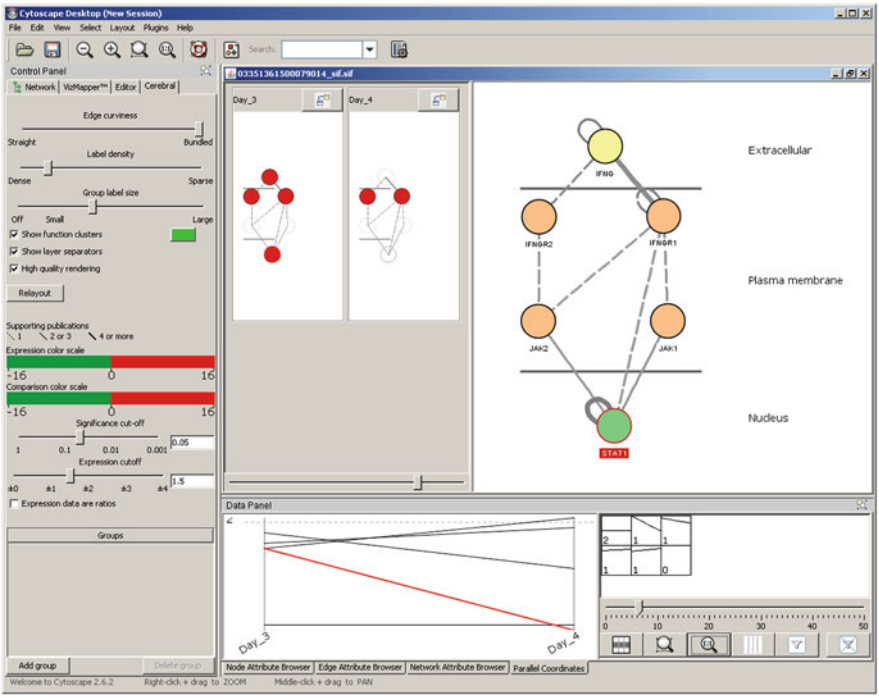


Fig. 22.9 The over-represented interferon-gamma pathway visualized in Cytoscape with the Cerebral plugin. The larger network view to the right shows the pathway laid out according to subcellular localization information derived from InnateDB. To the left of the main view, one mini-view for each gene expression condition (two, in this case) colors the network according to the quantitative data provided. Sliders and coloring schemes in the Control Panel area allow the user to adjust the look and feel of their diagram, while a profile view beneath the main view displays the quantitative data as a series of profiles. Here, the STAT1 node is selected in the main view, and its expression profile is highlighted in the profile view. Edge widths indicate the number of publications in InnateDB supporting a particular interaction – the more publications the thicker the line. Dashed edges have only one supporting publication

3. In the main display window, a large view of the pathway is shown with nodes colored according to their subcellular localization as provided from InnateDB. To improve image quality, select the *High quality rendering* button from the Cerebral panel at left. To the left are two smaller views – one for each time point of expression data we have provided. In these views, the nodes are colored according to their expression value, and the coloring scheme can be changed by clicking on the *Expression* and *Comparison color scale* tabs in the Cerebral panel. To promote the coloring of a mini-view into the main view, simply click anywhere in the mini-view (clicking on the mini-view's highlighted names will permit a return to default coloring). Note that the mini-views and main view are linked – zooming, highlighting, and selecting in one view will propagate that action across all views. The bottom panel shows the profile view – each

gene's expression profile is represented as a line, and a *k*-means clustering tool at the right-hand side allows users to quickly cluster genes with similar expression profiles.

4. Cerebral views can be exported as graphics suitable for inclusion in a presentation or publication by choosing *Plugins* > *Create Cerebral View*, while the complete Cytoscape session can be saved by choosing *File* > *Save As*. If Cytoscape is installed locally, these session files can be accessed anytime, allowing one to easily return to a given analysis.

By examining data overlaid onto individual pathways of interest using Cerebral, trends in the data that were not obvious from the table-format results alone can be readily observed. In this example, for instance, we see that while the expression of the interferon-gamma receptors increases slightly from day 3 to day 4, expression of the interferon-gamma ligand returns to baseline levels on day 4, as do levels of the STAT1 transcription factor, indicating a possible abatement of the interferon-gamma response after an early peak post-infection.

22.4.7 Generating and Exploring Molecular Interaction Networks Using InnateDB

InnateDB pathway and Gene Ontology analyses can be very powerful in determining which annotated pathways and biological processes are significantly associated with a data set of genes. Such analyses, however, rely on using the association of genes to known biological pathways or Gene Ontology terms. Annotation of pathways and Gene Ontology terms is far from complete and pathways are often annotated as relatively simple linear cascades. Network analysis has the ability to move the investigation from this simple view of the signaling response to a more comprehensive analysis of the molecular interactions between genes of interest and their encoded proteins and RNAs, potentially allowing one to uncover as yet unknown signaling cascades or pathways, functionally relevant sub-networks and the central molecules, or hubs, of these networks.

InnateDB is one of the most comprehensive databases of all human and mouse experimentally supported molecular interactions (~130,000) but also specifically includes annotation on more than 12,900 manually curated human and mouse innate immunity-relevant interactions, many of which are not present in any other database. InnateDB allows one to upload a gene list of interest along with associated gene expression data and returns this data integrated in a molecular interaction network context for visualization and further interrogation and analysis.

1. Return to the *Data Analysis* page and this time, select *Return a list of interactions*. This will bring up the interaction filtering dialog box. Three options are available. *Do not filter the results* will display all of the interactions that all of the uploaded genes participate in. By investigating networks, such as this,

that include interactions between differentially expressed genes and their non-differentially expressed interacting partners, one has the potential to identify key regulators of gene expression, even though these regulators themselves may not be differentially expressed but may be regulated at the post-transcriptional level.

For a gene list with hundreds of entries, however, this network can consist of several thousand interactions (Fig. 22.10). For this reason, with large gene lists it is often preferable to first create a more focused network in which only interactions between the genes in the uploaded gene list will be shown (for example, differentially expressed genes only). This filtering – *Only show interactions between uploaded molecules* – is demonstrated below. The third filtering option, *Filter for interactions in pathway*, provides an even more focused view, allowing one to display only those interactions that comprise a given pathway.

1. Select *Only show interactions between uploaded molecules* and execute the search for interactions between uploaded molecules. As in our earlier analyses,

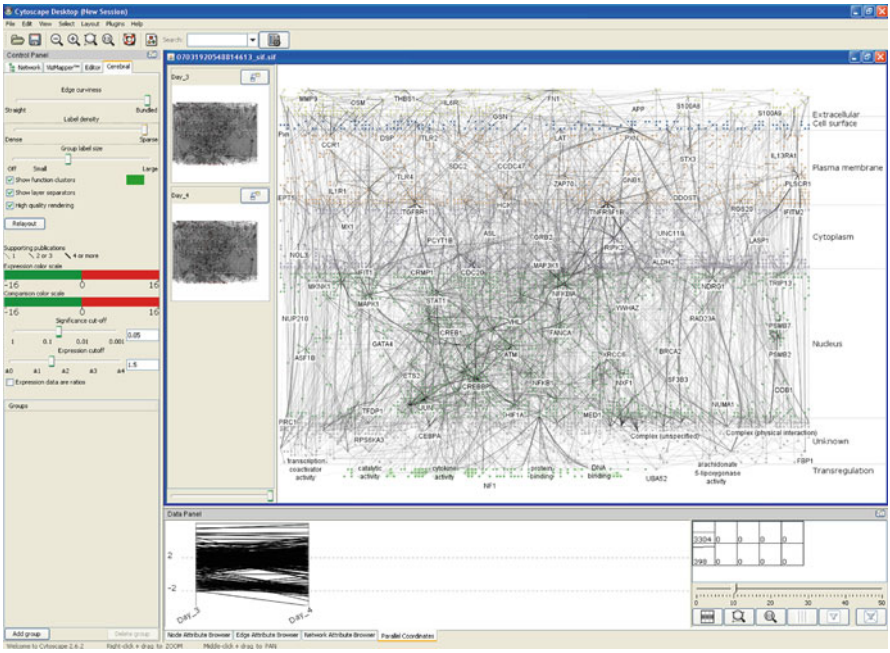


Fig. 22.10 Network of interactions between differentially expressed genes (and their encoded products) at day 3 and/or day 4 and all known interacting partners in InnateDB. The network was displayed in Cytoscape using the Cerebral plugin launched from InnateDB. Nodes encoded by up-regulated genes are shown in red, down-regulated in green. Analysis of this network enables the identification of central regulators (hubs/bottlenecks that are not necessarily regulated at the transcriptional level)

the search returns its results in a table format which can be edited, sorted and/or downloaded.

2. To visualize these retrieved interactions, click on the *Cerebral* button at the top of the page. A Cerebral view is launched in Cytoscape showing all of the interactions (Fig. 22.11).

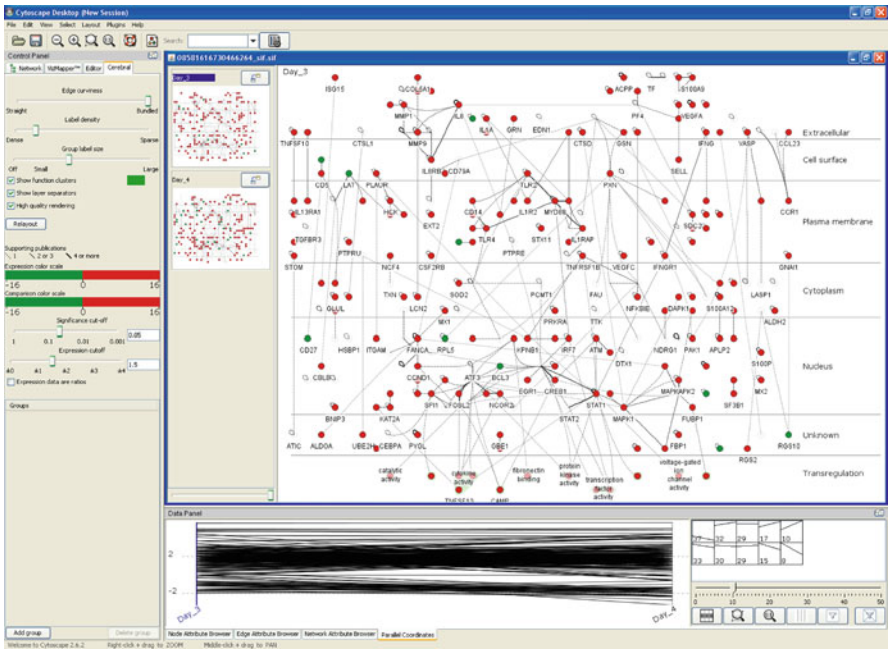


Fig. 22.11 A network of molecular interactions only between genes (and their encoded products) which were differentially expressed at day 3 and/or day 4. Interactions involving molecules that were not differentially expressed are not shown. The network was displayed in Cytoscape using the Cerebral plugin. Up-regulated genes at day 3 are shown in red and down-regulated genes in green. Un-shaded nodes were not differentially regulated at day 3. This network is useful to investigate molecular interactions between molecules encoded by differentially expressed genes

3. In this interaction-based analysis, it is often worthwhile to lay the data out in different formats for an alternative perspective, particularly when several hundred interactions are displayed. In the *Cytoscape Control Panel*, select the *Network* tab. The name of the network – a string of numbers automatically generated by InnateDB – will appear highlighted in green. Right-click this name and select *Destroy View*, then right-click the name again and select *Create View*. This will redraw the network in Cytoscape's default grid format.

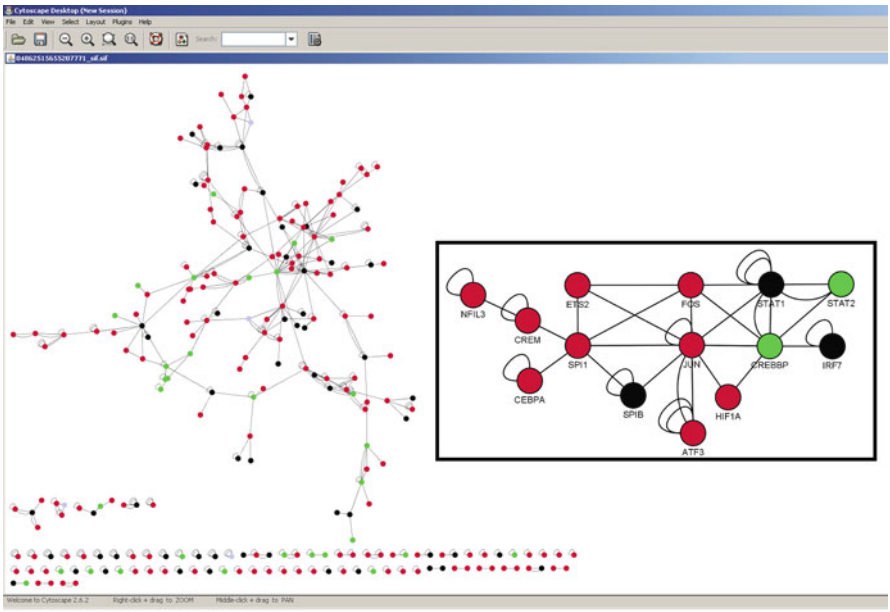


Fig. 22.12 An alternative layout of the network in Fig. 22.11. In this view, we have used one of Cytoscape's native layouts to visualize the relationships between the genes in our list, as the network structure is more apparent with this layout than with the Cerebral layout. The larger network shows all of the interactions between our genes of interest colored according to their expression level on day 4, while the inset view shows a network of transcriptional regulators extracted from the larger network. This type of analysis, which does not rely on pre-existing information such as GO or pathway annotation, can reveal novel processes, functions, or complexes active in a data set

4. From the *Cytoscape Layout* menu, select *yFiles* > *Organic* or any of the other available layout options. This will redraw the network in an alternative manner (Fig. 22.12).

At this point, you may wish to analyze the network using other tools and approaches that are presented in this book. To do this one may export the network and its attributes in a number of formats. From Cytoscape select *File* > *Export*. By examining the relationships between the nodes of the network, new insight into particular processes or protein complexes can be gained. As an example, the inset of Fig. 22.12 shows a network of transcription regulators extracted from the larger network in Fig. 22.12 and colored according to their expression at day 4. A number of transcription factors not identified through GO or pathway analysis are observed to be active, and the user may wish to follow up on this analysis by examining whether genes regulated by these transcription factors are enriched at subsequent time points in the experiment.

22.5 Conclusions and Future Directions

Systems biology approaches to investigating the innate immune response are beginning to provide novel insight and new understanding of the early host response to infectious disease. As we have shown, InnateDB greatly facilitates the interpretation of large-scale -omics data by allowing users to carry out a range of analyses on their data with just a few clicks of the mouse (Lynn et al. 2008). Uploading of data is simple, requiring only a spreadsheet containing the genes of interest, and from the Data Analysis page users can access tools ranging from network construction and visualization to powerful over-representation analyses. Display options customization and multiple download formats enable users to retrieve and store their data in the format of their choice, while visualizations through the Cytoscape/Cerebral tools (Shannon et al. 2003; Barsky et al. 2007; Barsky et al. 2008) allow for more intuitive approaches for analyzing data. Thus, in only a few steps a user can begin to interpret a gene list and generate specific testable biological hypotheses for follow-up.

Many challenges, however, remain. Despite the large number of interactions currently annotated in InnateDB and other databases, it is estimated that only approximately 15% of the human interactome is currently known (Bader et al. 2008). In addition, almost all of these interactions are protein–protein interactions with only a small fraction of potential transcription factor-DNA interactions currently experimentally validated and even fewer RNA interactions currently known. Fortunately, ChIP-chip methods are now enabling large-sale identification of transcription factor-DNA interactions (Ramsey et al. 2008) and new array platforms are allowing genome-wide profiling of microRNA expression.

Currently, however, the incomplete nature of the networks used for systems biology-oriented analyses undoubtedly means that important connections between signaling proteins and pathways are being missed. InnateDB curation efforts, whereby we have curated nearly 13,000 interactions of relevance to innate immunity, have assisted in providing a more complete picture of the innate immunity interactome based on data available in the biomedical literature. Large-scale interactome mapping efforts are essential however, to ensure that novel molecular interactions continue to be described to fill in the missing gaps in the interactome.

Another important issue moving forward is that although we are becoming closer to determining the entire human interactome, the interactome is not a static entity. The interactions that occur at any given time depend on the genes being expressed, post-translational modifications, the cell type or tissue type, exogenous and endogenous stimuli and the particular conditions being investigated. The interactome is thus a dynamic entity changing over time. Fortunately, gene and protein array technologies can assist us in determining which particular networks of interactions are likely most relevant to a given response, by providing quantitative data that can be analyzed and interpreted in the framework of the interactome. More detailed investigation and annotation of the context of particular interactions, such as the cell type in which they occur, will greatly help in moving from a static view of the interactome.

Better appreciation and understanding of host–pathogen interactions also needs to be accounted for in systems-based approaches. The host responses to disease

and the signaling networks involved can be actively manipulated by pathogens. For example, live (but not dead) *Mycobacterium tuberculosis* interfere with signaling in macrophages (Ehrt et al. 2001), while several viruses produce microRNAs to specifically modulate the host response by suppressing components of the innate immune response to inhibit apoptosis and promote virus latency (Pedersen and David 2008). Similarly, host factors influence the pathogens; interferon-gamma expressed by the host, for example, is sensed by *Pseudomonas aeruginosa* and causes expression of virulence factors (Wu et al. 2005). Several new databases that specialize in host–pathogen interactions provide valuable supplemental information to InnateDB including the VirusMINT (mint.bio.uniroma2.it/virusmint/) (Chatr-aryamontri et al. 2009) and Pathogen Interaction Gateway (PIG, molvis.vbi.vt.edu/pig) (Driscoll et al. 2009).

Despite these and many other challenges, systems biology approaches are already providing significant new insight into innate immunity (see for review Gardy et al. 2009) and promise a far deeper understanding of our first line of defense against invading pathogens than previously possible.

Acknowledgments The authors' systems biology research 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). DJL and JLG hold Postdoctoral Trainee Awards from the Michael Smith Foundation for Health Research (MSFHR) and JLG also holds a Sanofi Pasteur CIHR fellowship. FSLB is a Canadian Institutes of Health Research (CIHR) New Investigator and a MSFHR Senior Scholar. REWH holds a Canada Research Chair (CRC).

References

Abbas AR, Baldwin D, Ma Y et al (2005) Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. Genes Immun 6:319–331

Akira S (2006) TLR signaling. Curr Top Microbiol Immunol 311:1–16

Alibes A, Yankilevich P, Canada A et al (2007) IDconverter and IDClight: conversion and annotation of gene and protein IDs. BMC Bioinformatics 8:9

Andersen J, VanScoy S, Cheng TF et al (2008) IRF-3-dependent and augmented target genes during viral infection. Genes Immun 9:168–175

Angus DC, Linde-Zwirble WT, Lidicker J et al (2001) Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. Crit Care Med 29: 1303–1310

Ashburner M, Ball CA, Blake JA et al (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25–29

Bader S, Kuhner S and Gavin AC (2008) Interaction networks for systems biology. FEBS Lett 582:1220–1224

Barsky A, Gardy JL, Hancock REW et al (2007) Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. Bioinformatics 23:1040–1042

Barsky A, Munzner T, Gardy J et al (2008) Cerebral: visualizing multiple experimental conditions on a graph with biological context. IEEE Trans Vis Comput Graph 14:1253–1260

Benjamini Y and Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc. Series B 57:289–300

Bertrand MJ, Doiron K, Labbe K et al (2009) Cellular inhibitors of apoptosis cIAP1 and cIAP2 are required for innate immunity signaling by the pattern recognition receptors NOD1 and NOD2. Immunity 30:789–801

Bhoj VG and Chen ZJ (2009) Ubiquitylation in innate and adaptive immunity. Nature 458:430–437

Bi Y, Liu G and Yang R (2009) MicroRNAs: novel regulators during the immune response. J Cell Physiol 218:467–472

Brownstein BH, Logvinenko T, Lederer JA et al (2006) Commonality and differences in leukocyte gene expression patterns among three models of inflammation and injury. Physiol Genomics 24:298–309

Chatr-aryamontri A, Ceol A, Peluso D et al (2009) VirusMINT: a viral protein interaction database. Nucleic Acids Res 37:D669–673

Chen XM, Splinter PL, O'Hara SP 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

Chuang T and Ulevitch RJ (2001) Identification of hTLR10: a novel human Toll-like receptor preferentially expressed in immune cells. Biochim Biophys Acta 1518:157–161

Chuang TH and Ulevitch RJ (2000) Cloning and characterization of a sub-family of human toll-like receptors: hTLR7, hTLR8 and hTLR9. Eur Cytokine Netw 11:372–378

Cohen J and Enserink M (2002) Public health. Rough-and-tumble behind Bush's smallpox policy. Science 298:2312–2316

Collas P and Dahl JA (2008) Chop it, ChIP it, check it: the current status of chromatin immunoprecipitation. Front Biosci 13:929–943

Dimitriou ID, Clemenza L, Scotter AJ et al (2008) Putting out the fire: coordinated suppression of the innate and adaptive immune systems by SOCS1 and SOCS3 proteins. Immunol Rev 224:265–283

Driscoll T, Dyer MD, Murali TM et al (2009) PIG—the pathogen interaction gateway. Nucleic Acids Res 37:D647–650

Ehrt S, Schnappinger D, Bekiranov 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

Gardy JL, Lynn DJ, Brinkman FS et al (2009) Enabling a systems biology approach to immunology: focus on innate immunity. Trends Immunol 30:249–262

Gilchrist M, Thorsson V, Li B et al (2006) Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4. Nature 441:173–178

Heng TS and Painter MW (2008) The Immunological Genome Project: networks of gene expression in immune cells. Nat Immunol 9:1091–1094

Hermjakob H, Montecchi-Palazzi L, Bader G et al (2004) The HUPO PSI's molecular interaction format—a community standard for the representation of protein interaction data. Nat Biotechnol 22:177–183

Hijikata A, Kitamura H, Kimura Y 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

Honda K and Taniguchi T (2006) IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. Nat Rev Immunol 6:644–658

Hsueh RC, Natarajan M, Fraser I et al (2009) Deciphering signaling outcomes from a system of complex networks. Sci Signal 2:ra22

Hubbard TJ, Aken BL, Ayling S et al (2009) Ensembl 2009. Nucleic Acids Res 37:D690–697

Inohara N and Nunez G (2001) The NOD: a signaling module that regulates apoptosis and host defense against pathogens. Oncogene 20:6473–6481

Joshi-Tope G, Gillespie M, Vastrik I et al (2005) Reactome: a knowledgebase of biological pathways. Nucleic Acids Res 33:D428–432

Kanehisa M, Araki M, Goto S et al (2007) KEGG for linking genomes to life and the environment. Nucleic Acids Res 36:D480–484

Kanneganti TD, Lamkanfi M and Nunez G (2007) Intracellular NOD-like receptors in host defense and disease. Immunity 27:549–559

Kolchanov NA, Merkulova TI, Ignatieva EV et al (2007) Combined experimental and computational approaches to study the regulatory elements in eukaryotic genes. Brief Bioinform 8:266–274

Korb M, Rust AG, Thorsson V et al (2008) The Innate Immune Database (IIDB). BMC Immunol 9:7

Langlais D, Couture C, Balsalobre A 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

Lee MS and Kim YJ (2007) Signaling pathways downstream of pattern-recognition receptors and their cross talk. Annu Rev Biochem 76:447–480

Litvak V, Ramsey SA, Rust AG 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

Lynn DJ, Winsor GL, Chan C et al (2008) InnateDB: facilitating systems-level analyses of the mammalian innate immune response. Mol Syst Biol 4:218

MacLeod H and Wetzler LM (2007) T cell activation by TLRs: a role for TLRs in the adaptive immune response. Sci STKE 2007:pe48

Maglott D, Ostell J, Pruitt KD et al (2005) Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res 33:D54–58

Manicassamy S and Pulendran B (2009) Modulation of adaptive immunity with Toll-like receptors. Semin Immunol. doi: 10.1016/j.smim.2009.05.005

Medzhitov R, Janeway CA Jr (1997) Innate immunity: the virtues of a nonclonal system of recognition. Cell 91:295–298

Medzhitov R, Preston-Hurlburt P and Janeway CA Jr (1997) A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. Nature 388:394–397

Mookherjee N, Hamill P, Gardy J et al (2009) Systems biology evaluation of immune responses induced by human host defence peptide LL-37 in mononuclear cells. Mol Biosyst 5:483–496

Nilsson R, Bajic VB, Suzuki H et al (2006) Transcriptional network dynamics in macrophage activation. Genomics 88:133–142

Oda K and Kitano H (2006) A comprehensive map of the toll-like receptor signaling network. Mol Syst Biol 2:2006 0015

Okabe Y, Sano T and Nagata S (2009) Regulation of the innate immune response by threonine-phosphatase of Eyes absent. Nature 460:520–524

Orchard S, Salwinski L, Kerrien S et al (2007) The minimum information required for reporting a molecular interaction experiment (MIMIx). Nat Biotechnol 25:894–898

Pedersen I and David M (2008) MicroRNAs in the immune response. Cytokine 43:391–394

Pruitt KD, Tatusova T, Klimke W et al (2009) NCBI Reference Sequences: current status, policy and new initiatives. Nucleic Acids Res 37:D32–36

Ramsey SA, Klemm SL, Zak DE et al (2008) Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics. PLoS Comput Biol 4:e1000021

Robertson G, Hirst M, Bainbridge M et al (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods 4:651–657

Rock FL, Hardiman G, Timans JC et al (1998) A family of human receptors structurally related to Drosophila Toll. Proc Natl Acad Sci USA 95:588–593

Rubins KH, Hensley LE, Jahrling PB et al (2004) The host response to smallpox: analysis of the gene expression program in peripheral blood cells in a nonhuman primate model. Proc Natl Acad Sci USA 101:15190–15195

Seet BT, Johnston JB, Brunetti CR et al (2003) Poxviruses and immune evasion. Annu Rev Immunol 21:377–423

Shannon P, Markiel A, Ozier O et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504

Smith B, Ashburner M, Rosse C et al (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol 25:1251–1255

Taganov KD, Boldin MP, Chang KJ et al (2006) NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. Proc Natl Acad Sci USA 103:12481–12486

Takeuchi O, Kawai T, Sanjo H et al (1999) TLR6: A novel member of an expanding toll-like receptor family. Gene 231:59–65

Tegner J, Nilsson R, Bajic VB et al (2006) Systems biology of innate immunity. Cell Immunol 244:105–109

The UniProt Consortium (2008) The universal protein resource (UniProt). Nucleic Acids Res 36:D190–195

Thompson AJ and Locarnini SA (2007) Toll-like receptors, RIG-I-like RNA helicases and the antiviral innate immune response. Immunol Cell Biol 85:435–445

von Bernuth H, Picard C, Jin Z et al (2008) Pyogenic bacterial infections in humans with MyD88 deficiency. Science 321:691–696

Wall EA, Zavzavadjian JR, Chang MS et al (2009) Suppression of LPS-induced TNF-alpha production in macrophages by cAMP is mediated by PKA-AKAP95-p105. Sci Signal 2:ra28

Wu L, Estrada O, Zaborina O et al (2005) Recognition of host immune activation by Pseudomonas aeruginosa. Science 309:774–777

Yoneyama M, Kikuchi M, Natsukawa T et al (2004) The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. Nat Immunol 5:730–737