Published in final edited form as:

Nat Rev Immunol. ; 11(12): 873–880. doi:10.1038/nri3109.

Strategies to discover regulatory circuits of the mammalian immune system

Ido Amit^{1,2}, Aviv Regev^{1,3}, and Nir Hacohen^{1,4,5}

¹Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA

²Weizmann Institute, Department of Immunology, Rehovot 76100, Israel

³Howard Hughes Medical Institute, Department of Biology, MIT, Cambridge, MA 02142, USA

⁴Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, MA 02129, USA

⁵Department of Medicine, Harvard Medical School, Boston, MA 02115, USA

Abstract

Recent advances in technologies for genome- and proteome-scale measurements and perturbations promise to accelerate discovery in every aspect of biology and medicine. Although such rapid technological progress provides a tremendous opportunity, it also demands that we learn how to use these tools effectively. One application with great potential to enhance our understanding of biological systems is the unbiased reconstruction of genetic and molecular networks. Cells of the immune system provide a particularly useful model for developing and applying such approaches. Here, we review approaches for the reconstruction of signalling and transcriptional networks, with a focus on applications in the mammalian innate immune system.

With the rapid development of genomic and proteomic tools, global profiling (BOX 1) of cellular states in terms of gene and protein expression has become a routine activity in immunology. This approach has been used to study T-cell activation signatures¹, blood cell states in patients with autoimmunity², the responses of host cells to infection^{3, 4} and expression patterns that predict vaccine efficacy⁵. Although these datasets have provided important starting points for specific studies of the individual components in these processes, there have been few, if any, attempts to both generate and functionally test the large number of hypotheses that can be derived from genome-scale datasets. Here, we describe systematic functional strategies for the purpose of reconstructing the regulatory networks that underlie any immune process (Fig. 1).

Box 1

Tools of systems immunology

Global profiling

In this approach, the goal is to measure particular parameters, such as transcript levels or phosphorylated proteins, at the global scale of the entire genome or proteome. Microarrays have been the mainstay of global profiling for more than a decade, but new methods are rapidly shifting the focus towards unbiased sequencing methods that do not depend on predetermined nucleotide probes. These new techniques include transcriptome sequencing by RNA-seq, as well as chromatin immunoprecipitation followed by

sequencing (ChIP-seq) for studying the binding of proteins to DNA. In addition, mass spectrometry can now be used to profile thousands of peptides in a quantitative manner⁶⁰ — for example, using SILAC (stable isotope labelling with amino acids in cell culture), iTRAQ (isobaric tags for relative and absolute quantification) or other labelling methods — and this allows the discovery of protein modifications and protein–protein interactions. Although these new methods are extremely informative, they can only be used to profile a limited number of samples owing to their cost (US\$500–2,000 per sample) and the requirement for a large number of input cells.

Mesoscale profiling

It is not yet practical to study many conditions or perturbations using global profiling, but several affordable mesoscale profiling methods are available or under development. In such approaches, information obtained from global profiling is used to derive a smaller gene or protein expression signature that can be quantified across many more samples. An optimal methodology would accurately measure many parameters (for example, 100–1,000 distinct parameters, such as transcripts, proteins, phosphopeptides and metabolites) using minimal cell input (for example, 1–1,000 cells per well in multiwell plates) and at low cost (US\$1–50 per sample). Such methods have been developed for nucleic acid detection (for example, the technologies available from NanoString³¹, Fluidigm⁶¹ and Luminex^{62,63}) and for the identification of protein–DNA interactions (for example, ChIP–String)⁶⁴. Techniques are still under development for proteomics, but include multiplex protein detection by antibodies bound to a solid support^{63,65}, microwesterns⁶⁶ and mass cytometry³⁵.

Network Reconstruction

In a post-genomic world, a key goal in immunology should be to define all of the active components ('nodes'), their interactions with each other ('edges') and their functional roles in any immunological process. Components include proteins, coding and non-coding RNAs, metabolites, DNA motifs and many other bioactive molecules. Interactions can be direct (such as protein–protein or protein–nucleic acid binding, and enzyme–substrate reactions) or indirect (such as when a protein is required for the activity of another protein through intermediate components). A regulatory network focuses on the subset of components that primarily control the activities of other components; such networks include the transcriptional and signalling networks discussed here.

Defining the problem: systematic reconstruction of regulatory networks

A regulatory circuit combines *trans*-inputs (such as the levels and activities of transcription factors, non-coding regulatory RNAs and signalling molecules) and *cis*-inputs (such as regulatory sequences in the promoter and enhancer of a gene) to determine the level of mRNA produced from a gene. The central goal of reconstruction of a regulatory circuit is to identify all of the inputs (for example, proteins, non-coding RNAs and *cis*-regulatory elements), their physical 'wirings' and the transcriptional functions that they implement to regulate the level of mRNA^{6–10} (Fig. 1A). Thus, a successful model should simultaneously address two issues. First, it should provide a functional description of the input-output relationships (for example, if regulator A is induced, then target gene B is repressed to a particular extent). Second it should provide a physical description of the circuit (for example, regulator A binds to the promoter of gene B in sequence Y, modifies its chromatin and leads to repression). Although regulatory networks control complex downstream cellular phenotypes (such as cell death, proliferation and migration), here we focus on the methods that can be used to reconstruct the connectivity of a network through the monitoring of hundreds to thousands of cellular parameters, such as the levels of mRNAs.

Pros and cons of genomic approaches

Regulatory circuits can be studied and reconstructed on either a small scale or a large (genomic) scale¹⁰. Small-scale studies typically focus on mechanistic and quantitative models that are examined in detail^{11, 12}. For example, a recent study modelled and tested a three-component circuit - consisting of nuclear factor- κ B (NF- κ B), activating transcription factor 3 (ATF3) and CCAAT/enhancer-binding protein- δ (C/EBPD) - that is involved in the response to Toll-like receptor 4 (TLR4) ligation in macrophages¹³. Large-scale genomic approaches typically rely on statistical association in genomic datasets, resulting in models that are less quantitative and less mechanistic. For example, one study in macrophages¹⁴ used genome-wide transcriptional profiles to identify clusters of co-expressed — and thus potentially coregulated — genes. The authors then searched for transcription factors with an expression pattern that predicts that of the genes within a cluster (a ‘*trans* approach’) or for *cis*-regulatory elements that are enriched in the promoters of the genes within a cluster (a ‘*cis* approach’). Such a *cis* approach led to the discovery of a novel role for ATF3 in controlling the response of macrophages to lipopolysaccharide (LPS)¹⁵. In addition, both *cis* and *trans* approaches have recently led to the identification of many potential new regulators of human haematopoietic cell differentiation¹⁶ or of the differentiation of myeloid progenitor cells *in vitro*¹⁷; several of these potential regulators have been experimentally validated^{16, 17}.

Although both *cis* and *trans* genomic approaches provide important insights and have a broad scope, they also have substantial limitations, which have not been typically addressed in a systematic way. On the one hand, *cis*-regulatory elements in mammals are notoriously difficult to identify computationally — owing to a high falsepositive background (within a genome) and a lack of conservation in transcription factor binding sites (across multiple genomes)¹⁸ — and it is often not possible to distinguish between similar binding sites of related transcription factors¹⁹. Furthermore, even the presence of a ‘real’ *cis*-regulatory element in the regulatory region of a gene does not necessarily imply that it is functionally bound and has a regulatory capacity in the relevant condition^{20,21}. Moreover, many of the sequences that regulate mammalian genes are found in gene enhancers, which are probably numerous, function in a tissue-specific manner and are hard to assign to their target genes because they can be located at variable and substantial distances from their target genes^{22,23}. Finally, we understand little about the way in which different *cis* elements in a promoter or enhancer function together to determine the level of mRNA transcribed from a particular gene^{24,25}.

On the other hand, *trans* models that rely mainly on correlations between the expression of regulators and the expression of putative target genes can be affected by both false positives, owing to regulators that are coexpressed with genes that they do not regulate, and false negatives, owing to regulators that are not differentially expressed at the mRNA level. Physical interactions between a transcription factor and a gene sequence can be identified in direct binding assays such as chromatin immunoprecipitation followed by sequencing (ChIP-seq), in which bound transcription factors are immunoprecipitated and the associated DNA fragments are then sequenced. However, these methods cannot distinguish between functional binding that affects gene expression and binding that has little if any effect²⁰, nor do they predict the sign or magnitude of the effect of individual bound factors or their combinations.

Addressing these limitations of genomic approaches requires the systematic integration of more than a single strategy and data type. Until recently, progress towards such systematic network reconstruction has been most rapid in unicellular model organisms, most notably the baker’s yeast *Saccharomyces cerevisiae*¹⁰. Studies in *S. cerevisiae* have led the way in

integrative strategies that combine computational discovery of evolutionarily conserved *cis*-regulatory elements, transcriptional profiling of target genes, large-scale assays for detecting the binding of transcriptional regulators and systematic perturbation of regulators. For example, studies combining genetic perturbations, gene expression measurements and global ChIP profiling under physiological conditions²⁶ have uncovered the logic of the specific transcriptional networks that underlie several environmental responses, such as the response to DNA-damaging agents²¹ or high osmolarity²⁰. These studies show how gene promoters integrate information from the environment to generate a context-dependent response^{20,21}. Genetic interaction studies, in which multiple components of a system are perturbed together^{27,28}, have allowed researchers to further dissect the effects of multiple inputs in a regulatory circuit²⁰. For example, one study has determined how the signal from Hog1 (a mitogen-activated protein kinase) is spread out to five transcription factors that are then combined in different logic gates on various promoters to control the response of yeast to osmotic stress²⁰.

However, achieving similar success in mammals has proven much more challenging, owing to the larger number of transcription factors and *cis*-regulatory sites, the greater complexity of regulatory logic in enhancers and promoters and the technical limitations in mammalian genetic perturbation and manipulation¹⁰. Using genome-scale genetic and physical evidence to derive models of regulatory networks in mammals thus remains a significant open problem.

Reconstructing immune networks

Several studies have attempted to develop models of transcriptional networks in the immune system^{13–15,17,29,30}. In such studies, the first step is to profile on a global scale the dynamic changes in cellular transcripts that occur after stimulation with a particular physiological ligand. This is followed by a second step of devising a provisional network model based on the co-expression of transcription factors and their putative targets and/or on the enrichment of *cis*-regulatory elements in these targets. In the third step, some of the predictions of the resulting model are experimentally tested using either genetic approaches to perturb specific transcription factors or physical approaches to measure the binding of transcription factors to promoters. In some cases, the model is then updated based on validation data (FIG. 2). These studies have led to useful models of regulatory networks and a deeper understanding of the specific components in such networks.

Case study: the innate immune response of DCs through TLRs

One example of a method to derive *trans* models in mammals was used in our study of the responses of dendritic cells (DCs) to LPS²⁹. This method has three unique aspects. First, representative ‘mesoscale’ gene signatures (BOX 1) are selected for assessing the output and state of the circuit. This involves choosing genes with an expression pattern that can capture the main variations in the data or that can accurately classify a cell’s state. Second, a relatively broad set of potential transcriptional regulators are nominated in the provisional model (this does not require the identification of *cis* binding sites for these transcription factors in target genes). Third, RNA interference (RNAi) is used to perturb each candidate regulator, followed by quantitative monitoring of the resulting changes in gene expression signatures at the mesoscale (rather than genome-wide profiling, which, with the currently available technology, is less practical for testing a large number of samples). We have used this profiling and perturbation strategy for an initial reconstruction of the TLR circuit in DCs²⁹ (FIG. 1b). Using primary mouse DCs derived from the bone marrow as our experimental system, we profiled genome-wide mRNA levels over time in DCs stimulated with TLR ligands and nominated candidate regulators with expression patterns (based on microarray profiles) that could predict changes in the expression of modules of co-expressed

genes (at the same time or at a later time point). We then selected a TLR-induced mRNA signature for monitoring gene expression at the mesoscale. This consisted of multiple sets of genes that could discriminate between the different TLR agonists used and between different time points after TLR stimulation. We used lentiviral small hairpin RNAs (shRNAs) to silence the expression of 144 candidate transcriptional regulators and a multiplex mRNA detection system (the NanoString nCounter³¹) to quantify the effect of each gene knockdown on the expression of ~120 signature genes. From these expression data, we generated a network model that linked candidate regulators with specific target genes (relying on both control shRNAs and control genes to calculate a false discovery rate and identify reliable relations²⁹).

The reconstructed network is consistent with prior knowledge of TLR signalling (thus validating its accuracy) and also reveals new network components together with their roles in TLR responses, which demonstrates the power of such an unbiased approach. For example, we correctly assigned 32 known regulators — including NF- κ B, interferon response factors (IRFs) and signal transducer and activator of transcription (STAT) proteins — to their target genes. In addition, we identified another 68 regulators that have not been previously implicated in the response of DCs to TLR agonists. These included regulators of the cell cycle, such as retinoblastoma-like 1 (RBL1), retinoblastoma 1 (RB1), E2F transcription factor 5 (E2F5), E2F8, N-MYC interactor (NMI), fused in sarcoma (FUS) and TIMELESS. We quantified the contribution of each regulator to the control of two major transcriptional programmes in DCs (the inflammatory and antiviral responses), and this identified a core network of key regulators, as well as many fine-tuners. These regulators used feed-forward circuits, dominant activation and cross-inhibition to control response specificity. Most importantly, the regulatory network that we derived can help to explain the magnitude, direction and kinetics of gene expression in response to pathogen stimulation. For example, we found that the chromatin suppressor CBX4 is upregulated in DCs by stimulation with LPS but not by stimulation with polyinosinic–polycytidylic acid (polyI:C), and this might explain why there is a strong and sustained induction of interferon-beta in response to viral component mimetics (such as polyI:C), but a lower and more transient induction in response to bacterial components or their mimetics (such as LPS and Pam3CSK4). We expect that mining of our dataset by other groups will uncover additional regulators and their roles in the circuit.

What the future holds

We propose four major directions for extending the network reconstruction strategies that we and others have described. First, it should be possible to expand the scope of circuit components, by studying a more broadly defined set of ‘regulators’. For example, the basic approach that we used for transcription factors in DCs can be naturally expanded to study signalling molecules. Indeed, we recently used transcriptional profiles to nominate candidate signalling proteins and then used RNAi to discover new signalling genes used by the TLR pathway, leading us to define a new arm of the antiviral pathway mediated by the Polo-like kinases³⁰. This is consistent with the large number of cell cycle transcriptional regulators in the antiviral transcriptional network²⁹. Furthermore, by using proteomics and phosphoproteomics, rather than mRNA profiles, it should be possible to further expand this scope to identify candidate regulators with mRNA levels that are unchanged in the response, as we demonstrated in the same study³⁰.

Second, researchers could monitor additional types of output signature (in terms of reduced complexity signatures or mesoscale profiles (BOX 1)). This would involve measuring not only ‘steady-state’ mRNA levels, as typically occurs in profiling studies, but also specific steps in the life cycle of an mRNA or a protein. Such approaches have been effective in

several studies of mammalian cells^{32,33}. In our work, we recently showed how to use metabolic labelling and computational modelling to measure the levels of newly transcribed RNA, distinguish preprocessed immature RNA from mRNA and infer RNA transcription, processing and degradation rates³⁴. By coupling such measurements to perturbations, we should be able to distinguish the specific effects of different regulators on different steps in the mRNA life cycle and beyond. Other output signatures — such as phosphorylation patterns across a specific set of proteins³⁰, metabolite levels or secreted cytokines — can enhance models of network structure and dynamics. In addition, merging these disparate datasets (for example, using Bayesian networks that can incorporate any type of data) will help to generate more predictive and mechanistic models. Recent technological advances in mass spectrometry methods have enabled the profiling of protein signatures from single immune cells³⁵; this technique can identify molecular or regulatory relations that are obscured by population level measurements. Furthermore, there are now more efficient mesoscale methods for measuring peptides, such as selective reaction monitoring (SRM) mass spectrometry³⁶. These developments are expected to provide more sensitive approaches for proteomic analysis of more samples with fewer cells.

Third, advances should be made in deciphering the physical interactions that are involved in a regulatory circuit and the complexity of regulatory logic, by combining measurements of physical interactions in native and perturbed states. For *trans* inputs, combining ChIP-seq with perturbation and expression studies would help to distinguish between the direct and indirect effects of binding to a DNA sequence, as well as between functional and non-functional binding. This has been done successfully for yeast²⁰, but not yet for primary mammalian cells perturbed using RNAi (partly owing to the high cost). Furthermore, using ChIP-seq of one factor as a readout following perturbation of another factor, it should be possible to dissect complex feedback and feedforward pathways within a regulatory circuit. For *cis* inputs, the recent development of improved technologies to synthesize complex nucleotide sequences^{37–40} and to deliver them in targeted ways to cells^{41,42} is likely to allow systematic dissection and perturbation of *cis*-regulatory elements, and the resulting data could be coupled to either mRNA profiles or measurements of physical interactions.

Fourth, we should expand the scope of perturbations and the complexity of external signals. In addition to RNAi-mediated knockdown of gene expression, we can introduce other genetic perturbations to the circuit, such as overexpression (gain-of-function perturbations), mutations (including those modelling natural variants) and multiple simultaneous knockdowns. Furthermore, we can use small molecule inhibitors of circuit components, thus greatly improving our ability to quantitatively determine the effect of a component of interest. In addition, to decipher the complex functions that biological circuits compute we can manipulate and monitor the effects of more complex inputs, including combinations of signals, complex pathogens, signals of different magnitudes or temporally varying signals. Advances in microfluidics devices will be instrumental in achieving this goal⁴³.

Applications to other immune networks

The immune system is ideal for applying this paradigm of network reconstruction. Cells of the blood and lymphoid tissues are readily available from mouse models and from humans, and these cells can be used to examine numerous developmental and physiological phenomena *ex vivo*. These include DC maturation (FIG. 1b), DC-mediated activation of T cells and T cell polarization (FIG. 1c).

As an illustrative example, consider the precise balance of cytokines produced by T helper (TH) cells⁴⁴; this balance is crucial for protection against pathogens, as well as for the progression of autoimmunity and inflammation (FIG. 1c). We expect the regulatory

networks underlying naive CD4⁺ T cell polarization to be intricate, dynamic, multilayered and context dependent, as we found for TLR pathways. An unbiased analysis of TH1, TH2, TH17 and regulatory T (TReg) cell polarization could uncover the factors and circuits that either regulate or are independent of the key transcription factors that are known for each subset, namely T-bet, GATA-binding factor 3 (GATA3), retinoic acid receptor-related orphan receptor- γ t (ROR γ t) and forkhead box P3 (FOXP3), respectively.

The approach that we describe here could be naturally applied to this problem by measuring genome-wide mRNA expression levels (or another global parameter as described above) along the time course of naive CD4⁺ T cell polarization (for example, following stimulation with a CD3-specific antibody and a well-defined cocktail of cytokines). The mRNA expression profiles could then be used to select candidate regulators and a representative signature of the output (together with appropriate time points and ligands) for perturbation studies. Candidate regulators could be perturbed in primary T cells (for example, using a validated RNAi lentiviral construct)⁴⁵, and the resulting quantitative changes in gene expression signatures could be measured using a multiplex detection method. The end result would be the generation of a network model that could provide specific new predictions for further studies. Importantly, the polarization of T cells occurs over several days through a cascade of known and still-unknown regulators, and therefore provides an exciting opportunity for the discovery of new network components.

Another example is cell differentiation in the haematopoietic lineage (FIG. 1d). Haematopoietic cell differentiation — starting from haematopoietic stem cells and ending with all blood cells — occurs over a long time period and requires the stable rewiring of transcriptional networks in distinct ways in different lineages. Indeed, in a recent study, we measured global gene expression profiles of many cell types at different stages during human haematopoietic cell development and generated a pair of provisional models (one *cis* and one *trans*) to nominate candidate regulators (a few of which were validated by either perturbation or ChIP-seq)¹⁶. An extensive effort is also underway in mice to profile gene expression in a comprehensive set of resting and activated immune cells⁴⁶. In the future, it should be possible to select representative gene expression signatures for each haematopoietic cell type, and then to perturb each gene and measure the resulting gene signatures to produce a functional map between regulators and cell types.

A third illustrative direction for future research would be the roles of large noncoding RNAs (lncRNAs) in transcriptional circuits in the immune system^{47–49}. Recent studies indicate that lncRNAs affect regulatory circuits through various mechanisms, for example by recruiting chromatin modifiers^{50–52} and by mobilizing transcription factors and the transcriptional machinery^{52,53}. Moreover, specific lncRNAs have been shown to be involved in regulating signalling and transcription in immune cells^{47,53}. A recent study used a profiling and perturbation strategy to dissect the roles of 200 lncRNAs that are expressed in embryonic stem cells⁵⁴ and, in a similar manner, this strategy could be applied to determine the roles of lncRNAs in immune regulatory circuits.

Thus, the strategy we describe — in which mesoscale profiling is used to cost-effectively monitor the effects of many perturbations and thus reveal network structure — is generic and can be applied to decipher regulatory circuits for almost any type of immune activation or differentiation phenomenon. Although genome-wide small interfering RNA (siRNA) screens are also gaining popularity, mesoscale approaches are likely to be more feasible and cost effective for most problems. The mesoscale approach is also likely to yield more detailed and mechanistic models (with fewer false positives and artefacts) compared with large-scale RNAi screens (which rely on one or two outputs and test many hypotheses). This is because the mesoscale approach relies on complex molecular signatures as readouts for

the initial discovery of candidate factors and for deeper monitoring of molecular changes after perturbation. Nevertheless, for identifying the set of crucial genes that affect a cellular phenotype (such as resistance to infection), large-scale RNAi screens are likely to be more effective owing to their much larger search space⁵⁵.

Translational implications

Some of the genes and processes that go awry in human immune disorders have been revealed in the past two decades using unbiased genetic approaches, such as genome-wide association studies⁵⁶. However, these causal genetic alleles (which may lie in either coding or non-coding elements⁵⁷) do not function in isolation; instead, these genes and their encoded products are embedded within complex molecular and cellular networks. Therefore, each disease-associated allele exhibits a ‘ripple’ effect, leading to the dysregulation of multiple circuits⁵⁸.

Thus, a comprehensive understanding of the molecular changes associated with human traits and diseases will only be possible when unbiased genomic approaches are used routinely (FIG. 3). In addition to direct profiling of mRNA and protein expression patterns, network reconstruction methods will need to be extended to primary human cells. To achieve this aim, researchers will need to take into account all of the challenges posed by inter-individual variability, the limitations in cell and tissue procurement, the complex interplay with the environment and the limitations in carrying out whole-organism experiments.

It is useful to speculate how this might be done. First, to preserve precious cell samples and carry out studies across many individuals, we envision miniaturization of the needed experimental techniques for studies of small numbers of cells, allowing effective studies of each immune cell type across many individuals. Second, cell–cell interactions cannot be ignored and will need to be modelled using more sophisticated cell culture methods based on advances in tissue engineering and microfluidics devices. Third, the use of human induced pluripotent stem cells that can be differentiated *in vitro* into different types of immune cell will be essential to minimize the effects of uncontrolled and often unknown environmental variables (by deriving naive starting cells that have not been exposed to different environmental conditions *in vivo*). Finally, more sophisticated computational methods for statistically associating specific network behaviours with upstream gene alleles and downstream clinical outcomes will help to define causality and candidate targets for therapeutics⁵⁹. Solving these and additional unanticipated problems will form the foundation for an unbiased functional approach to human immunology that can pinpoint the cells and networks that contribute to human traits and diseases.

Acknowledgments

The authors would like to thank the NIH, HHMI, HFSP and the Broad Institute for funding the work presented here.

References

1. Haining WN, Wherry EJ. Integrating genomic signatures for immunologic discovery. *Immunity*. 2010; 32:152–61. [PubMed: 20189480]
2. Pascual V, Chaussabel D, Banchereau J. A genomic approach to human autoimmune diseases. *Annu Rev Immunol*. 2010; 28:535–71. [PubMed: 20192809]
3. Berry MP, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature*. 2010; 466:973–7. [PubMed: 20725040]

4. Diamond DL, et al. Temporal proteome and lipidome profiles reveal hepatitis C virus-associated reprogramming of hepatocellular metabolism and bioenergetics. *PLoS Pathog.* 2010; 6:e1000719. [PubMed: 20062526]
5. Querec TD, et al. Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. *Nature Immunol.* 2009; 10:116–25. [PubMed: 19029902]
6. Alon, U. An introduction to systems biology: design principles of biological circuits. Chapman & Hall/CRC; Boca Raton, FL: 2007.
7. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nature Rev Genet.* 2004; 5:101–13. [PubMed: 14735121]
8. Hyduke DR, Palsson BO. Towards genome-scale signalling network reconstructions. *Nature Rev Genet.* 2010; 11:297–307. [PubMed: 20177425]
9. Kitano H. Biological robustness. *Nature Rev Genet.* 2004; 5:826–37. [PubMed: 15520792]
10. Kim HD, Shay T, O'Shea EK, Regev A. Transcriptional regulatory circuits: predicting numbers from alphabets. *Science.* 2009; 325:429–32. [PubMed: 19628860]
11. Istrail S, Davidson EH. Logic functions of the genomic cis-regulatory code. *Proc Natl Acad Sci U S A.* 2005; 102:4954–9. [PubMed: 15788531]
12. Mason PB Jr, Lis JT. Cooperative and competitive protein interactions at the hsp70 promoter. *J Biol Chem.* 1997; 272:33227–33. [PubMed: 9407112]
13. Litvak V, et al. Function of C/EBPdelta in a regulatory circuit that discriminates between transient and persistent TLR4-induced signals. *Nature Immunol.* 2009; 10:437–43. [PubMed: 19270711]
14. Ramsey SA, et al. Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics. *PLoS Comput Biol.* 2008; 4:e1000021. [PubMed: 18369420]
15. Gilchrist M, et al. Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4. *Nature.* 2006; 441:173–8. [PubMed: 16688168]
16. Novershtern N, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell.* 2011; 144:296–309. [PubMed: 21241896]
17. Suzuki H, et al. The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nature Genet.* 2009; 41:553–62. [PubMed: 19377474]
18. Schmidt D, et al. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science.* 2010; 328:1036–40. [PubMed: 20378774]
19. Badis G, et al. Diversity and complexity in DNA recognition by transcription factors. *Science.* 2009; 324:1720–3. [PubMed: 19443739]
20. Capaldi AP, et al. Structure and function of a transcriptional network activated by the MAPK Hog1. *Nature Genet.* 2008; 40:1300–6. [PubMed: 18931682]
21. Workman CT, et al. A systems approach to mapping DNA damage response pathways. *Science.* 2006; 312:1054–9. [PubMed: 16709784]
22. Ernst J, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature.* 2011; 473:43–9. [PubMed: 21441907]
23. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnol.* 2010; 28:495–501. [PubMed: 20436461]
24. Raveh-Sadka T, Levo M, Segal E. Incorporating nucleosomes into thermodynamic models of transcription regulation. *Genome Res.* 2009; 19:1480–96. [PubMed: 19451592]
25. Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, Gaul U. Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature.* 2008; 451:535–40. [PubMed: 18172436]
26. Hu Z, Killion PJ, Iyer VR. Genetic reconstruction of a functional transcriptional regulatory network. *Nature Genet.* 2007; 39:683–7. [PubMed: 17417638]
27. Bandyopadhyay S, et al. Rewiring of genetic networks in response to DNA damage. *Science.* 2010; 330:1385–9. [PubMed: 21127252]
28. Costanzo M, et al. The genetic landscape of a cell. *Science.* 2010; 327:425–31. [PubMed: 20093466]
29. Amit I, et al. Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science.* 2009; 326:257–63. [PubMed: 19729616]

30. Chevrier N, et al. Systematic Discovery of Signaling Components Identifies New Branches in Viral-Sensing Pathways. *Cell*. 2011; XX:X–X.
31. Geiss GK, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnol*. 2008; 26:317–25. [PubMed: 18278033]
32. Dolken L, et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA*. 2008; 14:1959–72. [PubMed: 18658122]
33. Schwanhauss B, et al. Global quantification of mammalian gene expression control. *Nature*. 2011; 473:337–42. [PubMed: 21593866]
34. Rabani M, et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nature Biotechnol*. 2011; 29:436–42. [PubMed: 21516085]
35. Bendall SC, et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science*. 2011; 332:687–96. [PubMed: 21551058]
36. Picotti P, Bodenmiller B, Mueller LN, Domon B, Aebersold R. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell*. 2009; 138:795–806. [PubMed: 19664813]
37. Cox RS 3rd, Dunlop MJ, Elowitz MB. A synthetic three-color scaffold for monitoring genetic regulation and noise. *J Biol Eng*. 2010; 4:10. [PubMed: 20646328]
38. Gertz J, Siggia ED, Cohen BA. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature*. 2009; 457:215–8. [PubMed: 19029883]
39. Cox RS 3rd, Surette MG, Elowitz MB. Programming gene expression with combinatorial promoters. *Mol Syst Biol*. 2007; 3:145. [PubMed: 18004278]
40. Zeevi D, et al. Compensation for differences in gene copy number among yeast ribosomal proteins is encoded within their promoters. *Genome Res*. 2011 In press.
41. Le Provost F, et al. Zinc finger nuclease technology heralds a new era in mammalian transgenesis. *Trends Biotechnol*. 2010; 28:134–41. [PubMed: 20015561]
42. Zhang F, et al. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature Biotechnology*. 2011; 29:149–53.
43. Szita N, Polizzi K, Jaccard N, Baganz F. Microfluidic approaches for systems and synthetic biology. *Curr Opin Biotechnol*. 2010; 21:517–23. [PubMed: 20829028]
44. Zhou L, Chong MM, Littman DR. Plasticity of CD4+ T cell lineage differentiation. *Immunity*. 2009; 30:646–55. [PubMed: 19464987]
45. Astier AL, et al. RNA interference screen in primary human T cells reveals FLT3 as a modulator of IL-10 levels. *J Immunol*. 2010; 184:685–93. [PubMed: 20018615]
46. Heng TS, Painter MW. The Immunological Genome Project: networks of gene expression in immune cells. *Nature Immunol*. 2008; 9:1091–4. [PubMed: 18800157]
47. Guttman M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009; 458:223–7. [PubMed: 19182780]
48. Lee JT. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev*. 2009; 23:1831–42. [PubMed: 19684108]
49. Pang KC, et al. Genome-wide identification of long noncoding RNAs in CD8+ T cells. *J Immunol*. 2009; 182:7738–48. [PubMed: 19494298]
50. Huarte M, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*. 2010; 142:409–19. [PubMed: 20673990]
51. Khalil AM, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A*. 2009; 106:11667–72. [PubMed: 19571010]
52. Wang KC, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*. 2011; 472:120–4. [PubMed: 21423168]
53. Willingham AT, et al. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science*. 2005; 309:1570–3. [PubMed: 16141075]
54. Guttman MDJ, Carey BW, Garber M, Grenier J, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE, Lander ES. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*. 2011 In press.

55. Carette JE, et al. Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature*. 2011; 477:340–3. [PubMed: 21866103]
56. Xavier RJ, Rioux JD. Genome-wide association studies: a new window into immune-mediated diseases. *Nature Rev Immunol*. 2008; 8:631–43. [PubMed: 18654571]
57. Musunuru K, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*. 2010; 466:714–9. [PubMed: 20686566]
58. Heinig M, et al. A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature*. 2010; 467:460–4. [PubMed: 20827270]
59. Schadt EE. Molecular networks as sensors and drivers of common human diseases. *Nature*. 2009; 461:218–23. [PubMed: 19741703]
60. Cox J, Mann M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem*. 2011; 80:273–99. [PubMed: 21548781]
61. Spurgeon SL, Jones RC, Ramakrishnan R. High throughput gene expression measurement with real time PCR in a microfluidic dynamic array. *PLoS One*. 2008; 3:e1662. [PubMed: 18301740]
62. Peck D, et al. A method for high-throughput gene expression signature analysis. *Genome Biol*. 2006; 7:R61. [PubMed: 16859521]
63. Dunbar SA. Applications of Luminex xMAP technology for rapid, high-throughput multiplexed nucleic acid detection. *Clin Chim Acta*. 2006; 363:71–82. [PubMed: 16102740]
64. Ram O, et al. Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell*. 2011 In Press.
65. Wolf-Yadlin A, Sevecka M, MacBeath G. Dissecting protein function and signaling using protein microarrays. *Curr Opin Chem Biol*. 2009; 13:398–405. [PubMed: 19660979]
66. Ciaccio MF, Wagner JP, Chuu CP, Lauffenburger DA, Jones RB. Systems analysis of EGF receptor signaling dynamics with microwestern arrays. *Nature Methods*. 2010; 7:148–55. [PubMed: 20101245]

Figure 1. Immune cell activation, cell states and network reconstruction

A. We first define the goal of network reconstruction with an abstract example of a simple circuit. When resting immune cells encounter a specific ligand (such as LPS, which signals through TLR4, or peptide–MHC, which signals through TCR), they make a transition from a resting to an activated state through a series of intermediate states (where a ‘state’ is defined by specific values of measurable parameters, such as transcript levels or phosphorylated proteins, that are reproducibly observed under certain experimental conditions at specific timepoints). Such state transitions occur as a result of a temporal cascade of signalling and transcriptional events. In this simplified example, protein A is initially modified (for example, by phosphorylation) in response to a ligand and induces transcription of gene B by binding to its promoter. Proteins B and A then form a complex that stimulates transcription of C, a key protein in the final activated state. The challenge of network reconstruction is to identify components A, B and C, their direct and indirect interactions and their roles in regulating the parameters that define each cellular state. This can be carried out using an experimental strategy as described in our studies of dendritic cells (DCs)²⁹. **B–D.** Physiological examples of immune phenomena occurring at three timescales (from rapid to slow) that can be studied by network reconstruction: innate immune responses of DCs to pathogens (**B**), T cell responses to antigen (**C**) and haematopoiesis (**D**). **B.** Dendritic cell gene expression profiles can be measured in response to distinct ligands (such as lipopolysaccharide (LPS), double-stranded RNA (dsRNA), beta-glucan and phosphatidyl serine (PtdSer)). Each of the resulting four states of DCs has distinct effects on an individual’s innate and adaptive immune responses. Many additional states can be studied given the number of ligands affecting DCs. **C.** T cell polarization into specific states is a crucial decision point in the mammalian immune response, and is controlled by cytokines and other molecules that are present when T cells are activated by peptide–MHC complexes. **D.** Haematopoiesis is a well-studied developmental process in which haematopoietic stem cells (HSCs) develop into precursors (such as common lymphoid progenitors (CLPs) and common myeloid progenitors (CMPs)) that further develop into mature immune cells. Note that only a few cell types are shown for illustration.

Figure 2. Overview of the proposed network reconstruction strategy

Top, Four major steps in network reconstruction. **Bottom,** examples of technologies that could be used at each step.

Figure 3. Future clinical applications

Immune cells are isolated from patient blood or other tissues and studied *ex vivo* using global and focused profiling approaches (for example, as described in Figs.1 and 2 for studying the resting state or the responses to ligands and perturbations) to derive network models that help explain observed cellular states. If alterations in network behavior can be associated with particular disease phenotypes, then individual-specific network models will be helpful in the diagnosis, prognosis and selection of therapeutic interventions for immune disorders.