Christophe Benoist
Ronald N. Germain
Diane Mathis

A *Plaidoyer* for 'Systems Immunology'

Authors' address
Christophe Benoist[1], *Ronald N. Germain*[2], *Diane Mathis*[1]
[1]Section on Immunology and Immunogenetics, Joslin Diabetes Center, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[2]Lymphocyte Biology Section, Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

Correspondence to:
Christophe Benoist
Department of Medicine
Section on Immunology and Immunogenetics
Joslin Diabetes Center
Brigham and Women's Hospital
Harvard Medical School
One Joslin Place
Boston, MA 02215
USA
Tel.: +1 617 264 2745
Fax: +1 617 264 2744
E-mail: cbdm@joslin.harvard.edu

Acknowledgements
This text benefited greatly from discussion with and contribution to references from Reinhard Obst, John Luckey, and Linh Nguyen. This work was supported in part by grants from the Intramural Research Program of the NIH, NIAID, and by the Young Chair.

Immunological Reviews 2006
Vol. 210: 229–234

Immunological Reviews
0105-2896

Summary: A complete understanding of the immune system will ultimately require an integrated perspective on how genetic and epigenetic entities work together to produce the range of physiologic and pathologic behaviors characteristic of immune function. The immune network encompasses all of the connections and regulatory associations between individual cells and the sum of interactions between gene products within a cell. With 30 000+ protein-coding genes in a mammalian genome, further compounded by microRNAs and yet unrecognized layers of genetic controls, connecting the dots of this network is a monumental task. Over the past few years, high-throughput techniques have allowed a genome-scale view on cell states and cell- or system-level responses to perturbations. Here, we observe that after an early burst of enthusiasm, there has developed a distinct resistance to placing a high value on global genomic or proteomic analyses. Such reluctance has affected both the practice and the publication of immunological science, resulting in a substantial impediment to the advances in our understanding that such large-scale studies could potentially provide. We propose that distinct standards are needed for validation, evaluation, and visualization of global analyses, such that in-depth descriptions of cellular responses may complement the gene/factor-centric approaches currently in favor.

' . . . l'étude de la nature suppose dans l'esprit deux qualités qui paraissent opposées: les grandes vues qui embrassent tout d'un coup d'oeil, et les petites attentions qui ne s'attachent qu'à un seul point.'

Buffon, Histoire Naturelle, 1749

' . . . the study of natural history calls for a mind with apparently opposite qualities: wide views that grasp all at a glance, and the detailed care that focuses on a single point.'

Any given cell utilizes a significant proportion of the genome, and its manifest or potential phenotype is determined by the concerted activities of the products encoded by these thousands of transcribed genes. The activities are coordinated within a highly complex network, entailing several levels of regulation that help maintain the cell's homeostasis or direct its further differentiation. Extracellular signals or other perturbations of the network can have a strong and unforeseen impact on system behavior within such a network, effects that can appear surprisingly distant on the signaling maps that adorn review articles or lab corridors. If understanding a biological system is defined as the capacity to accurately predict the behavior of that system when perturbed, then it is imperative not only to identify all of the potentially relevant

cellular and molecular players but also to portray how they interact both qualitatively and quantitatively to effect the behavior of the system. This approach is particularly true of the immune system, with its rich cast of players.

Because of this problem's dimension even at the level of an individual cell, immunologists might be considered in the position of the proverbial three blind men exploring an elephant, each with a detailed sense of a limited part of the whole. In the last decade, a number of high-throughput techniques have appeared, each of which has the potential to describe some aspect of the biochemical or gene regulatory network on a large scale. Any given technology reveals only one layer of the network: steady-state levels of gene transcripts of cellular proteins, the set of genes bound by a transcription factor family. Yet, because of their global scope, these methods do have the ability to encompass the entirety of molecule classes in a cell (we will refer to this potential as 'genome-scale' in a generic sense, to denote the ability to describe the globality of a class of cellular elements, to reveal the elephant as a whole). The genomic and proteomic signatures that correspond to defined functional states comprise the lists of elements whose reverse engineering and ordering into a consistent network will be required to decipher cellular physiology and immune responses.

Complexity is discouraged

Genome-scale methods are no longer such a novel technology, especially gene-chips/microarrays. Although still improving in reliability, power and cost, microarray analysis of the entire transcriptome is by now quite well codified and is in fairly routine practice. But despite their integrative potential, these tools so far are having a surprisingly limited impact in guiding us to a higher order understanding of immune cell function. Most existing microarray publications are focused on a tightly defined immunological question, on a single pathway and very often utilize only a minute fraction of the total data collected. Very few studies have utilized genome-wide analytical techniques to address larger questions or have used their potential for reverse engineering of the immune network.

At least in part, this deficiency appears to arise, because genome-scale technologies depart from an experimental model that very strongly dominates current scientific practice. A semantic analysis of recently published papers shows that to a surprising degree, they conform to a very strict model: 'Factor X interacts with factor Y to effect mechanism Z', where X and Y are single elements (genes or proteins, mainly), and Z is an integrated function, such as 'antigen presentation' or 'anaphylaxis'. There are some variants of the formula (either of Y or Z can be skipped, as in 'X phosphorylates Y', or 'Mice deficient in X are defective in function Z'), but, by and large, this 'one factor/one event' (we will call it the x <> y > z) format has come to dominate the immunological literature. We picked at random past issues of three leading immunology journals to analyze in this regard (all papers were of very high quality, some articles of truly seminal nature, and some are contributions by the authors of the present piece). As illustrated in *Table 1*, the x <> y > z model was fit precisely by 23 of the 28 papers. Only five had a different structure. We have no reason to think that this example is an exception, and browsing the current literature shows that this dominance of the x <> y > z scheme is the norm.

This situation seems somewhat paradoxical, because immunology actually has a long history of non-reductionist experimentation. Prior to the advent of gene cloning and of cluster of differentiation and cytokine classifications, a significant fraction of immunological research was performed at the organismal level

***Table 1*. A semantic analysis of papers in randomly picked issues of three leading immunology journals**

Factor X	Factor Y	Phenotype/function Z
KIR and HLA-C		Pre-eclampsia
Blimp-1		Plasma cell differentiation
Helicobacter	DC-SIGN	Th1/2 balance
LAB/NTAL	FcεRI signals	
LAB/NTAL	Mast cell signals	
Aire		organ-specific autoimmunity
CTLA-4-Ig	FKHR	DC apoptosis
SLP-76	SMAC	
CMV p155	H60-NKG2D	NK activity
CD24		homeostatic activation
Gabp	IL7Ra	
RAPL		Lymphocyte traffic
A20	TLR signaling	
IRF7	MyD88 and TRAF6	IFN induction
EBF	Runx1	
CD22		B cell activation/function
Il17	Il23	Granulopoiesis
Sema4a		T priming and differentiation
Runx		T lineage commitment
FoxP3		Treg specification
Rgs1	Gnai2	B lymphocyte traffic
Granzyme A		Mitochondrial damage
Locus1	Locus2	Susceptibility to thymic tolerance

where X interacts with Y to modify (enhance, inhibit) phenotype Z. There were only a minority of articles in these issues that did not exactly fit this framework: Which thymic medullary cells effect clonal deletion, Axonal remodeling influences EAE recovery, Long-range intrachromosomal interactions in the Th2 cytokine locus, Tolerance to Ins2 reduces but does not abolish type 1 diabetes, A Human $CD34^+$ LN Subset Differentiates into $CD56^+$ NK Cells.

and was highly integrative. Because of the multicellular nature of the phenomena being investigated, immunology has had a tradition of seeking to understand how diverse elements function together as a whole. One might say that immunologists have been practicing 'systems biology' well before the term became a fad.

What has led the field away from this heritage? To a great extent, we immunologists have been the victims of our own success, especially the development of monoclonal antibodies for cell phenotyping and protein characterization, together with the early adoption and widespread use of the reverse genetics tools that have emerged in the past 15–20 years. Gene transduction, transgenic mice, gene knockouts and more recently knockdowns, and blockade or activation of single molecules by monoclonal antibodies have allowed immunologists to query the role of individual factors in physiological pathways and to focus on a single factor without explicit regard for the meshwork of elements within which this component carries out its function.

There is certainly great value to the x <> y > z approach. It is a classical reductionist model with full Popperian rigor, whose fruits have provided strong anchors for pathway dissection. Among such studies are many with elegant conceptual foundations and others that through iterative description bring profound insight into particular pathways, revealing the sheer beauty of natural biochemical or gene regulatory architecture, albeit typically only of local extent. Ultimately, given time, the x <> y > z approach would allow us to connect all the dots we need to connect to fully comprehend the immune system (i.e. enough people would see enough parts of the elephant).

So why question this well-accepted method for unraveling nature's secrets? A major problem is that the reductionist approach can promote overly simple thinking, with a focus on the single connection under study that ignores the multiplicity of other influences impinging on the pathway in question and the modulation of distant network properties when the chosen element is manipulated. For instance, the summary of a recent paper in a leading immunology journal and its accompanying commentary proposed a role for a kinase in concert with nuclear factor-κB (NF-κB) and cAMP responsive element-binding protein in the control of T-cell phenotypic differentiation. Yet, is it really possible to consider these connections independently of the numerous other partners with which the kinase interacts, all of which are quite likely to impact the phenomena analyzed? Ultimately, overly narrow viewpoints may give way to broader perspectives, but there is a great deal of inertia in our desire for simple pictures that are not made more difficult to grasp by a spider web of intersecting pathways. A retort to those objecting to the x <> y > z approach and seeking to go 'global' is that experience has validated the utility of the x <> y > z paradigm when the conclusions are supported by rigorous experimental tests of the underlying (possibly oversimplified) hypothesis. It seems fair to argue that no less should be acceptable from those drawing conclusions involving whole sets of genes. The difficulty is that the direct experimental demonstrations that would be optimal in an x <> y > z format are technically impossible or unrealistic when dealing with tens or hundreds of transcripts or proteins. One simply cannot achieve validation of the overall network of gene/protein interactions that emerges from a large-scale microarray or proteomics analysis by the straightforward application of the same experimental tests used to probe a single connection in that network. Even for the simpler notion of a genetic signature identified by the popular clustering algorithms in current use, verifying the functional role of one or two components of that signature by direct test does not establish the validity of the entire signature, or provide information on the boundary conditions that delineate when it can be properly employed in a diagnostic manner.

Should one just ignore the higher order conclusions one can derive from large-scale analyses, simply on the grounds that they cannot be validated by today's experimental conventions? To us, the obvious answer is 'No'. Yet, it is currently very difficult, seemingly impossible, to convince the immunological mainstream of the value of such observations. When they are employed, large scale microarray/proteomic analyses are seen as useful only in so far as they generate a hypothesis that can then be subjected to conventional experimental (read 'reductionist') examination. Typically, a microarray analysis opens the study, resulting in a ranked list of differential gene expression; the experimenter picks one or two of the top candidates, sometimes based on the fold-change in expression between two states of cell differentiation, at other times based on a 'best guess' that a gene product is in a category of potential importance to the behavior being considered (a kinase in the case of receptor-driven signaling, or an anti-apoptotic protein for memory cell maintenance). The functional relevance of the selected candidates is then assessed using various means such as gene knockout, immunoprecipitation, antibody blockade, short interfering RNA knockdown, etc. This is indeed a strong experimental strategy and one that brings us back to the intellectual snugness of the x <> y > z format. But it also leaves aside the higher complexity information that could have been gleaned from the initial data set, had

the urge to 'pick a gene' not been so strong. We believe it impedes the field's intellectual progress to dismiss such observations as mere starting points for further study and not to credit them with having an intrinsic value of their own. Indeed, provided that appropriate tools of statistical validation are employed, there is no reason why global analyses should not be used in a hypothesis-testing mode, to assess similarities between otherwise unrelated populations.

Another reason for the current reluctance to accept genome-scale analyses is, in many cases, that the results from such approaches suffer from the label 'descriptive', the kiss of death for any manuscript or grant application in the hands of a reviewer. The final result of an analysis of the proteome by high-resolution mass spectrometry, a microarray determination of gene expression, or a whole-genome chromatin immunoprecipitation may at times be only a more inclusive and extensive description of a cell state previously analyzed by an incomplete and punctate process. Although accurate descriptions of biological phenomena have proved of immense value in the past, such outcomes conflict with the dominant mindset in current scientific practice, which strongly favors testing a hypothesis defined *a priori*, eschewing more descriptive studies or open explorations that are dismissed as 'fishing expeditions'. Describing the gene expression signatures that define the different states adopted by B cells is indeed descriptive and could be compared with zoological studies of the 18–19th century, to the description of the different beak shapes and sizes among highly related birds. Just such solid descriptions of the structure and connectivity of biological systems seem to us a basic requirement for developing a deep understanding of their operation. An initial phase of descriptive biology is often (always?) required to generate the substrate from which 'interventionist' techniques can create functional insights and testable hypotheses. Didn't Darwin extract some of the most important insights in biology from what at their core are descriptions of this type? Closer to our frame of reference, the role of gene rearrangement in producing complete immunoglobulin protein chains was writ large in the descriptions of immunoglobulin sequences. In anatomy and histology, descriptions of tissues and cell types predated the elucidation of their function. The initial description of dendritic cells was based on their adhesive, morphological, and kinetic features that distinguished them from macrophages and lymphocytes but was entirely uninformed on their key role in the immune response. If we bring a similar level of insight to considerations of large data sets, significant advances in true understanding, beyond mere reporting, might also be achieved.

One can also make a good parallel between the structures generated by global analyses and those that derive from crystallographic analysis of proteins or macromolecules. Both start with long rows of obscure numbers that are processed to arrive at descriptions of objects best visualized by computer graphics. Both describe contacts and connections. In some cases, the structures can provide an immediate insight (e.g. the groove of MHC molecules), and in others, they form the basis for later extensive investigations (e.g. influenza hemagglutinin). But one does not routinely ask crystallographers for a functional followup to 'validate' the significance of the structures.

Complexity is difficult

One can certainly understand the resistance to genome-scale techniques. Many such studies can be conceptually inelegant, intellectually uncomfortable, involve the use of daunting computational methods, and be obscured by a lack of tools for comprehensible representations of the findings.

Inelegant, because they largely omit the intuition, lateral thinking, and scholarly insight that lead to the formulation of a hypothesis and seed an x <> y > z project. Not that all x <> y > z projects really have as rational an inception as recounted in publications, and there is much room to serendipity there as well. 'Let's look at all the genes and see what changes' is not very intellectually demanding as a premise, somewhat on par with 'Let's knock it out and see what happens' (except that the latter can subsequently be dressed up as hypothesis-driven for presentation). But the real added value of a scientist's input in these cases will be on the tail end of the work, in the choice of analytical tools employed or in the biological insight brought to interpretation of the patterns and signatures that emerge from the results.

Uncomfortable, because of the sheer magnitude and complexity of the data generated. An experimentalist is quite comfortable when the microarray analysis extracts a 'Top-30' list of the genes varying between two conditions; in contrast, the human brain is poorly equipped to handle the multidimensionality that results from broader meta analyses, to apprehend the full 45 000 datapoints of a DNA chip. With humbling ease, our machines can perform complex computations on large matrices of microarray data, but we are largely unable to fully grasp and conceptualize the results. When handed their first microarray data sets, investigators exhibit a rather stereotyped behavioral pattern, not knowing how to tackle the sheer scope of the data, unsure of from which end to begin to bite the large data file. This situation is reminiscent of Adams' 'Total Perspective Vortex', the imaginary machine

in which individuals become insane, because they are exposed to the full complexity and unimaginable infinity of the universe. Similarly, we have difficulty in formulating the actual questions to ask of our computers; we are conscious that there must be a deeper truth to extract from the mountains of data, but we do not know how to phrase the query to extract it. Analyses often abort with the production of a paltry gene list.

Daunting, because the tools required for global analyses are arcane and unfamiliar. This description does not refer to the analytical techniques themselves, which are little more than massively parallel versions of the basic tools of biochemistry and molecular biology with which immunologists have long been comfortable, but refers to the computational tools that are most foreign, requiring a background in mathematics, statistics, and computer science that are very much absent from the usual immunology curriculum. The obscurities of the bioinformatics discipline remain largely closed to us (perhaps a just payback, because immunologists have long been reputed for their own arcane jargon).

Results from broad analyses of complex data sets require representations that cannot be captured on paper, or even within simple electronic data files. The paper printout of a complex network, once the first awed response has passed, is of limited value. At present, most scientific tools used for data display are still primitive when compared with the computer graphics used in other walks of life. The three dimensional rendering of even the simplest of computer games far outpaces in its visual richness the usually static and awkward illustrations of the conventional scientific literature, a limitation that web-based publishing is only slowly rectifying.

Finally, it should be admitted that practitioners of global analyses can have a hazy view of where their explorations are taking them. The experimental and intellectual path of these explorations cannot always be charted as predictably as a gene-knockout experiment. The computational tools are evolving, and with this very evolution, we are beginning to get a better sense of the questions to ask. Yet, initial haziness is an intrinsic element of scientific research, as opposed to laboratory courses. Even Popper, who repeatedly pointed out the heuristic value of explicit hypotheses, admitted to their occasionally tenuous nature: 'I am inclined to think that scientific discovery is impossible without faith in ideas which are of a purely speculative kind, and sometimes even quite hazy' (The Logic of Scientific Discovery, 1959).

Validation and evaluation

We propose that it is important for the immunological community to support such integrative work through its initial growth, particularly when the techniques, concepts, and intellectual framework for operating in this sphere are still uncertain and a work-in-progress. The new concepts and exploratory modes should not be thrown out automatically for non-conformity to traditional x <> y > z hypothesis testing. Rather, the field should look for creative ways to support the publication of this work in mainstream immunological forums, not just in specialized bioinformatics journals that focus on the technical aspects of the algorithms that process the data rather than in the kernels of biological truth that result from this processing. Nor should such publication be confined to the small niche of dedicated systems biology journals. Integrative approaches should be part of the general discourse and debate in our field, not only because of the significance of the results *per se* but also by the educational value of exposing trainees and senior investigators alike to this new way of assessing biological systems. If this opening is to be done properly, so that the chaff does not outweigh the wheat, we do need appropriate rules for the validation and evaluation of such studies.

One might argue that if the broad conclusions of integrative immunology cannot be validated by the direct experimental techniques commonly applied in x <> y > z projects, and hence are not falsifiable, systems immunology might not constitute 'real' science, in the Popperian sense. The resolution of this dilemma seems to be to think more broadly about what constitutes 'validation' or 'falsifiability'. First, within the integrative mode itself, one can apply alternate computational strategies and algorithms to determine whether a particular outcome is a 'robust' result or an artifact of the particular analytic method employed. Likewise, statistical validity, always a thorny issue with large data sets, can be established rigorously by resampling techniques or estimates of false-discovery probabilities. Second, independent data sets can be used to examine whether a global conclusion is a 'one-off' or reflects the inherent behavior of the system. As just one example, the expression signature of memory $CD8^+$ T cells has now been reproduced in several studies. Finally, while it may not be always feasible, investigators can formulate and test experimentally predictions whose outcome relies on the global interpretations they have drawn from the data. Note that an appropriate test is not the rote assessment of the functional contributions of a few selected genes.

Criteria will need to be applied to evaluate grant applications or manuscripts that propose or present 'integrative' analyses or meta analyses. It is clear to all that a paper or application whose end result is 'Here is the list of genes that distinguishes A from B' should have low priority. The

distinguishing criterion should be whether the analysis arrives at novel information that provides new insight into the structure of complex ensembles and/or the (regulatory) connections among them: the 'complex ensemble' in question could be a pathway, gene signature, or cell (in a sense, this comes back to an x <> y > z paradigm, except that factors x and y are now complex objects, rather than a single gene or gene product). It is still important, even essential, to ask: what have we learned about the relationships between immune players and about how a differentiated state is globally achieved? What insight have we gained into immune function?

The field will also need standard formats for the description and reporting of the new objects it describes. We do not mean here the deposition of raw data in public databases, a minimum condition that should of course be enforced but which does not carry with it the full value of integrative explorations of the data; rather, we need to arrive at standard vocabularies and metrics to describe signatures, relationships, or network structures. These are the core objects that will be exchanged and compared between investigators, built upon in future explorations.

In conclusion, genome-scale explorations of the immune network need better ideas, tools, and modes of representation. These are emerging, even if the procedures to be used and mindsets for evaluating the advances provided by such work are still unsettled and uncodified. We must avoid suppressing integrative explorations by requiring them to fit the common mold of single gene/protein hypothesis testing. The danger would be to hide from the very complexity that gives rise to the biology we so much desire to understand.