Title

Consortium biology in immunology: the perspective from the Immunological Genome Project.

Journal

Nature reviews. Immunology, 12(10)

ISSN

1474-1733

Authors

Benoist, Christophe
Lanier, Lewis
Merad, Miriam
et al.

Publication Date

2012-10-01

DOI

10.1038/nri3300

Peer reviewed

39. Kalliomaki, M. *et al.* Probiotics in primary prevention of atopic disease: a randomised placebo-controlled trial. *Lancet* **357**, 1076–1079 (2001).
40. Taylor, A. L., Dunstan, J. A. & Prescott, S. L. Probiotic supplementation for the first 6 months of life fails to reduce the risk of atopic dermatitis and increases the risk of allergen sensitization in high-risk children: a randomized controlled trial. *J. Allergy Clin. Immunol.* **119**, 184–191 (2007).
41. Abrahamsson, T. R. *et al.* Probiotics in prevention of IgE-associated eczema: a double-blind, randomized, placebo-controlled trial. *J. Allergy Clin. Immunol.* **119**, 1174–1180 (2007).
42. Wickens, K. *et al.* A differential effect of 2 probiotics in the prevention of eczema and atopy: a double-blind, randomized, placebo-controlled trial. *J. Allergy Clin. Immunol.* **122**, 788–794 (2008).
43. Hurree, A. *et al.* Impact of maternal atopy and probiotic supplementation during pregnancy on infant sensitizations: a double-blind, placebo controlled study. *Clin. Exp. Allergy* **38**, 1342–1348 (2008).
44. Dotterud, C. K., Storro, O., Johnsen, R. & Oien, T. Probiotics in pregnant women to prevent allergic disease: a randomized, double-blind trial. *Br. J. Dermatol.* **163**, 616–623 (2010).
45. Boyle, R. J. *et al. Lactobacillus* GG treatment during pregnancy for the prevention of eczema: a randomized controlled trial. *Allergy* **66**, 509–516 (2011).
46. Niers, L. *et al.* The effects of selected probiotic strains on the development of eczema (the PandA study). *Allergy* **64**, 1349–1358 (2009).
47. Soh, S. E. *et al.* Probiotic supplementation in the first 6 months of life in at risk Asian infants – effects on eczema and atopic sensitization at the age of 1 year. *Clin. Exp. Allergy* **39**, 571–578 (2009).
48. West, C. E., Hammarstrom, M. L. & Hernell, O. Probiotics during weaning reduce the incidence of eczema. *Pediatr. Allergy Immunol.* **20**, 430–437 (2009).
49. Kim, J. Y. *et al.* Effect of probiotic mix (*Bifidobacterium bifidum*, *Bifidobacterium lactis*, *Lactobacillus acidophilus*) in the primary prevention of eczema: a double-blind, randomized, placebo-controlled trial. *Pediatr. Allergy Immunol.* **21**, e386–e393 (2009).
50. Gruber, C. *et al.* Reduced occurrence of early atopic dermatitis because of immunoactive prebiotics among low atopy risk infants. *J. Allergy Clin. Immunol.* **126**, 791–797 (2010).
51. Moro, G. *et al.* A mixture of prebiotic oligosaccharides reduces the incidence of atopic dermatitis during the first six months of age. *Arch. Dis. Child.* **91**, 814–819 (2006).
52. Gounaris, E. *et al.* T regulatory cells shift from a protective anti-inflammatory to a cancer-promoting proinflammatory phenotype in polyposis. *Cancer Res.* **69**, 5490–5497 (2009).
53. Blatner, N. R. *et al.* In colorectal cancer mast cells contribute to systemic regulatory T cell dysfunction. *Proc. Natl Acad. Sci. USA* **107**, 6430–6435 (2010).
54. Khazaie, K. *et al.* The significant role of mast cells in cancer. *Cancer Metastasis Rev.* **30**, 45–60 (2011).
55. Claes, I. J. J., De Keersmaecker, S. C. J., Vanderleyden, J. & Lebeer, S. Lessons from probiotic–host interaction studies in murine models of experimental colitis. *Mol. Nutr. Food Res.* **55**, 1441–1453 (2011).
56. Sartor, R. B. Efficacy of probiotics for the management of inflammatory bowel disease. *Gastroenterol. Hepatol.* **7**, 606–608 (2011).
57. Claes, I. J. *et al.* Impact of lipoteichoic acid modification on the performance of the probiotic *Lactobacillus rhamnosus* GG in experimental colitis. *Clin. Exp. Immunol.* **162**, 306–314 (2010).
58. Khazaie, K. *et al.* Abating colon cancer polyposis by *Lactobacillus acidophilus* deficient in lipoteichoic acid. *Proc. Natl Acad. Sci. USA* **109**, 10462–10467 (2012).
59. Ritchie, M. L. & Romanuk, T. N. A meta-analysis of probiotic efficacy for gastrointestinal diseases. *PLoS ONE* **7**, e34938 (2012).
60. Ligaarden, S. C., Axelsson, L., Naterstad, K., Lydersen, S. & Farup, P. G. A candidate probiotic with unfavourable effects in subjects with irritable bowel syndrome: a randomised controlled trial. *BMC Gastroenterol.* **10**, 16 (2010).
61. Pelucchi, C. *et al.* Probiotics supplementation during pregnancy or infancy for the prevention of atopic dermatitis: a meta-analysis. *Epidemiology* **23**, 402–414 (2012).
62. Arroyo, R. *et al.* Treatment of infectious mastitis during lactation: antibiotics versus oral administration of lactobacilli isolated from breast milk. *Clin. Infect. Dis.* **50**, 1551–1558 (2010).
63. Pineda Mde, L. *et al.* A randomized, double-blinded, placebo-controlled pilot study of probiotics in active rheumatoid arthritis. *Med. Sci. Monit.* **17**, CR347–CR354 (2011).
64. Lewis, S., Burmeister, S. & Brazier, J. Effect of the prebiotic oligofructose on relapse of *Clostridium difficile*-associated diarrhea: a randomized, controlled study. *Clin. Gastroenterol. Hepatol.* **3**, 442–448 (2005).
65. Depeint, F., Tzortzis, G., Vulevic, J., I'anson, K. & Gibson, G. R. Prebiotic evaluation of a novel galactooligosaccharide mixture produced by the enzymatic activity of *Bifidobacterium bifidum* NCIMB 41171, in healthy humans: a randomized, double-blind, crossover, placebo-controlled intervention study. *Am. J. Clin. Nutr.* **87**, 785–791 (2008).
66. Silk, D. B., Davis, A., Vulevic, J., Tzortzis, G. & Gibson, G. R. Clinical trial: the effects of a *trans*-galactooligosaccharide prebiotic on faecal microbiota and symptoms in irritable bowel syndrome. *Aliment. Pharmacol. Ther.* **29**, 508–518 (2009).
67. Besselink, M. G. *et al.* Probiotic prophylaxis in predicted severe acute pancreatitis: a randomised, double-blind, placebo-controlled trial. *Lancet* **371**, 651–659 (2008).
68. DuPont, A. W. & DuPont, H. L. The intestinal microbiota and chronic disorders of the gut. *Nature Rev. Gastroenterol. Hepatol.* **8**, 523–531 (2011).
69. Gareau, M. G., Sherman, P. M. & Walker, W. A. Probiotics and the gut microbiota in intestinal health and disease. *Nature Rev. Gastroenterol. Hepatol.* **7**, 503–514 (2010).
70. Guigoz, Y., Doré, J. & Schiffrin, E. J. The inflammatory status of old age can be nurtured from the intestinal environment. *Curr. Opin. Clin. Nutr. Metab. Care* **11**, 13–20 (2008).
71. Atarashi, K. *et al.* Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* **331**, 337–341 (2011).
72. Kleerebezem, M. *et al.* The extracellular biology of the lactobacilli. *FEMS Microbiol. Rev.* **34**, 199–230 (2010).
73. Khazaie, K. *et al.* Abating colon cancer polyposis by *Lactobacillus acidophilus* deficient in lipoteichoic acid. *Proc. Natl Acad. Sci. USA* **109**, 10462–10467 (2012).

Acknowledgements
M.R. is supported by the European Research Council, the European Commission (FP7: IBDase, MetaHIT), the Italian Ministry of Health, the Association for International Cancer Research and the Italian Association for Cancer Research.

Competing interests statement
The authors declare competing financial interests: see Web version for details.

FURTHER INFORMATION
Todd R. Klaenhammer's homepage: http://www.cals.ncsu.edu/food_science/KlaenhammerLab
Michiel Kleerebezem's homepage: http://www.hmi.wur.nl
Matthias Volkmar Kopp's homepage: http://www.kinderklinik-luebeck.de/kinderpneumologie
Maria Rescigno's homepage: http://www.ifom-ieo-campus.it/research/rescigno.php
ALL LINKS ARE ACTIVE IN THE ONLINE PDF

SCIENCE AND SOCIETY

Consortium biology in immunology: the perspective from the Immunological Genome Project

Christophe Benoist, Lewis Lanier, Miriam Merad, Diane Mathis and the Immunological Genome Project

Abstract | Although the field has a long collaborative tradition, immunology has made less use than genetics of 'consortium biology', wherein groups of investigators together tackle large integrated questions or problems. However, immunology is naturally suited to large-scale integrative and systems-level approaches, owing to the multicellular and adaptive nature of the cells it encompasses. Here, we discuss the value and drawbacks of this organization of research, in the context of the long-running 'big science' debate, and consider the opportunities that may exist for the immunology community. We position this analysis in light of our own experience, both positive and negative, as participants of the Immunological Genome Project.

At the outset, we should define what is meant here by 'consortium biology': a research programme led by a complementary set of laboratories or institutions, all working towards a common and well-defined goal. This common goal could not be achieved by any one participant, either because of its magnitude or because it requires multidisciplinary input. This definition does not encompass the more common collaborative groupings, in which independent projects are linked under a common thematic umbrella but remain mostly autonomous. Nor does it include the groups that coalesce to conform to politically or administratively imposed frameworks, rather than from objective scientific justification. In general, consortium biology can be considered to be 'big science' (BOX 1) and usually corresponds to discovery-led

research, which is distinct from hypothesis-driven investigations that are performed in individual laboratories and have formed much of the biology research enterprise of the past century. This correlation is not always true, and one could argue that the search for the Higgs boson — possibly the largest and most expensive example of consortium science — was testing the hypothetical existence of the particle.

Because of its ultimate significance, the Human Genome Project (1990–2003) is probably the pinnacle of consortium biology and the most obvious and profound success of this approach. Its results permeate and fertilize most of modern biology. Other examples of successful genome-oriented projects of well-defined scope are the HapMap Project, which aims to catalogue genetic differences in individuals through the systematic discovery of single-nucleotide polymorphisms or other variations in the human genome, as well as the large consortia that use this knowledge to chart genetic susceptibility to specific diseases. Similarly, the International Cancer Genome Consortium coordinates large-scale genomic studies of diverse cancer types[1]. The Allen Brain Atlas is another example, which integrates extensive gene expression and neuroanatomical data[2].

Immunology, as a whole, has perhaps engaged in consortium science less than other fields. Genetics has been far more active, or at least more visible, in this respect, partly because its object of study — the genome — can be parsed into discrete objects (such as genes, base pairs and mutations), giving it a natural scope for descriptive mapping that is missing from the more physiological branches of biology. Indeed, it is more straightforward to split between participants a project to identify, map and/or sequence specific genes than to organize a concerted approach to a complex process such as immune tolerance. However, there are precedents in immunology (TABLE 1). The Histocompatibility Workshops in the 1960s, 1970s and 1980s (now continued as the International Histocompatibility Working Group) were of extraordinary importance in deciphering the bewildering allelic variation in MHC loci and in defining a coherent framework for MHC nomenclature, a particular feat given that the serological and cellular reagents available at that time were far from specific. The accuracy of the data generated by this consortium was later confirmed when sequence data became available. The success came from an organization that fostered community participation, structured reagent exchange, carefully optimized operating protocols and the validation of the results from each laboratory using common sets of reagents. One can only imagine what might have occurred if (and this is not an implausible 'if') the work had been left to individual laboratories, using different sets of typing cell lines and alloreactive sera, competing for publications and naming rights.

Modelled on the Histocompatibility Workshops, the Human Leucocyte Differentiation Antigens (HLDA) Workshops (1982–present) sorted out the confusion that existed in the early 1980s when it was unclear which monoclonal antibodies were detecting the same cell-surface antigens and each laboratory held steadfastly to a different nomenclature for these antigens. Although the staid CD nomenclature enforced by the HLDA workshop lacked the poetic creativity of *Drosophila* geneticists, the effort provided a rational framework for the identification of the antigens expressed on cells of the immune system, before the cloning of the corresponding genes.

The Immune Epitope Database is a coordinated effort to define and catalogue in a systematic manner the B and T cell epitopes that dominate responses to microorganisms and other immunological insults, in relation to MHC alleles[3]. More recent has been the initiation of the Immunological Genome Project (ImmGen), a large resource effort aimed at deciphering gene expression and its regulation across the entire immune system, of which we are participants. It is from this perspective that we discuss various aspects of consortium science in general, and its potential in immunology.

Box 1 | **Big science: the pros and cons**

The pros

- For some large questions, big science is simply the only way of getting to the result.
- It promotes a culture of openness, through data and material sharing.
- Collaboration between groups makes, through collective wisdom, for much sounder decisions than would be made by any one individual.
- It promotes uniform standards, data and data formats that are compatible between laboratories, as well as the use of common nomenclature and shared reagents.
- Its results can diffuse broadly, and extend and enrich the 'little science' that follows.
- "Because it's there" (George Mallory). Several years before the lunar landings, Weinberg decried the productions of big science (large accelerators and space exploration) as costly monuments of our era, akin to the Pyramids or cathedrals of past civilizations. Such monuments were products of the culture of their times, but they also contributed, through the economic distortion that they caused, to the decline of the civilizations. But, surely, finding a fundamental particle of the Universe or deciphering the human genome has inspirational value at the individual and societal level that transcends any usual science project.
- It is economical. Through direct economies of scale, high-throughput approaches are less costly than smaller operations. In addition, accelerated technological change can result in more favourable data to cost ratios for all (as was the case for DNA sequencing).

The cons

- Big science competes for limited funding.
- Big science stifles scientific creativity by promoting 'science by committee', fads and 'herd thinking'.
- It can lead to 'closed-shop' situations, in which the existence of a large collaborative group effectively blocks alternative initiatives and actually restricts the spread of technology.
- It establishes self-perpetuating structures that tend to create projects to ensure their own survival, rather than for clear scientific need.
- It increases the number and importance of science administrators, creating "science understood by administrators"[4] who "love them [big science projects] because they produce results that can be easily summarized to politicians"[6], to the detriment of the creativity of individual scientists.
- More data are generated than can be analysed. Data are underexploited and not analysed as carefully as they might be[23], as is certainly the case today for the flood of genomic data.
- It is wasteful ("spending money instead of thought"[4]). In scoping large projects, less attention is paid to the value of some analyses than would be the case in small operations.

Big science, little science

As noted above, consortium biology naturally falls into the category of big science. The debate over the advantages and dangers of big science relative to the more individual

Table 1 | **Past and current consortium immunology projects**

Project	Dates	Participants and structure	Core aims	Key achievements
Histocompatibility Workshops	1964–present	Up to 350 laboratories; early leaders included: B. Amos, R. Payne, J. Dausset, R. Ceppellini, J. Bodmer, W. Bodmer, P. Terasaki, J. van Rood and many others	Uniformly define human histocompatibility antigens and their influence; continues today as the IHWG, mapping KIRs and other receptor families	Identification and cataloguing of HLA molecules and their allelic variants, through serology, T cell alloactivation and DNA sequencing
Human Leucocyte Differentiation Antigens (HLDA) Workshops	1982–present	>40 international laboratories	Characterize the structure, function and distribution of leukocyte surface molecules of the immune system	CD nomenclature and the harmonization of markers on immune cells (as of 2012, up to CD360)
Immune Tolerance Network	2002–present	14 core facilities for sample tracking and analysis, 175 participating centres	Accelerate the clinical development of immune tolerance therapies through collaborative multicentre trials and a rigorous support infrastructure	Conducting >40 clinical trials with mechanistic studies encompassing transplantation, allergy and autoimmune diseases
Human Microbiome Project	2005–present	200 laboratories, 80 institutions, 5 sequencing centres	Define the composition of the human microbiome and its role in health and disease	In progress; setting standards for microbiome analysis was an important first step
The Immunological Genome Project	2007–present	15 immunology and computational biology laboratories	Chart gene expression and its regulation across the immune system	In progress
Euroflow	2007–present	8 European laboratories supported by the EU	Develop and standardize flow cytometric tests for diagnosis and prognostic classification of haematological malignancies	In progress
Human Immunology Project Consortium	2010–present	7 laboratories sponsored by NIAID	Create a new public resource of data of different types that characterize diverse states of the human immune system	In progress

EU, European Union; FOCIS, Federation of Clinical Immunology Societies; IHWG, International Histocompatibility Working Group; KIR, killer cell immunoglobulin-like receptor; NIAID, US National Institute of Allergy and Infectious Diseases.

practice of research has been a constant thread for the past 50 years. The term 'big science' may have been coined in an editorial by Alvin Weinberg that was published in 1961 in the journal *Science*[4], in which he raised concerns regarding the financial and educational implications of big science approaches for particle physics and manned space exploration. The arguments raised then were echoed in the 1990s against the Human Genome Project[5], and more recently against systematic structural biology programmes[6]. These arguments — the advantages and drawbacks of big science relative to the more conventional practice of research (BOX 1) — have remained strikingly relevant over the past 50 years (with the exception, perhaps, of Weinberg's fear that the escalating costs of scientific research would derail national finances).

The main concerns with big science include: the stifling of scientific creativity; the establishment of self-perpetuating institutions, clubs or bureaucracies; and wasteful spending. At the same time, opponents of big science have usually agreed that it is here to stay and that it does have justification in some instances, if it is carefully balanced such that the creativity of 'little science' is not curtailed. As stated by Gregory Petsko[6], "the best kind of big science is the kind that supports and generates lots of good little science". The interface between big science projects and the wider scientific community may be, if properly managed, an important means to avoid several of the pitfalls of big science.

The big science debate also speaks to what it means to be a research biologist and what personal quest motivates the passion and the long hours in the laboratory or at the computer. To a large extent, biological research is still the quest for 'discovery', that climactic moment in which a veil is lifted, the puzzle pieces fall into place and the inherent beauty of the natural order is revealed. Eureka moments are rare in consortium biology. No one ever 'discovered' the human genome or the HLA alleles. Rather, the personal sense of accomplishment for participants comes from having contributed to a body of knowledge that is greater than that derived from any one experiment.

An issue with consortium biology is how to acknowledge participants' contributions to the project in publications, as this has a direct impact on careers and funding. In the scientific literature, the semantics of first, last and middle authorships (and their asterisked variants) are well understood, and provide clear indicators of each individual's contribution for evaluation bodies. But in a multi-author publication, the 23rd position may actually denote a seminal and creative contribution by one of the programme's teams, a contribution essential to the success of the whole. The interpretation of authorship needs revisiting in the context of consortium biology. Interestingly, the US National Institutes of Health (NIH) intramural programme has written into its career promotion rules that equivalent credit can be received from a senior-authored paper and from co-authorship on a consortium project. In addition, large-scale collaborative programmes have generated novel career paths for biologists who work as staff scientists rather than as principal investigators leading a small laboratory. Contributions from these scientists are essential to the viability of the programmes and the performance and continued improvement of their scientific platforms.

The Immunological Genome Project

Aims. ImmGen, initiated in 2007, is a collaborative group of 15 immunology and computational biology laboratories that have sought to perform, under carefully

standardized conditions, a thorough dissection of gene expression in the immune system[7]. In the first phase, the group focused on generating gene expression profiles for primary haematopoietic cells from different anatomical locations that were analysed directly *ex vivo* at finely defined stages of differentiation. Each participating laboratory was responsible for the definition and preparation of the cell populations within their lineage of interest and expertise. In the second phase, ImmGen computational biologists applied reverse engineering techniques to define modules of co-regulated genes and the regulatory programmes that drive the changes in their expression during immune cell differentiation. These baseline data are now being extended through the analysis of immune cells in challenged conditions (such as infections and tumours), and by testing the behaviour of the regulatory network after genetic or pharmacological perturbations.

Achievements. Much of the value of a project of this nature lies in generating data that are consistent across laboratories, and the organization of this project was defined from the start with this goal in mind. With only a few exceptions, all of the mice used were genetically identical and sourced from a single location, and tissue was harvested at a fixed time to avoid circadian variation. All RNAs were processed for profiling at a single location, and bioinformatic quality control was centralized. Cell preparation for sorting primarily involved reagents from a single source and common standard operating procedures (with particular attention to the preparation time between the tissue harvest and the final sort). All preparations were double-sorted, and qualification analyses were performed for all laboratories (although telling investigators with >20 years of experience in flow cytometry that their sorts would need vetting did raise a few hackles).

Overall, the expression datasets generated by ImmGen provide an extraordinarily broad picture of the immune system, the relationships between its lineages and the distribution of functions between the members of large multigene families. Only in the field of immunology could such relationships have been analysed, as no other biological field has resolved the differentiation steps of its component cell lineages to such a degree and possesses sets of reagents for such parsing (although, it must be noted that immunology benefits from dealing with cells that are mostly unattached and readily accessible). In addition, the high granularity of the ImmGen data (>250 cell types were analysed) has illuminated the distinctions between related cell subsets, working from the assumption that cells closely positioned in a lineage differentiation pathway are also closely related from a transcriptional standpoint. These results are discussed in detail in a series of reports from the ImmGen consortium (REFS 8–11 and manuscripts submitted for publication) and are not detailed here.

As examples, the positions in lineage differentiation pathways of controversial B cell intermediates have been clarified, well-distinguished subpopulations of natural killer T (NKT) cells and γδ T cells have been defined, and unexpected relationships between NKT cells, γδ T cells and memory $CD8^+$ T cells have been uncovered[8]. The results have perhaps been most illuminating for myeloid cells, for which lineage relationships between subsets of dendritic cells (DCs)[10], macrophages and monocytes have long been elusive, in part owing to the use of limited panels of cell-surface markers. Indeed, from the ImmGen data, some populations were reclassified between DCs and macrophages, and more selective markers were identified.

The analyses by each laboratory in the consortium, focused on their lineage of interest, benefited greatly from the uniformity of the data and from the opportunity to determine cell type-specific signatures against the backdrop provided by all the other lineages. It is clear that no single laboratory could have managed, or would have ever wanted to undertake, such a task on its own. The success of the group is derived from this combination of particular interests, and it is while pursuing, and motivated by, their individual research areas that ImmGen investigators collectively assembled the whole.

The extensive documentation of cell lineage differentiation pathways achieved by ImmGen provided a unique setting for computational analysis and required new algorithms to exploit it. The juxtaposition of bench researchers and computational biologists in the project made for enriching experiences, with each side acquiring a new understanding of a field that was *a priori* quite foreign. Indeed, some of the analysis 'jam sessions' even inspired some immunology fellows to pursue computational career tracks. But there was also a surprising realization of the depth of differences in mindset and thought processes between the disciplines. Indeed, some of the exchanges between computational biologists and immunologists proved trickier than anticipated, in part because what seemed like simple requests were actually far more complex than perceived. The semantic rigour of computational processing is not a natural trait for immunologists (for example, there was some difficulty in adopting a common and coherent construction for the 'code names' of the cell types profiled). This incomprehension was probably more common for immunologists than for the geneticists who work with computational biologists in other large projects, as geneticists have long applied mathematical and computational frameworks in their studies. This realization reinforced the importance of training immunology students in computational approaches, tools and mindset.

Opportunities and implications. ImmGen data and metadata are proving to be valuable resources for the immunology community. In addition to the pre-publication release of basic data through public repositories (such

Glossary

Big science
Scientific research that involves larger instruments or groups of scientists than that more commonly practised in individual laboratories.

Crowdsourcing
A process in which a task is performed, typically in small subfractions, by a large group of people who are *a priori* undefined and not affiliated to the initiating entity.

Glycomes and lipidomes
The complete sets of polysaccharides (free or complexed) and lipids expressed in a cell or organism.

Proteome
By analogy to the genome (the complete set of genes), the proteome is the complete set of proteins expressed in a cell or organism, and their post-translational modifications.

Reverse engineering
The process of discovering the operational principles of a device or system of unknown structure through analyses of its function and operation. In the analysis of genetic regulatory networks, one starts from the end result of the regulatory network (a large number of measures of gene expression in different cells, with or without perturbation) and computationally infers which regulatory inputs can generate these results. It often involves taking a system apart and analysing its workings with the aim of making a new device or programme that does the same thing without using any physical part of the original.

Text mining
Deriving information from computational analyses of patterns in texts. In biology, text mining refers to discovering relationships between biological objects from the patterns of co-occurrence in abstracts or texts of published articles.

as the National Center for Biotechnology Information (NCBI)), ImmGen has developed and implemented several new tools to publicly display the data, through dedicated web and smartphone interfaces. These portray the basic expression data, the relationships between genes, the transcripts that distinguish cell types or groups thereof, and coordinated gene clusters and their regulators. The ImmGen data browsers allow cross-lineage comparisons that would not have been possible as a result of any single project. The true impact of the ImmGen data is hard to assess, as the usual metrics of citation number or impact factor do not apply. Database downloads, web traffic and application software usage constitute the new metrics of impact (as of mid-2012, ~300 independent visits are made to the ImmGen server every weekday, from origins that correlate with the worldwide density of immunology research). From these statistics, it is plausible that some fruitful hypotheses have been generated in the community, and that a number of exploratory studies by quantitative PCR have been avoided. In this sense, Petsko's requirement that big science supports little science is perhaps fulfilled.

Consortium biology in immunology?

If we accept the notion that consortium biology is desirable for some questions or explorations, while recognizing and limiting the risks associated with big science, are there areas of immunology in which it would be fruitful? What are the best bets for new opportunities in 'big immunology'? One might delineate three types of endeavour: comprehensive mapping, comprehensive information and comprehensive resources.

Comprehensive mapping. A consortium approach could be used to establish complete charts of all the molecular components that are active in cells of the immune system.

Beyond the definition of coding mRNAs, which has been the focus of the ImmGen project to date, there are many other molecular components — such as microRNAs and non-coding RNAs — that need to be measured to obtain a complete immune regulatory map and to provide a true understanding of the events that underlie immune responses. The main characteristic of the immune system is its capacity to respond to diverse challenges, and a system-wide exploration of responses to pathogens and other triggers, using concerted and highly parallel analyses, will be essential.

Beyond the transcriptome, recent advances in the sensitivity and resolution of proteomic analyses make similar system-wide analyses of the proteome plausible in the near future, with the glycomes and lipidomes perhaps more distant frontiers. However, such extensive system-wide analyses raise the question of how much data we really need, or how much is too much (BOX 2).

Comprehensive mapping of human genetic diversity has been tackled by HapMap, ENCODE (Encyclopedia of DNA Elements) and other large-scale genetic mapping projects. More specific to immunology is the analysis of variation in immune receptor structure and expression, both on an evolutionary timescale for innate receptors (such as NK cell receptors) and on an organismal timescale for T cell receptors and B cell receptors. Analyses of the complexity of copy-number and allelic variations in NK cell receptors are already underway and are being tackled as part of the International Histocompatibility Working Group and HLDA consortia. For adaptive lymphocytes, the potential offered by high-throughput sequencing to analyse B and T cell repertoires would greatly benefit from coordinated and controlled efforts. It will be essential that we follow the example of the microbiome sequencing centres[12,13], which invested many months in collectively deriving robust and reproducible techniques[14] after realizing that their early sequencing data for bacterial community profiling were very divergent. The field of antigen receptor repertoire sequencing will collectively need to define optimized and standardized procedures that avoid the technical traps (such as uneven amplification, the generation of chimeric sequences and other PCR and sequencing errors masquerading as diversity) and generate robust data on somatically rearranged antigen receptor repertoires that are as comparable between laboratories and centres as genomic DNA sequences.

There also has been much recent interest in using the potential of systems-level analyses to bolster the study of human immunology[15,16]. Much of our knowledge of immune system organization and function stems from experimentation in mice, and although dissimilarities between human and mouse systems may have been overstated, there is a consensus that the field needs a better and more direct handle on human immunology. The application of high-information technologies might, to some extent, help to alleviate the key hurdles faced by human immunology, such as high genetic and environmental variability, restrictions to tissue access and ethical limitations to possible experimentation. Systems-level data might provide signatures that help to smooth out inter-individual variation and to establish blood biomarkers for events occurring deep in tissues that cannot be accessed. But achieving these goals will require a level of coordination, protocol sharing and standardization that the field has not yet reached. Such standards are not straightforward to implement in multicentre contexts, as shown by the experience at the Immune Tolerance Network[17], which coordinates clinical trials for immune and autoimmune diseases and which had to surmount daunting challenges with procedural reproducibility between study sites, even for such apparently trivial procedures as freezing peripheral blood mononuclear cells (PBMCs). In addition, there are currently no generally accepted reagent panels

Box 2 | **How much data is enough, or too much?**

In a forward-looking opinion piece that was written before the completion of the Human Genome Project, Eric Lander argued for a global view of biology, with a prescription for a complete and coordinated view of DNA, RNA and protein for every one of the 100,000 genes expected to be discovered[24]. But how much data do we truly need? To understand a biological function, is it necessary to know the behaviour of every single member of a multigene family? Are the broad outlines of the molecular changes that take place not sufficient to correctly understand a biological phenomenon, or alter it therapeutically? In addition, the levels of complexity that would be achieved by such a complete and quantitative description would be far beyond the capacity of the human mind to encompass. Only a supercomputer (or a cluster thereof) would have the power to do so. But, then, do we "know" something, if this knowledge is only within our machines?

However, the goal of a global view is probably as unavoidable in biology as in other fields of human discovery. Would ancient explorers have been content to sit by their seas, rivers and mountains, on the grounds that it would just be more of the same elsewhere? Would astronomers have been satisfied with mapping the major constellations, ignoring what lies beyond?

Curiosity and an innate need to reveal the hidden will always fuel the quest to unturn every stone, and the amount of information considered 'reasonable' obviously evolves over time. The right amount of information may be the amount that allows general conclusions at the level of the entire system, at the highest level of resolution, without leading to distraction over minute details.

Table 2 | **Past and current consortium immunology resources and databases**

Resource or database	Core aims	Web link
International Histocompatibility Working Group (IHWG)	Reference panels of typing cell lines and DNA	http://ihwg.org/index.html
Tetramer Core Facility at Emory	Panels of peptide–MHC multimer reagents to identify antigen-specific T cells	http://www.tetramer.yerkes.emory.edu
The Biodefense and Emerging Infections Research Resources Repository (BEI)	Cultures, reagents and information for microbiology and infectious disease research	http://www.beiresources.org
Mutagenetix	Information and sourcing for ENU-induced mouse mutants	http://www.mutagenetix.org
Immune Epitope Database	Catalogues of antibody and T cell epitopes related to infectious diseases, allergens and autoimmune diseases	http://www.immuneepitope.org
The Immunological Genome Project	Gene expression profiles for >200 cell types of immunological interest, allowing for the identification of genes distinguishing cells or groups of cells, and of co-regulated genes and their predicted regulators	http://www.immgen.org

ENU, *N*-ethyl-*N*-nitrosourea.

for basic immune profiling of blood lymphocyte populations[18]. Consortium approaches would be very valuable here, and several efforts — Euroflow, the FOCIS Human ImmunoPhenotyping Consortium and the Human Immunology Project Consortium — are underway, which will hopefully be harmonized.

Comprehensive information. Data are not knowledge. The translation and aggregation of molecular or functional data (from systematic data collections or by extracting data from the published literature) into a form that is both comprehensible and pertinent to immunology could also be a collective project. A goal might be to add a layer of immunological relevance that is usually missing from the excellent data aggregation that is performed generically by the NCBI, European Bioinformatics Institute (EBI) and RIKEN Institutes Database. For instance, the summary on the entry for *IL2RA* (interleukin-2 receptor α-chain; also known as CD25) in the NCBI gene database states that: "The interleukin 2 (IL2) receptor alpha (IL2RA) and beta (IL2RB) chains, together with the common gamma chain (IL2RG), constitute the high-affinity IL2 receptor". Although this is correct, this description misses the rich knowledge of the role of CD25 in memory cell formation, regulatory T cell differentiation and NK cell activation. Higher-level integration is harder than simple molecular descriptions. Creating such immunological annotations, which is potentially a monumental task, will need creative solutions, perhaps combining directed curation with text mining, Wikipedia-like crowdsourcing and/or competitions.

Comprehensive resources. The third application for consortium approaches is the coordinated generation of sets of common research tools. Several efforts of this nature exist already (TABLE 2). Beyond the mutant mice maintained by the generic international repositories, more specific to immunology are the *N*-ethyl-*N*-nitrosourea (ENU)[19,20] and gene-trap mutagenesis programmes that target immune phenotypes. These programmes (such as Mutagenetix and the Centre for Modelling Human Disease) are beginning to make available the sperm of mice with identified mutations in a wide array of immunologically relevant genes. The Biodefense and Emerging Infections Research Resources Repository (BEI) provides microorganisms and reagents for research into host–microorganism interactions. On the reagent side, the NIH Tetramer Core Facility at Emory University, Atlanta, Georgia, USA, provides a well-controlled and evolving set of MHC tetramer reagents for specific T cell detection.

One might contend that the immunology community missed the opportunity to fully exploit the potential of monoclonal antibodies as shared research tools. Attempts in academic laboratories in the early 1980s to generate comprehensive panels were quickly swamped by the magnitude of the task or were not appropriately funded, and it is only after a significant delay that the for-profit sector saw the potential of large monoclonal antibody panels. Much waste could have been avoided if a strong NIH-sponsored programme, emulating the Histocompatibility Workshops, had coordinated the assembly and distribution of comprehensive collections of monoclonal antibody reagents. Such a need for community-based coordination may become even more important given the ever-increasing capability of cytometry techniques for multiplex analyses[21,22].

Concluding remarks

Overall, the feeling from the experience of the ImmGen group is that, provided that the shoals of big science are kept in mind and carefully navigated, there is valuable potential in projects of this nature, and we aim to decipher the diversity of the immune system, at the level of genes, molecules, cells and individuals.

Christophe Benoist and Diane Mathis are at the Division of Immunology, Department of Microbiology and Immunobiology, Harvard Medical School, Boston, Massachusetts 02115, USA.

Lewis Lanier is at the Department of Microbiology and Immunology and the Cancer Research Institute, University of California, San Francisco, San Francisco, California 94143, USA.

Miriam Merad is at the Department of Oncological Sciences, 1425 Madison Avenue, Mount Sinai School of Medicine, New York, New York 10029, USA.

The Immunological Genome Project: Jeff Ericson, Michio Painter, Scott Davis, Catherine Laplace, Gordon Hyatt, Henry Paik, Katie Rothamel, Richard Cruse, Graeme Doran, Tracy Heng, Natasha Asinovski, Adriana Ortiz-Lopes, Ayla Ergun, Daniel Gray, Ei Wakamatsu, Jonathan Hill, Michael Mingueneau, Daniela Cipoletta, Hideyuki Yoshida, Christophe Benoist, Diane Mathis, Nadia Cohen, Edy Kim, Patrick Brennan, Lydia Lynch, Michael Brenner, James Costello, Jim J. Collins, David Blair, Michael Dustin, Jamie Knell, Edward Yang, Adam Best, Laura Shaw, Andrew Doedens, Ananda Goldrath, Susan Shinton, Yan Zhou, Randy Hardy, Vladimir Jojic, Sara Mostafavi, Daphne Koller, Radu Jianu, David Laidlaw, Natalie Bezman, Joseph Sun, Yanan Zhu, Deborah Hendricks, Yosuke Kamimura, Gundula Min-Oo, Deborah Hendricks, Maelig Morvan, Yosuke Kamimura, Tsukasa Nabekura, Viola Lam, Charles Kim, Lewis Lanier, Melanie Greter, Julie Helft, Andrew Chow, Milena Bogunovic, Arthur Mortha, Jeremy Price, Daigo Hashimoto, Jennifer Miller, Priyanka Sathe, Aleksey Chudnovskiy, Yonit Lavin, Juliana Idoyaga, Miriam Merad, Emmanuel Gautier, Claudia Jakubzick, June D'Angelo, Gwendolyn Randolph, Tal Shay, Aviv Regev, Roi Gazit, Derrick Rossi, Taras Kreslawsky, Harald von Bohmer, Angelique Bellemare-Pelletier, Kutlu Elpek, Lotte Spelv, Anne Fletcher, Deepali Malhotra, Viviana Cremasco, Shannon Turley, Francis Kim, Tata Nageswara Rao and Amy Wagers

Correspondence to C.B. and D.M.
e-mail: cbdm@hms.harvard.edu

doi:10.1038/nri3300
Published online 7 September 2012

1. Hudson, T. J. *et al.* International network of cancer genome projects. *Nature* **464**, 993–998 (2010).
2. Jones, A. R., Overly, C. C. & Sunkin, S. M. The Allen Brain Atlas: 5 years and beyond. *Nature Rev. Neurosci.* **10**, 821–828 (2009).
3. Vita, R. *et al.* The immune epitope database 2.0. *Nucleic Acids Res.* **38**, D854–D862 (2010).
4. Weinberg, A. M. Impact of large-scale science on the United States: big science is here to stay, but we have yet to make the hard financial and educational choices it imposes. *Science* **134**, 161–164 (1961).
5. Roberts, L. The human genome. Controversial from the start. *Science* **291**, 1182–1188 (2001).
6. Petsko, G. A. 'Big science, little science'. *EMBO Rep.* **10**, 1282 (2009).
7. Heng, T. S. & Painter, M. W. The Immunological Genome Project: networks of gene expression in immune cells. *Nature Immunol.* **9**, 1091–1094 (2008).
8. Narayan, K. *et al.* Intrathymic programming of effector fates in three molecularly distinct γδ T cell subtypes. *Nature Immunol.* **13**, 511–518 (2012).
9. Malhotra, D. *et al.* Transcriptional profiling of stroma from inflamed and resting lymph nodes defines immunological hallmarks. *Nature Immunol.* **13**, 499–510 (2012).
10. Miller, J. C. *et al.* Deciphering the transcriptional network of the dendritic cell lineage. *Nature Immunol.* 15 Jul 2012 (doi:10.1038/ni.2370).
11. Bezman, N. A. *et al.* Molecular definition of the identity and activation of natural killer cells. *Nature Immunol.* 19 Aug 2012 (doi:10.1038/ni.2395).
12. Chain, P. S. *et al.* Genomics. Genome project standards in a new era of sequencing. *Science* **326**, 236–237 (2009).
13. Peterson, J. *et al.* The NIH Human Microbiome Project. *Genome Res.* **19**, 2317–2323 (2009).
14. Jumpstart Consortium Human Microbiome Project Data Generation Working Group. Evaluation of 16S rDNA-based community profiling for human microbiome research. *PLoS ONE* **7**, e39315 (2012).
15. Davis, M. M. A prescription for human immunology. *Immunity.* **29**, 835–838 (2008).
16. Germain, R. N. & Schwartzberg, P. L. The human condition: an immunological perspective. *Nature Immunol.* **12**, 369–372 (2011).
17. Asare, A. L. *et al.* Differential gene expression profiles are dependent upon method of peripheral blood collection and RNA isolation. *BMC Genomics* **9**, 474 (2008).
18. Maecker, H. T., McCoy, J. P. & Nussenblatt, R. Standardizing immunophenotyping for the Human Immunology Project. *Nature Rev. Immunol.* **12**, 191–200 (2012).
19. Hoyne, G. F. & Goodnow, C. C. The use of genomewide ENU mutagenesis screens to unravel complex mammalian traits: identifying genes that regulate organ-specific and systemic autoimmunity. *Immunol. Rev.* **210**, 27–39 (2006).
20. Beutler, B. & Moresco, E. M. The forward genetic dissection of afferent innate immunity. *Curr. Top. Microbiol. Immunol.* **321**, 3–26 (2008).
21. Chattopadhyay, P. K., Hogerkorp, C. M. & Roederer, M. A chromatic explosion: the development and future of multiparameter flow cytometry. *Immunology* **125**, 441–449 (2008).
22. Bendall, S. C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687–696 (2011).
23. Sagoff, M. Data deluge and the Human Microbiome Project. *Issues Sci. Technol.* **28**, 71–78 (2012).
24. Lander, E. S. The new genomics: global views of biology. *Science* **274**, 536–539 (1996).

Acknowledgements
We thank R. Germain, B. Malissen and the reviewers for comments and suggestions. The ImmGen programme is supported by grant R24-AI072073 from the US National Institute of Allergy and Infectious Diseases, National Institutes of Health, and is grateful to eBioscience, Affymetrix and Expression Analysis for sponsorship.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION

Biodefense and Emerging Infections Research Resources Repository: http://www.beiresources.org
Centre for Modelling Human Disease: http://www.cmhd.ca/
Euroflow: http://www.euroflow.org
FOCIS Human ImmunoPhenotypong Consortium: http://www.focisnet.org/FOCIS/index.php?option=com_content&view=article&id=81&Itemid=11
Human Immunology Project Consortium: http://www.immuneprofiling.org/hipc/page/show
Human Leucocyte Differentiation Antigens (HLDA) Workshops: http://www.hcdm.org/AboutHCDM/tabid/56/Default.aspx#history
Immune Epitope Database: http://www.immuneepitope.org
Immunological Genome Project: http://www.immgen.org
International Histocompatibility Working Group: http://ihwg.org/index.html
International Histocompatibility Workshops: http://ihwg.org/about/history.html
Mutagenetix: http://www.mutagenetix.org
NIH Tetramer Core Facility: http://tetramer.yerkes.emory.edu

ALL LINKS ARE ACTIVE IN THE ONLINE PDF