Toward effective sharing of high-dimensional immunology data

Berend Snijder[1,2], Richard Kumaran Kandasamy[1,2] & Giulio Superti-Furga[1]

Immunology is on the cusp of a 'big data'–driven breakthrough, but strategies for standardizing and sharing high-dimensional data from independent laboratories are needed to ensure that data support the formation of new and robust hypotheses.

High-dimensional data generated by new technologies, including 'omics' approaches (genomics, transcriptomics, proteomics, etc.) are providing great insight into the molecular makeup of cells, and ongoing technological developments are enabling the characterization of large sets of individual cells. However, during the development of new technologies, the lack of standardization can make it challenging to directly compare datasets generated in independent laboratories and to integrate different types of datasets together[1]. Community efforts to increase standardization of both experimental and analytical procedures improve the quality and reproducibility of omics data—as has been shown for microarray data[2] and the identification of protein–protein interactions by mass spectrometry[3], for example. In the immunology community in particular, multilaboratory collaborative projects, such as ImmGen (http://www.immgen.org/) and the Human Immunology Project Consortium (HIPC) (http://www.immuneprofiling.org), establish standardization of protocols and data annotations, while characterizing various aspects of the immune system at a high resolution and in different conditions[4–7]. In this Commentary, we lay out the advantages of a decentralized and integrative approach for interrogating high-dimensional (or 'large-scale') immunology data and discuss some of the challenges the community faces in fully embracing such an approach.

[1]The CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria. [2]These authors contributed equally to this work. Correspondence should be addressed to B.S. bsnijder@cemm.oeaw.ac.at or G.S.-F. gsuperti@cemm.oeaw.ac.at

Divide and conquer

Despite their many benefits, centralized multilaboratory collaborations for data creation and integration fail to scale well and make suboptimal use of the potential output of the full immunology community. In some cases, a more decentralized approach where independent laboratories generate and publish data, and individual laboratories access and integrate these datasets, may be preferable. Several recent immunology studies integrating large transcriptional datasets justify the value of just such a strategy[8–11]. One study identified gene-pair expression signatures as putative cell-fate determinants of the hematopoietic system[9], another pinpointed a set of genes (called the 'common rejection module') whose expression is linked to organ-transplant rejection[10], a third revealed the immune cell subsets present in human colorectal tumors that correlate with patient survival[8] and a fourth identified common transcriptional signatures of antibody responses to different vaccines[11].

A common feature of these four studies is their use of hundreds or thousands of publicly available microarray datasets that had been independently generated in multiple laboratories; the teams renormalized and reanalyzed the collective datasets, and then used this information to generate new hypotheses. Together, these studies highlight the huge potential of large-scale and decentralized data creation for the generation of a systems perspective on immunology, and they show that results from different research operations and experimental settings can be integrated to reveal novel properties of the immune system that would have been hard to identify in individual laboratories or even by consortia.

An advantage of decentralized data creation is the diversity of experimental conditions used to generate datasets in different laboratories. At first glance such diversity might be viewed as an obstacle to interlaboratory data sharing and integration. But careful analysis and data curation and annotation can convert it into a strength, as results based on highly divergent datasets are less likely to overfit on individual experimental systems, are more resistant to the bias introduced by publication of selected observations from any one laboratory and are likely to be more easily reproduced in other laboratories as the original finding itself is built on reproduced observations from independent laboratories.

Lastly, such decentralized data creation and sharing may be particularly well-suited to studies of human immunology because of the extraordinary diversity of the various components of the human immune system. For example, thanks to the somatic rearrangement and further diversification of the large number of variable segments in the immune receptor loci, the number of different T cell receptors or antibodies present in a single human can be enormous. The highly polymorphic nature of the human leukocyte antigen (HLA) locus also contributes to this interindividual diversity, as does the influence of pathogen encounters to shaping the immune system. Detecting trends or rare shared clones among this sea of diversity may necessitate examination of data from many individuals, exceeding what is practical for researchers in a single laboratory to generate.

But what obstacles stand currently in the way of wide adoption of such a decentralized approach? Are there simple measures that could enable full deployment of the community's integrative research potential?

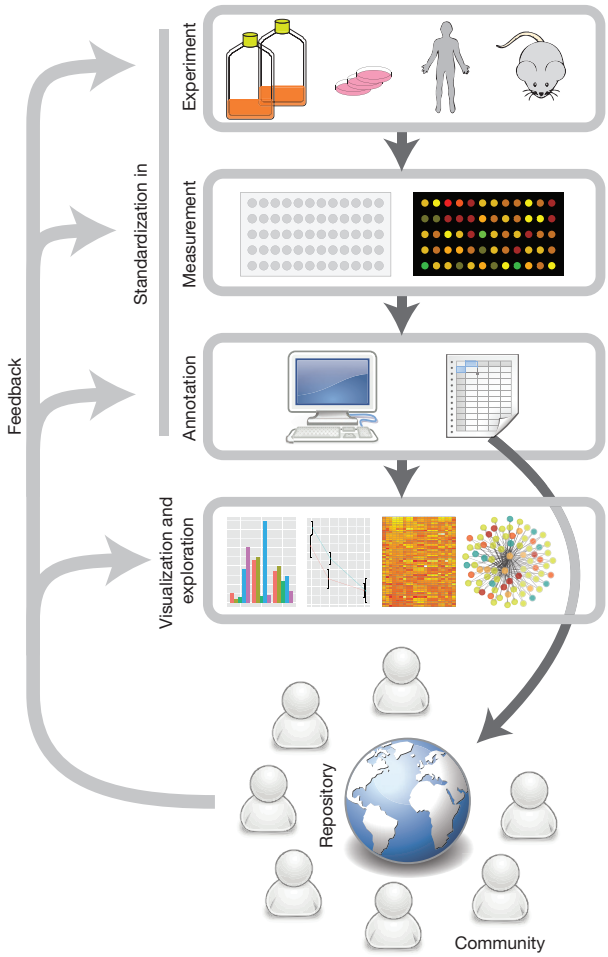

Figure 1 Procedures needed for effective decentralized large-scale data sharing. Efficient sharing and reuse of omics data requires standardization of experimental procedures used to generate data, methods of measuring data and data annotation. It also requires widely available and easy-to-use tools for visualization, exploration and analysis of data, and repositories into which properly annotated data can be deposited. These repositories will ideally enable researchers in other laboratories to easily search and identify datasets of interest, whereas standardization will enable integration and reanalysis of individual datasets into larger heterogeneous datasets. Early development and wide adoption of all of these procedures—which depends largely on the user community—create a positive feedback loop, which eventually leads to increasingly efficient data sharing and re-use. The figure shows aspects of the process relevant to microarray data; thanks to efforts by the user community, technology producers, funders and journals, these types of data are now easily shared and integrated on a large scale, leading to exciting new discoveries.

Below we discuss these issues, focusing where possible on high-dimensional immunological data generated using the new technologies described in a recent Focus (http://www.nature.com/focus/high_dimensional_immune_analysis).

A community-driven process

Microarray data represent a model data type with regard to our current ability to integrate separate datasets on a large scale. In the past 15 years or so that microarrays have been used, the user community has engaged in an extensive effort to enable data standardization, annotation, visualization, exploration and deposition in publicly accessible repositories (**Fig. 1**). This resulted in a broad consensus on how to annotate microarray data (Minimum Information of a Microarray Experiment; MIAME[12]) and where to deposit them (the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus[13] and European Bioinformatics Institute (EBI) ArrayExpress[14]). Notably, deposition of microarray data into these repositories is required by many journals, and repository-generated accession codes must be added to the final published paper. Together, these processes have greatly contributed to the computer-readable resource of microarray datasets and the subsequent capability to integrate and reanalyze the vast troves of results.

Even so, this approach to experimental analysis is at a much earlier stage for many of the newer technologies that are generating high-dimensional measurements of the human immune system today, including RNA sequencing of individual immune cells[15], single-cell mass cytometry, and antibody[16] and T cell receptor[17] repertoire analysis. The initiation of collaborative and cross-laboratory research in any area inevitably lags behind the introduction and refinement of new measurement techniques because such research is largely driven by the community of users of the technique and because the requirements evolve with an improved understanding of the methods involved. For instance, in measuring antibody and T cell receptor repertoire data, the field has not coalesced around a single 'best' experimental or analytical process[18], with experimental differences centering on the sequencing of genomic DNA or mRNA: some methods retain heavy-light chain pairing information whereas others do not, and different bioinformatics approaches are used to analyze the data. Comparative studies between laboratories will be required to identify best experimental practices[3]. Furthermore, crowd-sourced challenges, such as those organized by the Sage Bionetworks (Seattle, WA; http://sagebase.org/) and Dialogue on Reverse Engineering Assessments and Methods (DREAM; http://www.the-dream-project.org/), can be very instrumental to identify the analysis methods that best predict clinical outcome from the large datasets that measurements of the immune-cell repertoire produce. That said, collaborative projects are on the way that are specifically focused on developing the analysis tools, repositories and standards required for successful sharing and integration of these new techniques[19].

Standardization of annotation

Standardization of the way data are described, or annotated, greatly helps in the unambiguous interpretation of data between laboratories and as a consequence increases the potential impact of shared data. Minimal annotation standards have now been developed for many data types, including proteomics experiments and protein-protein interactions[20] (e.g., Minimal Information about a Proteomics Experiment; MIAPE[21]). Additionally, Minimum Information about a Flow Cytometry Experiment (MIFlowCyt; http://flowcyt.sourceforge.net/miflowcyt/) and the Minimal Information About T cell Assays (MIATA; http://miataproject.org/) have been developed for the annotation of

high dimensional (immunology) data coming from flow cytometry experiments. Many more efforts to standardize data reporting and annotation are underway, as illustrated by the BioSharing initiative (http://www.biosharing.org/), which currently lists over 500 reporting standards all dedicated to the sharing of various biological experimental data types. Like MIAME standards for gene expression data, these standards specify the sort of metadata that must be reported to properly describe an experimental dataset. These standards do not dictate how to perform the experiment, which is often a continued point of discussion, as can be seen, for instance, in the experimental standardization of flow cytometry for immunological measurements[22–24]. In flow cytometry experiments, sample handling, the reagents used for staining, instrument setup and data analysis are all topics for standardization, although in our opinion not all are required for successful or useful data sharing. As suggested by the above-cited microarray studies, it is important not only that the raw (single-cell) data are in a readable and annotated format but also that the staining reagents and experimental conditions employed are properly annotated and the hardware are correctly set up. Beyond that, data analysis (e.g., gating strategies) are likely to be done differently by the labs that reanalyze the various flow cytometry datasets, potentially using automated gating strategies that find optimal settings across the different datasets. Such strategies may allow the formation of hypotheses that are robust to the differences in sample handling between laboratories and build on the strength of the increased data sizes.

Successful annotation standards are typified by an initial phase of broad voluntary adoption of the standard and ultimately enforcement of the standard for publication of data once consensus has arisen in the community for that standard. As such, the success of an annotation standard depends on a complex mix of dynamic criteria. Successful annotations implement the proper balance between information content and freedom to warrant ease of use and ideally are compatible with a broad range of technical variations, keeping in mind backward compatibility. Successful adoption further depends on the engagement of the various stakeholders, not just the producers and users of equipment, but also funding agencies and scientific journals.

Visualization and analysis

With the development of new technologies comes the need for new methods for data analysis and visualization, which is a considerable challenge, especially for methods that produce single-cell data. For example, both viSNE[25] and SPADE[26] have been developed for the analysis and visualization of mass cytometry data, and these methods will likely find applications in other areas of high-content data analysis, such as in the analysis of single-cell measurements from image-based screens. We refer to a recent review for an overview of the computational methods developed in the field of high-dimensional immunological data[19].

The existence of multiple methods for analysis and visualization of a single data type has both advantages and disadvantages. If faced with choosing among several analytical methods, users may find it difficult to determine which analytical methods are best for which experiments. Unbiased benchmarking of new analytical methods can help. For example, Flowcap (Flow Cytometry: Critical Assessment of Population Identification Methods; http://flowcap.flowsite.org/), a project aiming to develop computational methods for the identification of cell populations of interest from flow cytometry data, offers such a test suite for the unbiased benchmarking of flow cytometry data analytical methods. The aforementioned DREAM challenges and similar comparative analyses may also be instrumental in identifying the strengths and weaknesses of the various computational methods for analyzing and visualizing high-dimensional immunological data, as they have been for instance in the identification of biological network features from large transcriptional and other aggregated experimental datasets[27,28].

It is likely that despite experimental and computational standardization, high-dimensional immunology data will always display more variability than, for instance, shared microarray datasets, as a result of human-to-human variability or differences in the experimental setups in different laboratories. Successful algorithms will therefore need to adopt specific strategies to account for the abundant variability in the data, for instance, by turning human-to-human variability into a signal that can be mined for inference of interesting trends.

Data repositories

Repositories for the deposition of experimental data are essential infrastructure for successful data sharing and integration. Without these, too often primary data described in publications cannot easily be accessed, or online resources are lost over time. The BioSharing initiative lists over 600 specialized repositories, most of which are dedicated to the sharing of specific experimental data types. With regard to repositories for high dimensional immune data in particular, progress is steady. Repositories for proteomics and protein–protein interaction data (e.g., European Bioinformatics Institute's Proteomics Identifications; PRIDE[29]) and flow cytometry and mass cytometry data (http://flowrepository.org/ and http://www.cytobank.org/) have been developed. And consortia and funders have been establishing immunology-focused portals offering access to, and sometimes analysis of, various related datasets. For instance, the HIPC is developing the immunespace.org repository (http://immunespace.org/), which will host a variety of high-dimensional immunologic data types, two other repositories (http://www.systemsimmunology.org/ and http://www.systemsimmunity.org/) offer genetic, genomic and proteomic data from long-running collaborative projects, and ImmPort (http://immport.niaid.nih.gov/) is an immunology database and analysis portal that provides an archive of data from all US National Institute of Allergy and Infectious Diseases (NIAID)-funded research.

On the other hand, the increasing number of dedicated data repositories can lead to fractionation and makes finding related data types across different repositories a considerable challenge. Topic-oriented portals do help to link related datasets that are stored in different repositories, and general-purpose repositories (e.g., http://datadryad.org/) offer another potential solution by storing diverse data types in the same place. We expect, however, that a big breakthrough in the integrative analysis of high-dimensional immunology data will come from the development of cross-repository data search engines—think a PubMed for datasets—that allows users to find relevant data generated and stored across platforms and repositories. However, such data search engines are currently largely a dream because they may require some form of uniform reporting of datasets, including basic descriptions and online location.

To address this bottleneck, we outline a proposal for the introduction of a uniform data description for the diverse omics datasets generated during immunological and other studies. This would help enable the creation of scientific data search engines that can direct researchers to relevant datasets in various online locations, identify complementary datasets such as replicated proteomic and transcriptomic analyses of the same cell types in the same conditions generated in different laboratories. In addition, as individual datasets can be linked to the researchers that have contributed to the data (ideally by the Open Researcher or Contributor ID; http://www.orcid.org/), recognition can be

Box 1 DEDALO: uniform description of data and their location

As research becomes increasingly data-intensive, the need is growing for a scheme to annotate datasets in a uniform, machine-readable format that is more amenable to discovery and indexing by search engines. We propose the location, contents and properties of each dataset belonging to a publication might be formalized by machine-readable minimal 'descriptor of dataset properties and location' (DEDALO). DEDALOs would not replace repositories but would instead combine a single DOI for each dataset with a description of data properties (author information, basic description or keywords of the experiment, dataset identifiers from public repositories (if any) and DOI of the corresponding publication) and the online location (e.g., a URL pointing to either data on an FTP server or a public repository) of the dataset. If such minimal information on all published experimental datasets were kept in an open-access database (which we imagine would be publicly funded and community-driven), it could give a great impulse to large-scale data integration in day-to-day research. Notably, it would be a single place to also document the storage of datasets that do not currently have standardized repositories, by linking directly to academic FTP servers or other online locations, avoiding a major loss of data during the period when standards are still developing.

For example, upon publication of data in a journal, the researcher will be asked to (voluntarily) fill in the DEDALO form online. After minimal validation, this process would result in the creation of a unique DOI for each individual dataset, which would link to an open-access description of the minimal annotation of the data. This minimal annotation would include information about the location(s) of the dataset (e.g., URL of an FTP server, URL and/or accession code of a repository), the DOI of any linked journal publication and a text description of the actual data (e.g., parameter(s) measured, file types, biological sample and treatments). To maximize flexibility, DEDALOs would not have extensive data annotation specific to the type of data but enough information to allow search engines to match them to user queries. Experimental details and information could be further pulled in by such search engines from the repositories that actually house the data, if present, as well as from the publication(s) linked to a DEDALO. Given the essentiality of such infrastructure we would imagine maintenance of this infrastructure to be best housed by big public institutes such as the EBI and NCBI.

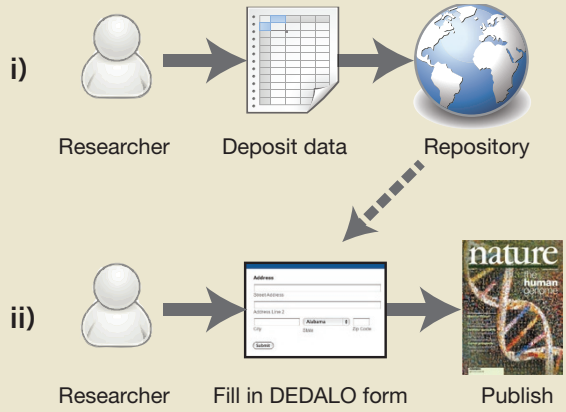

Figure 2 Suggested workflow for 'descriptors of dataset properties and location' (DEDALOs).

If DEDALOs were required by journals, the online location of all datasets in a uniform, machine-readable format would be guaranteed. The status of existing DEDALOs (e.g., whether the linked data are still accessible or whether a URL is still working) could be checked (and potentially updated) automatically, similar to how internet search engines crawl online URLs, without needing to update the data reference in the publication. DEDALOs could further spur the creation of scientific data search engines. These could allow researchers to more easily find various datasets matching their search but also suggest related datasets to researchers based on similarities in experimental description (keywords), on contributions by the same researchers or based on datasets that are often used together in publications. We realize such a system would take a huge effort and considerable funding, for which we offer no solution. The potential benefits are big though. Such a system would greatly improve the usability and integration of the entire output of the research community, while adding a more formal infrastructure for the recognition of datasets contributed by individual researchers.

given to researchers that freely share their data, moving away from over-reliance on impact factors for the evaluation of research output[30]. Even relatively small but valuable datasets generated in smaller laboratories can become highly visible, and the researchers involved be given their due credit. A particular dataset might be identified and tracked as 'hot', even years after is was described and interpreted in a publication that may have become obsolete. Moreover, improved versions of the original datasets can be easily linked and found, providing for the 'evergreening' of data. We provide additional details and discuss this proposal in **Box 1** and **Figure 2**.

Expand the incentives

Even in scenarios where applicable standards-compliant repositories already exist, it can be a challenge to prompt the researchers who generate and publish data (in journals) to also properly annotate and make the data available. Enforcement by journals and funding agencies is currently by far the single most effective incentive for data sharing. Beyond that, few incentives are established in the current system by which researchers that do share their data are rewarded or recognized for their openness. As a result, valuable experimental datasets are being lost daily.

Even the best working method, enforcement, has its shortcomings. Data deposition into annotation standards–compliant repositories before publication is currently often required for gene expression data, protein and DNA sequence data, and protein structures. Techniques for which the repositories and annotation standards are well-developed and broadly embraced by the community. However, it may be impossible to enforce data deposition before the community has agreed on what repositories and standards to embrace. In our opinion, although standardization of annotation is indeed a great aid to the ease of use of a dataset, functional data sharing can work in the absence of formal annotation standards, and general-purpose data repositories or even institute file servers can be useful in serving data to the online researchers community, as long as their location is properly annotated (**Box 1**). Lack of community consensus regarding what repository and annotations to use, or even how to precisely interpret biological results (such as in the discussion on what markers immune cell subsets might express), should not stand in the way of data sharing.

An important upcoming way to incentivize researchers to properly annotate and deposit data is to make appropriately annotated

datasets citable on their own, and include metrics of shared data in the evaluation of individual researchers by funding agencies and academic institutions. The recognition of research contribution and success beyond the publication record is, for instance, argued by the San Francisco Declaration on Research Assessment (DORA, http://am.ascb.org/dora/) and since 2013 explicitly implemented by the US National Science Foundation[30]. Technically, dataset popularity can be tracked by making datasets citable on their own, which in turn can be accomplished by adding a unique digital object identifier (DOI) to datasets. Such solutions are implemented, for instance, in the general-purpose repository DataDryad.org (http://www.datadryad.org/) and is the idea behind the recently launched Scientific Data project (http://www.nature.com/scientificdata/). There, individual datasets are published (complementary to the corresponding research papers) in both a human-readable and a computer-readable format, and uploaded datasets may become citable in scientific publications.

Finally, adoption of the practices required for large-scale immunological data sharing and integration may be propelled by success stories of the early adopters. As more and more papers realize the potential of cross-platform data integration in high-dimensional immunology data, both funders and researchers will inevitably be pushed to adopt the cultural change required for large-scale data standardization, annotation and sharing. The key will be to tag these smaller contributions such that they can be found and used.

Conclusions

Profiling of the immune system repertoire, single-cell sequencing and mass spectrometry are transforming immunology. Whether we adopt a culture of standardization and data sharing will determine whether we can maximize the impact of these high-dimensional data on our understanding of immunology at the molecular, single-cell and whole-organism level. It will also determine whether 'big immunology' will be the prerogative of big laboratories, institutions and centralized consortia or can instead be divided up into small parts where individual research groups contribute meaningfully to answering big immunological questions.

ACKNOWLEDGMENTS

We thank the members of the Superti-Furga lab, and S.H. Friend for critically reading the manuscript and helpful discussions. This work was supported by a Swiss National Science Foundation fellowship (P300P3_147897) to B.S., by a European Molecular Biology Organization long-term fellowship to R.K.K. (ALTF 314-2012), and by the Austrian Academy of Sciences and the European Research Council grant iFIVE to G.S.-F.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Baker, M. Big biology: the 'omes puzzle. *Nature* **494**, 416–419 (2013).
2. Bammler, T. Standardizing global gene expression analysis between laboratories and across platforms. *Nat. Methods* **2**, 351–356 (2005).
3. Varjosalo, M. *et al.* Interlaboratory reproducibility of large-scale human protein-complex analysis by standardized AP-MS. *Nat. Methods* **10**, 307–314 (2013).
4. Gautier, E.L. *et al.* Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat. Immunol.* **13**, 1118–1128 (2012).
5. Miller, J.C. *et al.* Deciphering the transcriptional network of the dendritic cell lineage. *Nat. Immunol.* **13**, 888–899 (2012).
6. Han, A. *et al.* Dietary gluten triggers concomitant activation of CD4$^+$ and CD8$^+$ alphabeta T cells and gammadelta T cells in celiac disease. *Proc. Natl. Acad. Sci. USA* **110**, 13073–13078 (2013).
7. Obermoser, G. *et al.* Systems scale interactive exploration reveals quantitative and qualitative differences in response to influenza and pneumococcal vaccines. *Immunity* **38**, 831–844 (2013).
8. Bindea, G. *et al.* Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **39**, 782–795 (2013).
9. Heinäniemi, M. *et al.* Gene-pair expression signatures reveal lineage control. *Nat. Methods* **10**, 577–583 (2013).
10. Khatri, P. *et al.* A common rejection module (CRM) for acute rejection across multiple organs identifies novel therapeutics for organ transplantation. *J. Exp. Med.* **210**, 2205–2221 (2013).
11. Li, S. *et al.* Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat. Immunol.* **15**, 195–204 (2013).
12. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat. Genet.* **29**, 365–371 (2001).
13. Barrett, T. *et al.* NCBI GEO: mining tens of millions of expression profiles—database and tools update. *Nucleic Acids Res.* **35**, D760–D765 (2007).
14. Parkinson, H. *et al.* ArrayExpress update—from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res.* **37**, D868–D872 (2009).
15. Chattopadhyay, P.K., Gierahn, T.M., Roederer, M. & Love, J.C. Single-cell technologies for monitoring immune systems. *Nat. Immunol.* **15**, 128–135 (2014).
16. Georgiou, G. *et al.* The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat. Biotechnol.* **32**, 158–168 (2014).
17. Newell, E.W. & Davis, M.M. Beyond model antigens: high-dimensional methods for the analysis of antigen-specific T cells. *Nat. Biotechnol.* **32**, 149–157 (2014).
18. Woodsworth, D.J., Castellarin, M. & Holt, R.A. Sequence analysis of T-cell repertoires in health and disease. *Genome Med.* **5**, 98 (2013).
19. Kidd, B.A., Peters, L.A., Schadt, E.E. & Dudley, J.T. Unifying immunology with informatics and multiscale biology. *Nat. Immunol.* **15**, 118–127 (2014).
20. Kerrien, S. *et al.* The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* **40**, D841–D846 (2012).
21. Taylor, C.F. *et al.* The minimum information about a proteomics experiment (MIAPE). *Nat. Biotechnol.* **25**, 887–893 (2007).
22. Maecker, H.T., McCoy, J.P. & Nussenblatt, R. Standardizing immunophenotyping for the human immunology project. *Nat. Rev. Immunol.* **12**, 191–200 (2012).
23. Pachón, G., Caragol, I. & Petriz, J. Subjectivity and flow cytometric variability. *Nat. Rev. Immunol.* **12**, 396–396 (2012).
24. Valle, A., Maugeri, N., Manfredi, A.A. & Battaglia, M. Standardization in flow cytometry: correct sample handling as a priority. *Rev. Immunol.* **12**, 191–200 (2012).
25. Amir, E.-D. *et al.* viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.* **31**, 545–552 (2013).
26. Qiu, P. *et al.* Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat. Biotechnol.* **29**, 886–891 (2011).
27. Marbach, D. *et al.* Wisdom of crowds for robust gene network inference. *Nat. Methods* **9**, 796–804 (2012).
28. Snijder, B., Liberali, P., Frechin, M., Stoeger, T. & Pelkmans, L. Predicting functional gene interactions with the hierarchical interaction score. *Nat. Methods* **10**, 1089–1092 (2013).
29. Vizcaíno, J.A. *et al.* The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41**, D1063–D1069 (2013).
30. Piwowar, H. Altmetrics: value all research products. *Nature* **493**, 159 (2013).