Molecular networks in context

Andrew M Gross & Trey Ideker

Network biology is beginning to tackle the complexities of multicellular systems and disease associations.

As a young science, network biology still lacks the tools to capture the rich diversity of complex biological phenomena such as multicellular organization and disease processes. But recent developments in the field point to the emergence of more-sophisticated approaches to network reconstruction. In a new study in *Nature Genetics*, Greene *et al.*[1] build a library of tissue-specific networks from diverse data sets—including tissue-specific gene-expression profiles, general and tissue-specific protein interactions, and maps of tissue development—and apply the library to predict how gene function changes across tissues. The tissue-specific networks provided in the study are the most comprehensive to date. They will aid efforts to model interactions of different cell types in tissues and whole organisms and to understand how these interactions are altered in disease.

The past two decades have seen an explosion in high-throughput 'omics' data sets on different biological systems. These data have been integrated and assembled into large gene networks through various computational methods that identify interactions among genes and correlations among gene profiles[2]. Many online resources for such analyses are now available, including FunctionalNet, STRING, GeneMania and bioPIXIE, and these sites make available large networks of gene interactions for humans and model species. However, the networks are typically presented as generic models without contextual information on the particular tissue, cell type, disease state, developmental stage or time point of the system or its external stresses and stimuli—all of which may have a strong influence on gene and protein interactions[3].

A handful of studies have attempted to address this limitation by constructing tissue-specific networks. Magger *et al.*[4] generated networks for 60 tissues and applied the networks to prioritize disease-gene associations. Cahan *et al.*[5] assembled expression data from 56 papers into networks for 22 tissues, which were used to guide cellular engineering[6]. The scale of the work of Greene *et al.*[1] far exceeds that of these earlier studies. Combining data from more than 14,000 publications, they construct networks for an unprecedented 144 human tissues and cell types (**Fig. 1**). This large number of networks allows gene function to be inferred even in tissues for which little experimental data exists, because biological information from functionally related tissues can be used to enhance the signal.

Greene *et al.*[1] use their networks to predict how certain cell lineages respond to perturbations and to investigate changes in gene function across different tissues. For example, they identify interactions that are specifically active in blood vessels in response to interleukin-1B. This result reflects that protein neighbors of interleukin-1B in the blood-vessel network are more likely to be perturbed than in other tissue types or in a tissue-naive network. In another example, the authors explore the ability of the LEF1 transcription factor to mediate signal transduction through interactions with various different factors, depending on the tissue. These cases show the potential of tissue-dependent networks to reveal diverse context-dependent functions for a protein.

In a very different application, Greene *et al.*[1] reinterpret the data from a published genome-wide association study (GWAS). These findings add to a growing body of work demonstrating the utility of network-based prior knowledge in improving the search for disease-associated genes[2,4,7]. GWAS often reveal marginally significant associations between various single-nucleotide polymorphisms (SNPs) and a disease of interest. However, if several identified genes are found to interact within the same network neighborhood, this event can be more significant than any of the individual SNPs on which it is based. The result not only improves statistical power but also helps home in on the mechanisms underlying the SNPs by isolating the pathways involved.

Using such an approach to reanalyze the genes identified in a GWAS of hypertension, Greene *et al.*[1] identify known disease-causing genes as well as new candidate disease-associated

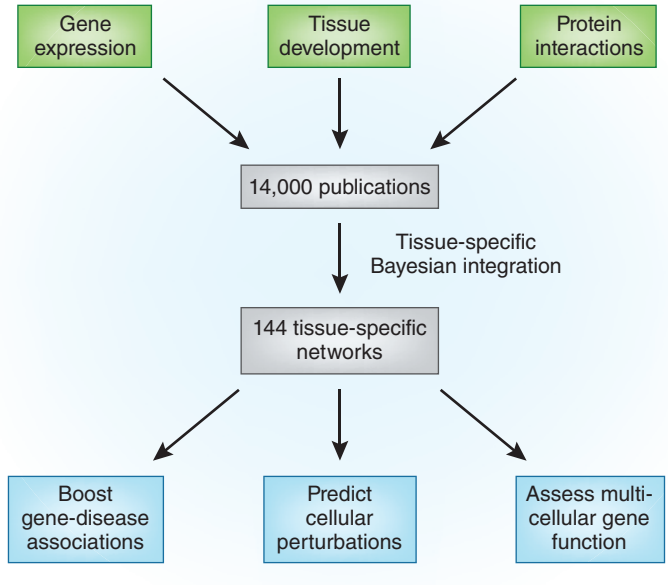

Figure 1 Overview of the approach by Greene *et al.*[1] to construct tissue-specific networks. Data from diverse experiments (green) from over 14,000 publications were used to construct interaction networks for 144 human tissues and cell types. Tissue-specific networks were applied to answer a range of biological questions (blue).

Andrew M. Gross and Trey Ideker are at the Bioinformatics Program and Department of Medicine, University of California, San Diego, San Diego, California, USA.
e-mail: tideker@ucsd.edu

genes that may control blood pressure. The availability of such a large number of tissue-specific maps will expand the phenotypic scope for investigating candidate-gene function in disease-relevant tissues.

Although the tissue-specific networks in Greene *et al.*[1] represent an important advance, improving accuracy and including more tissues and conditions remain key goals for the field. As in all data-driven approaches, differences in data quality and size can have a large impact on the ability to predict interactions. In the present study, some tissues had higher predictive power than others, and cross-validation of even the best network maps showed plenty of room for improvement (area under the receiver operating curve <0.65). Nonetheless, such networks may still prove extremely valuable in understanding how genes contribute to disease.

Beyond its technical achievements, the work of Greene *et al.*[1] is notable for distilling large and diverse data sets into a form accessible to a wide scientific audience. The authors' tissue-specific maps are available for download and can be interrogated directly through a web portal, which will facilitate rapid adoption. As studies such as these are extended to an even wider range of cell types and conditions, network biology may play a greater role in personalized medicine by aiding the interpretation of patients' genetic and phenotypic information. Only by defining the plasticity of the interactome in its many contexts can we truly begin to understand the functioning of complex organisms in health and disease.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Greene, C.S. *et al. Nat. Genet.* **47**, 569–576 (2015).
2. Lee, I., Blom, U.M., Wang, P.I., Shim, J.E. & Marcotte, E.M. *Genome Res.* **21**, 1109–1121 (2011).
3. Ideker, T. & Krogan, N.J. *Mol. Syst. Biol.* **8**, 565 (2012).
4. Magger, O., Waldman, Y.Y., Ruppin, E. & Sharan, R. *PLoS Comput. Biol.* **8**, e1002690 (2012).
5. Cahan, P. *et al. Cell* **158**, 903–915 (2014).
6. Morris, S.A. *et al. Cell* **158**, 889–902 (2014).
7. Ganegoda, G., Wang, J., Wu, F.-X. & Li, M. *BMC Syst. Biol.* **8**, S3 (2014).

Research Highlights

Papers from the literature selected by the Nature Biotechnology editors. (Follow us on Twitter, @NatureBiotech #nbtHighlight)

Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets
Macosko, E.Z. *et al. Cell* **161**, 1202–1214 (2015)

Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells
Klein, A.M. *et al. Cell* **161**, 1187–1201 (2015)

Effect of predicted protein-truncating genetic variants on the human transcriptome
Rivas, M.A. *et al. Science* **348**, 666–669 (2015)

Comprehensive serological profiling of human populations using a synthetic human virome
Xu, G.J. *et al. Science* doi:10.1126/science.aaa0698 (5 June 2015)

A microfluidic device for label-free, physical capture of circulating tumor cell clusters
Sarioglu, A.F. *et al. Nat. Methods* doi:10.1038/nmeth.3404 (18 May 2015)