A blueprint of cell identity

Avi Ma'ayan & Qiaonan Duan

Research on converting one cell type to another will be aided by systematic mapping of the gene-regulatory networks in mammalian cells.

Recipes for generating pure populations of specific cell types are undergoing rapid development for a diversity of applications, from cell therapies to disease modeling and drug screening. The main approaches used are differentiation of pluripotent stem cells, reprogramming of somatic cells to induced pluripotent stem cells and transdifferentiation of one type of somatic cell to another. But how can one be sure that the resulting cells are true exemplars of the desired cell type and that the conversion strategy is optimal? Two recent publications[1,2] in *Cell* describe an important step toward answering these questions. They present CellNet, a map of the gene-regulatory networks that define many different mammalian tissues and cell types. As the authors show, CellNet not only improves the assessment of similarity between cell types but can also be used to predict the transcription-factor perturbations needed to optimize cell-type conversion methods.

Over the last decade the stem cell field has made substantial progress on developing protocols for producing various types of mammalian cells *in vitro*, but so far computational systems biologists have been only marginally involved in these efforts. This is surprising because mapping of the regulatory networks that control cellular phenotypes and the transitions between cell types is at the heart of computational systems biology, and systems-level data on stem cells and differentiated cells—generated with RNA interference screens, ChIP-seq for many transcription factors and histone modifications, gene-expression profiling and protein interaction studies—are abundant. Even before many components of gene-regulatory networks were discovered, theoretical biologists had long been contemplating the nature of transitions between cellular phenotypes. The best-known example is Waddington's theory of the epigenetic landscape, which makes an analogy between cellular differentiation and a ball rolling down a hill that ends at different valleys representing different cell types[3].

Avi Ma'ayan & Qiaonan Duan are in the Department of Pharmacology and Systems Therapeutics, Icahn School of Medicine at Mount Sinai, New York, New York, USA. e-mail: avi.maayan@mssm.edu

In recent years, cross-fertilization between stem cell biologists and computational systems biologists has begun to show how computational models that integrate various types of genome-wide data[4–6] can complement and enhance experimental work on creating cell-conversion protocols and mapping the regulatory networks that govern cell-type transitions. In general, however, these models have covered only one network in one cell type. A more global attempt to map the mammalian signalome, which has some conceptual similarities with CellNet, was demonstrated with a method for defining cell identity called SPADE[7]. However, this work was limited to analysis of the hematopoietic system.

The CellNet studies[1,2] were carried out in a collaboration between an experimental stem cell biology laboratory and a computational systems biology laboratory, and they underscore the power of such cross-disciplinary research. The first step in creating CellNet was to build a global gene-regulatory network of cellular identity across many tissues by compiling data from thousands of published studies that profiled mammalian cells with cDNA microarrays. This network was then broken up into specific gene-regulatory networks for ~20 mammalian cell types. To achieve this, the global network was divided into clusters by identifying highly connected regions, and each cluster was assigned a cell type using enrichment analysis applied to the genes in each cluster. This enabled the development of a metric that characterizes the similarity between cell type–specific gene-regulatory networks and of a method to predict the transcription-factor perturbations that are likely to shift cells toward a desired new phenotype.

Experimental validation demonstrated that CellNet can accurately assess the purity of cell types that have been generated in many recent cellular engineering studies and can guide efforts to correct incomplete cell-type conversions. For example, the method correctly predicted that the conversion of B cells to macrophages would be improved by knockdown of the transcription factors Pou2af1 and Ebf1 to reduce the remaining B-cell identity of the transdifferentiated cells. CellNet was also applied to better characterize induced hepatocytes—fibroblasts converted

to hepatocyte progenitors by expression of the transcription factors Hnf4a and Foxa1[8]. Induced hepatocytes had been shown to engraft in the liver. Analysis with CellNet revealed that these cells actually have broader potential than previously believed and should be considered induced endoderm progenitors. The CellNet method predicted—and experiment subsequently validated—that induced hepatocytes can produce cells that resemble intestinal progenitors and engraft in the colon (**Fig. 1**).

More generally, analysis with CellNet revealed that direct lineage conversion between somatic cell types often partly retains the identity of the original cells, resulting in hybrid cell types. This may explain why in some cases GFP
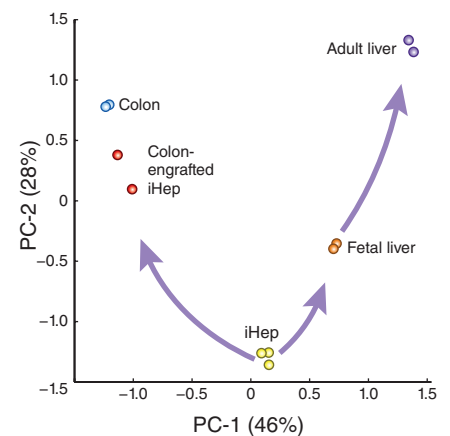

Figure 1 Gene-expression data from 11 samples corresponding to five cell types—colon, fetal liver, adult liver, induced hepatocyte (iHep) and colon-engrafted induced hepatocyte—profiled in one of the CellNet studies[2] were downloaded from GEO (GSE59037). The downloaded data were subjected to quantile normalization; the top 10% of genes with the most variance were retained for further analysis ($n = 4{,}510$), and principal component analysis (PCA) was applied on this subset of the data. The first two principal components (PC-1 and PC-2) were used to generate the plot. Numbers in parentheses indicate the percentage of variance captured by each component. Distance between points on the plot is an estimate of expression vector similarity between the samples. The plot supports the finding that induced hepatocytes are similar to liver progenitors but can also give rise to cells that engraft the colon. This finding was discovered though CellNet analysis and validated experimentally[2].

reporter genes suggest phenotypic conversion, whereas functional assays show otherwise.

The authors showed that CellNet outperforms standard correlation-based methods for classifying cell identity with genome-wide gene-expression data. The most likely explanation is that CellNet focuses only on transcription factors and their interactions, and computes correlations between these transcripts with a method called context likelihood of relatedness (CLR)[9]. The CLR algorithm identifies mutual information (a correlation method that can capture nonlinearities) between pairs of genes; CLR then boosts pairs of genes that have high mutual information scores within a specific window of samples to reduce false positives. This is relevant to the problem of constructing gene-regulatory networks from expression data to define cell identity because two genes may be unrelated in most cases but have a high mutual information score in certain specific cell types.

In contrast with standard unsupervised clustering methods, such as hierarchical clustering, which use the entire gene-expression space or a subspace defined by a crude global filtering method (for example, filtering by variance), CellNet strengthens the influence of transcription factors and their interactions—known to be critical for cell identity—and ignores less-critical variables.

The other important element of CellNet is that it can be benchmarked and integrated with gene-regulatory networks created from ChIP-seq and DNase-seq data. With the emergence of ChIP-seq and DNase-seq, gene-regulatory co-expression networks extracted from mRNA expression data alone can be cross-validated with other independent data sets[10]. This allows methods for constructing gene-regulatory networks to be compared in order to assess their quality.

The analysis for validating CellNet showed that the CLR method recovered many interactions between transcription factors and their targets independently identified by ChIP-seq data from the ENCODE project[1]. Integration of the gene-regulatory networks extracted from mRNA expression data with DNase-seq data collected from embryonic stem cells differentiated to neurons showed that open chromatin genes are more easily induced than chromatin-protected neuronal genes. Such integrative analysis is powerful because it points to specific epigenetic barriers that may prevent complete differentiation into neurons.

One of the remaining challenges with integrative analyses such as the one introduced by CellNet is the differences between the various platforms and technologies for profiling genome-wide gene expression. To build CellNet, the authors used only a few types of microarray platforms. It is difficult to consolidate results from multiple platforms because different platforms have different biases and levels of noise. It has yet to be determined whether using RNA-seq rather than microarrays would improve or hinder meta-analyses such as CellNet.

In addition, the current CellNet map is of low resolution, containing only ~20 generic tissue or cell types. More cell types are needed to properly classify engineered cells. Integrative analyses that synthesize many prior studies to obtain a global picture of the regulatory networks that define mammalian cell types in development, normal physiology and disease are expected to gradually increase in resolution and accuracy in the coming years.

An important avenue of research for reprogramming, directed differentiation and direct lineage conversion strategies is the use of small molecules rather than transcription factor induction or silencing. Small molecules would simplify cell engineering and be more clinically relevant because they are easily applied and removed *in vitro*. Few specific small-molecule inhibitors for transcription factors have been identified. But many inhibitors of kinases are available, and it would be interesting to apply CellNet to develop a regulatory map of the mammalian kinome. Other possibilities for CellNet-type maps are paracrine and autocrine networks that include receptors and their ligands. Perhaps the most effective cell type–specific map would include all of these regulatory layers. However, experimental data on the human kinome and on paracrine and autocrine networks are relatively sparse.

Global computational models such as CellNet can be used to evaluate the many studies being published in the field of cellular engineering. This approach allows more rational predictions about network perturbations that have a greater chance of success than *ad hoc* expert guesses. Beyond providing a more logical and comprehensive synthesis of existing knowledge, mapping the regulatory networks in mammalian cells should eventually allow us to fully control cells and organisms with small molecules in normal physiology and in disease.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Cahan, P. *et al. Cell* **158**, 903–915 (2014).
2. Morris, S.A. *et al. Cell* **158**, 889–902 (2014).
3. Waddington, C.H. *The Strategy of the Genes* (Routledge, 2014).
4. Xu, H., Schaniel, C., Lemischka, I.R. & Ma'ayan, A. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2**, 708–733 (2010).
5. Xu, H., Ang, Y.-S., Sevilla, A., Lemischka, I.R. & Ma'ayan, A. *PLOS Comput. Biol.* **10**, e1003777 (2014).
6. Dunn, S.-J., Martello, G., Yordanov, B., Emmott, S. & Smith, A. *Science* **344**, 1156–1160 (2014).
7. Qiu, P. *et al. Nat. Biotechnol.* **29**, 886–891 (2011).
8. Sekiya, S. & Suzuki, A. *Nature* **475**, 390–393 (2011).
9. Faith, J.J. *et al. PLoS Biol.* **5**, e8 (2007).
10. Ma'ayan, A. *et al. Trends Pharmacol. Sci.* **35**, 450–460 (2014).

Miniaturizing wireless implants

Henry Mei & Pedro P Irazoqui

A breakthrough in wireless power transfer enables smaller implantable devices, opening the door to new applications.

Since the first pacemaker was implanted over 50 years ago[1], implantable medical devices have undergone rapid development, resulting in smaller, more capable and less invasive systems for patients. Yet technologies to power these devices, such as wireless power transfer from an external source, have lagged behind. In *The Proceedings of the National Academy of Sciences*, Ho *et al.*[2] now present an approach to wirelessly transfer sufficient power to operate a

Henry Mei & Pedro P. Irazoqui are at the Center for Implantable Devices, Weldon School of Biomedical Engineering, Purdue University, West Lafayette, Indiana, USA.
e-mail: pip@purdue.edu

pacemaker the size of a grain of rice implanted deep in the chest cavity of a rabbit. This major advance hinges on a novel wireless electromagnetic energy transmitter that can be placed on the skin or perhaps integrated in clothing. The transmitter overcomes the interference of biological tissues to effectively focus electromagnetic energy through the body. The findings have clear clinical implications for the development of small devices implanted deep in the body—for example, in blood vessels.

Implantable medical devices are now available for a wide range of disorders. They include retinal and cochlear implants, deep brain stimulators for epilepsy and Parkinson's disease, pacemakers and brain-machine