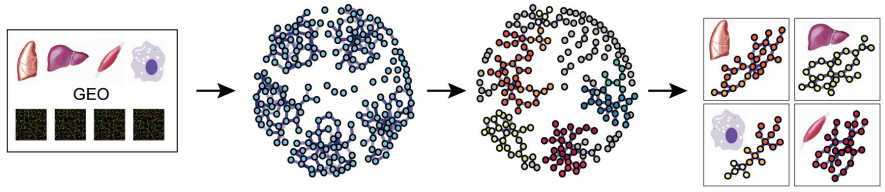
STEM CELLS

A systems view of cellular reprogramming

The CellNet platform, based on cell- and tissue-specific gene regulatory networks, is used to evaluate cells converted to a particular fate by various methods.



CellNet identifies cell- and tissue-specific gene regulatory networks. Image adapted from Cahan *et al.*, Elsevier.

Much of the promise of stem cell research lies in the ability to convert pluripotent or multipotent cells to any other desired cell type. A major goal of cellular reprogramming, whether it involves the reprogramming of somatic cells to induced pluripotency, followed by differentiation, or direct fate conversion between somatic cell types, is to generate cell types of interest for research, drug or toxicity screening, or therapy. Whatever the application, the identity and function of the resulting cells are critical. A differentiation or reprogramming method is only as good as the cells that come out at the end.

Assessing engineered cell types is not trivial. Cells are typically evaluated for the expression of key markers or for the ability to perform surrogate functional tests *in vitro*. In some cases, it is possible to rigorously test for *in vivo* function—for the ability of the cells to complement a deficiency or a disease in an animal model, for instance. But failure to fully complement *in vivo* may be attributable to reasons other than cell identity, and this approach is moreover limited in assessing the fate of human engineered cells. What, then, is one to do?

In two recent papers, George Daley of Harvard Medical School and Boston Children's Hospital, James Collins of Boston University and their colleagues have taken a systemic approach to the problem. By using global gene expression profiles to define cell type–specific regulatory networks rather than restricting the analysis to a few specific markers, the researchers have established the CellNet platform to assess how close an engineered cell is to a specific cell type of interest (Cahan *et al.*, 2014; Morris *et al.*, 2014).

The researchers began with publicly available microarray-based gene expression data. They identified 20 murine and 16 human cell types for each of which there were at least 60 expression profiles, with at least 10 of these representing different conditions or contexts. From these data, the researchers inferred gene regulatory networks (GRNs), which they validated against profiles of transcription factor binding and function, and then went on to define components of the network specific to particular cell or tissue types. Finally, they used these cell- or tissue-type networks to train classifiers that can distinguish cell types. The classifiers perform at nearly 100% sensitivity for a <5% false positive rate for mouse cells, though human cell classification does not quite reach this accuracy.

Using CellNet, Daley, Collins and colleagues investigated whether the cell types that emerge from different engineering approaches are equivalent. For murine neurons or cardiomyocytes generated by directed differentiation versus by direct conversion from fibroblasts, the analysis indicates not. For both cell types, the researchers found that cells generated by direct conversion suppressed the original (fibroblast) GRN less well, as has been previously suggested, and that GRNs matching alternative cell fates could also be identified in these directly converted cells. Indeed, their analysis of all published reprogramming and directed differentiation studies for which CellNet-compatible expression data were available led to the conclusion that many methods generate cells that are distinct in their GRNs from the *in vivo* correlate.

Not only can CellNet quantify how close a cell is to a desired type, it can also identify good candidate genes for further modulation to improve reprogramming outcomes. Daley, Collins and colleagues illustrated this function in the transcription factor–mediated conversion of murine B cells to macrophages. Finally, they used CellNet to argue that previously reported murine iHep cells should be reclassified as endodermal progenitors rather than as liver cells.

As the researchers recognize, there are many potential confounding effects in analyses with CellNet. The gene expression data used to train the classifiers are mainly derived from *in vivo* tissue, which contains many primary cell types, whereas CellNet is used to assess single cell types, albeit also of unknown functional homogeneity, generated and cultured *in vitro*. Although the authors rule out crippling effects of these confounders for some cell types, these may prove more problematic in other cases. Also, CellNet is still only available for a limited number of cell types and for analysis of data obtained with microarrays; the researchers plan to extend it to RNA-seq data in the future.

Even with a cell-specific GRN in hand, these studies illustrate that engineering cell identity is far from easy, not surprisingly given the molecular complexity involved. Its limitations notwithstanding, CellNet should be a very useful tool to evaluate fate conversion methods and the cells that result.

Natalie de Souza

RESEARCH PAPERS

Cahan, P. *et al.* CellNet: network biology applied to stem cell engineering. *Cell* **158**, 903–915 (2014).

Morris, S.A. *et al.* Dissecting engineered cell types and enhancing cell fate conversion via CellNet. *Cell* **158**, 889–902 (2014).