Cell Cycle 13:21, 3313–3314; November 1, 2014; © 2014 Taylor & Francis Group, LLC

Defining cellular identity through network biology

Patrick Cahan[1,2], Samantha A Morris[1,2], James J Collins[3,4], and George Q Daley[1,2,*]

[1]Stem Cell Transplantation Program; Division of Pediatric Hematology/Oncology; Manton Center for Orphan Disease Research; Howard Hughes Medical Institute; Children's Hospital Boston and Dana Farber Cancer Institute; Division of Hematology; Brigham and Women's Hospital; Boston, MA USA; [2]Department of Biological Chemistry and Molecular Pharmacology; Harvard Medical School; Harvard Stem Cell Institute; Boston, MA USA; [3]Howard Hughes Medical Institute; Department of Biomedical Engineering and Center of Synthetic Biology; Boston University; Boston, MA USA; [4]Wyss Institute for Biologically Inspired Engineering; Harvard University; Boston, MA USA

Mechanisms of development rank among the great conundrums of biology: how are the diverse cell types that comprise an adult organism generated from a single genome? While countless labs around the world are attacking different aspects of this question, a related question remains poorly defined: what is a 'cell type'? In other words, by what criteria should we categorize cells into classes? Answering this question is not a mere academic exercise but has deep implications for both developmental biology and for engineering cell fate. Generating specific cell types in the dish, either through directed differentiation of stem cells or through direct conversion between somatic cells, is an area of intense interest because engineered cells can serve as a platform to screen drugs, to study otherwise inaccessible cell populations, and for regenerative medicine. Much effort in this field is centered on finding appropriate culture conditions and factors that can guide cell populations toward a desired target cell type. However, two substantial barriers to the faithful recapitulation of cell types in vitro are intimately related to the concept of cell type or cell identity. First, we lack a quantitative metric to assess the identity of engineered populations. Second, we are in need of a rational, hypothesis-driven means to improve the identity of the engineered cells. We have recently developed a computational platform, CellNet, that addresses these issues,[1] and we have used CellNet to discover a previously unappreciated potential (and thus identity) of induced hepatocytes to functionally engraft the colon, and to improve the function of directly converted macrophages.[2]

Cells are typically categorized by their physiologic function: red blood cells transport oxygen and motor neurons transmit signals between skeletal muscle and the central nervous system. Therefore, determining whether engineered cells are equivalent to haematopoietic stem cells, for example, requires transplanting them into an immune-compromised mouse and assessing long-term, multi-lineage engraftment. However, functional assessments of identity can be technically challenging, may not sample all the desired functions of the target cell type, and are of limited value if they fail. Rather than a phenomenological test of identity, we sought to develop a test that decodes the essence of a cell type. Therefore, to build CellNet we took as our starting point the observation that gene regulatory networks (GRNs) determine both the steady-state transcriptional profile of a cell and its response to most perturbations (e.g., stress, lack or surfeit of nutrients, disease, age, etc.). In this way, GRNs determine the cell's behavior, and thus its identity. We used publicly available gene expression data to reconstruct human and mouse cell and tissue type specific GRNs, and used them to make cell type classifiers. Several important insights emerged from our application of the classifier to gene expression data of all compatible cell engineering experiments in the public domain. First, we found that directly differentiated cells approached their in vivo target types more closely than directly converted cells. Second, we found that the GRNs of the starting cell type frequently are maintained in the engineered populations (**Fig. 1**). Finally, we found that aberrant GRNs of other cell types (neither the starting nor the target) were partially established in the engineered cells.

We used two systems to test how this approach can be applied to cell engineering. First, we applied the classifier to hepatocyte-like cells (iHeps) induced through the expression of Foxa1 and Hnf4α, and found that GRNs of the colon were partially established. Hypothesizing that iHeps might have broader endoderm potential, we transplanted them into superficially damaged mouse colon and found that they were able to integrate into the colonic epithelium to support long-term functional engraftment. In the second system, we used CellNet to identify critical B cell transcription factors that were maintained in induced macrophages, knocked their expression down, and showed that resulting cells had improved macrophage function. These studies demonstrate the utility of CellNet for improving direct conversion and uncovering unappreciated features of engineered cells.

There are several areas in which our network biology approach to improve cell engineering can be substantially improved, and adapted and applied to other questions. First, we learned from

*Correspondence to: George Q Daley; Email: george.daley@childrens.harvard.edu
Submitted: 09/17/2014; Accepted: 09/25/2014
http://dx.doi.org/10.4161/15384101.2014.972918
Comment on: Cahan P, et al. Cell 2014; 158:903–15; Morris SA, et al. Cell 2014; 158:889–902.

our knock-down experiments in the induced macrophage system that parts of starting cell type GRNs are locked in or buffered by feedback loops. As the persistence of the starting cell type GRN was commonly found across cell engineering paradigms, learning how to overcome this barrier will be crucial. A combination of experimental screens and more sophisticated computational modeling should reveal principles to help decommission the starting cell type GRN. The second major area for future advancement will be to reconstruct truly cell (and not tissue) type specific GRNs using single cell RNA-Seq data. While this endeavor will require the coordinated input and efforts of many labs across disciplines, based on what we have already achieved with population-based microarray data, we believe that such a strategy will pay off in the long term. And, while the transcriptome is an excellent reflection of cellular state, incorporating epigenome information will further improve the predictive capacity of the platform. Finally, we speculate that a similar application of network biology will enable us to explore how mutations impact the GRNs in specific tumors, and how GRNs can serve to categorize the cell of origin of diverse cancers.

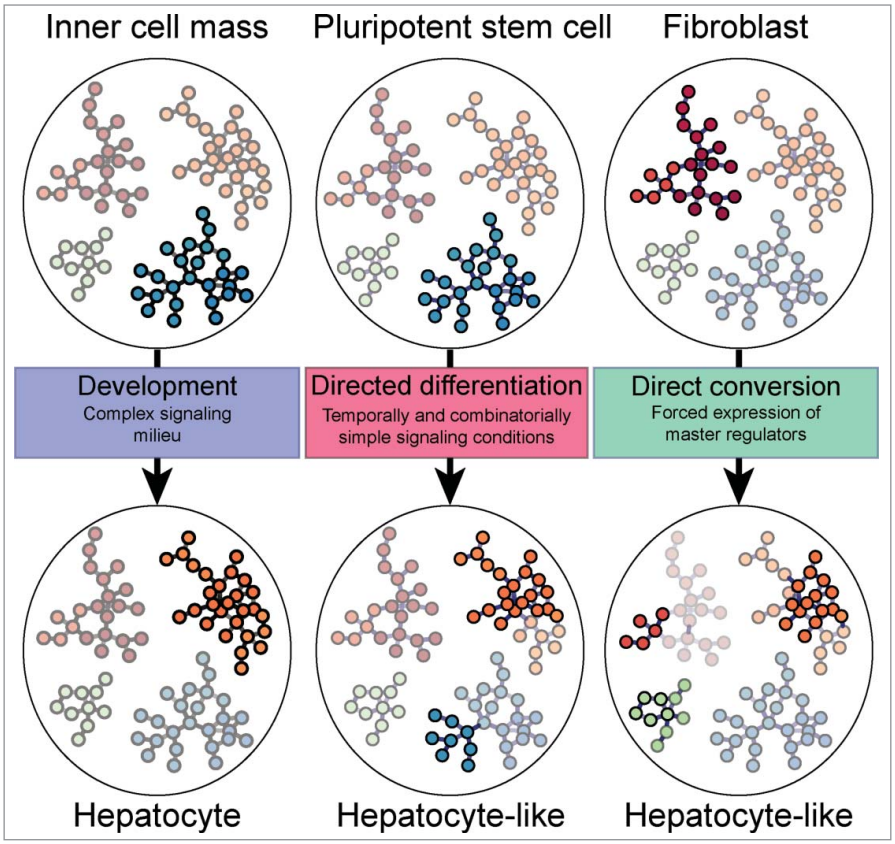

Figure 1. The establishment of a lineage specific gene regulatory network (GRN) in development (left), in directed differentiation of pluripotent stem cells (middle), and in direct conversion (right). Cell engineering approaches fail to fully install the hepatocyte GRN, fail to decommission the starting cell type GRN, and partially establish alternate fate GRNs.

References

1. Cahan P, et al. 2014; 158:903–15; PMID:25126793
2. Morris SA, et al. Cell 2014; 158:889–902; PMID:25126792