Directed differentiation through combinatorial transcription factor activation and gene regulatory network quantification

Gleb Kuznetsov — Biophysics 205 Proposal Draft — March 29, 2015

Brief Summary: Improving strategies to create mammalian cell types of interest will expand the potential of stem cell and directed differentiation-based therapies. Differentiation occurs through a cascade of gene expression and epigenetic changes within a cell. Rational analysis followed by low throughput experimentation has led to numerous demonstrations of derivations of specific cell types that resemble target types morphologically and functionally through ectopic over-expression of certain transcription factors. However, on a molecular level these induced cell and tissue types still harbor signature of their originating cells [1], which may ultimately limit their stability and therapeutic potential. I propose leveraging advances in reading and writing cell state to combinatorially explore directed differentiation pathways and identify new combinations that are less constrained by assumptions and knowledge a priori. The core idea is to use a catalytically dead Cas9 (dCas9) fused with an activating domain, along with a library of arrays of multiple guide RNAs simultaneously targeting activation of different combinations of transcription factors.

Background: In order to better engineer cells, we need a quantitative measure of cell state and techniques for updating the state of cells to more closely achieve the targeted state. Cell reprogramming has traditionally been accomplished through ectopic over-expression of transcription factors. The requirement of delivering entire transcription factor constructs limits the throughput of testing combinations of TFs. Recently, a complementary pair of studies showed that engineered cells still carry significant signatures from their cells of origin before reprogramming to iPS state and subsequent directed differentiation [2], [1]. These studies develop and demonstrate the use of a platform CellNet that infers gene regulatory networks (GRNs) [3] from cell and tissue expression data, which serves as a quantitative readout of cell state.

In addition to its utility as a way to introduce genomic mutations, CRISPR/Cas9, specifically in a catalytically dead form (dCas9), has been fused with transcription activating domains to yield a powerful tool for modulating activation. Several recent advances have made orders of magnitude of activation possible [4], [5]. Additionally, new tricks for better delivery of gRNA arrays are continuing to be reported, such as [6].

Finally, technology for synthesis of oligonucleotides on DNA microchips has reached an appropriate price and quality [7], enabling myriad high-throughput perturbative experiments to explore the space of biological possibilities through experiments such as saturation mutagenesis of genes [8] or combinatorial explorations of regulatory elements [9].

Impact: These technological developments present an opportunity to to extend our capabilities in directed differentiation. Through a high-throughput, combinatorial, data-driven approach to reprogramming cells, I believe it's possible to obtain better final products and collect more data and insight into the cell state modification process.

Aim 1: Develop system for combinatorially exploring space of transcription factor activation to bring cell state closer to goal.

Start with a culture of differentiated cells, say induced hepatocytes (iHeps) as in [2]. Two main components need to be introduced: an activating dCas9 such as in [4] or [5], and an array of CRISPR guide RNAs (gRNAs) that target some combination of transcription factor promoters. The dCas9 can be introduced early on with all experiments carried on clonally and libraries of different combinations of gRNA array cassettes can be introduced after. Upon lentiviral infection, the culture is left for several days to incubate. It may be desirable to come up with a strategy for splitting cells part way through incubation to prevent cells that diverge too far from the desired cell type from affecting their neighbors. After a week or so of incubation, there are a few ways to get quantitative readout.

To quantify results, RNA-seq will be used in combination with computational methods that assess state of GRN [1]. A potential variation on this is to do single-cell RNA-seq which will allow identifying individual cells that mostly resemble the target GRN network, and should pick up the inserted array that corresponds to this ideal transition.

Aim 2: Demonstrate that system can be used to generate cell types that are closer to targets. (Model system TBD) Following closely from Aim 1, we hope to show that this method of combinatorial TF activation actually yields cells that are closer to their target type.

Variations:

- Leverage orthogonal Cas9s for combiantions of simultaneous activation and repression of transcription factors.

- Target other epigenetic modifiers in addition to transcription factors.

- Measure more than expression data, including other epigenetic data such as methylation and chromatin modifications.

References

1. Patrick Cahan, Hu Li, Samantha A Morris, Edroaldo Lummertz da Rocha, George Q Daley, and James J Collins. Cellnet: network biology applied to stem cell engineering. *Cell*, 158(4):903–915, 2014.

2. Samantha A Morris, Patrick Cahan, Hu Li, Anna M Zhao, Adrianna K San Roman, Ramesh A Shivdasani, James J Collins, and George Q Daley. Dissecting engineered cell types and enhancing cell fate conversion via cellnet. *Cell*, 158(4):889–902, 2014.

3. Eric H Davidson and Douglas H Erwin. Gene regulatory networks and the evolution of animal body plans. *Science*, 311(5762):796–800, 2006.

4. Alejandro Chavez, Jonathan Scheiman, Suhani Vora, Benjamin Pruitt, Marcelle Tuttle, Eswar Iyer, Shuailiang Lin, Samira Kiani, Christopher Guzman, Daniel Wiegand, and et al. Highly efficient cas9-mediated transcriptional programming. *Nature Methods*, 0, 2015.

5. Silvana Konermann, Mark D Brigham, Alexandro E Trevino, Julia Joung, Omar O Abudayyeh, Clea Barcena, Patrick D Hsu, Naomi Habib, Jonathan S Gootenberg, Hiroshi Nishimasu, et al. Genome-scale transcriptional activation by an engineered crispr-cas9 complex. *Nature*, 2014.

6. Kabin Xie, Bastian Minkenberg, and Yinong Yang. Boosting crispr/cas9 multiplex editing capability with the endogenous trna-processing system. *Proceedings of the National Academy of Sciences*, 112(11):3570–3575, 2015.

7. Emily M LeProust, Bill J Peck, Konstantin Spirin, Heather Brummel McCuen, Bridget Moore, Eugeni Namsaraev, and Marvin H Caruthers. Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic acids research*, 38(8):2522–2540, 2010.

8. Gregory M Findlay, Evan A Boyle, Ronald J Hause, Jason C Klein, and Jay Shendure. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature*, 513(7516):120–123, 2014.

9. Sriram Kosuri, Daniel B Goodman, Guillaume Cambray, Vivek K Mutalik, Yuan Gao, Adam P Arkin, Drew Endy, and George M Church. Composability of regulatory sequences controlling transcription and translation in escherichia coli. *Proceedings of the National Academy of Sciences*, 110(34):14024–14029, 2013.