

# Epigenetic landscapes provide insight into cellular reprogramming

Alex H. Lang<sup>1,2</sup>, Katherine A. Benson<sup>2,3</sup>, Hu Li<sup>4,5</sup>, Darrell N. Kotton<sup>2,3</sup>, James J. Collins<sup>4,5</sup>, and Pankaj Mehta<sup>1,2</sup>

**Short Abstract — A common metaphor for describing development is a rugged epigenetic landscape where cell fates are represented as attracting valleys resulting from a complex regulatory network. Here, we introduce a framework for explicitly constructing epigenetic landscapes that combines genomic data with techniques from physics. Each cell fate is a dynamic attractor, yet cells can change fate in response to external signals. The landscapes suggest an explanation for partially-reprogrammed cells. In addition, our model reproduces known reprogramming protocols and identifies candidate transcription factors for reprogramming to novel cell fates, suggesting epigenetic landscapes are a powerful paradigm for understanding cellular identity.**

## I. PURPOSE

CELLULAR reprogramming and direct conversion holds enormous therapeutic potential for studying and treating human disease. Over the last several years, it has become possible to reprogram[1] differentiated cells into stem cells as well as directly transdifferentiating[2] cells into other specialized cell types by overexpressing specific transcription factors. Present cellular reprogramming and direct conversion strategies design protocols by selecting transcription factors in an *ad hoc* fashion through a combination of biological knowledge, intuition, and experimental trial and error. Consequently, conversion tends to be inefficient and incomplete. Furthermore, direct conversion protocols only exist between a few select cell types[3].

## II. RESULTS

We propose to overcome these limitations by inventing a new theoretical framework[4] to describe cellular identity that leads to predictions of transcription factors to use in reprogramming and direct conversion protocols. Our framework provides a method to turn the common developmental metaphor of a rugged epigenetic landscape (Waddington’s Landscape)[5] into a mathematically precise

landscape. Specifically, we combine genomic data (Affymetrix Mouse microarrays for 95 cell fates) and histone modification data with techniques from physics (extensions of Hopfield neural networks)[6]. This allows us to explicitly construct a mathematical representation of the epigenetic landscapes underlying cellular identity based on experimental data. In the model, each cell fate is a dynamic attractor, yet cells can change fate in response to external signals, thus capturing the main biological features of development. Our model provides a unified framework to discuss differentiation and reprogramming. It also naturally explains the existence of partially reprogrammed cell fates (stable cell fates found in reprogramming experiments but not in vivo) as ‘spurious’ attractors resulting from the high dimensionality of the landscape. Our model predicts, and we verify, that partially reprogrammed cells are hybrids that co-express TFs of multiple naturally occurring cell fates.

## III. CONCLUSION

To test our theory, we developed a novel tool for generating sets of genes that are “predictive” of specified cell fates, and we tested this tool’s potential for predicting reprogramming transcription factors or determining the cell types of a reprogrammed cell populations. Using predictivity and log-normal z-score gene expression, we were able to correctly predict the set of published transcription factors able to reprogram fibroblasts into iPSCs, hepatocytes, and cardiomyocytes. There is ongoing experimental work to validate the direct conversion candidates. Overall, this work suggests epigenetic landscapes are a powerful paradigm for understanding cellular identity and may be useful for rationally designing reprogramming and direct conversion protocols to novel cell fates.

## REFERENCES

- [1] Takahashi K, Yamanka S (2006) Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors *Cell* **126**:4 663-676.
- [2] Vierbuchen, T et al (2010) Direct conversion of fibroblasts to functional neurons by defined factors *Nature* **463**:7284 1035-1041.
- [3] Vierbuchen, T, Wernig M (2012) Molecular Roadblocks for Cellular Reprogramming *Molecular Cell* **47**:6, 827-838.
- [4] Lang AH et al (2012) Epigenetic landscapes explain partially reprogrammed cells and identify key reprogramming genes *arXiv* 1211.3133.
- [5] Waddington CH (1957) *The Strategy of the Genes* Allen&Unwin, London.
- [6] Kanter I, Sompolinsky H (1987) Associative recall of memory without errors. *Phys Rev A* **35**:1 380-392.

Acknowledgements: AHL was supported by a NSF GRFP.

<sup>1</sup>Department of Physics, Boston University, Boston, MA USA. E-mail: [alexlang@bu.edu](mailto:alexlang@bu.edu), [pankajm@bu.edu](mailto:pankajm@bu.edu)

<sup>2</sup>Center for Regenerative Medicine, Boston University, Boston, MA USA. E-mail: [kbenson@bu.edu](mailto:kbenson@bu.edu), [dkotton@bu.edu](mailto:dkotton@bu.edu)

<sup>3</sup>Pulmonary Center, Boston University School of Medicine, Boston, MA USA.

<sup>4</sup>Department of Biomedical Engineering, Boston University, Boston, MA USA. E-mail: [jcollins@engc.bu.edu](mailto:jcollins@engc.bu.edu)

<sup>5</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA USA. Email: [hu.li@wyss.harvard.edu](mailto:hu.li@wyss.harvard.edu)