# CHROM NET–Chromatin-based cellular phenotype classification

Caleb Lareau, BST 290 Midterm

Prepared for 5 November 2016

## 1 ABSTRACT

Recent advances in stem cell technology have enabled the isolation of induced pluripotent stem cells from committed cell lineages, providing a powerful framework for novel therapeutics and investigating cell-fate decisions. To assess factors relevant to inducing pluripotency, bioinformatics approaches have been proposed using transcriptional regulatory networks with some success. However, as transcriptional data can fail to clearly distinguish cellular phenotypes, we hypothesized that these methodological advancements could be improved. Thus, we introduce ChromNet, a novel computational approach that assigns cellular phenotypes to new samples using a reference panel of chromatin accessibility from ATAC-Seq data. We apply ChromNet to induced pluripotent cell states from blood and identify novel transcription factors associated with development cascades and cell-fate decision.

## 2 INTRODUCTION

The discovery of induced pluripotent stem cells (iPSC) from committed cell lineages has revolutionized cell biology.[1] In particular, the ability to induce pluripotent cell states from committed lineages has afforded a remarkable experimental program to discern
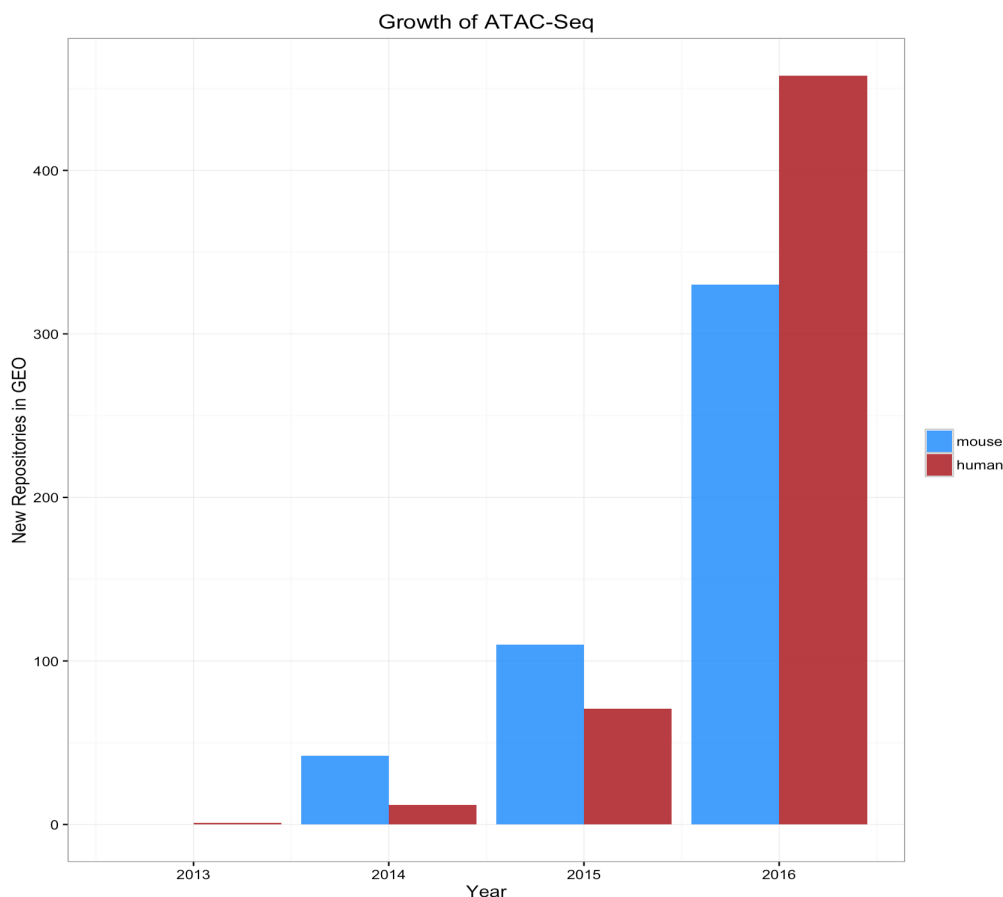


Figure 1: Growth of ATAC-Seq. Since the original description of the protocol in 2013, more than 2,000 samples are publicly available on GEO in both mouse and human. Additionally, the growth of this assay is increasing exponentially. New assays added each year are noted in the bar graphs stratified by organism. These numbers only consider bulk samples that could be downloaded from GEO determined by HTML scrubbing, but additional samples may be available through other public resources and in single cells.

| Cluster purities of hematopoietic cells | | | | |
|---|---|---|---|---|
| Data Type | RNA-Seq | ATAC-Seq | Promoter ATAC | Distal Enhancer ATAC |
| Cluster Purity | 0.776 | 0.857 | 0.675 | 0.909 |

Table 1: Cluster purities of hematopoietic cells from a previous manuscript.[4] While these cellular phenotypes are closely related, variable open chromatin at distal enhancer regions enables a strong separation of these phenotypes in the samples considered. Transcriptomic data as well as open chromatin at promoters were less effective in distinguishing cellular phenotypes in unsupervised clustering.

ID factors.

## 3 METHODS

(brief)
First,
To assess the efficacy of our proposed framework,

## 4 RESULTS

(brief) Data analysis results
(brief) Validation results

## 5 DISCUSSION

(brief) Interpretation:
(brief)

## 6 REFERENCES

[1] Yu, Junying, et al. "Induced pluripotent stem cell lines derived from human somatic cells." Science 318.5858 (2007): 1917-1920.

[2] Cahan, Patrick, et al. "CellNet: network biology applied to stem cell engineering." Cell 158.4 (2014): 903-915.

[3] Buenrostro, Jason D., et al. "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." Nature methods 10.12 (2013): 1213-1218.

[4] Corces, M. Ryan, et al. "Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution." Nature Genetics (2016).

[5] Graf, Thomas, and Tariq Enver. "Forcing cells to change lineages." Nature 462.7273 (2009): 587-594.
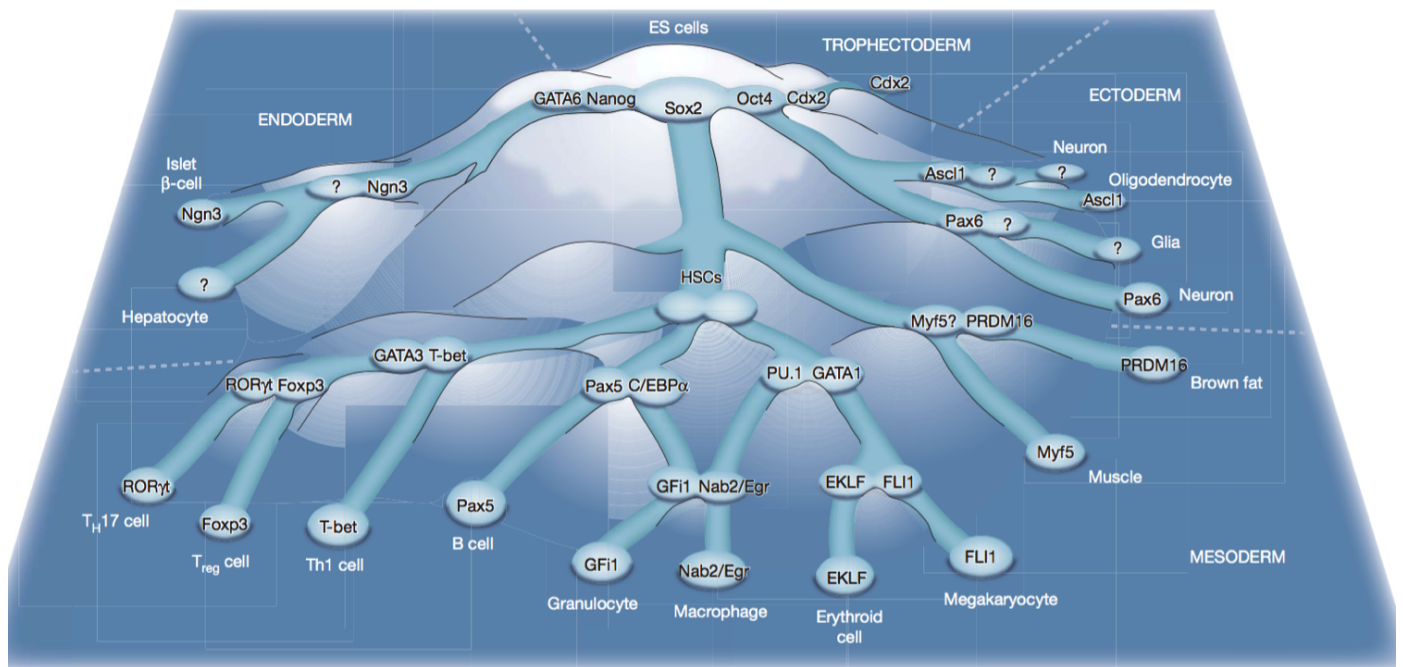
Figure 2: Transcription factors in a cascading landscape of cell states. Differentiated cell types are represented as basins whereas unstable theoretical cell states correspond to ridges and slopes in the landscape. The latter types of cells are observed infrequently during development and thus unlikely to correspond to observable cell types. Characterizing this landscape and the factors associated in moving from one type to another is vital for characterizing epigenetic plasticity. Image reproduced from a previous manuscript.[5]