Inference of TF-mediated epigenetic control of gene expression from 10x multiome data

Luan Quang Le, Ruslan Soldatov

Computational Oncology, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, 321 E 61st St, New York, NY, 10065

Introduction

Transcription factors (TF) play an important role in epigenetic control of gene expression. They can interact with regulatory proteins to modulate chromatin accessibility and transcription initiation. Multi-ome technology, which can concurrently sequence the same group of cells with scATAC-seq and scRNA-seq (Fig. 1), can help elucidate how TFs utilize epigenetic condition to mediate gene expression. The multi-ome data provide simultaneous information on chromatin accessibility and gene expression, which is more advantageous than disparate scRNA-seq or scATAC-seq on different cell groups. Using scATAC-seq data can identify TF motif enrichment; but it faces technical biases, sparsity of data and degenerate motifs from related TFs. Using scRNA-seq data can infer gene regulatory network; but it is static and unable to capture condition-specific rewiring via epigenetic changes. Concurrent collection of the two datasets can overcome these limitations, thus potentially unravelling how TF can channel signals from epigenetic changes to modulate gene expression.

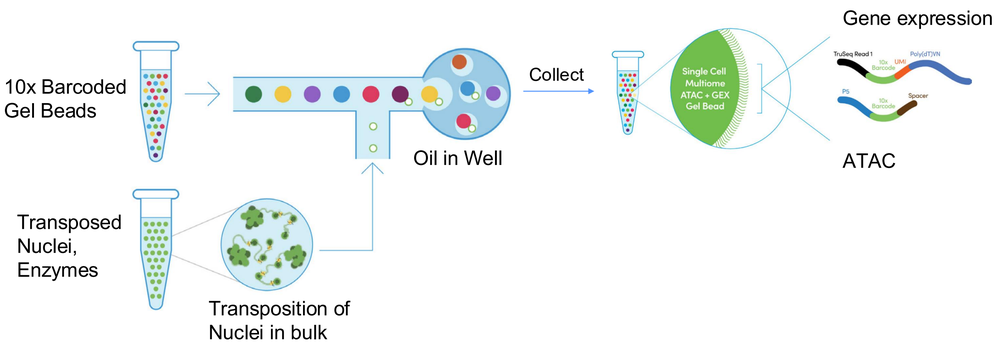


Figure 1. Illustration of multi-omics technology [1]. The technology can simultaneously incorporate barcode and transposase into single cells that are encapsulated by oil droplets.

Methods

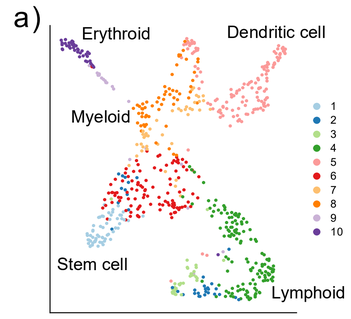
We hypothesize that TF activity can be inferred from peak accessibility and TF motifs presence in the peaks. Letting peak accessibility matrix $P \in \mathbb{R}^{p \times n}$ with p peaks and n cells, and peak-motif association matrix $M \in \mathbb{R}^{p \times m}$ with p peaks and m motifs, we infer TF activity as a low-rank, nuclear norm regularization regression as follows [2]:

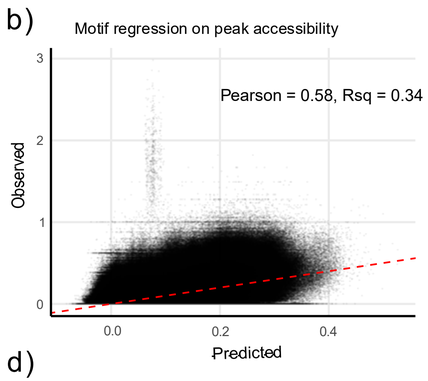
$$P = MA \quad \text{where rank(A)} \leq \min(m, n)$$

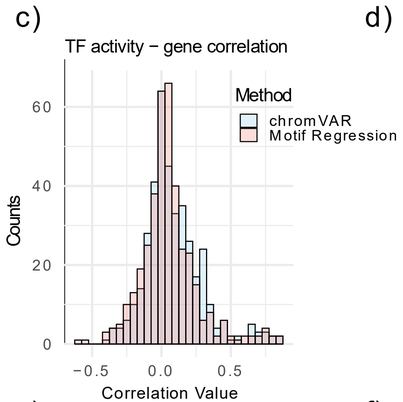
Future plans

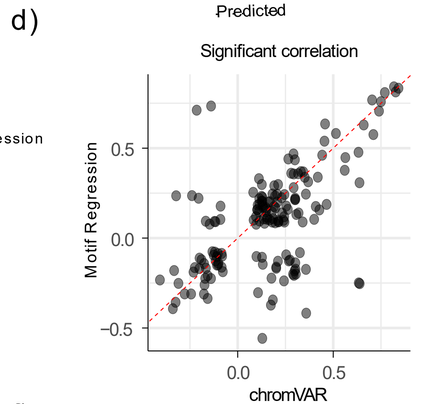
We aim to infer peak-TF-gene relation, in particular the ordering of how cis-element priming and TF transcription modulate gene expression and dictate the differentiation process, especially in the context of clonal hematopoeisis with mutations in DNMT3A and TET2.

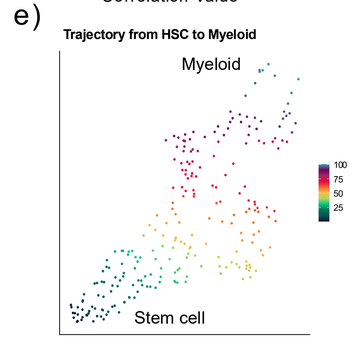
Results

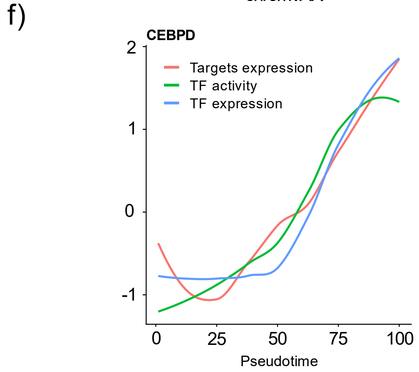


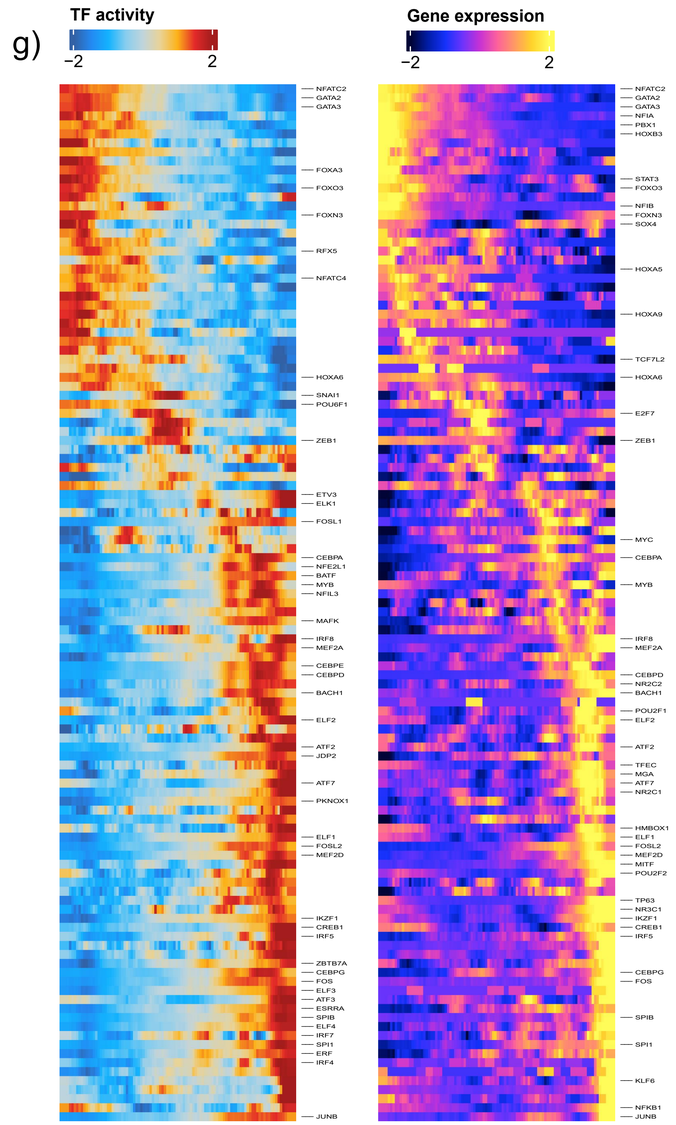












References

1. Simultaneous profiling of the transcriptome and epigenome at single cell resolution - 10x Genomics
2. Chen et al., Biometrika, 2013