E Pluribus Unum: United States of Single Cells

Joshua D. Welch[1(✉)], Alexander Hartemink[2], and Jan F. Prins[1]

[1] Department of Computer Science, The University of North Carolina,
Chapel Hill, USA
{jwelch,prins}@cs.unc.edu
[2] Department of Computer Science, Duke University, Durham, USA

Extended Abstract

Single cell genomic techniques promise to yield key insights into the dynamic interplay between gene expression and epigenetic modification. However, the experimental difficulty of performing multiple measurements on the same cell currently limits efforts to combine multiple genomic data sets into a united picture of single cell variation [1, 2]. The current understanding of epigenetic regulation suggests that any large changes in gene expression, such as those that occur during differentiation, are accompanied by epigenetic changes. This means that if cells undergoing a common process are sequenced using multiple genomic techniques, examining any of the genomic quantities should reveal the same underlying biological process. For example, the main difference among cells undergoing differentiation will be the extent of their differentiation progress, whether you look at the gene expression profiles or the chromatin accessibility profiles of the cells.

We reasoned that this property of single cell data could be used to infer correspondence between different types of genomic data. To infer single cell correspondences, we use a technique called manifold alignment. Intuitively, manifold alignment constructs a low-dimensional representation (manifold) for each of the observed data types, then projects these representations into a common space (alignment) in which measurements of different types are directly comparable [3, 4]. To the best of our knowledge, manifold alignment has never been used in genomics. However, other application areas recognize the technique as a powerful tool for multimodal data fusion, such as retrieving images based on a text description, and multilingual search without direct translation [4].

We show for the first time that it is possible to construct cell trajectories, reflecting the changes that occur in a sequential biological process, from single cell epigenetic data. In addition, we present an approach called MATCHER that computationally circumvents the experimental difficulties of performing multiple genomic measurements on a single cell by inferring correspondence between single cell transcriptomic and epigenetic measurements performed on different cells of the same type. MATCHER works by first learning a separate manifold for the trajectory of each kind of genomic data, then aligning the manifolds to infer a shared trajectory in which cells measured using different techniques are directly comparable. Because there is, in general, no actual cell-to-cell correspondence

S.C. Sahinalp (Ed.): RECOMB 2017, LNBI 10229, pp. 400–401, 2017.
DOI: 10.1007/978-3-319-56970-3

between datasets measured with different experimental techniques, MATCHER *generates* corresponding measurements by predicting what each type of measurement *would* look like at a given point in the process. Using scM&T-seq data, we confirm that MATCHER accurately predicts true single cell correlations between DNA methylation and gene expression without using known cell correspondence information.

We also downloaded publicly available single cell genomic data from a total of 4,974 single mouse embryonic stem cells grown in serum. Each cell in this dataset was individually assayed using one of four experimental techniques: RNA-seq, scM&T-seq, ATAC-seq, or ChIP-seq. We used MATCHER to infer correlations among these four measurements. This analysis gave novel insights into the changes that cells undergo as they transition from pluripotency to a differentiation primed state.

We found three main results. First, chromatin accessibility and histone modification changes largely fall into two anti-correlated categories: silencing of pluripotency factor binding sites and repression of lineage-specific genes by chromatin remodeling factors. Second, the action of pluripotency transcription factors is gradually removed by both transcriptional silencing of the genes and epigenetic silencing of the binding sites for these factors. In contrast, regulation of chromatin remodeling factor activity occurs primarily at the epigenetic level, largely unaccompanied by changes in the expression of the chromatin remodeling factors. Third, DNA methylation changes are strongly coupled to gene expression changes early in the process of differentiation priming, but the degree of coupling drops sharply later in the process.

Our work is a first step toward a united picture of heterogeneous transcriptomic and epigenetic states in single cells. MATCHER promises to be a powerful tool as single cell genomic approaches continue to generate revolutionary discoveries in fields ranging from cancer biology and regenerative medicine to developmental biology and neuroscience.

References

1. Bock, C., Farlik, M., Sheffield, N.C.: Multi-omics of single cells: strategies and applications. Trends Biotechnol. **34**(8), 605608 (2016)
2. Macaulay, I.C., Ponting, C.P., Voet, T.: Single-cell multiomics: multiple measurements from single cells. Trends Genet. **33**(2), 155–168 (2017)
3. Ham, J., Lee, D.D., Saul, L.K.: Semisupervised alignment of manifolds. In: AISTATS, p. 120127 (2005)
4. Wang, C., Mahadevan, S.: A general framework for manifold alignment. In: AAAI (2009)