

# Effects of Pseudotime Uncertainty in Downstream Analysis for Single-cell Data: A Case Study

Sumon Ahmed

Institute of Information Technology  
University of Dhaka, Dhaka - 1000, Bangladesh

## ABSTRACT

The analysis of high-throughput single-cell genomics data creates opportunities to uncover the temporal dynamics of cell populations undergoing some complex biological processes such as differentiation and cell division, where the development of bulk-based time course experiments is challenging or technically impossible. One popular approach is to learn a lower-dimensional manifold or trajectory called pseudotime through the data capturing major sources of variation. Most of the current approaches of trajectory inference only provide a point estimate of pseudotime, concealing the consequence of inherent biological and technical noise in single-cell data. Therefore, the intrinsic uncertainty in pseudotime and trajectory inference is not propagated to the downstream and its effects remain unknown. Here three commonly used pseudotime and trajectory inference methods have been discussed along with an important downstream analysis of trajectory inference, i.e. identification of gene-specific branching locations. The alteration of gene-specific branching locations in downstream is illustrated for different pseudotime and trajectory inference methods used in upstream. Finally, A set of recommendations is made that the downstream results or hypotheses should be confirmed using several pseudotime and trajectory inference methods.

## General Terms

Bioinformatics, Computational biology, Systems biology

## Keywords

Single-cell, gene expression, pseudotime, trajectory, branching

## 1. INTRODUCTION

In recent years, the field of functional genomics has advanced very rapidly and the analysis of single-cell data is now playing a very important role. The assaying of the single-cell transcriptome has gained widespread popularity over the last few years. The first single-cell RNA-sequencing (RNA-seq) study was introduced ten years back by [1], where a single 4-cell mouse blastomere was isolated manually. The RNA sequencing procedure was carried out for each cell individually. The motivation of the study was to produce embryonic samples capturing the richness of gene expression profiles compared to microarray techniques. Since this first study, there has been continuously growing interests in

assaying single-cells at higher resolution and on a larger scale. Nowadays it is possible to assay tens of thousands of cells by using droplet based techniques such as Drop-seq [2] as well as the massively parallel 10x platform [3].

This mass production of single-cell data has opened the doors to examining complex biological systems more closely, ranging from the microbial ecosystem to the genomics of human cancer [4]. In single-cell technology, the transcriptome of each cell is measured individually, in contrast to the bulk RNA-seq technology where the measurement is performed by averaging gene expression across a cell population. Averaging transcriptomes across a cell population fails to capture transcriptomic variation across individual cells. Recent studies have shown that many questions of biology can be answered in a more refined way at the single-cell level see e.g. [4, 5]. For instance, during differentiation, each cell defines its fate based on the signal received from other cells and other stimuli. Moreover, the developmental rate is not the same in each cell across a cell population, hence similar changes in transcriptomes can be observed at varying time scales for different cells. Therefore, averaging the expression profiles across a population of cells in a bulk analysis fails to mimic the true picture of the developmental and differential processes at the cellular level.

Examining the differences in gene expression at the single-cell level can facilitate the identification of novel cell types which is not possible by analysing bulk gene expression. Single-cell data have been used to identify multiple distinct as well as rare cell types [6, 7]. Single-cell sequencing has been used extensively to study differentiating cells, tracking changes in gene expression as cells progress. Early examples include identifying switch-like changes in gene expression profiles by reconstructing the trajectory of differentiating primary human myoblasts [8], as well as the identification of bi-potent progenitor cells by analysing the lineage structure of differentiating alveolar cells from mouse [9]. Single-cell sequencing has been used to study differentiating hematopoietic stem cells (HSCs) [10, 11, 12], differentiating CD4<sup>+</sup> T cells [13, 14] and more recently [15] provide a global view of mammalian organ development by using single-cell data from mice.

While studying single-cell data promises to facilitate the investigation of novel and unknown states of complex biological processes such as cell development and differentiation, it is challenging due to inherent biological and technical noise. Most of the cases, it is found useful to reduce the higher-dimensional single-cell genomics data into a lower-dimensional latent space that can capture major sources of variation in the data. Principal

and Independent Component Analysis (P/ICA) are the prevailing techniques of dimension reduction and has been used extensively in single-cell data analysis [8, 16, 17]. Multi-dimensional scaling (MDS) is another popular method of dimension reduction that tries to project the data into a lower dimension so that the distance between the original sample data is preserved as much as possible. One specific single-cell application of MDS can be found in [18] where the primary MDS axis represented the temporal order of cell progression and the secondary axis of MDS corresponded to the early responses of the cells. Gaussian Process Latent Variable Model (GPLVM) [19] is a non-linear probabilistic dimension reduction method and has been used to project single-cell data into lower dimensional space capturing non-linearity as well as inherent biological and technical noise [20, 21, 22, 23]. In some cases, the high dimensional data is reduced to a single dimension called pseudotime which represents the trajectory of cells going through some dynamic processes such as differentiation, cell division, etc. Based on the underlying biological process, the trajectory can be linear, branching, or even cyclic.

### 1.1 Pseudotime and trajectory inference

The developmental process starts from a single cell even in the most complex organisms. In biological systems, more potent cells are continuously developed and differentiated into more functionally restricted terminal cells. For example, stem cells are differentiated into neurons or skin cells [24] and a single hematopoietic stem cell (HSC) can reproduce the entire blood system [25, 26]. During development and differentiation, cells receive signals from other cells and stimuli and do not progress at the same rate defining a continuous relationship between maturing cell subsets. But this temporal information is lost during the cell capturing process. Therefore, to understand the individual cells' dynamic molecular mechanisms or to investigate the changes in gene expression levels over time it is needed to restore the ordering of cells across cell progression. One useful approach is to assign each cell a pseudotime which is a numeric value in arbitrary units [8]. Pseudotime points do not represent the actual cell capture times rather describes individual cells' progress through some dynamic processes such as development and differentiation. Gene expression dynamics then can be analysed based on the information where each cell resides in some pseudotemporal ordering which promises not only to help understand normal cell activities but also to identify and treat potential perturbation of cell developmental which may be the reason of various physical disorders [22].

In literature, a pseudotime trajectory has been represented using different formalisms. There are graph-based approaches such as Monocle [8], Wanderlust [27] and TSCAN [17] as well as curve-fitting approaches such as Embeddr [28] and Slingshot [29]. Depending on the underlying biological process the pseudotime trajectory can be linear, branching, or even cyclic. While the early methods of pseudotime estimation were limited to inferring linear trajectory only, recently developed methods such as Monocle 2 [30], Slingshot [31] and Wishbone [32] offer more flexibility and are able to identify trajectories with multiple bifurcations, where smooth or cyclic functions can be used to represent individual lineages. However, expression profiles are inherently noisy at the single-cell level. Therefore, the assignment of pseudotime points to individual cells possesses a significant amount of uncertainty which can be examined by re-estimating the pseudotime trajectory multiple times using different parameter settings and samples of the original expression data. For example, Campbell and Yau [22] have applied the Monocle algorithm on random subsets of the data. They

have shown that the pseudotime assigned by the Monocle algorithm for the same cell differs significantly across different subsets of the data used.

Since most of the pseudotime and trajectory inference methods developed to date only provide a single point estimate of pseudotime ignoring the associated uncertainty, therefore this uncertainty is not propagated and its effects in the downstream are not well known. The downstream of trajectory inference is to identify genes that are differentially expressed across different lineages in the trajectory. It is also important to identify the gene-specific branching or bifurcation points. Branching points are the locations in pseudotime where the expression profiles of individual genes start to diverge for the first time. Genes can be ranked based on the branching points information which promises to help to identify early and late branching genes and to understand each gene's role during cellular progression. Here, the downstream effects of pseudotime and trajectory inference uncertainty in identifying gene-specific branching points have been investigated. Three well known and popular trajectory inference algorithms Wishbone, Monocle 2, and Slingshot are compared, which finds substantial differences in gene order based on branching points for different upstream methods.

## 2. METHODS

Gene expression data are intrinsically noisy at the single-cell level which leads to the uncertainty in pseudotime and trajectory inference. Therefore, different pseudotime and trajectory inference methods can provide different results in downstream analysis. To investigate this, three widely used methods of trajectory inference Wishbone, Monocle 2, and Slingshot are compared. These methods have been applied on a publicly available single-cell RNA-seq dataset from mouse [10] to generate pseudotime and lineage-specific assignments of individual cells. The information gained by using three trajectory inference methods has been used in downstream to identify gene-specific branching locations.

### 2.1 Pseudotime and trajectory inference methods

**Wishbone** is a graph-based algorithm and can identify cellular bifurcation. It assumes that the trajectory has only one cellular branching point. Wishbone uses cell progression information to estimate a pseudotime for each cell as well as to identify the cellular branching point. At first, a K Nearest Neighbours (KNN) graph of cells is constructed and a shortest path algorithm is used to find an initial cell ordering. The cell ordering is iteratively re-estimated using a set of randomly selected points through the pseudotime trajectory termed waypoints. The inconsistencies among the waypoints indicate the presence of a cellular bifurcation or branching points. Finally, Wishbone uses t-Distributed Stochastic Neighbour Embedding (tSNE) to project the inferred trajectory in a lower-dimensional space where each cell is assigned to either pre-branching state or post-branching state based on their position in pseudotime trajectory.

**Monocle 2** is an unsupervised algorithm and does not need to know beforehand the marker genes or the number of lineages or branches describing the biological process. It uses reversed graph embedding (RGE) and identifies a global topology with multiple lineages by learning a principal graph [33]. At first, genes that are differentially expressed among cell clusters are selected where the cell clusters are generated using the density peak clustering algorithm. After the genes are selected, the algorithm finds a lower-dimensional representation using the DDRTree algorithm.

Monocle 2 then selects a set of centroids of the data and constructs a spanning tree by joining them. The algorithm iteratively refines the tree by adjusting each node's position as well as reconstructing the spanning trees. Finally, when the tree is learned, the user needs to select a root. Each cell is assigned a pseudotime value based on its distance from the root and a branch label or state based on its position on the principal graph.

**Slingshot** is a Euclidean distance-based curve-fitting approach of trajectory estimation. It does not need the number of branches or lineages to be pre-specified although it allows optional incorporation of prior information about the terminal states of lineages. Incorporating terminal cell clusters information does not restrict the number of lineages rather it imposes a local constraint on the minimum spanning tree algorithm and helps the algorithm to identify a more biologically meaningful global structure. The algorithm works in two steps. At first, a global lineage topology is constructed. Cells are grouped into clusters and a minimum spanning tree is constructed connecting these clusters. Cell clusters are then ordered based on a pre-defined root cluster. Thus, a global lineage structure of clusters is generated where each lineage has a unique terminal cluster but shares a common initial cluster. Finally, Slingshot uses simultaneous principal curves to smooth all lineages and refines each cell's assignment to individual lineages. Each cell is assigned a lineage-specific pseudotime value as well as lineage-specific weights representing cells belonging to different branches or lineages.

## 2.2 Identification of gene-specific branching location

To identify gene-specific branching locations the Detime model [34] is used. Detime is a Gaussian Process (GP) [35] based approach developed to identify the first point where two time course profiles start to diverge from one another. The idea is to design a joint covariance or kernel function  $k$  for two latent GP functions  $g(t) \sim \mathcal{GP}(0, k)$  and  $h(t) \sim \mathcal{GP}(0, k)$  constrained to intersect at the branching point  $t_b$ .

Let  $y^t(\mathbf{T})$  and  $y^b(\mathbf{T})$  are the noisy measurements of data from two lineages representing the trunk (baseline) and the diverging branch respectively. Before the branching time  $t_b$ , both  $y^t(\mathbf{T})$  and  $y^b(\mathbf{T})$  are considered as the noisy versions of the same underlying mean function  $g(t)$ ,

$$y^t(t_n) = g(t_n) + \epsilon, \quad (1)$$

$$y^b(t_n) = g(t_n) + \epsilon \quad \text{for } t_n \leq t_b, \quad (2)$$

where  $\epsilon \sim \mathcal{N}(0, \sigma_{\text{noise}}^2)$  is the i.i.d. Gaussian noise with variance  $\sigma_{\text{noise}}^2$ . After the branching point  $t_b$ , the mean function of the trunk  $y^t$  remains the same and follows  $g(t)$  while the mean function of the diverging branch  $y^b$  changes to follow  $h(t)$ ,

$$y^t(t_n) = g(t_n) + \epsilon, \quad (3)$$

$$y^b(t_n) = h(t_n) + \epsilon \quad \text{for } t_n > t_b. \quad (4)$$

The branching time  $t_b$  becomes a hyperparameter of the joint covariance function of this model along with the hyperparameters of the GP functions, i.e. lengthscale  $l$  and process variance  $\sigma^2$  (see [35] for details about kernel hyperparameters of GP functions). The lengthscale and process variance hyperparameters can be estimated prior to inferring the branching time. These hyperparameters are shared across the data from both lineages and can be estimated reasonably by fitting two separate GP regression models to the data from both lineages. To do so, the likelihood has been maximised for the case where branching time  $t_b$  approaches to  $-\infty$ , which corresponds to two independent GP regression models for the data

from both lineages [34],

$$\hat{\theta} = \operatorname{argmax}_{\theta} \left( \lim_{t_p \rightarrow -\infty} p_{\theta}(y^t(\mathbf{T}), y^b(\mathbf{T})|t_b, \theta) \right), \quad (5)$$

where  $\theta = \{l, \sigma^2\}$ . This reduces the problem to estimate a single parameter branching time  $t_b$  only, which can be done easily by using a simple histogram approach [34]. As in [34], a simple discretisation  $t_b \in [t_{\min}, t_{\min} + \delta, t_{\min} + 2\delta, \dots, t_{\max}]$  is used and the posterior distribution over the branching location is estimated using the normalised likelihood evaluated at each grid point,

$$p(t_b|y^t(\mathbf{T}), y^b(\mathbf{T})) \simeq \frac{p(y^t(\mathbf{T}), y^b(\mathbf{T})|t_b)}{\sum_{t=t_{\min}}^{t=t_{\max}} p(y^t(\mathbf{T}), y^b(\mathbf{T})|t)}. \quad (6)$$

As the posterior distribution of the branching time  $t_b$  is represented using a simple histogram approach, the point estimate of the branching time can easily be calculated by using mean, median, or mode (MAP) [34].

## 3. RESULTS

Three trajectory inference algorithms Wishbone, Monocle 2, and Slingshot have been applied on the single-cell RNA-seq of haematopoietic stem cells (HSCs) from mouse [10] to infer pseudotemporal ordering of cells as well as cell assignment to individual branches or lineages. This dataset contains cells that have been differentiated into two precursor cell types namely myeloid and erythroid. Paul et al. [10] have tracked gene expression changes for myeloid progenitors and produced a reference compendium of marker genes related to the progression of myeloid progenitors into erythrocytes and some other types of leukocytes. As this dataset has two major lineages, the cells that do not belong to either of the major branches have not been considered in our analysis. Therefore, outcomes of applying trajectory inference algorithms on the data are the pseudotemporal ordering of cells as well as the assignment of cells to either of the lineages.

To investigate the trajectory inference uncertainty, top 30 differentially expressed genes that show the highest level of branching evidence, have been selected [36]. The expression profiles of each gene along with the inferred pseudotime and lineage-specific cell assignment (using Wishbone, Monocle 2, and Slingshot) have been fed to the Detime model to identify gene-specific branching locations for these 30 genes. Mode (MAP), median, and mean of the posterior distribution of branching point are used to calculate a point estimate of the branching time  $t_b$ . Finally, genes are ranked based on the point estimate of the branching time for all cases. Fig. 1, shows the comparison of gene rankings for each pair of trajectory inference methods, where different branching point based gene rankings for different trajectory inference methods used in upstream are very evident. For instance, consider the bio-marker *Car1*, it has been identified as an early branching gene (rank 1) by Slingshot where both Wishbone and Monocle 2 have identified *Car1* as a late branching gene (rank 21). Similar comparisons can be made for other genes. Overall, a better similarity of gene rankings between Wishbone and Monocle 2 (Fig. 1 (a)) is found than between one of them with Slingshot (Fig. 1 (b) and (c)). The reason may be the case that both Wishbone and Monocle 2 are graph-based algorithms whereas Slingshot is a curve-fitting based approach. It is very evident from Fig. 1 that the intrinsic stochasticity of single-cell RNA-seq data and uncertainty in the pseudotemporal ordering of cells through different lineages

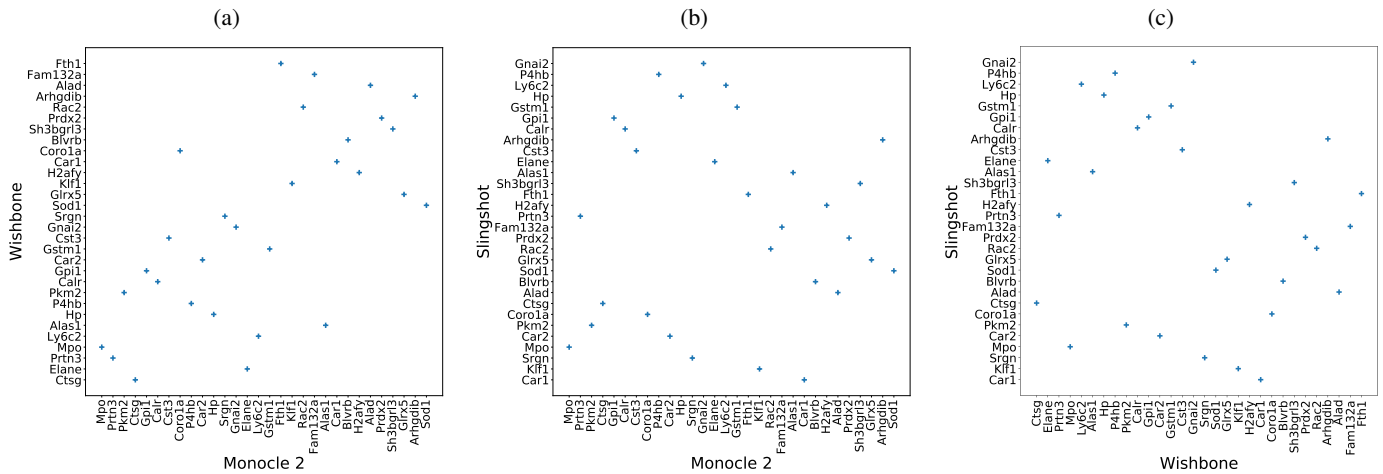


Fig. 1. Impacts of pseudotime and trajectory inference uncertainty in the downstream analysis of the single-cell RNA-seq of haematopoietic stem cells (HSCs) from mouse [10]. Three popular and commonly used pseudotime and trajectory inference methods Wishbone, Monocle 2, and Slingshot have been used to generate pseudotime as well as lineage-specific cell assignments to be used in downstream for identifying gene-specific branching locations for top 30 differentially expressed genes selected based on their branching evidence. Genes are ranked according to their branching points for all three trajectory inference methods. The gene rankings, i.e. gene-specific branching locations are altered in downstream for different pseudotime and trajectory inference methods used in upstream. The inconsistency in gene rankings for each pair of trajectory inference methods illustrates the effects of pseudotime uncertainty in downstream analysis. (a) Wishbone vs. Monocle 2, (b) Monocle 2 vs. Slingshot, (c) Wishbone vs. Slingshot.

have altered the ranking of differentially expressed genes, hence affected the downstream analysis.

#### 4. CONCLUSION

Gene expression profiles can be extremely noisy at the single-cell level imposing a number of challenges on the pseudotime and trajectory inference algorithms. Both biological and technical variability introduces a significant amount of uncertainty, therefore point estimate based trajectory inference algorithms may fail to mimic the true picture of the underlying biological process in downstream analysis. Here an investigation has been carried out on how the choice of a particular trajectory inference method may affect the downstream analysis. One of the important downstream analyses of trajectory inference is to identify genes that are differentially expressed across different lineages or branches of the global topology. It is also necessary to infer the gene-specific branching locations where the expression profiles of genes start to diverge for the first time. Further genes can be ranked based on the gene-specific branching location indicative of the role of individual genes during development and differentiation processes.

The impacts of three widely used pseudotime and trajectory inference algorithms Wishbone, Monocle 2, and Slingshot in downstream analysis are investigated. These methods are applied on the single-cell RNA-seq of haematopoietic stem cells (HSCs) from mouse to generate lineage-specific cell assignment as well as pseudotemporal ordering of cells to be examined in downstream. Gene-specific branching locations have been identified for top 30 differentially expressed genes that show the highest level branching evidence. These genes have been ranked based on their branching points for all three methods. Different gene-rankings are found for different trajectory inference methods used in upstream indicative of the effects of pseudotime and trajectory inference uncertainty in downstream analysis.

Single-cell data are continuously growing and are becoming available in larger volumes, opening new opportunities to examine the complex dynamics of biological processes where the production of time-course experiments is technically impossible or at least challenging. While a flurry of methods has been developed to model dynamic processes, the effects of trajectory inference uncertainty in the downstream analysis are not well studied. The performance of a pseudotime and trajectory inference method heavily depends on the type of trajectory under consideration and the choice of a particular method is primarily motivated by the anticipated trajectory structure in the data. But in most of the cases, the users know very little about the underlying trajectory structure that properly describes the data. Therefore, it is suggested that inferred trajectory, as well as downstream results for the data, should be confirmed by multiple pseudotime and trajectory inference methods. An interesting future extension of the current work is to apply other pseudotime and trajectory inference algorithms on a number of datasets having a different topological structure such as disconnected trajectories, cycles, etc., hence developing a comprehensive framework that will facilitate the researchers to examine and confirm their hypotheses about the underlying biological processes using several pseudotime and trajectory inference algorithms.

#### 5. REFERENCES

- [1] Fuchou Tang, Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, Nanlan Xu, Xiaohui Wang, John Bodeau, Brian B Tuch, Asim Siddiqui, et al. *mrna-seq whole-transcriptome analysis of a single cell. Nature methods*, 6(5):377, 2009.
- [2] Evan Z Macosko, Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, Allison R Bialas, Nolan Kamitaki, Emily M Martersteck, et al. *Highly parallel genome-wide expression profiling*

- of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214, 2015.
- [3] Grace XY Zheng, Jessica M Terry, Phillip Belgrader, Paul Ryvkin, Zachary W Bent, Ryan Wilson, Solongo B Ziraldo, Tobias D Wheeler, Geoff P McDermott, Junjie Zhu, et al. Massively parallel digital transcriptional profiling of single cells. *Nature communications*, 8:14049, 2017.
- [4] Charles Gawad, Winston Koh, and Stephen R Quake. Single-cell genome sequencing: current state of the science. *Nature Reviews Genetics*, 17(3):175, 2016.
- [5] Byungjin Hwang, Ji Hyun Lee, and Duhee Bang. Single-cell rna sequencing technologies and bioinformatics pipelines. *Experimental & molecular medicine*, 50(8):1–14, 2018.
- [6] Dominic Grün, Anna Lyubimova, Lennart Kester, Kay Wiebrands, Onur Basak, Nobuo Sasaki, Hans Clevers, and Alexander van Oudenaarden. Single-cell messenger rna sequencing reveals rare intestinal cell types. *Nature*, 525(7568):251, 2015.
- [7] Aleksandra A Kolodziejczyk, Jong Kyoung Kim, Valentine Svensson, John C Marioni, and Sarah A Teichmann. The technology and biology of single-cell rna sequencing. *Molecular cell*, 58(4):610–620, 2015.
- [8] Cole Trapnell, Davide Cacchiarelli, Jonna Grimsby, Prapti Pokharel, Shuqiang Li, Michael Morse, Niall J Lennon, Kenneth J Livak, Tarjei S Mikkelsen, and John L Rinn. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology*, 32(4):381–386, 2014.
- [9] Barbara Treutlein, Doug G Brownfield, Angela R Wu, Norma F Neff, Gary L Mantalas, F Hernan Espinoza, Tushar J Desai, Mark A Krasnow, and Stephen R Quake. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell rna-seq. *Nature*, 509(7500):371, 2014.
- [10] Franziska Paul, Yaara Arkin, Amir Giladi, Diego Adhemar Jaitin, Ephraim Kenigsberg, Hadas Keren-Shaul, Deborah Winter, David Lara-Astiaso, Meital Gury, Assaf Weiner, et al. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell*, 163(7):1663–1677, 2015.
- [11] Monika S Kowalczyk, Itay Tirosh, Dirk Heckl, Tata Nageswara Rao, Atray Dixit, Brian J Haas, Rebekka K Schneider, Amy J Wagers, Benjamin L Ebert, and Aviv Regev. Single-cell rna-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome research*, 25(12):1860–1872, 2015.
- [12] Fan Zhou, Xianlong Li, Weili Wang, Ping Zhu, Jie Zhou, Wenyan He, Meng Ding, Fuyin Xiong, Xiaona Zheng, Zhuan Li, et al. Tracing haematopoietic stem cell formation at single-cell resolution. *Nature*, 533(7604):487, 2016.
- [13] Tapio Lönnberg, Valentine Svensson, Kylie R James, Daniel Fernandez-Ruiz, Ismail Sebina, Ruddy Montandon, Megan SF Soon, Lily G Fogg, Arya Sheela Nair, Urijah Liligeto, et al. Single-cell rna-seq and computational analysis using temporal mixture modelling resolves th1/tfh fate bifurcation in malaria. *Science immunology*, 2(9), 2017.
- [14] Michael JT Stubbington, Tapio Lönnberg, Valentina Proserpio, Simon Clare, Anneliese O Speak, Gordon Dougan, and Sarah A Teichmann. T cell fate and clonality inference from single-cell transcriptomes. *Nature methods*, 13(4):329, 2016.
- [15] Junyue Cao, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M Ibrahim, Andrew J Hill, Fan Zhang, Stefan Mundlos, Lena Christiansen, Frank J Steemers, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, 566(7745):496, 2019.
- [16] Guoji Guo, Mikael Huss, Guo Qing Tong, Chaoyang Wang, Li Li Sun, Neil D Clarke, and Paul Robson. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Developmental cell*, 18(4):675–685, 2010.
- [17] Zhicheng Ji and Hongkai Ji. TSCAN: Pseudo-time reconstruction and evaluation in single-cell rna-seq analysis. *Nucleic acids research*, 44(13), 2016.
- [18] Tsukasa Kouno, Michiel de Hoon, Jessica C Mar, Yasuhiro Tomaru, Mitsuoaki Kawano, Piero Carninci, Harukazu Suzuki, Yoshihide Hayashizaki, and Jay W Shin. Temporal dynamics and transcriptional control using single-cell gene expression analysis. *Genome biology*, 14(10):R118, 2013.
- [19] Neil Lawrence. Probabilistic non-linear principal component analysis with gaussian process latent variable models. *Journal of Machine Learning Research*, 6(Nov):1783–1816, 2005.
- [20] John E Reid and Lorenz Wernisch. Pseudotime estimation: deconfounding single cell time series. *Bioinformatics*, 32(19):2973–2980, 2016.
- [21] Florian Buettner and Fabian J Theis. A novel approach for resolving differences in single-cell gene expression patterns from zygote to blastocyst. *Bioinformatics*, 28(18):i626–i632, 2012.
- [22] Kieran Campbell and Christopher Yau. Order under uncertainty: robust differential expression analysis using probabilistic models for pseudotime inference. *PLoS Computational Biology*, 12(11), 2016.
- [23] Sumon Ahmed, Magnus Rattray, and Alexis Boukouvalas. Grandprix: scaling up the bayesian gplvm for single-cell data. *Bioinformatics*, 35(1):47–54, 2019.
- [24] John T Dimos, Kit T Rodolfa, Kathy K Niakan, Laurin M Weisenthal, Hiroshi Mitsumoto, Wendy Chung, Gist F Croft, Genevieve Saphier, Rudy Leibel, Robin Goland, et al. Induced pluripotent stem cells generated from patients with als can be differentiated into motor neurons. *science*, 321(5893):1218–1221, 2008.
- [25] Ryohichi Sugimura, Deepak Kumar Jha, Areum Han, Clara Soria-Valles, Edroaldo Lummertz Da Rocha, Yi-Fen Lu, Jeremy A Goettel, Erik Serrao, R Grant Rowe, Mohan Malleshaiah, et al. Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature*, 545(7655):432, 2017.
- [26] Raphael Lis, Charles C Karrasch, Michael G Poulos, Balvir Kumar, David Redmond, Jose G Barcia Duran, Chaitanya R Badwe, William Schachterle, Michael Ginsberg, Jenny Xiang, et al. Conversion of adult endothelium to immunocompetent haematopoietic stem cells. *Nature*, 545(7655):439, 2017.
- [27] Sean C Bendall, Kara L Davis, El-ad David Amir, Michelle D Tadmor, Erin F Simonds, Tiffany J Chen, Daniel K Shenfeld, Garry P Nolan, and Dana Peer. Single-cell trajectory detection uncovers progression and regulatory coordination in human b cell development. *Cell*, 157(3):714–725, 2014.
- [28] Kieran Campbell, Chris P Ponting, and Caleb Webber. Laplacian eigenmaps and principal curves for high resolution

- pseudotemporal ordering of single-cell rna-seq profiles. *bioRxiv preprint*, page 027219, 2015.
- [29] Jaehoon Shin, Daniel A Berg, Yunhua Zhu, Joseph Y Shin, Juan Song, Michael A Bonaguidi, Grigori Enikolopov, David W Nauen, Kimberly M Christian, Guo-li Ming, et al. Single-cell rna-seq with waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell*, 17(3):360–372, 2015.
- [30] Xiaojie Qiu, Andrew Hill, Yi-An Ma, and Cole Trapnell. Single-cell mrna quantification and differential analysis with census. *Nat Meth*, pages 309–315, 2016.
- [31] Kelly Street, Davide Risso, Russell B Fletcher, Diya Das, John Ngai, Nir Yosef, Elizabeth Purdom, and Sandrine Dudoit. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics*, 19(1):477, 2018.
- [32] Manu Setty, Michelle D Tadmor, Shlomit Reich-Zeliger, Omer Angel, Tomer Meir Salame, Pooja Kathail, Kristy Choi, Sean Bendall, Nir Friedman, and Dana Pe’er. Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nature biotechnology*, 34(6):637, 2016.
- [33] Alexander N Gorban and Andrei Y Zinovyev. Principal graphs and manifolds. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 28–59. IGI Global, 2010.
- [34] Jing Yang, Christopher A Penfold, Murray R Grant, and Magnus Rattray. Inferring the perturbation time from biological time course data. *Bioinformatics*, 32(19):2956–2964, 2016.
- [35] Carl Edward Rasmussen and Christopher KI Williams. *Gaussian processes for machine learning*. MIT press Cambridge, 2006.
- [36] Alexis Boukouvalas, James Hensman, and Magnus Rattray. BGP: identifying gene-specific branching dynamics from single-cell data with a branching gaussian process. *Genome biology*, 19(1):65, 2018.