SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

BIO304: SYSTEMS BIOLOGY

Single-cell Trajectory Inference Approaches

Author:

Siyuan GUO 11611118

Supervisor:

Prof. Wei HUANG

Department of Biology

May 30, 2019

Abstract

Since the discovery of the cell in 1665, scientists have been working to classify cells and build cell lineage. In the process, scientists have developed a variety of methods and made some great achievements. In recent years, single-cell sequencing technology has developed rapidly, and researchers have developed a number of single-cell trajectory inference tools that can use single-cell sequencing data to classify cells and build cell lineage, turning research of cellular dynamic processes into analysis of the pseudo-time series and greatly improved analysis throughput. Some ambitious scientists also hope to use this technology to construct "The Human Cell Atlas". The existing scTI methods have achieved many achievements. In some simple cell differentiation systems, scientists have used these methods to find answers to many important biological problems. The paper also lists 20 commonly used scTI methods and 7 basic topologies in the cell lineage, and organizes and summarizes the basic information of these methods and their adaptability to various basic topologies. At the same time, in fact, cell lineage studies based on single-cell biology are still in the early stages of development, and scientists need more powerful sequencing techniques and better-performing core algorithms to further develop these scTI research strategies. In future biological research, especially after scientists have successfully mapped the complete "The Human Cell Atlas" and the cell lineage of various model animals, single-cell-based cell lineage and classification studies will provide a powerful boost to the development of biology.

Contents

1 Background

1.1 Cell

Modern biology theory defines the cells discovered by Hooke in 1665 as the most basic unit of organic organisms. Modern cell theory consists of three parts, all living things are composed of cells and cell products; cells are the most basic unit of life structure and function; all new cells are derived from old cells.

1.1.1 Cell Lineage

Cell lineage refers to the developmental history of blastomeres from the time of the first cleavage until the final differentiation into tissues and organ cells. The division of fertilized eggs in many animals is carried out in a strict format. In this process, the moment, the order, and the location in the space of the splitting balls' generation are specified. This kind of inter-cell relationship in the developmental generation is like the lineage of the human family, so it is called the cell lineage [36, 39].

The study of cell lineage plays an important role in understanding the relationship between the uneven distribution of egg quality and the developmental fate of blastomeres, as well as the evolutionary relationship between the early development of different species of animals. In the past many years, researchers have also begun to focus on the pedigree research of abnormal developmental cells such as cancer cells, in order to systematically understand its occurrence process and mechanism [39].

1.1.2 Cell Classification

For a long time, scientists have been working to discover and classify a wide variety of cells. At the same time, they are also committed to a detailed description of the characteristics of each cell type. However, due to the limitations of technology, scientists can only distinguish different types of cells by their shape, their position in the organism and their relative relationship with other cells.

After various chemical stains were discovered, scientists began to classify and differentiate them using the performance of various cells under different stains. The 1906 Nobel Prize in Physiology or Medicine was awarded to two scientists who used staining techniques and anatomical techniques to study the diverse structures of the brain.

Due to the imaging limitations of optical microscopy, in order to better study cell structure and other characteristics, scientists began to seek more powerful microscopic imaging tools. Since the 1930s, electron microscopy, which has been able to provide imaging magnifications of up to 5000 times, has become one of the important tools for biological research. In the rapid development phase of subsequent science, various modern biotechnologies such as FACS (fluorescence activated cell sorting) and FISH (fluorescence in situ hybridization) help researchers distinguish various cell types.

Today, we have so many techniques and methods to classify cells, and it seems that scientists are about to realize their ultimate pursuit of cell sorting. But the facts are not always satisfactory. In fact, the various results we have achieved on cell classification can only be considered to be fragmented and one-sided. The various classification models we have are often based on different detection indicators and evaluation criteria, and the correlation between these indicators is difficult to be strictly defined and evaluated, so the integration of different classification models becomes an extremely difficult task. So far, we have established a complete cell lineage of fertilized eggs to mature bodies only in *C.elegan*. It is a kind

of extremely simple creature and its mature individuals have only about 1000 cells. At the same time, the number of cells in different mature individuals of the same sex is the same. This is a rare character which is very meaningful in building cell lineage.

At the same time, due to the confusion of various detection methods and evaluation methods, we lack a strict definition of some basic concepts in the cell classification problem, which may cause confusion and trouble in some cases. The booming single-cell omics technology in recent years provides researchers with a comprehensive set of comprehensive indicators. Many researchers believe that this set of techniques can provide data-based definitions for different cell types, which greatly increases the accuracy of the definition. Also, single-cell multi-omics technology provides a platform on which scientists can re-integrate the various cell maps we know into a huge one with adding more single-cell level details to it. And they can also find that some cell types and cell development pathways which have not been discovered.

1.2 Single-cell Biology and Cell Lineage

1.2.1 Single-cell Omics

As one of the most basic unit concepts of life, cells are the cornerstone of life activities. Although biologists have been working under a microscope for nearly 180 years, we still don't know much about cells. We expect effective technical means to completely examine the composition of individual cells to identify and treat diseases at the cellular and even molecular levels [37, 41].

Single-cell sequencing refers to the technique of obtaining data and analyzing information by sequencing the levels of genomes, transcriptomes and epigenetic-genomes of individual cells. Single-cell sequencing can solve the problem of cellular heterogeneity that cannot be solved by traditionally sequencing mixed tissue samples, and provides a new method for

analyzing the behavior, mechanism and relationship between individual cells and the body [37, 39].

In conventional tissue RNA sequencing methods, the signals produced by different kinds of cells are averaged. Throughout the process, the difference between cells and cells, also known as cytoplasmic heterogeneity, was ignored. We can't find two identical cells in nature, and scRNA-seq provides us with the opportunity to discover the tiny differences between these cells. Using a scRNA-seq data-driven approach, researchers also have the chance to discover some new cell types [39]. Single-cell genome and epigenetic genome sequencing can identify cell genomes. The purpose of the genome approach is to identify the entire genome or capture a specific predefined region. Epigenetic methods can capture specific predefined sequences based on unique histone modification (scChIP-Seq), genome openness (scATAC-Seq), or the same identification of DNA methylation patterns (scDNAme-Seq) or 3D chromosome-structures (scHi-C) [36]. A combination of barcode-strategies is now used to capture tens of thousands of single cells. Single-cell epigenetic genomics methods usually study only the nucleus, so frozen or certain fixed samples could be used.

1.2.2 The Human Cell Atlas

Scientists have long recognized the need for a deeper understanding of cells, but until recently, with the rapid development of single-cell omics technology, the creation of a complete and systematic human cell map has become a viable goal. In 2017, researchers from different countries proposed an initiative to open a project called "The Human Cell Atlas" [39]. This is an ambitious project that, like the Human Genome Project, attempts to provide data support for fundamentally solving biological problems.

1.2.3 Single-cell Trajectory Inference (scTI)

In traditional studies of cellular dynamic processes (such as cell cycle, cell differentiation, and cell activation), scientists rely mainly on tracking cells calibrated in multiple sets of parallel experiments at different time nodes, a method that is undoubtedly very complex and time-consuming. Cells reach the cell types they finally differentiate to by distinguishing between asynchronous branching pathways. The material basis of this differentiation process is the change of molecular characteristics within cells, especially the regulation of different gene expression levels by cells at different times. The method based on single-cell omics (such as transcription, proteomics, and epigenetic genomics) is actually a pseudo-time sequence analysis strategy, which means that researchers can analyze the entire cellular dynamic process with few static samples [7, 42].

These dynamic process can be inferred by a trajectory inference (TI) method [36, 39]. It sorts cells along the trajectory according to the similarity of cell expression patterns. Through this technique, single cell cytology can reconstruct the evolution process of cells into a trajectory of high-dimensional space [20, 21, 48]. In general, trajectories produced by this method are linear, tree-shaped or bifurcated. But some of the latest methods also support the formation of trajectories consisting of more complex topologies, such as ring structures [35] and structures containing broken connections [52]. TI methods can help us to fully understand the dynamics of cells from a single-cell omics perspective [42], objectively identify different cell subtypes and subsets of cells [18] and map the corresponding cell differentiation trees [40, 43], and infer cell-cell interactions regulating differentiation bifurcations between cells [32].

If researchers had large enough single-celled sample sets, they would also have the opportunity to discover phenomena that could not be observed in normal cellular development dynamics systems. The more complex the cell lineage path, the more intersections there are, the larger the sample set is needed to study the above anomalies. At the same time, reasonable selection of clustering or regression algorithm to classify cells is beneficial for

researchers to observe rare cell intermediate states and cell subtypes. When the cellular lineage of the entire sample set is constructed, researchers can analyze the distribution of cells on different pathways along the lineage direction to obtain the relative duration and clustering size of each cell development stage, and analyze the subset of progenitor cells in the whole dataset [39].

2 Current Research Achievements

2.1 ScTI Methods are Constantly Being Developed

In the past few years, single-cell omics technology has flourished, and more and more scTI methods have been invented. Every month, new scTI methods are published, and researchers from around the world are continually experimenting with these new methods to get the most complete cell map. In the repository of commonly used single-cell omics tools [9, 45, 50], it is not difficult to find that the scTI tool is one of the largest categories of current single-cell omics tools. The core algorithms of each scTI are different, which means that the prior knowledge they rely on and their inferable trajectory structures are dissimilar. Also, different methods often have their own unique output structure.

2.2 ScTI Methods Analyzing Different Cell Lineage

Current computational methods have proven useful for analyzing cell lineage and corresponding trajectories based on large numbers of single-cell omics data, but these strategies still have limitations on many issues. We need better algorithms to derive multi-branched structures, to achieve more efficient extraction of cell features, and to take into account multiple pathways in order to show the fact that the same cell in a cell lineage may follow multiple dynamic paths simultaneously [3]. In the study of cell lineage with simple topologies, researchers have achieved many results, such as inferring cell lineage during differentiation of B cells by single-cell proteomic data [5], and studying the lineage of nervous system development [19, 23, 25] and early hematopoietic process [28] with single-cell transcriptome

data.

Of course, in more complex cell differentiation systems, cell lineage constructed using single-cell omics can also reveal the answers to important biological questions. Studies of embryonic stem cells have helped us understand embryonic development at the cellular level and find marker molecules for different cells at specific stages of embryonic development [15, 26]. Researchers who focus on bone marrow cells solve the problem that has plagued the academic world for many years through the method of single-cell omics: whether hematopoietic stem cells in the bone marrow have differentiated preferences after maturity and tend to differentiate to a certain kind of cell ? [17, 29]

2.2.1 Hematopoietic System

Previous studies have shown that in the hematopoietic system, the use of scTI methods to infer cell lineage is quite appropriate. In the single-cell data on hematopoietic cells, the researchers accurately isolated hematopoietic stem cells and progenitor cells (HSPCs) from single-cell data from acute myeloid leukemia by analyzing data from normal hematopoietic cells. This method is more accurate than traditional methods. Because traditional methods are based on classical cell surface markers, in some cases, such strategies do not accurately identify diseased cells in a disease. Single-cell omics data provides ultra-high-dimensional feature information, which makes feature-based recognition more accurate [16].

2.2.2 Cancer Research

Single-cell omics has revolutionized the entire field of cancer research. In the field of single cells just introduced into cancer research, qPCR-based single cell methods have been used to study radiation resistance of cancer cells and heterogeneity of colon cancer tissues at the cellular level [1, 2]. With the rise of second-generation sequencing technology, single-cell omics analysis provides new tools for researchers studying breast cancer and acute lymphoblastic

leukemia [8, 13]. On this basis, researchers can also infer the order in which various mutations lead to cell carcinogenesis [4, 6].

Analysis of single-cell RNA-seq data from some fresh tumor tissues can distinguish epithelial cells, immune cells, stromal cells and cancer cells. This method has achieved very good results in melanoma [30], myeloproliferative neoplasms [34] and glioblastoma [11]. Among the identified cancer cells, single-cell transcriptome data can also be used to distinguish cancer cells of different states, such as cancer stem cells [11, 31] and resistant cancer cells [30]. In cancer stem cells, cells in an active value-added state and cells in a relatively static state can also be identified [11, 30, 31].

2.3 Summery of Commonly Used scTI Methods

In the next sections, I have selected 20 commonly used scTI methods and divided them into different groups according to the different characteristics of their core algorithms. These are all based on *Python* or *R*. In Table 2.1 below, I listed the priori requirements, basing platform, topology features and references of these methods. In Table 2.2, seven basic inferable trajectory types of scTI methods are defined, however, not all scTI methods are applicable to all of these topologies. When it comes to Figure 2.1, these inferable trajectory types of scTI methods are represented in cartoons. Last, I show the inferable trajectory types of every scTI method in Table 2.3.

These methods are different from each other and have their own characteristics. In the course of practical research, researchers tend to combine results from multiple methods in order to achieve a satisfying end result.

2.3.1 Commonly Used scTI Methods

TABLE 2.1: Commonly Used scTI Methods

Method	Topology	Priori	Platform	Reference
Tree				
Monocle_1	Flexible	△	*R*	[12]
Monocle_2	Unfettered		*R*	[38]
Slingshot	Unfettered		*R*	[49]
MST	Unfettered		*R*	*R**
SCUBA	Unfettered		*Python*	[10]
pCreode	Unfettered		*Python*	[47]
Linear				
Embeddr	Constant		*R*	[14]
TSCAN	Constant		*R*	[27]
SCORPIUS	Constant		*R*	[22]
Component_1	Constant		*R*	*R**
MATCHER	Constant		*Python*	[44]
Multi-diverging				
STEMNET	Flexible	▲	*R*	[43]
MFA	Flexible	△	*R*	[33]
FateID	Flexible	▲	*R*	[46]
Bi-diverging				
DPT	Constant		*R*	[26]
Wishbone	Flexible	△	*Python*	[5]
Graph				
RaceID	Unfettered		*R*	[24]
PAGA	Unfettered	△	*Python*	[52]
Cyclic				
ElPiGraph	Constant		*R*	**
Angle	Constant		*R*	*R**

△ : Needing priori information like start or end cells in lineage.
▲ : Needing priori information like cell clustering or time series.
Unfettered : Topological structures deduced form data are free.
Constant : Topological structures deduced form data are constant.
Flexible : Topological structures deduced form data depend on the parameters.
*R** : This method could be implemented with a bit of R code.
** : Only in `github.com/Albluca/ElPiGraph.R`.

2.3.2 Basic Trajectory Structure Types

TABLE 2.2: The definition of basic trajectory structure types [51]

Types	Definition
Linear	A graph that every node in this graph has an in-degree and an out-degree not higher than 1 and 2 nodes in this graph have degrees equal to 1.
Ring	A graph that every node in this graph has an in-degree and an out-degree equal to 1.
Tree	A graph that every node in this graph has an in-degree lower than 1.
Mutil-diverging	A tree graph that every node except one in this tree graph has a degree not higher than 1.
Bi-diverging	A mutil-diverging graph where a node with its degree equal to 3.
Unconnected	A graph where not all nodes are connected.
Connected	A graph where all nodes are connected.

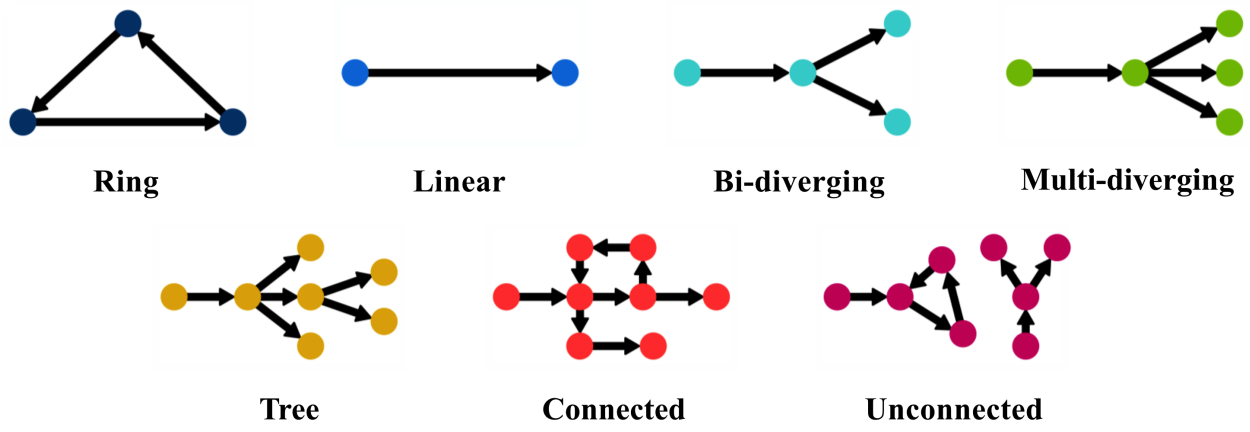

FIGURE 2.1: Seven basic trajectory structure types

2.3.3 Inferable Trajectory Structures of scTI Methods

TABLE 2.3: Inferable Trajectory Structures of scTI Methods

Method	R	L	B	M	T	C	U
Tree							
Monocle_1	□	■	■	■	■	□	□
Monocle_2	□	■	■	■	■	□	□
Slingshot	□	■	■	■	■	□	□
MST	□	■	■	■	■	□	□
SCUBA	□	■	■	■	■	□	□
pCreode	□	■	■	■	■	□	□
Linear							
Embeddr	□	■	□	□	□	□	□
TSCAN	□	■	□	□	□	□	□
SCORPIU	□	■	□	□	□	□	□
Component_1	□	■	□	□	□	□	□
MATCHER	□	■	□	□	□	□	□
Multi-diverging							
STEMNET	□	□	■	■	□	□	□
MFA	□	■	■	■	□	□	□
FateID	□	□	■	■	□	□	□
Bi-diverging							
DPT	□	□	■	□	□	□	□
Wishbone	□	■	■	□	□	□	□
Graph							
RaceID	■	■	■	■	■	■	■
PAGA	■	■	■	■	■	■	■
Cyclic							
EIPiGraph	■	□	□	□	□	□	□
Angle	■	□	□	□	□	□	□

R : Ring structure, **L** : Linear structure,
B : Bi-diverging structure, **M** : Multi-diverging structure,
T : Tree structure, **C** : Connected structure, **U** : Unconnected structure
■ : This method is able to infer this kind of trajectory structure.
□ : This method is not able to infer this kind of trajectory structure.

3 Future Directions

3.1 Single-cell Omics

Today, scientists have made great achievements in the study of single-cell omics, but many problems still exist. Solving these problems will require the joint efforts of relevant researchers around the world. There are many aspects to these issues, and we'll focus on the data quality of single-cell systematics, our processing strategies for these data with high rate of missing data, and the development of detection methods for obtaining multiple-omics data from a single cell.

3.1.1 Data Quality and Sequencing Depth

At present, due to the extremely low nucleic acid content in a single cell, which is hard to be detected by instruments and easy to dissolve in reaction buffer in the washing processes, there is often a very high percentage of data loss in the single-cell omics data we can obtain, which leads to the fact that the true single-cell study does not actually exist. The conventional way in which the scientific community deals with this problem is to generate single-cell data by cluster analysis to generate multiple cell clusters, and then cluster the cells as the basic unit for studying single-cell omics. It is foreseeable that in the future, scientists will be able to obtain omics data with lower missing-data concentration or even "perfect" single-cell omics data. At that time, a single cell would become the real basic unit in single-cell research, marking a new phase in this single-cell biology. On the other hand, subject to the properties of various enzymes used in single-cell omics detection techniques, current single cells are not deep sequencing in the traditional sense. If researchers can find better biochemical enzymes with better performance, they will be able to obtain single-sequence

techniques with higher sequencing depth, which will make the relevant research more simple and credible.

3.1.2 Data Preprocessing Technology

Single-cell systematics data are often ultra-high dimensions and containing high concentration of missing data, which is not easy to analyze. So the dimensionality reduction of these data in the data preprocessing link has become an essential part of the study of single cell biology. At the same time, multiple sources of noise are widely found in today's single-cell omics data. Also, there will be data drift between the data generated in different sequencing batches, this phenomenon is known as batch error. The more effective methods are expected to be developed to do the dimensionality reduction and noise-signal cancellation of single-cell omics data, which can help researchers extract the corresponding characteristic form data.

3.1.3 Multi-omics Analysis in One Cell

Nowadays, single-cell omics detection techniques can only obtain a single type of omics data from one cell. Although some published methods can obtain multi-omics data at the single cell level, they are subject to various factors. These methods have not been adopted by the industry to produce stable kits. Multi-omics data based on the same batch of cells can show us the same biological process in multiple aspects during the multi-omics comprehensive analysis process, which can greatly enhance our overall understanding of life activities at the cellular level. We have reason to believe that in the future, various omics technologies will be integrated, and researchers will be able to apply multiple omics techniques in a single cell.

3.2 Core Algorithms of scTI Methods

There is no end to the optimization of the algorithm, let alone the scTI research is still in its infancy and will flourish. Not only do scientists need algorithms that can produce better results, but as the throughput of single-cell sequencing continues to increase, they also need algorithms that can handle larger amounts of data. In the current situation, the scTI method's ability to analyze different topologies is unsatisfactory. To solve this problem, researchers in this field may need to cooperate with researchers in the field of graph theory.

3.2.1 Advanced Algorithms

Although researchers have developed so many tools, there is still no single tool that can achieve better analysis results than other tools in all situations. In this case, computational biology or bioinformatics researchers may need the help of scientists from the computer science field to develop better-performing scTI tools with more advanced algorithms. At the same time, researchers can try to introduce artificial intelligence algorithms to help them solve data analysis problems that are difficult to mathematically model. In addition, more advanced algorithms can be used to identify cell subtypes and cell status. The suitability of the scTI tool for different topologies is essentially a graph theory problem. Various topologies are possible within real cell lineage of living organisms. Mathematical or computer scientists who study graph theory may provide academic and technical support and advice to solve this problem.

3.2.2 Mathematical Models

At the current stage, most of the scTI methods are actually based on scRNA-seq data, and the relationship between development and differentiation between cells is judged by analyzing the similarity in gene expression between cells. This actually only uses a very simple mathematical model, its prediction effect, versatility and stability are relatively weak. We can foresee that in future research, scientists will use mathematical models that are more efficient and have higher interpretability. In addition, with the development of single-cell

multi-omics technology, scientists should establish mathematical models based on single-cell multi-omics data, which will greatly enhance our understanding of cell lineage, cell development and cell differentiation.

3.3 Biological Analysis Related to csTI Methods

In the future, better csTI methods will be used to analyze complex cell lineage, which often contain extremely complex topologies and more types of cell subtypes and cell states. Research on complex cell differentiation systems can help the scientific community solve some of the long-standing problems in biology. This technology can also be used to analyze the source of cancer cells in patients and related cellular carcinogenesis, to find mutation sites that lead to cancer, and provide basic information for precision medicine. High-sensitivity detection methods can also be used to analyze trace amounts of diseased cells in patient tissue samples, which could be applied to detecting cancers or other serious diseases at their early stages.

However, limited by the total number of cells in a single batch of single-cell sequencing, there are some rare cell subtypes that may be difficult to detect, which affects the inferred cell lineage integrity and sensitivity. This problem can be solved in two ways. The first method is to use data cleaning methods to eliminate the bias between different batches of single cell sequencing data, and the second method is to increase the throughput of single batch single cell sequencing. As methods continue to improve and sequencing costs decrease, scientists will have the opportunity to build complete species cell lineage.

3.4 Establishment of the complete cell lineage of species

Researchers' goals should not stay at relatively simple cell lineage, and their ultimate goal should be to obtain the complete cellular developmental pathways from the processes that fertilized eggs grow and develop to adult individuals. In those process, scientists should

analyze the characteristics of the single-cell omics data of each cell and classify and define the cells. If scientists succeed in building "The Human Cell Atlas" [39], they can also use the technology and experience accumulated in the process to establish the cell lineage of various model animals. If such a great cause is really realized, biological research will enter a new phase based on single-cell omics.

Bibliography

[1] Maximilian Diehn et al. "Association of reactive oxygen species levels and radioresistance in cancer stem cells". In: *Nature* 458.7239 (Apr. 9, 2009), pp. 780–783. ISSN: 1476-4687. DOI: 10.1038/nature07733.

[2] Piero Dalerba et al. "Single-cell dissection of transcriptional heterogeneity in human colon tumors". In: *Nat. Biotechnol.* 29.12 (Nov. 13, 2011), pp. 1120–1127. ISSN: 1546-1696. DOI: 10.1038/nbt.2038.

[3] James E. Ferrell. "Bistability, bifurcations, and Waddington's epigenetic landscape". In: *Curr. Biol.* 22.11 (June 5, 2012), R458–466. ISSN: 1879-0445. DOI: 10.1016/j.cub.2012.03.045.

[4] Max Jan et al. "Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia". In: *Sci Transl Med* 4.149 (Aug. 29, 2012), 149ra118. ISSN: 1946-6242. DOI: 10.1126/scitranslmed.3004315.

[5] Sean C. Bendall et al. "Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination in Human B Cell Development". In: *Cell* 157.3 (Apr. 2014), pp. 714–725. ISSN: 00928674. DOI: 10.1016/j.cell.2014.04.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0092867414004711 (visited on 05/22/2019).

[6] M. Ryan Corces-Zimmerman et al. "Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission". In: *Proc. Natl. Acad. Sci. U.S.A.* 111.7 (Feb. 18, 2014), pp. 2548–2553. ISSN: 1091-6490. DOI: 10.1073/pnas.1324297111.

[7] Martin Etzrodt, Max Endele, and Timm Schroeder. "Quantitative Single-Cell Approaches to Stem Cell Research". In: *Cell Stem Cell* 15.5 (Nov. 2014), pp. 546–558. ISSN: 19345909. DOI: 10.1016/j.stem.2014.10.015. URL: https://linkinghub.elsevier.com/retrieve/pii/S1934590914004664 (visited on 05/22/2019).

[8] Charles Gawad, Winston Koh, and Stephen R. Quake. "Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics". In: *Proc. Natl. Acad. Sci. U.S.A.* 111.50 (Dec. 16, 2014), pp. 17947–17952. ISSN: 1091-6490. DOI: 10.1073/pnas.1420822111.

[9] V. J. Henry et al. "OMICtools: an informative directory for multi-omic data analysis". In: *Database* 2014.0 (July 14, 2014), bau069–bau069. ISSN: 1758-0463. DOI: 10.1093/database/bau069. URL: https://academic.oup.com/database/article-lookup/doi/10.1093/database/bau069 (visited on 05/24/2019).

[10] Eugenio Marco et al. "Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape". In: *Proc Natl Acad Sci USA* 111.52 (Dec. 30, 2014), E5643–E5650. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1408993111. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1408993111 (visited on 05/22/2019).

[11] Anoop P. Patel et al. "Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma". In: *Science* 344.6190 (June 20, 2014), pp. 1396–1401. ISSN: 1095-9203. DOI: 10.1126/science.1254257.

[12] Cole Trapnell et al. "The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells". In: *Nat Biotechnol* 32.4 (Apr. 2014), pp. 381–386. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/nbt.2859. URL: http://www.nature.com/articles/nbt.2859 (visited on 05/22/2019).

[13] Yong Wang et al. "Clonal evolution in breast cancer revealed by single nucleus genome sequencing". In: *Nature* 512.7513 (Aug. 14, 2014), pp. 155–160. ISSN: 1476-4687. DOI: 10.1038/nature13600.

[14] Kieran Campbell, Chris P Ponting, and Caleb Webber. *Laplacian eigenmaps and principal curves for high resolution pseudotemporal ordering of single-cell RNA-seq profiles*. preprint. Bioinformatics, Sept. 18, 2015. DOI: 10.1101/027219. URL: http://biorxiv.org/lookup/doi/10.1101/027219 (visited on 05/22/2019).

[15] Laleh Haghverdi, Florian Buettner, and Fabian J. Theis. "Diffusion maps for high-dimensional single-cell analysis of differentiation data". In: *Bioinformatics* 31.18 (Sept. 15, 2015), pp. 2989–2998. ISSN: 1367-4803, 1460-2059. DOI: 10.1093/bioinformatics/btv325. URL: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btv325 (visited on 05/23/2019).

[16] Jacob H. Levine et al. "Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis". In: *Cell* 162.1 (July 2, 2015), pp. 184–197. ISSN: 1097-4172. DOI: 10.1016/j.cell.2015.05.047.

[17] Franziska Paul et al. "Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors". In: *Cell* 163.7 (Dec. 17, 2015), pp. 1663–1677. ISSN: 1097-4172. DOI: 10.1016/j.cell.2015.11.013.

[18] Andreas Schlitzer et al. "Identification of cDC1- and cDC2-committed DC progenitors reveals early lineage priming at the common DC progenitor stage in the bone marrow". In: *Nat Immunol* 16.7 (July 2015), pp. 718–728. ISSN: 1529-2908, 1529-2916. DOI: 10.1038/ni.3200. URL: http://www.nature.com/articles/ni.3200 (visited on 05/22/2019).

[19] Jaehoon Shin et al. "Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis". In: *Cell Stem Cell* 17.3 (Sept. 2015), pp. 360–372. ISSN: 19345909. DOI: 10.1016/j.stem.2015.07.013. URL: https://linkinghub.elsevier.com/retrieve/pii/S1934590915003124 (visited on 05/23/2019).

[20] Cole Trapnell. "Defining cell types and states with single-cell genomics". In: *Genome Res.* 25.10 (Oct. 2015), pp. 1491–1498. ISSN: 1088-9051, 1549-5469. DOI: 10.1101/gr.190595.115. URL: http://genome.cshlp.org/lookup/doi/10.1101/gr.190595.115 (visited on 05/22/2019).

[21] Robrecht Cannoodt, Wouter Saelens, and Yvan Saeys. "Computational methods for trajectory inference from single-cell transcriptomics". In: *Eur. J. Immunol.* 46.11 (Nov. 2016), pp. 2496–2506. ISSN: 00142980. DOI: 10.1002/eji.201646347. URL: http://doi.wiley.com/10.1002/eji.201646347 (visited on 05/22/2019).

[22] Robrecht Cannoodt et al. *SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development*. preprint. Bioinformatics, Oct. 6, 2016. DOI: 10.1101/079509. URL: http://biorxiv.org/lookup/doi/10.1101/079509 (visited on 05/22/2019).

[23] Jinmiao Chen et al. "Mpath maps multi-branching single-cell trajectories revealing progenitor cell progression during development". In: *Nat Commun* 7.1 (Sept. 2016), p. 11988. ISSN: 2041-1723. DOI: 10.1038/ncomms11988. URL: http://www.nature.com/articles/ncomms11988 (visited on 05/23/2019).

[24] Dominic Grün et al. "De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data". In: *Cell Stem Cell* 19.2 (Aug. 2016), pp. 266–277. ISSN: 19345909. DOI: 10.1016/j.stem.2016.05.010. URL: https://linkinghub.elsevier.com/retrieve/pii/S1934590916300947 (visited on 05/22/2019).

[25] Naomi Habib et al. "Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons". In: *Science* 353.6302 (Aug. 26, 2016), pp. 925–928. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aad7038. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.aad7038 (visited on 05/23/2019).

[26] Laleh Haghverdi et al. "Diffusion pseudotime robustly reconstructs lineage branching". In: *Nat Methods* 13.10 (Oct. 2016), pp. 845–848. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.3971. URL: http://www.nature.com/articles/nmeth.3971 (visited on 05/22/2019).

[27] Zhicheng Ji and Hongkai Ji. "TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis". In: *Nucleic Acids Res* 44.13 (July 27, 2016), e117–e117. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkw430. URL: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw430 (visited on 05/22/2019).

[28] S. Nestorowa et al. "A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation". In: *Blood* 128.8 (Aug. 25, 2016), e20–e31. ISSN: 0006-4971, 1528-0020. DOI: 10.1182/blood-2016-05-716480. URL: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2016-05-716480 (visited on 05/23/2019).

[29] Andre Olsson et al. "Single-cell analysis of mixed-lineage states leading to a binary cell fate choice". In: *Nature* 537.7622 (Sept. 2016), pp. 698–702. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature19348. URL: http://www.nature.com/articles/nature19348 (visited on 05/23/2019).

[30] Itay Tirosh et al. "Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq". In: *Science* 352.6282 (Apr. 8, 2016), pp. 189–196. ISSN: 1095-9203. DOI: 10.1126/science.aad0501.

[31] Itay Tirosh et al. "Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma". In: *Nature* 539.7628 (2016), pp. 309–313. ISSN: 1476-4687. DOI: 10.1038/nature20123.

[32] Sara Aibar et al. "SCENIC: single-cell regulatory network inference and clustering". In: *Nat Methods* 14.11 (Nov. 2017), pp. 1083–1086. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.4463. URL: http://www.nature.com/articles/nmeth.4463 (visited on 05/22/2019).

[33] Kieran R Campbell and Christopher Yau. "Probabilistic modeling of bifurcations in single-cell gene expression data using a Bayesian mixture of factor analyzers". In: *Wellcome Open Res* 2 (Mar. 15, 2017), p. 19. ISSN: 2398-502X. DOI: 10.12688/wellcomeopenres.11087.1. URL: https://wellcomeopenresearch.org/articles/2-19/v1 (visited on 05/22/2019).

[34] Vladimir Yu Kiselev et al. "SC3: consensus clustering of single-cell RNA-seq data". In: *Nat. Methods* 14.5 (May 2017), pp. 483–486. ISSN: 1548-7105. DOI: 10.1038/nmeth.4236.

[35] Zehua Liu et al. "Reconstructing cell cycle pseudo time-series via single-cell transcriptome data". In: *Nat Commun* 8.1 (Dec. 2017), p. 22. ISSN: 2041-1723. DOI: 10.1038/s41467-017-00039-z. URL: http://www.nature.com/articles/s41467-017-00039-z (visited on 05/22/2019).

[36] Anna Nowogrodzki. "How to build a human cell atlas". In: *Nature* 547.7661 (July 2017), pp. 24–26. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/547024a. URL: http://www.nature.com/articles/547024a (visited on 05/22/2019).

[37] Jeffrey M. Perkel. "Single-cell sequencing made simple". In: *Nature* 547.7661 (July 2017), pp. 125–126. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/547125a. URL: http://www.nature.com/articles/547125a (visited on 05/22/2019).

[38] Xiaojie Qiu et al. "Reversed graph embedding resolves complex single-cell trajectories". In: *Nat Methods* 14.10 (Oct. 2017), pp. 979–982. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.4402. URL: http://www.nature.com/articles/nmeth.4402 (visited on 05/22/2019).

[39] Aviv Regev et al. "The Human Cell Atlas". In: *eLife* 6 (Dec. 5, 2017), e27041. ISSN: 2050-084X. DOI: 10.7554/eLife.27041. URL: https://elifesciences.org/articles/27041 (visited on 05/22/2019).

[40] Peter See et al. "Mapping the human DC lineage through the integration of high-dimensional techniques". In: *Science* 356.6342 (June 9, 2017), eaag3009. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aag3009. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.aag3009 (visited on 05/22/2019).

[41] "Single-cell biology". In: *Nature* 547.7661 (July 2017), pp. 19–19. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/547019a. URL: http://www.nature.com/articles/547019a (visited on 05/22/2019).

[42] Amos Tanay and Aviv Regev. "Scaling single-cell genomics from phenomenology to mechanism". In: *Nature* 541.7637 (Jan. 2017), pp. 331–338. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature21350. URL: http://www.nature.com/articles/nature21350 (visited on 05/22/2019).

[43] Lars Velten et al. "Human haematopoietic stem cell lineage commitment is a continuous process". In: *Nat Cell Biol* 19.4 (Apr. 2017), pp. 271–281. ISSN: 1465-7392, 1476-4679. DOI: 10.1038/ncb3493. URL: http://www.nature.com/articles/ncb3493 (visited on 05/22/2019).

[44] Joshua D. Welch, Alexander J. Hartemink, and Jan F. Prins. "MATCHER: manifold alignment reveals correspondence between single cell transcriptome and epigenome dynamics". In: *Genome Biol* 18.1 (Dec. 2017), p. 138. ISSN: 1474-760X. DOI: 10.1186/s13059-017-1269-0. URL: http://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1269-0 (visited on 05/22/2019).

[45] Sean Davis et al. *Seandavi/Awesome-Single-Cell: 2018-06-20-1*. Version 2018-06-20-1. June 20, 2018. DOI: 10.5281/zenodo.1117762. URL: https://zenodo.org/record/1117762 (visited on 05/24/2019).

[46] Josip S Herman, Sagar, and Dominic Grün. "FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data". In: *Nat Methods* 15.5 (May 2018), pp. 379–386. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.4662. URL: http://www.nature.com/articles/nmeth.4662 (visited on 05/22/2019).

[47] Charles A. Herring et al. "Unsupervised Trajectory Analysis of Single-Cell RNA-Seq and Imaging Data Reveals Alternative Tuft Cell Origins in the Gut". In: *Cell Systems* 6.1 (Jan. 2018), 37–51.e9. ISSN: 24054712. DOI: 10.1016/j.cels.2017.10.012. URL: https://linkinghub.elsevier.com/retrieve/pii/S2405471217304490 (visited on 05/22/2019).

[48] Kevin R. Moon et al. "Manifold learning-based methods for analyzing single-cell RNA-sequencing data". In: *Current Opinion in Systems Biology* 7 (Feb. 2018), pp. 36–46. ISSN: 24523100. DOI: 10.1016/j.coisb.2017.12.008. URL: https://linkinghub.elsevier.com/retrieve/pii/S2452310017301877 (visited on 05/22/2019).

[49] Kelly Street et al. "Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics". In: *BMC Genomics* 19.1 (Dec. 2018), p. 477. ISSN: 1471-2164. DOI: 10.1186/s12864-018-4772-0. URL: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-4772-0 (visited on 05/22/2019).

[50] Luke Zappia, Belinda Phipson, and Alicia Oshlack. "Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database". In: *PLoS Comput Biol* 14.6 (June 25, 2018). Ed. by Dina Schneidman, e1006245. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006245. URL: https://dx.plos.org/10.1371/journal.pcbi.1006245 (visited on 05/24/2019).

[51] Wouter Saelens et al. "A comparison of single-cell trajectory inference methods". In: *Nat Biotechnol* 37.5 (May 2019), pp. 547–554. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/s41587-019-0071-9. URL: http://www.nature.com/articles/s41587-019-0071-9 (visited on 05/23/2019).

[52] F. Alexander Wolf et al. "PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells". In: *Genome Biol* 20.1 (Dec. 2019), p. 59. ISSN: 1474-760X. DOI: 10.1186/s13059-019-1663-x. URL: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1663-x (visited on 05/22/2019).