 DOI: 10.1002/eji.201646347 Eur. J. Immunol. 2016. 46: 2496–2506

European Journal of Immunology

HIGHLIGHTS

REVIEW

Computational methods for trajectory inference from single-cell transcriptomics

Robrecht Cannoodt[*1,2,3,4], *Wouter Saelens*[*1,2] *and Yvan Saeys*[1,2]

[1] Data Mining and Modelling for Biomedicine group, VIB Inflammation Research Center, Ghent, Belgium
[2] Department of Internal Medicine, Ghent University, Ghent, Belgium
[3] Center for Medical Genetics, Ghent University, Ghent, Belgium
[4] Cancer Research Institute Ghent (CRIG), Ghent, Belgium

Recent developments in single-cell transcriptomics have opened new opportunities for studying dynamic processes in immunology in a high throughput and unbiased manner. Starting from a mixture of cells in different stages of a developmental process, unsupervised trajectory inference algorithms aim to automatically reconstruct the underlying developmental path that cells are following. In this review, we break down the strategies used by this novel class of methods, and organize their components into a common framework, highlighting several practical advantages and disadvantages of the individual methods. We also give an overview of new insights these methods have already provided regarding the wiring and gene regulation of cell differentiation. As the trajectory inference field is still in its infancy, we propose several future developments that will ultimately lead to a global and data-driven way of studying immune cell differentiation.

Keywords: Bioinformatics · Cell differentiation · Single-cell transcriptomics

Introduction

Recent technological advances in single-cell analysis are currently revolutionizing biological and medical sciences. In immunology, multicolor flow cytometry has been the major workhorse for high-throughput single-cell analysis [1, 2], and current innovations in the field, such as mass cytometry [3], allow characterizing up to 50 parameters per cell, providing high-dimensional descriptions at single-cell level. In parallel, single-cell transcriptomics technologies have greatly matured during the past decade [4, 5], now allowing the use of highly multiplexed assays measuring whole transcriptomes at the single-cell level, and moreover for thousands of cells. Other types of "omics" approaches, such as genomics [6], metabolomics [7], chromatin accessibility, and methylome analysis [8, 9] at the single-cell level have also recently been pioneered.

The analysis of all these cellular aspects at the single-cell level offers a number of advantages compared with the traditional bulk analysis of cells [10, 11]. While bulk analysis averages out the response of all cells in the sample, single-cell analysis can distinguish between a high response in a small subset of cells versus a medium response in all cells. Even if the majority of cells within a population respond to a particular stimulus, the type of response and its intensity can still be very heterogeneous, and this heterogeneity can be better described using single-cell analyses [12, 13]. Single-cell transcriptomics methods also allow the generation of a comprehensive overview of a whole cell population, enabling data-driven identification of new subpopulations within a heterogeneous population [12, 14, 15], as well as rare subsets [14, 16]. In immunology, single-cell based methods have also been shown to be very promising when studying

Correspondence: Dr. Yvan Saeys
e-mail: yvan.saeys@ugent.be

*These authors contributed equally to this work.

dynamic cellular processes, including cell differentiation and maturation [17–19], activation and responses to stimuli [12], and cyclic processes such as circadian rhythms and the cell cycle [13].

The study of cellular dynamic processes from single-cell transcriptomics data has opened up a whole new niche of computational techniques and modeling approaches that aim to automatically reconstruct a dynamic process from single-cell data. As an example, it is now possible to automatically reconstruct cell differentiation pathways from single-cell transcriptomics and proteomics data using computational techniques, ordering individual cells in the experiment according to their progression along the differentiation pathway. These computational methods, referred to as *trajectory inference* (TI) methods in this review, assign to every cell a so-called *pseudotime*, a numeric value in arbitrary units which measures how far a particular cell is within a dynamic process of interest. By ordering the cells according to this pseudotime, it becomes possible to define the different transition stages through which a cell progresses during its dynamic process. The initial dataset can be either a single snapshot of a mixture of cells in different stages (e.g., immune cells in the BM), or a set of samples collected at different timepoints (Fig. 1A). Starting from such a dataset describing high-dimensional, single-cell data, TI methods aim to order the cells with respect to an underlying dynamic process that explains the cell heterogeneity in the sample. The structure of the dynamic process can either be linear (Fig. 1B) or nonlinear, an example of the latter being a branching process (Fig. 1C), as often encountered in cell differentiation.

In this review, we provide an overview of the latest developments in the field of TI, highlighting the potential of these novel computational tools in immunology. While several methods for TI have been developed in recent years, we introduce here a unifying framework that organizes the main steps of these methods into a modular framework. We also discuss advantages and shortcomings of individual TI approaches regarding scalability, parameterization, and prior assumptions. Subsequently, we show how these methods can be used to gain new biological insights in cellular dynamic processes. We conclude by summarizing some of the future challenges and opportunities these methods might offer for addressing critical questions within immunology. Undoubtedly, advances in the field, both at the level of the technology of single-cell transcriptomics as well as at the algorithmic level will only expand the possibilities in this research field.

A modular framework to characterize TI methods

TI methods aim to automatically reconstruct a cellular dynamic process by structuring individual cells sampled and profiled from that process along a trajectory. This trajectory is subsequently used to identify the different stages in the dynamic process and their interrelationships. As such, TI methods go one step further than alternative techniques for clustering and visualizing single-cell data, such as SPADE [20], viSNE [21], and FlowSOM [22], as they explicitly order the cells along a continuous path. They do so without requiring a time series model, as time series frequently have issues with time resolution, cellular heterogeneity, and good *in vivo* synchronization conditions [23].

As an example, consider a cellular differentiation process where cells start in an immature state, then transition through a transitional state, and finally end up in a mature end state. When observing a mixed population containing cells at different stages of this differentiation process, a TI algorithm will attempt to find an ordering of all cells in the sample, ideally sorting the immature cells first, followed by the cells in the transitional stage, and finally the mature cells. By constructing such a gradient, novel intermediate states might also be discovered, potentially leading to novel cell subsets [18].

While this example depicts a simple, linear trajectory consisting of a sequence of single states, many cellular dynamic processes are nonlinear, including branching processes (e.g., differentiation of immune cells) or cyclic patterns (e.g., cell cycle, circadian rhythm processes, and activation/deactivation cycles). The latter types of processes in particular are challenging from a modeling point of view, as both the correct nonlinear structure needs to be inferred from the data, as well as the cells that need to be mapped to that structure.

TI methods typically consist of two main parts: a dimensionality reduction step and a trajectory modeling step (Fig. 2). In the dimensionality reduction step the high-dimensional cellular space (with dozens to thousands of genes or protein markers) is converted to a simplified representation using dimensionality reduction methods, clustering or graph-based methods. Subsequently, a trajectory modeling step then operates in this reduced space, aiming to identify cell states, constructing a trajectory through the different states, and projecting the cells back to the trajectory. Each of the two parts can be broken down further into modular and interchangeable components, and the choices each TI method makes directly determine its principal characteristics, such as accuracy, computational complexity (a measure for computation time), and robustness.

Dimensionality reduction step

While whole-transcriptome analyses avoid the bias of starting from a prespecified gene set, the dimensionality of such datasets is typically too high for most modeling algorithms to tackle directly. Moreover, in reality, the intrinsic dimensionality of biological systems is probably much lower. For example, a differentiating hematopoietic cell could be described by two or more dimensions: a dimension that indicates how far it has progressed in its differentiation toward a particular cell type and at least one other dimension indicating its current cell-cycle stage. Each dimension frequently corresponds to one or more modules of coexpressed genes. The reduced space can be interpreted in ways analogous to Waddington's epigenetic landscapes [24–26], where the

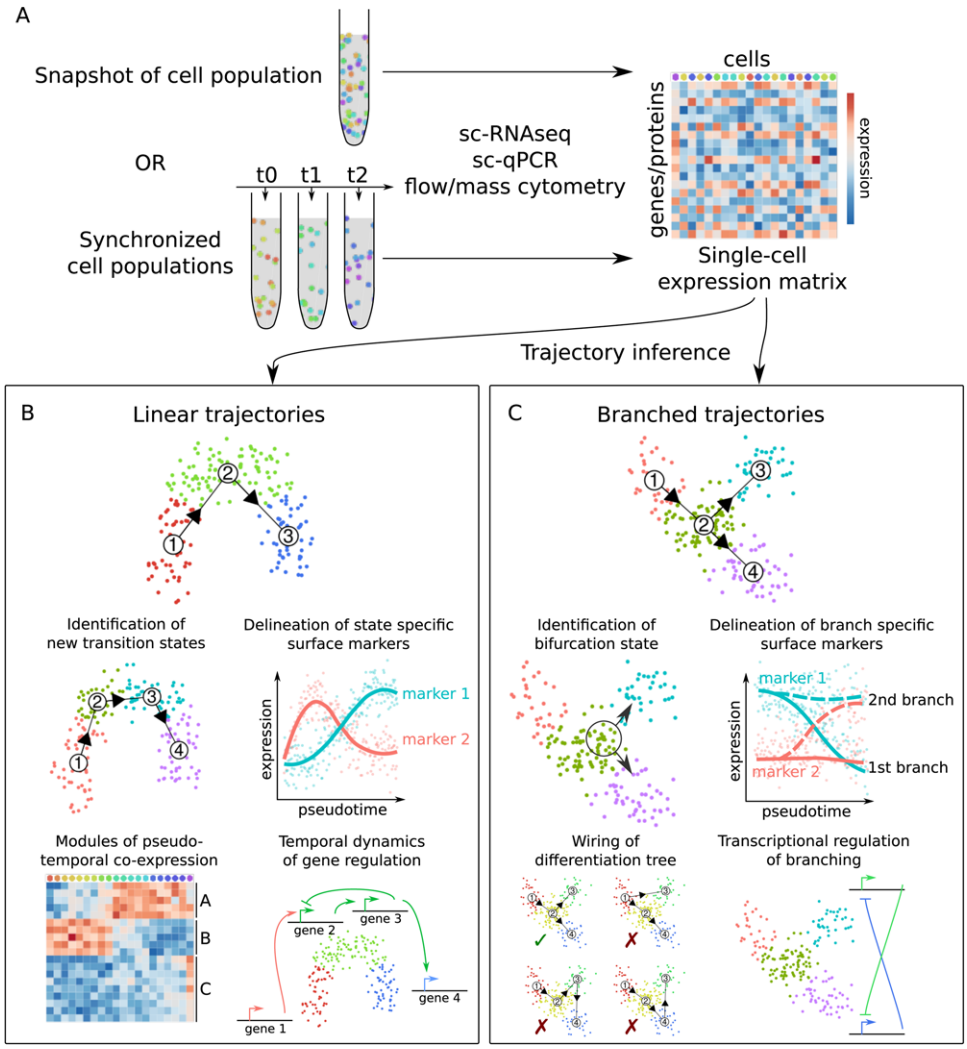

Figure 1. Applications of single-cell trajectory inference methods. (A) Single-cell data appropriate for TI can be both obtained from an unsynchronized population of single cells (snapshot data) but also from synchronized cell populations. Single-cell expression data can be obtained genomewide (sc-RNAseq), for a subset of genes (sc-qPCR) or for a set of proteins (flow/mass cytometry). (B) Linear trajectories can be used to find new transition states, try to find new markers to better separate transition states, identify coexpression modules, and investigate the dynamics of gene regulation underlying the dynamic process. (C) In addition to all applications of linear trajectories, branched trajectories can also be used to identify bifurcation states, find protein markers to better separate specific branches of the developmental tree, identify how cells move from one transition state to another in vivo, and understand how the commitment to a particular branch works at the level of the gene regulatory network.

landscape dictates the possible states a cell can reside in, and transitions between states correspond to dynamic cellular processes, such as cell differentiation. To this end, TI methods convert high-dimensional data to a more simplified representation by reducing the gene (feature) space, the cell (sample) space, or both, while maintaining the main characteristics of the data in the original space.

Dimensionality reduction techniques project the original, high-dimensional dataset onto a lower dimensional subspace. A wide variety of such techniques has been used in single-cell transcriptomics studies, the most popular of which are principal component analysis (PCA [27–29]), independent component analysis (ICA [30]), *t*-stochastic neighbor embedding (*t*-SNE [31–33]), and diffusion maps [34]. PCA performs a linear projection of the data such that the variance is preserved in the new space. While it is a relatively fast and easy to implement method, it makes certain assumptions, such as linearity, about the underlying structure of the data [35]. ICA, diffusion maps, and *t*-SNE are more general, with the latter two even being able to detect nonlinear relationships between cells [34, 36]. We refer the reader to other reviews for a thorough discussion of the advantages and disadvantages of different dimensionality reduction methods [37, 38].

Graph-based techniques reduce the feature space by representing the cells as nodes in a graph, where transcriptionally similar cells are linked together [39, 40]. By only retaining the most important edges in the graph, such techniques scale well to larger numbers of cells ($n > 10\,000$), and can be compared to a sparse distance matrix.

Clustering techniques can also be used to reduce the sample space, the feature space, or both [27, 41–43]. In this case, similar

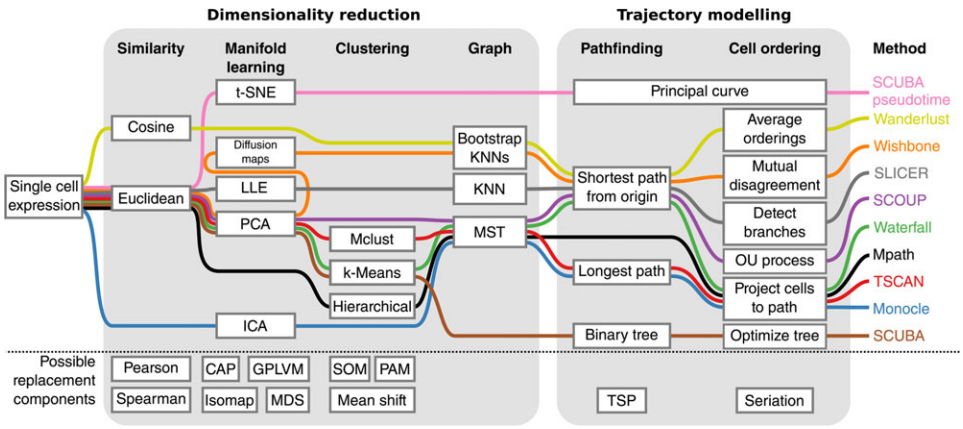

Figure 2. TI methods use several common building blocks and can be organized in a unifying modular framework. Overall, TI consists of two steps. In the first step dimensionality reduction techniques such as manifold learning, clustering, or graph-based methods are used to convert the dataset to a more simplified representation. This representation of the data then allows the trajectory itself to be more easily modelled in a second step. In this step, the trajectory itself is found within the data using both graph-based and curve-based approaches, after which the cells themselves can be ordered using a variety of methods.

sets of genes or cells are grouped together and subsequently simplified to one representative point per group. These representative point sets provide a good approximation of the different cells or genes in the dataset, speed up the downstream analysis and can even increase the interpretability of the results by reducing biological and technical noise.

Trajectory modeling step

A trajectory can be represented using several formalisms. Many TI methods use graph-based techniques, where a simplified graph representation is used as input to find a path through a series of nodes in the graph. These nodes can correspond to individual cells or groups of cells, and different path-finding algorithms are used by different algorithms. A number of methods rely on the prior definition of a "starting cell" by the user [27, 29, 39, 40, 44]. This cell is supposed to be representative for cells at the start of the underlying dynamic process (e.g., the most immature cell in the case of a cell developmental process), and is then used as a reference cell to compare all other cells against. Other methods [30, 42] look for the longest connected path in a sparsified graph, and then project all cells onto that path to order cells according to the underlying dynamic process.

A second way to represent a trajectory is by constructing a curve [31, 41], or a series of curves, such that the whole trajectory optimally explains the variation between the individual cells. An example of such an algorithm is principal curves, which extends PCA by looking for nonlinear principal directions in the data such that the curve is the average of all points that are projected onto it.

Characterization of TI methods

TI methods make use of a wide variety of components (Fig. 2), each with its own advantages and shortcomings (Fig. 3), an overview of which is given in the next sections.

Wanderlust [39] is one of the pioneering TI methods. While originally developed for cytometry data, it is also adopted for single-cell transcriptomics usage. Wanderlust first creates a graph consisting of edges between each cell and its neighbors by finding the k nearest neighbors (KNN) for each cell (Fig. 2). For cytometry data the Euclidean distance is used, whereas for transcriptomics data the cosine distance is used, as it is scale independent. Many subgraphs are bootstrapped, by sampling L out of K edges per node, after which the shortest path distance between each cell and a starting cell is calculated. Waypoints are used to align the orderings of the cells in the different bootstraps, and a consensus distance value is calculated for each of the cells. Because the shortest path algorithm is sensitive to noise in the expression data, Wanderlust uses an ensemble of trajectories to average out the noise and thus increase its robustness (Fig. 3). As additional input, Wanderlust requires specifying a starting cell, which is used as a prior to further improve the robustness of the method, although this introduces a bias.

Wishbone [40] is based on Wanderlust, but is able to detect a bifurcation point in the trajectory. It also requires the end user to specify a starting cell from which the algorithm will infer the trajectory. Wishbone uses a combination of PCA and diffusion maps to reduce the dimensionality of the dataset to a few components (Fig. 2). The distance between each cell is used to construct a KNN graph, which is repeatedly sampled by selecting L out of K edges. The shortest path distance from each cell to the starting cell is calculated, and waypoints are used to align the orderings of the cells in the different bootstraps. The disagreement between pairs of trajectories is investigated to identify a bifurcation event, and each cell is mapped to the bifurcating trajectory (Fig. 3). Wishbone uses several strategies to improve its robustness: (i) use of an ensemble of inferred trajectories, (ii) use of a starting cell, and (iii) use of the nonlinear dimensionality reduction technique diffusion maps [45]. In theory, the detection of bifurcations could be applied iteratively in order to produce a multiple branching trajectory instead of a bifurcating one, although this would increase the complexity of the problem at hand.

SLICER [44] is a KNN-based TI method that is able to infer branched trajectories while requiring little input from the end user. It uses an alpha-hull to estimate the k parameter for locally linear embedding dimensionality reduction. Locally linear embedding is used to compute the KNN graph, and to find the most "extreme" cells (Fig. 2). The user then has to specify which of these extreme cells is a starting cell, and from that cell the shortest

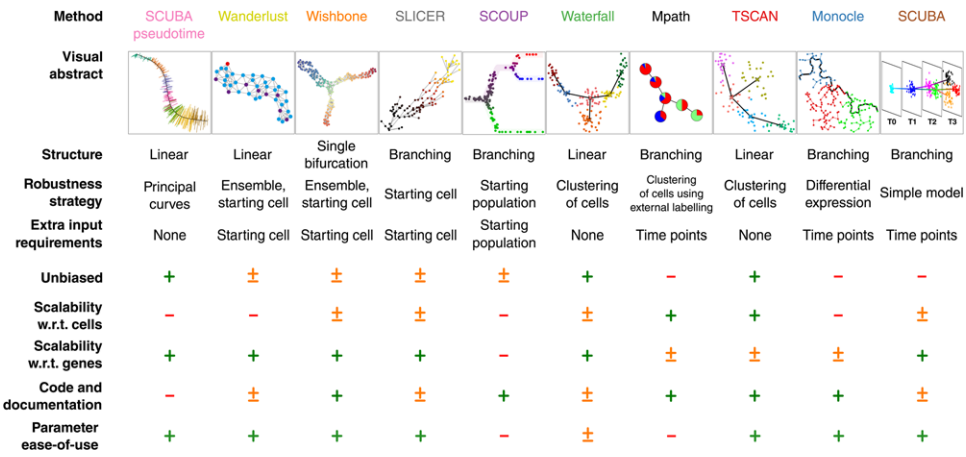

Figure 3. Each TI method has a unique set of strengths and weaknesses. The level of bias of each method indicates the amount of external information each method requires: ,+ none; ±, some; -, a lot. The scalability was determined by executing each method on an in silico dataset containing 100 cells (resp. genes) and an increasing number of genes (resp. cells) until the execution time exceeded 10 s: +, >10 000; ±, <10 000; -, <1000. Code and documentation denotes the quality of the code and documentation thereof provided. The parameter ease-of-use denotes how much is expected of the end user to tune the parameters.

path to all other cells is determined. The geodesic entropy is used to find branches; high values thereof indicate the existence of a branching point. SLICER helps the end user by automatically estimating its parameters and by determining good candidate starting cells. Allowing the end user to select a starting cell through visual inspection introduces a minimal bias effect in the results (Fig. 3). As for now, it is the only technique to offer fully branched TI with minimal bias in the outputted results, which can also scale well to thousands of genes and cells.

Monocle [30] is an alternative method to detect branched trajectories. Monocle uses ICA as its initial dimensionality reduction step (Fig. 2), but first performs a differential expression test between the different cell populations, as ICA does not scale well with an increasing number of genes. After the dimensionality has been reduced, a minimal spanning tree (MST) is calculated, after which the longest connected path(s) is determined within the graph. Each cell is then assigned to the nearest point in the inferred trajectory. Calculating an MST between individual cells is typically very sensitive to noise (Fig. 3). By reducing the number of genes using a differential expression test, the execution time of ICA is decreased and potentially improves the robustness of the MST. However, this introduces a heavy bias toward existing population groupings and away from possible heterogeneities within subpopulations. In a recent paper, Monocle was used to successfully recover the continuous development from endothelial cells to hematopoietic stem cells from scRNA-seq data [46].

Waterfall [27] is also an MST-based method, but calculates an MST between clusters of cells instead of individual cells. Waterfall uses PCA to reduce the dimensionality, and clusters the cells using *K*-means. An MST is calculated between the cluster centers, and the cluster with the lowest value in the first component is selected as the starting node. The longest path from the starting node is used as a trajectory, and each cell is perpendicularly projected to the path, creating the final ordering of the cells (Fig. 2). Waterfall is one of the few completely unsupervised methods (Fig. 3). The cluster centers are used to reduce the effect of noise while calculating the MST, although slight changes in the center locations might have a large impact on the MST and therefore the inferred trajectory.

Another MST-based method is **TSCAN** [42], which is conceptually very similar to Monocle and Waterfall. TSCAN also uses PCA to reduce the dimensionality, but uses Mclust, a clustering algorithm that models the datasets as a mixture of normal distributions, to cluster the cells while automatically determining the number of clusters using the Bayesian information criterion (Fig. 2). It calculates an MST through the cluster centers from which a trajectory is inferred by determining the longest connected path through the tree. Finally, all the cells are projected to the nearest point on the trajectory, creating the final ordering. TSCAN is also one of the few completely unsupervised methods, as it automatically determines both start and end cell clusters like Monocle does, while avoiding any prior filtering in the gene dimension. It has a high ease of use, as all of its parameters are automatically determined (Fig. 3). Like Monocle, it also finds the longest path in an MST, but does so more reliably since the cells were first grouped into clusters. However, similar to Waterfall, slight changes in center locations might have a large impact on the trajectory found.

SCUBA [41] infers branching trajectories from time-series experiments. For whole-transcriptome expression data, it first applies PCA. The cells in the initial cell state are clustered using *K*-means, with the number of clusters automatically determined using the gap statistic. For each consecutive time point, the cells are mapped to one of the clusters in the earlier time points. Each group of cells is then clustered into two clusters with *K*-means, and the gap statistic is used to determine whether a bifurcation occurs at this stage or not (Fig. 2). Finally, a penalized likelihood function is used to optimize the trajectory by reassigning cells to more fitting clusters. The model generated by SCUBA is unique in that it does not directly estimate a pseudotime for every cell but instead

infers the cellular states present in a time series experiment and reconstructs the hierarchical relationships between those states.

The authors of SCUBA acknowledged that this temporal information might be difficult to obtain experimentally, and provided an extension coined **SCUBA pseudotime** [41]. While this method was introduced as an extension of SCUBA, it has successfully been used as a separate TI method [31, 47]. The dimensionality is first reduced using *t*-SNE (Fig. 2). The principal curves algorithm is then used to infer a trajectory by starting from the first principal component as an initial path, and subsequently iteratively fitting the curve better to the data using a mean-shift procedure. The cells are orthogonally projected onto the principal curve to create the final ordering. The method does not require the specification of a start or an end cell, is completely unsupervised, and does not require any parameters to be specified (Fig. 3). The principal curves algorithm infers a trajectory and orders the cells at the same time, and is designed to be robust as it was originally designed to be a noise reduction algorithm [48].

SCOUP [29] uses expectation-maximization techniques to perform branched TI. As input the user is required to give an estimation of the normal distributions for each of the genes for cells that lie at the start of the dynamic process of interest. In practice this means that the user will designate a group of cells as the starting state, instead of a single cell as in other methods, and use these to estimate these distributions. It adds a dummy cell to the dataset with the mean of these distributions, uses PCA to reduce the dimensionality, calculates the MST, and uses the shortest path distance from the dummy cell to all other cells to provide an initial ordering of the cells (Fig. 2). It then uses an Ornstein–Uhlenbeck process to refine this ordering, in which a fitness criterion is defined for how well the cells are ordered with respect to the expression data, after which an expectation-maximization process is used to optimize this ordering. In addition, a mixture of Ornstein–Uhlenbeck processes can be used to infer a branching trajectory. While this approach is very promising, its current computational complexity does not allow it to scale up to higher numbers of cells or genes, and would therefore need more optimization (Fig. 3). The initial ordering is potentially also sensitive to noise, but could be easily replaced with a different approach. Finally, the requirement to designate a group of cells as the starting state could introduce a bias in the inferred trajectory.

Mpath [43] generates a branched trajectory by selecting landmarks and finding a minimum spanning tree. The landmarks are generated by hierarchically clustering the cells, iterating breadth-first over the resulting dendrogram, and checking whether a subtree passes empirically determined cutoffs with respect to its size and Shannon's index purity using the a priori cell population information (Fig. 2). A complete graph is created where each node represents a waypoint. Each cell is assigned to the edge between the two waypoints that are closest to the cell. The final trajectory is determined by calculating the minimum spanning tree, using the negative cell count per edge as weights. Cells can optionally be ordered by choosing a linear path on the trajectory and projecting the cells that are assigned to the landmarks on the path. Mpath makes no prior assumptions about the structure and the number of branches within the trajectory and does not require a prior filtering of the genes, although the requirement for initial knowledge about the cellular states could bias the inferred trajectory (Fig. 3).

Best practices for TI methods

At the methodological level, currently available TI methods frequently make assumptions about the structure of the trajectory and the parameterization of the method, or require external information such as known cellular states. While these assumptions are usually based on established biological knowledge, they still introduce a bias that could in some cases prevent the method from finding the correct but unexpected trajectory in the data. One can imagine for example that cells have the capacity to follow several alternative routes to a particular end state [49], but such branching and subsequent convergence is rarely covered by current methods. Moreover, a crucial assumption shared by every current TI method is that there is a trajectory present in the data, and as such these methods will always return a trajectory, even if it is not well supported by the data. In such cases alternative methods to cluster and/or visualize the data such as SPADE [20], viSNE [21], or FlowSOM [22] could be better suited. It is therefore crucial that the models generated by TI methods are seen as hypotheses, which require further experimental validation to establish the correct developmental relationships between cells.

Developers of new TI methods should strive to make their methods as unbiased as possible, by (i) avoiding the use of external information pertaining to the dataset (i.e., making the TI method truly data driven), (ii) avoiding any prior assumptions about the structure of the trajectory (i.e., generalized TI), (iii) automatically estimate parameters from the data itself, and (iv) estimate the confidence of how well the assumptions made by the method are supported by the data. Following these four rules will allow truly unexpected findings to be uncovered by the TI method, ultimately leading to a fully data-driven understanding of cellular development.

Each TI method uses a unique combination of common building blocks. While we classified the different TI methods with respect to their methodologies, this does not reflect the accuracy of these methods. Although each study proposing a new TI method provides a methodological validation thereof, these are often anecdotal [27, 30, 39–41, 44] or use a limited number of datasets [29, 42]. A large-scale comparative evaluation is currently lacking, and crucial to be able to characterize TI methods in terms of their overall performance, robustness to problem complexity or noise, parameter tuning complexity or best-performing scenario. Such an evaluation could for example be organized in the form of a community-wide challenge [50], which has proven very useful for a broad range of computational biology problems. The most challenging part of such an evaluation will probably be the correct definition of the gold standard and will possibly require a combination of datasets from well-established dynamic processes and simulated data, as has been proven

useful in past evaluation studies of network inference models [51].

Applications of single-cell TI

Several groundbreaking studies have already shown the potential of single-cell trajectory approaches to identify new transition states, unravel the wiring of cell differentiation, and infer the gene regulation events underlying the dynamic process. Here, we summarize some of these studies, describe their immunological relevance, and highlight future applications of single-cell trajectory approaches.

Identifying transition states using single-cell trajectories

During a dynamic process cells pass through several transitional states, characterized by different waves of transcriptional, morphological, epigenomic, and/or surface marker changes [49]. In the past, identification of cellular transition states has mostly relied on the identification of cell surface markers which can distinguish one state from the others. While this approach has undeniably proven useful, the discovery of new transition stages has remained challenging, as the identification of separating surface markers requires extensive trial-and-error and is not necessarily successful because transition stages are not always separable by phenotypes alone. Trajectory analysis of single-cell data therefore provides now a much more straightforward and unbiased approach to identifying and correctly ordering different transitional stages. New transition states correspond to dense clusters of cells within the developmental continuum, which can be identified using unsupervised clustering techniques [19]. Because whole-transcriptome data are available, subsequent analyses can be performed to identify surface markers to easily separate the transitional stages for further testing, but also to identify modules of coexpressed genes to further investigate the functional relevance of the transition state (Fig. 1B). Several recent success stories illustrate the power of TI approaches for identifying new transition states. In the study first proposing the Wanderlust method, a new population of human B-cell precursors was described based on the combined protein expression of CD34, CD38, CD24, and TdT [39]. Another study was able to divide mouse precursor dendritic cells in two consecutive stages, with each having a distinct gene expression signature [18].

TI unravels the wiring of dynamic processes

During development, cells progressively restrict their differentiation capabilities to specific subsets of cell types until they become committed to differentiate into a particular cell type. In hematopoiesis in particular, this progressive commitment has been modeled as a binary tree starting from the long-term hematopoietic stem cells that consecutively split into more and more restricted intermediate cellular states. Over the years, the hematopoietic tree has seen several revisions [52, 53] and the exact wiring of the differentiation tree is to this day still controversial [54]. Recently, high-throughput single-cell studies have been used to address this controversy, and have shown how the commitment to a particular differentiation branch happens much earlier than previously thought (Fig. 1C). In human hematopoiesis, for example, the common myeloid progenitor was shown to be highly heterogeneous with most cells already being committed to a particular myeloid branch [55]. So-called priming of progenitors to particular branches was already described earlier in murine hematopoiesis using both single-cell transcriptomics approaches [17, 56] and cellular barcoding [57]. Further downstream within hematopoietic differentiation it was demonstrated that certain pre-DCs are already primed to either the cDC1 or cDC2 lineage prior to exiting the BM [18]. In the lymphoid lineage single-cell trajectory approaches were used to identify the bifurcation stage between the lymphoid tissue-inducer cells and other innate lymphoid cells [19].

While these studies nicely demonstrate that the bifurcating step lies earlier than previously thought, they are still unable to pinpoint the reasons why certain cells choose one branch over another. Does the stochastic expression of one or more branch-determining transcription factors sway the cell to a particular lineage [58]? Or do spatial factors and cell–cell interactions influence the commitment, as has been shown for B-cell progenitors [59]? Furthermore, it is possible that there are earlier cellular events prior to transcriptional heterogeneity that could already be predictive for a preference to a particular branch. For example differential activity of the CARM1 histone methyltransferase has been shown to determine cell fate decision in 4-cell mouse embryos [60]. Combining single-cell gene expression data with single-cell protein [61], epigenome [62, 63] and spatial data [64, 65] could therefore provide a more complete view on the differentiation trajectories and indicate the earliest cellular events that sway a cell toward a particular branch.

Elucidating network dynamics along transition states and branch points

One of the central cellular processes underlying development is transcriptional regulation. During development, changes in transcription factor activity induce chromatin modifications, chromatin remodeling, and ultimately a differential recruitment of the basal transcriptional machinery [66]. Modeling the dynamics of gene regulation is therefore essential to better understand why a cellular dynamic processes progress through several steps, and what goes wrong in the case of disease.

The dynamics of gene regulation has classically been studied using time series data [23]. When dynamic processes progress asynchronously, such as in hematopoiesis, time series data are usually obtained by sorting different transition states and assessing

bulk gene expression and transcription factor binding within the population [67–70]. Alternatively, time series data can also be generated by synchronizing the dynamic process between cells. However, issues with time resolution, heterogeneity and good in vivo synchronization models can often limit the predictive power of the dynamic models of gene regulation that can be constructed [23]. Combining single-cell snapshot data with TI can now in principle construct an accurate time-series dataset for every cell population both in vivo and in vitro, while avoiding issues with heterogeneity. Initially, several pioneering studies demonstrated the potential of this approach in single cells using single-cell qPCR data for selected sets of 10 to 50 transcription factors [71–75]. With the advent of genome wide single-cell technologies, the models of dynamic gene regulation can now be made even more complete while avoiding a prior bias to well-studied transcription factors [17, 30, 76, 77] and several methods have already been developed specifically tuned toward single-cell data [78–80].

The power of a dynamic network approach in single cells is not only that it allows the identification of transcription factors specifically regulating a specific state, but also how these factors then induce sets of genes necessary for regulating the next wave of transcription changes (Fig. 1B). Such a gene regulatory cascade was described during cellular reprogramming from mouse embryonic fibroblasts to induced pluripotent stem cells, where SOX2 was shown to activate a series of other transcription factors which, together with SOX2, induces pluripotency [72]. Transition states are frequently associated with particular subnetworks of transcriptional regulators, as was shown in a direct reprogramming setting from fibroblasts to neurons [77]. Several studies have highlighted how some regulatory links can be very dynamic while others show evidence of being static during consecutive developmental stages [74, 75]. Regulatory network motifs can also be extracted from the regulatory network [74], some of which can have interesting properties in dynamic systems [81, 82]. In branching trajectories dynamic regulatory networks can also be used to study the regulators important for a particular branch [17, 75, 79] (Fig. 1C).

Another advantage of using ordered datasets, either based on bulk time-series or single-cell trajectories, is the improved accuracy of inferring true causative regulatory interactions. Assume that during development the expression of regulator A increases and in turn this regulator activates the expression of regulator B and gene C. Between the upregulation of regulator A and the upregulation of B and C we expect a time delay, equal to the time it takes for regulator A to be translated, translocated to the nucleus, and induce a detectable increase in the expression of B and C. On the other hand the expression of regulator B and gene C will be very strongly correlated, because they are both activated at the same time. Incorporating this time delay into the gene regulatory network inference could therefore be used to better discern the true causative regulatory edge between A and B/C with a spurious correlative edge between B and C [83].

Conclusions and future perspectives

TI is currently transforming the way dynamic cellular processes are studied. As technology matures, the cost to sequence large populations at the single-cell level will rapidly decrease [84], which will increase the power to identify rare transition states while avoiding overfitting on outlier cells. We expect single-cell omics technologies to be of extreme importance to better resolve both controversial and established developmental trajectories in the immune system and beyond. In this regard, single-cell omics approaches are very complementary with other emerging technologies, such as lineage tracing and time-lapse imaging [85, 86]. Ultimately, these approaches will be able to reconstruct the developmental hierarchies not only of individual cell lineages, but also of complete tissues and organisms.

We believe however that current studies have only explored the tip of the iceberg of what TI can offer. While most single-cell datasets currently only look at gene expression or a set of protein markers, several new technologies have recently been proposed to investigate other omics data types at the single-cell level. Integrating this data during TI could both improve the accuracy of the ordering of single cells, but also allow the study of the dynamics of transcriptional regulation in conjunction with gene expression changes. Another area where TI shows a lot of potential is for the relationship between cell development and disease. Many diseases are known to be caused by defects in cellular development pathways (e.g. defects in immune cell development can lead to disorders such as lymphocytopenia, hypogammaglobulinemia [87, 88], or myeloproliferative syndrome [89]), but the mechanisms by which a genetic defect leads to problems in cellular differentiation are not always well understood [90]. By comparing different single-cell trajectories from control and diseased patients, it will be possible to not only determine the exact point at which the cellular development becomes perturbed, but also how this perturbation affects the downstream developmental pathways. New methods will have to be developed to integrate different data types within the TI, and align these trajectories between samples to make them comparable.

Different dynamic processes can happen in parallel in the same cell. Probably the most obvious examples of such parallel dynamic processes are the cell cycle and cell differentiation [91]. One study proposed a method to extract and remove this cell-cycle effect, as it can influence the clear separation of distinct cell types [92]. This method works by decomposing the expression matrix into different components, each related to a particular biological process. However, decomposing the expression dataset in different components is not necessarily the correct approach, as these different dynamic processes could interact with each other. It is for example clear that cell differentiation and the cell cycle are strongly interlinked [91, 93]. In the ideal case, a TI method should include different parallel dynamic processes and their interactions within one common model, which will not be an easy feat. This will probably require either a very large number of cells or the inclusion of prior biological knowledge, such as known sets of genes involved

in a particular biological process as has already been used to discern heterogeneity at the single-cell level [92, 94].

In conclusion, several seminal studies have provided a glimpse of how TI can lead to a more unbiased and data-driven knowledge of biological dynamic processes. Methods for inferring linear trajectories have already reached some stage of maturity, with a broad set of different techniques offering a robust and unbiased TI. On the other hand, methods for inferring bifurcating and branching trajectories are currently a very active research topic. While certainly innovative, the many assumptions current methods make could be a hindrance to correctly determine the wiring of the differentiation tree. Furthermore, it still has to be seen how well all these methods will scale not only with the increasing number of cells, but also with increasingly complex models to include several parallel dynamic processes, their interactions, and different data types within one model of cell differentiation.

Acknowledgments: This work is supported by the Fund for Scientific Research FWO Flanders (R.C. and W.S.). Y.S. is an ISAC Marylou Ingram Scholar.

Conflict of interest: The authors declare no financial or commercial conflict of interest.

References

1 **Robinson, J. P. and Roederer, M.**, Flow cytometry strikes gold. *Science* 2015. **350**: 739–740.

2 **Mair, F., Florian, M., Hartmann, F. J., Dunja, M., Vinko, T., Carsten, K. and Burkhard, B.**, The end of gating? An introduction to automated analysis of high dimensional cytometry data. *Eur. J. Immunol.* 2015. **46**: 34–43.

3 **Bandura, D. R., Baranov, V. I., Ornatsky, O. I., Antonov, A., Kinach, R., Lou, X., Pavlov, S. et al.**, Mass cytometry: technique for real time single-cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal. Chem.* 2009. **81**: 6813–6822.

4 **Wang, Y. and Navin, N. E.**, Advances and applications of single-cell sequencing technologies. *Mol. Cell.* 2015. **58**: 598–609.

5 **Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. and Teichmann, S. A.**, The technology and biology of single-cell RNA sequencing. *Mol. Cell* 2015. **58**: 610–620.

6 **Gawad, C., Koh, W. and Quake, S. R.**, Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* 2016. **17**: 175–188.

7 **Zenobi, R.**, Single-cell metabolomics: analytical and biological perspectives. *Science* 2013. **342**: 1–10.

8 **Smallwood, S. A., Lee, H. J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S. R. et al.**, Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* 2014. **11**: 817–820.

9 **Farlik, M., Sheffield, N. C., Nuzzo, A., Datlinger, P., Schönegger, A., Klughammer, J. and Bock, C.**, Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep.* 2015. **10**: 1386–1397.

10 **Bacher, R. and Kendziorski, C.**, Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biol.* 2016. **17**: 1–10.

11 **Stegle, O., Teichmann, S. A. and Marioni, J. C.**, Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* 2015. **16**: 133–145.

12 **Shalek, A. K., Satija, R., Shuga, J., Trombetta, J. J., Gennert, D., Lu, D., Chen, P. et al.**, Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* 2014. **510**: 363–369.

13 **Trapnell, C.**, Defining cell types and states with single-cell genomics. *Genome Res.* 2015. **25**: 1491–1498.

14 **Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H. et al.**, Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* 2015. **525**: 251–255.

15 **Wen, L. and Tang, F.**, Single-cell sequencing in stem cell biology. *Genome Biol.* 2016. **17**: 1–12.

16 **Jiang, L., Chen, H., Pinello, L. and Yuan, G.-C.**, GiniClust: detecting rare cell types from single-cell gene expression data with Gini index. *Genome Biol.* 2016. **17**: 1–13.

17 **Paul, F., Arkin, Y., Giladi, A., Jaitin, D. A., Kenigsberg, E., Keren-Shaul, H., Winter, D. et al.**, Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell* 2016. **164**: 1663–1677.

18 **Schlitzer, A., Sivakamasundari, V., Chen, J., Sumatoh, H. R. B., Schreuder, J., Lum, J., Malleret, B. et al.**, Identification of cDC1- and cDC2-committed DC progenitors reveals early lineage priming at the common DC progenitor stage in the bone marrow. *Nat. Immunol.* 2015. **16**: 718–728.

19 **Ishizuka, I. E., Chea, S., Gudjonson, H., Constantinides, M. G., Dinner, A. R., Bendelac, A. and Golub, R.**, Single-cell analysis defines the divergence between the innate lymphoid cell lineage and lymphoid tissue-inducer cell lineage. *Nat. Immunol.* 2016. **17**: 269–276.

20 **Qiu, P., Simonds, E. F., Bendall, S. C., Gibbs, K. D. Jr., Bruggner, R. V., Linderman, M. D., Sachs, K. et al.**, Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat. Biotechnol.* 2011. **29**: 886–891.

21 **Amir, E.-A. D., Davis, K. L., Tadmor, M. D., Simonds, E. F., Levine, J. H., Bendall, S. C., Shenfeld, D. K. et al.**, viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.* 2013. **31**: 545–552.

22 **Van Gassen, S., Callebaut, B. and Van Helden, M. J.**, FlowSOM: using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry* 2015. **87**: 636–645.

23 **Bar-Joseph, Z., Gitter, A. and Simon, I.**, Studying and modelling dynamic biological processes using time-series gene expression data. *Nat. Rev. Genet.* 2012. **13**: 552–564.

24 **Waddington, C. H.**, *The strategy of the genes. A discussion of some aspects of theoretical biology*, George Allen & Unwin, Ltd., London, 1957.

25 **Ferrell, J. E. Jr**, Bistability, bifurcations, and Waddington's epigenetic landscape. *Curr. Biol.* 2012. **22**: R458–R466.

26 **Rebhahn, J. A., Deng, N., Sharma, G., Livingstone, A. M., Huang, S. and Mosmann, T. R.**, An animated landscape representation of CD4+ T-cell differentiation, variability, and plasticity: insights into the behavior of populations versus cells. *Eur. J. Immunol.* 2014. **44**: 2216–2229.

27 **Shin, J., Berg, D. A., Zhu, Y., Shin, J. Y., Song, J., Bonaguidi, M. A., Enikolopov, G. et al.**, Single-cell RNA-Seq with waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell* 2015. **17**: 360–372.

28 **Arsenio, J., Boyko, K., Metz, P. J., Kim, S. H., Yeo, G. W. and Chang, J. T.**, Early specification of CD8 T lymphocyte fates during adaptive immunity revealed by single-cell gene-expression analyses. *Nat. Immunol.* 2014. **15**: 365–372.

29 **Matsumoto, H. and Hisanori, K.**, SCOUP: a probabilistic model based on the Ornstein–Uhlenbeck process to analyze single-cell expression data during differentiation. *BMC Bioinformatics* 2016. **17**: 1–16.

30 **Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J. et al.**, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 2014. **32**: 381–386.

31 **Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S. P., Codeluppi, S., Plaza Reyes, A. et al.**, Single-Cell RNA-Seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* 2016. **165**: 1012–1026.

32 **Zeisel, A., Muñoz-Manchado, A. B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S. et al.**, Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 2015. **347**: 1138–1142.

33 **Becher, B., Andreas, S., Jinmiao, C., Florian, M., Sumatoh, H. R., Teng, K. W. W., Donovan, L. et al.**, High-dimensional analysis of the murine myeloid cell system. *Nat. Immunol.* 2014. **15**: 1181–1189.

34 **Haghverdi, L., Buettner, F. and Theis, F. J.**, Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* 2015. **31**: 2989–2998.

35 **Shlens, J.**, A tutorial on principal component analysis. 2014. Available at: http://arxiv.org/abs/1404.1100.

36 **van der Maaten, L. and Hinton, G.**, Visualizing data using t-SNE. *J. Mach. Learn. Res.* 2008. **9**: 2579–2605.

37 **van der Maaten, L. and Postma, E., van den Herik, J.**, Dimensionality reduction: a comparative review. *J. Mach. Learn. Res.* 2009. **10**: 66–71.

38 **Lin, B., He, X. and Ye, J.**, A geometric viewpoint of manifold learning. *Appl Inform.* 2015. **2**: 1–12.

39 **Bendall, S. C., Davis, K. L., Amir, E.-A. D., Tadmor, M. D., Simonds, E. F., Chen, T. J., Shenfeld, D. K. et al.**, Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* 2014. **157**: 714–725.

40 **Setty, M., Tadmor, M. D., Reich-Zeliger, S., Angel, O., Salame, T. M., Kathail, P., Choi, K. et al.**, Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat. Biotechnol.* 2016. **34**: 637–645.

41 **Marco, E., Karp, R. L., Guo, G., Robson, P., Hart, A. H., Trippa, L. and Yuan, G.-C.**, Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. *Proc. Natl. Acad. Sci. U. S. A.* 2014. **111**: E5643–E5650.

42 **Ji, Z. and Ji, H.**, TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res.* 2016. **44**: e117.

43 **Chen, J., Schlitzer, A., Chakarov, S., Ginhoux, F. and Poidinger, M.**, Mpath maps multi-branching single-cell trajectories revealing progenitor cell progression during development. *Nat. Commun.* 2016. **7**: 1–15.

44 **Welch, J. D., Hartemink, A. J. and Prins JF**, SLICER: inferring branched, nonlinear cellular trajectories from single cell RNA-seq data. *Genome Biol.* 2016. **17**: 1–15.

45 **Coifman, R. R., Lafon, S., Lee, A. B., Maggioni, M., Nadler, B., Warner, F. and Zucker, S. W.**, Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proc. Natl. Acad. Sci. U. S. A.* 2005. **102**: 7426–7431.

46 **Zhou, F., Fan, Z., Xianlong, L., Weili, W., Ping, Z., Jie, Z., Wenyan, H. et al.**, Tracing haematopoietic stem cell formation at single-cell resolution. *Nature* 2016. **533**: 487–492.

47 **Kim, T.-H., Saadatpour, A., Guo, G., Saxena, M., Cavazza, A., Desai, N., Jadhav, U. et al.**, Single-cell transcript profiles reveal multilineage priming in early progenitors derived from Lgr5(+) intestinal stem cells. *Cell Rep.* 2016. **16**: 2053–2060.

48 **Hastie, T. and Stuetzle, W.**, Principal curves. *J. Am. Stat. Assoc.* 1989: 502–516.

49 **Enver, T., Pera, M., Peterson, C. and Andrews, P. W.**, Stem cell states, fates, and the rules of attraction. *Cell Stem Cell* 2009. **4**: 387–397.

50 **Bender, E.**, Crowdsourced solutions. *Nature* 2016. **533**: S62–S64.

51 **Stolovitzky, G., Monroe, D. and Califano, A.**, Dialogue on reverse-engineering assessment and methods: the DREAM of high-throughput pathway inference. *Ann. N. Y. Acad. Sci.* 2007. **1115**: 1–22.

52 **Adolfsson, J., Månsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C. T., Bryder, D. et al.**, Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell* 2005. **121**: 295–306.

53 **Månsson, R., Hultquist, A., Luc, S., Yang, L., Anderson, K., Kharazi, S., Al-Hashmi, S. et al.**, Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. *Immunity* 2007. **26**: 407–419.

54 **Ceredig, R., Rolink, A. G. and Brown, G.**, Models of haematopoiesis: seeing the wood for the trees. *Nat. Rev. Immunol.* 2009. **9**: 293–300.

55 **Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O. I., Wilson, G., Kaufmann, K. B. et al.**, Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* 2016. **351**: aab2116.

56 **Guo, G., Luc, S., Marco, E., Lin, T.-W., Peng, C., Kerenyi, M. A., Beyaz, S. et al.**, Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell* 2013. **13**: 492–505.

57 **Perié, L., Duffy, K. R., Kok, L., de Boer R. J. and Schumacher, T. N.**, The branching point in erythro-myeloid differentiation. *Cell* 2015. **163**: 1655–1662.

58 **Enver, T., Heyworth, C. M. and Dexter, T. M.**, Do stem cells play dice? *Blood* 1998. **92**: 348–351. discussion 352.

59 **Ding, L. and Morrison, S. J.**, Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. *Nature* 2013. **495**: 231–235.

60 **Goolam, M., Scialdone, A., Graham, S. J. L., Macaulay, I. C., Jedrusik, A., Hupalowska, A., Voet, T. et al.**, Heterogeneity in Oct4 and Sox2 targets biases cell fate in 4-cell mouse embryos. *Cell* 2016. **165**: 61–74.

61 **Bendall, S. C., Simonds, E. F., Qiu, P., Amir, E.-A. D., Krutzik, P. O., Finck, R., Bruggner, R. V. et al.**, Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* 2011. **332**: 687–696.

62 **Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., Chang, H. Y. et al.**, Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 2015. **523**: 486–490.

63 **Rotem, A., Ram, O., Shoresh, N., Sperling, R. A., Goren, A., Weitz, D. A. and Bernstein, B. E.**, Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* 2015. **33**: 1165–1172.

64 **Giesen, C., Wang, H. A. O., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., Schüffler, P. J. et al.**, Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* 2014. **11**: 417–422.

65 **Angelo, M., Bendall, S. C., Finck, R., Hale, M. B., Hitzman, C., Borowsky, A. D., Levenson, R. M. et al.**, Multiplexed ion beam imaging of human breast tumors. *Nat. Med.* 2014. **20**: 436–442.

66 **Coulon, A., Chow, C. C., Singer, R. H. and Larson, D. R.**, Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat. Rev. Genet.* 2013. **14**: 572–584.

67 **Novershtern, N., Subramanian, A., Lawton, L. N., Mak, R. H., Haining, W. N., McConkey, M. E., Habib, N. et al.**, Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* 2011. **144**: 296–309.

68 **May, G., Soneji, S., Tipping, A. J., Teles, J., McGowan, S. J., Wu, M., Guo, Y. et al.**, Dynamic analysis of gene expression and genome-wide transcription factor binding during lineage specification of multipotent progenitors. *Cell Stem Cell* 2013. **13**: 754–768.

69 **Jojic, V., Shay, T., Sylvia, K., Zuk, O., Sun, X., Kang, J., Regev, A. et al.**, Identification of transcriptional regulators in the mouse immune system. *Nat. Immunol.* 2013. **14**: 633–643.

70 **Goode, D. K., Obier, N., Vijayabaskar, M. S., Lie-A-Ling, M., Lilly, A. J., Hannah, R., Lichtinger, M. et al.**, Dynamic gene regulatory networks drive hematopoietic specification and differentiation. *Dev. Cell* 2016. **36**: 572–587.

71 **Guo, G., Huss, M., Tong, G. Q., Wang, C., Li, S. L., Clarke, N. D. and Robson, P.**, Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* 2010. **18**: 675–685.

72 **Buganim, Y., Faddah, D. A., Cheng, A. W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S. L. et al.**, Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* 2012. **150**: 1209–1222.

73 **Kouno, T., de Hoon M., Mar, J. C., Tomaru, Y., Kawano, M., Carninci, P., Suzuki, H. et al.**, Temporal dynamics and transcriptional control using single-cell gene expression analysis. *Genome Biol.* 2013. **14**: R118.

74 **Moignard, V., Macaulay, I. C., Swiers, G., Buettner, F., Schütte, J., Calero-Nieto, F. J., Kinston, S. et al.**, Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat. Cell Biol.* 2013. **15**: 363–372.

75 **Pina, C., Teles, J., Fugazza, C., May, G., Wang, D., Guo, Y., Soneji, S. et al.**, Single-Cell Network Analysis Identifies DDIT3 as a Nodal Lineage Regulator in Hematopoiesis. *Cell Rep.* 2015. **11**: 1503–1510.

76 **Shalek, A. K., Satija, R., Adiconis, X., Gertner, R. S., Gaublomme, J. T., Raychowdhury, R., Schwartz, S. et al.**, Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 2013. **498**: 236–240.

77 **Treutlein, B., Lee, Q. Y., Camp, J. G., Mall, M., Koh, W., Shariati, S. A. M., Sim, S. et al.**, Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature* 2016. **534**: 391–395.

78 **Ocone, A., Haghverdi, L., Mueller, N. S. and Theis, F. J.**, Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data. *Bioinformatics* 2015. **31**: i89–i96.

79 **Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A. J., Tanaka, Y., Wilkinson, A. C., Buettner, F. et al.**, Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* 2015. **33**: 269–276.

80 **Chen, H., Guo, J., Mishra, S. K., Robson, P., Niranjan, M. and Zheng, J.**, Single-cell transcriptional analysis to uncover regulatory circuits driving cell fate decisions in early mouse development. *Bioinformatics* 2015. **31**: 1060–1066.

81 **Prill, R. J., Iglesias, P. A. and Levchenko, A.**, Dynamic properties of network motifs contribute to biological network organization. *PLoS Biol.* 2005. **3**: e343.

82 **Alon, U.**, Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 2007. **8**: 450–461.

83 **Sima, C., Hua, J. and Jung, S.**, Inference of gene regulatory networks using time-series data: a survey. *Curr. Genomics* 2009. **10**: 416–429.

84 **Linnarsson, S. and Teichmann, S. A.**, Single-cell genomics: coming of age. *Genome Biol.* 2016. **17**: 1–3.

85 **Etzrodt, M., Endele, M. and Schroeder, T.**, Quantitative single-cell approaches to stem cell research. *Cell Stem Cell* 2014. **15**: 546–558.

86 **Sun, J., Ramos, A., Chapman, B., Johnnidis, J. B., Le, L., Ho, Y.-J., Klein, A. et al.**, Clonal dynamics of native haematopoiesis. *Nature.* 2014. **514**: 322–327.

87 **Vale, A. M. and Schroeder, H. W. Jr**, Clinical consequences of defects in B-cell development. *J. Allergy Clin. Immunol.* 2010. **125**: 778–787.

88 **Cunningham-Rundles, C. and Ponda, P. P.**, Molecular defects in T- and B-cell primary immunodeficiency diseases. *Nat. Rev. Immunol.* 2005. **5**: 880–892.

89 **Levine, R. L. and Gilliland, D. G.**, Myeloproliferative disorders. *Blood* 2008. **112**: 2190–2198.

90 **Crotty, S.**, T follicular helper cell differentiation, function, and roles in disease. *Immunity* 2014. **41**: 529–542.

91 **Kueh, H. Y., Champhekar, A., Nutt, S. L., Elowitz, M. B. and Rothenberg, E. V.**, Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. *Science* 2013. **341**: 670–673.

92 **Buettner, F., Natarajan, K. N., Casale, F. P., Proserpio, V., Scialdone, A., Theis, F. J., Teichmann, S. A. et al.**, Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* 2015. **33**: 155–160.

93 **Li, V. C. and Kirschner, M. W.**, Molecular ties between the cell cycle and differentiation in embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 2014. **111**: 9503–9508.

94 **Fan, J., Salathia, N., Liu, R., Kaeser, G. E., Yung, Y. C., Herman, J. L., Kaper, F. et al.**, Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat. Methods.* 2016. **13**: 241–244.

Abbreviations: **ICA**: independent component analysis · **KNN**: *k* nearest neighbors · **MST**: minimal spanning tree · **PCA**: principal component analysis · **TI**: trajectory inference · ***t*-SNE**: *t*-stochastic neighbor embedding

Full correspondence: Dr. Yvan Saeys, Data Mining and Modelling for Biomedicine group, VIB Inflammation Research Center, Ghent University, Technologiepark 927, 9052 Ghent, Belgium
e-mail: yvan.saeys@ugent.be

Received: 2/7/2016
Revised: 30/8/2016
Accepted: 26/9/2016
Accepted article online: 28/9/2016

