cmgh
CELLULAR AND MOLECULAR GASTROENTEROLOGY AND HEPATOLOGY

REVIEW

Single-Cell Computational Strategies for Lineage Reconstruction in Tissue Systems

Charles A. Herring,[1,2] Bob Chen,[1,3] Eliot T. McKinley,[1,4] and Ken S. Lau[1,2,3]

[1]Epithelial Biology Center, [4]Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee; [2]Program in Chemical and Physical Biology, [3]Department of Cell and Developmental Biology, Vanderbilt University School of Medicine, Nashville, Tennessee

SUMMARY

Recent developments in single-cell technologies have stimulated growth in analysis techniques, in particular, computational tools for ordering cell states as a function of pseudotemporal progression. We provide a review of current algorithms and a generalized single-cell workflow tailored for trajectory analysis, with a focus on underlying assumptions and caveats.

Function at the organ level manifests itself from a heterogeneous collection of cell types. Cellular heterogeneity emerges from developmental processes by which multipotent progenitor cells make fate decisions and transition to specific cell types through intermediate cell states. Although genetic experimental strategies such as lineage tracing have provided insights into cell lineages, recent developments in single-cell technologies have greatly increased our ability to interrogate distinct cell types, as well as transitional cell states in tissue systems. From single-cell data that describe these intermediate cell states, computational tools have been developed to reconstruct cell-state transition trajectories that model cell developmental processes. These algorithms, although powerful, are still in their infancy, and attention must be paid to their strengths and weaknesses when they are used. Here, we review some of these tools, also referred to as *pseudotemporal ordering algorithms*, and their associated assumptions and caveats. We hope to provide a rational and generalizable workflow for single-cell trajectory analysis that is intuitive for experimental biologists. *(Cell Mol Gastroenterol Hepatol 2018;5:539–548; https://doi.org/10.1016/j.jcmgh.2018.01.023)*

Keywords: Trajectory; Pseudotime; Single-Cell Analysis; Differentiation; Cell State Transition; Stem Cells.

Cellular heterogeneity, defined by a diversity of co-occurring cell types in a tissue, is characteristic of practically every organ in the human body. The organs of the digestive system also comprise specialized cell populations that play important but diverse roles in absorption, secretion, and barrier function. For instance, distinct cell types of the pancreatic islet secrete different hormones, including insulin-secreting β cells, glucagon-secreting δ cells, and somatostatin-expressing δ cells.[1] Likewise, the small and large intestines exist in a dynamic equilibrium of heterogeneous stem, transitional, and differentiated cell populations, with the latter responsible for nutrient absorption, antimicrobial peptide secretion, and formation and maintenance of the mucus layer in the gut.[2] A fundamental question in developmental biology is the origin of cellular heterogeneity, which arises from a specification process initiated from multipotent cells. Recent developments in multiplex single-cell experimental tools have greatly facilitated the interrogation of individual cells; data on single cells then can be grouped into relevant cell populations. In digestive organ systems, populational analysis of single-cell data has been used for discovering previously unidentified β cell subpopulations in the pancreatic islet,[1] novel markers of intestinal tuft cells,[2,3] endocrine progenitor cell heterogeneity,[4] and signaling mechanisms between neighboring intestinal epithelial cells,[5] among others. Populational analysis using single-cell tools is a powerful approach for dissecting tissue-level heterogeneity, and has been reviewed extensively elsewhere.[6,7] Beyond defining cell populations, such as stem and differentiated cell types, single-cell experimental tools also can be used to characterize transitional intermediate cell states in various tissues and organoid systems.[8] Thus, it theoretically should be possible, using single-cell data, to trace terminal cell types through intermediate cell states back to their roots of differentiation in a series of progenitor–progeny relationships. Here, we review current computational tools by which a "virtual lineage trace," also known as a *pseudotemporal order*, can be extracted from multidimensional single-cell data.

Single-Cell Experimental Technologies to Interrogate Cell States From Tissues

The theoretical basis of pseudotemporal ordering is that asynchronous sampling from multiple time points over development[9] or snap-shot sampling at a single time point

Abbreviations used in this paper: MST, minimum spanning tree; PCA, principal component analysis; scRNA-seq, single-cell RNA-sequencing; t-SNE, t-distributed stochastic neighbor embedding.

Most current article

2352-345X
https://doi.org/10.1016/j.jcmgh.2018.01.023

of a continually renewing tissue (such as the intestine)[10] can result in a dense sampling of transitional states that can be aligned to reflect a time course of state transitions (Figure 1). A cell state is represented by the position of a cell in a data space defined by multiple molecular markers that describe the identity and behavior of cells (Figure 1). Ordering is conducted on the basis of similarity in cell states; dense sampling of these states is required to obtain a continuum of data by which the relationship between cell states can be inferred. Because transitional cell states often are rare compared with differentiated cells in tissue, it is required for single-cell technologies to be able to query a large volume of data points, as well as simultaneously measure multiple markers, to fully depict a continuum of cell states. Here, we briefly review common single-cell tools that can evaluate many cells in a multiplex fashion in the context of their classification into either suspension approaches or in situ approaches.

Suspension approaches involve cellular dissociation and then separate processing and analysis of individual cells, with the major caveat that the spatial context of the tissue is lost. Suspension approaches include protein-based techniques such as mass and multiparameter flow cytometry, and transcript-based techniques such as single-cell RNA-sequencing (scRNA-seq) and gene expression assays.[11] The advantage of these approaches is in their high-throughput capacity to produce data. Flow and mass cytometry can analyze hundreds of thousands of cells in a multiplex fashion (20–40 protein analytes per cell) on the order of minutes,[12] while scRNA-seq can quantify gene expression in an unbiased, genome-wide manner (thousands of gene analytes).[13] Multiple platforms of scRNA-seq exist, with variations in cell-containment strategies ranging from microwells[14–16] to liquid-oil emulsion droplets[17–19]; many of the current iterations can query up to thousands of cells.

A factor to consider when applying suspension approaches, especially on organs of the digestive system, is the perturbation imposed on cells when they are disaggregated from tissue. Cells of the hematopoietic system exist either as single-cell suspensions or in loosely connected tissues, which are readily amenable to single-cell analysis.[12] For intestinal cells, specifically for those in the lamina propria, protocols have been developed such that the correct numbers and types of cells can be retrieved for single-cell analysis, providing critical insights into biological and disease processes.[20] For epithelial tissues that are tightly connected, additional factors must be considered so as to not introduce technical artifacts during the single-cell dissociation process.[21] Disaggregation for Intracellular Signaling of Single Epithelial Cells from Tissue was developed as a fixation approach for preserving the intact state of epithelial cells for single-cell signaling analysis using mass and flow cytometry.[5] Disaggregation for Intracellular Signaling of Single Epithelial Cells from Tissue can be applied to formalin-fixed paraffin-embedded tissues, for instance, to observe signaling state alterations in human colorectal cancer specimens.[22] On the scRNA-seq side, Adam et al[23] adapted psychrophilic proteases for single-cell dissociation in the cold, which drastically reduces artifacts and maintains native cell states. Adaptation of a similar strategy to fixed[24,25] or frozen tissues[26] may enable scRNA-seq of preserved cell states. For cells that cannot be dissociated without compromising integrity, such as neurons with long and fragile axonal processes, single nucleus profiling of fresh and preserved tissues is a viable strategy to obtain a glimpse of cell state.[27–29] It should be noted, however, that transcriptomes obtained from the nucleus may be drastically different from those obtained from the entire cell. These efforts highlight recent developments into suspension approaches to enable high-throughput evaluation of native cell states for characterizing cellular heterogeneity and developmental events.

Unlike suspension approaches, in situ imaging techniques allow cells and their niche components to be analyzed in their native spatial context. Because of the lack of tissue dissociation, communication mechanisms between niche cells and epithelial cells can be directly visualized and quantified. Recent advances have improved the multiplex capabilities of microscopy approaches, enabling detection and quantification of dozens of markers leading to accurate identification of cell types that reside within certain niches. Current multiplex imaging technologies for proteins can be classified either as mass-based or iterative. Mass-based imaging approaches,

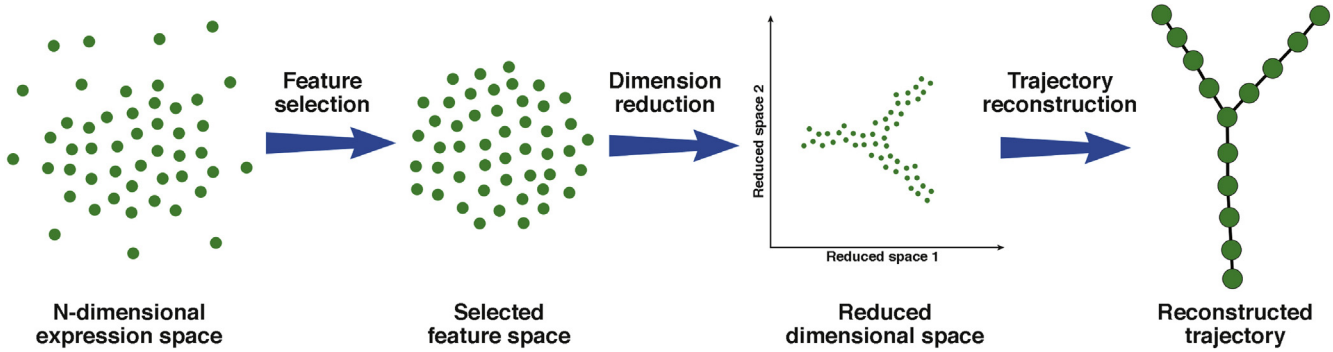

Figure 1. General workflow of trajectory analysis algorithms. Beginning with data in multidimensional space, feature selection is first performed to include relevant analytes and exclude noise. From the selected feature set, dimension reduction is applied to best emphasize the part of the data most relevant to cell-state transitions. Trajectories then are reconstructed in this reduced space and analyzed as pseudotime courses.