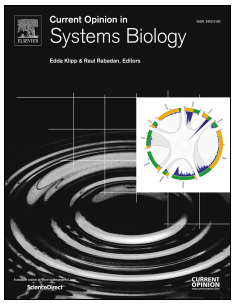
Accepted Manuscript



Revealing routes of cellular differentiation by single-cell RNA-seq

Dominic Grün

PII: S2452-3100(18)30013-1

DOI: 10.1016/j.coisb.2018.07.006

Reference: COISB 174

To appear in: *Current Opinion in Systems Biology*

Received Date: 22 March 2018

Accepted Date: 17 July 2018

Please cite this article as: Grün D, Revealing routes of cellular differentiation by single-cell RNA-seq, *Current Opinion in Systems Biology* (2018), doi: 10.1016/j.coisb.2018.07.006.

Revealing routes of cellular differentiation by single-cell RNA-seq

Dominic Grün

Abstract
Differentiation of multipotent stem cells is controlled by the intricate regulatory interactions of thousands of genes. It remains one of the major challenges to understand how nature has designed such robust and reproducible regulatory mechanisms. Knowing the detailed structure of the underlying lineage trees is the basis for investigating the molecular control of this process. The recent availability of large-scale sensitive single-cell RNA-seq protocols has enabled the generation of snapshot data covering the entire spectrum of cell states in a system of interest. Consequently, a large number of computational methods for the reconstruction of cellular differentiation trajectories have been developed. Here, I will provide a detailed overview of the concepts and ideas behind some of these algorithms and discuss the particular aspects addressed by each method.

Address
Max Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg, Germany

Corresponding author: Grün, Dominic (gruen@ie-freiburg.mpg.de)

Introduction

The emergence and maintenance of a complex multicellular organism requires a multitude of cellular differentiation decisions. These have to be executed with spatial and temporal precision in order to ensure that the appropriate number of cells of each type is produced in a tissue at any developmental stage. Given the stochasticity of the molecular processes underlying the interactions of thousands of genes in each cell, it is remarkable that stem cell differentiation is extremely robust and reproducible, always giving rise to the same complex organismal architecture even under highly variable external conditions [1–4]. It is one of the major goals of contemporary molecular biology to decipher the cellular processes underlying stem cell differentiation [5–7]. Although scientists have explored stem cell differentiation in great detail for decades, e.g. the early embryonic development of mammals [8] or the development and the homeostasis of the immune systems [6], fundamental aspects of cell fate decisions in these and other systems remain to be elucidated.

The recent establishment of a number of advanced single-cell RNA-sequencing protocols [9–18] for the large-scale analysis of single-cell transcriptomes [19–23] has begun to reveal unprecedented insight into the heterogeneity of organs and tissues and spawned the endeavor to characterize every cell type in the human body [24]. Seminal studies have resolved cell types in a variety of tissues, including lung [25], brain [26], skin [27], intestine [28] and bone marrow [29]. One of the main goals is to understand the process by which mature cell types are generated from multipotent cells during development or tissue homeostasis in the adult organism, requiring the inference of cellular differentiation trajectories. It comes as a major disadvantage of single-cell RNA-seq that the tissue has to be dissociated at a particular point in time and cells are lysed during the process and cells are lysed during the process, prohibiting the direct inference of ancestral relations between cells at distinct stages of differentiation. Although cutting-edge multiplexed lineage-tracing techniques utilizing CRISPR/Cas9 [30–33], recombination-based approaches [34,35] or lentiviral barcoding [36] allow the indirect inference of this information, these methods are challenging to establish, depend on the availability of cell type-specific inducible markers, and are limited in resolution.

As a common alternative approach, pseudo-temporal ordering permits the inference of differentiation trajectories from single-cell RNA-seq snapshot data of a given tissue. Here, cells are ordered by transcriptome similarity on a continuous trajectory or on a branched structure representing a lineage tree. Such methods assume continuity of transcriptome changes during differentiation and rely on the presence of all intermediate stages connecting