Unbiased modelling of dynamic processes with single-cell RNA-sequencing

Robrecht Cannoodt ROBRECHT.CANNOODT@UGENT.BE
Data Mining and Modelling for Biomedicine group, VIB Inflammation Research Center, Ghent, Belgium

Wouter Saelens WOUTER.SAELENS@UGENT.BE
Data Mining and Modelling for Biomedicine group, VIB Inflammation Research Center, Ghent, Belgium

Katleen De Preter KATLEEN.DEPRETER@UGENT.BE
Center for Medical Genetics, Ghent University, Ghent, Belgium

Yvan Saeys YVAN.SAEYS@UGENT.BE
Data Mining and Modelling for Biomedicine group, VIB Inflammation Research Center, Ghent, Belgium

Keywords: trajectory inference, single-cell transcriptomics

Abstract

Recent advances in RNA sequencing enable the generation of expression data at the single-cell level, opening up new avenues for transcriptomics and systems biology. A new application in the field of single-cell transcriptomics is the unbiased ordering of single cells along a developmental timeline. In this study, we developed SCORPIUS, a novel unsupervised technique to automatically infer developmental chronologies from single-cell RNA sequencing data. Upon comprehensive evaluation of SCORPIUS with other methods using ten datasets, we show that it outperforms and is able to extract novel biological knowledge from single-cell expression profiles.

1. Introduction

Recent advances in the field of high-throughput, single-cell technologies have generated a wealth of high-dimensional data that describe individual cells into unprecedented detail. More specifically, single-cell RNA sequencing is able to measure thousands of RNA transcripts for hundreds of cells in highly multiplexed assays. Computational methods that analyze single-cell RNA sequencing (scRNA-seq) allows not only to model well-defined cellular states, but also the transitions between them.

While several works have introduced TI methods for scRNA-seq data, the evaluation of these methods suffer from two main drawbacks: 1) they are only evaluated on a very limited number of datasets (mostly only one) and 2) they have not been evaluated in a quantitative way. As part of this study, we benchmarked existing approaches using ten publicly available datasets, and observed that none of the existing approaches perform well across all datasets.

2. Results

In this work we present SCORPIUS, a novel and fully unsupervised technique to investigate transcriptional changes in dynamic cellular processes (Figure 1). It starts from snapshot data that contains a mixture of cells along different states in a dynamic process (e.g. differentiation or cell cycle). It first denoises the data, projects the cells into a three-dimensional space, and constructs a model for the dynamic process by inferring a trajectory through the cells.

SCORPIUS is able to accurately reconstruct trajectories for a wide variety of dynamic cellular processes, including cell differentiation and cellular responses to external stimuli. We evaluated its performance in a comprehensive benchmark, providing for the first time a quantitative study of the performance of trajectory inference algorithms. The benchmark dataset consisted

Appearing in *Proceedings of Benelearn 2016*. Copyright 2016 by the author(s)/owner(s).

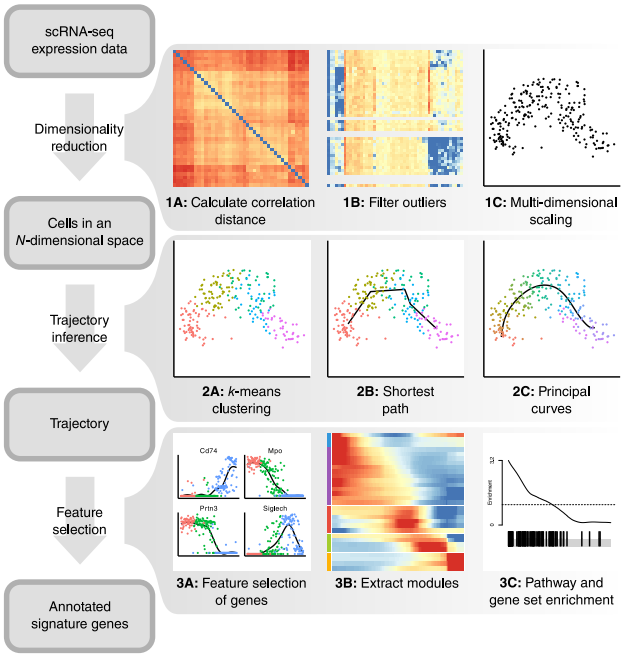


Figure 1. SCORPIUS infers a trajectory in three steps. 1) Dimensionality reduction involves calculating the correlation distance, optionally filtering out outliers, and performing multi-dimensional scaling. 2) Trajectory inference creates an initial path by calculating the shortest path through k cluster centers, which is then iteratively fitted to the data using the principal curves algorithm. 3) A feature selection technique selects the most important genes with respect to the inferred trajectory, which are then clustered and checked for gene set enrichment.

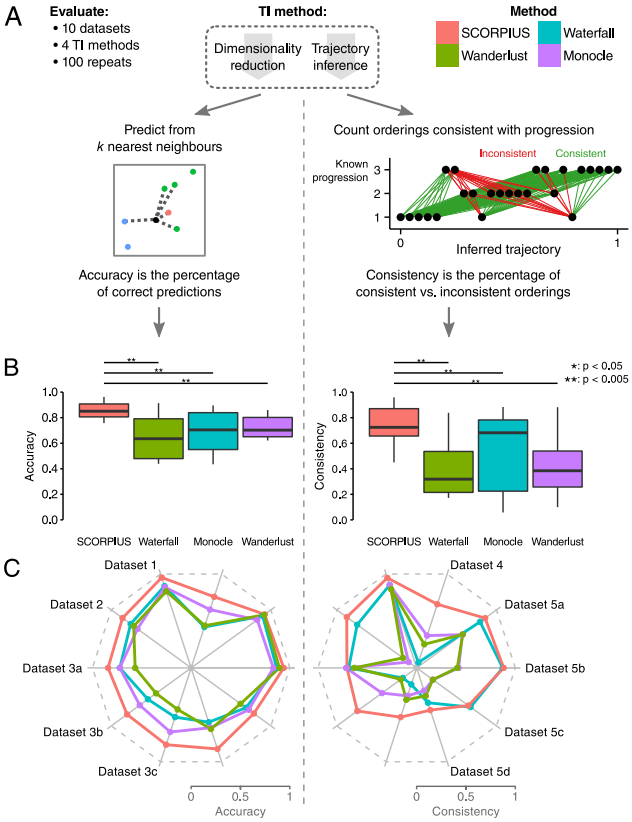


Figure 2. A) For each of the TI approaches, we evaluated two common steps amongst all approaches using 10 different datasets using the accuracy and consistency metrics. B) SCORPIUS outperforms other state-of-the-art approaches, both in dimensionality reduction as well as the ordering of the cells (*: p-value < 0.05, **: p-value < 0.005). C) Accuracy and consistency scores for every method and dataset.

of ten datasets, and evaluated four different TI methods: Wanderlust, Monocle, Waterfall and SCORPIUS. These methods each have a dimensionality reduction and trajectory inference step, therefore these are evaluated separately (Figure 2A). We show that SCORPIUS significantly outperforms existing TI approaches (Figure 2B), both for the dimensionality reduction as well as the trajectory inference steps. Upon further investigation, we see its dimensionality reduction step outperforms all other approaches on all other datasets, whereas its trajectory inference step is slightly outperformed by other approaches on datasets 3a and 5c (Figure 2C).

3. Discussion

Trajectory inference methods have recently been introduced to order single cells along a pseudo-temporal timeline implicitly present an scRNA-seq dataset. The resulting ordering can subsequently be used to infer novel dynamical properties of processes such as cellular development, differentiation and responses to stimuli. In this work, we introduced SCORPIUS, a new method that is able to accurately infer trajectories using single-cell transcriptomics data. We benchmarked its performance on 10 publicly available scRNA-seq data sets, comparing it to the state-of-the-art. Using a novel quantitative assessment, we showed that SCORPIUS consistently outperformed other methods, paving the way towards objective benchmark settings for trajectory inference methods to be developed in the future.

Acknowledgments

This work is supported by Fund for Scientific Research FWO Flanders (R.C. and W.S.).