

Simulation of Datasets with Splatter

ScMaSigPro Supplementary Material-II

Priyansh Srivastava, ... Ana Conesa

2023-08-16

Introduction

To evaluate the effectiveness of the scMaSigPro in controlling the False Positive Rates (FPR) and False Negatives Rates (FNR), we have evaluated its application on a wide variety of simulated scRNA-datasets. As in simulated scRNA-datasets, the ground truth is well established; therefore, simulations serve as the basis to benchmark the accuracy and precision of the method.

We have used Splatter, which uses a Gamma-Poisson distribution to simulate scRNA-Seq data [Zappia et al.]. We simulated Splatter's `splatSimulate(method = "pathS")` function to simulate a differentiation process where one cell type changes into another [Luke Zappia, Belinda Phipson and Alicia Oshlack]. Splatter approximates this process by simulating a series of steps between two groups and randomly assigning each cell to a **Step**. Since Pseudotime values are arbitrary, “**Step**” can be treated as Pseudotime. Additionally, just like Pseudotime, the **steps** simulated by Splatter also starts from 0, denoting immature cells first with a pseudotime of 0, followed by the cells in the transitional stage $\text{pseudotime} > 0$, and finally, the mature cells. [Deconinck et al.].

General Parameters for Datasets

We simulated bifurcating topologies (One cell type divides into two different cell types) with different parameters. 5000 features/Genes were simulated across 3000 cells/samples for all the simulations along two **paths/groups**. This produces a scRNA-Dataset where 1500 cells exist in each **paths/groups** following **Steps/Pseudotime** 0 to 1500. This represents an ideal case in which one cell state is associated with one value of **Step/Pseudotime**. This reflects an ideal sequencing experiment in which all the cell states are captured and sequenced, and the inferred **Steps/Pseudotime** orders all the cells in the native biological order. See Figure 1.

Annotation of Differential Genes (Ground Truth)

Let's assume a simple bifurcating trajectory with 4 genes *geneA*, *geneB*, *geneC* and *geneD*. Splatter, simulated the differential expression by simulating fold-change per **path/group** which reflect the resulting change at the end of the **path/Group** in relation to the start i.e. $Gene_{i_{base}}$. The final expression of any $gene_i$ is the product of $Path_{i_{foldChange}} * Gene_{i_{base}}$. the effective change in expression across the **Steps/Pseudotime** can be obtained by taking the difference between the start $Gene_{i_{base}}$ and end $Path_{i_{foldChange}} * Gene_{i_{base}}$.

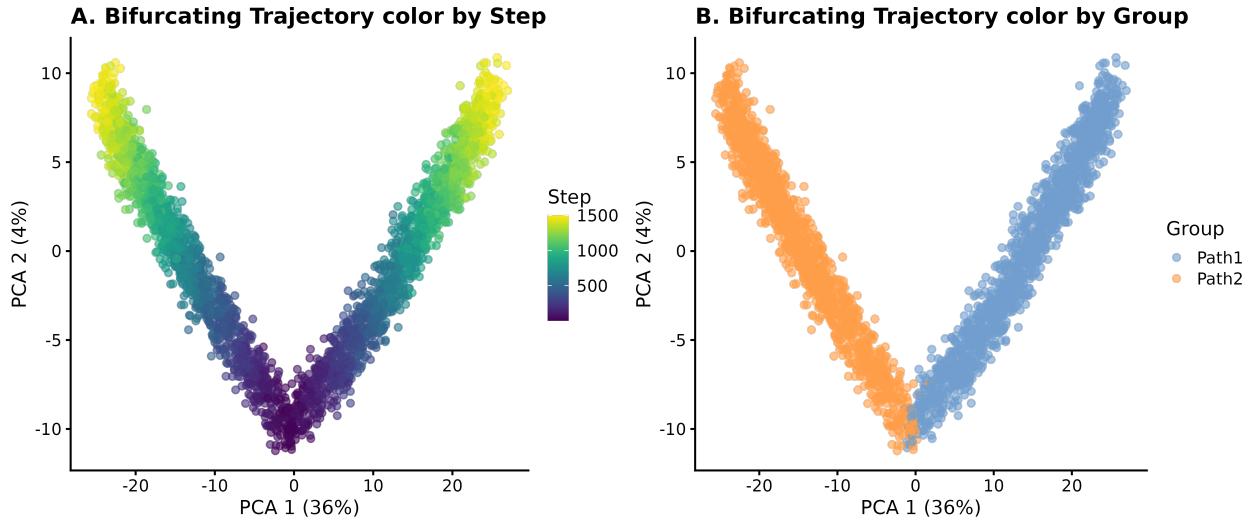


Figure 1: Ideal representation of native biological order. (A) Principal Components of bifurcating trajectory, where each cell is coloured by the associated Steps/Pseudotime. The length of the one path/group 1500 where each cell/sample is associated with one Steps/Pseudotime. (B) Each is coloured to represent a bifurcating trajectory

Specific Parameters for benchmarking

1. Zero-Inflation/Drop-out

References

Louise Deconinck, Robrecht Cannoodt, Wouter Saelens, Bart Deplancke, and Yvan Saeys. Recent advances in trajectory inference from single-cell omics data. 27:100344. ISSN 24523100. doi: 10.1016/j.coisb.2021.05.005. URL <https://linkinghub.elsevier.com/retrieve/pii/S2452310021000299>.

Luke Zappia, Belinda Phipson and Alicia Oshlack. Introduction to splatter. URL https://www.bioconductor.org/packages/release/bioc/vignettes/splatter/inst/doc/splatter.html#62_Simulating_paths.

Luke Zappia, Belinda Phipson, and Alicia Oshlack. Splatter: simulation of single-cell RNA sequencing data. 18(1):174. ISSN 1474-760X. doi: 10.1186/s13059-017-1305-0. URL <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1305-0>.

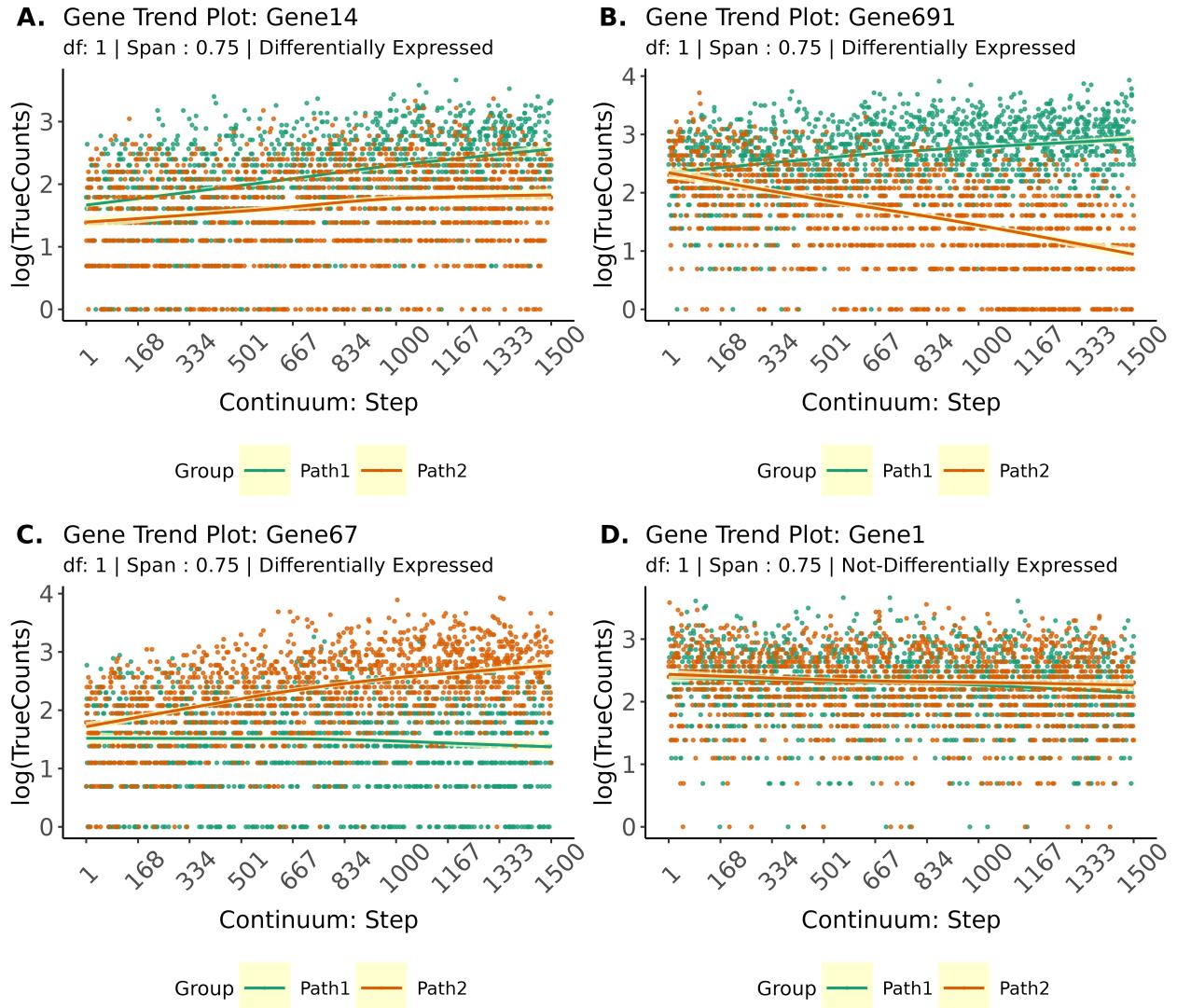


Figure 2: Ideal representation of native biological order. (A) Principal Components of bifurcating trajectory, where each cell is coloured by the associated Steps/Pseudotime. The length of the one path/group 1500 where each cell/sample is associated with one Steps/Pseudotime. (B) Each is coloured to represent a bifurcating trajectory