Author Contributions Checklist Form

This form documents the artifacts associated with the article (i.e., the data and code supporting the computational findings) and describes how to reproduce the findings.

# Part 1: Data

This paper **does not** involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

## Abstract

We have included the real data set in the “data.Rdata” file. It can be loaded via load(“../Data-and-Results/data.Rdata") The data set is described in Section 8 of the main paper. It is in the format of a 747x2 matrix, where

* **Rows**: cells (statistical units);
* **Columns**: named “MDS1”, “MDS2” represent the two biomarkers describing the geniting expressions.

## Availability

Data **are** publicly available

Data **cannot be made** publicly available

If the data are publicly available, see the *Publicly available data* section. Otherwise, see the *Non-publicly available dat*a section, below.

### Publicly available data

Data are available online at: The Github pages of the authors as a GitHub repository (now anonymized).

Data are available as part of the paper’s supplementary material.

Data are publicly available by request, following the process described here:

Click or tap here to enter text.

Data are or will be made available through some other mechanism, described here:

Click or tap here to enter text.

### Non-publicly available data

Discussion of lack of publicly available data:

Click or tap here to enter text.

## Description

### File format(s)

CSV or other plain text:

Software-specific binary format (.Rda, Python pickle, etc.): .Rdata

Standardized binary format (e.g., netCDF, HDF5, etc.):

Other (described here):

Click or tap here to enter text.

### Data dictionary

Provided by the authors in the following file(s): README

Data file(s) is (are) self-describiing (e.g., netCDF files)

Available at the following URL:

The Github pages of the authors as a GitHub repository (now anonymized).

### Additional information (optional)

Click or tap here to enter text.

# Part 2: Code

## Abstract

The codes are written in R.

The codes comprise two main programs:

1. a) **GARP\_main.R**: This is the main script that runs the analysis by calling the functions included in the other files. This script produces the MCMC samples of the single-cell RNA data analysis and reproduces the results of the analysis summarized in the main manuscript.
2. b) **GARP\_fcts.R:** This file contains all the R functions needed to run the main script, including the MCMC function to implement the sampler described in Section 6 of the main manuscript.

## Description

### Code format(s)

Script files

R  Python  Matlab

Other:

Package

R  Python  MATLAB toolbox

Other:

Reproducible report

R Markdown  Jupyter notebook

Other: Word

Shell script

Other (described here):

A repository is available at the Github pages of the authors as a GitHub repository (now anonymized).

### **Supporting software requirements**

Version of primary software used

R version 4.2.3

Libraries and dependencies used by the code

library(rstudioapi) # version 0.14

library(MASS) # version 7.3-58.2

library(ggplot2) # version 3.4.2

library(viridis) # version 0.6.2

library(salso) # version 0.3.29

library(reshape2) # version 1.4.4

library(Cairo) # version 1.6-0

library(scales) # version 1.2.1

library(plyr) # version 1.8.8

library(mvtnorm) # version 1.1-3

library(LaplacesDemon) # version 16.1.6

library(bayesm) # version 3.1-5

library(mniw) # version 1.0.1

library(xtable) # version 1.8-4

library(dplyr) # version 1.1.1

library(ggpubr) # version 0.6.0

library(clusterExperiment) # version 2.18.2

### Supporting system/hardware requirements (optional)

Click or tap here to enter text.

### Parallelization used

No parallel code used

Multi-core parallelization on a single machine/node

Number of cores used:

Multi-machine/multi-node parallelization

Number of nodes and cores used:

### License

MIT License (default)

BSD

GPL v3.0

Creative Commons

Other (described here):

Click or tap here to enter text.

### Additional information (optional)

Click or tap here to enter text.

# Part 3: Reproducibility workflow

## Scope

The provided workflow reproduces:

Any numbers provided in text in the paper

The computational method(s) presented in the paper (i.e., code is provided that implements the method(s))

All tables and figures in the paper

Selected tables and figures in the paper, as explained and justified here:

## Workflow details

### Location

The workflow is available:

As part of the paper’s supplementary material

In this Git repository: A repository is also available at the Github pages of the authors as a GitHub repository (now anonymized).

Other:

Click or tap here to enter text.

### Format(s)

Single master code file

Wrapper (shell) script(s)

Self-contained R Markdown file, Jupyter notebook, or other literate programming approach

Text file (e.g., a readme-style file) that documents workflow

Makefile

Other (more detail in 'Instructions' below)

### Instructions

Click or tap here to enter text.

Expected run-time

Approximate time needed to reproduce the analyses on a standard desktop machine:

<1 minute

1-10 minutes

10-60 minutes

1-8 hours

>8 hours

Not feasible to run on a desktop machine, as described here:

### Additional documentation (optional)

The MCMC algorithm takes around 30 minutes on a Lenovo machine with 32 Gb RAM. Alternatively, the user can load the results from a previous run of the algorithm (instructions in the body of the R script). Additional descriptions and instructions are included as detailed comments in the body of the R scripts.

# Notes (optional)

A repository is available at the Github pages of the authors as a GitHub repository (now anonymized).