

# 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

The data used in this paper is from a population scale single cell RNA-sequencing study from Perez et al 2022 on eQTLs for patients with the autoimmune disease SLE. The data includes over 1.2 million PBMCs with over 250 total individuals. Our paper focuses on three specified cell types using the annotations used from the original paper: CD4+ T-cells, CD8+ T-cells, and classical monocytes. We primarily use .RDS files to use Seurat and focus on cell-type specific signatures.

## 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 data* section, below.

### Publicly available data

- ☒ Data are available online at:  
[https://zenodo.org/records/17402494?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjM0MzNkNDE1LWY4ZTctNDVhYi1hODk5LWJmNzhjNzg4MDUxNyIsImRhdGEiOiJ9LCJyYW5kb20iOiI1OTVhOGVjZTBkYmZkZjBjMDA2ZTY4ZTBmNmVjN2Q3NiJ9.ONISAR5Zgx5GZ0odRZKmfSKmKTzBUTRyZ250S-hCc18EzXopSveq12rdOqvJt\\_VgHZaHObG8x909Sya\\_av9CVQ](https://zenodo.org/records/17402494?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjM0MzNkNDE1LWY4ZTctNDVhYi1hODk5LWJmNzhjNzg4MDUxNyIsImRhdGEiOiJ9LCJyYW5kb20iOiI1OTVhOGVjZTBkYmZkZjBjMDA2ZTY4ZTBmNmVjN2Q3NiJ9.ONISAR5Zgx5GZ0odRZKmfSKmKTzBUTRyZ250S-hCc18EzXopSveq12rdOqvJt_VgHZaHObG8x909Sya_av9CVQ)  
☐ Data are available as part of the paper's supplementary material.

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

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

## Non-publicly available data

Discussion of lack of publicly available data:

## Description

### File format(s)

- ☒ CSV or other plain text:
- ☒ Software-specific binary format (.Rda, Python pickle, etc.):
- ☐ Standardized binary format (e.g., netCDF, HDF5, etc.):
- ☐ Other (described here):

The file from the original study is of .h5ad format. Because we use Seurat objects, we use .RDS and .csv file formats for our processed objects.

### Data dictionary

- ☐ Provided by the authors in the following file(s):
- ☐ Data file(s) is (are) self-describing (e.g., netCDF files)
- ☒ Available at the following URL:

[https://github.com/qz91/sef\\_deDist](https://github.com/qz91/sef_deDist)

### Additional information (optional)

## Part 2: Code

### Abstract

The code is segmented into two directories (simulations and RDA), each with multiple .R files. For both directories, we provide a main file declaring each function used. We include code files that are reproducible for our key results as well.

### Description

#### Code format(s)

- ☒ Script files
- ☒ R   ☐ Python   ☐ Matlab
  - ☐ Other:
- ☐ Package
- ☐ R   ☐ Python   ☐ MATLAB toolbox
  - ☐ Other:
- ☐ Reproducible report
- ☐ R Markdown   ☐ Jupyter notebook
  - ☐ Other: .R file
- ☐ Shell script
- ☐ Other (described here):

We use R in the form of .R files, of which most of them can be run on any R-compatible environment. Permutation tests should be run on the command line.

### Supporting software requirements

#### Version of primary software used

R: 4.5.1, Rstudio 2025.9.1.401

#### Libraries and dependencies used by the code

R:  
dplyr

ggplot2  
tidyverse  
ggplot2  
mvtnorm  
MASS  
patchwork  
grid  
parallel  
truncnorm  
Seurat  
pbmcapply  
parallel  
doSNOW  
foreach  
doParallel  
Ake  
Matrix  
Hmisc  
clusterProfiler  
enrichplot  
org.Hs.eg.db

### Supporting system/hardware requirements (optional)

### Parallelization used

- ☐ No parallel code used
- ☒ Multi-core parallelization on a single machine/node  
Number of cores used: 2-8
- ☐ Multi-machine/multi-node parallelization  
Number of nodes and cores used:

### License

- ☐ MIT License (default)
- ☐ BSD
- ☐ GPL v3.0
- ☐ Creative Commons

☐ Other (described here):

Additional information (optional)

## 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:

Code to reproduce figures are listed in the data dictionary in the GitHub repository. The files listed denote which figures they correspond to.

### Workflow details

#### Location

The workflow is available:

- ☐ As part of the paper's supplementary material
- ☒ In this Git repository: [https://github.com/qz91/sef\\_deDist](https://github.com/qz91/sef_deDist)
- ☐ Other:

The data dictionary in the GitHub repository provides details regarding the contents of each file.

#### 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

The workflow is segmented by figure and idea. These are described clearly in the Data Dictionary in the README on the repository.

## 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:

We make a note that the permutation script specifically is extremely computationally expensive, and we list this for >8 hours.

## Additional documentation (optional)

## Notes (optional)