The Journal of Open Source Software

# Lazy matrix evaluation in R

**Eric Kernfeld** [1]

**1** Department of Biomedical Engineering, Johns Hopkins University, USA

**DOI:** 10.xxxxxx/draft

**Software**

- Review
- Repository
- Archive

**Editor:** Open Journals

**Reviewers:**

- @openjournals

**Submitted:** 01 January 1970
**Published:** unpublished



## Summary

Modern single-cell genomics technologies hold tremendous promise for cancer genomics, developmental biology, and cellular engineering (Griffiths et al., 2018; Lei et al., 2021; Rackham et al., 2021; Teichmann & others, 2017). Due to the breadth of applications for single-cell genomics and due to the idiosyncracy of individual biological systems, single-cell genomics admits no one-size-fits-all approach to quality control or analysis. Rather, the rapid adoption of these measurement devices (Svensson et al., 2020) has suddenly conscripted an entire cohort of computational biologists into labor-intensive exploration and bespoke analysis of large, sparse count matrices.

Individual datasets with minimal biological replication easily reach tens of thousands of observations across tens of thousands of genes or hundreds of thousands of functional genome elements, and RAM is a ubiquitous bottleneck. Preserving the initial sparsity of the data is a natural solution. Alas, this sparsity is often annihilated within the first few steps of a typical analysis – for example, by centering each gene to have a mean of 0. There is a need for software that can perform common statistical analysis of matrices, such as centering, scaling, and PCA, while preserving sparsity (Booeshaghi et al., 2022).

## Statement of need

`MatrixLazyEval` is an `R` package for sparse matrix operations that prioritizes simplicity and ease of use, anticipating and avoiding typical misunderstandings. For ease of use, it is compatible with `R`'s native `matrix` class as well as the popular `Matrix` and `DelayedArray` packages (see https://cran.r-project.org/web/packages/Matrix/index.html and https://bioconductor.org/packages/release/bioc/html/DelayedArray.html). Its workhorse objects use overloaded versions of typical `R` matrix operations such as `%*%`, `t`, and `tcrossprod`. Since this leaves some ambiguity about whether a given operation will be performed immediately or put off, `MatrixLazyEval` explicitly documents every function as lazy or eager.

The package provides implementations of:

- centering,
- scaling,
- replacing a matrix with its residuals after regression, and
- randomized SVD or PCA.

These operations can be easily chained/combined without loss of sparsity. The result is a viable drop-in replacement for matrix objects in single-cell genomics workflows.

## Acknowledgements

The author is grateful to Dr. René Maehr for access to single-cell genomics datasets at cutting-edge scale and complexity.

# References


Booeshaghi, A. S., Hallgrimsdottir, I. B., Galvez-Merchan, A., & Pachter, L. (2022). Depth normalization for single-cell genomics count data. *bioRxiv*.

Griffiths, J. A., Scialdone, A., & Marioni, J. C. (2018). Using single-cell genomics to understand developmental processes and cell fate decisions. *Molecular Systems Biology*, *14*(4), e8046.

Lei, Y., Tang, R., Xu, J., Wang, W., Zhang, B., Liu, J., Yu, X., & Shi, S. (2021). Applications of single-cell sequencing in cancer research: Progress and perspectives. *Journal of Hematology & Oncology*, *14*(1), 1–26.

Rackham, O., Cahan, P., Mah, N., Morris, S., Ouyang, J. F., Plant, A. L., Tanaka, Y., & Wells, C. A. (2021). Challenges for computational stem cell biology: A discussion for the field. *Stem Cell Reports*, *16*(1), 3–9.

Svensson, V., Veiga Beltrame, E. da, & Pachter, L. (2020). A curated database reveals trends in single-cell transcriptomics. *Database*, *2020*.

Teichmann, S., & others. (2017). The human cell atlas. *eLife*, *6*, e27041.