

MitoHEAR

Gabriele Lubatti^{*1,2,3}, Elmir Mahammadov^{†1,2,3}, and Antonio Scialdone^{1,2,3¶}

¹ Institute of Epigenetics and Stem Cells, Helmholtz Zentrum München, Munich, Germany ² Institute of Functional Epigenetics, Helmholtz Zentrum München, Neuherberg, Germany ³ Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany ¶ Corresponding author

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Charlotte Soneson](#) ↗

Reviewers:

- [@batoolmm](#)
- [@sahilseth](#)
- [@juanvillada](#)

Submitted: 21 February 2022

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

To produce the energy they need, eukaryotic cells rely on mitochondria, organelles that are equipped with their own DNA (mtDNA). Each cell includes multiple mtDNA copies that are not perfectly identical but have differences in their sequence; such sequence variability is called heteroplasmy. mtDNA heteroplasmy has been associated with diseases ([Nissanka & Moraes, 2020](#)), can affect cellular fitness and have an impact on cellular competition ([Lima et al., 2021](#)). Several single-cell sequencing protocols provide the data to estimate mtDNA heteroplasmy, including single-cell DNA-seq, RNA-seq and ATAC-seq, in addition to dedicated protocols like MAESTER ([Miller et al., 2021](#)). Here, we provide MitoHEAR (Mitochondrial Heteroplasmy Analyzer), a user-friendly software written in R that allows the estimation as well as downstream statistical analysis of the mtDNA heteroplasmy calculated from single-cell datasets. MitoHEAR starts from FASTQ files, computes the frequency of each allele and, starting from these, estimates the mtDNA heteroplasmy at each covered position for each cell. The parameters of the analysis (e.g., the filtering of the mtDNA positions based on read quality and coverage) are easily tuneable in a very user-friendly way. Moreover, statistical tests are available to explore the dependency of the mtDNA heteroplasmy on continuous or discrete cell covariates (e.g., culture conditions, differentiation states, etc), as extensively shown in the detailed tutorials we include.

Statement of need

Despite mtDNA heteroplasmy has important consequences on human health ([Stewart, 2015](#)) and embryonic development ([Floros VI, 2019](#)), there are still many open questions on how heteroplasmy affects cell's ability to function and regarding the mechanisms cells use to keep it under control. With the increasing availability of single-cell data, many questions can begin to be answered, but it is fundamental to have efficient and streamlined computational tools enabling researchers to estimate and analyse mtDNA heteroplasmy. Our package MitoHEAR covers all steps of the analysis, instead of focussing only on a few specific steps (as in, e.g., ([Huang & Huang, 2021](#)),([Prashant et al., 2021](#))): from the processing of raw FASTQ files to the estimation of the heteroplasmy and downstream statistical analysis, with user-friendly functions that are highly customizable.

Key functions

The two main functions of MitoHEAR are:

*first author
†co-author

1. `get_raw_counts_allele`: a parallelized function that relies on Rsamtools and generates the raw counts matrix starting from FASTQ files, with cells as rows and bases with the four possible allele as columns.
2. `get_heteroplasmy`: Starting from the output of `get_raw_counts_allele`, it computes the matrix with heteroplasmy values (defined as 1 minus the frequency of the most common allele) and the matrix with allele frequency values, for all the cells and bases that pass a filtering step procedure.

Among the downstream analyses implemented in the package there are: 1. Several statistical tests (e.g., Wilcoxon rank-sum test) for the identification of the mtDNA positions with the most different levels of heteroplasmy between discrete groups of cells or along a trajectory of cells (i.e., cells sorted according to a diffusion pseudo time); 2. Plotting functions for the visualization of heteroplasmy and the corresponding allele frequency values among cells; 3. Hierarchical clustering of cells based on a distance matrix defined from the angular distance of allele frequencies that could be relevant for lineage tracing analysis (Ludwig et al., 2019)

The package has been used in a recently published paper (Lima et al., 2021), where we revealed that cells with higher levels of heteroplasmy are eliminated by cell competition in mouse embryos and are characterized by specific gene expression patterns.

References

- Floros VI, D. S., Pyle A. (2019). Segregation of mitochondrial DNA heteroplasmy through a developmental genetic bottleneck in human embryos. *Nat Cell Biol.* <https://doi.org/10.1038/s41556-017-0017-8>
- Huang, X., & Huang, Y. (2021). Cellsnp-lite: an efficient tool for genotyping single cells. *Bioinformatics.* <https://doi.org/10.1093/bioinformatics/btab358>
- Lima, A., Lubatti, G., Burgstaller, J., Hu, D., Green, A., Gregorio, A. D., Zawadzki, T., Pernaute, B., Mahammadov, E., Dore, M., Sanchez, J. M., Bowling, S., Sancho, M., Karimi, M., Carling, D., Jones, N., Srinivas, S., Scialdone, A., & Rodriguez, T. A. (2021). Cell competition acts as a purifying selection to eliminate cells with mitochondrial defects during early mouse development. *Nature Metabolism.* <https://doi.org/10.1038/s42255-021-00422-7>
- Ludwig, L. S., Lareau, C. A., Ulirsch, J. C., Christian, E., Muus, C., Li, L. H., Pelka, K., Ge, W., Oren, Y., Brack, A., Law, T., Rodman, C., Chen, J. H., Boland, G. M., Hacohen, N., Rozenblatt-Rosen, O., Aryee, M. J., Buenrostro, J. D., Regev, A., & Sankaran, V. G. (2019). Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell*, 176(6), 1325–1339.e22. <https://doi.org/10.1016/j.cell.2019.01.022>
- Miller, T. E., Lareau, C. A., Verga, J. A., Ssozi, D., Ludwig, L. S., Farran, C. E., Griffin, G. K., Lane, A. A., Bernstein, B. E., Sankaran, V. G., & Galen, P. van. (2021). Mitochondrial variant enrichment from high-throughput single-cell RNA-seq resolves clonal populations. *bioRxiv.* <https://doi.org/10.1101/2021.03.08.434450>
- Nissanka, N., & Moraes, C. T. (2020). Mitochondrial DNA heteroplasmy in disease and targeted nuclease-based therapeutic approaches. *EMBO Reports*, 21(3), e49612. <https://doi.org/10.15252/embr.201949612>
- Prashant, N., Alomran, N., Chen, Y., Liu, H., Bousounis, P., Movassagh, M., Edwards, N., & Horvath, A. (2021). SCReadCounts: Estimation of cell-level SNVs from scRNA-seq data. *bioRxiv.* <https://doi.org/10.1101/2020.11.23.394569>
- Stewart, C., J. (2015). The dynamics of mitochondrial DNA heteroplasmy: Implications for human health and disease. *Nat Rev Genet.* <https://doi.org/doi.org/10.1038/nrg3966>