Concomitant sequencing of accessible chromatin and mitochondrial genomes in single cells using mtscATAC-seq

Leif S. Ludwig[1,2] and Caleb A. Lareau[3]

1. Berlin Institute of Health at Charité Universitätsmedizin Berlin, 10178 Berlin, Germany
2. Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association, Berlin Institute for Medical Systems Biology, 10115 Berlin, Germany
3. Departments of Genetics and Pathology, Stanford University, Stanford, CA 94305, USA

Abstract

Mitochondria are unique organelles of eukaryotic cells that carry their own multi-copy number and circular genome. In most mammals, including humans and mice, the size of the chromosome is ~16,000 base pairs and unlike nuclear DNA, mitochondrial DNA (mtDNA) is not densely compacted. This results in mtDNA to be highly accessible for enzymes such as the Tn5 transposase, commonly used for accessible chromatin profiling of nuclear chromatinized DNA. Here, we describe a method for the concomitant sequencing of mtDNA and accessible chromatin in thousands of individual cells via the mitochondrial single-cell assay for transposase accessible chromatin by sequencing (mtscATAC-seq). Our approach extends the utility of existing scATAC-seq products and protocols as we 1) fix cells using formaldehyde to retain mitochondria and its mtDNA within its originating cell; 2) modify lysis conditions to permeabilize cells and mitochondria; and 3) optimized bioinformatic processing protocols to collectively increase mitochondrial genome coverage for downstream analysis. Here, we discuss the essentials for the experimental and computational methodologies to generate and analyze thousands of multi-omic profiles of single cells over the course of a few days, enabling the profiling of accessible chromatin and mtDNA genotypes to reconstruct clonal relationships and studies of mitochondrial genetics and disease.

Key Words Single cell multi-omics, accessible chromatin profiling, mitochondrial DNA, somatic mutation, lineage tracing, pathogenic mutation, mitochondrial disease

1. Introduction

Single cell genomic approaches have revolutionized our ability to comprehensively characterize cellular states and cell types, revealing a previously unappreciated heterogeneity and diversity of human cells in health and pathology. More recently, these efforts have focused on the integration and/or simultaneous detection of multiple high dimensional data types from the same single cells, thereby enabling an unprecedented depth of phenotyping of the building blocks of our organs [1]. Within this realm, the single cell Assay for Transposase Accessible Chromatin by Sequencing (scATAC-seq), has enabled the characterization of cell type and state-specific gene regulatory elements and transcription factor activities that orchestrate gene expression activity. Notably, the initial versions of the bulk ATAC protocol yielded up to 50% of sequencing reads that mapped to mitochondrial DNA (mtDNA) and which were originally perceived as a nuisance of the assay [2, 3]. Indeed, mtDNA is a relatively small circular genome (~16,591bp in humans) that is present in multiple copies per mitochondria and thus per cell and is readily tagmented by the Tn5 transposase used for ATAC-seq.

Leveraging these features of mtDNA and the ATAC-seq workflow, we recently described mtscATAC-seq, which enables the massively parallel single cell mitochondrial DNA genotyping and concomitant accessible chromatin profiling across thousands of cells. Conceptually, mtscATAC-seq extends the typical scATAC-seq workflow by using whole, but fixed and permeabilized cells as input, though the overall protocol remains largely unchanged (**Figure 1**). Due to the ease and power of this multi-modal method, mtscATAC-seq thereby enables multiple avenues of research. One area includes the study of fundamentals of mitochondrial genetics, mutational patterns and their consequences on cellular function, including those of pathogenic mtDNA mutations associated with congenital mitochondrial disorders, such as myoclonic epilepsy with ragged red fibers (MERRF) or mitochondrial encephalomyopathy with lactic acidosis and stroke-like episodes (MELAS) [4, 5]. Moreover, somatic mtDNA mutations may be used as clonal markers to reconstruct cellular relationships and dynamics, thereby enabling lineage and clonal tracing studies of human cells *in vivo* in combination with cell (pheno)typing as demonstrated in human hematopoiesis, different types of leukemia and solid cancers [6–8].

Here, we outline the essentials of the experimental workflow, which we extensively tested with primary human hematopoietic cells and the 10x Genomics scATAC v1 and NextGEM v1.1 kits. We describe best practices that should in principle translate to other cell types and organs of interest and important modifications to standard scATAC-seq protocols, which primarily isolate

nuclei and/or specifically deplete mitochondria. We illustrate the (pre-)processing of the resulting sequencing data to enable mtDNA genotyping alongside accessible chromatin profiling and outline avenues for further analysis.

2. Materials

2.1 Cell processing, fixation and lysis

1. Phosphate buffered saline (PBS)
2. FACS buffer: PBS with 1% FBS. Filtered at 0.45µm, store at 4°C.
3. Formaldehyde, 16%.
4. Glycine solution, 2.5M.
5. Lysis buffer: 10mM Tris-HCl pH 7.4, 10 mM NaCl, 3mM $MgCl_2$, 0.1% NP40, 1% bovine serum albumine (BSA). Prepare fresh, keep on ice until use (*see* **Note 1**).
6. Wash buffer: 10mM Tris-HCl pH 7.4, 10 mM NaCl, 3mM $MgCl_2$, 1% BSA. Prepare fresh, keep on ice until use (*see* **Note 1**).
7. Flowmi Cell Strainer 40µm

2.2 mtscATAC-seq library preparation

1. 10x Genomics Chromium Next GEM Single Cell ATAC Library & Gel Bead Kit, 16 or 4 rxns (*see* **Note 2**)
2. 10x Genomics Chromium Next GEM Chip H Single Cell Kit, 48 or 16 rxns
3. 10x Genomics Single Index Kit N, Set A, 96 rxns

2.3 Quality control and sequencing

1. Qubit dsDNA HS Assay Kit
2. Agilent Bioanalyzer High Sensitivity DNA Analysis Kit
3. Illumina NovaSeq or NextSeq reagent kits (*see* **Note 3**)

2.4 Computational resources

1. 10x Genomics Cell Ranger ATAC package (*see* **Note** 4) (https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/what-is-cell-ranger-atac)
2. mgatk package and dependencies (https://github.com/caleblareau/mgatk) (*see* **Note** 5)

3. Methods

The protocol described here has been optimized for the use of hematopoietic cell lines and primary human hematopoietic cells, including peripheral blood or bone marrow derived mononuclear cells that have been obtained via the use of standard approaches such as ficoll-based gradient centrifugation. Specific populations of interest may be enriched for example via flow cytometry based sorting. A high viability of cells (>95%) and a low residual granulocyte/neutrophil content (<3%) is essential to obtain high quality mtscATAC-seq data. For primary hematopoietic cells, we have processed these fresh or cryopreserved them using standard practices (e.g. in 90% FBS with 10% DMSO) with no significant loss of data quality following thawing and processing of cells when combined with sorting to ensure high viability of the input cell population. For the centrifugation of cells we recommend the use of DNA lobind microcentrifuge tubes and favour swinging-bucket centrifuges compared to fixed angle rotors. For mtscATAC-seq library preparation, we follow the Chromium Next GEM Single Cell ATAC Reagent kits v1.1 user guide from 10x Genomics (CG000209 Fev F) and only briefly describe modified steps as outlined in 3.2.

3.1 Cell processing, fixation and lysis

1. Transfer $1x10^5$ to $1x10^6$ live cells to a 1.5 ml microcentrifuge tube and spin for 5 minutes at 400 g at 4°C. Discard the supernatant without disrupting the cell pellet, resuspend and wash cells in 1-1.5 ml FACS buffer and spin for 5 minutes at 400 g at 4°C.
2. Discard the supernatant without disrupting the cell pellet, gently flick the tube to loosen the cell pellet and carefully and completely resuspend the cells in 450 µl of room temperature PBS.
3. Fix cells by adding formaldehyde to a final concentration of 1% (e.g. by adding 30 µl of 16% formaldehyde), followed by inversion of the tube for complete mixing. Incubate at room temperature for 10 minutes and occasionally invert the tube.
4. Quench the fixation reaction by adding a glycine solution to a final concentration of 0.125 M and invert the tube for complete mixing. Add 950 µl PBS or FACS buffer, invert the tube 2-3x times and spin for 5 minutes at 400 g at 4°C. Discard the supernatant without disrupting the cell pellet, gently flick the tube to loosen the cell pellet, resuspend cells and repeat the wash with 1-1.5 ml FACS buffer and spin for 5 minutes at 400 g at 4°C.
5. Discard the supernatant without disrupting the cell pellet, gently flick the tube to loosen the cell pellet and add 200-300 µl ice-cold lysis buffer. Gently pipette up and down 3x

times to completely resuspend the cells. Incubate on ice for 3 minutes, before adding 1 ml of ice-cold wash buffer and spin for 5 minutes at 500 g at 4°C (*see* **Note 6**).

6. Discard the supernatant without disrupting the cell pellet, gently flick the tube to loosen the pellet and resuspend cells in freshly prepared 1x nuclei buffer provided by the 10x Genomics scATAC-seq kit. Aim for a concentration of 2,000-7,500 cells/µl, as validated by counting an aliquot of cells mixed with trypan blue using a hematocytometer (e.g. Neubauer Improved) or a ThermoFisher Countess II or III automated cell counter (*see* **Note 7**). If cell clumps are abundant, the cell suspension may be filtered using for example 40 µm Flowmi cell strainers. Immediately proceed with the next steps of the protocol.

3.2 mtscATAC-seq library preparation

1. Adjust the cell concentration as desired with 1x nuclei buffer, following the recommendations by 10x Genomics. We typically aim for a concentration of about 2,500 cells/µl and note that only 5 µl of cell suspension may be used for the tagmentation reaction.
2. The cells are mixed with the transposition mix on ice in a suitable PCR tube, followed by transposition at 37°C before proceeding with GEM generation and barcoding using linear PCR and post GEM incubation cleanup. Please follow the detailed instructions of the 10x Genomics user guide for these steps without modifications (*see* **Note 8**).
3. For the library construction step involving the index PCR of the mtscATAC-seq sample, we typically conduct 1-2 additional cycles of PCR (*see* **Note 9**), before cleaning up of the libraries as described.

3.3 Quality control and sequencing

1. The yield of the mtscATAC-seq libraries is determined using a Qubit dsDNA HS Assay kit following the manufacturer's recommendations. We typically use 1 µl of the library and typically yield 5-20 ng/µl depending on cell type and used cell input.
2. The size distribution of the mtscATAC-seq library is assessed using an Agilent 2100 Bioanalyzer system and a High sensitivity DNA analysis kit using 1-10 ng of the library. Typical bioanalyzer traces of libraries prepared with the original 10x scATAC-seq and the modified mtscATAC-seq protocols are shown in **Figure 2** (*see* **Note 10**). We typically set region gates at 100 and 9000 bp to assess the molarity and yield in ng/µl of the mtscATAC-seq library via the Agilent Bioanalzyer software and compare the yields to the Qubit results, which should be largely concordant (+/-10-20% variance, *see* **Note 11**).

3. For sequencing we follow the recommendations by 10x Genomics and typically use Illumina NovaSeq and NextSeq reagent kits using paired-read sequencing (Read 1 and 2: 50-100 bp, Index Read 1: 8 bp, Index Read 2: 16 bp). Longer reads will increase mitochondrial genome coverage to more comprehensively capture genomic variants along sequenced DNA fragments. We note that reads derived from mtDNA have a median insert size of 120 bases (see **Figure 3**), and fully covering the majority of the bases in a molecule on a sequencing read will be advantageous to detect low frequency / heteroplasmy variants.
4. We typically recommend sequencing ~50,000 reads / cell yielded in the library capture. This value may vary depending on the cell input type, the overall viability of the input sample, and the depth of the desired analysis (*see* **Note 12**).

3.4. Computational processing and analyses

A feature of the mtscATAC-seq protocol that we emphasize is that each of the multiple modalities (mtDNA genotypes and accessible chromatin) is contained within a single library, which provides a convenient bioinformatics workflow and mitigates the risk of uncoupling modalities from a single capture. Furthermore, we've designed the mtscATAC-seq workflow to capitalize on multiple cores / threads via parallel computing.

1. Demultiplex mtscATAC-seq sequencing data using cellranger-atac mkfastq (*see* **Note 13**).

   $ cellranger-atac mkfastq --id=mtscatac_seq_fastqs --run=/path/to/flow_cell --csv=sample_sheet.csv
2. Download a NUMT modified reference genome (*see* **Notes 14, 15**). This approach results in more uniform coverage of the mitochondrial genome by eliminating multi-mapping biases with NUMT regions in the nuclear genome that are highly homologous to sequences within mtDNA (*see* **Note 16**).
3. Align the full mtscATAC-seq libraries to the blacklisted reference genome using the cellranger-atac `count` function (*see* **Note** 17).

   $ cellranger-atac count --reference /path/to/masked/reference --sample mtscatac_sample_output --fastq mtscatac_seq_fastqs/flow_cell
4. Genotype the mtDNA data using mgatk (*see* **Note** 18).

   $ mgatk tenx -i mtscatac_sample_output/outs/possorted_bam.bam -n mtscatac_sample_mgatk -o mtscatac_sample_mgatk -bt CB -b mtscatac_sample_output/outs/filtered_peak_bc_matrix/barcodes.tsv

5. Collect processed data files necessary for downstream analysis. These include relevant accessible chromatin summary files from the cellranger-atac execution available in the **mtscatac_sample_output/outs** directory and relevant mgatk output files in **mtscatac_sample_mgatk/final** (*see* **Note 19**).
6. Interpret the presence and abundance of variable heteroplasmic variants for downstream analyses in the ***.vmr_strand_plot.png** file (*see* **Note 20**). An example plot derived from a typical and high-quality mtscATAC-seq library is shown in **Figure 5**.
7. For biological samples that span multiple-libraries, one often uses the cellranger-atac aggr function to generate a comprehensive peak set, peaks-by-cells matrix, and total fragments file. However, the .bam file is not aggregated, meaning that the mgatk output must come from the per-library processing as described in **Step 4** of this section.
8. Perform interactive analyses using widely-used tools such as Signac [9]. We note that while other tools such as SnapATAC [10] and ArchR [11] have full suites of interactive functionality for chromatin accessibility data, Signac currently has the only built-in functionality for directly working with mtscATAC-seq data, in particular including the mitochondrial DNA genotyping described here (*see* **Notes 21, 22**).
9. Identify high-quality cells for downstream analysis. We note that mtDNA-specific per-cell quality control metrics, such as mean mtDNA coverage, can be incorporated for quality control and analysis of populations. In general and in our experiences, cells that have high-quality accessible chromatin profiles also have well-captured amounts of mtDNA (**Figure 6**, *see* **Note 23**) .
10. Identify lineage-biased mtDNA variants. The power of the mtscATAC-seq workflow is the concomitant inference of cell state via chromatin accessibility and clonal relationships via the mtDNA mutation profiles. Appropriate statistical tests that we've used in the past include the Kruskal-Wallis and Chi-Squared tests of association (*see* **Note 24**).

We emphasize that the biological questions of interest within a specific library will largely dictate the exact computational workflow after **Step 6**. Examples of custom scripts used for varied downstream analyses, including nucleotide enrichment analyses, trajectory inferences, clonal bias, and longitudinal sampling are available online: https://github.com/caleblareau/mtscATACpaper_reproducibility.

4. Notes

1. Note that the lysis and wash buffer do not contain Tween 20 as is being used in many scATAC-seq workflows, including the standard protocol by 10x Genomics. We omit Tween

20 as it depletes mitochondria and mtDNA within [3, 4], thereby preventing the sequencing and identification of mtDNA variants.

2. For mtscATAC-seq library preparation, we refer the reader to the detailed instructions of the user guide by 10x Genomics, which further includes a detailed list of reagents, consumables and best practices required to successfully complete the protocol.
3. For sequencing we have successfully worked with the Illumina NextSeq and NovaSeq reagents kits and respective sequencing platforms. We typically have used kits with 150 to 200 cycles to obtain high coverage of the mitochondrial genome for variant calling enabled by the longer read lengths.
4. The 10x Genomics cellranger-atac software comes as a stable binary that requires no installation aside from untarring the requisite files and placing them in a stable directory for execution.
5. A complete discussion of dependencies and installation instructions is available here: https://github.com/caleblareau/mgatk/wiki/Installation.
6. Lysis time may need to be optimized depending on cell type. Lysis efficacy may be assessed via the quantification of live/dead cells and should be performed on unfixed cells. Please also see the 10x Genomics demonstrated protocol *Nuclei isolation for single cell ATAC sequencing* (CG000169 Rev D) and the troubleshooting section within.
7. To obtain a sufficiently high cell concentration for the tagmentation reaction we initially resuspend the cells in a small volume of 1x nuclei buffer to avoid the need to concentrate further via an additional centrifugation step. The initial volume is dependent on the starting cell number and for 300,000 cells we would typically resuspend in 20-30 µl from which to obtain a first cell count using 5 µl of the cell suspension. One typically loses some cells during the upstream processing and it is advisable to be more conservative before diluting the cell concentration too much. For overloading of 10x channels, for example when pooling multiple cell lines or cells of multiple donors or when applying hashing-based approaches [12] to enable downstream computational demultiplexing of the sample origin, a higher cell concentration will be required.
8. Given the fixation of cells prior to lysis/permeabilization, we have attempted to incorporate a decrosslinking step as part of the GEM incubation PCR program. Extended decrosslinking at 60°C or 72°C for 1 h to 12 h did not appear to significantly improve mtscATAC-seq data quality. As such and given the high temperatures during cycling we recommend working with the standard PCR conditions recommended by 10x Genomics.

9. The number of cycles for the index PCR are a function of cell input as well as cell type. Cell lines typically yield a higher number of accessible chromatin fragments, and cells with higher mtDNA content will also show higher yield, while the fixation appears to diminish nuclear library complexity. As such, the optimal numbers of cycles during the index PCR may need to be determined empirically. A yield of 10-15 ng/µl in an elution volume of 20 µl usually provides sufficient material for downstream library preparation for sequencing.
10. For mtscATAC-seq bioanalyzer traces, the shape and size distribution is also a function of cell type and mtDNA content. We have observed the first nucleosome free peak (200-300 bp) to also be lower or even higher than the second mono-nucleosome peak (300-400 bp) as depicted in **Figure 2**. Ultimately, only sequencing enables one to obtain proper quality control metrics to assess the success of the experiment.
11. For mtscATAC-seq library quantification for sequencing we prefer the outlined method and have found it to be sufficiently reliable and less time consuming than the qPCR based approaches that are being recommended by 10x Genomics or other protocols.
12. As an alternative strategy, we perform a two-pass sequencing strategy where we first sequence ~25,000 reads / cell and then examine the library complexity of the overall run based on the CellRanger-ATAC quality control file. Based on this estimated complexity and the percent duplicates, we will perform additional sequencing, typically equalling or slightly surpassing the overall complexity and surpassing ~40% duplicates.
13. cellranger-atac mkfastq wraps the Illumina bcl2fastq software, which must also be in your environment. A description of this software and support for building the sample sheet based on the supplied indices is available here: https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/using/mkfastq
14. The widely-used pre-built references are available on the 10x Genomics website: https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/advanced/references.
15. We recommend generating a custom reference using a pre-generated blacklist of NUMTs and other blacklist features. This is achieved by hard-masking the .fasta reference file in the cellranger-atac reference as we describe here: https://github.com/caleblareau/mgatk/wiki/Increasing-coverage-from-10x-processing. A repository for handling .bed files of NUMT regions for common reference genomes is available https://github.com/caleblareau/mitoblacklist as well as custom code to generate these reference files for uncommon reference genomes.

16. We note that this approach is highly efficacious because NUMTs are present in individual cells at a much lower copy number than mtDNA and are not necessarily accessible to the Tn5 enzyme used for scATAC-seq. Other approaches for mitigating mapping biases in the mtDNA genome include re-mapping of multi-mapping reads or mapping to a shifted reference. Our recommended approach of mapping to the modified reference provides a one-step execution that is very parsimonious.
17. This is the most computationally intensive part of the protocol. We recommend specifying a minimum of 12 cores for efficient processing. Further, we recommend using cellranger-atac version 2.0+ that has been greatly improved for computational efficiency and runtime performance.
18. The main inputs into this execution include the .bam file form the cellranger-atac processing, this list of barcodes that are to be analyzed downstream (we use all cells that pass the knee threshold via the **-b** flag and specification, but any user-provided list would suffice. A full list of user parameters for mgatk is provided by the **mgatk --help** option. We note that by default PCR duplicates (fragments that share the same cell barcode and both transposition events) are removed by default, which can be retained with the **--keep-duplicates** flag. Retaining PCR duplicates could be useful in settings where the mtDNA copy number is particularly high and the chances of multiple duplicate transposition events is a possibility. Further, the number of cores available for parallel processing can be specified using the **--ncores** flag.
19. From the cellranger output folder, we typically recommend retaining the fragments.tsv.gz (+ index) and single-cell .csv files. However, the .bam file from this run is required for initial genotyping with mgatk. For the mgatk output, we typically only keep the *.rds file, which summarizes the rest of the contents. Notably, the outputs of this folder are redundant, and retaining the .A/.C/.G/.T plain text files represent sufficient summary of the data for downstream analyses though may be less convenient than the other files in the output folder.
20. The x-axis depicts the strand correlation in per-cell heteroplasmy across all cells in the experiment. Mutations separated on this axis are typically low quality (low strand correlation) versus high quality (high strand correlation), and we employ a density-based threshold to make these calls. The y-axis depicts the per-variant variance-mean-ratio (VMR) across all cells. Mutations separated on the y-axis represent heteroplasmic (high VMR) versus homoplasmic allele frequencies. The population of mutations high for both values represent high-quality variants used for downstream analyses. However,

homoplasmic mutations may also be useful in certain analysis settings, such as identification of the mtDNA haplogroup.

21. For interactive analyses using Signac, we recommend the following online vignette: https://satijalab.org/signac/articles/mito.html. Many accessory functions have been built into Signac since version 1.0.0 to facilitate the importing, analysis, and interpretation of mgatk data for multi-modal analyses.
22. For use in ArchR, we recommend using the addCellColData() function to append per-cell heteroplasmy values associated with the mgatk output from the ***.cell_heteroplasmic_df.tsv.gz** file.
23. In our experience, we typically recommend thresholding on a minimum 10x or 20x mean per-cell coverage alongside other commonly used metrics, such as transcription start site (TSS) score, # unique nuclear fragments, fraction of reads in peaks, and/or nucleosome score. The mean coverage per cell metric is computed automatically in the mgatk output, but is simply defined as the mean per-base coverage of the mgatk output for a given cell.
24. From our current analyses, Kruskal-Wallis is a non-parametric test of association between clusters (identified via the accessible chromatin cell state) and a continuous value (per-cell heteroplasmy). If a relevant heteroplasmy cutoff can be drawn (at say 5%, 10%, etc.), the association between presence/absence of a variant and the cell state clusters can be best modeled using a Chi-squared test of association. These statistical tests are performed independently via a loop over all variants in the high-quality set identified by mgatk.

Acknowledgements

L.S.L. is supported by an Emmy Noether fellowship by the German Research Foundation (DFG, LU 2336/2-1) and a Hector Research Career Development Award. C.A.L. is supported by a Stanford Science Fellowship and a Parker Institute of Cancer Immunotherapy Scholarship. We are grateful to the Ludwig lab for useful feedback in the generation of this protocol.

Figures

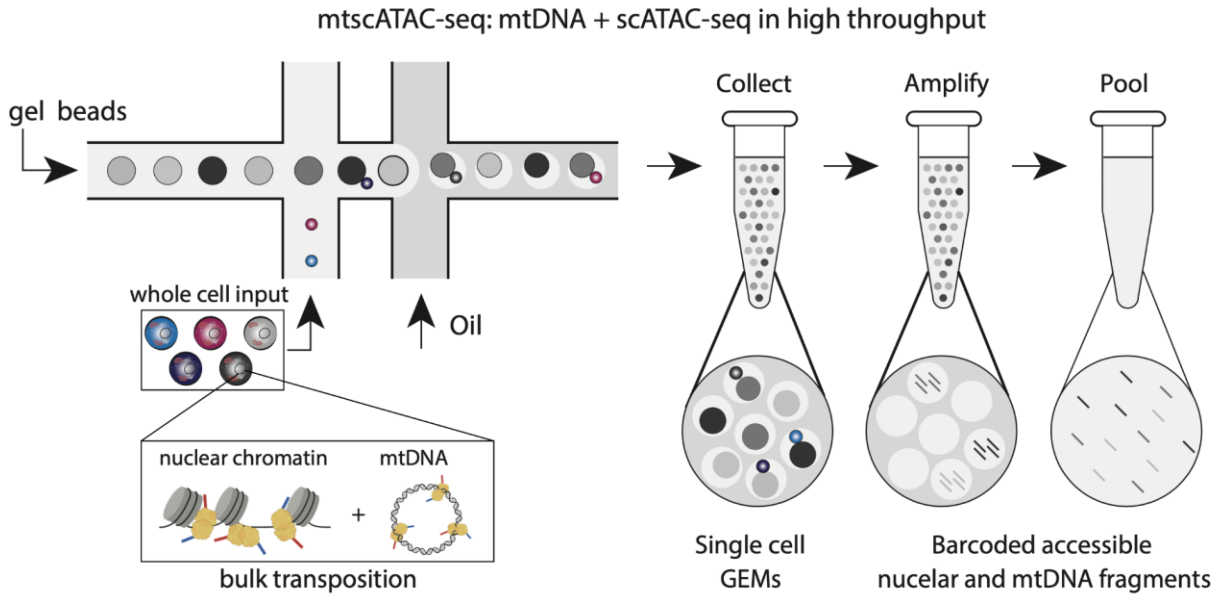

Figure 1. Schematic of the mtscATAC-seq reaction. Fixed and permeabilized whole cells are used as input into the Tn5 transposition reaction wherein both mtDNA and accessible chromatin are transposed within cells before being input into the 10x Chromium controller microfluidic device. These key differences are colored in the schematic. The remaining steps (greyscale) closely match the standard 10x scATAC-seq workflow.

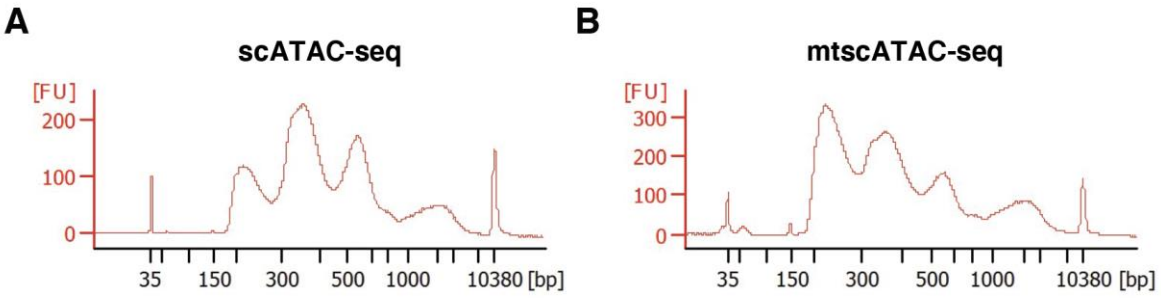

Figure 2. Representative bioanalyzer traces of sequencing libraries. (A) Original scATAC-seq and **(B)** mtscATAC-seq library of human peripheral blood mononuclear cells. We note the increased abundance of the nucleosome free (size <300bp) region in the mtscATAC-seq library relative to the scATAC-seq library that corresponds to the significant increase of captured mtDNA fragments.

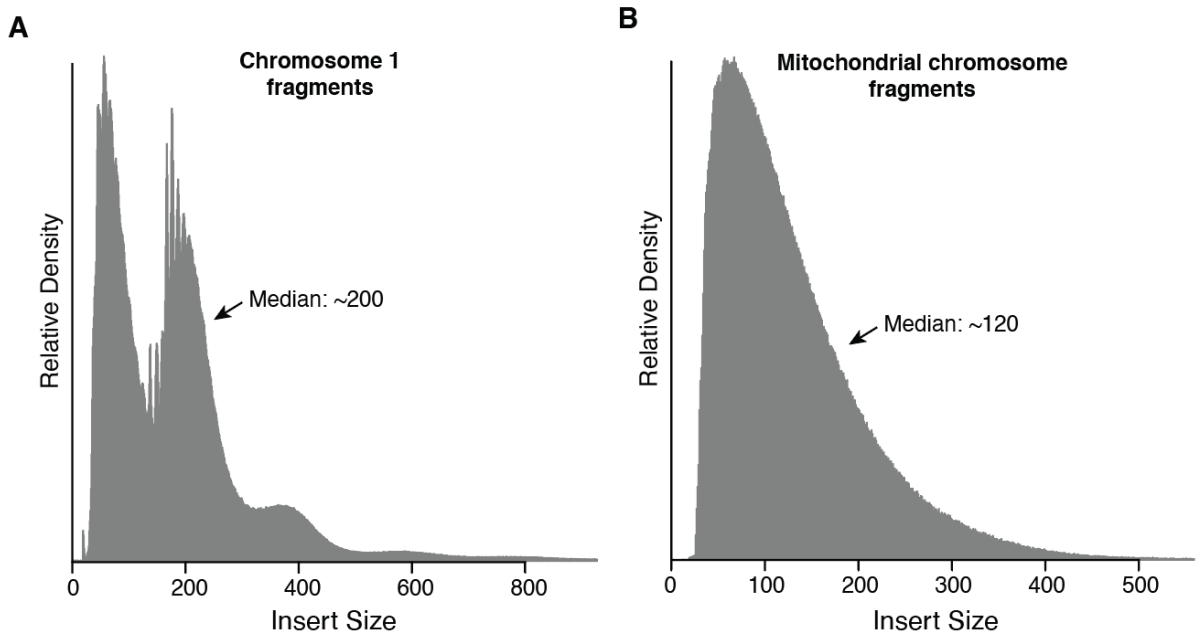

Figure 3. Insert size distribution of mtscATAC-seq libraries. (A) Representative size distribution of accessible fragments mapping to a nuclear chromosome and **(B)** the mitochondrial DNA chromosome. The distinctive mono- and di-nucleosome peaks from ATAC-seq data appear only in the nuclear genome as mtDNA is not compacted into nucleosomes. The median length of the fragment distribution lengths are indicated.

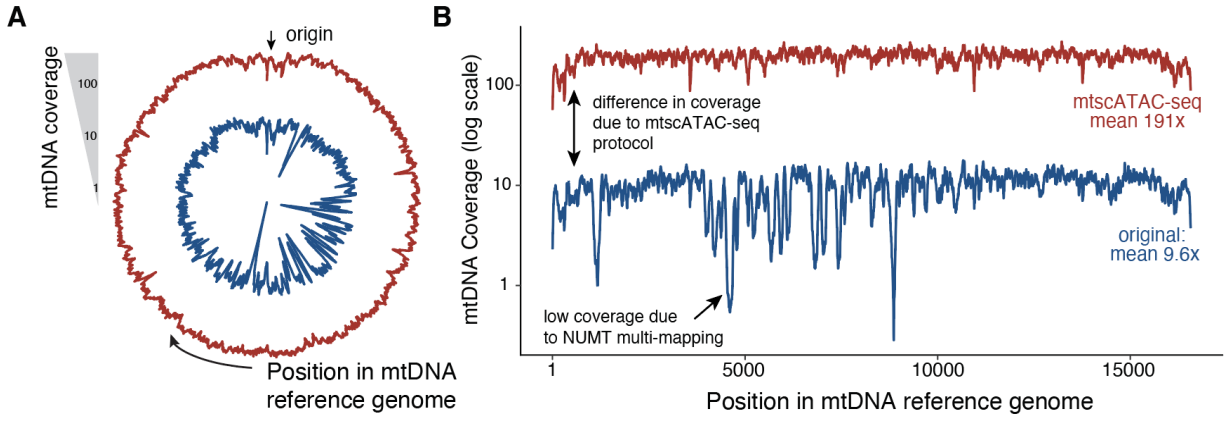

Figure 4. Mitochondrial genome coverage. (A) circular and **(B)** linear representations of the mitochondrial genome showing the differences in coverage of mtscATAC-seq (red) and scATAC-seq (blue). The same mtscATAC-seq data obtained from hematopoietic cell lines [4] is shown in both panels. Notable differences in the coverage plots are highlighted. Note that the resulting coverage is a function of cell type and sequencing depth. Primary hematopoietic cells tend to have lower mean mtDNA coverage.

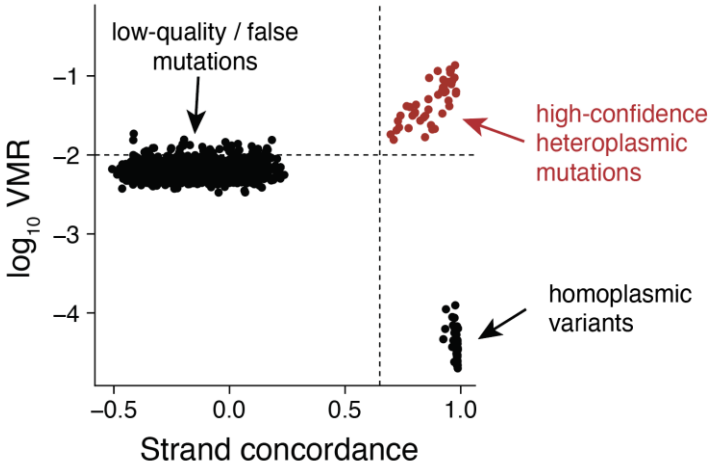

Figure 5. **Example of variant calling output from mgatk**. Each dot is a distinct mtDNA mutation. Heteroplasmic, low-quality, and homoplasmic mutations are separated by these two dimensions (x axis: heteroplasmy correlation between strands; y-axis: variance-mean-ratio (VMR) of the allele frequencies for all cells in the analysis).

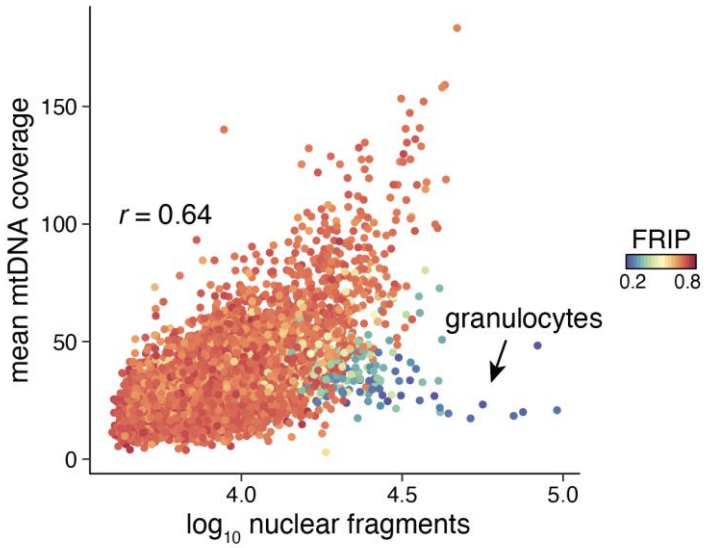

Figure 6. Quality control metrics for mtscATAC-seq data. The correlation between the log10 number of nuclear chromatin accessibility fragments and the mean mtDNA coverage per cell is shown for peripheral blood mononuclear cells. The cells with low FRIP and proportionally lower mtDNA abundance present residual granulocytes.

References

1. Nam AS, Chaligne R, Landau DA (2021) Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. Nat Rev Genet 22:3–18

2. Buenrostro JD, Giresi PG, Zaba LC, et al (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10:1213–1218

3. Corces MR, Trevino AE, Hamilton EG, et al (2017) An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods 14:959–962

4. Lareau CA, Ludwig LS, Muus C, et al (2021) Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. Nat Biotechnol 39:451–461

5. Walker MA, Lareau CA, Ludwig LS, et al (2020) Purifying Selection against Pathogenic Mitochondrial DNA in Human T Cells. New England Journal of Medicine 383:1556–1563

6. Ludwig LS, Lareau CA, Ulirsch JC, et al (2019) Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. Cell 176:1325–1339.e22

7. Penter L, Gohil SH, Lareau C, et al (2021) Longitudinal single-cell dynamics of chromatin accessibility and mitochondrial mutations in chronic lymphocytic leukemia mirror disease history. Cancer Discov. https://doi.org/10.1158/2159-8290.CD-21-0276

8. Lareau CA, Ludwig LS, Sankaran VG (2019) Longitudinal assessment of clonal mosaicism in human hematopoiesis via mitochondrial mutation tracking. Blood Adv 3:4161–4165

9. Stuart T, Srivastava A, Madad S, et al (2021) Multimodal single-cell chromatin analysis with Signac. Nat Meth. *In press*.

10. Fang R, Preissl S, Li Y, et al (2021) Comprehensive analysis of single cell ATAC-seq data with SnapATAC. Nat Commun 12:1–15

11. Granja JM, Corces MR, Pierce SE, et al (2021) ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat Genet 53:403–411

12. Mimitou EP, Lareau CA, Chen KY, et al (2021) Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. Nat Biotechnol. https://doi.org/10.1038/s41587-021-00927-2