



Mitochondrial single-cell ATAC-seq for high-throughput multi-omic detection of mitochondrial genotypes and chromatin accessibility

Caleb A. Lareau^{1,2}✉, Vincent Liu¹, Christoph Muus^{3,4}, Samantha D. Praktijn^{5,6}, Lena Nitsch^{5,6,7}, Pauline Kautz^{5,6,8}, Katalin Sandor¹, Yajie Yin¹, Jacob C. Gutierrez¹, Karin Pelka⁹, Ansuman T. Satpathy^{1,2,9}, Aviv Regev¹⁰✉, Vijay G. Sankaran^{3,11}✉ and Leif S. Ludwig^{5,6}✉

Natural sequence variation within mitochondrial DNA (mtDNA) contributes to human phenotypes and may serve as natural genetic markers in human cells for clonal and lineage tracing. We recently developed a single-cell multi-omic approach, called ‘mitochondrial single-cell assay for transposase-accessible chromatin with sequencing’ (mtscATAC-seq), enabling concomitant high-throughput mtDNA genotyping and accessible chromatin profiling. Specifically, our technique allows the mitochondrial genome-wide inference of mtDNA variant heteroplasmy along with information on cell state and accessible chromatin variation in individual cells. Leveraging somatic mtDNA mutations, our method further enables inference of clonal relationships among native ex vivo-derived human cells not amenable to genetic engineering-based clonal tracing approaches. Here, we provide a step-by-step protocol for the use of mtscATAC-seq, including various cell-processing and flow cytometry workflows, by using primary hematopoietic cells, subsequent single-cell genomic library preparation and sequencing that collectively take ~3–4 days to complete. We discuss experimental and computational data quality control metrics and considerations for the extension to other mammalian tissues. Overall, mtscATAC-seq provides a broadly applicable platform to map clonal relationships between cells in human tissues, investigate fundamental aspects of mitochondrial genetics and enable additional modes of multi-omic discovery.

Introduction

Single-cell multi-omics approaches enable the study of cellular heterogeneity and states within complex populations as well as the interdependence of features between different data modalities. Within this realm, clonal and lineage tracing approaches that aim to reconstruct the ancestral relationship between cells have advanced our understanding of fundamental processes of development, cell differentiation, lineage commitment and organ regeneration^{1,2}. In model organisms or in vitro systems, many modern lineage-tracing methodologies rely on engineered genetic labels to tag individual cells and their progeny with a heritable mark^{3,4}. With the rare exception of gene therapy trials^{5,6}, these concepts generally cannot be applied to human tissue samples to discern cell lineage, fate specification or cellular dynamics in vivo. As a consequence, most efforts to perform clonal or lineage tracing in the context of human biology and pathology rely on the occurrence of somatic DNA mutations that are propagated during cell division to daughter cells, thereby enabling the retrospective inference of cellular relationships and phylogenies. Because these somatic events occur naturally early in life and throughout adulthood, they may serve as genetic ‘barcodes’, and their

¹Department of Pathology, Stanford University, Stanford, CA, USA. ²Parker Institute for Cancer Immunotherapy, San Francisco, CA, USA. ³Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁴Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. ⁵Berlin Institute of Health at Charité-Universitätsmedizin Berlin, Berlin, Germany. ⁶Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin Institute for Medical Systems Biology (BIMSB), Berlin, Germany. ⁷Department of Biology, Chemistry, Pharmacy, Freie Universität Berlin, Berlin, Germany. ⁸Institute of Biotechnology, Technische Universität Berlin, Berlin, Germany. ⁹Gladstone-UCSF Institute of Genomic Immunology, San Francisco, CA, USA. ¹⁰Genentech, South San Francisco, CA, USA. ¹¹Division of Hematology/Oncology, Boston Children’s Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. ✉e-mail: clareau@stanford.edu; aviv.regev.sc@gmail.com; sankaran@broadinstitute.org; leif.ludwig@bih-charite.de

detection and tracking have enabled reconstruction of early human developmental events^{7,8}, of the developing brain⁹, of clonal dynamics during steady-state hematopoiesis¹⁰ and of healthy as well as cirrhotic human liver^{10,11}.

Although whole-genome sequencing-based approaches may be superior in detecting the entire breadth of somatic mutations arising in our cells, these methods are currently still relatively expensive to apply at scale, technically challenging at the single-cell level and generally not paired with a high-dimensional characterization of cellular states¹². Using an orthogonal approach, we and others have demonstrated that naturally occurring somatic mutations in the mitochondrial genome can analogously be leveraged to infer cellular relationships and that these variants may be readily detected via widely used single-cell genomics technologies. These technologies include RNA-based approaches such as Smart-seq2¹³; assay for transposase-accessible chromatin with sequencing (ATAC-seq)-based technologies, including the Fluidigm C1 system^{13,14}; and 10x Genomics droplet-based techniques^{15,16}. These sequencing-based approaches have complemented existing efforts that rely on the indirect detection of somatic mutations. For example, mutations may lead to the loss of protein expression of mtDNA-encoded genes such as cytochrome C oxidase (*MT-CO1*), which can be used to track clonal processes in situ via immunohistochemical staining^{17,18}. As such, the utility of somatic mtDNA variant-based lineage tracing has been demonstrated through multiple molecular and cellular approaches. Notably, the detection of somatic mtDNA mutations may also pose several advantages over the use of nuclear genomic variants. First, the mitochondrial genome is small (16.6 kb) but sufficiently large to provide substantial genetic diversity. Second, it has a higher mutation rate than nuclear DNA (up to 10–100×)^{19,20}. Third, mitochondrial genomes have a high copy number per cell (hundreds to thousands depending on cell type), and mutations in mtDNA often reach high levels of heteroplasmy (the proportion of mitochondrial genomes containing a specific mutation) or homoplasmy (the mutation is present in all copies of the mitochondrial genome)^{19,20}. This multi-copy number significantly facilitates the detection of variants of higher heteroplasmy at the single-cell level. Fourth, the simultaneous detection of cell state information (e.g., transcriptome and/or accessible chromatin profiling) is readily compatible with mtDNA genotyping, providing invaluable additional information to enable cell typing and genomic characterization of cellular (sub-)clones. Thereby, these single cell-based mtDNA genotyping approaches have proven effective in overcoming hurdles relating to the study of mitochondrial genetics and the presence of pathogenic mtDNA at variable heteroplasmy levels^{15,21}.

Development and overview of the protocol

To facilitate effective mitochondrial whole-genome sequencing in individual cells, we have established a high-throughput approach to genotype mtDNA in thousands of single cells alongside cell state readouts in an efficient, easy-to-use protocol using the widely available 10x Genomics platform^{15,16}. Early bulk ATAC-seq protocols used whole cells for combined lysis and tagmentation, during which Tn5 transposase tagmented into accessible nuclear chromatin as well as mtDNA derived from mitochondria. Because mtDNA was initially considered an experimental nuisance, various strategies have been devised to deplete it^{22–24}, and these mtDNA depletion strategies have been adapted by most single-cell assay for transposase-accessible chromatin with sequencing (scATAC-seq) protocols that focus on isolating nuclei devoid of mitochondria. To retain mitochondria and mtDNA within their host cell, we combine fixation with milder lysis conditions to enable combined intranuclear and mitochondrial access of the Tn5 transposase¹⁵. The workflow is readily compatible with different cell enrichment strategies such as sorting, and fixation of cells may be conducted at various stages of the process. After concomitant tagmentation of nuclear accessible chromatin and mtDNA, single-cell compartmentalization and library generation are subsequently achieved via the commercially available Chromium Next GEM Single Cell ATAC platform by 10x Genomics (Fig. 1). Library generation is followed by sequencing, data processing and analysis. Specific analysis will be context-dependent but may, for example, aim to investigate the clonal relationships among human cells on the basis of somatic mtDNA mutational profiles. The accessible chromatin data serve to identify cell types and states and further facilitate the study of genomic consequences of pathogenic mtDNA mutations affecting mitochondrial and cellular metabolism and function, leading to adaptive changes in gene regulation (Fig. 2). Because the modified cell-processing workflow further retains plasma membrane integrity, extensions of the method enable surface protein and intracellular protein staining via proteogenomics approaches, namely ATAC with select antigen profiling by sequencing and PHAGE-ATAC^{25,26}. In this protocol, we provide a detailed overview of the mitochondrial single-cell assay for

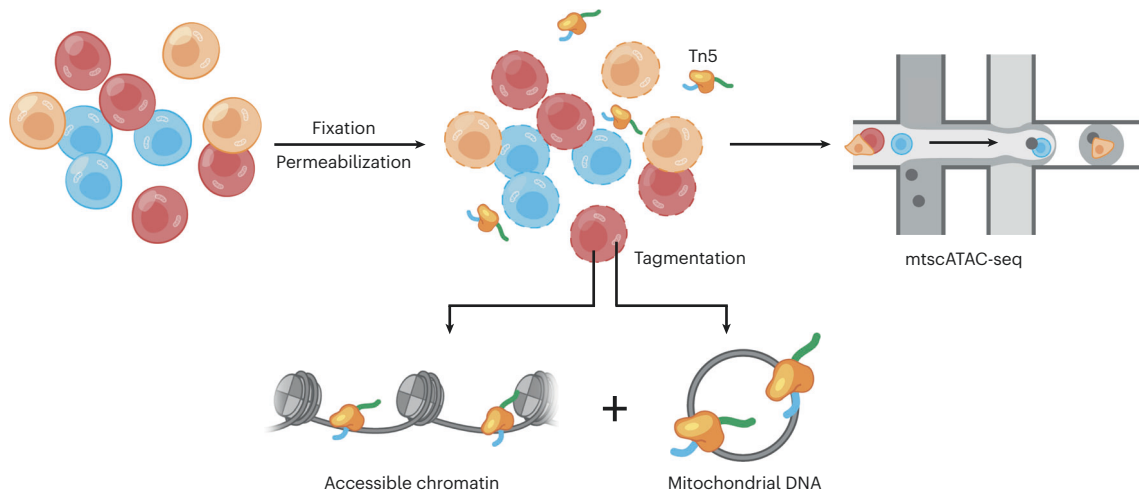


Fig. 1 | Schematic of the mtscATAC-seq experimental workflow. To retain mtDNA within their host cells, whole cells are fixed before mild lysis and permeabilization. Permeabilization enables access of the Tn5 transposase for transposition of nuclear accessible chromatin and mtDNA. Tagmented cells are encapsulated into droplets by using the Chromium Next GEM Single Cell ATAC platform by 10x Genomics to achieve single-cell compartmentalization and library generation.

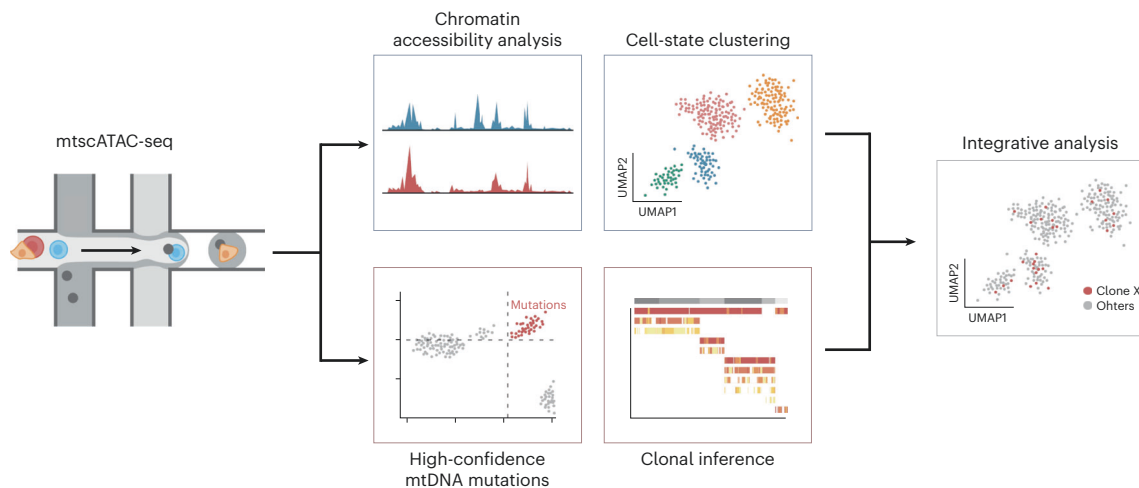


Fig. 2 | Schematic of the computational mtscATAC-seq pipeline. mtscATAC-seq library generation is followed by sequencing, data processing and analysis. Cell types and states can be identified on the basis of the chromatin accessibility data. Leveraging mtDNA reads, high-confidence (somatic) mutations and genotypes can be identified and used, for example, for clonal inferences. Both data modalities may be readily integrated for in-depth downstream analysis.

transposase-accessible chromatin with sequencing (mtscATAC-seq) technique, including best practices for cell isolation, single-cell library preparation and sequencing, as well as possible extensions of the methodology. Furthermore, we demonstrate the utility of the mitochondrial genome analysis toolkit (mgatk), a user-friendly, command-line enabled workflow that enables rapid bioinformatic processing of mtscATAC-seq data to identify somatic mtDNA mutations with a frequency as low as 0.005–0.01% from the overall (pseudobulk) population in native human samples.

Computational preprocessing

Various challenges in identifying high-confidence mtDNA variants have been well documented^{27–29}. These include variable mtDNA copy number (e.g., often >100 copies in lymphocytes and up to thousands of copies in skeletal muscle cells) and resulting heteroplasmy of mtDNA variants, because both variability in copy number and heteroplasmy are features that differ vastly in comparison to the relatively stable diploid structure in the mammalian nuclear genome. Consequently, the distribution of mtDNA-derived variants can be highly variable in single cells, confounding the assumptions made in most popular variant callers. Furthermore, many regions in the mitochondrial genome have high homology with nuclear mitochondrial DNA segments

(NUMTs). As a consequence, reads that truly originate from mtDNA are often marked as multi-mapping to NUMTs and subsequently discarded by most alignment tools using default parameters, including CellRanger-ATAC³⁰, a widely-used preprocessing pipeline for scATAC-seq data. As such, and to complement the high-throughput mtscATAC-seq assay, we have established and benchmarked computational tools that facilitate the handling of complications of mtDNA variant calling while capitalizing on the unique nature of the mtscATAC-seq data. Our protocol yields thousands of cells with high coverage per position throughout the mitochondrial genome, readily enabling the genotyping of highly heteroplasmic or homoplasmic variants. In hematopoietic cells, we have shown a mean coverage of 55× per cell and routinely threshold cells such that a minimum of 20× coverage is present; however, these values may vary across biological systems as a function of the mtDNA copy number underlying the tissue and present cell types. We have previously shown that for mtscATAC-seq, reads mapping to both the mitochondrial genome and NUMTs almost exclusively originate from mtDNA, because of high mtDNA copy number and markedly greater accessibility of mtDNA for Tn5 than nuclear DNA, ultimately estimating that only ~1 in 1,000 reads derive from NUMTs in mtscATAC-seq¹⁵. Therefore, we devised a strategy to recover multi-mapped reads in our computational framework by using a custom reference genome that masks NUMTs, providing a flexible approach that is agnostic to the desired reference genome and ensures that multi-mapped reads are appropriately assigned to the mitochondrial genome. Following sequencing read alignment, we have established a sensitive (can detect variants with allele fractions of ~1 in 1,000 in pseudobulk) yet computationally efficient (runs in 1–2 h) framework in *mgatk* for identifying and optimally calling somatic variants in the mitochondrial genome to infer clonal relationships, which is freely available at <https://github.com/caleblareau/mgatk>. Using the .bam file in CellRanger-ATAC output, *mgatk* first tallies the observed count of each nucleotide at every position in the mitochondrial genome for every cell. Importantly, *mgatk* tabulates these observed nucleotides in a strand-specific manner, because we determined that a major source of false-positive heteroplasmy (i.e., photobleaching of Illumina sequencing reads) could be detected and filtered systematically¹⁵. As a natural addition to the CellRanger output, *mgatk* generates four CSV files (one each for each of the four nucleotides) in which the four columns respectively correspond to position in the mitochondrial genome, cell barcode, count from forward strand and count from reverse strand. Furthermore, a coverage CSV file is generated in which the three columns correspond to position, cell barcode and total count across all four nucleotides and two strands. Finally, a depth table is generated to encode the average coverage across all positions for each cell barcode.

Applications of the method

mtscATAC-seq provides a unique combination of data, enabling the studies of diverse biological systems and questions. Like all scATAC-seq approaches, it enables the profiling of genome-wide chromatin accessibility from single cells derived from a complex population of cells or tissue. As such, the resulting data provide information about cell type and state and usage of gene regulatory elements and enable the inference of gene and transcription factor activities to resolve cellular heterogeneity. In addition, the whole-mtDNA genotyping provides information about mtDNA haplotypes, the presence of germline and somatic variants and their levels of heteroplasmy and allows inference of mtDNA copy number. Because somatic mtDNA variants may be stably propagated across cell divisions, we and others have used these as natural genetic barcodes to reconstruct cellular relationships and clonal dynamics across diverse biological contexts^{13–16,26}. Generally, the ability to trace the output of individual cells enables the investigation of fundamental questions, such as which (stem) cells give rise to an organ, subsequently maintain it or repair the tissue upon injury. In the context of hematopoiesis, this enabled us to demonstrate the highly polyclonal nature of blood production¹⁵. In leukemias, we were able to clonally trace leukemic cells over extended periods of time and identify clones that were sensitive or resistant to treatment¹⁶. Here, the chromatin accessibility-derived data have proven invaluable for further characterizing subclonal populations and the evolution of their genomic heterogeneity over time. This additional layer of information may further be invaluable in the study of cell fate decision making, for example, in revealing what are the genomic properties of individual (hematopoietic stem) cells that may govern production of myeloid versus lymphoid cells. In the context of mitochondrial pathologies, mtscATAC-seq has enabled the investigation of cellular and genomic consequences of pathogenic mtDNA^{15,21} and, more generally, of features of mitochondrial genetics³¹.

Comparison with other methods

mtscATAC-seq is currently unique in its ability to concomitantly profile accessible chromatin profiles alongside mitochondrial genotypes in a scalable manner. Standard scATAC-seq protocols using droplet-based techniques, including from 10x Genomics or Bio-Rad, or combinatorial indexing-based approaches tend to isolate nuclei and capture only negligible amounts of mtDNA^{30,32,33}. Furthermore, we note that additional modalities can be measured via augmenting or modifying the mtscATAC-seq workflow, including in ATAC with select antigen profiling by sequencing (adding protein measurements by using barcoded antibodies via oligonucleotide conjugation)²⁶, PHAGE-ATAC (protein detection via nanobodies displayed on phages)²⁵ or DOGMA-seq (including transcriptome and surface proteome measurements)²⁶. Although we and others have previously used the C1 Fluidigm platform, which produces comparable data to mtscATAC-seq, the limited throughput and high costs per cell are unfavorable compared to droplet-based platforms^{13,34}. Split-pool-based combinatorial indexing approaches^{32,35} may provide a viable alternative but await adaptation to co-capture mtDNA genotypes, which presents a potential challenge given the primary use of isolated nuclei in such techniques.

Because ~92% of the mitochondrial genome is transcribed³⁶, single-cell RNA-seq approaches can also be used for genotyping from mtRNA in a multi-omic assay alongside transcriptional profiling of the same cell. However, droplet-based scRNA-seq approaches tend to sequence transcripts from the 5' or 3' end, thereby limiting the coverage one can obtain from mtRNA. In our experience, ~15% of the mitochondrial transcriptome is sufficiently well covered for variant calling and analysis in standard droplet-based chemistries. Conversely, full-length RNA-seq approaches (e.g., Smart-seq2) provide uniform coverage of mtRNAs, allowing for more robust genotyping of sequenced cells, but suffer from throughput limitations¹³. Finally, a recent method called 'MAESTER'³⁷ has been introduced that amplifies mtRNAs from high-throughput scRNA-seq technologies and is a plausible alternative to mtscATAC-seq both in terms of throughput of cells and overall coverage of potential variants. All of these RNA-based genotyping methods suffer from two key limitations in that (i) they do not cover non-transcribed regions or tRNA loci (the latter of which are prevalently mutated in mitochondrialriopathies) and (ii) suffer from a high false-positive rate that we have attributed in part to the high error rate of the mitochondrial RNA polymerase¹³. We note that errors due to RNA polymerase would be stochastic in nature, rather than informative for clonal lineage-tracing applications.

Finally, although (single-cell) exome or whole-genome sequencing approaches also co-capture mtDNA sequences, cell state information is not directly integrated and thus requires orthogonal workflows¹². Together, the combination of high-throughput (via droplet microfluidics); near-uniform coverage of the mtDNA genome; amplification of DNA (rather than RNA), allowing for high-confidence variant identification; and concomitant high-dimensional cell state inference establishes mtscATAC-seq as a powerful method in the single-cell multi-omics toolkit.

Strengths and limitations of the protocol

Advantages of mtscATAC-seq include the following: (i) relatively uniform coverage across the mitochondrial genome because of the high accessibility and only modest bias of the Tn5 transposase on native mtDNA; (ii) high overall enrichment of mtDNA fragments per cell-derived sequencing library (~10–40% of fragments depending on cell type); (iii) concomitant detection of epigenomic measurements for the paired analyses of cell lineage (via mtDNA mutations) with cell state (accessible chromatin) from the same single cell; (iv) parsimonious design, in which the paired measurements of both cells are within the same sequencing library; and (v) negligible contamination of NUMT, given their relative depletion in regions of accessible chromatin¹⁵.

mtscATAC-seq is compatible with rather low input cell requirements (<100,000 cells) and enables the sequencing of 5,000–10,000 cells per 10x Genomics channel, with overloading and demultiplexing approaches being able to increase these numbers further²⁶. The combination of combinatorial indexing via Tn5-based barcoding is expected to further substantially increase cell throughput³². The number of nuclear and mtDNA fragments retrieved from these approaches may further be enhanced via Tn5 adapter replacement strategies³⁸. In the current implementation of mtscATAC-seq, a limitation is the requirement for a 10x Genomics Chromium controller and sequencing platforms, but we envision that the experimental adaptations described for mtscATAC-seq may readily translate to alternative scATAC-seq protocols.

With respect to leveraging somatic mutations for clonal or lineage tracing applications, several considerations have to be made when applying mtscATAC-seq. In principle, different types of

somatic mutations may be leveraged for such applications, and these may be detected by various genomic techniques. Although the nuclear genome provides a highly diverse reservoir for somatic mutations to occur, they are primarily present in the heterozygous state (thereby only in a single copy of DNA), which along with the large nuclear genomic DNA sequence space, make their detection technically difficult as well as costly in assays such as single-cell whole-genome sequencing. Moreover, the combination with readouts of cellular state, such as RNA-seq or ATAC-seq, is also technically challenging, does not scale well and is not readily feasible yet. Targeted mutation detection (requiring a priori knowledge) using single-cell RNA-seq approaches has been described but is often limited to mutations in (highly) expressed genes and/or near the 3' end of a transcript^{39,40}. By contrast, mtscATAC-seq enables single-cell whole-mitochondrial genome sequencing with relatively uniform coverage, which is further readily compatible with accessible chromatin profiling to enable cell typing and provide readouts of cellular state. Furthermore, in a variant assay, our application of mtscATAC-seq modifications to the 10x Genomics Multiome kit yielded DOGMA-seq²⁶, which enabled the additional profiling of transcriptome and surface markers (using an analogous strategy to CITE-seq, cellular indexing of transcriptomes and epitopes by sequencing⁴¹), yielding a four-modality readout for single cells.

One limitation of the approach is that it relies on natural mtDNA sequence variation that the experimental biologist cannot directly modulate, alongside the dynamics of mitochondrial genetics still being less well understood than their nuclear counterpart. Moreover, mainly because of a lack of data availability on a single-cell level, we currently are less aware of mtDNA mutational diversity at a younger age (e.g., pediatric samples), although bulk genomic approaches in specific contexts suggest somatic mtDNA variation may be already substantial⁴². Nevertheless, the reconstruction of early human developmental events may be more difficult as opposed to investigating the output of adult hematopoietic stem cells, which, as we demonstrated, show substantial mtDNA diversity in humans >20 years of age^{13,15,26}. In the context of malignancies, we have demonstrated the utility of mtscATAC-seq to co-detect chromosomal copy number variants, which together with somatic mtDNA mutations enable the joint reconstruction of clonal and lineage relationships based on information derived from both genomes^{15,16}.

Expertise needed to implement the protocol

Experience in cell biology will be an asset for processing cells of diverse cellular origins. We have extensively applied mtscATAC-seq to human blood and immune cells (e.g., peripheral blood, bone marrow, lymph nodes and tumor-derived cells), and expertise in handling these materials and enriching live cells (e.g., via cell sorting or magnetic enrichment protocols) will be helpful. For solid tissues, additional steps to obtain high-quality single-cell suspensions will be required and may have to be determined empirically. Because mtscATAC-seq involves fixation, particularly sensitive cells may be fixed before additional and potentially stressful cell-processing steps such as sorting. Expertise in single-cell genomics will be helpful, because library preparation strategies often follow similar principles, while obtaining a high-quality single-cell suspension is a prerequisite to successfully execute the protocol. Experience with the 10x Genomics platform will be highly beneficial, and access to a sequencing platform will be required (e.g., Illumina NextSeq 500/550 or NovaSeq 6000). As opposed to scRNA-seq-based technologies, we have found mtscATAC-seq very straightforward to adopt, given the smaller number of steps required to obtain libraries. For the processing from raw sequencing reads to count tables, existing software packages are readily accessible, and basic command line skills are needed. Additional expertise in computational biology and data science are required to perform more in-depth data analyses.

Experimental design

Sample preparation

It is important to work quickly and wear personal protective equipment when working with human cells or patient material. Cell preparation is exemplified in the procedure for flow cytometry-sorted human peripheral blood mononuclear cells (PBMCs). When starting with whole blood, perform a density gradient centrifugation to isolate PBMCs, for example, by using Ficoll-Paque-based approaches such as SepMate-50 (STEMCELL Technologies, cat. no. 85450) with Ficoll-Paque PLUS or Ficoll-Paque PREMIUM centrifugation media (Cytiva, cat. nos. 17-1440-02, 17-5442-02 and 17-5446-02). In addition, red blood cell lysis may be considered (e.g., by using Gibco ACK (ammonium-chloride-potassium) lysing buffer (Thermo Fisher Scientific, cat. no. A1049201)). This is

Box 1 | Solid tissue dissociation

Here, we describe a procedure for dissociation of fresh human ovarian and endometrial tumors and matching normal tissue specimens.

Procedure.

- 1 Mince specimens on ice to pieces of <math><1\text{ mm}^3</math> and transfer to 5 ml of digestion medium containing DNase I (100 $\mu\text{g}/\text{ml}$) and collagenase P (2 mg/ml) in Advanced DMEM/F-12.
- 2 Perform dissociations in C-tubes by using the gentleMACS Octo Dissociator system at 37 °C at 20 rpm for 20 min. The composition of the digestion medium and the program might require additional optimization depending on the histology type of the tissue.
- 3 After digestion, filter the cell suspension through a 70- μm filter and wash the strainer with an additional 10 ml of DMEM/F-12 and centrifuge the sample at 400g at 4 °C for 5 min. Residual undigested tissue may be further digested for an additional 20-min incubation as described above with additional medium.
- 4 After centrifugation, discard the supernatant, resuspend the pellet in 500 μl of ACK red blood cell lysis buffer and incubate for 1 min on ice, followed by the addition of two volumes of ice-cold PBS.
- 5 Immediately assess cell count and viability by trypan blue staining by using a Countess II FL automated cell counter or a manual hemocytometer.
- 6 Immediately process the cells for cell sorting or cryopreserve the cells for future use.

We note that for solid human lung, ovarian and endometrial tumors, as well as healthy adjacent tissues, we were able to cryopreserve dissociated single cells, which upon thawing, were processed analogously to what we describe for PBMCs. However, we currently do not recommend mtscATAC-seq on flash-frozen tissues. Given the loss of membrane integrity upon rapid freezing and the relatively small size of mitochondria, we anticipate significant cross-cellular exchange of mitochondria and mtDNA variants. For other solid tissues, we recommend optimizing single-cell dissociation protocols inspired by prior works^{43–45} as needed for the specific tissue of interest.

important to deplete undesired cell populations, including erythrocytes and granulocytes, that do not yield genomic data or might interfere with downstream analysis. We recommend the use of swinging bucket centrifuges for all downstream steps to maximize cell recovery. After sorting, we recommend using DNA LoBind tubes (1.5 or 2 ml) for cell-processing steps and the use of low-retention tips. The protocol can be readily adapted to other suspension cell types and cell lines but may require additional sample-specific optimization.

Isolation of single cells

mtscATAC-seq requires the generation of a single-cell suspension. We routinely process suspension cells such as primary human hematopoietic cells, including peripheral blood and bone marrow-derived mononuclear cells and hematopoietic and lymphoblastoid cell lines, that do not require dissociation other than the removal of clumps via gentle pipetting or application of cell strainers (e.g., 40–70 μm). Solid tissue specimens may require more extensive and tissue-specific processing and dissociation protocols that may need to be optimized to obtain a single-cell suspension as described^{43–45}. In Box 1, we provide an exemplary workflow, which we have successfully tested on human ovarian and endometrial cancer specimens. High viability (>90%) and depletion of granulocytes (e.g., via density gradient centrifugation (Ficoll) and/or sorting) are essential for high-quality data. We recommend obtaining >100,000 viable single cells for cell processing and mtscATAC-seq library generation and obtaining mtscATAC-seq data for 5,000–10,000 cells. Cells should be processed and stored on ice (4 °C) unless indicated otherwise, and the protocol should be followed without extensive delays. The total number of cells to be profiled will ultimately depend on the context and question of interest. Profiling of a few thousand cells will enable identification of major cell types and states on the basis of differences in their chromatin accessibility landscapes, whereas the annotation of rare cell populations will be facilitated with increasing number of cells sequenced. In an analogous manner, for mtDNA-based clonal tracing, the number of cells to be sequenced will be a function of the ‘clonality’ of the sample. Specifically, dominant clones are more readily detected, whereas the detection of lower-frequency clones ultimately improves with an increasing number of cells profiled. On the basis of our previous experiences, as few as ~1,000 cells can highlight subclonal structures in human malignancies^{15,16}. For steady-state hematopoiesis, ~10,000 cells have provided initial informative insights¹⁵.

FACS

Flow cytometry is used to generate single-cell suspensions and enrich specific cell populations of interest by using fluorophore-conjugated antibodies targeting surface antigens. We exclude granulocytes, given their very accessible and non-condensed chromatin structure, which depending on their proportion in the population of interest, can negatively affect scATAC-seq data quality^{46,47}. Minimally, we use a combination of forward scatter, side scatter, a live or dead cell stain (e.g., SYTOX Blue, Ghost Dye Violet or Zombie Violet) and an anti-human CD66b antibody to obtain live cells and exclude granulocytes (Fig. 3). Because cells are fixed for mtscATAC-seq by using formaldehyde, the

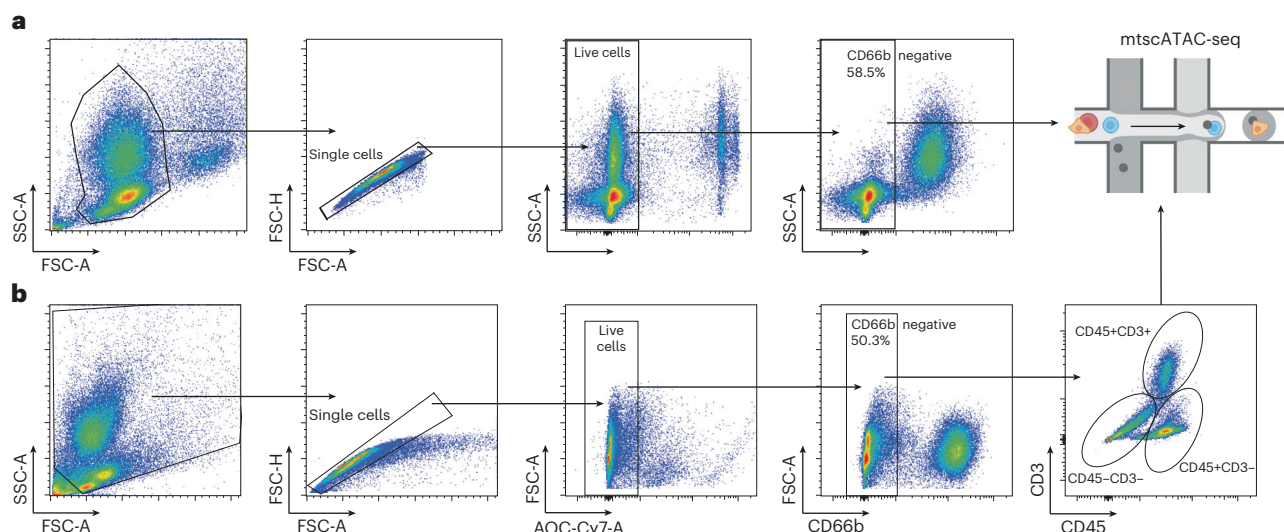


Fig. 3 | Flow cytometry cell sorting strategies. **a**, Human PBMCs were isolated from whole blood by using a density gradient followed by red blood cell lysis and stained with a live or dead cell marker (SYTOX Blue) and anti-CD66b. The gating strategy is shown and aims to exclude cell doublets (forward scatter area (FSC-A):side scatter area (SSC-A)), dead cells (SYTOX Blue:SSC-A) and granulocytes (CD66b:SSC-A). Note that when using Ficoll density-gradient centrifugation, most granulocytes will have been depleted. **b**, Gating strategy to obtain various (non)-immune cells from a human ovarian tumor. Additional surface markers CD45 and CD3 were stained to enrich indicated cell populations via sorting for the downstream mtscATAC-seq workflow. FSC-H, forward scatter height.

fixation may occur before or after cell sorting, with no discernible difference in genomics data quality. We have worked successfully with several different types of sorters such as BD FACSAria and Sony SH800 devices.

Generation of mtscATAC-seq libraries

For the generation of mtscATAC-seq libraries we have adapted the 10x Genomics scATAC-seq platform and we have successfully worked with the v1 and the NextGEM v1.1 kits. We outline a modified version of the “Nuclei Isolation for Single Cell ATAC Sequencing” (CG000169 Rev D) user guide, where we fix and permeabilize cells to retain mitochondria and mtDNA within their host cell, which are otherwise depleted using the standard protocol. For the library preparation we follow the “Chromium Next GEM Single Cell ATAC Reagent Kits v1.1” (CG000209 Rev F) user guide with only minor modifications as described and highlighted below and otherwise refer the reader to the original and highly detailed workflow by 10x Genomics.

Quality control of libraries before sequencing

After library preparation, the yield of the mtscATAC-seq libraries is quantified by using a Qubit high-sensitivity double-stranded DNA (dsDNA) kit, followed by resolving the size distribution by using an Agilent high-sensitivity DNA kit and BioAnalyzer-based electrophoresis. Together, these provide a first indication of the success of the protocol and, in our hands, enable reliable quantification of libraries before sequencing without qPCR-based quantification.

Sequencing and depth

To obtain high coverage of the mitochondrial genome, we recommend paired-end sequencing and 150 cycle kits or longer, specifying at least 70 cycles for both Reads 1 and 2. Specifying these read lengths ensures full coverage of most fragments for optimal sensitivity of detection of mtDNA variants from mtscATAC-seq data. For sequencing, we have routinely worked with the Illumina NextSeq 550 and NovaSeq 6000 platforms. We have found that any of the major Illumina sequencing platforms are appropriate for high-confidence heteroplasmy estimation with mtscATAC-seq, given their overall low error rates of $\sim 0.1\text{--}0.6\%$ ⁴⁸. Notably, our prior work has shown that false-positive heteroplasmy often derives from stretches of homopolymer sequences in the mtDNA genome (e.g., poly-G sequences) that can lead to errant heteroplasmy calling^{15,16}. On the basis of our careful examination of these data, we developed a sensitive variant calling platform (within the mgatk package; discussed at Step 46) that accounts for potential systematic errors by considering the strand

orientation of heteroplasmy when identifying high-confidence variants. Conceptually, heteroplasmy derived from technical errors from sequencing is more likely to (i) have lower variance computed over cells compared to true clonal variants and (ii) be detected disproportionately from one sequencing direction relative to the other. By quantifying these effects and thresholding appropriately, we have been successful in minimizing the effects of sequencing errors in our analyses even on the most error-prone Illumina sequencers. Thus, users can proceed with confidence on any current Illumina sequencing platform.

For library loading and pooling, we follow the recommendations provided by 10x Genomics. We generally aim to allocate between 30,000 and 35,000 paired-end reads (read pairs consisting of Read 1 and Read 2) per cell in a library, typically resulting in a sequencing saturation depth of >50%. In our experience, high-quality libraries yield >20× mitochondrial genome coverage after removal of PCR duplicated reads to enable confident mtDNA mutation calling for primary human hematopoietic cells. We emphasize that deeper sequencing may further improve mitochondrial genome coverage and the detection of low heteroplasmic mtDNA mutations as well as recovery of unique nuclear accessible chromatin fragments. The relative utility of additional sequencing depth may be assessed from the saturation metrics. Moreover, mtDNA content may vary from tissue and cell type, requiring adjustment of sequencing depth.

Materials

Biological materials

- *Cells from primary tissues or cell lines.* We have successfully used diverse types of cells, including human and murine PBMCs, human and murine bone marrow mononuclear cells, cells from human colorectal cancer and ovarian cancer and human lung, human thymus and other cell lines such as HEK293T (RRID: [CVCL_0063](#)), TF1 (RRID: [CVCL_0559](#)) and lymphoblastoid cell lines (e.g., GM11906 (RRID: [CVCL_IN29](#))). **!CAUTION** All applications involving primary human cells and tissues from healthy donors or patients should be performed in accordance with relevant guidelines and regulations, including obtaining informed consent. Implement additional cautionary measures when handling potentially infectious material. **!CAUTION** If using cell lines, they should be regularly checked to ensure that they are authentic and are not infected with mycoplasma.

Reagents

- RPMI 1640 (GIBCO, cat. no. 11875-119)
- FBS (Atlanta Biologicals, cat. no. S11150)
- PBS (GIBCO, cat. no. 10010-023)
- Nuclease-free water (Thermo Fisher Scientific, cat. no. AM9937)
- Tris-HCl (pH 7.4, 1 M) (Sigma-Aldrich, cat. no. T2194)
- NaCl (5 M) (Sigma-Aldrich, cat. no. 59222C)
- MgCl₂ solution (1 M) (Sigma-Aldrich, cat. no. M1028)
- NP-40 Surfact-Amps detergent solution (Thermo Fisher Scientific, cat. no. 28342) **!CAUTION** NP-40 can cause serious eye damage and is harmful to aquatic life, with long-lasting effects. Wear protective gloves, protective clothing, eye protection and face protection and avoid release to the environment.
- IGEPAL CA-630 (Sigma, cat. no. 9002-93-1)
- Collagenase P, from *Clostridium histolyticum* (Sigma-Aldrich, cat. no. 11249002001)
- DNase I, from bovine pancreas (Sigma-Aldrich, cat. no. 11284932001)
- Advanced DMEM/F-12 (Thermo Fisher Scientific, cat. no. 12-634-028)
- Gibco ACK (ammonium-chloride-potassium) lysing buffer (Thermo Fisher Scientific, cat. no. A1049201)
- BSA, 10% (wt/vol) (Miltenyi Biotech, cat. no. 130-091-376)
- EB buffer (10 mM Tris-HCl, pH 8.5) (Qiagen, cat. no. 19086)
- Tween 20, 10% (wt/vol) (Bio-Rad, cat. no. 1610781)
- Human TruStain FcX (Fc receptor blocking solution; BioLegend, cat. no. 422301)
- Anti-human CD66b-PE antibody (BioLegend, Clone G10F5, cat. no. 305105; RRID: [AB_10550093](#))
- SYTOX Blue (Thermo Fisher Scientific, cat. no. S34857) **▲CRITICAL** Store aliquots at -20°C protected from light.
- Ghost Dye Violet 510 (Tonbo Biosciences, cat. no. 13-0870-T100) **▲CRITICAL** Store aliquots at -20°C protected from light.
- Zombie Violet (BioLegend, cat. no. 423113) **▲CRITICAL** Store aliquots at -20°C protected from light.

- Formaldehyde, 16% (wt/vol) (Thermo Fisher Scientific, cat. no. 28906) **! CAUTION** Formaldehyde is harmful if swallowed or inhaled or when in contact with skin and causes skin irritation. It may cause an allergic skin reaction or respiratory irritation. It causes serious eye irritation. Formaldehyde is suspected of causing genetic defects and may cause cancer. Obtain special instructions before use and avoid breathing dust, fumes, gas, mist, vapors or spray. Wear protective gloves, protective clothing, eye protection and face protection and wash your hands thoroughly after handling. Do not eat, drink or smoke when using. Store in a well-ventilated place in a tightly closed container.
- Glycine (2.5 M; Boston Bioproducts, cat. no. C43755)
- Trypan blue solution (Thermo Fisher Scientific, cat. no. 15250061) **! CAUTION** Trypan blue may cause cancer and is suspected of damaging fertility or the unborn child. Obtain special instructions before use and wear protective gloves, protective clothing, eye protection and face protection.
- Ethanol, Pure (200 proof, anhydrous) (Millipore Sigma, cat. no. E7023-500ML) **! CAUTION** Pure ethanol is toxic and highly flammable. Store in a fireproof cabinet and use it with caution.
- Chromium Next GEM single cell ATAC library & gel bead kit, 16 or 4 reactions (10x Genomics, cat. no. PN-1000175 or PN-1000176). The kits include:
 - 20× nuclei buffer (10x Genomics, cat. no. PN-2000207)
 - Amp mix (10x Genomics, cat. no. PN-2000047 or PN-200103) **! CAUTION** Amp mix causes skin and eye irritation. Wear protective gloves and wash thoroughly after handling.
 - ATAC buffer B (10x Genomics, cat. no. PN-2000193) **! CAUTION** ATAC buffer B is a combustible liquid and causes serious eye irritation. It may cause cancer and may damage fertility or the unborn child. Obtain special instructions before use and keep away from flames and hot surfaces. No smoking. Wear protective gloves, protective clothing, eye protection and face protection and wash hands thoroughly after handling. Store in a well-ventilated place and keep cool. Store locked up.
 - ATAC enzyme (10x Genomics, cat. no. PN-2000123 or PN-2000138)
 - Barcoding enzyme (10x Genomics, cat. no. PN-2000125 or PN-2000139)
 - Barcoding reagent B (10x Genomics, cat. no. PN-2000194) **! CAUTION** Barcoding reagent B causes eye irritation. Wash hands thoroughly after handling.
 - Reducing agent B (10x Genomics, cat. no. PN-2000087) **! CAUTION** Reducing agent B is harmful if swallowed. It causes skin irritation and serious eye irritation. Wear protective gloves, eye protection and face protection and wash hands thoroughly after handling. Do not eat, drink or smoke when using this product.
 - Cleanup buffer (10x Genomics, cat. no. PN-2000088) **! CAUTION** Cleanup buffer is toxic if swallowed or upon contact with skin. It is harmful if inhaled. Avoid breathing dust, fumes, gas, mist, vapors or spray and wash hands thoroughly after handling. Do not eat, drink or smoke when using this product and use only in a well-ventilated area. Wear protective gloves and protective clothing. Store locked up.
 - SI-PCR primer B (10x Genomics, cat. no. PN-2000128)
 - Single cell ATAC gel beads v1.1 (10x Genomics, cat. no. PN-2000210)
 - Dynabeads MyOne SILANE (Thermo Fisher Scientific, cat. no. PN-2000048)
- Chromium Next GEM Chip H single cell kit, 48 or 16 reactions (10x Genomics, cat. nos. PN-1000161 and PN-1000162). The kits include:
 - Partitioning oil (10x Genomics, cat. no. PN-2000190) **! CAUTION** Partitioning oil causes eye and skin irritation and may be harmful to aquatic life, with long-lasting effects. Wear protective gloves, protective clothing, eye protection and face protection and avoid release to the environment. Do not breathe in dust or fumes.
 - Recovery agent (10x Genomics, cat. no. PN-220016) **! CAUTION** Recovery agent is a combustible liquid and causes serious eye irritation and skin irritation. It might cause respiratory irritation. Keep away from hot flames and surfaces and avoid breathing dust or fumes. Wear protective gloves and eye protection and wash hands thoroughly after handling.
 - Chromium Next GEM Chip H (10x Genomics, cat. no. PN-2000180)
 - Gasket, 2- or 6-pack (10x Genomics, cat. no. PN-370017 or PN-3000072)
- Single Index Kit N, Set A, 96 reactions (10x Genomics, cat. no. PN-1000212) containing the Single Index Plate N Set A (10x Genomics, cat. no. PN-3000427)
- SPRIselect reagent kit (Beckman Coulter, cat. no. B23318)
- Qubit dsDNA high-sensitivity assay kit (Thermo Fisher Scientific, cat. no. Q32854)
- Bioanalyzer high-sensitivity DNA analysis kit (Agilent, cat. no. 5067-4626)
- NextSeq 500/550 or NovaSeq 6000 reagent kits (150–200 cycles; Illumina)

Equipment

- Conical tubes, 15 ml (Falcon, cat. no. 352196)
- Conical tubes, 50 ml (Falcon, cat. no. 352070)
- DNA LoBind tubes, 1.5 ml (Eppendorf, cat. no. 022431021)
- DNA LoBind tubes, 2.0 ml (Eppendorf, cat. no. 022431048)
- gentleMACS C tubes (Miltenyi Biotech, cat. no. 130-093-237)
- Falcon round-bottom polystyrene test tubes with cell strainer snap cap for FACS, 5 ml (Thermo Fisher Scientific, cat. no. 10585801)
- Flowmi cell strainer, 40 μ m (Bel-Art, cat. no. H13680-0040)
- PCR 8-tube strips (USA Scientific, cat. no. 1402-4700)
- Low-retention tips LTS 20UL filter RT-L10FLR (Rainin, cat. no. 30389226)
- Low-retention tips LTS 200UL filter RT-L200FLR (Rainin, cat. no. 30389240)
- Low-retention tips LTS 1ML filter RT-L1000FLR (Rainin, cat. no. 30389213)
- Microcentrifuge (VWR, cat. no. 521-2319)
- Minicentrifuge 5425R (Eppendorf, cat. no. 5406000518)
- Centrifuge 5810R (Eppendorf, cat. no. 5811000015)
- gentleMACS Octo dissociator (Miltenyi Biotech, cat. no. 130-095-937)
- Countess II FL automated cell counter (Thermo Fisher Scientific, cat. no. A27974)
- Hemacytometer, enhanced Neubauer rulings (Bright-Line, cat. no. Z359629)
- Bright-field microscope (Leica DM IL LED)
- BD FACS Aria sorter (BD Biosciences)
- C1000 touch thermal cycler with 96-deep well reaction module (Bio-Rad, cat. no. 1851197)
- 10x Chromium Controller (10x Genomics, cat. no. 1000202 or 1000204)
- 10x magnetic separator (10x Genomics, cat. no. 12050 or 230003)
- 10x vortex adapter (10x Genomics, cat. no. 120251 or 330002)
- Chromium Next GEM secondary holder (10x Genomics, cat. no. 10000195 / cat. no. 3000332)
- Thermomixer C (Eppendorf, cat. no. 5382000015)
- Bioanalyzer 2100 system (Agilent, cat. no. G2939BA)
- NextSeq 500/550 sequencer (Illumina) or NovaSeq 6000 sequencer (Illumina)

Software

For comprehensive analysis of next-generation sequencing data from mtscATAC-seq libraries, we note three components of required software: CellRanger-ATAC³⁰, mgatk¹⁵, and Signac⁴⁹. For optimal mitochondrial genome coverage, we recommend using bedtools⁵⁰ to hard-mask NUMTs in the reference genome. Other optional software for generating custom blacklists is noted below.

- CellRanger-ATAC, v2.0+ (<https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/installation>)
- mgatk, v0.6.4+ (<https://github.com/caleblareau/mgatk>)
- Signac, v1.1+ (<https://satijalab.org/signac/>)
- bedtools, v2.25+ (<https://bedtools.readthedocs.io/en/latest/content/installation.html>)
- (Optional) ART read simulator, v2016.06.05+ (<https://www.niehs.nih.gov/research/resources/software/biostatistics/art/index.cfm>)
- (Optional) Bowtie2, v2.4.0+ (<http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml>)

Reagent setup

Cell-thawing medium

Cell-thawing medium consists of RPMI + 10% (vol/vol) FBS, 0.45- μ m filtered; store at 4 °C for \leq 2 months and pre-warm to 37 °C before use.

FACS buffer

FACS buffer consists of 1% (vol/vol) FBS in PBS, 0.45- μ m filtered; store at 4 °C for \leq 2 months.

2 \times resuspension buffer (2 \times RSB)

2 \times RSB consists of 20 mM Tris-HCl pH 7.4, 20 mM NaCl and 6 mM MgCl₂, 0.45- μ m filtered. Store at 4 °C for \leq 2 months or at -20 °C until further use.

Lysis buffer

Lysis buffer consists of 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% (vol/vol) NP-40 and 1% (vol/vol) BSA. To obtain the 1× working solution, dilute 2× RSB with nuclease-free water and supplement with 1% (vol/vol) BSA and 0.1% (vol/vol) NP-40. The solution should be made fresh and kept on ice until used. IGEPAL CA-630 at 0.1% (vol/vol) may be used to substitute for NP-40.

Wash buffer

Wash buffer consists of 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂ and 1% (vol/vol) BSA. To obtain the 1× working solution, dilute 2× RSB with nuclease-free water and supplement with 1% (vol/vol) BSA. The solution should be made fresh and kept on ice until used. **▲ CRITICAL** The lysis and wash buffer preparation differ from the one recommended in the ‘Nuclei Isolation for Single Cell ATAC Sequencing’ user guide from 10x Genomics (CG000169 Rev D). We omitted Tween 20, which depletes mitochondria and mtDNA.

1× nuclei buffer

1× nuclei buffer consists of 20× nuclei buffer diluted 1:20 (vol/vol) in nuclease-free water. The solution should be made fresh and kept on ice until used.

Elution solution I

For 100 µl, mix 98 µl of EB buffer, 1 µl of 10% (vol/vol) Tween 20 and 1 µl of reducing agent B. The solution should be made fresh and kept at room temperature (20–22 °C) until used.

Procedure**Cell thawing ● Timing 25–35 min**

- 1 For cryopreserved cells, remove the cryovial from liquid nitrogen storage and quickly thaw in a water bath at 37 °C until a small ice crystal remains. Remove from the water bath, spray down with ethanol, transfer to a clean tissue culture hood and continue with Step 2. For freshly isolated single-cell suspensions, directly continue with cell staining (Step 9).
- 2 Assuming a volume of 1 ml of cryopreserved cells (cell numbers may vary), use a P1000 pipette to carefully transfer the thawed cells to a 50-ml Falcon tube. Rinse the cryovial with 1 ml of prewarmed cell-thawing medium and add the rinse dropwise to the 50-ml conical tube while gently shaking the tube.
- 3 Sequentially dilute cells in the 50-ml conical tube by incremental 1:1 volume additions of thawing medium at a speed of 1 ml/3–5 s to the tube and swirl. Repeat for a total of five times with ~1-min waits between additions. Starting with 1 ml, add 2 ml the next time, followed by 4 ml, 8 ml and 16 ml. In the final step, top up to 50 ml.
▲ CRITICAL STEP Serial dilution helps to avoid osmotic shock of cells when resuspending in a protein-containing medium.
- 4 Centrifuge at 400g for 5 min at room temperature.
- 5 Carefully remove the supernatant, carefully flick the cell pellet to loosen it and resuspend the cells in 25 ml of cell-thawing medium by gentle pipetting.
- 6 Determine the cell concentration and viability of the suspension by using a Countess II FL automated cell counter or a manual hemocytometer. For primary human PBMCs, we typically observe high viability (>90%) after proper cryopreservation.
? TROUBLESHOOTING
- 7 Centrifuge at 400g for 5 min at room temperature.
- 8 Resuspend the cells in FACS buffer and transfer to a 5-ml FACS tube with cell strainer snap cap to generate a single-cell suspension at a concentration of $\sim 1 \times 10^6$ cells per ml.

Cell staining for flow cytometry ● Timing 35–40 min

- ▲ CRITICAL** We exemplify sorting of live PBMCs by using a live or dead cell stain and an hCD66b antibody to exclude residual granulocytes. Note that any type of antibody combination amenable to flow cytometry-based enrichment or exclusion may be used depending on the context of interest.
- 9 Thaw SYTOX Blue and dilute 1:1,000 in ice-cold FACS buffer (FACS-SYTOX). Alternatively, Ghost Dye Violet may be used according to the manufacturer’s instructions.
 - 10 Transfer $\sim 1 \times 10^6$ cells to a 1.5-ml microcentrifuge tube.
▲ CRITICAL STEP Antibody volumes may need to be adjusted in the case of staining a higher number of cells.

- 11 Add 1 ml of FACS buffer to wash cells and spin for 5 min at 400g at 4 °C.
- 12 Discard the supernatant and resuspend the cells in 100 µl of FACS buffer.
- 13 Add 5 µl of human TruStain FcX for blocking, mix and incubate for 5 min on ice.
- 14 Add 5 µl of CD66b-PE, mix and incubate for 15 min on ice, protected from light.
- 15 Wash the cells two times with 1 ml of FACS buffer and spin for 5 min at 400g at 4 °C.
- 16 Resuspend the cells in 400–600 µl of FACS-SYTOX and mix. Keep at 4 °C and protected from light until sorting.

Cell sorting ● Timing 30 min–1 h (or longer depending on the frequency of the target cell population)

- 17 Sort ~100,000 to 1×10^6 SYTOX Blue and CD66b-PE double-negative cells (Fig. 3) into a 15-ml Falcon tube with 7 ml of chilled cell-thawing medium. Note that the number of cells to sort depends on the number of cells one intends to sequence. We typically sort 200,000–400,000 cells when running one to two 10x Genomics channels (yielding 5,000–10,000 cells of data per channel). In all subsequent steps after centrifugation to pellet cells, we recommend gently flicking the cell pellet (at the bottom of the tube) to gently loosen the cells before careful resuspension via pipetting. Extensive pipetting of cells should be avoided.

? TROUBLESHOOTING

- 18 Top up the cell suspension to a volume of 14 ml with chilled PBS and invert to mix.
- 19 Centrifuge the cells for 5 min at 400g at 4 °C.
- 20 Remove the supernatant without disrupting the pellet.
- 21 Resuspend the cells in 450 µl of room temperature PBS, transfer to a 1.5-ml DNA Lobind tube and either determine the cell concentration and viability after resuspension to validate successful recovery of cells after sorting or immediately proceed with cell fixation.

Cell fixation ● Timing 25 min

- 22 Fix cells in 1% (vol/vol) formaldehyde for 10 min at room temperature (e.g., by using a 16% (wt/vol) stock concentration, add 30 µl to the cell suspension), swirl the tube and invert occasionally. Box 2 outlines considerations for fixing cells before sorting when this might be desirable. Box 3 discusses effects of fixation on accessible chromatin data quality and de-cross-linking during the GEM incubation (Step 33).
- 23 Quench with glycine solution to a final concentration of 0.125 M (e.g., by using a stock of 2.5 M glycine, add 25 µl to the solution).
- 24 Wash cells two times in 1 ml of PBS or FACS buffer via centrifugation at 400g for 5 min at 4 °C.
- 25 Remove the supernatant and immediately proceed with cell lysis.

Box 2 | Fixation of cells before cell sorting

For certain workflows, including working with fragile and shear-stress-sensitive cells, it might be preferable to fix cells before sorting. In this instance, we recommend using the fixable dye Zombie Violet (Biolegend, cat. no. 423113) for live or dead cell staining (perform Steps 22–24 before Step 11). Note that the Zombie dye staining should be conducted in the absence of protein, and as such, cells should be resuspended in 100 µl of PBS, followed by washing, antibody staining, fixation and cell sorting into FACS buffer or cell thawing medium. We recommend testing the compatibility of antibodies with fixation beforehand. Note that forward and side scatter properties as well as antibody staining intensities of fixed cells may differ compared to native cells.

! CAUTION With the current fixation conditions, we do not suggest extensive storing of cells similar to what is being done for some flow cytometry workflows, because this significantly diminished scATAC-seq data quality (e.g., transcription start site (TSS) scores <2). As such, we recommend immediately proceeding with cell lysis and permeabilization (Step 26).

Box 3 | Fixation and de-cross-linking

Fixation of cells presents a primary feature of mtscATAC-seq for effective mtDNA genotyping. Although fixation may negatively affect nuclear chromatin fragment complexity, we have found 1% (vol/vol) formaldehyde to be a suitable concentration to balance mtDNA genotyping and accessible chromatin profiling efforts. We tested extended incubation of GEMs (30 min to 12 h) at 60 °C to further facilitate de-cross-linking before the first 72 °C elongation PCR step. This did not improve library complexity or other quality-control metrics, and we recommend using the PCR conditions specified in the 10x Genomics scATAC-seq protocol.

Table 1 | Expected cell recovery as a function of cell stock concentration

Number of cells to be recovered	Stock concentration (cells/ μ l)
3,000	925–2,300
4,000	1,230–3,075
5,000	1,540–3,850
6,000	1,850–4,600
7,000	2,150–5,400
8,000	2,460–6,150
9,000	2,770–6,900
10,000	3,080–7,700

Cell lysis and permeabilization ● **Timing 60 min**

▲ **CRITICAL** The following steps are performed as outlined in the ‘Nuclei Isolation for Single Cell ATAC Sequencing’ protocol from 10x Genomics (CG000169 Rev D) with the following modifications: lysis and wash buffer preparation differs from the original protocol, because we found that Tween 20 significantly decreases mtDNA yield and mitochondrial genome coverage.

- 26 Add 100 μ l of chilled lysis buffer to the loosened cell pellet and pipette gently up and down three times. Recommended cell concentrations for lysis range from 100,000–300,000 cells per 100 μ l of lysis buffer.
- 27 Incubate the cells on ice (3 min for primary cells and 5 min for cell lines). The lysis time may need to be optimized depending on the organ and cell type of interest.
- 28 Wash cells one time by gently adding 1 ml of chilled wash buffer and invert the tube multiple times.
- 29 Centrifuge at 500g for 5 min at 4 °C.
- 30 Discard the supernatant gently without disrupting the pellet.
- 31 Resuspend loosened cells in chilled 1 \times diluted nuclei buffer at the desired concentration depending on the number of cells to be recovered after sequencing (Table 1). We typically observe a loss of cells after the lysis and wash steps and recommend initially resuspending the cells in a small volume of 10–20 μ l and counting the number of cells in an aliquot, followed by additional dilution to achieve the desired cell stock concentration for tagmentation. Cell loss is usually not a problem when starting with larger cell numbers of >100,000 cells. We typically aim for a stock concentration of \geq 2,000–3,000 cells/ μ l, by using \leq 5 μ l of cell suspension in the tagmentation reaction.
- 32 Keep the cell suspension on ice and take an aliquot of cells to mix 1:1 with trypan blue (e.g., 5 + 5 μ l or 10 + 10 μ l) for counting using a Countess II FL automated cell counter. When counting manually by using a Neubauer hemocytometer: Concentration [cells/ml] = (total number of cells \times 10,000/ number of squares) \times dilution factor. Cells should not clump together but be evenly dispersed (Fig. 4). If large cell clumps are observed, use a 40- μ m Flowmi cell strainer. Note that this may substantially reduce the cell concentration and volume available for downstream processing and may not be a suitable strategy in all instances.

▲ **CRITICAL STEP** After counting, proceed immediately with the tagmentation reaction.

? TROUBLESHOOTING**Tagmentation and GEM generation** ● **Timing 150–180 min**

- 33 Perform tagmentation and GEM generation as outlined in the ‘Chromium Next GEM Single Cell ATAC Reagent Kits v1.1’ user guide from 10x Genomics (CG000209 Rev F, Steps 1 and 2). We typically forward 10,000–17,000 cells in a single 15- μ l tagmentation reaction and tend to recover data for 60–65% of the cells. We did not observe differences in cell recovery rates compared to the standard protocol. Should significantly lower cell recovery rates be observed, we recommend contacting 10x Genomics for guidance and troubleshooting. It is recommended to consider the multiplet rate, which increases with loading higher cell numbers to the chip, because it is more likely that two or more cells will be captured together in the same GEM (e.g., multiplet rate is 8.0% when loading 16,500 cells, typically recovering 10,000 cells).

■ **PAUSE POINT** After PCR of generated GEMs, the tube may be stored at 4 °C or –20 °C for \leq 1 week. Emulsion breakage might occur at lower temperatures and after thawing. However, this is

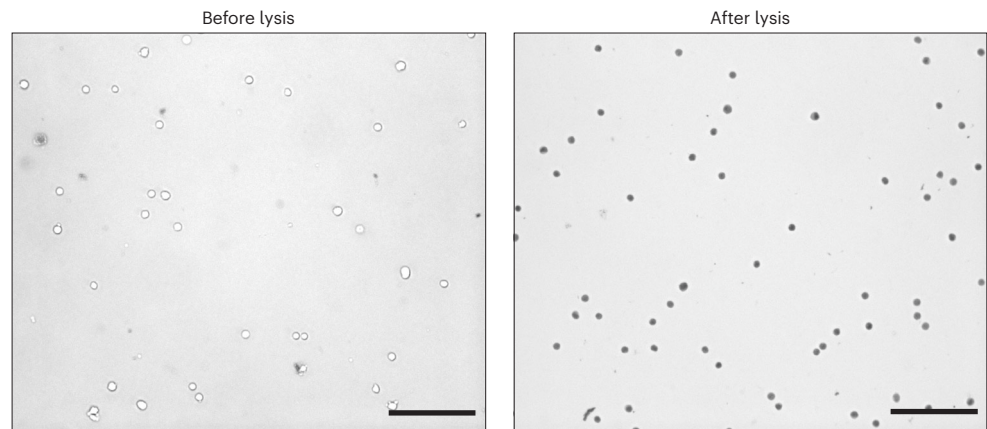


Fig. 4 | Images of human PBMCs before and after lysis. PBMCs were stained with trypan blue before fixation and lysis (left) and after lysis (right) to validate viability and successful lysis and permeabilization with no clumping of cells. Scale bar = 100 μ m. Images were obtained by using a Leica DFC3000 G microscope (20 \times magnification).

Table 2 | PCR cycle number suggestions

Cell number	Number of PCR cycles
500-2,000	12-13
2,001-6,000	11-12
6,001-10,000	10-11

not a problem, because the PCR reaction and cell barcoding of accessible chromatin and mtDNA fragments are complete at this point.

Post-GEM incubation cleanup and library construction ● Timing 90 min

34 Perform cleanup of the post-GEM pre-amplified PCR products and subsequent library construction (index PCR to add unique barcodes to each sample preparation) for sequencing as outlined in the ‘Chromium Next GEM Single Cell ATAC Reagent Kits v1.1’ user guide from 10x Genomics (CG000209 Rev F, Steps 3 and 4.1). Note: For mtscATAC-seq, we typically conduct one (rarely up to two) PCR cycle(s) more than recommended during the sample index PCR, depending on the cell number (Table 2 and Step 32).

■ **PAUSE POINT** After index PCR, store libraries at 4 °C for \leq 72 h or proceed to the next step.

Post-sample index PCR and double-sided size selection ● Timing 20-30 min

35 Follow the steps outlined in the ‘Chromium Next GEM Single Cell ATAC Reagent Kits v1.1’ user guide from 10x Genomics (CG000209 Rev F, Step 4.2). We typically elute libraries in 20–30 μ l of EB buffer. We recommend using DNA LoBind tubes for (long-term) storage of mtscATAC-seq sequencing libraries.

■ **PAUSE POINT** After clean-up, store libraries at 4 °C for \leq 72 h or at –20 °C for long-term storage.

mtscATAC-seq library quality control and quantification ● Timing 60-75 min

36 Quantify libraries by using a Qubit dsDNA high-sensitivity assay kit. We typically use 1 μ l of the eluted library and perform quantification in duplicates. A 10–25-ng/ μ l yield can be expected, but this will be a function of cell type and the number of cells forwarded to tagmentation and assuming a recovery of 60–65% of cells.

37 Check the fragment size distribution with a high-sensitivity DNA chip run on an Agilent Bioanalyzer 2100 system by using 1–10 ng of the mtscATAC-seq library (Fig. 5). Compared to the standard 10x Genomics Chromium scATAC-seq protocol, we observe a relative increase of the first peak, which typically contains nucleosome-free nuclear chromatin fragments (>150–300 bp) but is now further enriched with mtDNA fragments. Note that the relative ratio is a function of the

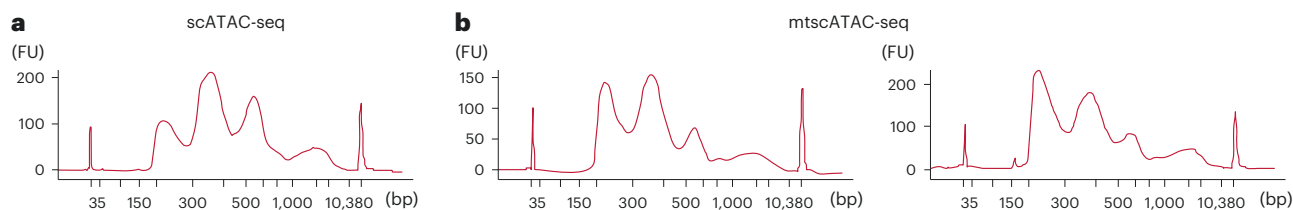


Fig. 5 | mtscATAC-seq library size distribution of human blood and immune cells. **a**, Typical fragment size distribution of a conventional 10x Genomics scATAC-seq library with prominent peaks indicative of the nucleosomal banding of chromatin. **b**, Typical fragment size distribution of mtscATAC-seq libraries, with the first peak (nucleosome-free DNA) often, but not always, being the most prominent, given the additional enrichment of mitochondrial DNA. Libraries were run on a high-sensitivity DNA chip and the Agilent Bioanalyzer 2100 system. FU, fluorescence units.

mtDNA content of cells and that the height of the first peak may be variable. See Fig. 6 for examples of mtscATAC-seq library size distribution problems that you might encounter.

? TROUBLESHOOTING

Sequencing ● Timing 1 d

38 Sequence the libraries by using, for example, the Illumina NextSeq 500/550 or NovaSeq 6000 platforms. We recommend longer read lengths for Read 1 and Read 2 to improve the coverage of mtDNA sequencing, thereby facilitating mutation calling (Table 3). For the NextSeq, we recommend the high-output cartridge kits (150 cycles) and paired-end sequencing (2 × 72 cycles for Reads 1 and 2). For the NovaSeq, we typically use SP, S1 or S2 reagent kits v1.5 (200 cycles) depending on the number of libraries to be sequenced and paired-end sequencing (2 × 100 cycles for Reads 1 and 2). For a dataset of 10,000 cells, we recommend sequencing to a depth of 100–150 million paired-end reads to start.

Generating a custom reference genome ● Timing 2–3 h

▲ **CRITICAL** Many regions of the mitochondrial genome have high homology with regions in the nuclear genome (NUMTs). Default parameters of sequencing alignment tools will discard many true mtDNA reads because of this homology. To assign multi-mapped reads to the mitochondrial genome, we generate a custom reference genome by hard-masking these high-homology regions in the nuclear genome. Implementing this step may increase coverage of the mitochondrial genome per cell by as much as 30%. If using a common reference genome (i.e., hg19, hg38 and mm10), access the prebuilt blacklist at <https://github.com/caleblareau/mitoblacklist> and skip to Step 42. Otherwise, simulate reads from the mitochondrial genome and align to the full reference genome (Steps 39–41).

39 Simulate short reads of length 20 from chrM. This requires the use of a fasta file of only the mitochondrial chromosome (chrM.fa) for the organism or reference genome desired for the mtscATAC-seq analysis. The Illumina ART software can generate these data with a single command:

```
art_illumina -ss NS50 -c 10000000 -i chrM.fa -l 20 -o simulated_chrM_dat
```

40 Map simulated chrM reads to the reference genome. Here, we use a bowtie2 indexed reference genome, but similar results can be obtained by using bwa⁵¹ or any related aligner.

```
bowtie2 -x path/to/bowtie2/reference -U simulated_chrM_dat | samtools view -bS - | \
```

```
samtools sort -o chrM.multimapped.bam
```

41 Make a blacklist bed file. Here, we identify bases that have non-zero coverage from the alignments at contigs that are not the mitochondrial chromosome (despite all reads from the simulation having originated at the mitochondrial chromosome). With the steps specified below, this can be readily exported into a .bed file that will be the input for the blacklist.

```
bedtools genomecov -ibam chrM.multimapped.bam -bg | grep -v chrM | \
```

```
awk '$4 > 0 {print $0}' | bedtools merge > mtDNA_blacklist.bed
```

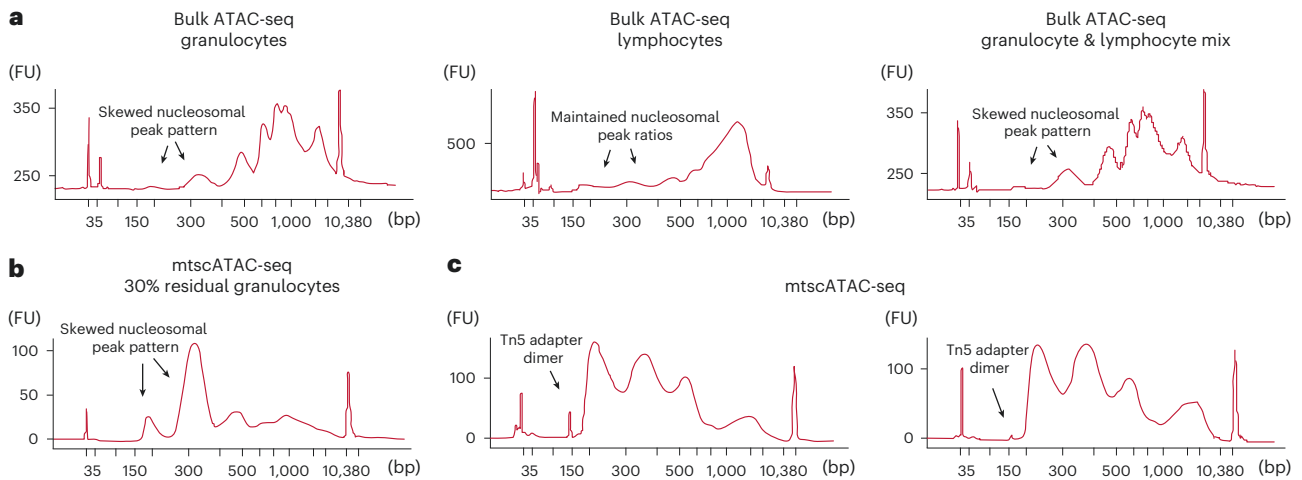



Fig. 6 | Troubleshooting mtscATAC-seq library size distribution. **a,b**, (Residual) granulocytes may significantly alter the ATAC-seq library size distribution as shown in bulk ATAC-seq (**a**) and *mtscATAC-seq* (**b**) libraries, resulting in substantially altered sequencing data quality. Note that bulk ATAC-seq libraries in **a** were not size selected, leading to larger genomic fragments between 1,000 and 10,000 bp being retained. **c**, Putative Tn5 adapter dimers may be observed in *scATAC-seq* and *mtscATAC-seq* libraries and appear more pronounced when the number of input cells is low. When dimers are present at low fractions within the libraries (as shown), they do not appear to interfere with sequencing.

Table 3 | Sequencing run parameters

Read or index	Number of cycles
Index 1 (i7)	8
Read 1	72-100
Read 2	72-100
Index 2 (i5)	16

42 Next, hard-mask the existing reference genome to be used in the CellRanger ATAC reference. This will replace A/C/G/T bases in the nuclear reference genome with Ns so that reads will not align at those loci. In our previous work^{15,16}, we observed that for human *mtscATAC-seq* libraries, this leads to an errant alignment of <1 in 1,000 molecules for high-quality *mtscATAC-seq* data.

```
bedtools maskfasta -fi genome.fa -bed mtDNA_blacklist.bed -fo mask-ed_genome.fa
```

43 Rebuild the reference, specifying the new masked genome file in the .json file. This will produce a reference genome packet compatible with the CellRanger-ATAC kit from 10x Genomics. The path can be specified in a typical CellRanger-ATAC run (see Step 45). We also describe these steps in detail at <https://github.com/caleblareau/mgatk/wiki/Increasing-coverage-from-10x-processing>.

```
cellranger-atac mkref -c config.json
```

mtscATAC-seq alignment and processing ● **Timing 12-24 h**

▲ **CRITICAL** This section describes the processing of sequencing data and the use of the described custom reference genome to fully process *mtscATAC-seq* data. Figure 7a shows a complete overview of the critical pieces for run-time execution, noting a final step of multi-modal analysis that can be completed interactively as outlined in the following section.

44 Demultiplex raw sequencing data with CellRanger-ATAC `mkfastq`. A feature of *mtscATAC-seq* is that both modalities (accessible chromatin and mtDNA) are contained within the same library, so one set of .fastq files will be sufficient for each experiment. Furthermore, the *mtscATAC-seq* libraries will be appropriately indexed with the 10x Genomics barcode sequence.

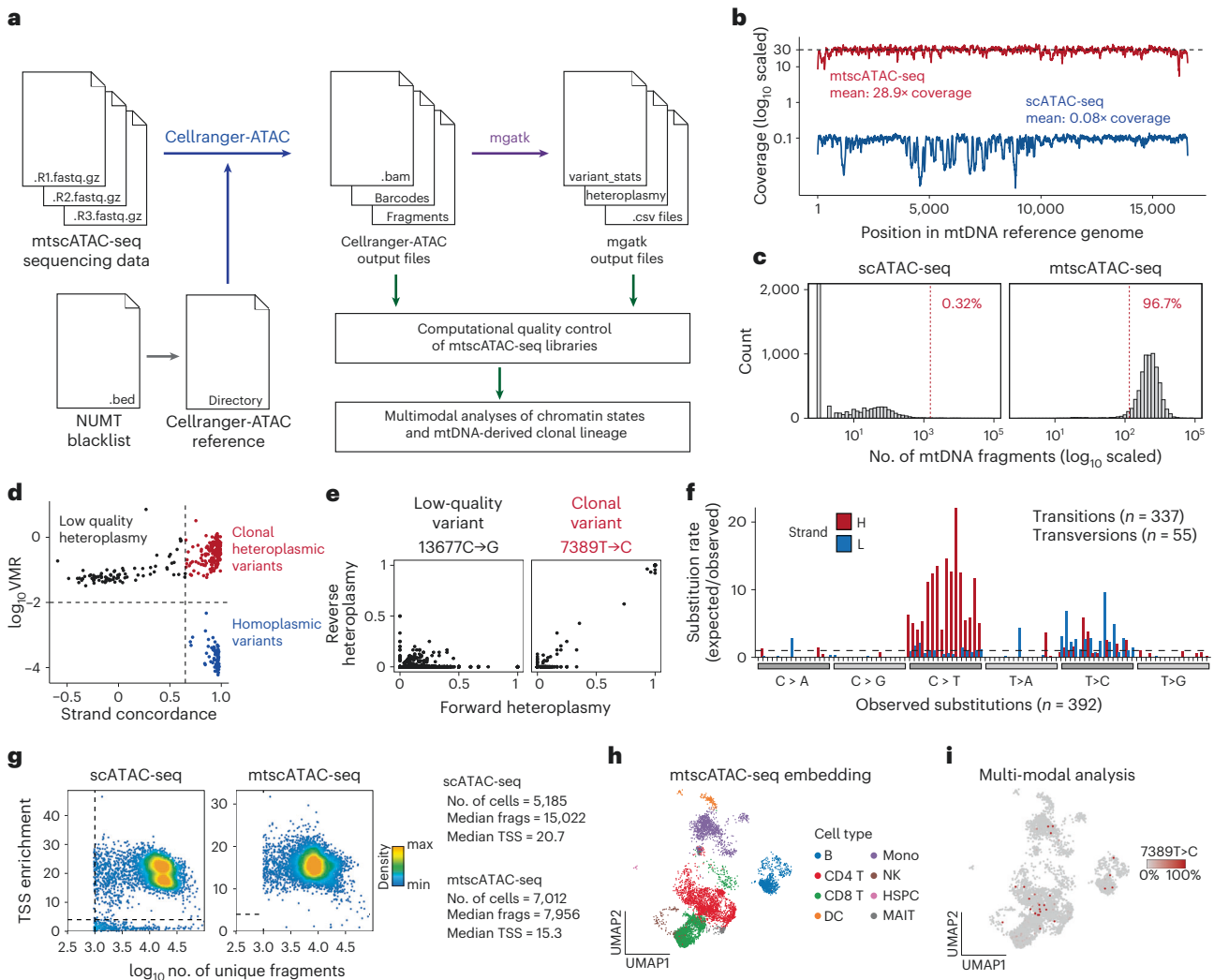


Fig. 7 | Overview of mtscATAC-seq computational workflow and quality-control metrics. a, Schematic of the computational processing workflow of mtscATAC-seq data using CellRanger-ATAC and mgatk. **b**, Visualization of mitochondrial genome coverage improvements in mtscATAC-seq (red) compared with the original scATAC-seq protocol (blue). The mean coverage per cell as called by CellRanger-ATAC is shown. The dotted line represents 30× coverage. **c**, Distribution of the number of mtDNA fragments (log₁₀) per cell as a function of library type. The red line shows -10× coverage and the proportion of cells per library that exceed this minimum coverage. **d**, Identification of high-confidence variants from high strand concordance in paired-end sequencing data and high variance-mean ratio (VMR). Homoplasmic and likely clonal heteroplasmic variants are noted in blue and red, respectively. **e**, Comparison of heteroplasmy estimated from reads aligned to either the forward or reverse strand for individual cells (dots). 13677C→G (left) is a low-quality variant (low strand correlation), whereas 7389T→C (right) is a high-quality variant (high strand correlation) as identified by mgatk. **f**, Mutation signature plot for all called variants from the default mgatk output. Shown is the substitution rate (observed over expected) of mutations (y axis) in each class of mononucleotide and trinucleotide change resolved by the heavy (H) and light (L) strands of the mitochondrial genome. **g**, ArchR quality-control metrics and thresholds for accessible chromatin data for indicated library types. Key summary metrics are summarized (right). **h**, Uniform manifold approximation and projection (UMAP) showing an example embedding of PBMCs annotated by accessible chromatin profiles. **i**, Projection of heteroplasmy of a high-quality variant (7389T→C; from **d** and **e**) in specific cell states (as in **h**), indicating a common clonal origin and showcasing the multi-modal nature of mtscATAC-seq data. DC, dendritic cell; HSPC, hematopoietic stem and progenitor cell; MAIT, mucosa-associated invariant T cell; Mono, monocyte; NK, natural killer.

```
cellranger-atac mkfastq -input raw_flow_cell_ref
```

45 Align sequencing data to the modified reference genome from Step 43 with CellRanger-ATAC. The custom CellRanger-ATAC reference facilitates a single-pass alignment to generate both high-quality chromatin accessibility data and approximately uniform coverage across the mitochondrial chromosome.

```
cellranger-atac count --fastqs path_to_fastqs --id sample_id --sample sample_directory \
```

```
--reference path_to_cellranger_atac_custom_reference
```

? TROUBLESHOOTING

- 46 Process sequencing data with `mgatk` to generate all possible variants for all cells identified by CellRanger-ATAC (specified in the filtered barcodes .tsv file). A full list of command line arguments for `mgatk` are available at <https://github.com/caleblareau/mgatk/wiki>.

```
mgatk tenx -i sample_directory/outs/possorted_bam.bam -n sample_id_mgatk -o sample_directory_mgatk \
```

```
-b sample_directory/outs/filtered_peak_bc_matrix/barcodes.tsv
```

! CAUTION We note that specific assumptions that hold true for mtscATAC-seq data are not necessarily true for other single-cell libraries. Specifically, high-quality variant inference from mtscATAC-seq libraries uses the principle that mutations are equally likely to be detectable on either strand during paired-end sequencing (Reads 1 and 2), whereas errant heteroplasmy will predominantly be detectable only from the forward or reverse strand. This errant heteroplasmy becomes more common in two-color sequencing chemistries, such as those used on the Illumina NextSeq and NovaSeq machines. In most scRNA-seq libraries, mitochondrial mRNA sequence information is extracted from only one of the sequenced strands, making the `mgatk` method not applicable. Furthermore, RNA-specific errors related to lower fidelity of the mtRNA polymerase and RNA-editing events probably occur in scRNA-seq libraries. Although mtRNA-based variants obtained from scRNA-seq data can indeed reveal true clonal lineages as we have previously shown^{13,37}, we recommend the use of a complementary (often bulk or single-cell) DNA-based genotyping assay, such as mtscATAC-seq, to corroborate such mutations.

? TROUBLESHOOTING

- 47 Verify the relative abundance of mtDNA fragments in the library and the coverage across the mitochondrial genome, as shown in Fig. 7b,c. Per-cell, per-position coverage estimates are reported in a plain text txt file (.coverage.txt.gz) as part of the default `mgatk` output. Summarizing the coverage by averaging over all cells in the capture yields the plots shown in Fig. 7b, whereas the count of mtDNA fragments (Fig. 7c) can be assessed from the `singlecell.csv` file.

? TROUBLESHOOTING

Variant analyses and integrative multi-modal analyses ● Timing 1-2 h

▲ CRITICAL Here, we outline the state of the art for using data from mtscATAC-seq alignments and `mgatk` output to identify high-confidence variants that can inform subclonal structure from the cellular input population. Furthermore, we describe the conceptual steps for downstream analysis and key files used throughout the workflow. For examples of specific commands used for downstream analysis, we refer to our online resources, including comprehensive vignettes available at our GitHub repositories.

- 48 Examine mtDNA variants from `mgatk` output. A plot of strand correlation and variance-mean ratio is the most appropriate for identifying informative mtDNA variants (Fig. 7d) and is reported in the `'vmr_strand_plot.png'` plot as part of the default output. Specifically, the x axis represents the Pearson correlation between a variant's forward and reverse strand read counts across cells. This metric effectively separates low-quality variants from high-quality ones on the basis of the overall concordance of heteroplasmy between strands. Examples of a low- and high-quality variant and respective strand-specific variant heteroplasmy are shown in Fig. 7e, whereby the high-quality clonal variant (7389T→C) has a more concordant profile between the strands. Overall, we expect to identify a pattern of substitutions in which some variants are more common than others (Fig. 7f). This variant signature plot can be generated rapidly from the `'variant_stats.tsv.gz'` file returned in the `mgatk` output.
- 49 Assess single-cell chromatin accessibility data quality. Many popular tools like ArchR⁵² and Signac⁴⁹ provide user-friendly methods to quickly assess library quality. Typical quality-control metrics for a public scATAC-seq dataset and an mtscATAC-seq library, both profiling PBMCs, are shown in Fig. 7g. Cell lines may show similar performance between mtscATAC-seq and

Box 4 | Key quality-control metrics for mtscATAC-seq libraries

- *Number of mtDNA fragments* (see Fig. 7c). This can be directly measured from the singlecell.csv file from CellRanger-ATAC. A high-quality mtscATAC-seq library will have 80–90% of cells with $\geq 1,000$ mtDNA fragments sequenced.
- *TSS score* (see Fig. 7g). As recommended in the ArchR manuscript⁵², TSS scores computed at the single-cell level provide a peak-independent means to reproducibly quantify accessible chromatin enrichment. A median TSS >10 for cells called by ArchR processing represent high-quality nuclear accessible chromatin alongside mtDNA in the mtscATAC-seq assay. An alternative method to derive the TSS score is from the .html summary provided by CellRanger-ATAC, which typically yields a value of 6–10 or more, averaging over all fragments detected in the library. Together, these two metrics can be used in tandem to verify the quality of both arms of the multimodal mtscATAC-seq assay.

scATAC-seq, but an ~30–40% reduction in accessible chromatin fragment yield may be observed, most likely due to the fixation step. A more significant departure from standard scATAC-seq quality metrics, particularly from the transcription start site (TSS) score, should be examined carefully to assess potential optimization of cell-processing steps upstream of 10x Genomics-based library preparation. Box 4 summarizes key quality-control metrics that are characteristic of a successful mtscATAC-seq library preparation of human PBMCs.

- 50 Perform multi-modal analysis. By using the default CellRanger-ATAC analysis or facile representations of cell states via ArchR⁵² or Signac⁴⁹, heterogeneity of cell phenotypes may be revealed as demonstrated for a PBMC sample (Fig. 7h). By using the ‘cell_heteroplasmic_df.tsv.gz’ file that provides single-cell heteroplasmy for all variants called with the default mgatk parameters (see Step 48), multi-modal analysis with mtDNA genotypes may be performed and their distribution visualized across the landscape of accessible chromatin profiles (Fig. 7i). Finally, we note that many analyses may require custom code that extends beyond the steps provided here. For this, we recommend Signac⁴⁹, which has an mtDNA genotyping, clonotype calling and analysis methodology built in as core functions.

? TROUBLESHOOTING

- 51 (Optional) Aggregate data across multiple 10x Genomics channels. This step is required only if multiple libraries of mtscATAC-seq are processed, which is often applicable in settings with multiple libraries in the analytical workflow. Aggregating data can be achieved in two ways. First, mgatk (Step 46) can be run on the aggregated .bam file automatically produced from CellRanger-ATAC (by default, this aggregated file is excluded from the aggr function). Second, one can run the cbind() function on the SummarizedExperiment objects (.rds files) from the mgatk output interactively in an R session. We emphasize that unlike for scRNA-seq or scATAC-seq, it is generally not appropriate to merge cells derived from multiple different donors, because of the lack of direct clonal relationship between cells in such instances. For consistency of representation of cell states, we generally will project all mtscATAC-seq channels onto a common well-annotated reference map, as shown in Seurat3/Signac⁴⁹.

Troubleshooting

Troubleshooting advice can be found in Table 4.

Table 4 | Troubleshooting table

Step	Problem	Possible reason	Solution
6	Low cell viability after cell processing or thawing	Cryopreservation or cell dissociation steps may not be optimal	Implement gentle dissociation methods to ensure high viability before cryopreservation of cells Avoid excessive pipetting of cells After thawing, minimize osmotic pressure on cells Consider fixing cells in the presence of a fixable live or dead cell dye early in the process to maintain their integrity
17	Low cell yield after sorting	Too stringent sort parameters	Consult with the staff at your FACS facility. We often use settings designed to enable optimal yield of desired cell populations
	Low cell recovery after sorting	Electrical charge of cells may increase their stickiness to plastic consumables	Add or increase the amount of BSA or FBS in the medium or buffer into which one sorts the cells (e.g., 1–5%), because this can substantially improve recovery of cells for downstream processing Use swinging-bucket centrifuges for cell recovery

Table continued

Table 4 (continued)

Step	Problem	Possible reason	Solution
32	High proportion of live cells (>20%) after lysis	Insufficient lysis and/or permeabilization	Optimize lysis and/or permeabilization conditions for your cell type or tissue. The incubation time in lysis buffer may need to be adjusted and may be determined empirically by using non-formaldehyde-fixed cells
37	Abnormal library size distribution, such as prominent high-molecular-weight peaks	Double-sided solid-phase reversible immobilization (SPRI) may not have been optimal	Ensure that the correct SPRI ratios are used and that the SPRI solution has been properly mixed before use. Note that high-molecular-weight fragments are not sequenced by using the conventional sequencing approaches described here and thus may not interfere with the results
	Relatively small first peak (nucleosome-free DNA)	Depletion of mtDNA or contamination with granulocytes	Mitochondria may have been depleted during lysis. Do not add Tween 20 to the lysis or wash buffer, because this leads to depletion of mitochondria and mtDNA Alternatively, a large proportion of residual granulocytes (>5%) may lead to skewing of the library size distribution (Fig. 6a,b). Ensure the purity of your sample by more stringent gating when sorting and perform density gradient centrifugation (e.g., Ficoll) to remove granulocyte populations
	Peak at 150 bp	Tn5 adapter dimer	Sometimes peaks at 150 bp may be observed (Fig. 6c). We believe that these arise from unused Tn5 adapters, which may form primer dimers with oligos coming from the 10x Genomics gel beads. We have noted these peaks to be more prominent with lower input cell numbers and to essentially disappear when cell input is high, which probably saturates the use of all available Tn5 adapters; thus, no dimers can form. At low input levels (as shown), these peaks do not appear to interfere with sequencing. Should these dimers make up a significant proportion (>20%) of the libraries loaded on a sequencer, we would be concerned about possible interference, given their likely highly invariant nucleotide sequences. Consider increasing your input cell numbers and/or repeat a double-sided SPRI selection
	Low yield of final library as measured by Qubit	Ensure proper elution of DNA during cleanup steps	Carefully follow the outlined DNA purification steps by using SPRI select and be aware of when DNA is bound to beads or in the supernatant. Ensure careful removal of any residual ethanol during wash steps
45	Run-time errors with CellRanger-ATAC	Incorrectly specified file inputs; incompatible compute requirements	Obtain professional support from 10x Genomics staff: https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/troubleshooting
46	Errors with installation or setting up mgatk	First-time user	Follow the complete installation guide at https://github.com/caleblareau/mgatk/wiki
47	Run-time errors with mgatk	Incorrect reference genome specification; memory errors; file input error	Review over 50 previously solved issues or raise a new one at https://github.com/caleblareau/mgatk/issues
50	Advanced interactive analyses or run-time errors with Signac	Downstream function requirements	Follow the step-by-step guide for interactive analyses with Signac at https://satijalab.org/signac/articles/mito.html

For issues relating to GEM generation, including clogs, refer to the troubleshooting guide provided by 10x Genomics. We have not observed any recurrent issues related to these aspects with our modified cell-processing, fixation and lysis and permeabilization protocols.

Timing

Steps 1–25, cell preparation: 2.5 h (hands-on time: 1.5 h)—Steps 1–8, thawing: 25–35 min; Steps 9–16, staining: 35–40 min (hands-on time: 15 min); Steps 17–21, sorting: 30–60 min (hands-on time: 15 min); Steps 22–25, fixation: 25 min
Steps 26–32, cell lysis and permeabilization: 1 h
Step 33, tagmentation and GEM generation: 2.5–3 h (hands-on time: 1 h)
Steps 34–37, library preparation: 3 h (hands-on time: 2 h)—Step 34, Dynabeads cleanup: 35 min; Step 34, SPRI cleanup: 15 min; Step 34, indexing PCR: 40 min (hands-on time: 10 min); Step 35, post-sample index PCR and double-sided size selection: 20–30 min; Step 36, Qubit quantification: 10 min; Step 37,

Bioanalyzer: 60 min (hands-on time: 15 min)
Step 38, next-generation sequencing: ~1 d
Steps 39–43, generating a custom reference genome: 2–3 h
Steps 44–47, mtscATAC-seq alignment and processing: 12–24 h
Steps 48–51, mtDNA variant calling: 1–2 h (hands-on time: 5 min)

Anticipated results

For primary human hematopoietic and immune cells, we typically observe 10–25 ng/μl of product in 20–30 μl of elution volume for the final mtscATAC library (Steps 36 and 37), noting that this is a function of cell type, input cell amount, fraction of recovered cells and number of PCR cycles used. Typical mtscATAC-seq library size distributions are shown in Fig. 5, noting that some variability in distribution may be observed. Ultimately, only sequencing will yield reliable information about the success and quality of the experiment, as indicated by the outlined quality-control metrics (Box 4 and Fig. 7), including the number of cells with >1,000 fragments per cell, TSS score and abundance of mtDNA reads. Accessible chromatin-specific metrics for mtscATAC-seq data should align with best practices used for scATAC-seq from other platforms, including single-nucleus 10x Genomics scATAC-seq.

In addition to nuclear accessible chromatin fragments, our python package, *mgatk*, facilitates the rapid processing and analysis of mtDNA information into quality-control metrics as well as the identification of variants for functional analysis. Conceptually, *mgatk* enables the computationally efficient processing of sequencing reads stored in the CellRanger-ATAC output into plain text files that exhaustively enumerate per-variant, per-cell read counts. From these processed files, both quality control and variant calling can be performed efficiently. Specifically, high-confidence cells with an mtDNA coverage depth exceeding a user-specified threshold (default: 10) serve as input into the downstream variant calling. Thus, a common quality-control metric is to verify that ≥80–90% of cells have a minimum coverage of at least a certain value (e.g., 10 or 20) depending on the cell type and biological question of interest. In our experience with human immune cells derived from peripheral blood, a quality mtscATAC-seq library will have >95% of cells that have >10× coverage over the mitochondrial genome. Furthermore, a key measure of the overall protocol effectiveness is the identification of high-quality mtDNA variants, which can be done rapidly via the *mgatk* package. Here, we note that in our experience, >80% of the called variants will be transition mutations (Fig. 7f), including a disproportionate enrichment of C→T mutations on the heavy strand of the DNA. Users can proceed with confidence if the mutational signature analysis resembles the plot shown in Fig. 7f. Noting that transition mutations occur via common deamination, transitions comprise the vast majority of mutations in every mtscATAC-seq library that we have processed to date (primarily in hematopoietic cells).

An important advantage of the mtscATAC-seq workflow is the sensitivity of the variant identification, which we have established is accurate down to 1 mutated molecule per 1,000 mtDNA genomes in pseudobulk¹⁵. For each observed nucleotide at a position, *mgatk* sums across all cells separately for the forward and reverse strands, discarding any nucleotide or position pair if it is not observed in any cell on either of the two strands. After comparison to the reference genome, a list of potential variants is then scored for various metrics on the basis of heteroplasmy, strand concordance, average depth at the position per cell and within-library mean and variability of heteroplasmy. We have found that two metrics in particular consistently yield high-confidence variants. The first metric uses the Pearson correlation between counts from the forward and reverse strands. Similar to the Fisher's test used in GATK⁵³, our rationale behind this metric stems from our observation that many artefactual variants tend to be highly strand biased (due to differences in the proximal nucleic acid sequences near the variant from either direction), whereas true variants should be well supported by both strands. Second, *mgatk* computes the variance-mean ratio (VMR) of the per-cell heteroplasmy. We have found VMR useful for delineating homoplasmic from heteroplasmic variants. However, we note one exception, RNA editing events, that, although not meaningful in mtscATAC-seq data, provides a rationale for further development of between-cell metrics that may correctly identify more constant sources of false heteroplasmy between cells.

In the final step of calling high-confidence heteroplasmic variants, *mgatk* uses fixed default values for these metrics (0.65 for strand correlation and 0.01 for VMR) that we have found to be applicable across many different datasets, including distinct tissue types and cell yields varying by more than an order of magnitude. However, we anticipate that depending on the particular application in mind, users may choose to be more or less conservative with the selected value of either threshold. As such, *mgatk* produces a scatter plot of potential variants in which these two metrics are plotted, and the default values

are noted. Although we generally recommend using only variants passing both default thresholds for downstream analysis, *mgatk* outputs a variant statistics table as a .tsv file for all potential variants, enabling rapid updates to variant call sets (whereas other tools often require re-running one or more steps of a pipeline to change a statistical parameter).

In addition to the strand correlation and VMR statistics that have been discussed, we emphasize other summary statistics that we have found useful for understanding properties of somatic mtDNA variation in our studies. In 'variant_n_cells_conf_detected', the number of cells for which there are more than two counts of the variant in both strands (indicative of additional confidence that the variant is present in that individual cell) are indicated, and 'variant_n_cells_over_Z' contains the number of cells with variant heteroplasmy greater than Z% ($Z = \{5, 10, 20, 95\}$ in the current version); these are semi-arbitrary thresholds that we have found useful for downstream interpretation of heteroplasmic variants. Finally, the 'max_heteroplasmy' statistic is the maximum percentage of heteroplasmy detected in any cell. These additional summary statistics facilitate an understanding of the lineage dynamics, because we often observe somatic variants occurring at 100% homoplasmy in a small subset of cells. Lastly, a cell-by-variant percent heteroplasmy table is saved as a .tsv file for all variants with 'variant_n_cells_conf_detected' >2. In total, this summary statistics file provides an appropriate starting point for custom downstream analyses, including cell-cell similarity estimates and clone calling via graph-based clustering¹⁵.

Data availability

Raw mtscATAC-seq data for the demonstration of the analysis was downloaded from <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM4472967> and processed with Cell Ranger v2 to the hg38 reference genome before being processed with *mgatk* v0.6.1. The comparison single-nucleus ATAC-seq dataset is available at <https://www.10xgenomics.com/resources/datasets/5-k-peripheral-blood-mononuclear-cells-pbm-cs-from-a-healthy-donor-next-gem-v-1-1-1-1-standard-2-0-0>.

Code availability

A code resource for performing mtscATAC-seq analyses and reproducing the analysis figures in this work is available at https://github.com/caleblareau/mtscatac_protocol. The *mgatk* package is made freely available and is actively maintained at <https://github.com/caleblareau/mgatk>. Additional support for interactive variant calling in the R environment has been incorporated in the CRAN Signac package for single-cell chromatin analyses⁴⁹.

References

1. Baron, C. S. & van Oudenaarden, A. Unravelling cellular relationships during development and regeneration using genetic lineage tracing. *Nat. Rev. Mol. Cell Biol.* **20**, 753–765 (2019).
2. VanHorn, S. & Morris, S. A. Next-generation lineage tracing and fate mapping to interrogate development. *Dev. Cell* **56**, 7–21 (2021).
3. Wagner, D. E. & Klein, A. M. Lineage tracing meets single-cell omics: opportunities and challenges. *Nat. Rev. Genet.* **21**, 410–427 (2020).
4. Woodworth, M. B., Girsakis, K. M. & Walsh, C. A. Building a lineage from single cells: genetic techniques for cell lineage tracking. *Nat. Rev. Genet.* **18**, 230–244 (2017).
5. Biasco, L. et al. In vivo tracking of human hematopoiesis reveals patterns of clonal dynamics during early and steady-state reconstitution phases. *Cell Stem Cell* **19**, 107–119 (2016).
6. Ferrari, S. et al. Efficient gene editing of human long-term hematopoietic stem cells validated by clonal tracking. *Nat. Biotechnol.* **38**, 1298–1308 (2020).
7. Park, S. et al. Clonal dynamics in early human embryogenesis inferred from somatic mutation. *Nature* **597**, 393–397 (2021).
8. Coorens, T. H. H. et al. Extensive phylogenies of human development inferred from somatic mutations. *Nature* **597**, 387–392 (2021).
9. Lodato, M. A. et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94–98 (2015).
10. Lee-Six, H. et al. Population dynamics of normal human blood inferred from somatic mutations. *Nature* **561**, 473–478 (2018).
11. Brunner, S. F. et al. Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature* **574**, 538–542 (2019).
12. Macaulay, I. C. et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* **12**, 519–522 (2015).
13. Ludwig, L. S. et al. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* **176**, 1325–1339.e22 (2019).

14. Xu, J. et al. Single-cell lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial DNA. *Elife* **8**, e45105 (2019).
15. Lareau, C. A. et al. Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat. Biotechnol.* **39**, 451–461 (2021).
16. Penter, L. et al. Longitudinal single-cell dynamics of chromatin accessibility and mitochondrial mutations in chronic lymphocytic leukemia mirror disease history. *Cancer Discov.* **11**, 3048–3063 (2021).
17. Taylor, R. W. et al. Mitochondrial DNA mutations in human colonic crypt stem cells. *J. Clin. Invest.* **112**, 1351–1360 (2003).
18. Teixeira, V. H. et al. Stochastic homeostasis in human airway epithelium is achieved by neutral competition of basal cell progenitors. *Elife* **2**, e00966 (2013).
19. Stewart, J. B. & Chinnery, P. F. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat. Rev. Genet.* **16**, 530–542 (2015).
20. Kang, E. et al. Age-related accumulation of somatic mitochondrial DNA mutations in adult-derived human iPSCs. *Cell Stem Cell* **18**, 625–636 (2016).
21. Walker, M. A. et al. Purifying selection against pathogenic mitochondrial DNA in human T cells. *N. Engl. J. Med.* **383**, 1556–1563 (2020).
22. Corces, M. R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).
23. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
24. Montefiori, L. et al. Reducing mitochondrial reads in ATAC-seq using CRISPR/Cas9. *Sci. Rep.* **7**, 1–9 (2017).
25. Fiskin, E. et al. Single-cell profiling of proteins and chromatin accessibility using PHAGE-ATAC. *Nat. Biotechnol.* **40**, 374–381 (2022).
26. Mimitou, E. P. et al. Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat. Biotechnol.* **39**, 1246–1258 (2021).
27. Santibanez-Koref, M. et al. Assessing mitochondrial heteroplasmy using next generation sequencing: a note of caution. *Mitochondrion* **46**, 302–306 (2019).
28. Laricchia, K. M. et al. Mitochondrial DNA variation across 56,434 individuals in gnomAD. *Genome Res.* **32**, 569–582 (2022).
29. Maude, H. et al. NUMT confounding biases mitochondrial heteroplasmy calls in favor of the reference allele. *Front. Cell Dev. Biol.* **7**, 201 (2019).
30. Satpathy, A. T. et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019).
31. Tang, Z. et al. A genetic bottleneck of mitochondrial DNA during human lymphocyte development. *Mol. Biol. Evol.* **39**, msac090 (2022).
32. Lareau, C. A. et al. Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat. Biotechnol.* **37**, 916–924 (2019).
33. Domcke, S. et al. A human cell atlas of fetal chromatin accessibility. *Science* **370**, eaba7612 (2020).
34. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
35. Cusanovich, D. A. et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
36. Basu, U., Bostwick, A. M., Das, K., Dittenhafer-Reed, K. E. & Patel, S. S. Structure, mechanism, and regulation of mitochondrial DNA transcription initiation. *J. Biol. Chem.* **295**, 18406–18425 (2020).
37. Miller, T. E. et al. Mitochondrial variant enrichment from high-throughput single-cell RNA sequencing resolves clonal populations. *Nat. Biotechnol.* **40**, 1030–1034 (2022).
38. Mulqueen, R. M. et al. High-content single-cell combinatorial indexing. *Nat. Biotechnol.* **39**, 1574–1580 (2021).
39. Nam, A. S. et al. Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature* **571**, 355–360 (2019).
40. Rodriguez-Meira, A. et al. Unravelling intratumoral heterogeneity through high-sensitivity single-cell mutational analysis and parallel RNA sequencing. *Mol. Cell* **73**, 1292–1305.e8 (2019).
41. Stoekius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
42. Triska, P. et al. Landscape of germline and somatic mitochondrial DNA mutations in pediatric malignancies. *Cancer Res.* **79**, 1318–1330 (2019).
43. Slyper, M. et al. A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat. Med.* **26**, 792–802 (2020).
44. Qian, J. et al. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Res.* **30**, 745–762 (2020).
45. Wakiro, I. et al. HTAPP_Dissociation of human ovarian cancer resection to a single-cell suspension for single-cell RNA-seq. Available at <https://www.protocols.io/view/htapp-dissociation-of-human-ovarian-cancer-resecti-bbhj2j6> (2020).
46. Swanson, E. et al. Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq. *Elife* **10**, e63632 (2021).

47. Chen, X. et al. ATAC-seq reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* **13**, 1013–1020 (2016).
48. Stoler, N. & Nekrutenko, A. Sequencing error profiles of Illumina sequencing instruments. *NAR Genom. Bioinform.* **3**, lqab019 (2021).
49. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
50. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
51. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
52. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).
53. Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **43**, 11.10.1–11.10.33 (2013).

Acknowledgements

We are grateful to members of the Satpathy, Regev, Sankaran and Ludwig laboratories for helpful discussions. This research was supported by a Stanford Science Fellowship (to C.A.L.) and a Parker Institute of Cancer Immunotherapy Scholarship (to C.A.L.), a Stanford Artificial Intelligence in Medicine and Imaging seed grant (to C.A.L.), a BroadIgnite Award (L.S.L.), R01 DK103794 (to V.G.S.), R33 HL120791 (to V.G.S.), a gift from the Lodish Family to Boston Children's Hospital (to V.G.S.), the New York Stem Cell Foundation (NYSCEF; to V.G.S.) and the Howard Hughes Medical Institute and Klarman Cell Observatory (to A.R.). V.G.S. is an NYSCEF-Robertson Investigator. P.K. is an Associated Fellow of the Hector Fellow Academy. L.N. is funded by a fellowship from the MDC-NYU exchange program and is an Associated Fellow of the Hector Fellow Academy. L.S.L. is supported by the Berlin Institute of Health, an Emmy Noether fellowship by the German Research Foundation (DFG, LU 2336/2-1) and a Hector Research Career Development Award. A.T.S. is supported by the National Institutes of Health (NIH) U01CA260852, the Parker Institute for Cancer Immunotherapy, a Career Award for Medical Scientists from the Burroughs Wellcome Fund and a Pew-Stewart Scholars for Cancer Research Award. A.T.S., C.A.L. and L.S.L. are supported by NIH UM1HG012076.

Author contributions

C.A.L., L.S.L. and C.M. conceived and designed the protocol with supervision from A.R. and V.G.S. K.S. designed and optimized the solid tissue dissociation protocol. V.L. and J.C.G. refined the computational workflow with supervision from C.A.L. and A.T.S. L.N., P.K., K.S., Y.Y. and K.P. independently executed the protocol and provided critical feedback on the manuscript. C.A.L., S.D.P., A.T.S., A.R., V.G.S. and L.S.L. supervised various aspects of this study. All authors contributed to the writing and editing of the protocol.

Competing interests

The Broad Institute has filed a patent relating to the use of the technology described in this paper for which C.A.L., L.S.L., C.M., A.R. and V.G.S. are named as inventors (US Patent App. 17/251,451). A.R. is a founder and equity holder of Celsius Therapeutics, an equity holder in Immunitas Therapeutics and until 31 August 2020, was a scientific advisory board member of Syros Pharmaceuticals, Neogene Therapeutics, Asimov and Thermo Fisher Scientific. From 1 August 2020, A.R. has been an employee of Genentech. V.G.S. serves as an advisor to and/or has equity in Novartis, Forma, Cellarity, Ensoma and Polaris Partners. A.T.S. is a founder of Immunai and Cartography Biosciences and receives research funding from Merck Research Laboratories and Allogene Therapeutics.

Additional information

Correspondence and requests for materials should be addressed to Caleb A. Lareau, Aviv Regev, Vijay G. Sankaran or Leif S. Ludwig.

Peer review information *Nature Protocols* thanks Judith Zaugg and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Received: 8 March 2022; Accepted: 11 November 2022;

Published online: 15 February 2023

Related links

Key references using this protocol

Lareau, C. et al. *Nat. Biotechnol.* **39**, 451–461 (2021): <https://doi.org/10.1038/s41587-020-0645-6>

Walker, M. et al. *N. Engl. J. Med.* **383**, 1556–1563 (2020): <https://doi.org/10.1056/NEJMoa2001265>

Penter, L. et al. *Cancer Discov.* **11**, 3048–3063 (2021): <https://doi.org/10.1158/2159-8290.CD-21-0276>