Single-cell Full-length Nanopore Sequencing for Quantitative Variant Analysis of Native and Genome-Edited Mitochondria

Chongwei Bi[1], Lin Wang[1], Yong Fan[2], Yang Yu[3], Juan Carlos Izpisua Belmonte[4], and **Mo Li**[1]

1 Stem Cell and Regeneration Laboratory, King Abdullah University of Science and Technology (KAUST), Thuwal, Kingdom of Saudi Arabia. 2 The Third Affiliated Hospital of Guangzhou Medical University. 3 Peking University Third Hospital. 4 Altos Labs, San Diego, CA, USA. Contact: mo.li@kaust.edu.sa

hiring postdocs

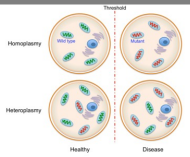
Abstract

The ontogeny and dynamics of mtDNA heteroplasmy remain unclear due to limitations of current mtDNA sequencing methods. We developed individual **Mi**tochondrial **G**enome **seq**uencing (iMiGseq) of full-length mtDNA for ultra-sensitive variant detection, complete haplotyping, and unbiased evaluation of heteroplasmy levels, all at the individual mtDNA molecule level. iMiGseq detected sequential acquisition of detrimental mutations in defective mtDNA in NARP/Leigh syndrome patient-derived induced pluripotent stem cells (iPSCs). iMiGseq identified unintended heteroplasmy shifts in mitoTALEN edited NARP/Leigh syndrome iPSCs. iMiGseq of mitochondrial base editor DdCBE-edited cells did not detect any appreciable level of unintended mutations in mtDNA. iMiGseq uncovered unappreciated levels of heteroplasmic variants in single healthy human oocytes well below the conventional NGS detection limit, of which numerous variants are deleterious and associated with late-onset mitochondrial disease and cancer. iMiGseq revealed dramatic shifts in variant frequency and clonal expansion of large structural variants during oogenesis and stable heteroplasmy levels during human blastoid generation. It showed the first haplotype-resolved mitochondrial genomes from single human oocytes and single human blastoids. Therefore, iMiGseq could not only help elucidate the mitochondrial etiology of diseases, but also enhance the precision of mitochondrial disease diagnosis.

Introduction

Challenges in mtDNA study

- Mutant mtDNA exhibits **uneven distribution** among different tissues and cells
- **Heteroplasmy** - coexistence of wild-type and mutant mtDNA in a single cell
- Amalgamation and fragmentation of mtDNA when using conventional NGS



What's needed?

- **single-cell** level manipulation
- **capture** the entire mtDNA
- **single reads** covering the entire mtDNA
- long reads avoid mapping of nuclear mitochondrial DNA-like sequences (**NUMTs**)

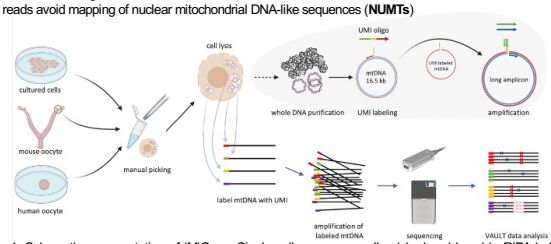


Figure 1. Schematic representation of iMiGseq. Single cells were manually picked and lysed in RIPA buffer. iMiGseq applies the IDMseq strategy to specifically label individual mtDNA with UMIs by a single round of primer extension. The UMI oligo contains a 3' gene-specific sequence (red), a UMI sequence (yellow), and a 5' universal primer sequence (green). The UMI-labeled mtDNA is further amplified by long-range PCR using the universal primer (green) and gene-specific reverse primer (purple) as a single amplicon. The sequencing of long amplicons is performed on long-read Oxford Nanopore MinION and PacBio sequencing platforms. The sequencing data were analyzed by a bioinformatic toolkit–VAULT–to identify UMI sequences, bin reads based on UMI and call variants.

Results

1. Validation of iMiGseq using 293T cells.

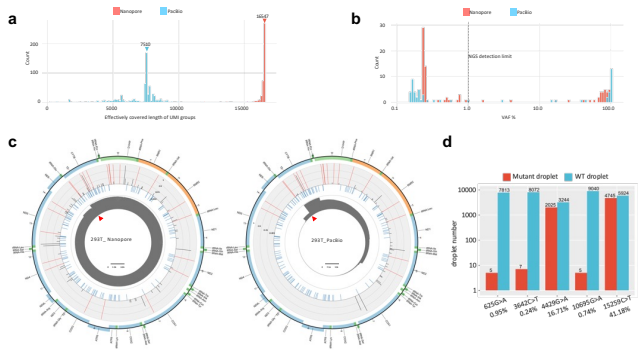


Figure 2. **a**, Distribution of effectively covered length of UMI groups detected by Nanopore and PacBio sequencing. Color-coded triangles indicate N50 values. **b**, VAF distribution of SNVs detected in iMiGseq of 293T cells. The majority of unique SNVs are below the current 1% detection limit (the vertical dotted line). **c**, Circular plots showing the distribution of SNVs in the mitochondrial genome of 293T cells determined by Nanopore and PacBio sequencing. The innermost circle (grey) shows the depth of reads of all detected UMI groups in linear scale as indicated by the scale bar in the center. The red triangle indicates the position of the primers. The middle circle (light blue) represents common SNPs from the human dbSNP-151 database. Individual SNVs are plotted in the outer barplot circle, in which the height of bar represents the VAF. Red color indicates VAF > 0.6. **d**, ddPCR results showing the detected positive events (droplets) for variants with different VAFs. The VAFs calculated by Nanopore iMiGseq are shown under the variants. Around 80 cells were used in ddPCR to ensure the generation of enough events for analysis, as compared to 5 cells in iMiGseq. The data are from two independent biological replicates.

2. iMiGseq of patient-derived hEPS$^{m.8993T>C}$ cells.

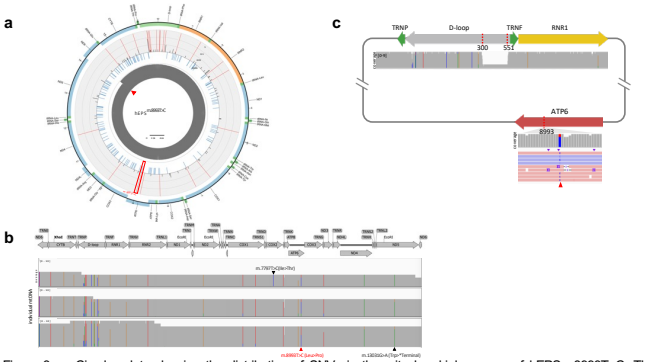
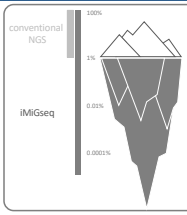


Figure 3. **a,** Circular plots showing the distribution of SNVs in the mitochondrial genome of hEPSm.8993T>C. The arrangement of the circular plot is similar to Fig. 2c and the m.8993T>C mutation is indicated by the red frame. **b**, Sequential acquisition of detrimental mutations illustrated by Integrative Genomics Viewer (IGV) plots. Three UMI groups (mtDNA) with the same pathogenic m.13031G>A mutation (indicated in black under the plot) are shown. The putative disease-causing mutation of m.8993T>C is indicated in red under the plot. A de novo m.7797T>C mutation is found in one UMI group and indicated in black above the plot. **c**, A schematic map of the mutant mtDNA harboring the 251-bp deletion and m.8993T>C mutation detected in hEPSm.8993T>C and the accompanying IGV tracks showing the alignment of Nanopore reads. The read triangle shows the position 8993 in the reads.

Take-Home Messages



- **Ultra-sensitive variant detection**, **complete haplotyping**, and **unbiased evaluation of heteroplasmy levels**, all at the **individual mtDNA molecule level**
- iMiGseq identified unintended heteroplasmy shifts in **mitoTALEN editing**.
- Many rare variants are deleterious or **associated with late-onset mitochondrial disease and cancer**
- iMiGseq revealed **dramatic shifts in variant frequency** and **clonal expansion of large structural variants during oogenesis** and stable heteroplasmy levels during **human blastoid generation.**

3. Characterization of mitochondrial genome editing by mitoTALEN and DdCBE.

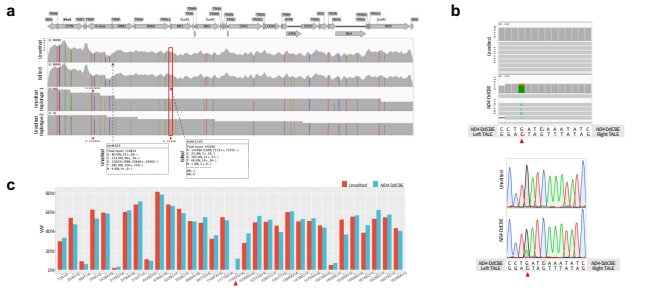


Figure 4. **a,** The read alignment of unedited (the top track) and mitoTALEN edited (the second track) cells by deep Illumina sequencing, and two haplotypes (the bottom two tracks) of unedited cells by iMiGseq illustrated by IGV plots. The mitoTALEN target position (m.3243) is labeled by red frame. Bulk Illumina sequencing confirms the complete elimination of the m.3243A>G mutation but fails to detect the rare m.625G>A SNV (VAF of 0.55%). The two haplotypes by iMiGseq show the predominant linkage between m.16289A and m.3243G, and the infrequent linkage of m.16289G and m.3243G (only one UMI group is shown for each haplotype). Bulk Illumina sequencing is unable to conduct haplotype analysis. **b**, The m.11922G>A SNV (on the top H-strand, i.e., C4>T on the bottom L-strand) (red triangle) installed by ND4-DdCBE is detected by iMiGseq (left, one UMI group is shown for each genotype) and confirmed by Sanger sequencing (right). A diagram of the ND4-DdCBE left and right TALEs and the target spacing region is shown below the sequencing results. The cytosines that are within the ND4-DdCBE editing window are highlighted. **c**, The VAF changes of primary SNVs (VAF > 1%) before and after ND4-DdCBE editing. No significant VAF shift is observed.

4. iMiGseq of single human oocytes.

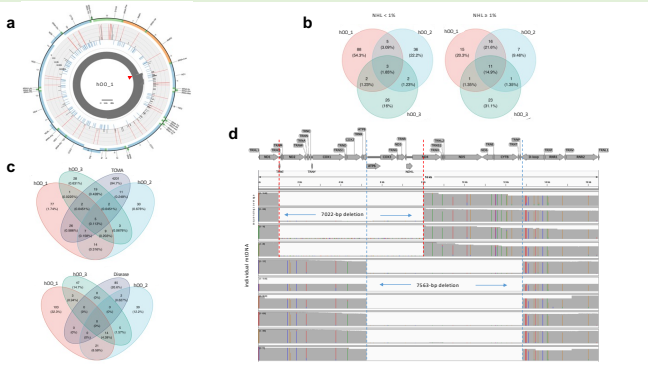


Figure 5. **a,** Circular plots showing the distribution of SNVs in the mitochondrial genome of hOO_1. **b**, Venn diagrams showing the overlap of mtDNA SNVs between the hOOs at different heteroplasmy levels. A higher proportion of SNVs with an NHL ≥ 1% (right) were shared among hOOs than those with an NHL < 1% (left). **c**, Venn diagrams showing the overlap of mtDNA SNVs in hOOs and cancers (TCMA, upper), or confirmed pathogenic mtDNA SNVs in mitochondrial diseases (Disease, lower). Cancer-related mtDNA SNVs exist in all three hOOs. **e**, A schematic map of the human mtDNA and the accompanying IGV tracks showing the alignment of Nanopore reads of ten UMI groups (mtDNA) harboring 7022-bp and 7563-bp deletions in SFhOO_2. Vertical dotted lines indicate deletion breakpoints. Colored bars in the read coverage track represent SNVs.

5. iMiGseq for tracking mtDNA heteroplasmy dynamics during the generation of human blastoid from niPSCs.

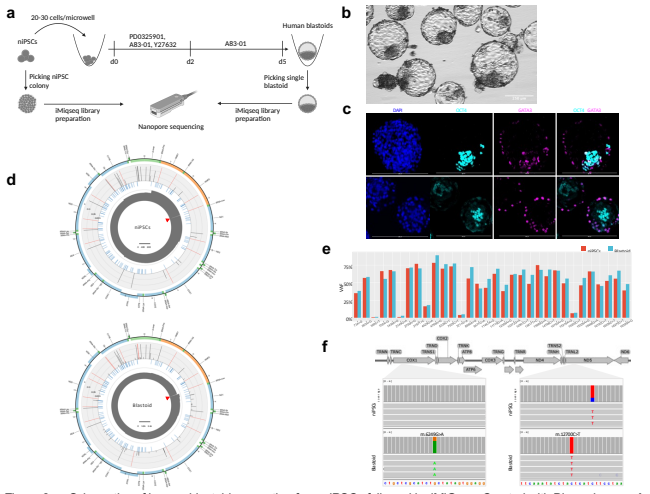


Figure 6. **a,** Schematics of human blastoid generation from niPSCs followed by iMiGseq. Created with Biorender.com. **b,** Representative widefield images for day 5 niPSC-derived blastoids. Scale bar=200 μm. **c**, immunofluorescence of naïve epiblast and trophectoderm marker genes in human blastoids. **d,** Circular plots showing the distribution of SNVs in the mitochondrial genome of niPSCs and single blastoid. **e**, Comparison of heteroplasmy levels of primary SNVs (VAF > 1%). No significant VAF shift is observed. **f**, Two rare amino acid-changing SNVs of m.6249G>A (p.Ala116Thr) and m.12700C>T (p.Leu122Phe) detected in the single blastoid but absent in niPSCs.

REFERENCES: 1 Bi, C., Wang, L., Fan, Y., Yuan, B., Alsolami, S., Zhang, Y., Zhang, P-Y., Huang, Y., Yu, Y., Belmonte, J.C.I. & Li, M. Quantitative haplotype-resolved analysis of mitochondrial DNA heteroplasmy in human single oocytes, blastoids and pluripotent stem cells. *Nucleic Acids Research* (2023). | article
2 Bi, C., Wang, L., Fan, Y., Yuan, B., Ramos-Mandujano, G., Zhang, Y., Alsolami, S., Zhou, X., Wang, J., Shao, Y., Reddy, P., Zhang, P-Y., Huang, Y., Belmonte, J.C.I. & Li, M. Single-cell individual full-length mtDNA sequencing by iMiGseq uncovers unexpected heteroplasmy shifts in mtDNA editing. *Nucleic Acids Research* (2023). | article