



Single-cell mtDNA heteroplasmy in colorectal cancer

João Almeida^a, Andrés Pérez-Figueroa^{a,b,c}, João M. Alves^{b,c}, Monica Valecha^{b,c},
Sonia Prado-López^{b,c,1}, Pilar Alvarino^{b,c}, José Manuel Cameselle-Teijeiro^{d,e},
Débora Chantada^f, Miguel M. Fonseca^a, David Posada^{b,c,g,*}

^a Interdisciplinary Centre of Marine and Environmental Research (CIIMAR), University of Porto, Portugal

^b CINBIO, Universidade de Vigo, 36310 Vigo, Spain

^c Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, Spain

^d Department of Pathology, Clinical University Hospital, Galician Healthcare Service (SERGAS), Santiago de Compostela, Spain

^e Medical Faculty, University of Santiago de Compostela, Santiago de Compostela, Spain

^f Department of Pathology, Hospital Álvaro Cunqueiro, Vigo, Spain

^g Department of Biochemistry, Genetics, and Immunology, Universidade de Vigo, 36310 Vigo, Spain

ARTICLE INFO

Keywords:

Single-cell mitochondrial genomics
mtDNA homoplasmy
Intracellular heteroplasmy
scDNA-seq

ABSTRACT

Human mitochondria can be genetically distinct within the same individual, a phenomenon known as heteroplasmy. In cancer, this phenomenon seems exacerbated, and most mitochondrial mutations seem to be heteroplasmic. How this genetic variation is arranged within and among normal and tumor cells is not well understood. To address this question, here we sequenced single-cell mitochondrial genomes from multiple normal and tumoral locations in four colorectal cancer patients. Our results suggest that single cells, both normal and tumoral, can carry various mitochondrial haplotypes. Remarkably, this intra-cell heteroplasmy can arise before tumor development and be maintained afterward in specific tumoral cell subpopulations. At least in the colorectal patients studied here, the somatic mutations in the single-cells do not seem to have a prominent role in tumorigenesis.

1. Introduction

As each human cell contains hundreds or thousands of mitochondria [34], wild-type and mutant mtDNA can co-exist in a state called *heteroplasmy*. Heteroplasmy levels can change within a given cell due to different processes like relaxed replication, degradation, de novo mutation, intercellular transfer, and recombination [15,33]. As mitochondria are randomly distributed to daughter cells during cell division, the levels of heteroplasmy among cells can also fluctuate over time [8] and modulate the potential phenotypic penetrance of associated diseases, including cancer [10,14,20,31–33,38]. Potentially, the proportion of mutant mtDNA in a tissue may drift toward fixation and reach homoplasmy. However, in cancer, the vast majority of the tumoral mtDNA mutations (>85%) in tissue samples seem to be heteroplasmic, with variant allele frequencies (VAFs) lower than 0.6 [41]. How the overall mtDNA heteroplasmy is structured among and within cells is unknown. A given amount of mtDNA heteroplasmy can be explained by a

population of cells carrying each a single but distinct mtDNA haplotype (*intercellular homoplasmy*) or by a population of more or less similar cells carrying several mtDNA haplotypes each (*intracellular heteroplasmy*). In other words, we do not know whether single cells typically have one or multiple mtDNA haplotypes [13]. Due to random drift and selection, within-cell homoplasmy is expected to take place in dividing cells eventually, but not necessarily in non-dividing cells [24].

On the other hand, the number of mtDNA molecules per cell fluctuates among human tissues [6]. In many types of cancer, tumor cells have fewer copies of mtDNA than the normal cells [20,26,37]. Variation in the number of mtDNA copies can serve as a potential biomarker for cancer, where a higher risk was associated with copy number variation [17,35]. However, how mtDNA copy number varies within and among cells has not yet been studied.

Finally, intratumor heterogeneity (ITH), particularly the spatial separation of somatic mutations within a tumor, is a determinant tumor characteristic in malignant growth, invasion, metastasis, and resistance

* Corresponding author at: CINBIO, Universidade de Vigo, 36310 Vigo, Spain.

E-mail address: dposada@uvigo.es (D. Posada).

¹ Current address: Institute of Solid State Electronics E362, Technische Universität Wien, Austria.

<https://doi.org/10.1016/j.ygeno.2022.110315>

Received 24 November 2021; Received in revised form 31 January 2022; Accepted 14 February 2022

Available online 16 February 2022

0888-7543/© 2022 The Authors.

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

acquisition [1,30]. Nonetheless, the levels of mtDNA ITH across different samples of the same tissue have not been studied yet at the single-cell level.

This study leveraged tumor multiregional single-cell whole-genome data from four colorectal cancer patients to overview how mtDNA mutation frequency and copy number are distributed among and within cells and across geographical space.

2. Results

2.1. Single-cell mtDNA coverage heterogeneity

The average sequencing depth, or coverage, across the mitochondrial genome, was $998\times$, $574\times$, $236\times$ and $881\times$ for CRC01, CRC07, CRC08, and CRC12, respectively. The coverage distribution was heterogeneous, with a consistent overrepresentation of particular regions across patients (Table S1, Fig. S1). The breadth of the coverage, i.e., the percentage of positions covered by at least one read, was 92–97%. Around 70% of the genomes were covered on average by at least 10 reads (only 57% for CRC08).

2.2. Single-cell mtDNA variants

We identified 3, 16, 23, and 20 mtDNA variants, primarily single nucleotide changes, in patients CRC01, CRC07, CRC08, and CRC12, respectively (Table S2). In CRC01, we found 3 germline variants. In CRC07, we identified 2 germline, 4 somatic, and 10 somatic/germline variants. In CRC08, we found 2 germline, 7 somatic, and 14 somatic/germline variants. Finally, in CRC12, we identified 13 germline and 7 somatic/germline variants. See Table S3 and Section 4.4 (Materials and Methods) for a definition of the somatic, germline, and somatic/germline categories.

2.3. Within-cell mtDNA heteroplasmy

We found no heteroplasmic variants in CRC01 (Fig. 1), but at least one in 60/105 cells (57.1%) in CRC07 (Fig. 2), 35/52 (67.3%) in CRC08 (Fig. 3), and 9/65 (13.8%) in CRC12 (Fig. 4). In CRC07, we observed within-cell heteroplasmy (VAF between 0.1 and 0.9) in 13/16 (81.3%) variant sites, and among-cell heteroplasmy (i.e., VAF less than 0.1 in some cells and more than 0.9 in others) in 8/16 (50%) (Fig. 2). At site 15,149, a somatic tumor variant appears at high frequency in the tumoral bulk and in specific tumor cells sampled from the distal and middle regions while is absent from the healthy cells. At site 6220, multiple normal and tumor cells showed a somatic variant at a shallow frequency –except for a normal cell, which shows a very high VAF– that was not detected in the corresponding bulk samples. At site 13,993, we see a similar pattern, with a few normal and tumor cells showing a fixed somatic variant. At site 10,704, a high-frequency variant was absent in some cells and fixed in others regardless of their normal/tumor status. At site 13,966, we observed a similar situation, although in this case, the variant was fixed in both normal and tumoral bulk samples.

In CRC08, we observed within-cell heteroplasmy in 17/23 (73.9%) variants and among-cell heteroplasmy in 8/23 (34.8%) (Fig. 3). At sites 64 and 664, the alternative allele only appears in tumoral cells sampled from the central region, reaching complete fixation at site 64. At site 15,059, all normal cells and the tumor cells sampled from the central area are homoplasmic for the reference allele, while most tumor cells from the proximal and distal regions show a high VAF. In CRC12, we observed within-cell heteroplasmy in 6/20 (30%) variants and among-cell heteroplasmy in 3/20 (15%) (Fig. 4). The variant at site 12,308 was absent in some cells and fixed in others regardless of their normal/tumor status.

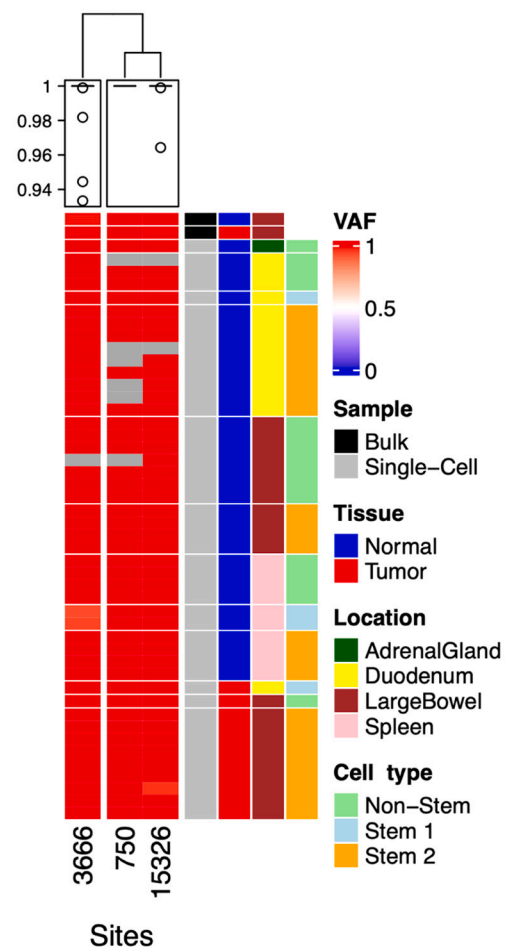


Fig. 1. Single-cell mtDNA variant allele frequencies for CRC01. The plot depicts the mtDNA variant allele frequencies (VAFs) for two bulk samples (on top, one normal and one tumoral) and 46 single-cells at three variable sites in patient CRC01. VAF values vary from blue to red, representing the extremes 0 and 1, and grey indicates missing data (no reads at that site). The four rightmost columns indicate which rows represent bulk or single-cell samples, normal or tumoral tissue, different anatomical locations, and distinct cell types. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.4. Functional impact of single-cell mtDNA variants

Heteroplasmic sites were significantly enriched in missense mutations in CRC07, stop gains in CRC08, and rRNA mutations in CRC12 (Table S4). The number of non-synonymous and synonymous mutations in normal and tumor cells was too small for a reliable estimation of the ratio of non-synonymous to synonymous rates (dN/dS) [22]. Non-synonymous variants did not have, overall, a higher VAF than synonymous variants in the tumors.

2.5. Single-cell mtDNA population structure

The F_{ST} statistic for cell differentiation was relatively high for CRC07, CRC08, and CRC12 (0.729, 0.667, and 0.878, respectively), but not for CRC01 (0.079). These F_{ST} values suggest that a larger fraction of the observed heteroplasmy results from differences among cells. Still, at the same time, there is a noticeable level of within-cell heteroplasmy, as we can appreciate in the VAF plots. The analysis of molecular variance (AMOVA) unveiled significant genetic differences among normal and

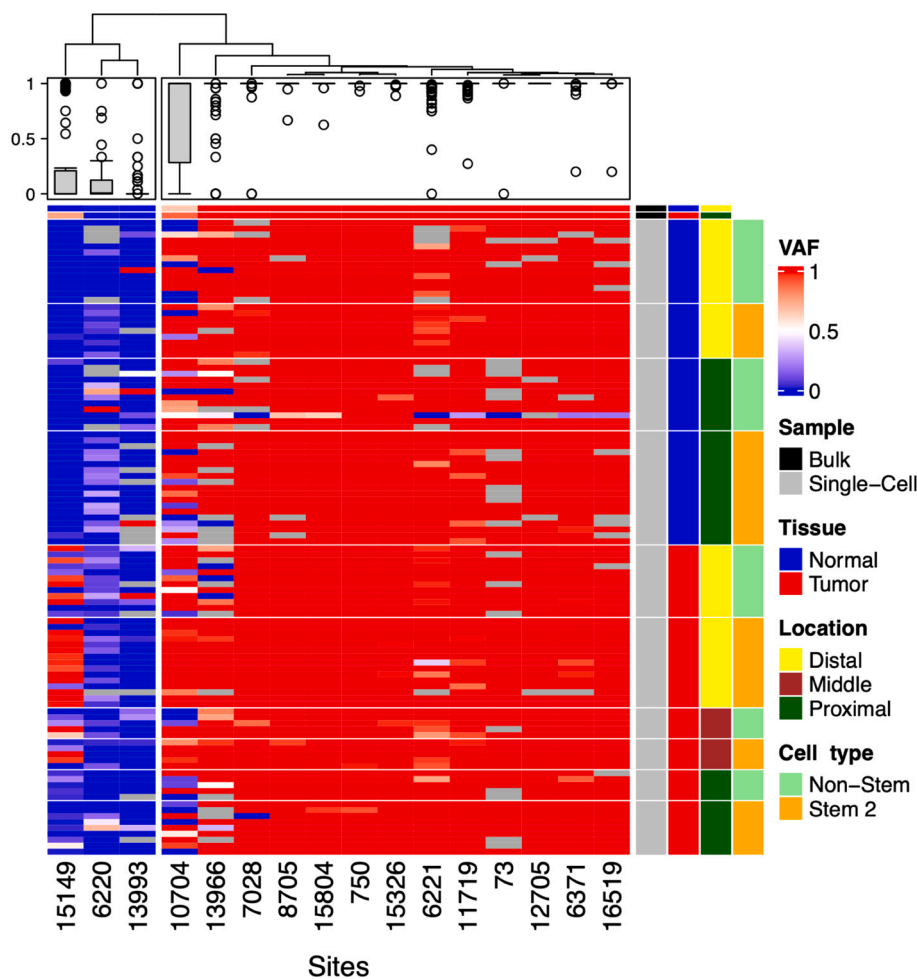


Fig. 2. Single-cell mtDNA variant allele frequencies for CRC07. The plot depicts the mtDNA variant allele frequencies (VAFs) for two bulk samples (on top, one normal and one tumoral) and 105 single-cells at 16 variable sites in patient CRC07. VAF values vary from blue to red, representing the extremes 0 and 1, and grey indicates missing data (no reads at that site). The four rightmost columns indicate which rows represent bulk or single-cell samples, normal or tumoral tissue, different anatomical locations, and distinct cell types. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tumor cell populations, among anatomical locations, and among cell types in CRC07 and among anatomical areas in CRC08 (Table 1). We could not run the AMOVA for patient CRC01 because of its low genetic variability.

In addition, we found significant pairwise genetic differences in CRC07 between all tumoral locations (Table S5) and between two of the tumoral areas and the normal locations. We also found significant differences in CRC08 between the normal proximal location and all three tumoral locations and between the central tumor location and the other two tumor locations. In CRC12, only the distance between the distal and proximal tumor locations was significant. Besides, we observed significant pairwise genetic differences in CRC07 among every cell type, except among tumor stem 2 (TS2) and tumor non-stem cells (TNS) (Table S6). In CRC08, we found significant differences between several non-stem and stem cell types in normal and tumor tissues. Finally, in CRC12, we only observed significant differences between TS2 and normal stem 2 (NS2) cells.

2.6. Single-cell mtDNA copy number variation

We estimated a statistically significant (p -value ≤ 0.05) higher number of mtDNA copies in tumor cells in patients CRC01 (16,507 copies on average in tumor cells versus 2197 in normal cells), CRC07 (302 vs. 181), and CRC12 (534 vs. 258), but not in CRC08 (135 vs 167;

p -value = 0.48) (Fig. 5). In CRC01, we observed a significantly higher number of mtDNA copies in the large bowel's normal cells than in the duodenum. In CRC07, we detected significantly more mtDNA copies in normal and tumor cells in the distal region than in the proximal region.

3. Discussion

In this study, we have leveraged single-cell whole-genome sequencing data to study the levels of mtDNA heteroplasmy within and between cells in four CRC patients. To our knowledge, this is the first study addressing mtDNA heteroplasmy at the single-cell level in normal and cancer tissues.

We identified a limited number of somatic mtDNA variants, consistent with previous studies [41]. Still, our results indicate that a single (normal or tumoral) cell can carry multiple mtDNA haplotypes and that these levels of intra-cell heteroplasmy can change from patient to patient.

Notably, at least in patients CRC07 and CRC08, some sites are heteroplasmic both in normal and tumor cells, suggesting that the genetic bottleneck produced during transformation (i.e., we assume that all the tumor cells descend from a single, ancestral tumor cell) does not necessarily eliminate the intra-cell mtDNA variation present in the transformed cell. In other words, this suggests not only that many mtDNA mutations do pre-date cancer itself [4] but also that these

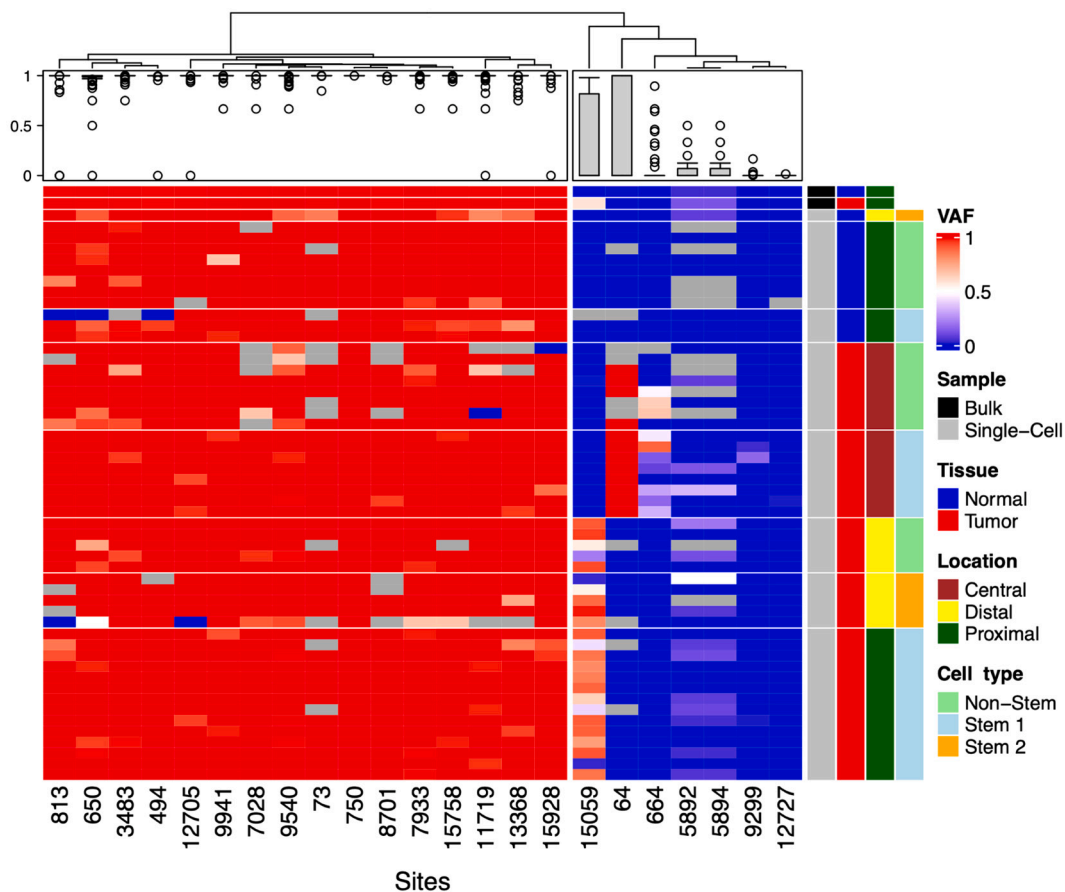


Fig. 3. Single-cell mtDNA variant allele frequencies for CRC08. The plot depicts the mtDNA variant allele frequencies (VAFs) for two bulk samples (on top, one normal and one tumoral) and 52 single cells at 23 variable sites in patient CRC08. VAF values vary from blue to red, representing the extremes 0 and 1, and grey indicates missing data (no reads at that site). The four rightmost columns indicate which rows represent bulk or single-cell samples, normal or tumoral tissue, different anatomical locations, and distinct cell types. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mutations can be heteroplasmic within the original tumor cell.

In our data set, heteroplasmy is mainly a result of differences among cells than within cells, as expected in a population of dividing cells where the mitochondrial population is subject to drift and, potentially, selection. Still, intracell heteroplasmy does occur. We observed some changes in within-cell heteroplasmy among normal and tumor cell populations, among distinct anatomical locations, or different cell types in some patients. In the VAF plots, we could identify clear groups of cells with consistent allele frequency changes at specific sites. For example, we observed three sites in which a variant was absent in the normal cells but frequent in at least some tumor regions (T15149C in CRC07 and at C64T and G15059A in CRC08). Curiously, in CRC08, these two variants reached fixation (or almost fixation) in different regions of the same tumor. Still, these VAF changes do not appear to have a clear functional relevance. Skonieczna et al. [29] found several mtDNA variants in a cohort of 100 CRC patients that were heteroplasmic in the normal tissue but homoplasmic in the tumors. Still, we did not observe cases like this, perhaps because of our limited sample size. Indeed, it is possible that we can fail to detect geographical structure because of lack of statistical power.

We estimated a higher mtDNA copy number in the tumoral than in normal cells. Despite mitochondrial copy numbers varying wildly within and across cancers [41], our numbers are consistent with the increased mtDNA copy number seen in several types of cancer [17,28,35,36].

In principle, mtDNA somatic mutations accumulate in a neutral, or nearly neutral, fashion [16,41]. In our single-cell datasets, non-synonymous variants did not reach a higher VAF than synonymous

variants, suggesting in any case that the variants we detected are not associated with tumor development.

4. Materials and methods

4.1. Sample collection

We obtained multiple tumoral and normal tissue samples from four colorectal cancer patients (Table S7). CRC01 samples were obtained during a warm autopsy, while the samples from the other patients were obtained from excess tumor tissue present in the colectomy specimens. All colorectal cancers were conventional adenocarcinomas (not otherwise specified), according to the criteria of the latest World Health Organization (WHO) digestive system tumors classification (Who Classification of Tumors Editorial Board [39]). Samples included in this study were provided by the Biobanks of the Health Research Institute of Santiago (PT13/0010/0068) and Galicia Sur Health Research Institute (B.0000802), both integrated into the Spanish National Biobank Network. Samples were processed following standard operating procedures with the approval of the Ethical and Scientific Committees (CAEI Galicia 2014/015). Written informed consents were provided by the patients or by their families.

4.2. Tumor disaggregation and sorting

We froze the tissue samples in liquid nitrogen, placed them in dry ice, and transported them to the laboratory. Next, we minced the samples

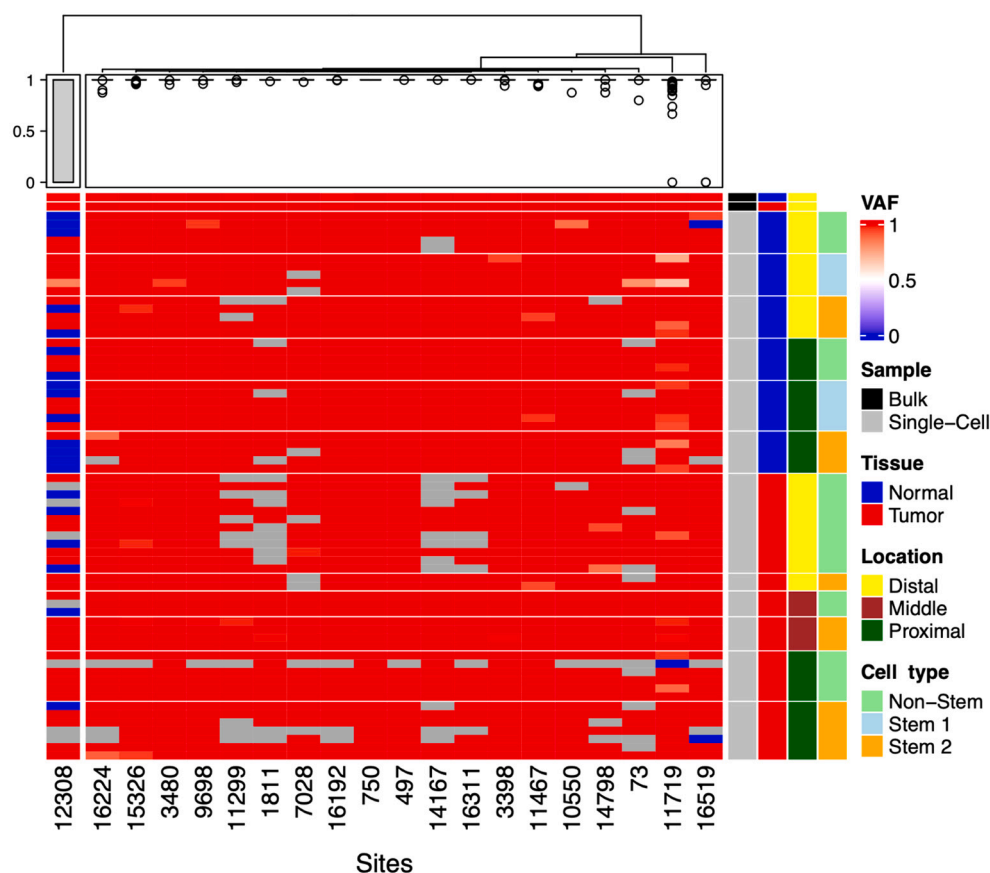


Fig. 4. Single-cell mtDNA variant allele frequencies for CRC12. The plot depicts the mtDNA variant allele frequencies (VAFs) for two bulk samples (on top, one normal and one tumoral) and 65 single cells at 20 variable sites in patient CRC12. VAF values vary from blue to red, representing the extremes 0 and 1, and grey indicates missing data (no reads at that site). The four rightmost columns indicate which rows represent bulk or single-cell samples, normal or tumoral tissue, different anatomical locations, and distinct cell types. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Analysis of Molecular Variance (AMOVA). P-values of the AMOVA for the Nei's genetic distances computed from the single-cell VAF values across all individual samples and for different hierarchical levels.

Group	CRC07	CRC08	CRC12
Tumor/Normal	0.000***	0.394	0.971
Location within Tumor/Normal	0.001***	0.000***	0.149
Cell type within Tumor/Normal	0.019*	0.055	0.535

* p-value ≤ 0.05.
*** p-value ≤ 0.001.

into pieces of 1 mm³ with a scalpel and digested by incubation in Accutase (LINUS) for 1 h at 37°C. After that, we filtered the cell suspension with a 70 µm cell strainer (FALCON) and assessed cell viability with Triptan Blue (Gibco). When the percentage of dead cells exceeded 30%, we did a Ficoll-Paque density gradient centrifugation to get rid of dead cells before sorting. We washed the cell pellets twice, suspended them in ice-cold phosphate-buffered saline (PBS), and stained them for 30 min with the following monoclonal antibodies: Anti-EpCAM (EBA1) (FITC)-conjugated, Anti-CD44 (APC)-conjugated, Anti-CD166 (PE)-conjugated, Anti-Lgr5 (VB 421)-conjugated. All antibodies were purchased from BD Biosciences. Following three successive washes in the PBS buffer, we added DRAQ5 and 7AAD dyes to select nucleated cells and exclude non-viable ones. We carried out flow cytometry analyses, and sorting of EpCAM+/CD44+/CD166-/Lgr5+ (tumoral stem 1 [TS1] and normal stem 1 [NS1] cells); EpCAM+/CD44+/CD166+/Lgr5- (tumoral stem 2 [TS2] and normal stem 2 [NS2]) and EpCAM+/CD44-/CD166-/Lgr5- (tumoral non-stem [TNS] and normal non-stem [NNS]) cell populations with a FACS ARIA III (BD Biosciences), and analyzed the data with the BD FACSDiva and Miltenyi Biotec Flowlogic software. The normal or tumoral status of the isolated cells was confirmed by looking

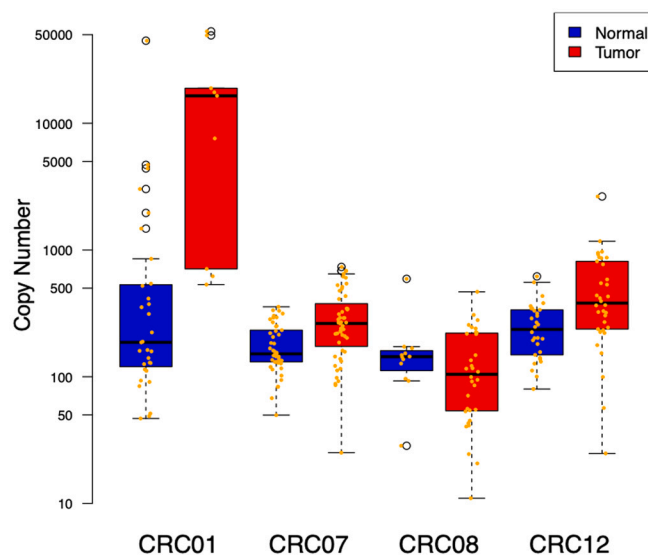


Fig. 5. Single-cell mtDNA copy number in normal and tumor cells. Copy number for some cells could not be calculated due to a lack of diploid regions (CRC01, n = 40; CRC07, n = 96; CRC08, n = 44; CRC12, n = 64).

at the copy-number profiles in the nuclear genome (data not shown). In total, we selected 268 cells for further analysis (46 for CRC01, 105 for CRC07, 52 for CRC08 and 65 for CRC12), from those, 133 (18 TS1 + 59 TNS + 56 TS2) were normal (healthy) and 135 (22 TS1 + 66 TNS + 47 TS) were tumor cells (Table S7).

4.3. Whole-genome single-cell and bulk sequencing

4.3.1. Single-cell whole genome amplification

To obtain enough DNA for sequencing from the individual cells, we carried out single-cell whole genome amplification (scWGA) with the Ampli1 Kit from Silicon Biosystems. To minimize potential contamination, we worked under a Biological Safety Cabinet, UV-irradiated all the plastic materials employed, and used a dedicated set of pipettes. In addition to patient cells, we included a positive (10 ng/μl REPLiG human control kit, QIAGEN) and negative (DNase/RNase free water) control in the amplification process. Next, we assessed the quality of the amplified DNA with the Ampli1 QC Kit and selected the positive samples for the 4 PCR DNA fragments. Then, we used the Ampli1 ReAmp/ds kit on the selected samples to increase the total double-stranded DNA. Later, we removed the kit adaptors adding 5 μl of NE Buffer 4 10× (New England Biolabs), 1 μl of *MseI* 50 U/μl (New England Biolabs), and 19 μl of nuclease-free water to every 25 μl of a sample. We introduced the resulting mix in a thermocycler. Next, we applied a program consisting of a step of 37 °C for 3 h, followed by 20 min at 65 °C for enzyme inactivation. Then, we purified the Ampli1 products using 1.8× AMPure XP beads (Agencourt, Beckman Coulter). Later, we measured DNA yield, integrity, and amplicon size distribution with a Qubit 3.0 fluorometer (ThermoFisher Scientific) and a 2200 TapeStation platform with the D5000 ScreenTape assay (Agilent Technologies).

4.3.2. Bulk genomic DNA (gDNA) isolation

We isolated the gDNA from the bulk samples using the QIAamp DNA Mini kit (QIAGEN) following the fabricant recommendations. Next, we estimated the yield and DNA integrity as described for the single cells above but using the Genomic DNA Screentape Assay instead.

4.3.3. Library construction and next-generation sequencing

We sent all samples to the Spanish National Center for Genomic Analysis (CNAG), where bulk and single-cell whole-genome sequencing libraries were built using the KAPA (Kapa Biosystems) library preparation kit with some modifications. Bulk and single-cell libraries were sequenced at ~ 39× (bulk) and ~ 6× (single cells) in an Illumina NovaSeq 6000 platform.

4.4. Mitochondrial variant calling

To avoid the detection of mtDNA mutations caused by pseudogenes or homologous sequences in nuclear DNA (NUMTs), we aligned the reads to both nuclear and mitochondrial reference genomes, treating NUMTs and mitochondrial specific mutations equally, and thus, artificially producing a high coverage on these nuclear positions while creating a deficit in the mtDNA ones. This method, which trades less interference of NUMTs in mtDNA for lack of detection power for those positions in the mitogenome, has already been used in mtDNA variant calling pipelines [12]. We mapped the reads to the Revised Cambridge Reference Sequence (rCRS) with SAMtools [18] to get the first set of reads ready for calling input.

We converted back mtDNA reads mapped to rCRS and unmapped reads into FASTQ, following existing pipelines [7], and remapped them to a “shifted” rCRS, for consideration on the circularity of the mitochondrial genome (mt-genome). This procedure avoids unaligned and discarded reads on the extremities of the reference resulting from the artificial breakpoint of the mt-genome on the replication control region, defining the “start” at position 1 and “end” at position 16,569. In this double-alignment method [7], the “shifted” reference is created by switching the positions of, roughly, the first and second half of the rCRS so that positions 8001–16,569 appear first, followed by positions 1–8000.

We applied quality control filters to avoid calling errors by selecting reads with a base quality score ≥ 20 , median depth ≥ 100 per individual, raw depth ≥ 40 , and depth after base quality score filter ≥ 10 , MAF \geq

4%, as in Ding et al. [7]. For variant calling, we used Mutect2 on multi-sample mode [5] and with option `-max-mnp-distance 0`. The resulting variants were filtered with `FilterMutectCalls`. We performed the variant calling using all the bulk and single-cell samples together. We used the reads mapped to rCRS to call the variants for coordinates 4000–12,000 and reads mapped to the “shifted” rCRS to call for coordinates 0–4000 and 12,000–16,000. Then, we merged the two sets of called variants. Finally, we removed multiallelic variants, read depth 1, and variants with more than 20% missing data.

We classified the variants into somatic, germline, or somatic/germline, and as tumor or normal (Table S3). Germline variants are those fixed for the alternative allele in all bulk or single-cell samples. Somatic variants are those present in some single-cells, and not fixed in the bulk samples. Germline/somatic variants are germline variants –therefore fixed for the alternative allele in the normal bulk– that also display the reference allele in some cells.

4.5. Estimation of the mtDNA copy number

We used fastMitoCalc [7,25] to estimate the mtDNA copy number as the depth ratio between mtDNA and nuclear DNA:

$$\text{mtDNA copy number} = \frac{\text{average number of mtDNA reads}}{\text{average number of DNA reads}} \times P$$

where P is the ploidy. This procedure assumes that regions of the genome of equal ploidy have the same depth [3,6,7,12,26,27,42,43]. For normal diploid cells, P is 2; however, cancer cells can exhibit large-scale genomic amplifications and deletions, altering their ploidy [26]. Therefore, we calculated the mtDNA copy number using only diploid nuclear regions [3,42]. We called single-cell copy-number variants (CNVs) with Ginkgo [11] to identify these regions using variable-length bins of around 500 kb. After binning, data for each cell was normalized and segmented using default parameters.

4.6. Identification of within-cell heteroplasmy

We calculated the variant allele frequency (VAF) by dividing the allelic depth of the alternative allele by the total read depth. We classified as somatic those variants present in at least one single (normal or tumor) cell and absent or not fixed in the normal bulk, as germline those variants fixed in the normal bulk. Variants fixed in the normal bulk and present in at least one single (normal or tumoral) cell but not fixed were classified as germline and somatic. We consider a variant to be heteroplasmic for a given cell if its VAF was between 0.1 and 0.9. Otherwise we classified it as homoplasmic ($0.1 > \text{VAF} > 0.9$).

4.7. Population genetic structure

To quantify the structuration of heteroplasmy within and among cells, we computed the haploid equivalent of the F_{ST} statistic [40], whose values go from 0 to 1. An F_{ST} of 0 will indicate that all heteroplasmy occurs because of differences among cells. A value of 1 will suggest that all heteroplasmy occurs because of differences within cells.

In addition, we used Nei’s [21] distances from the VAFs with an in-house R script (<https://github.com/anpefi/sc-mtDNA>) to describe the mitochondrial population structure at the single-cell level. We carried out an analysis of molecular variance (AMOVA) [9] using the R package *pegas* [23] across different hierarchical levels: tumor vs. normal, location within tumor/normal, and cell type within tumor/normal. Additionally, a pairwise AMOVA, using the same method as the hierarchical AMOVA, was applied to the different groups at the location per tissue and cell type per tissue level. To evaluate the null hypothesis of no population structure, we used permutation tests with 10,000 replicates.

4.8. Functional analyses

To look for differences between the functional characterization of tumor and normal cells, we looked for alternatives to “typically normal allele” homoplasmic variants. We identified these “typically normal allele” homoplasmic variants as the variants that differ from those considered as individually fixed in normal cells and distributed them into functional groups: synonymous, missense, or stop gain, for protein-coding, and intergenic or rRNA for non-protein-coding. We annotated and prioritized the variants using Ensemble’s Variant Effect Predictor (VEP) [19]. In addition, we annotated each variant using the online SNV query of MITOMASTER [2].

Data availability

Data and scripts are available at <https://github.com/anepfi/sc-mtDNA>.

CRediT authorship contribution statement

João Almeida: Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Andrés Pérez-Figueroa:** Methodology, Formal analysis, Writing – review & editing, Visualization. **João M. Alves:** Data curation. **Monica Valecha:** Data curation. **Sonia Prado-López:** Methodology, Investigation. **Pilar Alvarino:** Investigation. **José Manuel Cameselle-Teijeiro:** Resources. **Débora Chantada:** Resources. **Miguel M. Fonseca:** Methodology, Writing – review & editing, Supervision, Funding acquisition. **David Posada:** Conceptualization, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

This work was supported by project No. 32030 awarded to M.F., co-financed by COMPETE 2020, Portugal 2020 and the EU through the ERDF, and by FCT (PTDC/BIA-EVL/32030/2017) through national funds, and by the European Research Council (ERC-617457-PHYLO-CANCER awarded to D.P.) and Spanish Ministry of Science, Innovation, and Universities - MCIU (PID2019-106247GB-I00 awarded to D.P.). J.M. C.-T. is supported by grant no. ISCIII-PI19/01316 from Instituto de Salud Carlos III, State Research Agency and Ministry of Science and Innovation (Spain), with the participation of the European FEDER fund. We want to thank Sara Rocha for her helpful comments on this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ygeno.2022.110315>.

References

- [1] P.L. Bedard, A.R. Hansen, M.J. Ratain, L.L. Siu, Tumour heterogeneity in the clinic, *Nature* 501 (2013) 355–364.
- [2] M.C. Brandon, E. Ruiz-Pesini, D. Mishmar, V. Procaccio, M.T. Lott, K.C. Nguyen, S. Spolim, U. Patil, P. Baldi, D.C. Wallace, MITOMASTER: a bioinformatics tool for the analysis of mitochondrial DNA sequences, *Hum. Mutat.* 30 (2009) 1–6.
- [3] N. Cai, Y. Li, S. Chang, J. Liang, C. Lin, X. Zhang, L. Liang, J. Hu, W. Chan, K. S. Kendler, et al., Genetic control over mtDNA and its relationship to major depressive disorder, *Curr. Biol.* 25 (2015) 3170–3177.
- [4] P.F. Chinnery, D.C. Samuels, J. Elson, D.M. Turnbull, Accumulation of mitochondrial DNA mutations in ageing, cancer, and mitochondrial disease: is there a common mechanism? *Lancet* 360 (2002) 1323–1325.
- [5] K. Cibulskis, M.S. Lawrence, S.L. Carter, A. Sivachenko, D. Jaffe, C. Sougnez, S. Gabriel, M. Meyerson, E.S. Lander, G. Getz, Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples, *Nat. Biotechnol.* 31 (2013) 213–219.
- [6] A.M. D’Erchia, A. Atlante, G. Gadaleta, G. Pavesi, M. Chiara, C. De Virgilio, C. Manzari, F. Mastropasqua, G.M. Prazzoli, E. Picardi, et al., Tissue-specific mtDNA abundance from exome data and its correlation with mitochondrial transcription, mass and respiratory activity, *Mitochondrion* 20 (2015) 13–21.
- [7] J. Ding, C. Sidore, T.J. Butler, M.K. Wing, Y. Qian, O. Meirelles, F. Busonero, L. C. Tsoi, A. Maschio, A. Angius, et al., Assessing mitochondrial DNA variation and copy number in lymphocytes of ~2,000 Sardinians using tailored sequencing analysis tools, *PLoS Genet.* 11 (2015) 1–18.
- [8] J.L. Elson, D.C. Samuels, D.M. Turnbull, P.F. Chinnery, Random intracellular drift explains the clonal expansion of mitochondrial DNA mutations with age, *Am. J. Hum. Genet.* 68 (2001) 802–806.
- [9] L. Excoffier, P.E. Smouse, J.M. Quattro, Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data, *Genetics* 131 (1992) 479–491.
- [10] L. Fendt, F. Fazzini, H. Weissensteiner, E. Bruckmoser, S. Schönherr, G. Schäfer, J. L. Losso, G.A. Streiter, C. Lamina, M. Rasse, et al., Profiling of mitochondrial DNA heteroplasmy in a prospective oral squamous cell carcinoma study, *Cancers* 12 (2020), <https://doi.org/10.3390/cancers12071933>.
- [11] T. Garvin, R. Aboukhalil, J. Kendall, T. Baslan, G.S. Atwal, J. Hicks, M. Wiegler, M. C. Schatz, Interactive analysis and assessment of single-cell copy-number variations, *Nat. Methods* 12 (2015) 1058–1060.
- [12] Y. Guo, J. Li, C.I. Li, Y. Shyr, D.C. Samuels, MitoSeek: extracting mitochondrial information and performing high-throughput mitochondria sequencing analysis, *Bioinformatics* 29 (2013) 1210–1211.
- [13] Y. He, J. Wu, D.C. Dressman, C. Iacobuzio-Donahue, S.D. Markowitz, V. E. Velculescu, L.A. Diaz Jr., K.W. Kinzler, B. Vogelstein, N. Papadopoulos, Heteroplasmic mitochondrial DNA mutations in normal and tumour cells, *Nature* 464 (2010) 610–614.
- [14] J.F. Hopkins, V.Y. Sabelnykova, J. Weischenfeldt, R. Simon, J.A. Aguiar, R. Alkallas, L.E. Heisler, J. Zhang, J.D. Watson, M.L.K. Chua, et al., Mitochondrial mutations drive prostate cancer aggression, *Nat. Commun.* 8 (2017) 656.
- [15] I.G. Johnston, J.P. Burgstaller, Evolving mtDNA populations within cells, *Biochem. Soc. Trans.* 47 (2019) 1367–1382.
- [16] Y.S. Ju, L.B. Alexandrov, M. Gerstung, I. Martincorena, S. Nik-Zainal, M. Ramakrishna, H.R. Davies, E. Papaemmanuil, G. Gundem, A. Shlien, et al., Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer, *Elife* 3 (2014), <https://doi.org/10.7554/eLife.02935>.
- [17] Q. Lan, U. Lim, C.S. Liu, S.J. Weinstein, S. Chanock, M.R. Bonner, J. Virtamo, D. Albanes, N. Rothman, A prospective study of mitochondrial DNA copy number and risk of non-Hodgkin lymphoma, *Blood* 112 (2008) 4247–4249.
- [18] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map format and SAMtools, *Bioinformatics* 25 (2009) 2078–2079.
- [19] W. McLaren, L. Gil, S.E. Hunt, H.S. Riat, G.R.S. Ritchie, A. Thormann, P. Flicek, F. Cunningham, The ensembl variant effect predictor, *Genome Biol.* 17 (2016) 122.
- [20] S. McMahon, T. LaFramboise, Mutational patterns in the breast cancer mitochondrial genome, with clinical correlates, *Carcinogenesis* 35 (2014) 1046–1054.
- [21] M. Nei, Genetic distance between populations, *Am. Nat.* 106 (1972) 283–292, <https://doi.org/10.1086/282771>.
- [22] M. Nei, T. Gojbori, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions, *Mol. Biol. Evol.* 3 (1986) 418–426.
- [23] E. Paradis, pegas: an R package for population genetics with an integrated-modular approach, *Bioinformatics* 26 (2010) 419–420.
- [24] C.J. Pérez-Amado, A. Bazan-Cordoba, A. Hidalgo-Miranda, S. Jiménez-Morales, Mitochondrial heteroplasmy shifting as a potential biomarker of cancer progression, *Int. J. Mol. Sci.* 22 (2021), <https://doi.org/10.3390/ijms22147369>.
- [25] Y. Qian, T.J. Butler, K. Opsahl-Ong, N.S. Giroux, C. Sidore, R. Nagaraja, F. Cucca, L. Ferrucci, G.R. Abecasis, D. Schlessinger, et al., fastMitoCalc: an ultra-fast program to estimate mitochondrial DNA copy number from whole-genome sequences, *Bioinformatics* 33 (2017) 1399–1401.
- [26] E. Reznik, M.L. Miller, Y. Şenbabaoglu, N. Riaz, J. Sarungbam, S.K. Tickoo, H.A. Al-Ahmadie, W. Lee, V.E. Seshan, A.A. Hakimi, et al., Mitochondrial DNA copy number variation across human cancers, *Elife* 5 (2016) 1–20.
- [27] D.C. Samuels, L. Han, J. Li, S. Quanguo, T.A. Clark, Y. Shyr, Y. Guo, Finding the lost treasures in exome sequencing data, *Trends Genet.* 29 (2013) 593–599.
- [28] J. Shen, M. Platek, A. Mahasneh, C.B. Ambrosone, H. Zhao, Mitochondrial copy number and risk of breast cancer: a pilot study, *Mitochondrion* 10 (2010) 62–68.
- [29] K. Skonieczna, B. Malyarchuk, A. Jawień, A. Marszałek, Z. Banaszekiewicz, P. Jarmocik, T. Grzybowski, Mitogenomic differences between the normal and tumor cells of colorectal cancer patients, *Hum. Mutat.* 39 (2018) 691–701.
- [30] G. Stanta, S. Bonin, Overview on clinical relevance of intra-tumor heterogeneity, *Front. Med.* 5 (2018) 85.
- [31] G.B. Stefano, C. Bjerning, F. Wang, N. Wang, R.M. Kream, in: G. Santulli (Ed.), *Mitochondrial Heteroplasmy BT - Mitochondrial Dynamics in Cardiovascular Medicine*, Springer International Publishing, Cham, 2017, pp. 577–594.
- [32] G.B. Stefano, R.M. Kream, Mitochondrial DNA heteroplasmy in human health and disease (review), *Biomed. Rep.* 4 (2016) 259–262.
- [33] J.B. Stewart, P.F. Chinnery, The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease, *Nat. Rev. Genet.* 16 (2015) 530–542.
- [34] J.B. Stewart, P.F. Chinnery, Extreme heterogeneity of human mitochondrial DNA from organelles to populations, *Nat. Rev. Genet.* 22 (2021) 106–118.
- [35] B. Thyagarajan, R. Wang, H. Nelson, H. Barcelo, W.P. Koh, J.M. Yuan, Mitochondrial DNA copy number is associated with breast cancer risk, *PLoS One* 8 (2013), <https://doi.org/10.1371/journal.pone.0065968>.
- [36] S.K. Tickoo, M.W. Lee, J.N. Eble, M. Amin, T. Christopherson, R.J. Zarbo, M. B. Amin, Ultrastructural observations on mitochondria and microvesicles in renal oncocytoma, chromophobe renal cell carcinoma, and eosinophilic variant of conventional (clear cell) renal cell carcinoma, *Am. J. Surg. Pathol.* [Internet] 24

- (2000). https://journals.lww.com/ajsp/Fulltext/2000/09000/Ultrastructural_Observations_on_Mitochondria_and.8.aspx.
- [37] L.-M. Tseng, P.-H. Yin, C.-W. Chi, C.-Y. Hsu, C.-W. Wu, L.-M. Lee, Y.-H. Wei, H.-C. Lee, Mitochondrial DNA mutations and mitochondrial DNA depletion in breast cancer, *Genes Chromosom. Cancer* 45 (2006) 629–638.
- [38] D.C. Wallace, D. Chalkia, Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease, *Cold Spring Harb. Perspect. Biol.* 5 (2013) 1–47.
- [39] Who Classification of Tumours Editorial Board, in: F. Bosman, F. Carneiro, R. Hruban, N. Theise (Eds.), WHO Classification of Tumours: Digestive System Tumours (WHO Classification of Tumours, 1), World Health Organization, Lyon, 2019.
- [40] S. Wright, The genetical structure of populations, *Ann. Eugenics* 15 (1951) 323–354.
- [41] Y. Yuan, Y.S. Ju, Y. Kim, J. Li, Y. Wang, C.J. Yoon, Y. Yang, I. Martincorena, C. J. Creighton, J.N. Weinstein, et al., Comprehensive molecular characterization of mitochondrial genomes in human cancers, *Nat. Genet.* 52 (2020) 342–352.
- [42] P. Zhang, B.D. Lehmann, D.C. Samuels, S. Zhao, Y.Y. Zhao, Y. Shyr, Y. Guo, Estimating relative mitochondrial DNA copy number using high throughput sequencing data, *Genomics* 109 (2017) 457–462.
- [43] J. Zhao, J. Zhang, M. Yu, Y. Xie, Y. Huang, D.W. Wolff, P.W. Abel, Y. Tu, Mitochondrial dynamics regulates migration and invasion of breast cancer cells, *Oncogene* 32 (2013) 4814–4824.