

Single-molecule mitochondrial DNA sequencing shows no evidence of CpG methylation in human cells and tissues

Iacopo Bicci^{1,2,†}, Claudia Calabrese^{1,2,†}, Zoe J. Golder^{1,2}, Aurora Gomez-Duran^{1,2,3} and Patrick F. Chinnery^{1,2,*}

¹MRC-Mitochondrial Biology Unit, The Keith Peters Building, Cambridge CB2 0XY, UK, ²Department of Clinical Neurosciences, University of Cambridge, Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0XY, UK and ³Centro de Investigaciones Biológicas Margarita Salas. Spanish National Research Council, Madrid, Spain

Received September 16, 2021; Revised October 29, 2021; Editorial Decision November 05, 2021; Accepted November 18, 2021

ABSTRACT

Methylation on CpG residues is one of the most important epigenetic modifications of nuclear DNA, regulating gene expression. Methylation of mitochondrial DNA (mtDNA) has been studied using whole genome bisulfite sequencing (WGBS), but recent evidence has uncovered technical issues which introduce a potential bias during methylation quantification. Here, we validate the technical concerns of WGBS, and develop and assess the accuracy of a new protocol for mtDNA nucleotide variant-specific methylation using single-molecule Oxford Nanopore Sequencing (ONS). Our approach circumvents confounders by enriching for full-length molecules over nuclear DNA. Variant calling analysis against showed that 99.5% of homoplasmic mtDNA variants can be reliably identified providing there is adequate sequencing depth. We show that some of the mtDNA methylation signal detected by ONS is due to sequence-specific false positives introduced by the technique. The residual signal was observed across several human primary and cancer cell lines and multiple human tissues, but was always below the error threshold modelled using negative controls. We conclude that there is no evidence for CpG methylation in human mtDNA, thus resolving previous controversies. Additionally, we developed a reliable protocol to study epigenetic modifications of mtDNA at single-molecule and single-base resolution, with potential applications beyond CpG methylation.

INTRODUCTION

Cytosine methylation is an epigenetic modification of nuclear DNA (nDNA) that can regulate gene expression during development (1) and throughout life (2), but the presence of CpG methylation on the mitochondrial genome (mtDNA) is a matter of debate (3–5). This is an important issue to resolve given the pivotal role of mtDNA in cellular metabolism (6).

Whole genome bisulfite sequencing (WGBS) is the gold standard technique for detecting methylation across the nuclear genome (7–9), where sequencing before and after the chemical conversion of unmethylated cytosine to uracil allows the degree of methylation to be measured at single-base resolution. WGBS studies have reported methylation patterns across the mtDNA molecule in different biological contexts (10). However, recent studies suggest that these are influenced by technical artefacts (3–5,11). MtDNA has a purine-rich ‘Heavy’ (H-) and a pyrimidine-rich ‘Light’ (L-) strand (12), leading to a disproportionate fragmentation of the cytosine-rich L-strand by bisulfite treatment (13). Moreover, the presence of multiple mtDNA genotypes within mitochondria of the same cell (heteroplasmy (14)), and nuclear sequences originated from the mtDNA (NuMTs (15,16)) are potential confounders for mtDNA methylation detection.

To overcome these limitations, we set out to quantify CpG methylation of native mtDNA using long-read based Oxford Nanopore Sequencing (ONS) technology (17). We developed a new protocol enabling the assessment of methylation at nucleotide-level resolution (18), to study multiple human cell lines and in human tissues. We show that the population nucleotide sequence variants introduce artifacts giving the impression of mtDNA methylation which can be removed using an individual-specific mtDNA reference sequence. We also show that residual apparent low levels of mtDNA methylation fall below the detection threshold for

*To whom correspondence should be addressed. Email: pfc25@cam.ac.uk

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

ONS, providing independent evidence that significant levels of mtDNA methylation is unlikely to occur *in vivo*.

MATERIALS AND METHODS

Cell culture and DNA extraction from human cell lines

Cell lines used in this study are listed in Table 1. Cells were maintained in fibroblast medium [DMEM high glucose (Gibco) with 10% foetal bovine serum (Gibco) and no antibiotics] at 37°C in a humidified 5% CO₂ atmosphere. Cells were grown until ~80% confluence. When ready, cells were washed with PBS (Gibco), then incubated with 0.05% trypsin (Gibco) for 5 min at 37°C. Cells were collected by centrifugation (1500 rcf for 5 min) and pellets were washed once with PBS, before being snap frozen in liquid nitrogen and kept at -20°C until further use. All DNA from cell lines was extracted from snap-frozen pellets using the QIAmp DNA mini kit (QIAGEN) following the manufacturer's instructions. DNA was quantified using the Qubit dsDNA kit (Invitrogen) following the manufacturer's instructions.

DNA extraction from human tissues

Tissues used in this study are listed in Table 1. Tissues were obtained from 7 different healthy individuals. All DNA from human tissues was extracted using QIAmp Fast DNA Tissue Kit (QIAGEN), following the manufacturer's instructions. DNA was quantified using the Qubit dsDNA kit (Invitrogen) following the manufacturer's instructions.

Long-range polymerase reactions (LR-PCR)

LR-PCR amplification reaction was performed using PrimeSTAR GXL DNA Polymerase kit (Takara) according to manufacturer's instructions. The primers used are detailed in Supplementary Table S1. Product length encompasses most part of the mtDNA sequence. Amplification reactions were performed using the following cycling conditions: 94°C for 1 min, followed by 30 cycles of 98°C for 10 s, 55°C for 15 s and 68°C for 10 min.

Generation of negative and positive controls

Untreated LR-PCR amplicons were used as negative controls for methylation. To generate positive controls, the same amplicons were treated *in vitro* with the recombinant CpG methyltransferase M.SssI (NEB). Briefly, 1 µg of amplicon DNA per 50 µl reaction was treated for 4 h at 37°C with 50 units of M.SssI in the presence of 1× NEB buffer #2 and 160 µM of S-adenosylmethionine (SAM). To test the efficiency of the M.SssI reaction, 10 units of methylation-sensitive restriction enzyme BstUI were added at the end of the incubation. This was followed by a further incubation at 60°C for 1 h. Protection of the M.SssI-treated amplicons from BstUI digestion was assessed using the Genomic DNA ScreenTape System (Agilent) on an Agilent 2200 TapeStation platform following manufacturer's instructions (data not shown). Supplementary Figure S1. To generate positive controls with intermediate methylation levels, we mixed negative and positive controls according to Supplementary Table S2.

Mitochondrial DNA enrichment for single-molecule sequencing

1 µg of genomic DNA (nuclear + mitochondrial DNA) per 50 µl reactions was digested with 40 units of the recombinant restriction enzyme BamHI-HF (NEB) for 1 h at 37°C in the presence of CutSmart buffer (NEB). To achieve combined DNA purification and selection of high molecular weight fragments, DNA was purified using Monarch[®] PCR & DNA Cleanup Kit (NEB), using the following recommended protocol modification: 20 µl of elution buffer was heated to 50°C before the last elution step.

Quantification of mtDNA levels using ddPCR

ddPCR was used to quantify relative mtDNA enrichment following BamHI-HF (NEB) treatment of control DNA. To quantify relative mtDNA copy number, a mitochondrial and nuclear target (the genes *MT-ND1* and *RNASE P*, respectively) were amplified and fluorescent signal was generated using the primers and probes detailed in the Supplementary Table S1. ddPCR protocol was performed following manufacturer's instructions. Briefly, PCR reaction master mix was prepared in 1x (final concentration) ddPCR Supermix for Probes (no dUTP, BioRad), by adding 300nM of each primer and 200 nM of each probe in 19 µl final volume. 1 ng of sample DNA was then added to the mastermix. Droplets were generated using an Automated Droplet Generation instrument (BioRad) and were then subjected to PCR amplification, performed using the following cycling conditions: 95°C for 10 min, followed by 39 cycles of 94°C for 30 s and 58°C for 1 min, followed by a final stabilization step at 98°C for 10 min. Droplets were then loaded into a QX200 droplet reader (BioRad) and analysed using an absolute quantification protocol (ABS) to measure the absolute copy number of each probe. Droplet analysis was performed using the QuantaSoft analysis software (BioRad).

ONS library preparation and sequencing on the MinION instrument

Approximately 1 µg of native genomic DNA or purified LR-PCR amplicons were prepared for ONS sequencing on R9.4.1 flow cells using the Ligation Sequencing Kit SQK-LSK109 (Nanoporetech), in combination with the Native Barcoding Expansion Kit EXP-NBD114 (Nanoporetech). Genomic DNA was fragmented either through BamHI digestion (Materials and Methods) or sheared to 10 kb using g-tubes (Covaris), following manufacturers' instructions. Simultaneous DNA repairing, end-repairing and dA-tailing was achieved using the NEBNext FFPE Repair Mix (NEB) and the Ultra II end-repair module (NEB). Barcodes were ligated to individual samples using Blunt/TA Ligase Master Mix (NEB). Samples were then combined and AMII adapters containing the motor proteins needed for sequencing were ligated using NEBNext[®] Quick Ligation Module (NEB). AMPure XP beads (Beckman Coulter) at a concentration of 1x, 1x and 0.5x, respectively, were used to purify DNA between the library preparation steps. Final libraries were loaded onto R9.4.1 flow cells and samples were sequenced using a single MinION Mk 1B. To keep

Table 1. List of cells and tissues used in this study

Cell line studied					
Cell line description	Code	Used for	ONS protocol		
Human cybrid cell line - H haplogroup	613H	ONS library preparation, variant calling, methylation analysis	Fragmentation, BamHI-based		
Human cybrid cell line - J haplogroup	128J	ONS library preparation, variant calling, methylation analysis	Fragmentation, BamHI-based		
Human cybrid cell line - J2 haplogroup	135J2	ONS library preparation, variant calling, methylation analysis	Fragmentation, BamHI-based		
Human primary fibroblast cell line - Control	Control 1	Variant calling, methylation analysis	BamHI-based		
Human primary fibroblast cell line - Control	Control 2	Variant calling, methylation analysis	BamHI-based		
Human primary fibroblast cell line - MELAS mutation	m.3243A>G (1)	Variant calling, methylation analysis	BamHI-based		
Human primary fibroblast cell line - MELAS mutation	m.3243A>G (2)	Variant calling, methylation analysis	BamHI-based		
Human primary fibroblast cell line - MERRF cell line	m.8344A>G	Variant calling, methylation analysis	BamHI-based		
Human tissues studied					
Tissue type	ID code	Source	Gender/age	Used for	ONS protocol
Human Liver	TB15-0139	Addenbrooke's Tissue Bank	Male/36 years	Methylation analysis	BamHI-based
Kidney	TB12-1905	Addenbrooke's Tissue Bank	Male/60 years	Methylation analysis	BamHI-based
Human Kidney	TB15-153	Addenbrooke's Tissue Bank	Male/75 years	Methylation analysis	BamHI-based
Heart	TB12-2860	Addenbrooke's Tissue Bank	Male/28 years	Methylation analysis	BamHI-based
Skeletal Muscle	TB15-2606	Addenbrooke's Tissue Bank	Male/56 years	Methylation analysis	BamHI-based
Skeletal Muscle	TB13-1505	Addenbrooke's Tissue Bank	Male/40 years	Methylation analysis	BamHI-based
Skeletal Muscle	TB05-0578	Addenbrooke's Tissue Bank	Male/82 years	Methylation analysis	BamHI-based

the sequencing throughput consistent, six biological samples were always pooled together and sequenced for 24 h. LR-PCR amplicons were pooled together and sequenced for 6 h.

Illumina Miseq library preparation and sequencing

MiSeq libraries were prepared from genomic DNA by amplification of the mitochondrial DNA in two overlapping fragments, using the primers shown in Supplementary Table S1. Amplicons were individually purified, quantified, and then were pooled in equal amounts from each sample. Libraries were prepared using NEBNext Ultra library prep reagents (NEB) according to manufacturer's instructions and sequenced using a 2 × 250-cycle MiSeq Reagent kit v3.0 (Illumina, CA).

WGBS data analysis

Raw WGBS experiments part of the Roadmap Epigenome Project (19) were downloaded from the GEO Database. Downloaded files from single-ended WGBS sequencing experiments were converted from SRA format to fastq files using `fastq-dump` (Key Resources Table) with the following options: `--readids --skip-technical -W --read-filter pass --gzip`. Read quality of the converted fastq files was assessed with FastQC v0.11.5 (20). All of the reports generated from FastQC were manually checked to determine whether a trimming of low-quality reads and/or adapters was needed. Where trimming was deemed necessary, TrimGalore! v0.4.5 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used. The software automatically trims adapter sequences from the reads (if present) and retains those with an average Phred quality score ≤ 20 (before and/or after trimming). Reads shorter than 45 bp after trimming were discarded using `--length` option. Upon quality check and trimming, both alignment of the WGBS fastq files to the reference human genome sequence (GRCh38) and extraction of the methylation information were carried out

with bowtie2 v2.3.2 (21) and Bismark v0.19.0 (22), respectively. Coverage was calculated from BAM files using `samtools depth`. This was defined as the percentage of mtDNA genome in each strand covered by at least 5 reads. Methylation extraction was carried out using the `bismark_methylation_extractor` package with the following options: `--comprehensive --merge_non_CpG --gzip --bedGraph --CX_context`, but only CpG residues were considered for further analyses.

ONS data analysis

Base-calling of fast5 files containing raw electric current information was performed by the `guppy_basecaller` package of Guppy v3.2.2+9fe0a78 (Nanoporetech). Base-called, barcoded reads were de-multiplexed into individual samples using the `guppy_barcode` package of Guppy v3.2.2+9fe0a78 (Nanoporetech). In order to simultaneously enrich for linear full-length mitochondrial sequences, exclude ligation artifacts and minimise the presence of NuMTs, we applied a stringent filter on read sequence length (min: 4000 bp, max: 17 000 bp) and quality (Phred quality score ≥ 9) using NanoFilt v2.2.0 (23) on the bar-coded fastq files.

Minimap2 v2.10-r761 (24) with the `-x map-ont` option was used to perform the alignment of Nanopore reads onto the GRCh38 reference (which includes the mitochondrial rCRS reference sequence, NC_012920.1), and the option `-secondary = no` was used to exclude secondary alignments in the BAM output. Because minimap2 does not recognise circular reference sequences, reads spanning the D-loop are reported as supplementary alignments in the output BAM files. For this reason, we included in the final set of aligned reads also supplementary alignments aligning onto the mtDNA reference and spanning the D-loop, but only if they aligned in the same orientation on the same strand (H or L strand). Any other kind of supplementary alignment was excluded. Similarly, to avoid the same issue with reads spanning the BamHI cut site in the *ND6* gene (base 14 258–14 259 of the mtDNA reference sequence), we created an alternative GRCh38 reference sequence with a

modified mitochondrial reference starting at base 14 259 instead of base 1. All of the experiments where the samples were fragmented using BamHI were aligned to this alternative sequence (gene annotations were adapted accordingly). Quality control plots and sequencing statistics were automatically generated using NanoPlot v1.13.0 (23).

ROC curve generation

We calculated a ROC curve to assess the accuracy of our methylation calling, using a procedure previously adopted in Simpson et al (25). Briefly, we randomly chose 50,000 mtDNA CpG sites from positive and negative controls and classified each CpG call as true positive (TP) or false positive (FP), depending on which of the two controls each site came from and on whether methylation fell above or below a log-likelihood methylation threshold. We repeated the TP and FP calculation by varying log-likelihood threshold values within a range of -20 to 20 (to build the ROC curve) and 0 to 10 (to calculate accuracy, intended as the proportion of true calls, either TP or true negatives (TN)), with a step of 0.25 , as explained in Simpson *et al.* (25).

Dataset simulation and background noise modelling

To elucidate the relationship between the methylation levels and the read depth in ONS data, we generated *in silico* multiple datasets of simulated sequencing experiments, subsampling the negative control BAM file. We used `samtools -s (read fraction) -b BAM > simulated.sam`. We selected 30 different read fractions matching the read depths achieved with both the fragmentation and BamHI-based sequencing experiments on native DNA. Once the simulated SAM files were generated, we proceeded with the methylation calling using `Nanopolish`, following the same workflow used for cell lines, primary fibroblasts and tissues. Methylation levels calculated on the simulated data were therefore considered background noise introduced by either the ONS technique or the methylation calling procedure. We chose a function describing an exponential decay (1) to model the background noise, given the inverse relationship we observed in simulated data (high methylation levels corresponding to low read depth and vice versa).

$$Y = m * e^{(-t*x)} + b \quad (1)$$

The goodness of fit test showed that the exponential function in (1) well explained the variation of the simulated data ($R^2 = 0.94$), therefore we set out to use the estimated parameters (m , t and b) and the equation in (1) to calculate the background noise present in all downstream ONS sequencing experiments. The background noise model fitting was performed using the `optimize.curve_fit` function of the Scipy Python module. All analyses have been performed in Python 3.0 and code is available at https://github.com/ib361/scripts_paper.

Mitochondrial variant calling of ONS samples

Because Nanopore technology allows a simultaneous read of epigenetic modifications while sequencing the target

DNA, we performed a mitochondrial variant calling on the fastq files filtered with NanoFilt v2.2.0 (23). For this we used a modified version of the MToolBox pipeline (26), adapted to long-reads sequencing analysis (https://github.com/mitoNGS/MToolBox/tree/MToolBox_Nanopore).

Briefly, the main changes integrated into the MToolBox workflow are (i) the use of the Minimap2 aligner software (24) for long-reads mapping and (ii) additional parsing of SAM files to include reads uniquely mapped on the mtDNA reference and reads with supplementary alignments but only showing mtDNA mapping locations. These reads can be the results of the process of linearization of the circular molecule of mtDNA due to random fragmentation or to BamHI enzymatic cut. Reads with secondary or supplementary alignments on the nuclear genome were excluded and classified as possible NuMTs. For read mapping we used the GRCh38 human genome assembly (which includes rCRS as mitochondrial reference sequence). For variant calling, we set the quality score (QS) threshold to retain variants to 10 (changing the `-q` option of the `assemblyMTgenome.py` script). Variants with a read depth per position ≥ 30 and variant allele fraction $\geq 10\%$ were retained. Only single nucleotide variants (SNVs) were considered for comparison with Illumina Miseq sequencing. Haplogroup predictions were performed using both MToolBox and Haplogrep 2 v.2.1.1 (27). Haplogrep2 predictions were based on homoplasmic variants only (with variant allele fraction ≥ 0.9).

Mitochondrial variant calling of Illumina Miseq samples

Fastq files generated with Illumina Miseq were checked for quality using FastQC v0.11.5 (20). Illumina adapters and read ends showing poor *per-base* quality were trimmed using TrimGalore! v0.4.5 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), setting a minimum *per-base* QS = 20 and minimum read length after trimming = 35 bp. Mitochondrial variant calling was then performed with the standard MToolBox pipeline (26), which mapped reads to the human reference genome (GRCh38) with the two-mapping step protocol, to exclude possible NUMT. Single nucleotide variants with ≥ 5 reads of support (and at least 1 read of support on each strand) and minimum QS per base ≥ 25 were retained. Haplogroup predictions were performed using both MToolBox and Haplogrep 2 v.2.1.1 (27). Haplogrep2 predictions were based on homoplasmic variants only (with variant allele fraction ≥ 0.9).

CpG methylation detection in ONS samples

Detection of methylation in CpG context was carried out using Nanopolish v0.11.0 `call-methylation` package (25). Nanopolish utilises a trained Hidden Markov Model to detect modified bases by comparing raw electric signals of modified/unmodified cytosines with expected signal from a reference sequence. The methylation calling output is a log-likelihood ratio where a positive value indicates evidence supporting methylation. Nanopolish utilises `fast5` files containing raw electric signal information, basecalled fastq files and BAM alignment files to generate an index file used by the algorithm to deter-

mine methylation Log-likelihood ratios. Minimap2 alignments to reference sequences were performed with the same parameters described in the Nanopore Data Analysis section. Log-likelihood ratios were then converted to a binary methylated/unmethylated call for each read, then percentage of methylation was obtained by calculating the fraction of methylated reads, using the `calculate_methylation_frequency.py` script available in the package. The default calling threshold of ≥ 2.5 LLR was modified to a more stringent ≥ 5 LLR to increase the accuracy of the call. Since Nanopolish groups neighbouring CpG sites and calls them jointly, CpG sites in the same group were separated and assigned the same methylation frequency using the `-s` option.

CpG methylation analysis in ONS samples

We applied a series of stringent quality filters to remove possible artefacts of the CpG methylation calling and errors introduced by the Nanopolish algorithm. Then, we removed CpGs calls with a methylation frequency greater than two standard deviations from the mean in negative controls (false positives (Supplementary Table S3) and calls neighbouring any heteroplasmic nucleotide variant (heteroplasmy < 0.9) in a ± 5 nucleotides window. This last approach was deemed necessary after noticing that Nanopolish introduced a false methylation call every time a homoplasmic haplogroup-defining variant position fell within ± 5 nucleotides from a CpG. As 11 nucleotides is the kmer size that Nanopolish considers to calculate CpG LLR, we hypothesized that the introduction of a nucleotide variant within ± 5 nucleotides from the CpG altered the Nanopolish methylation determination, leading to an incorrect methylation call. To demonstrate this, we used MToolBox (26) to generate a consensus sequence from the Illumina data, carrying the major alleles at each position, and used this new sequence to perform the methylation calling again on ONS samples. As expected, this time no methylation was identified in the CpGs close to the haplogroup-defining variants. Differential methylation analysis was performed using the R package DSS (28), following the protocol detailed in Gigante et al (18). Differentially methylated mtDNA positions and regions (defined by overlapping tiles of 50nt) were deemed significant if False Discovery Rate was below 1%. For the comparison of the human cell lines and the primary fibroblasts we used as baseline the 613H cell line and the control fibroblasts, respectively.

Statistical tests

Each data distribution was checked for normality by using the Shapiro-Wilk test. For pairwise comparisons, we chose to use the parametric Student's *t*-test or Anova one-way test when values were normally distributed. When not stated, distributions were non-normal and a Wilcoxon two-tailed test was used instead. Spearman's rank test has been used to calculate correlation between variables.

RESULTS

CpG methylation analysis of mtDNA with WGBS

We sought independent evidence that WGBS has limitations for mtDNA by analysing data from 67 human

cell lines and tissues from the NIH Human Epigenome Roadmap Project (19). Fifty-five passed quality control (Materials and Methods) and were aligned to the human genome build GRCh38 (Supplementary Table S4). Analysis of the mtDNA-aligned reads revealed a pronounced *per*-strand mapping and coverage bias. By looking at the percentage of reads mapped *per*-mtDNA-strand, we arbitrarily divided the samples in two groups, depending on the proportion of reads mapping on the H strand ('Biased', where $\geq 55\%$ of the reads aligned to one strand, and 'Low Bias' where 50–55% of the reads aligned to one strand). The Biased (BG) group included 58.2% ($N = 32/55$) samples with the majority of reads mapped to the mitochondrial H-strand ($P \leq 0.0001$, Figure 1A, Supplementary Figure S2A) and a more pronounced *per*-strand coverage bias on the L-strand (L-strand coverage_{BG} = 6.2–88.3%; H-strand coverage_{BG} = 83.5–91.7%, Figure 1B top panel). The remaining data ($N = 23/55$, 'Low Bias' group, LBG), showed less mapping bias on the H-strand (between 50–55% reads; $P \leq 0.0001$, Figure 1A, Supplementary Figure S2A) but no coverage bias (Figure 1B, bottom panel). We observed differences ($P \leq 0.0001$) in the average read depth per position calculated in the two groups: $66.32 \pm 28.84 \times$ BG versus $148.77 \pm 55.45 \times$ LBG (group mean \pm sd; Figure 1C). We found higher apparent methylation levels in the L-strand compared to the H in all samples analysed (L-strand_{BG} = $4.97\% \pm 8.79$ versus H-strand_{BG} = $2.01\% \pm 1.92$ mean methylation \pm sd; L-strand_{LBG} = $1.43\% \pm 0.77$ versus H-strand_{LBG} = $1.39\% \pm 0.7$ mean methylation \pm sd; $P \leq 0.001$; Figure 1D, Supplementary Figure S2B). This is explained by a significant inverse correlation with the read depth per position, leading to the appearance of higher methylation levels where the read depth is low (Spearman's rank test $P < 2.2e-16$; average rho coefficient = -0.78 , Figure 1E). This holds true also for the Low Bias group (with less alignment bias), where local fluctuations in the read depth alter CpG methylation levels (Supplementary Figure S2C, D). This is consistent with previous observations indicating a bisulfite-related selective loss of the cytosine-rich L-strand (13).

Design and assessment of an ONS-based protocol for mtDNA enrichment and analysis

To overcome the problems intrinsic to the WGBS methylation determination, we set out to quantify mtDNA CpG methylation using ONS on native human gDNA. First, we developed a custom-made library preparation protocol (Supplementary Figure S3A) based on the simultaneous linearisation and enrichment of the native full-length mtDNA molecule (Supplementary Figure S3B) through BamHI restriction enzyme digestion (which usually cuts the mtDNA once). We tested the efficiency of our modified protocol over the standard ONS library preparation based on random fragmentation, by performing ONS on gDNA from 3 trans-mitochondrial osteosarcoma cybrid cell lines with known mtDNAs belonging to different mtDNA human haplogroups (29) and with an identical nuclear background (30) ($N = 5$ biological replicates of 3 independent cell lines with the mitochondrial haplogroup H1, J1c and J2, respectively; '613H', '128J', '135J2'; Table 1, Supplementary Table S5). Each gDNA was processed in parallel with both protocols. We further performed strict filtering on

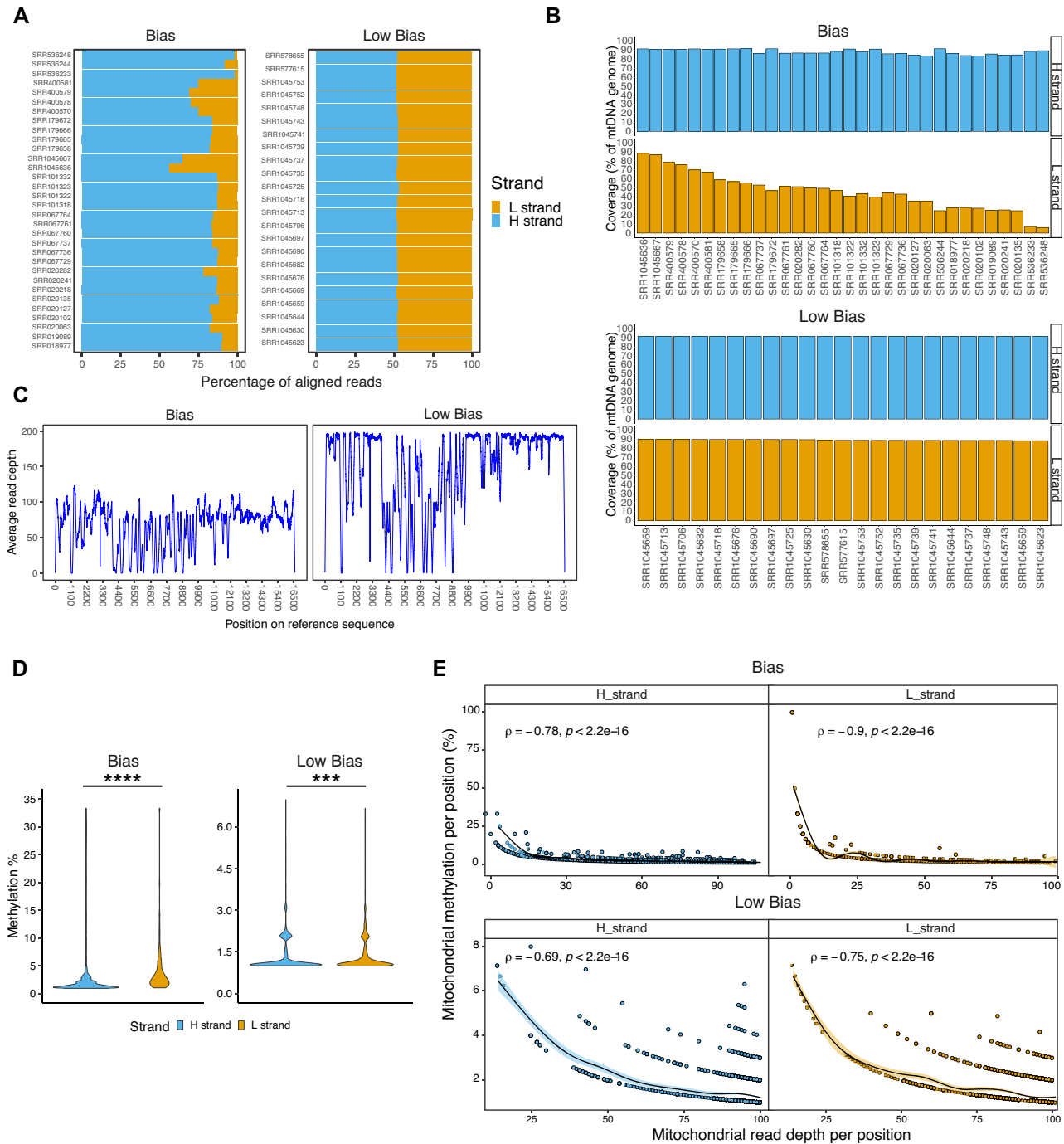


Figure 1. Quantification of WGBS alignment and coverage bias. (A) Percentage of reads aligned to the mtDNA reference per sample, identifying samples with a marked (bias, $N = 32/55$) or low (low bias, $N = 23/55$) per-strand-bias. (B) Percentage of mtDNA covered by at least 5 reads on the two mtDNA strands (H and L) in (top) bias and (bottom) low bias sample groups. (C) Distribution of the average read depth per mtDNA position in the two per-strand-bias groups. (D) Quantification of the average CpG methylation per strand (H and L), divided by per-strand-bias group. The lower and upper hinges of the violinplot correspond to the first and third quartile of the distribution, with median in the centre. Stars indicate significance (***: two-sided $P \leq 0.001$; two-sided **** $P \leq 0.0001$, Wilcoxon test). (E) Correlations between average read depth and average methylation percentage for every cytosine in CpG context, in the two per-strand-bias groups and mtDNA strands (H and L). Spearman's rank test correlation coefficient and two-sided P-values are shown. For all the plots in (D, E), average methylation is intended as the mean methylation value across all the WGBS samples analysed ($N = 55$).

read lengths (selecting between 4000 and 17 000 bp) and per read quality (Phred ≥ 9) before the alignment (Supplementary Figure S3C-D and S4A-B), followed by supplementary alignment removal. This filtering excludes the possibility of studying mitochondrial 7s DNA (31), as its average length (~650 bp) falls below our minimum read length threshold. While not altering quality parameters (percentage of identity and average base quality *per-read*, Supplementary Figure S4C, D), our filtering enriched for full length mtDNA sequences in all BamHI-treated samples. A higher percentage of reads aligned on mtDNA in the BamHI protocols compared to the fragmentation protocol, further confirming that the BamHI treatment enriched for mtDNA (Student's *t*-test $P \leq 0.05$, Supplementary Figure S5).

Under the conditions outlined above, the fragmentation-based method showed a mapping bias on the L-strand (L-strand_{FRAG} = 46.12% \pm 5.13, H-strand_{FRAG} = 53.87% \pm 5.13, percentage of aligned reads mean \pm sd; Anova one-way test $P \leq 0.001$, Figure 2A, Supplementary Figure S6A) with six samples having <100% coverage (Figure 2B). On the contrary, the BamHI-based protocol did not show significant mapping or coverage bias (L-strand_{BAMHI} = 50.67% \pm 4.07, H-strand_{BAMHI} = 49.32% \pm 4.07, percentage of aligned reads mean \pm sd; $P = 0.36$, Figure 2A, B, Supplementary Figure S6A). The average mtDNA read depth was higher in the samples processed with the BamHI-based protocol (Frag. = 23.83x \pm 4.33, BamHI = 131.73x \pm 8.15, mean \pm sd; $P \leq 0.0001$, Figure 2C, Supplementary Figure S6B), with almost half of the mitochondrial reads mapped as full-length molecules (≥ 15 000 bp; 42% \pm 12 of BamHI reads versus 2% \pm 2 of Frag. reads, Supplementary Figure S3D). Overall, these results suggest that our custom-made BamHI ONS protocol is more efficient in achieving full-length mtDNA enrichment and higher mtDNA read depths than the standard Nanopore library preparation.

MtDNA sequencing and the detection of heteroplasmic variants with ONS

Next, we set out to validate the BamHI-ONS protocol for mtDNA sequencing including heteroplasmy detection. We used high-depth Illumina MiSeq sequencing of mtDNA to determine the major alleles and accurately measure heteroplasmy levels (mean read depth = 2769x, min = 318x, max = 5559x; Supplementary Table S6) in the primary and cancer cell lines (Table 1, Supplementary Results). Variant calling with ONS detected 99.5% ($N = 739/743$) of the homoplasmic variants (het. $\geq 95\%$) also detected by the Illumina sequencing, enabling reliable haplogroup predictions (Supplementary Results, Table S6 and Figure S7). As reported previously for exome sequencing (32), higher ONS read-depths were required to reliably measure heteroplasmy levels detected by high-depth Illumina sequencing (Supplementary Figure S8).

Establishing the methylation detection strategy with ONS

We first assessed the accuracy of the methylation calling on mtDNA by sequencing a near complete PCR amplicon of human mtDNA (negative control, NC, 0% methylated) and a corresponding positive control generated *in vitro* with a recombinant CpG methyltransferase (PC, 100% methylated;

Supplementary Figure S1). We used Nanopolish software (25) to call methylation on PC and NC, which generated log-likelihood ratio (LLR) values of CpG methylation (Supplementary Figure S9A). A site is considered methylated when its LLR is above a certain threshold. To choose the most accurate methylation calling cut-off for mtDNA, we: (i) determined the ratios of true and false positives by varying LLR thresholds values (following previous procedures (25), Materials and Methods) and calculating a receiving operating characteristic (ROC) and (ii) methylation calling accuracy (intended as proportion of true calls; Supplementary Figure S9B-C). The ability to distinguish between mtDNA unmethylated and methylated sites was measured by the area under the ROC curve (AUC), which was equal to 0.97 (Supplementary Figure S9B). With the default Nanopolish LLR threshold (≥ 2.5), an accuracy of 97.7% could be achieved (Supplementary Figure S9C). Hence, we chose a more stringent methylation calling threshold (LLR ≥ 5) yielding an accuracy of 99% (Supplementary Figure S9C). Also, by looking at methylation profiles of NC, we identified 13 CpG positions that showed methylation levels above 2 standard deviations from the NC mean (likely false positives, Supplementary Table S3 and Figure S9D), which were investigated further.

We then performed CpG methylation calling followed by differential methylation (DM) analysis (18), in all the cell lines and primary fibroblasts (Table 1, Materials and Methods). First, we checked the methylation levels of the 13 likely false positives we had identified in the NCs, and we found them to be methylated consistently in all our samples analysed (Supplementary Table S3). Hence, we removed the 13 positions from all subsequent ONS sequencing experiments results.

In the cell lines, the analysis revealed several other possible differentially methylated CpGs (DM-CpGs; Supplementary Table S7). However, close scrutiny revealed that an haplogroup-defining mtDNA variant always fell within a ± 5 bp window from a possible DM-CpG, prompting us to hypothesize that the haplogroup variants influenced Nanopolish methylation calling (Materials and Methods). To test this, we generated a new reference for methylation calling based on a mtDNA consensus sequence of the major mtDNA alleles identified with Illumina MiSeq sequencing. DM analysis using sample-specific consensus sequences entirely removed the apparent methylation signal detected earlier (Figure 3A, Supplementary Table S7). Therefore, we set out to perform methylation calling always using a sample-specific consensus sequence. We also compared the level of DM on mtDNA molecules carrying the m.3243A>G mutation to wild-type molecules in the same two primary fibroblasts lines, but saw no difference in CpGs methylation between the two molecular species (Supplementary Table S7).

We observed a negative correlation between read depth and methylation signal in the samples used to test the enrichment protocol (Table 1), which suggested that the observed methylation was background noise intrinsic to the technology (Figure 3B, Supplementary Results). To test this hypothesis, we performed subsampling from the NC and generated simulated unmethylated ONS datasets ('Background'). These matched the read depths obtained with

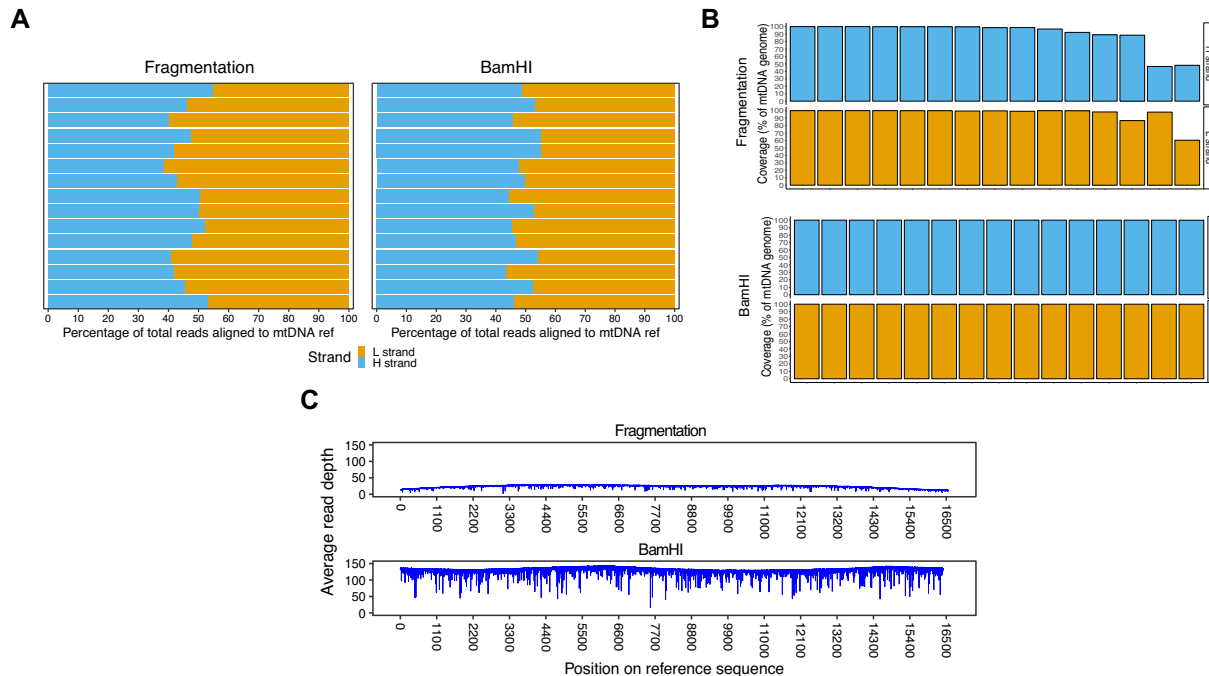


Figure 2. BamHI-based protocol improves mtDNA reads alignment over the standard ONS library preparation. **(A)** Percentage of reads aligned to the mtDNA reference per strand per biological replicate ($N = 15$ samples per protocol), in samples processed with (left) fragmentation protocol and (right) BamHI-based protocol. **(B)** Percentage of mtDNA covered by at least 5 reads on the two mtDNA strands (H and L) per biological replicate ($N = 15$ samples per protocol), in samples processed with (top) fragmentation protocol and (bottom) BamHI-based protocol. **(C)** Distribution of the average read depth per mtDNA position in samples processed with (top) fragmentation protocol and (bottom) BamHI-based protocol ($N = 15$ samples per protocol). Samples in (A–C) are the same processed in parallel with either protocol

both fragmentation and BamHI-based protocol-derived experiments (‘Observed’) (Figure 3C). We then inferred a model that best fitted the simulated data ($R^2 = 0.94$, Figure 3B, Materials and Methods) which we used to estimate the background noise in methylation calling of all the ONS experiments we have performed in this study.

ONS-based CpG methylation analysis of mtDNA in human cell lines and tissues

Using a sample-specific mtDNA reference sequence for methylation calling we looked for evidence of mtDNA methylation in all three cancer cell lines ($N = 5$ biological replicates) and three primary fibroblast lines ($N = 3$ biological replicates; Figure 4A, Supplementary Table S7; Methylation_{C_LINES/FIB} = 1.3–2%; min–max). In each case, the apparent methylation values were below the estimated background noise level.

To extend these findings to *in vivo* samples, we sequenced seven fresh human tissues of different healthy individuals (Table 1, Supplementary Table S5). Again, we observed that the apparent methylation levels were below the estimated background noise, even at higher read depths compared to the cell lines and primary fibroblasts (Figure 4B, Supplementary Table S7; Methylation_{TISSUES} = 0.6–0.8%; min–max).

Finally, we sought for conclusive evidence that ONS is capable of identifying methylation above the background level. To do that, we generated and sequenced with ONS 4 additional control samples with expected methylation levels

of 5%, 25%, 50% and 75%, by mixing the PC and NC (Materials and Methods). Results of this analysis revealed that the expected methylation levels were correctly measured with ONS ($\text{Rho} = 1$, $P = 0.003$, Spearman’s rank test, Figure 4C).

DISCUSSION

The discovery of mitochondrially-targeted methyltransferases (33–35) implied that mtDNA could be methylated. In addition to a role in modulating mitochondrial gene expression, multiple studies report mtDNA CpG methylation as a biomarker of ageing (36), environmental exposure to tobacco smoke (37), cancer (38) and neurological diseases (39,40). Currently, quantitative analysis of CpGs is mostly based either on mass spectrometry or on the bisulfite treatment of gDNA (bisulfite pyrosequencing and WGBS). While the first method is the most sensitive in determining the general CpG methylation level of a given sample, it lacks information about the position of individual methylated residues (41). On the other hand, while bisulfite-based technologies resolve the CpG methylation at a single-base level, they are susceptible to the introduction of biases due to the selective degradation of cytosine-rich sequences (both nuclear and on the L-strand of mtDNA) (13,42). Despite this, multiple groups have continued to investigate mtDNA methylation using bisulfite-based technologies, without accounting for presence of alignment biases (35,43).

In an attempt to resolve this controversy, we studied 55 publicly available WGBS datasets part of the Roadmap

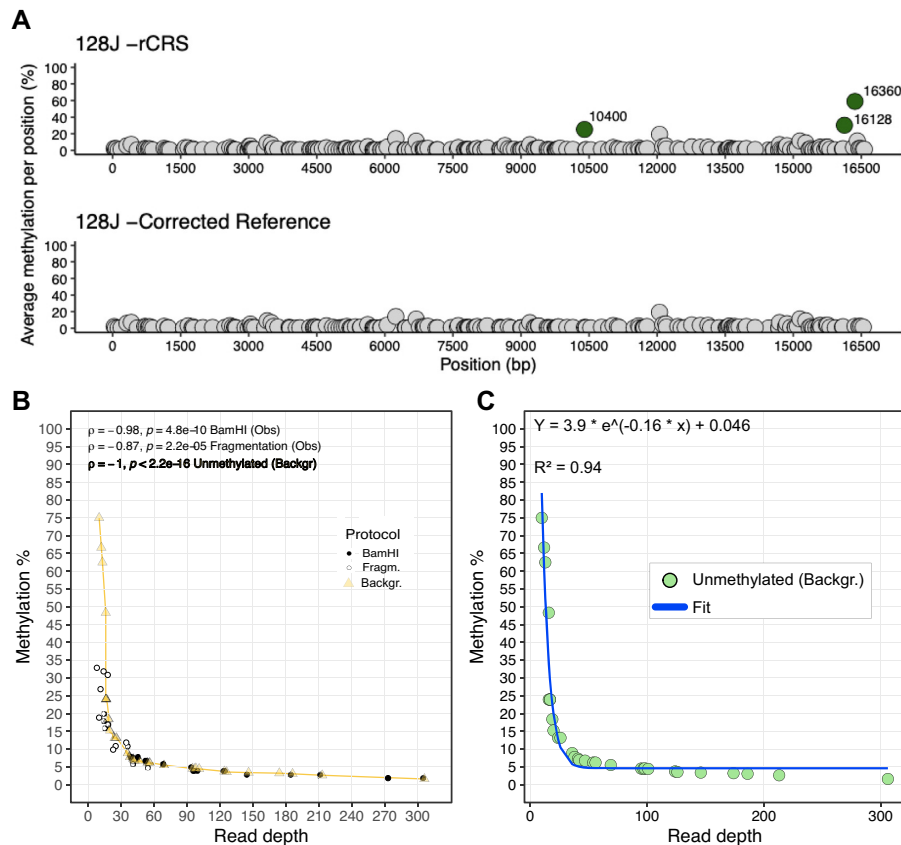


Figure 3. Methylation detection strategy with ONS. (A) Example of methylation calling artefacts introduced when using hg38 as reference (which includes the mitochondrial reference sequence rCRS) (top) instead of a sample-specific consensus sequence (bottom). In green are highlighted the sample-specific differentially methylated positions which disappear upon reference correction. (B) Correlation between average read depth and average methylation percentage in samples processed with fragmentation- and BamHI-based protocols (Obs) and in unmethylated datasets simulated from the negative control (Backgr). Circles represent a sample sequenced with either sequencing protocol ($N = 15$ per protocol). Triangles represent an unmethylated dataset simulated from the negative control. Spearman's rank correlation coefficient and two-sided P -values are shown. (C) Green circles represent correlation between average read depth and average methylation percentage in 30 unmethylated datasets simulated from negative controls. The blue solid line represents the fitted line (exponential decay function) that describes such distribution, corresponding to the methylation background noise (Backgr). The formula describing the fit and R^2 correlation calculated with the goodness of fit test are shown.

Epigenome Project (19), focussing on describing per-strand sequencing metrics and how these affect the methylation profile of mtDNA. Our analysis confirmed a marked per-strand bias with an impact on global mtDNA CpG methylation levels quantification.

To overcome WGBS limitations, we developed an accurate and reproducible protocol to investigate mtDNA methylation using ONS, and tested our method against the standard ONS library preparation protocol based on random fragmentation. Our protocol is based on selective restriction digestion by BamHI followed by selection of longer sequences, which results in an enrichment for native full-length mtDNA thereby minimising the potential for NuMTs contamination. Comparing our results with Illumina sequencing, we found that our protocol allows the correct calling of the vast majority of homoplasmic and high-heteroplasmy (>95%) mtDNA alleles, including pathogenic mutations.

Our analysis also revealed that the methylation calling with Nanopolish is influenced by the presence of mtDNA variants surrounding the CpG residue. In light of this, we

recommend a careful review of previously identified methylated positions and of differential methylation results (44) in the context of the nucleotide reference sequence. This is also likely to be an issue for nuclear DNA, although we have not formally studied this here. It is not clear why the 13 positions listed in Supplementary Table S3 appeared to be methylated in every one of the samples we analysed. However, local sequence context is known to influence methylation calling, as we have shown for mtDNA Figure 3, and has been attributed to systematic errors introduced by the Nanopolish software (45).

Finally, our study indicates that, after removing the technical biases and regressing out the background noise, no residual CpG mtDNA methylation could be identified with ONS across multiple human tissues and cell lines. These findings add to emerging evidence (3–5,11) that CpG methylation is not occurring on human mtDNA. Although our findings do not exclude the possibility of very low levels of mtDNA methylation, these are unlikely to be biologically relevant. The exact nature of this background noise is an interesting matter of debate. The residual signal could come

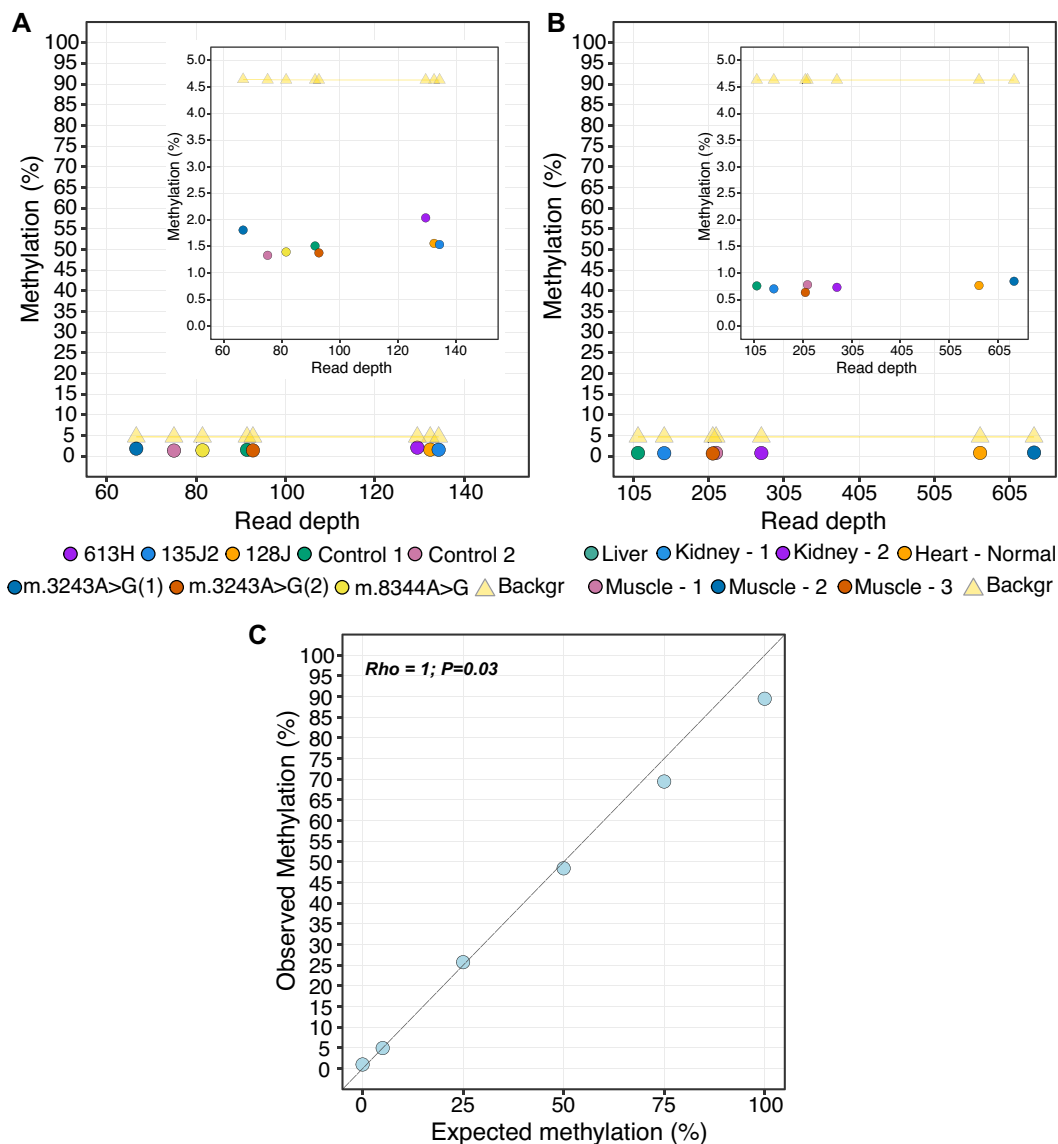


Figure 4. ONS analysis on human cell lines and tissues reveals absence of relevant CpG methylation levels on mtDNA. (A, B) Scatterplots showing the relationship between average read depth and average methylation percentage in samples processed with BamHI protocol. Circles in (A) represent an average of all mtDNA positions in either five (cell lines) or three (primary fibroblasts) biological replicates. Circles in (B) represent an average of all the mtDNA positions in human tissues from seven different individuals. Yellow triangles in (A) and (B) represent the background noise. Inset plots show magnification of the data shown. (C) Correlation between the expected and observed methylation levels calculated on methylated controls generated by mixing PC and NC. Spearman's test, P -value and rho are shown.

from methylated NuMTs sequences which aligned to the mtDNA reference despite our filtering applied before and after alignment. Against that, we saw the noise despite minimising the potential for NuMTs contamination by developing a protocol focussed on full-length mtDNA sequencing. It is also unlikely that spontaneous background methylation *in vivo* explains the residual signal, because the concentrations of SAM necessary to introduce mutations are ~16-fold greater than the levels measured in mitochondria (46,47). It therefore seems more likely that the noise arises from random variations in the ONS electric signal at the moment it is generated and registered. Such small variations would possibly be interpreted by Nanopolish as background methylation.

Importantly, the protocol we have developed has potential applications beyond CpG methylation, including mtDNA variant calling and measuring other types of epigenetic modifications (e.g. m6A methylation) at the single nucleotide level (48,49).

DATA AVAILABILITY

The raw DNA sequencing data in this manuscript has been deposited on the SRA archive and is accessible using the BioProject accession number PRJNA763486. The MToolBox pipeline used for mitochondrial variant calling and code for filtering VCF files are available as GitHub branch of the MToolBox repository: <https://github.com/mitoNGS/>

MToolBox/tree/MToolBox_Nanopore. Code used for plotting and data analysis is available at https://github.com/ib361/scripts_paper.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

I.B. is funded by the Cambridge Trust, the MRC-DTP PhD Program and Magdalene College, Cambridge; A.G.D. is an Atracción de Talento M1 Fellow from Comunidad de Madrid (Spain) [2019-T1BMD-14236] who receives support from the Spanish Ministry of Science [PID2020-114709RA-I00]; P.F.C. is a Wellcome Trust Principal Research Fellow [212219/Z/18/Z]; UK NIHR Senior Investigator, who receives support from the Medical Research Council Mitochondrial Biology Unit [MC.UU.00015/9]; Medical Research Council (MRC) International Centre for Genomic Medicine in Neuromuscular Disease [MR/S005021/1]; Leverhulme Trust [RPG-2018-408]; MRC research grant [MR/S035699/1]; Alzheimer's Society Project Grant [AS-PG-18b-022]; NIHR Cambridge Biomedical Research Centre [BRC-1215-20014]. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. Human tissues were sourced from the Human Research Tissue Bank, which is supported by the NIHR Cambridge Biomedical Research Centre. Funding for open access charge: Wellcome Trust.

Conflict of interest statement. None declared.

REFERENCES

- Smith, Z.D. and Meissner, A. (2013) DNA methylation: roles in mammalian development. *Nat. Rev. Genet.*, **14**, 204–220.
- Siegfried, Z. and Simon, I. (2010) DNA methylation and gene expression. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, **2**, 362–371.
- Matsuda, S., Yasukawa, T., Sakaguchi, Y., Ichiyonagi, K., Unoki, M., Gotoh, K., Fukuda, K., Sasaki, H., Suzuki, T. and Kang, D. (2018) Accurate estimation of 5-methylcytosine in mammalian mitochondrial DNA. *Scientific Rep.*, **8**, 5801–5801.
- Mechta, M., Ingerslev, L.R., Fabre, O., Picard, M. and Barrès, R. (2017) Evidence suggesting absence of mitochondrial DNA methylation. *Front. Genet.*, **8**, 166.
- Hong, E.E., Okitsu, C.Y., Smith, A.D. and Hsieh, C.-L. (2013) Regionally specific and genome-wide analyses conclusively demonstrate the absence of CpG methylation in human mitochondrial DNA. *Mol. Cell Biol.*, **33**, 2683–2690.
- Suomalainen, A. and Battersby, B.J. (2018) Mitochondrial diseases: the contribution of organelle stress responses to pathology. *Nat. Rev. Mol. Cell Biol.*, **19**, 77–92.
- Sirard, M.-A. (2019) Distribution and dynamics of mitochondrial DNA methylation in oocytes, embryos and granulosa cells. *Sci. Rep.*, **9**, 11937.
- Krueger, F., Kreck, B., Franke, A. and Andrews, S.R. (2012) DNA methylome analysis using short bisulfite sequencing data. *Nat. Methods*, **9**, 145–151.
- Wolters, J.E.J., van Breda, S.G.J., Caiment, F., Claessen, S.M., de Kok, T.M.C.M. and Kleinjans, J.C.S. (2017) Nuclear and mitochondrial DNA methylation patterns induced by valproic acid in human hepatocytes. *Chem. Res. Toxicol.*, **30**, 1847–1854.
- Devall, M., Smith, R.G., Jeffries, A., Hannon, E., Davies, M.N., Schalkwyk, L., Mill, J., Weedon, M. and Lunnon, K. (2017) Regional differences in mitochondrial DNA methylation in human post-mortem brain tissue. *Clin Epigenet.*, **9**, 47.
- Liu, B., Du, Q., Chen, L., Fu, G., Li, S., Fu, L., Zhang, X., Ma, C. and Bin, C. (2016) CpG methylation patterns of human mitochondrial DNA. *Sci. Rep.*, **6**, 23421.
- Anderson, S., Bankier, A.T., Barrell, B.G., Bruijn, M.H.L., de Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F. et al. (1981) Sequence and organization of the human mitochondrial genome. *Nature*, **290**, 457–465.
- Olova, N., Krueger, F., Andrews, S., Oxley, D., Berrens, R.V., Branco, M.R. and Reik, W. (2018) Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biol.*, **19**, 33.
- Stewart, J.B. and Chinnery, P.F. (2015) The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat. Rev. Genet.*, **16**, 530–542.
- Hazkani-Covo, E., Zeller, R.M. and Martin, W. (2010) Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet.*, **6**, e1000834.
- Dayama, G., Emery, S.B., Kidd, J.M. and Mills, R.E. (2014) The genomic landscape of polymorphic human nuclear mitochondrial insertions. *Nucleic Acids Res.*, **42**, 12640–12649.
- Jain, M., Olsen, H.E., Paten, B. and Akeson, M. (2016) The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.*, **17**, 239.
- Gigante, S., Gouil, Q., Lucattini, A., Keniry, A., Beck, T., Tinning, M., Gordon, L., Woodruff, C., Speed, T.P., Blewitt, M.E. et al. (2019) Using long-read sequencing to detect imprinted DNA methylation. *Nucleic Acids Res.*, **47**, e46.
- Consortium, Roadmap Epigenomics, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J. et al. (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317–330.
- Andrews, S. (2015) FASTQC a quality control tool for high throughput sequencing data. <https://www.bioinformatics.org/cwl/view/git/e0a30aa1ad516dd2ec0e9ce006428964b840daf4/workflows/fastqc.cwl>.
- Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
- Krueger, F. and Andrews, S.R. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, **27**, 1571–1572.
- De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M. and Van Broeckhoven, C. (2018) NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*, **34**, 2666–2669.
- Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094–3100.
- Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J. and Timp, W. (2017) Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods*, **14**, 407–410.
- Calabrese, C., Simone, D., Diroma, M.A., Santorsola, M., Guttà, C., Gasparre, G., Picardi, E., Pesole, G. and Attimonelli, M. (2014) MToolBox: a highly automated pipeline for heteroplasmy annotation and prioritization analysis of human mitochondrial variants in high-throughput sequencing. *Bioinformatics*, **30**, 3115–3117.
- Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A. and Schönherr, S. (2016) HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.*, **44**, W58–W63.
- Park, Y. and Wu, H. (2016) Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics*, **32**, 1446–1453.
- Gómez-Durán, A., Pacheu-Grau, D., Martínez-Romero, I., López-Gallardo, E., López-Pérez, M.J., Montoya, J. and Ruiz-Pesini, E. (2012) Oxidative phosphorylation differences between mitochondrial DNA haplogroups modify the risk of Leber's hereditary optic neuropathy. *Biochim. Biophys. Acta*, **1822**, 1216–1222.
- Chomyn, A., Lai, S.T., Shakeley, R., Bresolin, N., Scarlato, G. and Attardi, G. (1994) Platelet-mediated transformation of mtDNA-less human cells: analysis of phenotypic variability among clones from normal individuals—and complementation behavior of the tRNALys mutation causing myoclonic epilepsy and ragged red fibers. *Am. J. Hum. Genet.*, **54**, 966–974.
- Nicholls, T.J. and Minczuk, M. (2014) In D-loop: 40 years of mitochondrial 7S DNA. *Exp. Gerontol.*, **56**, 175–81.

32. Griffin,H.R., Pyle,A., Blakely,E.L., Alston,C.L., Duff,J., Hudson,G., Horvath,R., Wilson,I.J., Santibanez-Koref,M., Taylor,R.W. *et al.* (2014) Accurate mitochondrial DNA sequencing using off-target reads provides a single test to identify pathogenic point mutations. *Genet. Med.*, **16**, 962–971.
33. Shock,L.S., Thakkar,P.V., Peterson,E.J., Moran,R.G. and Taylor,S.M. (2011) DNA methyltransferase 1, cytosine methylation, and cytosine hydroxymethylation in mammalian mitochondria. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 3630–3635.
34. Wong,M., Gertz,B., Chestnut,B.A. and Martin,L.J. (2013) Mitochondrial DNMT3A and DNA methylation in skeletal muscle and CNS of transgenic mouse models of ALS. *Front. Cell. Neurosci.*, **7**, 279.
35. Patil,V., Cuenin,C., Chung,F., Aguilera,J.R.R., Fernandez-Jimenez,N., Romero-Garmendia,I., Bilbao,J.R., Cahais,V., Rothwell,J. and Herceg,Z. (2019) Human mitochondrial DNA is extensively methylated in a non-CpG context. *Nucleic Acids Res.*, **47**, 10072–10085.
36. D'Aquila,P., Giordano,M., Montesanto,A., De Rango,F., Passarino,G. and Bellizzi,D. (2015) Age- and gender-related pattern of methylation in the *MT-RNR1* gene. *Epigenomics*, **7**, 707–716.
37. Vos,S., Nawrot,T.S., Martens,D.S., Byun,H.-M. and Janssen,B.G. (2021) Mitochondrial DNA methylation in placental tissue: a proof of concept study by means of prenatal environmental stressors. *Epigenetics*, **16**, 121–131.
38. Dong,Z., Pu,L. and Cui,H. (2020) Mitopigenetics and Its emerging roles in cancer. *Front. Cell Dev. Biol.*, **8**, 4.
39. Blanch,M., Mosquera,J.L., Ansoleaga,B., Ferrer,I. and Barrachina,M. (2016) Altered mitochondrial DNA methylation pattern in Alzheimer disease-related pathology and in Parkinson disease. *Am. J. Pathol.*, **186**, 385–397.
40. Stocco,A., Siciliano,G., Migliore,L. and Coppè,F. (2017) Decreased methylation of the mitochondrial D-loop region in late-onset Alzheimer's disease. *JAD*, **59**, 559–564.
41. Song,L., James,S.R., Kazim,L. and Karpf,A.R. (2005) Specific method for the determination of genomic DNA methylation by liquid chromatography-electrospray ionization tandem mass spectrometry. *Anal. Chem.*, **77**, 504–510.
42. Ji,L., Sasaki,T., Sun,X., Ma,P., Lewis,Z.A. and Schmitz,R.J. (2014) Methylated DNA is over-represented in whole-genome bisulfite sequencing data. *Front. Genet.*, **5**, 341.
43. Dou,X., Boyd-Kirkup,J.D., McDermott,J., Zhang,X., Li,F., Rong,B., Zhang,R., Miao,B., Chen,P., Cheng,H. *et al.* (2019) The strand-biased mitochondrial DNA methylome and its regulation by DNMT3A. *Genome Res.*, **29**, 1622–1634.
44. Goldsmith,C., Rodríguez-Aguilera,J.R., El-Rifai,I., Jarretier-Yuste,A., Hervieu,V., Raineteau,O., Saintigny,P., Chagoya de Sánchez,V., Dante,R., Ichim,G. *et al.* (2021) Low biological fluctuation of mitochondrial CpG and non-CpG methylation at the single-molecule level. *Sci. Rep.*, **11**, 8032.
45. Yuen,Z.W.-S., Srivastava,A., Daniel,R., McNevin,D., Jack,C. and Eyras,E. (2021) Systematic benchmarking of tools for CpG methylation detection from nanopore sequencing. *Nat. Commun.*, **12**, 3438.
46. Farooqui,J.Z., Lee,H.W., Kim,S. and Paik,W.K. (1983) Studies on compartmentation of S-adenosyl-L-methionine in *Saccharomyces cerevisiae* and isolated rat hepatocytes. *Biochim. Biophys. Acta (BBA) - Gen. Subj.*, **757**, 342–351.
47. Rydberg,B. and Lindahl,T. (1982) Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. *EMBO J.*, **1**, 211–216.
48. Hao,Z., Wu,T., Cui,X., Zhu,P., Tan,C., Dou,X., Hsu,K.-W., Lin,Y.-T., Peng,P.-H., Zhang,L.-S. *et al.* (2020) N6-deoxyadenosine methylation in mammalian mitochondrial DNA. *Mol. Cell*, **78**, 382–395.
49. Willson,J. (2020) DNA 6mA in times of mitochondrial stress. *Nat. Rev. Mol. Cell Biol.*, **21**, 252–253.