**Effect of Mitochondrial DNA Variation on the Nuclear Epigenome and Transcriptome**

Tyler Shin Nagano

Western University

Schulich School of Medicine and Dentistry – Pathology and Laboratory Medicine

Castellani Lab

Dr. Christina Castellani

April 20th, 2022

**Abstract:**

**Introduction:** Mitochondrial DNA copy number (mtDNA-CN) is associated with several age-related chronic diseases and is a predictor of all-cause mortality. A previous study by Castellani et al. identified a role for mtDNA-CN variability in the regulation of nuclear gene expression via nuclear DNA (nDNA) methylation. What remains unknown are the underlying biological mechanisms controlling the effect of mtDNA-CN on nDNA gene expression. In this study, we hypothesized that site-specific differential nDNA methylation and differential gene expression resulting from in vitro reduction of mtDNA-CN would uncover shared genes and biological pathways important in the mechanisms mediating the effect of mtDNA-CN on disease.

**Methods:** To test this hypothesis, we generated epigenome and transcriptome profiles for three independent human embryonic kidney (HEK293T) cell lines harbouring a mitochondrial transcription A (TFAM) heterozygous knockout using CRISPR-Cas9, and matched control lines. Methylation data was generated using the Illumina Infinium Methylation EPIC BeadChip, RNA sequencing was performed using the Illumina HiSeq 2500 instrument and the data was analyzed to call differentially methylated sites, differentially methylated regions, and differentially expressed genes. Finally, we integrated the analyses and performed functional enrichment to determine genes and pathways which may facilitate mtDNA-CN effect on nDNA gene expression.

**Results:** Our results identified 4242 differentially methylated sites associated with mtDNA-CN at epigenome-wide significance (p < 1e-7). 228 differentially methylated regions associated with mtDNA-CN (p < 1.17e-5) and 179 differentially expressed genes were identified (p < 3.59e-6). Integration analysis identified 314 Gene-CpG pairs (FDR < 0.01). GO enrichment analyses demonstrated that GABAA receptor genes were overrepresented in all analyses. KEGG enrichment analysis demonstrated that 5 of 6 significant pathways contain GABAA receptor genes. Chromatin states relating to enhancers, polycomb repression and heterochromatin were overrepresented in our data.

**Discussion:**  These findings demonstrate that GABAA receptor genes in significant KEGG pathways may be related to the underlying biological mechanisms which facilitate the effect of mtDNA-CN on nDNA methylation. Other biological mechanisms have been identified from the Gene-CpG pairs which may facilitate mtDNA-CN effect on age-related chronic diseases. Further, these results suggest that mitochondrial DNA variation signals to the nuclear DNA epigenome and transcriptome and may lead to changes relevant to development, aging, and complex disease.

**Keywords:** Mitochondrial DNA Copy Number, Epigenomics, DNA Methylation, TFAM, Transcriptomics

**Overall Objective:**

The main objective of this thesis project is to identify some underlying biological mechanisms of the effect of mitochondrial DNA copy number (mtDNA-CN) variation on nuclear DNA (nDNA) methylation in a mitochondrial transcription factor A (TFAM) heterozygous knockout cell line model. The epigenome and transcriptome data generated from the cell line model will be analyzed using bioinformatics programs on the command line and R to determine sites of differential methylation and differential gene expression to determine relevant biological mechanisms which mediate the effect of mtDNA-CN on disease.

**Background:**

**Mitochondria Function and DNA:**

Mitochondria are cytoplasmic organelles which encode genes that are essential to cellular metabolism. Mitochondria are central to many cellular processes such as oxidative phosphorylation to produce ATP, apoptosis for proper cell death, and cell differentiation via cell signalling [1]. Previously, mitochondrial dysfunction was thought to be important only in rare diseases such as Kearns-Sayre syndrome, MELAS syndrome and Leber’s hereditary optic neuropathy because mitochondrial DNA (mtDNA) is inherited by different genetic rules and uses a modified genetic code compared to nDNA [2]. Recently, mtDNA mutations and polymorphisms have been shown to contribute to many common age-related diseases such as cardiovascular disease [3], obesity [4], and many cancers [5]. Despite the large evidence that mtDNA mutations affects many diseases, the biological mechanisms responsible for mitochondrial dysfunction leading to many age-related diseases are largely unknown.

The mitochondrial genome is circular and 16.6 kilobase pairs long in humans. The genome encodes for 37 genes which include 13 proteins, 22 tRNAs and 2 rRNAs and it is passed down to offspring through maternal inheritance. Unlike nDNA which comes in two copies per cell, mtDNA has between 2 and 10 copies and the number of copies per cell are highly variable depending on cell type [6]. Additionally, the number of copies of mitochondria per cell vary from tens to thousands depending on the energy demands of each cell type [6]. Although variation in mitochondrial DNA copy number (mtDNA-CN) is common between cell types, further evidence shows that mtDNA-CN levels decline with age, generally women have higher levels of mtDNA-CN and there is substantial variation between individuals [7].

Although there are many factors which impact mtDNA-CN variation, reductions in mtDNA-CN have been shown to have reduced respiratory enzyme activity, lower expression of proteins in oxidative phosphorylation and differences in cellular characteristics [8]. mtDNA-CN variation has been shown to be associated with many age-related diseases such as diabetes and cardiovascular disease and reduced mtDNA-CN has been shown to be associated with all-cause mortality [9]. Additionally, mtDNA-CN estimates can serve as a biomarker of mitochondrial function which can be estimated fairly easily using DNA from blood samples.

Chart

Description automatically generated

**Figure 1.** Kaplan-Meier survival curve estimates for all-cause mortality by quintiles of mtDNA-CN adapted from (Ashar et al., 2015) [9]

**Nuclear Epigenome:**

The nuclear epigenome consists of all the chemical modifications to DNA and histones. These modifications include histone methylation, histone acetylation, DNA methylation and many others currently emerging in the field [10]. Epigenetic marks are important for cellular function because the marks allow cells to dynamically regulate gene expression, temporally and spatially, in response to changes in the environment [10]. In addition, epigenetic marks can be preserved and passed down to offspring contributing to long-term modifications to cellular function[10].

The most well understood epigenetic mark is DNA methylation. DNA methylation is typically found on cytosine residues at CG dinucleotides called CpGs. The methylation of CpG sites is maintained during DNA replication by DNA methyltransferase [11]. In most cells, a significant portion of the genome is methylated which include regulatory regions, such as promoters and enhancers, motif sequences and gene-coding regions [11]. Clusters of CpGs are characterized by higher G and C base content and are called CpG islands. Many promoter regions are associated with CpG islands and the methylation status of the CpG islands can affect gene expression [12]. In many cases, when a promoter region is unmethylated transcription of the associated gene can occur, but when the region is methylated, the gene cannot be transcribed [12]. During normal development, methylation of specific promoter regions helps to suppress transcription of genes in processes such as genomic imprinting and embryonic development [12].

**Mitochondria Effect on Nuclear DNA Methylation:**

The bi-directional communication between mtDNA and nDNA is known to be essential to maintaining homeostasis, normal cell functioning and maintaining the structural integrity of the cell [13,14]. It is also well-known that disruptions to interactions between mtDNA and nDNA contribute to many diseases. Although we understand the importance of the bi-directional communication between mtDNA and nDNA the specific mechanisms of the relationship are not well understood.

Previous studies have investigated mtDNA variation effect on nDNA methylation and they have shown that there is an association between mtDNA and nDNA methylation [15]. In one study, variation in methylation of nDNA sites have been shown between normal and mitochondria-depleted cancer cell lines [16]. In a mouse model, it was shown that mice with identical nDNA, but different mtDNA variants demonstrated differences in nDNA methylation in brain tissue with matching differences in gene expression [17]. Other studies have investigated the effect of variation in mtDNA-CN on nDNA methylation which have shown that reductions in mtDNA-CN is associated with hypermethylation of nDNA promoters which further cancer progression. Additionally, in glioblastoma tumours variations in mtDNA-CN were shown to effect nDNA methylation and gene expression [18,19]. More recent work has shown that mtDNA-CN is causal of changes to nDNA methylation which therefore effect nuclear gene expression [20].

**Hypothesis:**

*In vitro* reduction of mtDNA-CN leads to differential nDNA methylation and differential gene expression in shared genes and biological pathways which mediate the effect of mtDNA-CN on age-related diseases. To test this hypothesis, epigenome and transcriptome data will be analyzed to uncover sites of differential methylation and differentially expressed genes. The significant results of the analyses will be integrated to determine relationships between CpGs and genes. In addition, functional enrichment analyses on the significant results will help to discover the shared genes and biological pathways important in facilitating mtDNA-CN effect on age-related diseases.

**Methods:**

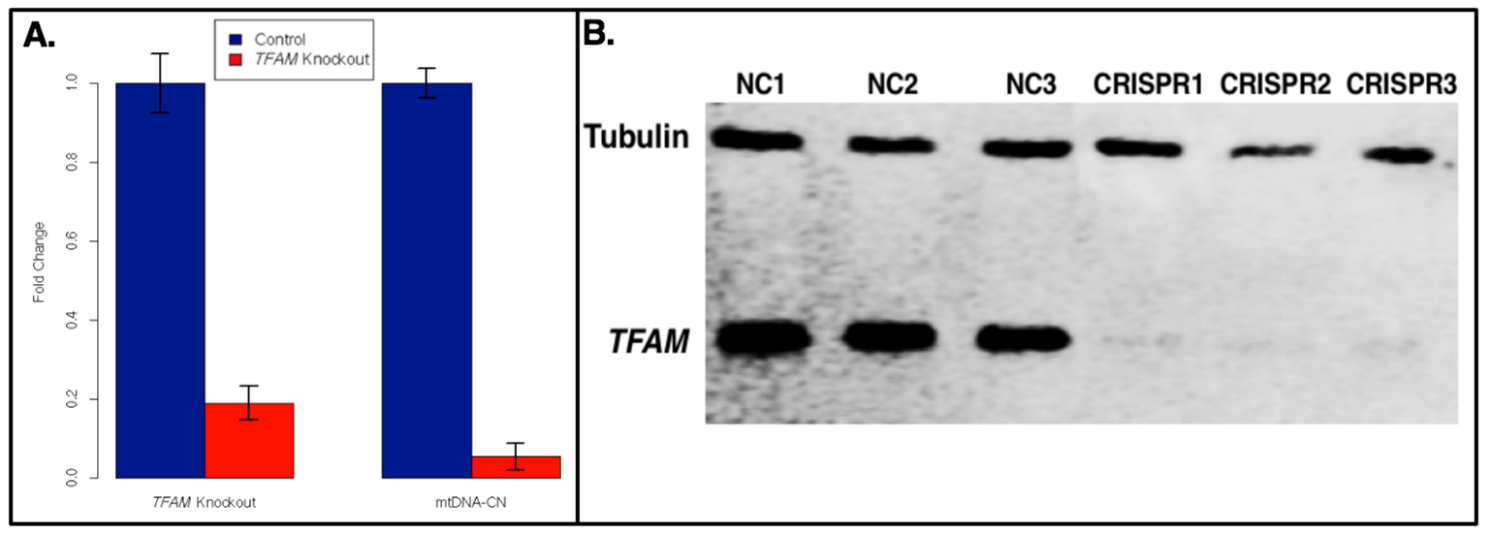
**TFAM Knockout Cell Line Model:**

The experimental model used in this study is a human embryonic kidney (HEK293T) cell line with mitochondrial transcription factor A (TFAM) heterozygous knockout using CRISPR-Cas9. This cell model was used because it has been shown that TFAM is an essential nuclear gene for facilitating human mtDNA transcription and is key regulator of mtDNA-CN in cells [21]. Therefore, this cell line model reduced mtDNA transcription and mtDNA-CN to allow for comparisons between the knockout (KO) cell lines and negative control (NC) cell lines. Additionally, this cell line model is the same model used in the recent study [20] which demonstrated the causal relationship of mtDNA-CN effect on nuclear gene expression through nDNA methylation.

The knockout procedure was performed independently for each of the 3 knockout (KO) cell lines. 3 negative control (NC) cell lines were generated using the same cell line without the TFAM knockout. Each of the cell lines were plated separately to create 6 samples (3 KO and 3 NC).

qPCR was used to measure mtDNA-CN which generated the mtDNA copy number relative to the nuclear DNA copy number. A Western Blot was performed to demonstrate the removal of the TFAM gene in the KO lines with Tubulin as the control. The KO lines demonstrated a 5-fold reduction of RNA expression and an 18-fold reduction in mtDNA-CN compared to the negative control lines

The generation of the TFAM knockout cell model, DNA/RNA/protein isolation, mtDNA-CN estimation, qPCR and Western Blotting were performed previously in this study [20].



**Figure 2.** CRISPR-Cas9-induced heterozygous knockout of TFAM. **A.** 80% reduction of RNA expression of TFAM KO compared to NC (left). 18-fold reduction of mtDNA-CN in KO cell lines compared to NC. **B.** TFAM KO and NC Western Blot demonstrating >81% reduction in TFAM protein. Control is Tubulin. Adapted from [20]

**Data Collection:**

The Illumina Infinium EPIC BeadChip was used to hybridize the TFAM KO and NC cell line following the methods in this study to determine the DNA methylation profile for >850,000 CpGs in the human genome [20]. All samples (3 KO and 3 NC) were run on the EPIC BeadChip at two separate time points which resulted in 12 .idat files (6 KO and 6 NC).

RNA quantification, quality control and library preparation and sequencing were performed using the Qiagen AllPrep DNA/RNA Mini Kit, Qubit 2.0 Fluorometer and Illumina’s TruSeq Stranded Total RNA kit as described in this paper [20] to generate gene expression profiles. Each of the 6 samples (3 KO and 3 NC) were sequenced twice at the same time point which resulted in 12 compressed FASTQ files.

The duplicate runs are technical replicates to improve the power of the study. The different time points of methylation data collection do not have a biological reason.

**Overview of Analyses:**

All analyses were run on Compute Canada using R version 4.1.2. All Bioconductor R packages used in the analysis are from Bioconductor version 3.14.

**Methylation Analysis:**

The minfi package was used for EPIC BeadChip analysis and quality control to remove poor quality probes [22]. The data was normalized using functional normalization [23]. Poor quality probes were removed if the probe had a detection p-value >0.01, the probe was a Pidsley cross-reactive probe so it binds to multiple spots on the genome [24], or the probe has known SNPs. Additionally, CpGs with >0.15 difference in mean beta value between the first and second runs for the NC cell lines have been removed. These CpGs wete removed because no difference in mean beta value is expected between the NC EPIC BeadChip runs so this filter accounts for measurement variation inherent in the EPIC BeadChip.

Differentially methylated sites were determined using a linear model with mtDNA-CN as the independent variable and methylation value as the dependent variable using DMPFinder in minfi [22]. A significance threshold of p < 1e-7 was used to determine differentially methylated sites because this is the accepted significance level for epigenome-wide association studies.

Differentially methylated regions were determined using a linear fixed model in DMRcate with treatment status (KO vs NC) as the independent variable and methylation beta value as the dependent variable [25]. The batch (1st and 2nd run) was used as a fixed effect. A significance threshold of p < 1e-7 was used to determine differentially methylated sites which are grouped together into differentially methylated regions. Regions are defined as having a minimum of 10 CpGs with at most 1000 bps between CpGs within the region. Additionally, regions will be removed if the region has a mean beta value <0.05 difference between KO and NC. A significance threshold of p < 1.17e-5 will be used to determine differentially methylated regions using a Bonferroni correction for multiple testing.

**RNA-Seq Analysis:**

Kallisto will be used to pseudo-align the compressed FASTQ files to the Genome Reference Consortium Human Build 37 to generate transcript level counts [26]. To account for uncertainties in transcript counts, a total of 100 bootstraps was performed using Kallisto. Transcript level counts were converted to gene level counts by taking the max transcript count if the gene has multiple transcript isoforms.

Normalization, quality control and differential expression analysis were performed using the EdgeR package [27]. The data was normalized to scale for effective library size of samples using the trimmed mean of M-values (TMM) [28]. Genes with low expression were removed if the gene had less than 15 reads across all samples. In addition, genes are removed if the sum of counts per million (CPM) across samples is < 6. The samples for a gene are included in the sum of CPM for a gene if the CPM for the sample is > 0.6.

Gene expression was modelled using a weighted likelihood empirical Bayes method described in this paper [29]. Differentially expressed genes were determined using a likelihood ratio test between the null and the observed model and the batch (1st and 2nd run) was used as a fixed effect. Genes with a log fold-change (log2FC) < +/-2 were removed. A significance threshold of p < 3.53e-6 was used to determine differentially expressed genes using a Bonferroni correction for multiple testing.

**Integrated Methylation and Gene Expression Analysis:**

The ELMERv2 [30] package was used to integrate differential methylation and differential gene expression results to find inverse associations between methylation and gene expression. This analysis used the supervised analysis option where the methylated and unmethylated groups will be split into KO and NC. The methylation data are the significant CpGs from the differentially methylated site analysis and the gene expression data is the significant genes from the differential expression analysis.

To test for inverse correlations between methylation and gene expression each CpG site was mapped to the nearest 20 genes (10 upstream and 10 downstream) in the genome. Each Gene-CpG pair was tested using a one-sided Mann-Whitney U test to check for difference in methylation between KO and NC. Gene-CpG pairs were removed from the analysis if the CpG is greater than 1Mbp from the transcriptional start site of the gene using the ENSEMBL [31] gene level annotations. A significance threshold of false discovery rate (FDR) < 0.01 is used to determine significant Gene-CpG pairs. In addition, genes in Gene-CpG pairs were tested for differential gene expression between the KO and NC using a student’s t-test. Gene-CpG pairs were removed if the gene did not meet the significance threshold of p < 0.001.

**Functional Enrichment Analyses:**

**GO/KEGG Functional Enrichment:**

Functional Enrichment Analyses were performed on the differential methylation and differential gene expression results on the Gene Ontology (GO) [32,33] database for functional gene groups, the Kyoto Encyclopedia of Genes and Genomes (KEGG) [34–36] for biological pathways, the Reactome Pathway and Transcription Factor gene sets from MSigDB [37,38] and MitoCarta3.0 [39] database for mitochondrial related genes and pathways.

The GO and KEGG functional enrichments were performed on the significant methylation site and methylation region results using missMethyl [40]. CpG sites are mapped to genes and bias from the varying number of CpGs per gene and that a single CpG can be annotated to multiple genes are taken into consideration. The GO and KEGG functional enrichment analyses were performed on the significant gene expression results using limma [41]. The top KEGG pathways were visualized using Pathview to show significant genes within the pathway [42].

**Transcription Factor/Reactome Pathway/MitoCarta3.0 Functional Enrichment:**

The Reactome Pathway, Transcription Factor and MitoCarta3.0 pathways functional enrichments were performed on the significant methylation and gene expression results. Overrepresentation of genes in gene sets was determined by using a student’s t-test on significant genes in the gene set against the genes not in the gene set. A significance threshold of p < 0.05 was used. In addition, the significant genes from the RNA analysis were tested for overlaps with genes in MitoCarta3.0 using a chi-square test. A significance threshold of p < 0.05 was used.

For the GO, KEGG, and Transcription Factor functional enrichments the methylation and gene expression results were combined using the Fisher’s combined probability test to combine p-values [43]. The methylation site and the methylation region results were separately combined with the gene expression results using this method. Functional groups with p < 0.05 in either the methylation or gene expression results were removed before combining p-values.

**DMRichR Functional Enrichment:**

Additional functional enrichment analyses were performed using the DMRichR [44] package to test for enrichment of CpG and gene regions. The enrichment testing for overrepresented CpG and gene regions used a Fisher’s Exact test and a significance threshold of FDR < 0.05 was used. The DMRichR package also includes a wrapper for the LOLA [45] package which tests for enrichment in 15-state chromatin state model from ChromHMM [46] and the 5 related core histone modifications from Roadmap Epigenomics [47]. The enrichment testing for chromatin states and histone modifications used a Fisher’s Exact test and a significance threshold of FDR < 0.05 was used.

**Results:**

**Differentially Methylated Site Analysis:**

We performed an epigenome-wide association study (EWAS) in DNA for 6 TFAM KO cell lines and 6 NC cell lines. After normalization and quality control steps were performed, 769026 CpG sites remain for the differential methylation testing. The linear model identified 4242 independent CpGs spread across the whole genome were identified to be differentially methylated with respect to mtDNA-CN (p < 1e-7) (Figure 3) (Table 1). Of these 4242 CpGs, 1113 CpGs (26%) are hypermethylated and 3129 CpGs (74%) are hypomethylated between the NC and KO groups.

A picture containing timeline

Description automatically generated

**Figure 3.** Manhattan plot of methylation sites across the genome. Significant sites are red (p < 1e-7)

**Table 1.** Top 10 significant CpGs from differentially methylated site analysis

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CpG | Intercept | beta | t | p-value | chr | pos |
| cg06161779 | -5.40932 | 1.054941 | 22.42515 | 7.22E-12 | chr21 | 25801042 |
| cg12369078 | 8.050477 | -1.30233 | -22.3165 | 7.68E-12 | chr13 | 111766257 |
| cg20352894 | -8.87455 | 1.323929 | 21.89414 | 9.82E-12 | chr21 | 36921419 |
| cg16722220 | -5.99367 | 0.936416 | 21.72589 | 1.08E-11 | chr17 | 25309058 |
| cg03670369 | 8.619966 | -1.04951 | -20.8328 | 1.86E-11 | chr6 | 170239250 |
| cg06499112 | -11.0764 | 1.587516 | 20.76871 | 1.93E-11 | chr18 | 64866949 |
| cg05403454 | 3.308109 | -0.73014 | -19.8829 | 3.37E-11 | chr18 | 59666022 |
| cg23865597 | 5.718608 | -0.66731 | -19.3895 | 4.64E-11 | chr18 | 46987786 |
| cg20792408 | 1.86806 | -0.77834 | -19.3793 | 4.67E-11 | chr22 | 50457208 |
| cg07921965 | 5.698131 | -1.1153 | -19.0262 | 5.90E-11 | chr11 | 57283680 |

Previous studies using DNA extracts from blood have identified multiple CpGs which are significantly associated with mtDNA-CN [20,48]. None of the significant CpGs identified in the previous analyses overlap with the significant CpGs identified in this analysis. 4 of the 6 CpGs identified in one of the previous analyses demonstrated the same direction of effect [20]. The remaining 2 CpGs (cg23513930, cg26094004) were removed in quality control or the CpG site is no longer tested on the Illumina EPIC Array. In the other analysis, 15 of the 21 CpGs demonstrated the same direction of effect [48]. 2 of the remaining CpGs (cg21848084, cg26094004) were removed in quality control. The 4 remaining CpGs (cg27187555, cg21393163, cg05673882, and cg27192248) demonstrated the opposite direction of effect (Table 2).

**Table 2**. CpGs identified in two epigenome-wide association studies with mtDNA-CN. \* Signifies opposite direction of effect compared to studies

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Paper | CpG | NCMeans | KOMeans | DeltaMeth | pval |
| Wang\* | cg21393163 | 0.5012 | 0.2233 | 0.2779 | 7.70E-07 |
| Wang | cg02980249 | 0.8706 | 0.7363 | 0.1342 | 1.25E-05 |
| Wang | cg10713715 | 0.7320 | 0.5118 | 0.2201 | 2.76E-05 |
| Wang | cg02194129 | 0.8648 | 0.7405 | 0.1243 | 2.07E-04 |
| Castellani/Wang | cg08899667 | 0.8104 | 0.6859 | 0.1245 | 5.26E-04 |
| Wang | cg00988037 | 0.5774 | 0.4459 | 0.1315 | 9.46E-04 |
| Wang\* | cg27192248 | 0.4646 | 0.1059 | 0.3586 | 3.47E-03 |
| Wang | cg20507228 | 0.8377 | 0.7636 | 0.0742 | 7.83E-03 |
| Wang\* | cg05673882 | 0.2448 | 0.1419 | 0.1029 | 8.18E-03 |
| Wang | cg17619755 | 0.9227 | 0.8604 | 0.0622 | 8.32E-03 |
| Wang | cg02597894 | 0.8694 | 0.7740 | 0.0954 | 9.20E-03 |
| Wang | cg22761205 | 0.7873 | 0.6965 | 0.0908 | 1.07E-02 |
| Wang | cg24420089 | 0.9410 | 0.9194 | 0.0216 | 2.29E-02 |
| Wang | cg09548275 | 0.9648 | 0.9461 | 0.0187 | 2.35E-02 |
| Wang | cg04018738 | 0.9750 | 0.9621 | 0.0130 | 4.89E-02 |
| Castellani | cg03964851 | 0.7937 | 0.7367 | 0.0570 | 6.30E-02 |
| Castellani | cg14575356 | 0.8612 | 0.8257 | 0.0356 | 7.45E-02 |
| Wang\* | cg27187555 | 0.0372 | 0.0565 | -0.0192 | 2.15E-01 |
| Wang | cg03732020 | 0.9817 | 0.9730 | 0.0086 | 2.19E-01 |
| Wang | cg04983687 | 0.9681 | 0.9571 | 0.0110 | 2.53E-01 |
| Wang | cg04368724 | 0.8453 | 0.8134 | 0.0319 | 3.42E-01 |
| Castellani | cg26563141 | 0.3001 | 0.3275 | -0.0274 | 5.13E-01 |

**Differentially Methylated Region Analysis:**

For the differential methylation region analysis, 9110 independent CpGs were identified to be differentially methylated and these CpGs were aggregated into regions. 228 regions were identified to be differentially methylated (p < 1.17e-5) (Table 3). Interestingly, most of the differentially methylated regions show decreased methylation in the KO group compared to the NC group (Figure 4).

**Table 3.** Top 10 differentially methylated regions and genes within region

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Chromosome | Start | End | Width | Num CpGs | Fisher | Genes in Region |
| chr3 | 44902933 | 44904059 | 1127 | 14 | 3.34E-81 | KIF15, TMEM42, MIR564 |
| chr5 | 139926773 | 139927754 | 982 | 14 | 3.02E-75 | ANKHD1, ANKHD1-EIF4EBP3, EIF4EBP3, SRA1 |
| chr20 | 61446962 | 61448534 | 1573 | 39 | 1.02E-72 | COL9A3 |
| chr4 | 154709378 | 154711519 | 2142 | 31 | 1.79E-71 | SFRP2 |
| chr1 | 218518317 | 218520959 | 2643 | 22 | 5.68E-71 | TGFB2, RP11-224O19.2 |
| chr10 | 135049598 | 135052004 | 2407 | 36 | 1.56E-69 | VENTX |
| chr6 | 33244752 | 33246488 | 1737 | 47 | 8.48E-65 | B3GALT4 |
| chr3 | 13589889 | 13591084 | 1196 | 14 | 4.54E-61 | FBLN2 |
| chr7 | 27190849 | 27193351 | 2503 | 15 | 9.46E-58 | HOXA-AS3, HOXA3, RP1-170O19.22, RP1-170O19.23, HOXA7 |
| chr1 | 92949337 | 92950836 | 1500 | 31 | 9.38E-56 | GFI1 |

Graphical user interface, application

Description automatically generated

**Figure 4.** Top 4 differentially methylated regions in order; Top Left, Top Right, Bottom Left, Bottom Right. Top 4 regions demonstrate reduced methylation in KO vs NC.

**Differentially Expressed Genes Analysis:**

We performed a genome-wide association study (GWAS) for 6 TFAM KO cell lines and 6 NC cell lines. After normalization and quality control steps were performed, 14149 genes remained to test for differential gene expression. 179 genes were identified to be differentially expressed (p < 3.53e-6) (Figure 5). Of the 179 genes, 41 genes are over expressed, and 138 genes are under expressed between the NC and KO groups. The top 10 genes in descending order of significance are SFRP2, KCTD8, FLRT2, FBXL7, ABCD1, RIOX1, PGR, EIF4EBP3, IGFBPL1, and LINGO2 (Table 4).

Chart, scatter chart

Description automatically generated

**Figure 5.** Manhattan plot of genes across the genome. Significant genes in red (p < 3.53e-6)

**Table 4.** Top 10 differentially expressed genes from RNA analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gene ID | Symbol | logFC | LR | p-value |
| ENSG00000145423 | SFRP2 | -9.93899 | 504.0956 | 1.22E-111 |
| ENSG00000183783 | KCTD8 | -6.66688 | 407.8657 | 1.07E-90 |
| ENSG00000185070 | FLRT2 | -6.64207 | 335.1203 | 7.37E-75 |
| ENSG00000183580 | FBXL7 | -4.96651 | 289.7323 | 5.69E-65 |
| ENSG00000101986 | ABCD1 | -3.99071 | 279.3776 | 1.03E-62 |
| ENSG00000170468 | RIOX1 | -7.78114 | 267.0318 | 5.03E-60 |
| ENSG00000082175 | PGR | -6.42846 | 249.955 | 2.66E-56 |
| ENSG00000243056 | EIF4EBP3 | -6.72159 | 224.2616 | 1.06E-50 |
| ENSG00000137142 | IGFBPL1 | -3.02228 | 223.9127 | 1.27E-50 |
| ENSG00000174482 | LINGO2 | -6.92435 | 205.5668 | 1.27E-46 |

**Integration Analysis**:

The integration analysis of the significant results from the methylation site and gene expression analysis tested 35347 Gene-CpG pairs for our 4242 significant CpGs. Of these Gene-CpG pairs, 314 pairs have the CpG site +/- 1 Mbp away from the gene transcriptional start site and meet the significance thresholds for differential methylation (FDR < 0.01) and differential expression (p < 0.001). There are 45 unique genes and 286 unique probes which make up the 314 Gene-CpG pairs and 70 pairs have the probe at the transcriptional start site (Figure 6). The top 10 genes from the RNA analysis are displayed with their associated CpG in Table 5.

Graphical user interface

Description automatically generated

**Figure 6.** Example Gene-CpG pairs. Top Left: GABRB1/cg00621837, Top Right: EIF4EBP3/cg00078221, Bottom Left: ABCD1/cg26149887, Bottom Right: SFRP2/cg05774801

**Table 5.** Top 10 significant genes from RNA analysis and associated CpG sites

|  |  |  |  |
| --- | --- | --- | --- |
| GeneID | Symbol | Num CpGs | CpGs |
| ENSG00000145423 | SFRP2 | 6 | cg05160592, cg05164933, cg05774801, cg09689478, cg15974487, cg23121156 |
| ENSG00000185070 | FLRT2 | 6 | cg03480299, cg09435541, cg12223091, cg15922990, cg16324934, cg16759787 |
| ENSG00000183580 | FBXL7 | 25 | cg00254064, cg00723919, cg02390954, cg02704181, cg05023415, cg07856071, cg10729001, cg10948524, cg11032415, cg11043073, cg11232953, cg11339964, cg11973206, cg12478554, cg16210328, cg16641358, cg18202928, cg18620300, cg20328399, cg21306464, cg21477068, cg22094823, cg23041295, cg24640610, cg26233084 |
| ENSG00000101986 | ABCD1 | 9 | cg00207916, cg01022618, cg01083397, cg05365121, cg15373098, cg19514407, cg20844535, cg24324483, cg26149887 |
| ENSG00000170468 | RIOX1 | 7 | cg04508532, cg05031854, cg06842537, cg07122798, cg08311250, cg15079605, cg18107818 |
| ENSG00000243056 | EIF4EBP3 | 17 | cg00078221, cg00275741, cg02434007, cg05197036, cg06048354, cg07124173, cg07810961, cg09574819, cg09936799, cg10731022, cg11639849, cg15923025, cg16316394, cg23247945, cg24897141, cg25870097, cg26084484 |
| ENSG00000163288 | GABRB1 | 13 | cg00621837, cg03095358, cg05629723, cg06259377, cg15294261, cg15452017, cg17680303, cg20111179, cg20544315, cg21472835, cg24949632, cg25463409, cg27184227 |
| ENSG00000137142 | IGFBPL1 | 2 | cg04550644, cg13219226 |
| ENSG00000174482 | LINGO2 | 6 | cg03011514, cg03142841, cg04566296, cg05117544, cg06916852, cg08154622 |

**Functional Enrichment Results:**

**GO/KEGG Functional Enrichment:**

To identify potential biological mechanisms underlying the associations, we performed GO and KEGG enrichment analyses. The GO Fisher’s combined results of the methylation site and gene expression data demonstrate groups relating to the cell membrane, types of ion channels and GABA/GABAA related genes are overrepresented (Table 6). The GO Fisher’s combined results of the methylation region and gene expression data demonstrate that similar groups are overrepresented and the GABA/GABAA related groups becoming the most overrepresented feature (Table 7).

**Table 6.** Top 10 GO Fisher’s Combined results from methylation site and gene expression

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| GO ID | ONT | TERM | p-value Met | p-value RNA | Fisher |
| GO:0031226 | CC | intrinsic component of plasma membrane | 2.61E-09 | 3.63E-08 | 3.58E-15 |
| GO:0005887 | CC | integral component of plasma membrane | 8.45E-09 | 2.90E-08 | 9.05E-15 |
| GO:0071944 | CC | cell periphery | 4.10E-06 | 9.59E-07 | 1.07E-10 |
| GO:0022851 | MF | GABA-gated chloride ion channel activity | 3.17E-05 | 4.32E-07 | 3.56E-10 |
| GO:0016917 | MF | GABA receptor activity | 1.61E-06 | 9.02E-06 | 3.78E-10 |
| GO:0034702 | CC | ion channel complex | 3.17E-07 | 4.88E-05 | 4.00E-10 |
| GO:0004890 | MF | GABA-A receptor activity | 1.16E-05 | 1.97E-06 | 5.84E-10 |
| GO:1902711 | CC | GABA-A receptor complex | 1.16E-05 | 1.97E-06 | 5.84E-10 |
| GO:0022824 | MF | transmitter-gated ion channel activity | 2.21E-06 | 1.63E-05 | 9.01E-10 |
| GO:0022835 | MF | transmitter-gated channel activity | 2.21E-06 | 1.63E-05 | 9.01E-10 |

**Table 7.** Top 10 GO Fisher’s Combined results from methylation region and gene expression

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| GO ID | ONT | TERM | p-value Met | p-value RNA | Fisher |
| GO:0022851 | MF | GABA-gated chloride ion channel activity | 4.49E-05 | 4.32E-07 | 4.98E-10 |
| GO:0016917 | MF | GABA receptor activity | 2.10E-05 | 9.02E-06 | 4.43E-09 |
| GO:0099095 | MF | ligand-gated anion channel activity | 1.41E-04 | 1.97E-06 | 6.40E-09 |
| GO:0004890 | MF | GABA-A receptor activity | 1.72E-04 | 1.97E-06 | 7.76E-09 |
| GO:1902711 | CC | GABA-A receptor complex | 1.72E-04 | 1.97E-06 | 7.76E-09 |
| GO:0005887 | CC | integral component of plasma membrane | 2.61E-02 | 2.90E-08 | 1.67E-08 |
| GO:1902710 | CC | GABA receptor complex | 2.37E-04 | 3.52E-06 | 1.83E-08 |
| GO:0007214 | BP | gamma-aminobutyric acid signaling pathway | 5.62E-05 | 1.91E-05 | 2.33E-08 |
| GO:0031226 | CC | intrinsic component of plasma membrane | 3.22E-02 | 3.63E-08 | 2.52E-08 |
| GO:0062023 | CC | collagen-containing extracellular matrix | 6.70E-05 | 4.60E-05 | 6.34E-08 |

The KEGG Fisher’s combined results of the methylation site with the gene expression results and methylation region with the gene expression results both show that the Nicotine Addiction (path:hsa05033, p < 6.93e-11, p < 1.36e-07), Neuroactive ligand-receptor interaction (path:hsa04080, p < 4.66e-8, p < 4.36e-7), Morphine Addiction (path:hsa05032, p < 1.52e-5, p < 8.53e-4), GABAergic Synapse (path:hsa04727, p < 2.40e-5, p < 2.47e-5), Protein Digestion and Absorption (path:hsa04974, p < 1.47e-4, p < 1.35e-4) and Retrograde Endocannabinoid signaling (path:hsa04723, p < 2.14e-4, p < 1.25e-3) KEGG pathways are overrepresented in both results. In the combined methylation site and gene expression results the Herpes simplex virus 1 infection (path:hsa05168, p < 3.34e-3) and Malaria (path:has:05144, p < 8.82e-3) pathways were overrepresented. In the combined methylation region and gene expression results the Cytokine-cytokine receptor interaction (path:hsa4060, p < 3.91e-3) pathway was overrepresented (Table 8, Table 9).

**Table 8.** KEGG Fisher’s Combined results from methylation site and gene expression

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| KEGG ID | Description | p-value Met | p-value RNA | Fisher |
| path:hsa05033 | Nicotine addiction | 1.70E-06 | 1.47E-06 | 6.93E-11 |
| path:hsa04080 | Neuroactive ligand-receptor interaction | 4.64E-04 | 4.80E-06 | 4.66E-08 |
| path:hsa05032 | Morphine addiction | 8.71E-05 | 1.18E-02 | 1.52E-05 |
| path:hsa04727 | GABAergic synapse | 1.85E-04 | 9.04E-03 | 2.40E-05 |
| path:hsa04974 | Protein digestion and absorption | 2.40E-02 | 4.97E-04 | 1.47E-04 |
| path:hsa04723 | Retrograde endocannabinoid signaling | 1.02E-03 | 1.75E-02 | 2.14E-04 |
| path:hsa05168 | Herpes simplex virus 1 infection | 1.13E-02 | 3.33E-02 | 3.34E-03 |
| path:hsa05144 | Malaria | 3.36E-02 | 3.37E-02 | 8.82E-03 |

**Table 9.** KEGG Fisher’s Combined results from methylation region and gene expression

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| KEGG ID | Description | p-value Met | p-value RNA | Fisher |
| path:hsa05033 | Nicotine addiction | 4.67E-03 | 1.47E-06 | 1.36E-07 |
| path:hsa04080 | Neuroactive ligand-receptor interaction | 4.89E-03 | 4.80E-06 | 4.36E-07 |
| path:hsa04727 | GABAergic synapse | 1.91E-04 | 9.04E-03 | 2.47E-05 |
| path:hsa04974 | Protein digestion and absorption | 2.19E-02 | 4.97E-04 | 1.35E-04 |
| path:hsa05032 | Morphine addiction | 6.94E-03 | 1.18E-02 | 8.53E-04 |
| path:hsa04723 | Retrograde endocannabinoid signaling | 7.14E-03 | 1.75E-02 | 1.25E-03 |
| path:hsa04060 | Cytokine-cytokine receptor interaction | 9.45E-03 | 4.75E-02 | 3.91E-03 |

Interestingly, the Neuroactive ligand-receptor interaction pathway is the second most overrepresented pathway in both methylation site and methylation region combined results. This is interesting because this is the same pathway identified in this previous study [20]. In addition, the GABAergic Synapse pathway in the methylation site combined results demonstrate that the GABAA gene is hypermethylated and the gene expression results demonstrate that the GABAA gene is under expressed (Figure 7). Furthermore, in the Protein Digestion and Absorption pathway the methylation site results demonstrate that the peptidase gene is hypomethylated and gene expression results demonstrate that the peptidase gene is over expressed (Figure 8).

Diagram

Description automatically generated

**Figure 7.** GABAA Receptor genes in the GABAergic Synapse KEGG pathway demonstrate hypermethylation (left) and under expression (right). 4 of the genes that make up the GABAA Receptor are differentially expressed.

Diagram, schematic

Description automatically generated

**Figure 8.** Peptidase genes in the Protein Digestion and Absorption pathway demonstrate hypomethylation (left) and over expression (right).

**Transcription Factor/Reactome Pathway/MitoCarta3.0 Functional Enrichment**

The Transcription Factor Fisher’s combined functional enrichments demonstrate that 11 of the 1128 Transcription Factor gene sets are overrepresented in the combined methylation and gene expression results (Table 10). For the Reactome Pathway and MitoCarta3.0 pathways, the Fisher’s combined functional enrichments did not have any overlapping pathways between the methylation and gene expression results so Fisher’s combined probability test could not be performed. Top individual results are reported instead (Table 11, Table 12). In addition, 5 (ABCD1, ABCD2, PCK2, SARDH, SDSL) of 179 significant genes from the RNA analysis are found in the MitoCarta3.0 database for mitochondrial related genes and all 5 genes are under expressed (Table 13). The Chi-Square test demonstrated that the mitochondrial related genes are underrepresented in the dataset (p < 0.0372).

**Table 10.** Transcription Factor Fisher’s Combined Results

|  |  |  |  |
| --- | --- | --- | --- |
| TF Set | p-value Met | p-value RNA | Fisher |
| HNF4ALPHA\_Q6 | 6.39E-03 | 1.52E-09 | 2.56E-10 |
| UBN1\_TARGET\_GENES | 3.13E-02 | 1.27E-07 | 8.06E-08 |
| GGATTA\_PITX2\_Q2 | 1.65E-03 | 3.53E-03 | 7.62E-05 |
| PAX2\_02 | 8.28E-03 | 2.57E-03 | 2.51E-04 |
| ZNF274\_TARGET\_GENES | 3.70E-02 | 1.75E-03 | 6.91E-04 |
| NFE2L1\_TARGET\_GENES | 5.22E-03 | 1.60E-02 | 8.70E-04 |
| FOXO4\_01 | 5.54E-03 | 4.20E-02 | 2.18E-03 |
| HMG20B\_TARGET\_GENES | 4.72E-02 | 1.12E-02 | 4.51E-03 |
| CHX10\_01 | 2.34E-02 | 2.43E-02 | 4.82E-03 |
| CACCCBINDINGFACTOR\_Q6 | 4.78E-02 | 1.46E-02 | 5.76E-03 |
| E2F2\_TARGET\_GENES | 3.09E-02 | 3.48E-02 | 8.42E-03 |

**Table 11.** MitoCarta3.0 enrichment results for methylation site and gene expression. (Meth is methylation results, RNA is gene expression results)

|  |  |  |  |
| --- | --- | --- | --- |
| Data | MT Path | T.test.pval | Num Genes |
| RNA | Metabolism | 5.11E-02 | 3 |
| Meth | Mitochondrial central dogma | 1.38E-03 | 7 |
| Meth | mtRNA metabolism | 2.76E-03 | 3 |
| Meth | Small molecule transport | 5.80E-03 | 5 |
| Meth | Calcium homeostasis | 1.56E-02 | 3 |
| Meth | Calcium cycle | 1.56E-02 | 3 |

**Table 12.** Reactome Pathway enrichment results for methylation site and gene expression. (Meth is methylation results, RNA is gene expression results)

|  |  |  |  |
| --- | --- | --- | --- |
| Data | RT Path | T-test p-value | Num Genes |
| Meth | REACTOME\_INWARDLY\_RECTIFYING\_K\_CHANNELS | 6.63E-06 | 10 |
| Meth | REACTOME\_CA2\_PATHWAY | 2.85E-05 | 4 |
| Meth | REACTOME\_ESTROGEN\_DEPENDENT\_GENE\_EXPRESSION | 3.15E-05 | 4 |
| Meth | REACTOME\_GASTRIN\_CREB\_SIGNALLING\_PATHWAY\_VIA\_PKC\_AND\_MAPK | 5.05E-04 | 3 |
| Meth | REACTOME\_SENSORY\_PERCEPTION | 6.48E-04 | 23 |
| Meth | REACTOME\_ACTIVATION\_OF\_MATRIX\_METALLOPROTEINASES | 8.76E-04 | 3 |
| Meth | REACTOME\_IMMUNOREGULATORY\_INTERACTIONS\_BETWEEN\_A\_LYMPHOID\_AND\_A\_NON\_LYMPHOID\_CELL | 1.30E-03 | 6 |
| Meth | REACTOME\_BINDING\_AND\_UPTAKE\_OF\_LIGANDS\_BY\_SCAVENGER\_RECEPTORS | 1.36E-03 | 5 |
| Meth | REACTOME\_SCAVENGING\_BY\_CLASS\_A\_RECEPTORS | 1.36E-03 | 5 |
| Meth | REACTOME\_TRANSPORT\_OF\_INORGANIC\_CATIONS\_ANIONS\_AND\_AMINO\_ACIDS\_OLIGOPEPTIDES | 2.60E-03 | 14 |
| RNA | REACTOME\_SLC\_MEDIATED\_TRANSMEMBRANE\_TRANSPORT | 3.19E-04 | 3 |
| RNA | REACTOME\_METABOLISM\_OF\_AMINO\_ACIDS\_AND\_DERIVATIVES | 3.21E-03 | 5 |
| RNA | REACTOME\_TRANSPORT\_OF\_SMALL\_MOLECULES | 1.25E-02 | 8 |
| RNA | REACTOME\_REPRODUCTION | 1.42E-02 | 3 |
| RNA | REACTOME\_SIGNALING\_BY\_RECEPTOR\_TYROSINE\_KINASES | 2.49E-02 | 11 |
| RNA | REACTOME\_SIGNALING\_BY\_ERBB4 | 4.55E-02 | 4 |

**DMRichR/LOLA Functional Enrichment:**

The enrichment testing for CpG, gene, and chromatin state regions demonstrate that the CpG shelves, 5’ UTR, Exon, 3’ UTR, Intergenic, Enhancers, ZNF Genes and Repeats, heterochromatin, Bivalent TSS, Flanking Bivalent TSS/Enh, Bivalent Enhancer, Repressed PolyComb, Weak Repressed PolyComb and Quiescent/Low group all meet the significance threshold FDR < 0.05 (Figure 9) (Table 11).

Chart, bar chart

Description automatically generated

**Figure 9.** Functional enrichment of CpG, gene, and chromatin state regions in methylation data. \* indicate significant enrichment (FDR < 0.05)

**Table 11.** Odds ratio and FDR of functional regions tested for enrichment

|  |  |  |
| --- | --- | --- |
| Functional Region | OR | FDR |
| Weak Repressed PolyComb | 2.403685 | 3.11E-72 |
| Heterochromatin | 2.199333 | 4.81E-71 |
| Repressed PolyComb | 2.003021 | 2.64E-61 |
| Quiescent/Low | 2.188101 | 4.83E-41 |
| Flanking Bivalent TSS/Enhancer | 1.778229 | 2.71E-40 |
| Bivalent Enhancer | 1.733487 | 2.82E-39 |
| Enhancers | 1.45079 | 1.31E-15 |
| Intergenic | 1.396322 | 5.51E-10 |
| 5' UTR | 0.415632 | 9.07E-05 |
| ZNF Genes & Repeats | 1.241154 | 2.36E-04 |
| Bivalent/Poised TSS | 1.149232 | 8.07E-04 |
| Exon | 0.782081 | 1.96E-02 |
| CpG Shelves | 0.819122 | 3.47E-02 |
| 3' UTR | 0.772858 | 4.75E-02 |
| Active TSS | 1.079673 | 5.03E-02 |
| Intron | 0.937317 | 2.43E-01 |
| Open Sea | 1.057879 | 3.90E-01 |
| CpG Shores | 1.046642 | 4.60E-01 |
| CpG Islands | 0.972323 | 5.37E-01 |
| Promoter | 0.971709 | 5.90E-01 |
| Weak Transcription | 0.984861 | 8.68E-01 |
| Downstream | 0.992659 | 1.00E+00 |
| Flanking Active TSS | 0.935926 | 1.00E+00 |
| Strong Transcription | 0.681818 | 1.00E+00 |
| Transcription at Gene 5' and 3' | 0.595448 | 1.00E+00 |
| Genic Enhancers | 0.568981 | 1.00E+00 |

Interestingly, the 5’ UTR and Intergenic regions from the gene region enrichment testing along with the different enhancer types, heterochromatin and repressed PolyComb regions from the chromatin state enrichment testing demonstrate that epigenetic marks in non-coding regions such as in enhancers are most overrepresented in the methylation data.

In the enrichment testing for the 5 related core histone modifications demonstrate that H3K4me3 and H3K36me3 histone markers are underrepresented which means that methylation in promoters and gene bodies is decreased (Figure 10). In addition, the H3K27me3 and H3K9me3 histone markers are overrepresented which means that methylation in polycomb repression and heterochromatin is increased.

Chart

Description automatically generated

**Figure 10.** Heatmap of 5 core histone modifications for 127 epigenomes from Roadmap epigenomics

**Discussion**:

To determine which genes and biological pathways mediate the effect of mtDNA-CN on differential nDNA methylation and differential gene expression leading to age-related diseases, we conducted an epigenome-wide association study (EWAS), genome-wide association study (GWAS) and functional enrichment analyses to investigate mtDNA-CN effect on nDNA methylation. Previous studies have proposed that metabolites such as Acetyl-CoA and Alpha-ketoglutarate important in facilitating nDNA methylation by mtDNA-CN [20]. These results do not directly demonstrate that the proposed metabolites are impacting nDNA methylation. Other potential biological mechanisms are proposed.

**KEGG Pathways and GABAA Receptor:**

Our results from the KEGG Fisher’s combined probability test demonstrate that the Nicotine Addiction, Neuro-Active Ligand Receptor Interaction, Morphine Addiction, GABAergic Synapse, Protein Digestion and Absorption and Retrograde Endocannabinoid Signalling pathways are overrepresented and these pathways are related to cell-signalling and the nervous system. Genes in the overrepresented pathways likely have importance in facilitating the effect of mtDNA-CN has on nDNA methylation.

Previous studies have shown that the Neuro-Active Ligand Receptor Interaction pathway is overrepresented across multiple independent cohorts [20]. The overrepresentation of this pathway in this analysis provides further evidence that it is important in understanding mtDNA-CN effect on nDNA methylation. In addition, the Cytokine-Cytokine Receptor Interaction pathway was also overrepresented in the methylation site and gene expression results. These two KEGG pathways both fall under the Environmental Information Processing and Signaling Molecules and Interaction KEGG category which further show that cellular signaling is important for mtDNA-CN effect on nDNA methylation.

5 of 6 KEGG pathways shown in both Fisher combined results contain GABA related genes. In addition, our GO results demonstrate the GABA and GABAA related gene groups are overrepresented in our dataset. The RNA analysis shows 4 of the subunit genes [49] (GABRB1, GABRB3, GABRE, GABRG1) that form the GABAA receptor are significant, and all of these genes are under expressed in the knockout. In addition, the integration analysis demonstrated that the GABRB1 gene is inversely related to methylation of 13 CpGs. The GABAA receptor functions as a ligand-gated ion channel and allows for fast inhibitory synaptic transmission [49]. With reduced expression of the GABAA receptor genes, fewer receptors will be present to bind GABA and therefore cannot initiate as many inhibitory synaptic signals. GABA has been shown to be important in a variety of cellular processes such as suppressing inflammatory immune responses [50] and binding to islet alphas-cells to inhibit secreting glucagon [51]. In addition, a study in a mouse model demonstrated that mice lacking GABA-synthesizing genes die at birth despite having no visible structural differences in their brain [49]. Therefore, one potential biological mechanism is that mtDNA-CN causes hypermethylation of GABAA receptor genes leading to fewer GABAA receptors to receive inhibitory signals for essential cellular processes ultimately leading to age-related chronic diseases such as diabetes [51].

**Mitochondria Related Genes: ABCD1/ABCD2**

The ABCD1 gene functions as an ATP-binding cassette transporter which helps move very long chain fatty acid (VLCFA)-CoA from the cytosol to the peroxisome [52]. This gene also regulates mitochondrial functions such as oxidative phosphorylation and helps to degrade and synthesize fatty acids [52]. It is also known that this gene is important in X-linked adrenoleukodystrophy which is a rare genetic peroxisome disease affecting the nervous system and kidney [52]. In our results, 5 of the 179 significant genes were part of MitoCarta3.0. The ABCD1 gene was the only one of these 5 genes which demonstrated an inverse association with methylation. Furthermore, this gene and the paralog, ABCD2, are both under expressed in our results and the metabolism pathway was the only overrepresented pathway using the RNA results. Therefore, mtDNA-CN may lead to hypermethylation of ABCD1 leading to decreased expression ultimately reducing mitochondrial function and energy metabolism by reducing the amount of VLCFA-CoA that can be used by the cell.

**Chromatin States and Histones (RIOX1):**

Our functional enrichment results for histone markers demonstrates that H3K27me3 and H3K9me3, which are related to polycomb repression and heterochromatin respectively, are hypermethylated while H3K4me3 and H3K36me3, which are related to promoters and gene bodies respectively, are hypomethylated across the 127 reference epigenomes [53]. Increased H3K27me3 and H3K9me3 methylation reduces transcription of genes by creating heterochromatin [53]. Decreased H3K4me3 and H3K36me3 methylation is associated with decreased transcription of genes and decreased methylation of gene bodies are associated with genes that change their expression during aging [53].

The RIOX1 gene is a ribosomal oxygenase which has been reported to demethylate H3K4me1/2/3 and H3K36me3/2 and preferentially demethylates H3K4me3 and H3K4me1 [54]. Our results demonstrate that the RIOX1 gene is significantly under expressed, but no significant Gene-CpG pairs were identified that included this gene. Therefore, mtDNA-CN may cause variable methylation of histones leading to overall reduced transcription and variable expression of age-related genes via RIOX1 ultimately disrupting various cellular processes leading to age-related chronic diseases.

**Protein Synthesis and Cancer: EIF4EBP3**

EIF4E is Eukaryotic initiation factor 4E which functions as the rate limiting step of protein synthesis [55]. This protein is part of a the EIF4F complex which recognizes the 7-methylguanosine cap at the 5’ end of messenger RNA (mRNA) to initiate translation [55]. EIF4EBP3 encodes a protein which binds to EIF4E to prevent EIF4E to assemble with other proteins to create the EIF4F complex [55]. Our results demonstrate that EIF4EBP3 is under expressed and this gene is inversely related to methylation at 17 CpG sites. With decreased expression of EIF4EBP3, reduced binding of this gene to EIF4E will occur leading to increased number of EIF4F complexes that can guide mRNA to be translated ultimately leading to global increased protein synthesis. Increased levels of EIF4E have been shown to facilitate cellular transformation, increase tumorigenesis and increase metastasis [56]. Therefore, mtDNA-CN may cause hypermethylation of EIF4EBP3 decreasing gene expression leading to increased protein synthesis and ultimately causing age-related chronic diseases such as cancer.

**Limitations:**

This study has some potential limitations. One possible limitation of the results is that the TFAM knockout reduced mtDNA-CN by a larger reduction than commonly seen in patients. This may have uncovered some relationships that would not be seen *in vivo* due to the larger variation leading to more significant nDNA methylation changes. The TFAM gene may also independently impact other pathways other than regulating mtDNA-CN and therefore could impact nDNA methylation through unknown biological mechanisms. In addition, the cell line used to create the KO and NC cell lines was HEK293T cells. These cells come from kidney and may have specific methylation patterns that may not be found in other cell types such as in blood or skeletal muscle. The methylation data was also generated at two separate time points which may have caused unknown changes to the cell characteristics. Furthermore, the integration analysis only checked for inverse relationships between methylation and gene expression so if there were any positive correlations between methylation and gene expression these would not be found.

**Future Steps:**

Our results also demonstrate significant enrichments in chromatin state regions such as enhancers are significantly overrepresented in our dataset. In addition, CpG related regions (CpG Islands, CpG Shores and Open Sea) are not overrepresented in our dataset and CpG shelves just meet our significance threshold. These results imply that mtDNA-CN not only affects CG rich regions, but also interacts with chromatin states to facilitate the effect mtDNA-CN has on disease. Therefore, future studies should investigate the associations between mtDNA-CN and chromatin states using ATAC-Seq or ChIP-seq.

**Conclusion:**

Many studies have demonstrated the relationship between variation in mtDNA-CN and nDNA methylation, but the mechanisms of this relationship are not well understood [18,19]. In this study we have identified associations between mtDNA-CN and nDNA methylation across the genome. We performed GO/KEGG functional enrichment analyses on the significant results that demonstrated that the GABAA receptor genes are overrepresented and may facilitate the link between mtDNA-CN and nDNA methylation. In addition, our integration analysis demonstrate multiple Gene-CpG pairs have an inverse relationship between methylation and gene expression and identifies some potential biological mechanisms which facilitate the effect of mtDNA-CN. Furthermore, enrichment of functional genomic regions demonstrated that chromatin states such as enhancers and heterochromatin are overrepresented in the methylation data implying that mtDNA-CN may also be impacting gene expression via chromatin states. These results demonstrate that mtDNA-CN impacts nDNA methylation contributing to age-related chronic diseases through multiple biological mechanisms.

**References:**

1. Modica-Napolitano JS, Singh KK. Mitochondrial dysfunction in cancer. Mitochondrion. 2004;4:755–62.

2. Taylor RW, Turnbull DM. Mitochondrial DNA mutations in human disease. Nature Reviews Genetics. 2005. p. 389–402.

3. Ballinger SW. Mitochondrial dysfunction in cardiovascular disease. Free Radical Biology and Medicine. Elsevier Inc.; 2005. p. 1278–95.

4. Arruda AP, Pers BM, Parlakgül G, Güney E, Inouye K, Hotamisligil GS. Chronic enrichment of hepatic endoplasmic reticulum-mitochondria contact leads to mitochondrial dysfunction in obesity. Nature Medicine. Nature Publishing Group; 2014;20:1427–35.

5. Dang C v. Links between metabolism and cancer. Genes and Development. 2012. p. 877–90.

6. Wai T, Ao A, Zhang X, Cyr D, Dufort D, Shoubridge EA. The role of mitochondrial DNA copy number in mammalian fertility. Biology of Reproduction. Society for the Study of Reproduction; 2010;83:52–62.

7. Knez J, Winckelmans E, Plusquin M, Thijs L, Cauwenberghs N, Gu Y, et al. Correlates of Peripheral Blood Mitochondrial DNA Content in a General Population. American Journal of Epidemiology. Oxford University Press; 2016;183:138–46.

8. Jeng J-Y, Yeh T-S, Lee J-W, Lin S-H, Fong T-H, Hsieh R-H. Maintenance of mitochondrial DNA copy number and expression are essential for preservation of mitochondrial function and cell growth. Journal of Cellular Biochemistry. 2008;103:347–57.

9. Ashar FN, Moes A, Moore AZ, Grove ML, Chaves PHM, Coresh J, et al. Association of mitochondrial DNA levels with frailty and all-cause mortality. Journal of Molecular Medicine. Springer Verlag; 2015;93:177–86.

10. Bernstein BE, Meissner A, Lander ES. The Mammalian Epigenome. Cell. Elsevier B.V.; 2007. p. 669–81.

11. Portela A, Esteller M. Epigenetic modifications and human disease. Nature Biotechnology. 2010. p. 1057–68.

12. Li E, Beard C, Jaenlsch R. Role of DNA methylation in genomic imprinting. Nature. 1993;366:362–5.

13. Cagin U, Enriquez JA. The complex crosstalk between mitochondria and the nucleus: What goes in between? International Journal of Biochemistry and Cell Biology. Elsevier Ltd; 2015. p. 10–5.

14. Delsite R, Kachhap S, Anbazhagan R, Gabrielson E, Singh KK. Nuclear genes involved in mitochondria-to-nucleus communication in breast cancer cells. Molecular Cancer. 2002;1:6.

15. Bellizzi D, Daquila P, Giordano M, Montesanto A, Passarino G. Global DNA methylation levels are modulated by mitochondrial DNA variants. Epigenomics. 2012;4:17–27.

16. Smiraglia D, Kulawiec M, Bistulfi GL, Ghoshal S, Singh KK. A novel role for mitochondria in regulating epigenetic modifications in the nucleus. Cancer Biology & Therapy. 2008;7:1182–90.

17. Vivian CJ, Brinker AE, Graw S, Koestler DC, Legendre C, Gooden GC, et al. Mitochondrial genomic backgrounds affect nuclear DNA methylation and gene expression. Cancer Research. American Association for Cancer Research Inc.; 2017;77:6202–14.

18. Sun X, Johnson J, St John JC. Global DNA methylation synergistically regulates the nuclear and mitochondrial genomes in glioblastoma cells. Nucleic Acids Research. Oxford University Press; 2018;46:5977–95.

19. Xie C, Naito A, Mizumachi T, Evans TT, Douglas MG, Cooney CA, et al. Mitochondrial regulation of cancer associated nuclear DNA methylation. Biochemical and Biophysical Research Communications. 2007;364:656–61.

20. Castellani CA, Longchamps RJ, Sumpter JA, Newcomb CE, Lane JA, Grove ML, et al. Mitochondrial DNA copy number can influence mortality and cardiovascular disease via methylation of nuclear DNA CpGs. Genome Medicine. BioMed Central Ltd; 2020;12.

21. Ekstrand MI, Falkenberg M, Rantanen A, Park CB, Gaspari M, Hultenby K, et al. Mitochondrial transcription factor A regulates mtDNA copy number in mammals. Human Molecular Genetics. 2004;13:935–44.

22. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. Oxford University Press; 2014;30:1363–9.

23. Fortin JP, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics. Oxford University Press; 2017;33:558–60.

24. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biology. BioMed Central Ltd.; 2016;17.

25. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, v Lord R, et al. De novo identification of differentially methylated regions in the human genome. Epigenetics & Chromatin. 2015;8:6.

26. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology. Nature Publishing Group; 2016;34:525–7.

27. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. Oxford University Press; 2009;26:139–40.

28. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology. 2010;11.

29. Chen Y, Lun ATL, Smyth GK. Differential Expression Analysis of Complex RNA-seq Experiments Using edgeR \* [Internet]. 2014. Available from: http://www.bioconductor.org.

30. Silva TC, Coetzee SG, Gull N, Yao L, Hazelett DJ, Noushmehr H, et al. ELmer v.2: An r/bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. Bioinformatics. Oxford University Press; 2019;35:1974–7.

31. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, et al. Ensembl 2016. Nucleic Acids Research. Oxford University Press; 2016;44:D710–6.

32. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: Tool for the unification of biology. Nature Genetics. 2000. p. 25–9.

33. Carbon S, Douglass E, Good BM, Unni DR, Harris NL, Mungall CJ, et al. The Gene Ontology resource: Enriching a GOld mine. Nucleic Acids Research. Oxford University Press; 2021;49:D325–34.

34. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes [Internet]. Nucleic Acids Research. 2000. Available from: http://www.genome.ad.jp/kegg/

35. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Science. Blackwell Publishing Ltd; 2019. p. 1947–51.

36. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: Integrating viruses and cellular organisms. Nucleic Acids Research. Oxford University Press; 2021;49:D545–51.

37. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. Cell Systems. Cell Press; 2015;1:417–25.

38. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–50.

39. Rath S, Sharma R, Gupta R, Ast T, Chan C, Durham TJ, et al. MitoCarta3.0: An updated mitochondrial proteome now with sub-organelle localization and pathway annotations. Nucleic Acids Research. Oxford University Press; 2021;49:D1541–7.

40. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina’s HumanMethylation450 platform. Bioinformatics. 2016;32:286–8.

41. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biology. 2010;11.

42. Luo W, Brouwer C. Pathview: An R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics. 2013;29:1830–1.

43. Dewey M. metap: meta-analysis of significance values. 2021.

44. Laufer BI, Hwang H, Jianu JM, Mordaunt CE, Korf IF, Hertz-Picciotto I, et al. Low-pass whole genome bisulfite sequencing of neonatal dried blood spots identifies a role for RUNX1 in down syndrome DNA methylation profiles. Human Molecular Genetics. Oxford University Press; 2020;29:3465–76.

45. Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinformatics. 2016;32:587–9.

46. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nature Methods. 2012;9:215–6.

47. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH roadmap epigenomics mapping consortium. Nature Biotechnology. 2010. p. 1045–8.

48. Wang P, Castellani CA, Yao J, Huan T, Bielak LF, Zhao W, et al. Epigenome-wide association study of mitochondrial genome copy number. Human Molecular Genetics. Oxford University Press; 2022;31:309–19.

49. Owens DF, Kriegstein AR. Is there more to GABA than synaptic inhibition? Nature Reviews Neuroscience. 2002. p. 715–27.

50. Tian J, Lu Y, Zhang H, Chau CH, Dang HN, Kaufman DL. γ-Aminobutyric Acid Inhibits T Cell Autoimmunity and the Development of Inflammatory Responses in a Mouse Type 1 Diabetes Model. The Journal of Immunology. The American Association of Immunologists; 2004;173:5298–304.

51. Soltani N, Qiu H, Aleksic M, Glinka Y, Zhao F, Liu R, et al. GABA exerts protective and regenerative effects on islet beta cells and reverses diabetes. Proc Natl Acad Sci U S A. 2011;108:11692–7.

52. Tawbeh A, Gondcaille C, Trompier D, Savary S. Peroxisomal abc transporters: An update. International Journal of Molecular Sciences. MDPI; 2021.

53. Zhang Y, Sun Z, Jia J, Du T, Zhang N, Tang Y, et al. Overview of Histone Modification. 2021. p. 1–16.

54. Sinha KM, Yasuda H, Coombes MM, Yr Dent S, de Crombrugghe B. Regulation of the osteoblast-specific transcription factor Osterix by NO66, a Jumonji family histone demethylase. EMBO Journal. 2010;29:68–79.

55. Batool A, Aashaq S, Andrabi KI. Eukaryotic initiation factor 4E (eIF4E): A recap of the cap‐binding protein. Journal of Cellular Biochemistry. 2019;120:14201–12.

56. de Benedetti A, Graff JR. eIF-4E expression and its role in malignancies and metastases. Oncogene. 2004. p. 3189–99.