Tissue-specific mtDNA abundance from exome data and its correlation with mitochondrial transcription, mass and respiratory activity

D'Erchia Anna Maria[1], Atlante Anna[2], Gadaleta Gemma[1], Pavesi Giulio[3], Chiara Matteo[3], De Virgilio Caterina[1], Manzari Caterina[2], Mastropasqua Francesca[1], Prazzoli Gian Marco[3], Picardi Ernesto[1], Gissi Carmela[3], Horner David[3], Reyes Aurelio[4], Sbisà Elisabetta[5], Tullo Apollonia[5], and Pesole Graziano[1,2] ^

[1] Dipartimento di Bioscienze, Biotecnologie e Biofarmaceutica, Università degli Studi di Bari Aldo Moro, Via Orabona 4, 70126 Bari, Italia

[2] Istituto di Biomembrane e Bioenergetica, CNR, via Amendola 165/A, 70126 Bari, Italia

[3] Dipartimento di Bioscienze, Università degli Studi di Milano, Via Celoria 26, 20133 Milano, Italia

[4] Mitochondrial Biology Unit, Medical Research Council, Wellcome Trust/MRC Building Hills Road, Cambridge CB2 0XY, United Kingdom

[5] Istituto di Tecnologie Biomediche- Sede di Bari, CNR, Via Amendola 122/D, 70126 Bari, Italia

^ *Corresponding Author:* via Amendola 165/A, 70125 Bari, Italy. Tel. +390805443588, Fax +390805443317, email graziano.pesole@uniba.it

Running Title: **Tissue-specific mtDNA abundance from exome data**

Abstract

Eukaryotic cells contain a population of mitochondria, variable in number and shape, which in turn contain multiple copies of a tiny compact genome (mtDNA) whose expression and function is strictly coordinated with the nuclear one. mtDNA copy number varies between different cell or tissues types, both in response to overall metabolic and bioenergetics demands and as a consequence or cause of specific pathological conditions. Here we present a novel and reliable methodology to assess the effective mtDNA copy number per diploid genome by investigating off-target reads obtained by whole-exome sequencing (WES) experiments. We also investigate whether and how mtDNA copy number correlates with mitochondrial mass, respiratory activity and expression levels. Analyzing six different tissues from three age- and sex-matched human individuals, we found a highly significant linear correlation between mtDNA copy number estimated by qPCR and the frequency of mtDNA off target WES reads. Furthermore, mtDNA copy number showed highly significant correlation with mitochondrial gene expression levels as measured by RNA-Seq as well as with mitochondrial mass and respiratory activity. Our methodology makes thus feasible, at a large scale, the investigation of mtDNA copy number in diverse cell-types, tissues and pathological conditions or in response to specific treatments.

Introduction

Mitochondria play a range of critical roles in the life of eukaryotic cells (Pesole et al., 2012). They are the commonly referred to as the "power-stations" of the cell—for their provision of ATP through oxidative phosphorylation (OXPHOS), but they are also responsible for the biosynthesis of numerous macromolecules (lipids, proteins and nucleic acids) and contribute to the regulation of apoptosis, cell proliferation and motility. Individual cells, tissues and organs have distinct metabolic profiles and energy demands, which can change in response to environmental stimuli, alteration of physiological status or the onset of pathological conditions (Kunz, 2003, Leary et al., 1998, Leverve and Fontaine, 2001, Pfeiffer et al., 2001). Variation in mitochondrial respiratory capacity between tissues is also related to mitochondrial function, protein composition and morphology (Benard et al., 2006, Johnson et al., 2007a, Johnson et al., 2007b, Mootha et al., 2003, Pagliarini et al., 2008).

Mitochondria are endowed of their own genetic system (mtDNA), a legacy of the endosymbiotic event that can be considered to represent the origin of eukaryotes and which occurred some 1.5–2 billion years ago (Lane and Martin, 2010). In mammals, the mitochondrial genome (mtDNA), a double stranded circular macromolecule, of around 16.5 kbp in length, is uniparentally inherited from the mother. Numerous mitochondria, each with multiple copies of the mtDNA, are present in each cell and form a highly dynamic network, the result of continuous fusion and fission processes (Rafelski, 2013). The function and activity of mitochondria is strictly regulated through the coordinated expression of mitochondrial and nuclear genomes. The mtDNA copy number per diploid nuclear genome correlates with ATP production and can range between 1000 and 5000 (Bogenhagen and Clayton, 1974, Shmookler Reis and Goldstein, 1983). mtDNA copy number is specific to tissue types and developmental stages, is under strict regulation (Clay Montier et al., 2009, Moraes, 2001), apparently a reflection of differing energy requirements. Altered mtDNA copy number is associated with oxidative stress and with several pathological conditions, including neuromuscular diseases, cardiomyopathy, type 2 diabetes and cancer (Clay Montier et al., 2009).

A comprehensive vision of tissue- and pathology-related variability in mtDNA copy number per cell is lacking and few studies have attempted to correlate mtDNA copy number with mitochondrial mass, respiratory activity or mitochondrial gene expression levels.

We previously demonstrated that off-target reads from human whole-exome sequencing (WES) can be used to assemble mitochondrial genomes (Picardi and Pesole, 2012). Here, we investigate if the relative abundance of mtDNA reads obtained by WES experiments may be a reliable indicator of the effective mtDNA copy number per diploid genome, and whether and how mtDNA copy number correlates with global mitochondrial gene expression levels measured by RNA-Seq.

In the current study, we have estimated mtDNA copy number by qPCR (Venegas et al., 2011) as well as by the proposed WES-derived protocol for liver, kidney, brain, lung, muscle and heart samples from three age- and sex-matched human individuals. All samples were also subjected to whole transcriptome sequencing as well as to citrate synthase and cytochrome oxidase assays.

A highly significant linear correlation between qPCR data and the frequency of mtDNA off target reads imply that, in addition to allowing the reconstruction of complete or nearly

complete mitochondrial genomes (Picardi and Pesole, 2012), WES data permit accurate quantification of mtDNA copy number. Furthermore, mtDNA copy number showed highly significant correlation with mitochondrial gene expression levels as measured using RNA-Seq as well as with key functional data. The relative abundance of distinct mtDNA-derived transcripts was tissue-specific and highly replicable.

Materials and Methods

Samples and nucleic acids extraction

Six different post-mortem human snap-frozen tissues (brain, liver, lung, striated muscle, kidney and heart) from three unrelated healthy Caucasian individuals (males, aged 47–54 years) were obtained from Cureline (South San Francisco, CA, USA).

The three sample IDs are S7/11, S12/12 and S13/12, and all sample details are reported in Supplemental Table S1.

DNA was purified using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions, quantified and qualitatively checked on NanoDrop 2000c (Thermo Fisher Scientific, USA).

Total RNA was purified using the RNeasy Plus Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. RNA quality was assessed on Agilent Bioanalyzer 2100, obtaining RIN (RNA Integrity Number) values ranging from 5 to 7, that were considered acceptable for RNA derived from post-mortem tissues.

Quantification of mtDNA content

Relative mtDNA copy number was measured by qPCR with SYBR detection using primers specific for the mitochondrial tRNA Leu$^{(UUR)}$ gene and the single copy nuclear ß-2-microglobulin (ß2M) gene (Venegas et al., 2011). All reactions were performed using 5 ng of total DNA as template on a ABI Prism 7000 sequence detector system (Applied Biosystems, Foster City, CA, USA), according to this two-step thermal cycling protocol: 50 °C for 2 min (UDG pre-treatment); 95 °C for 10 min (initial denaturation); 40 cycles at 95 °C for 15 s and 62 °C for 1 min, followed by a melting curve analysis (95 °C for 15 s, 60 °C for 30s and 95 °C for 15 s) to verify the specificity and identity of the PCR product. Primers sequence, amplicon size and annealing temperature are reported in the Supplemental Table S2.

The intensity of SYBR fluorescent signals were then analyzed by the SDS software (version 1.2.3), and the C_T value for each qPCR was used to calculate the mtDNA content by difference in C_T values between the tRNA Leu$^{(UUR)}$ and ß2M genes (ΔC_T). mtDNA content was obtained using the formula $2 \times 2^{-(\Delta CT)}$. Results represent the average of three independent experiments performed on the same DNA preparation and are shown with standard deviations.

Exome sequencing and mtDNA assembly

Exome capture was performed using the TruSeq Exome Enrichment Kit (Illumina, San Diego, CA), according to the manufacturer's instructions. Briefly, for each tissue, a DNA library, including inserts ranging in size from 200 to 400 bp approximately, was prepared using the TruSeq DNA Sample Prep kit (Illumina). Then, each library was hybridated with biotinylated probes targeting the exonic regions (about 200,000 exons, covering about 62 Mb

of the human genome). After two steps of enrichment with the probes, the captured exonic regions were sequenced on the Illumina HiSeq 2000 sequencer, at IGA Technology Services in Udine (Italy), generating for each tissue approximately 40 million of 100 bp paired-end reads.

Exome reads were mapped onto the Revised Cambridge Reference Sequence (rCRS with GenBank accession number NC_012920) of human mtDNA (Andrews et al., 1999) using GSNAP program version 2013-07-14 (Wu and Nacu, 2010) since it enables the handling of circular genomes.

Aligned reads were mapped again onto the complete human genome (assembly hg19 including the rCRS sequence) using GSNAP in order to exclude read pairs also mapping on nuclear mitochondrial DNA sequences (i.e., Numts) (for further details see Picardi and Pesole, 2012). The complete mapping procedure was automated by using a custom python script (mapExome.py), available upon request.

Mitochondrial reads in SAM format were converted into BAM and then pileup format using *samtools* (version 0.1.18) (Li et al., 2009). The final pileup file was parsed position-by-position in order to calculate the distribution of nucleotides aligned at each position, removing bases with a quality score less than 25. For each position of the rCRS sequence, supported by at least 5 independent reads, the consensus base was calculated using a minimum confidence level of 0.75.

Contiguous consensus positions were grouped in contigs and assembled into a final mitochondrial genome using a custom python script previously developed in our group (Picardi and Pesole, 2012).

The relative amounts of off-target mtDNA reads were calculated as the number of reads mapping on mtDNA per million mapped reads.

Strand-oriented RNA-Sequencing and analysis

For each tissue, a strand-oriented RNA library was prepared to preserve information about which DNA strand was the original template during the synthesis of transcripts, thus offering strand orientation for detection of antisense transcription and providing information about regulatory relationships.

The cytoplasmatic rRNA removal was performed for each total RNA sample using the Ribo-Zero rRNA removal Kit (Epicentre, Madison, WI, USA). The rRNA-depleted RNA was used to prepare the stranded-oriented RNA-seq library using the TruSeq Stranded Total RNA Sample Prep Kit (Illumina, San Diego, CA, USA), according to the manufacturer's instructions. Briefly, each RNA was chemically fragmented prior to the random priming reverse transcription reaction for first strand cDNA generation. The fragmentation step resulted in an RNA-seq library including inserts ranging in size from approximately 100–400 bp. During the second-strand synthesis, dUTP was incorporated in place of dTTP, thus preventing amplification of this strand during the subsequent PCR step and retaining strand information. cDNA libraries were sequenced on the Illumina HiSeq 2000 platform at IGA Technology Services in Udine (Italy), generating for each tissue sample approximately from 27 to 35 million 100 bp paired-end reads.

In order to investigate the expression level of mitochondrial-encoded transcripts, we used the reference annotation in Supplemental Table S3, which also includes long non-coding RNAs (Rackham et al., 2011).

No reads that could be attributed uniquely to mature tRNAs were mapped, as expected, as the procedure used for the RNA extraction does not provide an enrichment for transcripts < 200 nucleotides long (such as miRNAs and tRNAs).

Sequences were first mapped against the human genome (hg19) using Tophat with default parameters, with a percentage of mapped reads in the different samples ranging from 88% to 94%. Sequence read pairs that mapped on mtDNA as a correct forward-reverse pair, with a maximum insert size of 1000 bp were considered to be initial mitochondrial candidate reads.

As expected, a non-negligible fraction of the latter (about 8% of the sequence pairs) resulted to be mapped on Numt regions of the nuclear genome as well, either as a singleton (only one of the two reads mapping), as a chimeric pair mapping on different nuclear chromosomes, or as a pair on the same chromosome. Since having a correct estimation of the number of sequence reads that could be reliably assigned to mitochondrial genes was essential for this study, we further investigated this issue. A comparison of matching against the mitochondrial and the nuclear genomes revealed that for more of the 90% of the ambiguously mapped sequence pairs a match was found for only one of the two paired reads on the nuclear genome. Only about 1% of the reads mapping as paired on the mitochondrial genome was mapped on Numts as correct forward–reverse pair, however with a number of mismatches equal to or greater than the one resulting from the mapping on the mitochondrion.

All in all, these results show how mapping ambiguity between the Numt sequences and the mitochondrion can be resolved by employing paired-end sequences. All reads mapping as a correctly oriented pair on the mitochondrial sequence can be considered for further analyses independently of mapping on the nuclear genome.

Starting from reads mapped on the mitochondrial genome, expression of genes was estimated by using the annotation summarized in Supplemental Table S3. Sequence reads were assigned to a gene when both sequence reads were completely contained within the gene boundaries and were assigned by directional sequencing to the same strand. Reads per kilobase per million (RPKM) values (Mortazavi et al., 2008) were then computed starting from these counts.

RNA–DNA variations were detected using REDItools (Picardi and Pesole, 2013)

Tissue homogenate preparation

The PBI-Shredder, an auxiliary high-resolution respirometry (HRR) Tool, was used to prepare homogenate—in 0.2 M phosphate buffer (pH 8.0)—of frozen tissue specimens (Draxl et al., 2013), with high reproducibility of mitochondrial function as evaluated with HRR by means of Oxygraph-2 k OROBOROS®. Homogenate protein content was determined according to (Waddell, 1956) with bovine serum albumin used as a standard.

Enzymatic activity measurements

Citrate synthase (CS) and cytochrome c oxidase (COX) activities were measured by spectrophotometric standard methods. Each assay was performed at least in triplicate by using homogenate tissues subjected to three freeze–thaw cycles to disrupt membranes and expose mitochondrial enzymes.

The reduction of 5,5-dithiobis(2-nitrobenzoic acid) (DTNB) by CS at 412 nm (extinction coefficient is 13.6 mM^{-1} cm^{-1}) was followed in a coupled reaction with coenzyme A and oxaloacetate (Robinson and Srere, 1985). A reaction mixture of 0.1 mM acetyl-coenzyme A, 0.2% Triton-x-100, 0.1 mM DTNB and 20–40 μg of homogenate protein was incubated at 25 °C for 5 min. The reaction was initiated by the addition of 0.5 mM oxaloacetate. Results are expressed as nanomoles CoA formed per minute per mg homogenate protein.

COX activity was measured by following the decrease of absorbance at 548–540 nm (extinction coefficient is 19.1 mM^{-1} cm^{-1}), due to the oxidation of 40 μM ferrocytochrome c (reduced with substoichiometric concentrations of potassium ascorbate), for 90 s at 25 °C (see (Bobba et al., 2013)). Very stringent controls, including (i) test with cyanide, (ii) adequate homogenate proteins used for the reaction and (iii) gel-filtration of the reduced cyt c (in a superfine Sephadex G-25 column) to remove both excess of reductant and dimeric/multimeric form of cyt c, which per se can inhibit the reaction itself were performed to assure correct estimates of the Cox assay. Time-dependent absorbance changes were recorded with a Jasco double-beam/double-wavelength spectrophotometer UV-550. The rate of ferrocytochrome c oxidation, obtained as tangents to the initial part of the progress curves, is expressed as nanomoles cyt c_{ox} formed per minute per mg homogenate protein.

Data were expressed as means ± standard deviation (S.D.) ($n > 3$) and analyzed with SPSS software by 1-way analysis of variance (ANOVA) for repeated measures followed by the *post hoc* Bonferroni test for multiple comparisons. Statistical differences were determined at $P < 0.05$.

Results

mtDNA copy number per diploid nuclear genome determined by qPCR

We measured the relative mtDNA copy number in six tissue types (brain, lung, kidney, liver, heart and skeletal muscle) from three unrelated individuals (S7/11, S12/12 and S13/12) by qPCR. The highest mtDNA copy number was observed in heart tissue, followed by skeletal muscle, brain, liver, kidney and lung (Fig. 1 and Supplemental Table S4) consistent with the hypothesis that tissues with higher ATP requirement should show higher mitochondrial copy number. Differences in mtDNA content between the three subjects are particularly evident in brain and muscle. The low mtDNA copy number observed in the brain individual S13/12 could be related to the cause of death (asphyxia), while the differences observed in muscle might reflect heterogeneous relative content of nuclei and mitochondria between muscle fibers or to different levels of habitual physical activity among the three sampled subjects.

Correlation between mtDNA copy number and unspecific enrichment of mtDNA reads

To evaluate whether the relative abundance of mtDNA reads among off-target exome enrichment reads was a reliable and effective estimator of mtDNA copy number, we generated and analyzed whole exome sequencing (WES) data for each of the 18 samples (Supplemental Table S5). The relative number of mtDNA reads among off-target WES reads was calculated using previously published methods (Picardi and Pesole, 2012). Bivariate linear correlation between mtDNA copy number and the relative amount of off-target mtDNA reads (Fig. 2) was highly significant ($r^2 = 0.92$, $P < 0.0001$) confirming the effectiveness and reliability of the proposed approach for quantifying mtDNA copy number.

Nucleotide variants in assembled mtDNA sequences

For each tissue and individual, we assembled a complete mtDNA from whole exome reads (see Methods). Analysis of mtDNAs did not show significant somatic variations between different tissues of the same individual. Using stringent criteria (minimal coverage of 50 reads and a base variation frequency higher than 10%), we found heteroplasmy higher than 90% at position 310 of the D-Loop (T-to-C) in two of the three individuals. These observations were supported by the corresponding RNA-Seq data (Supplemental Table S6). The stringent filters employed were imposed to minimize false-discovery rate but might lead to the exclusion of poorly supported heteroplasmic sites. Relaxing the variation frequency, we also detected low frequency (5%) heteroplasmy at position 310 in the third individual as well as an additional low-frequency heteroplasmic position at D-Loop position 72 (T-to-C) in one individual (Supplemental Table S6). Both of these D-Loop variant positions are annotated as heteroplasmic sites in MITOMAP (Ruiz-Pesini et al., 2007).

Sample matched comparisons between DNA (from whole exome) and RNA (from RNA-Seq) mitochondrial reads were used to identify potential post-transcriptional modification events. Only changes occurring at DNA homoplasmic sites with frequencies higher than 10% at the RNA level were considered. We observed tissue-specific nucleotide variations in four positions, three in heart and one in brain, consistently shared by all three individuals (Supplemental Table S7). Positions 1955 and 2617 (heart) were with the 16S rRNA, whereas site 905 (heart) fell in the 12S rRNA. Notably, at position 2617, we identified both A-to-U

and A-to-G changes as previously reported (Bar-Yaacov et al., 2013). Position 8303 falls within the lysine tRNA and is found here for the first time as a target of a post-transcriptional modification event supported by RNA-Seq reads likely derived from polycistronic pre-processed transcripts which might include tRNAs (Nardelli et al., 1994). In this site both A-to-U and A-to-G changes were observed, but only in brain tissue. An additional RNA–DNA difference at position 295, described by Bar-Yaacov et al. (2013), was also observed in our samples, although not in all individuals and tissues because of the stringent filters employed (data not shown).

Determination of mitochondrial mass and tissue energy requirement

Citrate synthase (CS), a component of the tricarboxylic acid cycle, is a stably expressed mitochondrial matrix enzyme. Its specific activity is frequently used as an indicator of total mitochondrial mass (Figueiredo et al., 2008; Sarnat and Marin-Garcia, 2005, Kirby et al., 2007). Cytochrome oxidase (COX) activity is often employed as a marker of OXPHOS activity since this enzyme (complex IV) constitutes the last step in the respiratory chain (RC), likely limiting its electron flux (Capaldi, 1990, Kunz et al., 2000, Larsen et al., 2012, Mazat et al., 2001, Villani and Attardi, 1997, Villani and Attardi, 2000, Villani et al., 1998). To evaluate variation in mitochondria content and respiratory capacity, CS and COX activities were determined spectrophotometrically in total homogenates from all 18 samples.

Significant variation in CS and COX activities (normalized by the total homogenate protein content) were observed between tissues as well as between individuals (ANOVA). Tissue-dependent differences again reflected known differences in energetic demands between tissues with heart, the organ that consumes most energy per mass unit (Goffart et al., 2004, Van den Bogert et al., 1993), showing the highest CS and COX specific activities with values 4- to 12-fold and 1.2- to 2fold respectively higher than other tissues (Table 1).

Knowing the amount of nuclear DNA (nDNA) and mtDNA per mg of protein lysate of the same tissue specimens, it was possible to normalize the CS and COX enzymatic activities per nDNA and mtDNA. Table 1 shows that CS activity/nDNA ratio varies greatly between tissues, again reflecting the tissue-specific differences in mitochondrial activities. As expected, the heart shows the highest ratio in all three individuals, in accordance with mtDNA relative abundance values and evidence that it has the highest mitochondrial content per cell (Fernandez-Vizarra et al., 2011). Moreover, the CS activity/mtDNA also differs markedly between samples, with higher values for heart, skeletal muscle and brain (Table 1).

The highest COX activity/nDNA values were also associated with high energy-requirement tissues. However, COX activity/mtDNA values were, while variable, somewhat more consistent.

Correlation between mtDNA copy number and mitochondrial mass and respiratory capacity

We then evaluated if a correlation exists between the mtDNA copy number and the mitochondrial mass and the respiratory capacity in the different tissues. As shown in Fig. 3A and B, a positive and highly significant correlation was also found between either CS or COX activity per cell and the amount of mtDNA per cell. This indicates that the tissues which have the highest COX and CS activities have also the highest cellular mtDNA amount.

Mitochondrial gene expression in different tissues evaluated by RNA-seq

We carried out a strand-oriented RNA-Seq (2×100 bp paired-end, with random hexamer priming and no mitochondrial rRNA depletion) in the six tissue samples from the three individuals. Special attention was paid to minimize the impact of reads that could generate read mapping artifact to Numt (nuclear mitochondrial DNA) sequences (Calabrese et al., 2012) and introduce biases into expression level estimates (see Methods). Analysis of cDNA reads that mapped to Numts revealed that while rRNA like reads were most predominant, around 90% of ambiguously mapped read pairs from protein coding ORFs derived from CO1 or CO3 genes.

Globally, from 4% to 27% of read-pairs (according to tissue) were of mitochondrial origin (see % mtDNA PE reads in Supplemental Table S8 and Methods).

We observed a remarkably imbalanced level of expression between the mtDNA plus and minus strand, with about one thousand fold more reads mapping to the former. Furthermore, only about 0.36% of the reads mapping on the mtDNA plus strand resulted to be originated from the precursor RNA in contrast to over 85% of reads mapping on the minus strand (Supplemental Table S8 and Methods). Mapping of reads derived from the heavy strand covered the whole mitochondrion DNA sequence in all tissues (data not shown). The largest fraction of PE reads, around 95%, originated from 12S and 16S rRNA genes in all tissues considered. Other reads originated, as expected, mostly on annotated protein coding genes, with however sharp enrichment peaks also in the antisense strand of 16S rRNA, CO1 and ND5 genes (data not shown). Antisense transcripts of the ND5, ND6 and CytB genes have been demonstrated to correspond to three lncRNAs (Rackham et al., 2011), and we hypothesize that other antisense ncRNAs could originate in correspondence with the other peaks observed.

Expression levels of the 11 mtDNA protein coding mature transcripts, ribosomal RNAs, and lncRNAs are reported in Supplemental Table S9, expressed as RPKM values as well as relative expression within the same tissue. The comparison of absolute RPKM values of the same gene across different samples shows high variability, reflecting the different concentrations of mitochondrial RNAs in each sample. The relative expression of genes within the same sample, instead, remains remarkably constant across the different tissues (Fig. 4). However, we observe a considerable variability of this measure, corresponding to a highly variable level of steady-state expression of the different mature transcripts, suggesting a remarkable and variable effect of post-transcriptional cleavage, processing and stability mechanisms in the regulation of gene expression, that are conserved across the different tissues investigated.

Correlation of between mtDNA copy number and mitochondrial gene expression

Finally, we sought to investigate potential correlation between mtDNA copy number and the expression level of mtDNA-encoded genes. The overall concentration of RNAs of mitochondrial origin in each tissue, expressed either as a percentage of non ribosomal mitochondrial reads on the overall number of mapped reads, or as the sum of RPKM values associated with non ribosomal genes, showed a remarkable correlation with the estimated mtDNA copy number, yielding a Pearson correlation of 0.81, as shown in Fig. 5 (Supplemental Table S10). Similar correlation values (always around 0.8) were obtained by employing different measures for mitochondrial RNA concentration, namely, sum of genes

RPKM values including rRNA genes, or fraction of mitochondrial RNA in the RNA-Seq sample, and also when computed against estimated mtDNA abundance derived from qPCR or exome sequencing (data not shown).

Discussion

The advent of high-throughput technologies for DNA and RNA sequencing has opened new avenues in biological research. A better understanding of the coordinated expression of the mitochondrial and nuclear genome will be critical for the characterization of novel processes underlying the functioning of eukaryotic cells. In particular, mtDNA copy number is a key functional parameter that varies greatly between different cell or tissues types, both in response to overall metabolic and bioenergetics demands and as a consequence or cause of specific physiological and pathological conditions. Alterations in mtDNA copy number have been related to aging (He et al., 2014), cancer (Zhang et al., 2013), neurodegenerative diseases (Podlesniy et al., 2013), diabetes (Chien et al., 2012) and other mitochondrial-related diseases (Liu et al., 2013).

We previously reported a simple methodology for the reconstruction of complete or nearly complete mitochondrial genomes from off-target reads generated in whole exome sequencing experiments (WES) (Picardi and Pesole, 2012).

Here, using WES data from six different tissues (brain, liver, lung, striated muscle, kidney and heart) from three unrelated healthy Caucasian individuals (males, aged 47–54) in conjunction with qPCR for experimental validation, we show that mtDNA copy number per cell can be reliably estimated from the relative amount of off-target reads of mitochondrial origin detected in WES data. As previously noted (Picardi and Pesole, 2012), it is worth underlining that the fraction of off-target mitochondrial reads is dependent on the exome enrichment protocol used, with the Illumina TruSeq Exome Enrichment Kit providing a higher amount of mitochondrial off-target reads than either Roche Nimblegen and Agilent SureSelect (Picardi and Pesole, 2012). We observed, as expected, strict tissue specificity of mtDNA copy number, reflecting the specific bioenergetics and metabolic demands of different tissues with heart and lung tissues showing the highest and lowest values, respectively. Inter-individual variations for the same tissues were also observed, and are likely related to the individual-specific physiological status and to the cause of death.

Interestingly, we did not detect somatic variations in mtDNA sequences in different tissues from the same individual. Furthermore, the accuracy of sequence data was demonstrated by the overall consistency of gene and transcript sequences obtained by WES and RNA-Seq, respectively. A few putative post-transcriptional changes were detected, most of them in rRNA and tRNA genes, at sites already reported in the literature as susceptible to base-specific post-transcriptional modification.

We also show, by analyzing RNA-Seq data generated from the 18 samples, that a significant correlation exists, between the mtDNA copy number and the expression level of mt protein coding genes. Indeed, the paired-end strand-specific transcriptome sequencing allowed us to obtain reliable estimates of mature and precursor transcripts in sense and antisense orientations. We observed considerable variability in the expression levels of different protein coding transcripts, with CO1 and CO3 significantly more abundant than other genes (Fig. 4). On the other hand, the ND6 gene, the only L-strand gene encoding a polyA-transcript, is expressed at very low levels, totaling on average only 0.22% of all protein coding transcripts. Our data are not in accordance with the expression profile of mitochondrial mRNAs observed in Hela (Piechota et al., 2006) or 143B cells (Mercer et al.,

2011). However, it is quite expected that mitochondria from cultured human cells, particularly HeLa cells, do not reflect the transcriptional pattern observed in human tissues.

The strand-specific sequencing performed also allowed the detection of stable antisense transcripts of ND5 and ND6 genes (Supplemental Table S9) as already identified in rat (Tullo et al., 1994) and human cell 143B cell line (Mercer et al., 2011). Interestingly, these antisense transcripts have different levels in the different tissues, with the highest levels in heart and kidney.

Moreover, we found a remarkably significant positive correlation between mtDNA copy number and the mitochondrial mass, i.e., the CS/nDNA ratio, as well as the respiratory capacity, i.e., the COX/nDNA ratio, when all the tissues were considered together (Fig. 3), indicating that tissues with the highest mtDNA content i.e., heart, muscle and brain, have also the highest CS and COX activities per cell. Indeed, it is known that different metabolic profiles and variable energetic demands of different tissues are due to the inherently different functions or, in a single tissue, to changes in ATP demand due to physiological or pathological conditions (Kunz, 2003, Leary et al., 1998, Leverve and Fontaine, 2001, Pfeiffer et al., 2001). While our data suggest that mtDNA copy number is a critical parameter in the definition of quantitative aspects of respiratory function, mitochondria also exhibit tissue related variation in functional capacity, protein composition and morphology (Benard et al., 2006, Fernandez-Vizarra et al., 2011, Johnson et al., 2007a, Johnson et al., 2007b, Mootha et al., 2003, Pagliarini et al., 2008).

An aspect that must not be neglected—in the light of different metabolic profile of the three individuals—is that cellular adaptation to environmental and physiological constraints necessitates a fine tuning of the control of mitochondrial respiration, in response to changes in energy demand and substrate delivery. Accordingly, different tissues present large differences in the composition of the OXPHOS machinery and the organization of mitochondria (Chan, 2006, Fernandez-Vizarra et al., 2011), which, in addition to mtDNA copy number (Di Mauro and Bonilla, 2004), could influence their physiological activity. Nonetheless, mtDNA copy number remains a good proxy for mitochondrial activity for a sound conceptual reason: the mitochondrial mass or the activity of a respiratory enzyme are being related to the unit of genome that is needed to produce them (Fernandez-Vizarra et al., 2008, Fernandez-Vizarra et al., 2011) as well as to its relevant nuclear counterparts (Mercer et al., 2011).

In conclusion, the methodology presented here demonstrates the feasibility of large-scale detection of mtDNA copy number in diverse cell-types, tissues and pathological conditions. Given that several thousand exome sequence data sets are available in public repositories from different sources, the approach presented here is expected to generate a wealth of information that may contribute to a better understanding of nucleo-mitochondrion cross-talk and its involvement in health and disease.

Funding

Data Access

All short read data are available to the users upon request and are being submitted at the dbGAP archive.

References

R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell **Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA** Nat. Genet., 23 (1999), p. 147

D. Bar-Yaacov, G. Avital, L. Levin, A.L. Richards, N. Hachen, B. Rebolledo Jaramillo, A. Nekrutenko, R. Zarivach, D. Mishmar **RNA-DNA differences in human mitochondria restore ancestral form of 16S ribosomal RNA** Genome Res., 23 (2013), pp. 1789-1796

G. Benard, B. Faustin, E. Passerieux, A. Galinier, C. Rocher, N. Bellance, J.P. Delage, L. Casteilla, T. Letellier, R. Rossignol **Physiological diversity of mitochondrial oxidative phosphorylation** Am. J. Physiol. Cell Physiol., 291 (2006), pp. C1172-C1182

A. Bobba, G. Amadoro, D. Valenti, V. Corsetti, R. Lassandro, A. Atlante **Mitochondrial respiratory chain Complexes I and IV are impaired by beta-amyloid via direct interaction and through Complex I-dependent ROS production, respectively** Mitochondrion, 13 (2013), pp. 298-311

D. Bogenhagen, D.A. Clayton **The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. Quantitative isolation of mitochondrial deoxyribonucleic acid** J. Biol. Chem., 249 (1974), pp. 7991-7995

F.M. Calabrese, D. Simone, M. Attimonelli **Primates and mouse NumtS in the UCSC genome browser** BMC Bioinforma., 13 (Suppl. 4) (2012), p. S15

R.A. Capaldi **Structure and function of cytochrome *c* oxidase** Annu. Rev. Biochem., 59 (1990), pp. 569-596

D.C. Chan **Mitochondrial fusion and fission in mammals** Annu. Rev. Cell Dev. Biol., 22 (2006), pp. 79-99

M.C. Chien, W.T. Huang, P.W. Wang, C.W. Liou, T.K. Lin, C.J. Hsieh, S.W. Weng **Role of mitochondrial DNA variants and copy number in diabetic atherogenesis** Genet. Mol. Res., 11 (2012), pp. 3339-3348

L.L. Clay Montier, J.J. Deng, Y. Bai **Number matters: control of mammalian mitochondrial DNA copy number** J. Genet. Genomics, 36 (2009), pp. 125-131

S. Di Mauro, E. BonillaA. Engel, C.F.-A. (Eds.), Myology, vol. II, Mc Grav Hill, Philadelphia (2004), pp. 1623-1676

A. Draxl, A. Eigentler, E. Gnaiger **PBI-Shredder HRR-Set: preparation of tissue homogenates for diagnosis of mitochondrial respiratory function** Mitochondrial Physiol. Netw., 17 (02) (2013), pp. 1-8

E. Fernandez-Vizarra, J.A. Enriquez, A. Perez-Martos, J. Montoya, P. Fernandez-Silva **Mitochondrial gene expression is regulated at multiple levels and differentially in the heart and liver by thyroid hormones** Curr. Genet., 54 (2008), pp. 13-22

E. Fernandez-Vizarra, J.A. Enriquez, A. Perez-Martos, J. Montoya, P. Fernandez-Silva **Tissue-specific differences in mitochondrial activity and biogenesis** Mitochondrion, 11 (2011), pp. 207-213

P.A. Figueiredo, R.M. Ferreira, H.J. Appell, J.A. Duarte **Age-induced morphological, biochemical, and functional alterations in isolated mitochondria from murine skeletal muscle** J. Gerontol. A Biol. Sci. Med. Sci., 63 (4) (2008), pp. 350-359

S. Goffart, J.C. von Kleist-Retzow, R.J. Wiesner**Regulation of mitochondrial proliferation in the heart: power-plant failure contributes to cardiac failure in hypertrophy** Cardiovasc. Res., 64 (2004), pp. 198-207

Y.H. He, X. Lu, H. Wu, W.W. Cai, L.Q. Yang, L.Y. Xu, H.P. Sun, Q.P. Kong **Mitochondrial DNA content contributes to healthy aging in Chinese: a study from nonagenarians and centenarians** Neurobiol. Aging, 35 (7) (2014), pp. 1779.e1-1779.e4

D.T. Johnson, R.A. Harris, P.V. Blair, R.S. Balaban **Functional consequences of mitochondrial proteome heterogeneity** Am. J. Physiol. Cell Physiol., 292 (2007), pp. C698-C707

D.T. Johnson, R.A. Harris, S. French, P.V. Blair, J. You, K.G. Bemis, M. Wang, R.S. Balaban **Tissue heterogeneity of the mammalian mitochondrial proteome** Am. J. Physiol. Cell Physiol., 292 (2007), pp. C689-C697

D.M. Kirby, D.R. Thorburn, D.M. Turnbull, R.W. Taylor **Biochemical assays of respiratory chain complex activity** Methods Cell Biol., 80 (2007), pp. 93-119

W.S. Kunz **Different metabolic properties of mitochondrial oxidative phosphorylation in different cell types—mportant implications for mitochondrial cytopathies** Exp. Physiol., 88 (2003), pp. 149-154

W.S. Kunz, A. Kudin, S. Vielhaber, C.E. Elger, G. Attardi, G. Villani **Flux control of cytochrome *c* oxidase in human skeletal muscle** J. Biol. Chem., 275 (2000), pp. 27741-27745

N. Lane, W. Martin **The energetics of genome complexity** Nature, 467 (2010), pp. 929-934

S. Larsen, J. Nielsen, C.N. Hansen, L.B. Nielsen, F. Wibrand, N. Stride, H.D. Schroder, R. Boushel, J.W. Helge, F. Dela, M. Hey-Mogensen **Biomarkers of mitochondrial content in skeletal muscle of healthy young human subjects** J. Physiol., 590 (2012), pp. 3349-3360

S.C. Leary, B.J. Battersby, C.D. Moyes **Inter-tissue differences in mitochondrial enzyme activity, RNA and DNA in rainbow trout (*Oncorhynchus mykiss*)** J. Exp. Biol., 201 (Pt 24) (1998), pp. 3377-3384

X.M. Leverve, E. Fontaine **Role of substrates in the regulation of mitochondrial function in situ** IUBMB Life, 52 (2001), pp. 221-229

H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin **The Sequence Alignment/Map format and SAMtools** Bioinformatics, 25 (2009), pp. 2078-2079

H. Liu, Y. Ma, F. Fang, Y. Zhang, L. Zou, Y. Yang, S. Zhu, S. Wang, X. Zheng, P. Pei, *et al*. **Wild-type mitochondrial DNA copy number in urinary cells as a useful marker for diagnosing severity of the mitochondrial diseases** PLoS One, 8 (2013), p. e67146

J.P. Mazat, R. Rossignol, M. Malgat, C. Rocher, B. Faustin, T. Letellier **What do mitochondrial diseases teach us about normal mitochondrial functions…that we already knew: threshold expression of mitochondrial defects** Biochim. Biophys. Acta, 1504 (2001), pp. 20-30

T.R. Mercer, S. Neph, M.E. Dinger, J. Crawford, M.A. Smith, A.M. Shearwood, E. Haugen, C.P. Bracken, O. Rackham, J.A. Stamatoyannopoulos, A. Filipovska, J.S. Mattick **The human mitochondrial transcriptome** Cell, 146 (2011), pp. 645-658

V.K. Mootha, J. Bunkenborg, J.V. Olsen, M. Hjerrild, J.R. Wisniewski, E. Stahl, M.S. Bolouri, H.N. Ray, S. Sihag, M. Kamal, N. Patterson, E.S. Lander, M. Mann **Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria** Cell, 115 (2003), pp. 629-640

C.T. Moraes **What regulates mitochondrial DNA copy number in animal cells?** Trends Genet., 17 (2001), pp. 199-205

A. Mortazavi, B.A. Williams, K. McCue, L. Schaeffer, B. Wold **Mapping and quantifying mammalian transcriptomes by RNA-Seq** Nat. Methods, 5 (2008), pp. 621-628

M. Nardelli, S. Tommasi, A.M. D'Erchia, F. Tanzariello, A. Tullo, A.T. Primavera, M. De Lena, E. Sbisa, C. Saccone **Detection of novel transcripts in the human mitochondrial DNA region coding for ATPase8-ATPase6 subunits** FEBS Lett., 344 (1994), pp. 10-14

D.J. Pagliarini, S.E. Calvo, B. Chang, S.A. Sheth, S.B. Vafai, S.E. Ong, G.A. Walford, C. Sugiana, A. Boneh, W.K. Chen, *et al*. **A mitochondrial protein compendium elucidates complex I disease biology** Cell, 134 (2008), pp. 112-123

G. Pesole, J.F. Allen, N. Lane, W. Martin, D.M. Rand, G. Schatz, C. Saccone **The neglected genome** EMBO Rep., 13 (2012), pp. 473-474

T. Pfeiffer, S. Schuster, S. Bonhoeffer **Cooperation and competition in the evolution of ATP-producing pathways** Science, 292 (2001), pp. 504-507

E. Picardi, G. Pesole **Mitochondrial genomes gleaned from human whole-exome sequencing** Nat. Methods, 9 (2012), pp. 523-524

E. Picardi, G. Pesole **REDItools: high-throughput RNA editing detection made easy** Bioinformatics, 29 (2013), pp. 1813-1814

J. Piechota, R. Tomecki, K. Gewartowski, R. Szczesny, A. Dmochowska, M. Kudla, L. Dybczynska, P.P. Stepien, E. Bartnik **Differential stability of mitochondrial mRNA in HeLa cells** Acta Biochim. Pol., 53 (2006), pp. 157-168

P. Podlesniy, J. Figueiro-Silva, A. Llado, A. Antonell, R. Sanchez-Valle, D. Alcolea, A. Lleo, J.L. Molinuevo, N. Serra, R. Trullas **Low cerebrospinal fluid concentration of mitochondrial DNA in preclinical Alzheimer disease** Ann. Neurol., 74 (2013), pp. 655-668

O. Rackham, A.M. Shearwood, T.R. Mercer, S.M. Davies, J.S. Mattick, A. Filipovska **Long noncoding RNAs are generated from the mitochondrial genome and regulated by nuclear-encoded proteins** RNA, 17 (2011), pp. 2085-2093

S.M. Rafelski **Mitochondrial network morphology: building an integrative, geometrical view** BMC Biol., 11 (2013), p. 71

J.B. Robinson Jr., P.A. Srere **Organization of Krebs tricarboxylic acid cycle enzymes in mitochondria** J. Biol. Chem., 260 (1985), pp. 10800-10805

E. Ruiz-Pesini, M.T. Lott, V. Procaccio, J.C. Poole, M.C. Brandon, D. Mishmar, C. Yi, J. Kreuziger, P. Baldi, D.C. Wallace **An enhanced MITOMAP with a global mtDNA mutational phylogeny** Nucleic Acids Res., 35 (2007), pp. D823-D828

H.B. Sarnat, J. Marin-Garcia **Pathology of mitochondrial encephalomyopathies** The Canadian journal of neurological sciences Le journal canadien des sciences neurologiques, 32 (2) (2005), pp. 152-166

R.J. Shmookler Reis, S. Goldstein **Mitochondrial DNA in mortal and immortal human cells. Genome number, integrity, and methylation** J. Biol. Chem., 258 (1983), pp. 9078-9085

A. Tullo, F. Tanzariello, A.M. D'Erchia, M. Nardelli, P.A. Papeo, E. Sbisa, C. Saccone **Transcription of rat mitochondrial NADH-dehydrogenase subunits. Presence of antisense and precursor RNA species** FEBS Lett., 354 (1994), pp. 30-36

C. Van den Bogert, H. De Vries, M. Holtrop, P. Muus, H.L. Dekker, M.J. Van Galen, P.A. Bolhuis, J.W. Taanman **Regulation of the expression of mitochondrial proteins: relationship between mtDNA copy number and cytochrome-c oxidase activity in human cells and tissues** Biochim. Biophys. Acta, 1144 (1993), pp. 177-183

V. Venegas, J. Wang, D. Dimmock, L.J. Wong **Real-time quantitative PCR analysis of mitochondrial DNA content** Jonathan L. Haines, *et al.* (Eds.), Current protocols in human genetics/editorial board (2011), p. 17

G. Villani, G. Attardi **In vivo control of respiration by cytochrome *c* oxidase in wild-type and mitochondrial DNA mutation-carrying human cells** Proc. Natl. Acad. Sci. U. S. A., 94 (1997), pp. 1166-1171

G. Villani, G. Attardi **In vivo control of respiration by cytochrome *c* oxidase in human cells** Free Radic. Biol. Med., 29 (2000), pp. 202-210

G. Villani, M. Greco, S. Papa, G. Attardi **Low reserve of cytochrome _c_ oxidase capacity in vivo in the respiratory chain of a variety of human cell types** J. Biol. Chem., 273 (1998), pp. 31829-31836

W.J. Waddell **A simple ultraviolet spectrophotometric method for the determination of protein** J. Lab. Clin. Med., 48 (1956), pp. 311-314

T.D. Wu, S. Nacu **Fast and SNP-tolerant detection of complex variants and splicing in short reads** Bioinformatics, 26 (2010), pp. 873-881

G. Zhang, Y. Qu, S. Dang, Q. Yang, B. Shi, P. Hou **Variable copy number of mitochondrial DNA (mtDNA) predicts worse prognosis in advanced gastric cancer patients** Diagn. Pathol., 8 (2013), p. 173

Fig 1

Relative mtDNA copy number in six tissues of three individuals (S7/11, S12/12 and S13/12) calculated by qPCR. Values represent the average of three independent experiments.
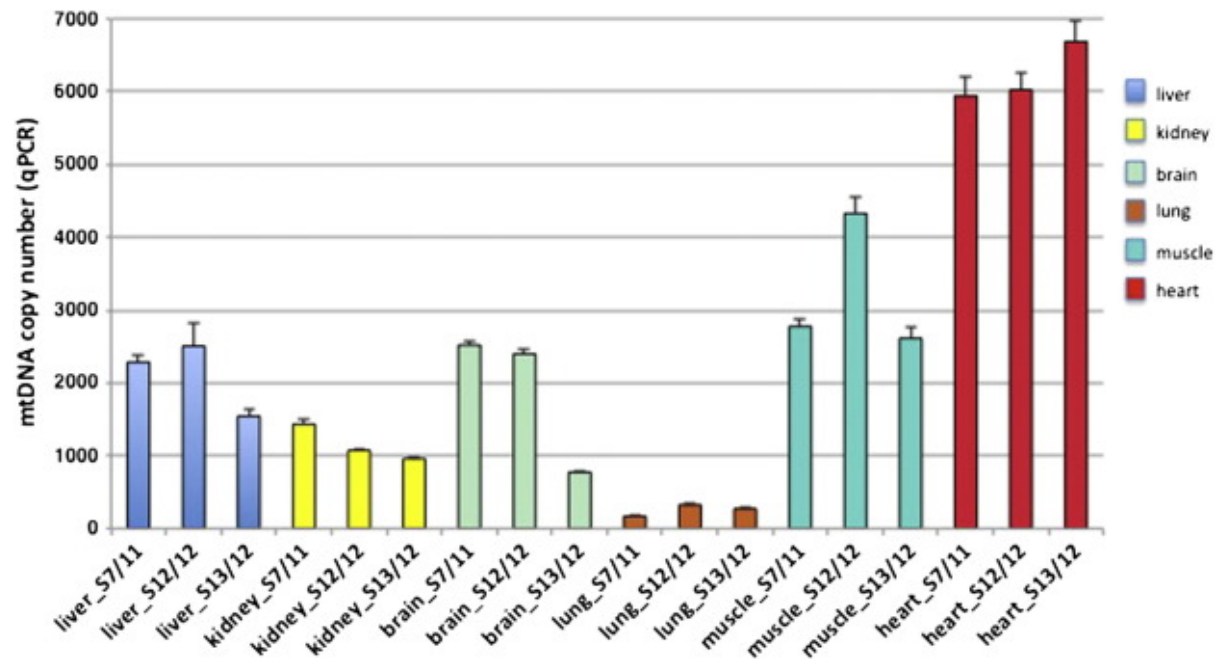
Fig 2

Linear correlation between the mtDNA copy number in the six tissues of three individuals (S7/11, S12/12 and S13/12) and off-target mtDNA reads of whole exome sequencing data, calculated performing a bivariate linear fit analysis ($P < 0.0001$). The amount of off-target mtDNA reads has been calculated as the number of mtDNA reads per million WES reads (see Supplemental Table S5).
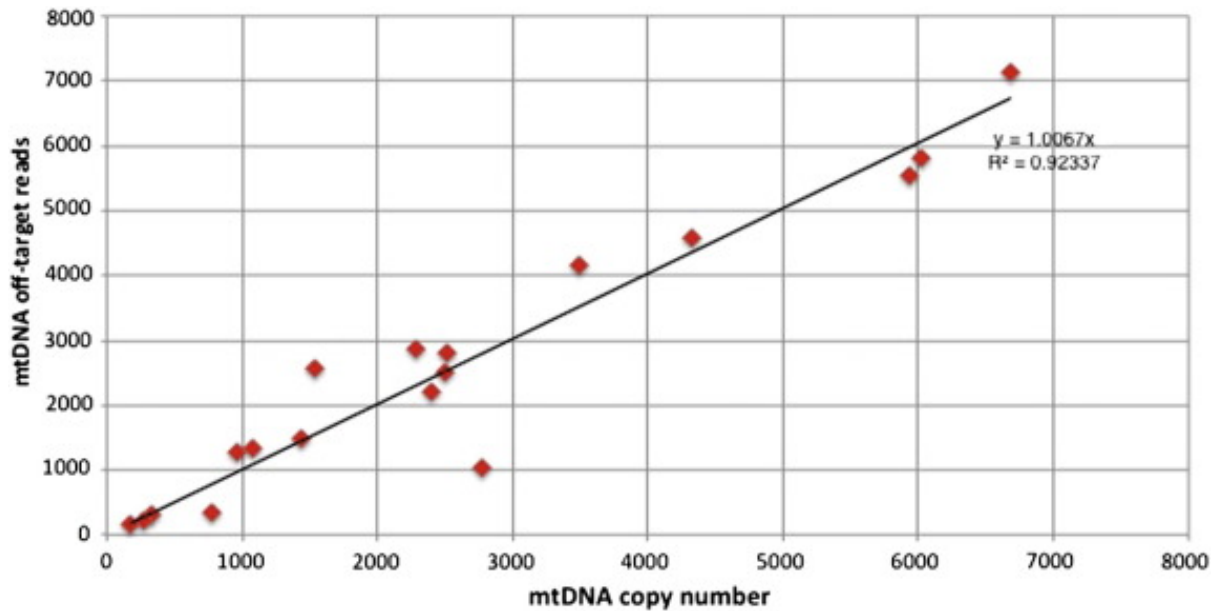
Fig 3

Linear correlation between mtDNA copy number and mitochondrial mass and respiratory capacity. (A) Linear correlation between COX/nDNA and the mtDNA copy number in six tissues of three individuals (S7/11, S12/12 and S13/12) ($P < 0.0001$). (B) Linear correlation between CS/nDNA and mtDNA copy number in six tissues of three individuals (S7/11, S12/12 and S13/12) ($P < 0.0001$).

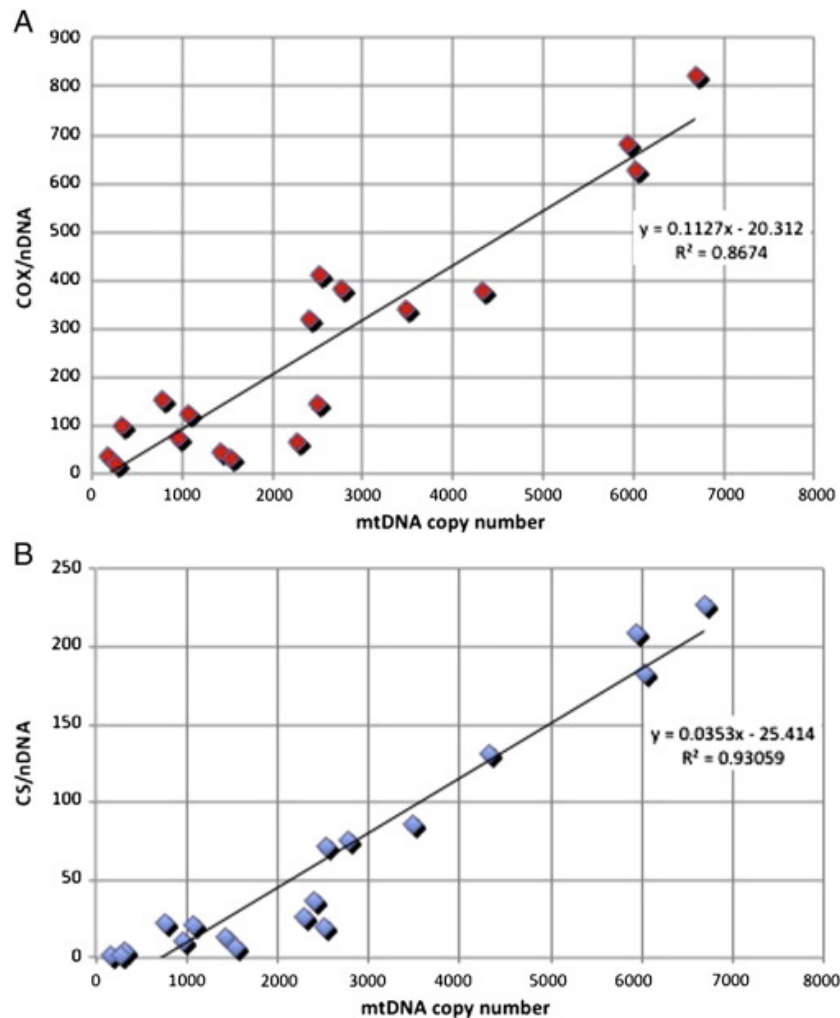Correlation was calculated using a bivariate linear fit analysis.

Fig 4

Relative levels of the mitochondrial mature mRNAs in the six examined tissues expressed as average for the three individuals (S7/11, S12/12 and S13/12). Values represent the ratio between the RPKM (reads per kilobase per million reads) of a specific mRNA and the total RPKM for all mature mRNAs (see Supplementary Table S9) calculated for each given tissue and individual.
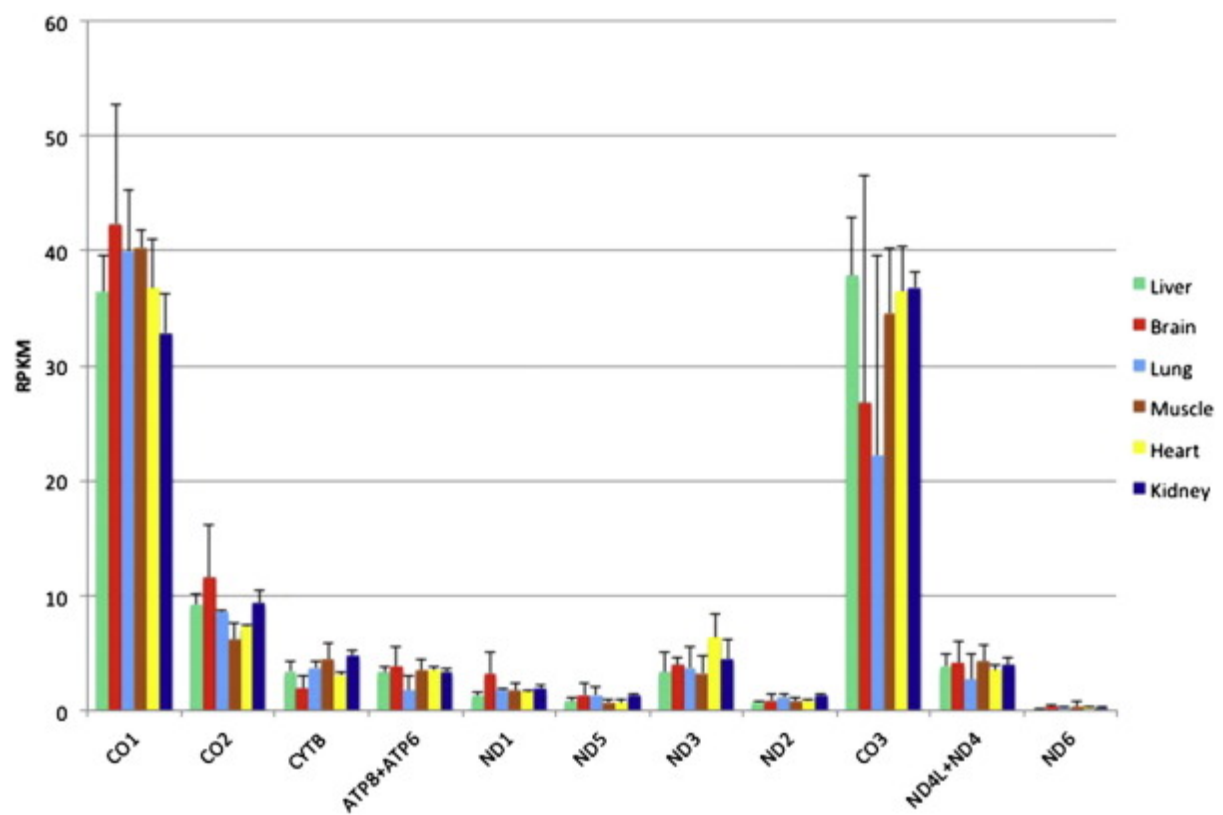
Fig 5

Linear correlation between estimated mtDNA copy number and relative levels of mt mature transcripts expressed as average of RPKM values in the three individuals (S7/11, S12/12 and S13/12). Correlation was calculated using a bivariate linear fit analysis ($P < 0.001$).
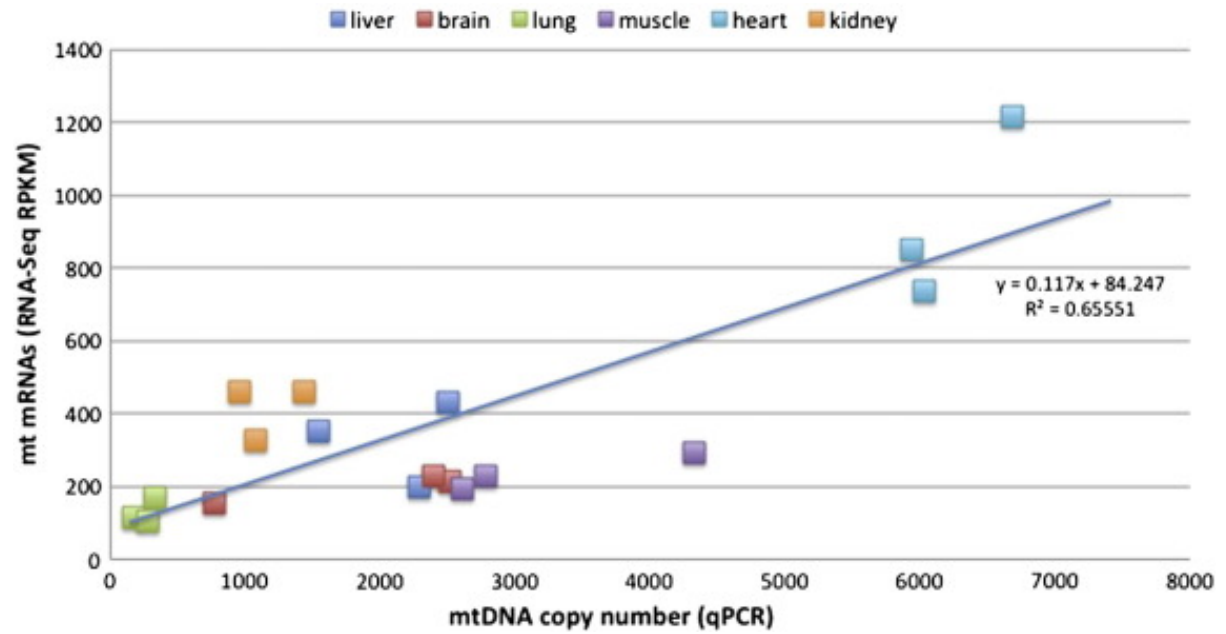
Table 1

Citrate synthase (CS) and cytochrome oxidase (COX)-specific activity measurements and normalization by mtDNA and nDNA content.

Data are the mean + standard deviation, with $n = 4$, obtained from six human tissues, liver, kidney, brain, lung, muscle and heart, from three age- and sex-matching individuals.

Tissue	CS-specific activity	COX-specific activity	CS/mtDNA	CS/nDNA	COX/mtDNA	COX/nDNA
liver_S7/11	530 ± 31	1357 ± 32	5196	27	13303	68
liver_S12/12	245 ± 14	1813 ± 47	4803	20	35549	144
liver_S13/12	297 ± 15	1411 ± 55	1650	7	7839	31
kidney_S7/11	367 ± 18	1260 ± 53	3 336	13	11454	144
kidney_S12/12	246 ± 11	1442 ± 44	6648	22	38972	126
kidney_S13/12	426 ± 21	2832 ± 62	4260	11	28320	75
brain_S7/11	164 ± 10	949 ± 37	10933	71	63266	413
brain_S12/12	166 ± 11	1400 ± 43	6384	37	53846	318
brain_S13/12	166 ± 9	1153 ± 30	12769	22	88693	154
lung_S7/11	76 ± 5	1436 ± 41	4750	2	89750	36
lung_S12/12	62 ± 5	1522 ± 42	5166	4	126833	98
lung_S13/12	132 ± 7	1445 ± 47	2870	2	31413	21
muscle_S7/11	196 ± 10	995 ± 28	8521	75	43260	383
muscle_S12/12	475 ± 22	1363 ± 31	10106	131	29000	378
muscle_S13/12	335 ± 12	1335 ± 33	14565	86	58043	342
heart_S7/11	564 ± 28	1836 ± 41	16114	209	52457	680
heart_S12/12	768 ± 37	2639 ± 55	10378	183	35662	628
heart_S13/12	524 ± 25	1890 ± 47	12780	227	46097	821