Deep sequencing reveals the mitochondrial DNA variation landscapes of breast-to-brain metastasis blood samples

Rhiannon E. McGeehan[1*$], Lewis A. Cockram[1,2$], D. Timothy J. Littlewood[3], Kathleen Keatley[1,4], Diana M. Eccles[5] and Qian An[2*]

[1] Brain Tumour Research Centre, Institute of Biomedical and Biomolecular Sciences, University of Portsmouth, Portsmouth, PO1 2DT, UK
[2] School of Pharmacy and Biomedical Sciences, Institute of Biomedical and Biomolecular Sciences, University of Portsmouth, Portsmouth, PO1 2DT, UK
[3] Department of Life Sciences, Natural History Museum, London, SW7 5BD, UK
[4] School of Biological Sciences, Institute of Biomedical and Biomolecular Sciences, University of Portsmouth, Portsmouth, PO1 2DT, UK
[5] Cancer Sciences Academic Unit, Faculty of Medicine, University of Southampton, Southampton, SO16 6YA, UK
[$] Joint first authors
[*] Correspondence: rhiannon.mcgeehan@port.ac.uk or qian.an@port.ac.uk; Tel.: +44 2392842556

Abstract: Breast-to-brain metastasis (BBM) often represents a terminal event, due to the inability of many systemic treatments to cross the blood-brain barrier (BBB), rendering the brain a sanctuary site for tumour cells. Identifying genetic variations that can predict the patients who will develop BBM would allow targeting of adjuvant treatments to reduce risk while disease bulk is minimal. Germ-line genetic variations may contribute to whether a BBM forms by influencing the primary tumour sub-type that presents, or by influencing the host-response to the tumour or treatment regimen, or by facilitating transition of tumour cells across the BBB and establish a viable brain metastasis. The role of mitochondrial DNA (mtDNA) variants specifically in BBM is underexplored. Consequently, using a sensitive deep-sequencing approach, we characterised the mtDNA variation landscapes of blood samples derived from 13 females who were diagnosed with early onset breast cancer and later went on to develop BBM. We also predicted the potential pathogenic significance of variations identified in all mtDNA-encoded oxidative phosphorylation (OXPHOS) proteins using 3D protein structural mapping and analysis, to identify variations worthy of follow-up. From the 70 variations found in protein coding regions, we reveal novel links between 3 specific mtDNA variations and altered OXPHOS structure and function in 23% of the BBM samples. Further studies are required to confirm the origin of mtDNA variations, whether they correlate with 1) the predicted alterations in mitochondrial function and 2) increased risk of developing breast brain metastasis using a much larger cohort of samples.

Keywords: mtDNA; breast cancer; breast to brain metastasis; 3D protein structural mapping and analysis; OXPHOS and long PCR

1. Introduction

Breast neoplasms are currently the second highest cause of cancer related death in women. Although brain metastases are less common than lesions in the bone or lungs, they often represent a terminal event with a projected survival of just 3 to 6 months from diagnosis, with less than 20% of patients surviving more than 1 year [1–3]. Poor clinical outcome is due, in part, to the difficulty in treating the metastatic site, with systemic treatments generally being unable to penetrate the blood brain barrier. Predicting clinical outcome is also difficult due to the complicated interactions of metastatic cells with host homeostatic mechanisms, which metastatic cells ultimately exploit for their own survival and proliferation [4].

The process of brain metastasis, including breast-to-brain metastasis (BBM), consists of a series of steps. In order that a clinically relevant brain lesion is observed, tumour cells must: 1) reach >1mm^3 by rapidly proliferating and establishing new vasculature, 2) invade the host cells and gain access to either the circulatory lymphatic or haematogenous systems, 3) survive the lymphatic or haematogenous systems and arrest in the capillary beds and extravasate into the brain, and 4) proliferate and form metastasis in the brain [4]. Despite this broad understanding of the metastatic process, the incidence of BBM is on the increase and there is consequently a need for better understanding of the BBM process as well as markers that could predict which patients will develop BBM, which in turn could improve current treatments for patients with BBM.

Patient-derived blood samples are often used to identify associations between genetic variations in the germ-line and cancer prognosis [5,6]. Identifying such variations could help predict which patients may develop BBM, allowing the targeting of adjuvant treatments aimed at reducing risk while the disease bulk is minimal.

MtDNA, which exists as multiple copies in the mitochondrial matrix, is 16,568 base pairs long and encodes 37 genes. These genes provide 13 key catalytic proteins of the oxidative phosphorylation (OXPHOS) system, as well as the 2 rRNAs (12S and 16S) and 22 tRNAs vital for their synthesis. Of the 13 mtDNA-encoded proteins, 7 contribute to complex I (NADH dehydrogenase), 1 contributes to complex III (cytochrome bc1 complex), 3 contribute to complex IV (cytochrome c oxidase) and 2 contribute to complex V (ATP synthase).

The OXPHOS system is composed of the mitochondrial respiratory chain (complexes I to IV) and ATP synthase (complex V). This system is vital for the production of cellular energy in the form of ATP and also has numerous secondary roles in other processes including calcium homeostasis [7], reactive oxygen species (ROS) production and signaling [8] and apoptosis [9,10]. Due to the importance of OXPHOS, variations in the mtDNA sequence, which can be germ-line or somatic, have the capacity to inflict profound effects on cellular function, contributing to or causing numerous human diseases (reviewed in [11]). As several hundred mtDNA molecules co-exist within a cell or tissue however, each with the capacity to hold different variations, a certain threshold must be met for a variation to exert an observable phenotypic effect [12]. The occurrence of multiple mtDNAs within a population is termed heteroplasmy.

Alterations in OXPHOS, as a consequence of mtDNA variations, have a proposed role in metastasis [13]. For example, transmitochondrial hybrids (cybrids), formed through the cell-cell fusion of enucleate highly metastatic lung cancer cells and mtDNA deplete lung cancer cells with low metastatic potential, enhanced the recipient cell's ability to form metastatic nodules in the lungs of mice following tail vein inoculation [14]. Consistent with this, the reciprocal mtDNA exchange suppressed the ability of the host cells to form metastatic nodules [14]. Despite the various studies that have looked into the incidence and predisposing roleof mtDNA variations in primary breast cancer [15–18] and the role of mtDNA variations in the metastatic process using *in vitro* cellular and mouse xenograft models [13-14, 46], no similar studies have been conducted in the context of BBM.

As a first step towards determining whether mtDNA variations in the germ-line could influence the likelihood of developing breast to brain metastasis, in this study we determined the blood mtDNA variation landscapes of 13 breast to brain metastasis patients using an established, sensitive deep sequencing approach [19]. As expected, numerous mtDNA variations were detected relative to the revised Cambridge reference sequence (rCRS [23] [Genbank: NC_012920]). Consequently, in order to draw up a short-list of candidates worthy of follow up investiagtion, all of the non synonymous variations identified in mtDNA-encoded OXPHOS proteins were then inspected in detail *in silico* using the latest, best quality and homologous 3D protein structure data available to analyse and predict their potential functional impact on the OXPHOS system [20] and therefore potential links with the BBM process. Until now, this has only been possible for mtDNA-variations in complex III and IV genes [19, 20]. However, recent publication of the cryo-EM 3D structures for complexes I and V [28,29, respectively] means, for the first time, we can also gain insight into the impact of mtDNA-variations at the level of protein structure in these complexes.

2. Materials and Methods

2.1 Breast to brain metastasis patients and DNA samples

Thirteen whole blood DNA samples were obtained from a subset of the Prospective study of Outcomes in Sporadic vs Hereditary breast cancer (POSH) [21,22]. Specifically, these patients had developed breast-to-brain metastasis. Patients were aged 40 or younger, and had been diagnosed with invasive breast cancer between January 2000 and January 2008. They also had varying NPI (Nottingham prognostic index) scores, hormone receptor expression and treatment exposure (Table S1). The NPI is used to calculate the prognosis of breast cancer patients following surgery, taking into consideration the size of the lesion, the number of lymph nodes affected and the tumour grade (I-III). As a collective, the patients in this study had an average NPI score of just above 5.4, meaning 70% of patients had a projected survival of 5 years.

2.2 Long PCR amplification and sequencing of complete mtDNAs

Based on the method of Lloyd et al. 2015 [19], total DNA was extracted from the total blood of each patient using DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions. 100ng of total whole blood DNA of each patient, was amplified using two overlapping mtDNA specific primer pairs: CytbF and HumanLongR (Amplicon 1) or HumanLongF and CytbR (Amplicon 2) to yield two halves of the mitochondrial genome in each case, which were then resolved by agarose gel electrophoresis, purified and quantified (Table S2). Amplicon 1 and 2 from each sample were pooled (0.65 nM each) and 1ng of DNA from each pool was used to construct 13 Truseq Nano libraries (Illumina) which were combined, sequenced with paired-end 2 X 150 reads on the Illumina MiSeq system (Department of Biochemistry, University of Cambridge).

2.3 Identification of variations, determination of heteroplasmy and annotation

Post sequencing, reads obtained from the Illumina Resequencing Workflow (http://supportres.illumina.com/documents/documentation/software_documentation/miseqreporter/miseq-reporter-user-guide-15042295-b.pdf), were aligned against the rCRS [23] [Genbank: NC_012920] using the Burrows Wheeler Aligner (Galaxy Tools Version (GTv) 1.2.3). The rCRS, which represents haplogroup H2a2, is currently the standard sequence used for identifying the presence of mtDNA variations. Refinement of the initial alignment was then conducted through the marking and removal of duplicate reads, using Picard beta (GTv 1.56.0), followed by setting up regions for indel realignment, realignment around indels, identification of co-variates and recalibration of quality scores using the GATK tools beta (GTv 0.0.4, 0.0.6, 0.0.5 and 0.0.6, respectively).

Variant calling was carried out on the processed alignment with the GATK unified genotyper (GTv 0.0.6) to identify regions of the sequences that diverged from the rCRS. Complete processing of reads through this workflow removed the presence of several suspected false variations (*G3483C, T3488A, A3492C* and *3492insC*), that frequently reside in the unprocessed variant files generated directly by the Illumina Resequencing Workflow.

To further minimise miscalling of variations, those which had previously been identified as "false" variants due to misalignments around homopolymer tracts or sequencing errors by Ju et al. (*A302C, C309T* and *C3106A*; [27]) were discarded. Finally, variants at sites in the rCRS which are known to contain rare alleles (*A263, A750, A1438, A8860* and *A15326*) were not considered in further analysis [23].

To confirm the presence and calculate the heteroplasmy of each of the remaining variations a minimum read quality (Phred) score of 30 (Q30) was applied, this means that there is a very low likelihood (just a 0.01% chance) that the variations detected were due to sequencing error. To be confident that the variants called were not due to sequencing error a stringent heteroplasmy cut off of ≥1% was applied, i.e. the number of mtDNA molecules harboring the variant had to be ≥1% or ≤99% to be considered heteroplasmic. If <1% then the variant would be considered homoplasmic wild-type. If >99% then the variant would be considered homoplasmic (pure) mutant. To determine heteroplasmy, 250 of the Q30-reads (with the duplicates removed) in the alignment were chosen at random and then heteroplasmy was calculated by dividing the depth of reads with the variant

nucleotide by 250, and then expressed as a percentage. The presence of variations was further confirmed by visually inspecting the full alignments using Tablet[47]. Annotation of the identified variations was conducted manually by inputting variations into Mitowheel (http://mitowheel.org/mitowheel.html) to ascertain their locus and precise position in the mitochondrial genome. All new mtDNAs generated have been submitted to GenBank with the following accession numbers (MF317865- MF317877).

2.4 Prevalence of variations in normal/healthy (control) bloods and pathologic tissues

Previous disease associations of the variations were established through MITOMAP (http://www.mitomap.org/MITOMAP) and the human mitochondrial database (HmtDB; http://www.hmtdb.uniba.it/hmdb/). The HmtDB was also used to determine the prevalence of variations within bloods obtained from normal, healthy individuals (as controls). The HmtDB contains over 28,196 and 3,539 mitogenomic sequences obtained from the tissues of normal, healthy subjects and patients, respectively. Variation identification by HmtDB is based on the Reconstructed Sapiens Reference Sequence (RSRS) and Phylotree mtDNA tree Build 17), this means certain variations identified in the BBM mtDNAs relative to the rCRS, do not generate hits using the HmtDB (a full list can be found at Phylotree: http://www.phylotree.org/resources/RSRS_vs_rCRS.htm).

2.5 Identifying functional candidates using three-dimensional protein structure mapping and analysis

Three-dimensional (3D) protein structure mapping and analysis was used to examine and predict the functional impact of all non synonymous variations identified in mitochondrial encoded proteins. This approach was developed and validated on a number of complex III and complex IV variations associated with a broad range of human diseases, and has more recently been used to identify mtDNA variations worthy of follow up investigations using *in vitro* functional assays in glioblastoma [19,20, respectively]. Subsequently, several variations found in GBM MT-CYBs and predicted to be of significance have been shown to influence complex III properties such as activity or drug-sensitivity, when individually introduced into yeast, validating the usefulness of the approach [48]. The recent high profile publications of both the mammalian complex I [28] and complex V [29] cryo-EM structures, means here, for the first time the pathogenicity of mtDNA variations within these complexes can also be predicted, something which precluded the earlier studies [19, 20].

In brief, the 13 human mitochondrial protein sequences were used to find the latest, best quality and most similar OXPHOS complex structures for complexes I, III, IV and V available from RCSB PDB (complex I – [PDB: 5LC5] (*B. taurus*) [28]; complex III – [PDB:1NTZ] (*B. taurus*) [30] & [PDB:1BE3] (*B. taurus*) [31] ; complex IV [PDB:2EIJ] (*B. taurus*) [32] and complex V [PDB:5ARA] (*B. taurus*) [29]. Similarity between human sequences and the new complex I and V models are shown in Table S3, while the others have been documented previously [20].

Relative location and conservation of variations between the human and model sequence alignments was identified using ClustalW2 [33] and then annotated with ESPript [34]. Variations were mapped to their most homologous associated structures and displayed in PyMOL (Schrödinger, LLC). In the case of complex I, maps were generated using the X-ray generated *T. thermophilus* [PDB:4HEA] model, as this was not possible using the cryo-EM generated *B. taurus* [PDB: 5LC5] model, which is not available with docked X-ray crystallographic data. Detailed analysis of the effect of variations, based on their proximity to important structural features within the models, was then performed on the most similar structures using COOT [35]. Following this, variations were classified into 1 of 5 distinct pathogenic categories: (i) frameshift, (ii) active site, (iii) binding pocket, (iv) protein interaction region or (v) non pathogenic.

3. Results and Discussion

3.1. Generation of new complete mitochondrial genomes

Long PCR and sequencing of whole blood DNA from BBM patients yielded 13 high quality new complete mitochondrial genomes (BBM1-13]. Each genome was compiled from 1,626,861 ± 196,610 reads with a mean length of 150 ± 0.3 nucleotides (Table 1). These reads were then mapped against the revised Cambridge Reference Sequence (rCRS [23]), forming alignments with an average total of 1,573,446 ± 196,129 reads and a mean depth of coverage per nucleotide of 14,224 ± 1,781 over the entire mitochondrial genome (Table 1).

3.2 Highly variable mtDNA profiles

In total, 330 variations were identified across the 13 mitochondrial genomes; this consisted of 160 nucleotide positions that were found to be altered in just one or multiple samples (detailed in Table S4). The vast majority of the total variations (274/330, 83%) were nucleotide substitutions with the remaining variations being indels, including 36 insertions and 20 deletions.

The average number of variations in each sample was 25.38 ± 15.98, with most variations occurring in BBM4 and BBM8, which contained 58 and 53, respectively. BBM6, BBM7 and BBM11 contained just 11 variations each (Figure 1A). The range of heteroplasmy for each variation detected in each sample was also diverse, with no significant differences in the average heteroplasmy (72.03% ± 34.84%) observed between samples (Figure 1B). The total number of variations and average heteroplasmy observed in this small cohort of individual samples did not appear to correlate with markers known to be associated with the likelihood of developing BBM, i.e. triple negative or Her2+, or ER- (see Rostami et al. for review [36]), nor whether the patient had been exposed to radio- or chemotherapy prior to the blood DNA being taken for analysis (Table S1), despite some studies showing chemotherapies can result in mtDNA instability [37]. This prompted us to look in more detail at the specific variations present.

3.3 D-loop, mt nd4, mt-nd5 and mt cyb are variation hotspots

Variations were identified in all regions of the mtDNA genome apart from *mt nd3* (Figure 2A). Variations were predominantly situated in the displacement loop (*D loop*), a non coding control region and known "hotspot" for mtDNA variation in many cancers [38,39], which contained 45% (149/330) of all variations.

In the coding regions, the largest number of variations (12%; 41/330) occurred in *mt rnr2*, which encodes for the mitochondrial 16S rRNA. Three other mitochondrial coding genes also contained a high number of variations: *mt nd4* and *mt nd5* (both encoding complex I subunits) each contained 5% (18/330) and *mt cyb* (encoding for a complex III subunit), which contained 6% (19/330) of the total variations. These patterns did not appear to be correlated with locus length (Figure 2B).

3.4 Heteroplasmy, prevalence and disease-association characteristics of recurrent variations

Seventy out of 160 variations appeared in 3 or more samples (Table S4). Thirty-two (46%) of the recurrent variations were identified in the *D-loop* and included: *A73G*, *T152C*, *C299A*, *A301C*, *302insC*, *T310C*, *310insC*, *A16183C*, *C16256T*, *T16311C*, *T16326C* and *T16519C*. The remaining 38 (54%) of the recurrent variations were present in the RNA-coding regions. The most frequently recorded variations were (in descending order of frequency): *A2706G* in *mt rnr2*, *C7028T* in *mt co1*, *G11719A* in *mt-nd4*, *A12308G* in *mt tl2*, *C14766T* in *mt cyb, A11467G* in *mt nd4*, *G12372A* in *mt nd5* and *G709A* in *mt rnr1*.

Broadly speaking, the recurrent variations could be categorised into two groups (Table S4):

(1) a large group of almost 100% heteroplasmy, probably reflecting a germ-line origin. Among these, three subgoups were identified. Those which were found at a similar frequency in the bloods of BBM patients and normal subjects (*T152C, T16311C* and *T16362C*) and therefore unlikely to be of significance, those which were enriched by 10% or more frequent in the bloods of BBM patients than normal subjects (*A11467G, G12372A, G709A, A12308G, T10463C* and *C16256T*) and therefore

potentially significant, and finally, those which were depleted by ~10% or more in the bloods of BBM patients than normal subjects (*G11719A, C14766T, A73G, A2706G, C7028T* and *T16519C*).

(2) a small group with reasonably low average heteroplasmy (34.5 ± 26 %), which could be of either germ-line or tumour cell origin. With the exception on *A16183C*, all of the variations in this group were enriched by ~10% or more in the bloods of BBM patients when compared to normal subjects (*C299A, A301C, 302insC, A3105T, 3105insAC, T310C* and *310insC*) and therefore of potential significance. Interestingly, among the somatic variations detected in primary breast cancer cells, those somatic variations detected in the "*D310* repeat region of the *D-loop*", of which *T310C* and *310insC* (this study) would be examples, have been promoted as marker for breast tumourigenesis [40]. *302insC, T310C, C16256T, T16362C, G709A, T10463C, A11467G, G12372A, A12308G* and *A16183C* have been detected in multiple tumour types, including breast and brain tumours previously; *T16519C* and *A73G* have been detected in multiple tumour types including brain tumours previously; and *T310C* and *T16311C* have been detected in melanoma and prostate, respectively, although not breast or brain tumours. *T152C* has been detected in ovarian and breast tumours. On the other hand, *A2706G, C7028T, G11719A, C14766T, C299A, A301C,* A*3105T* and *3105insAC* have not been reported in tissues of patients with pathology previously (Table S4).

Although, there is evidence that suggests mtDNA variations can play a role in primary breast cancer, e.g. the germ-line variations*: T16189C* [15], *G9055A, T16519C, T239C* and *C16207T* [16,17] are associated with breast cancer development susceptibility and *G10398A* is associated with higher breast cancer risk in African women [18], these were not among the recurrent variations in our cohort.

3.5 Global structural maps

Overall 116 variations were identified in protein coding regions, of these 79 (68%) were synonymous and 37 (32%) were non synonymous (Figure 3A). All non-synonymous variations in the mtDNA-encoded OXPHOS genes were mapped onto their corresponding structural homologs, yielding the first insight into their distribution across the complexes. Figure 4, which is a compendium of the 23 non-synonymous variations identified in all of the BBM patients, reveals a non-uniform distribution. Most variations were found in (in descending order): complex I, IV, III and V. The high frequency of variations in complexes I, IV and III can be accounted for by variation hotspots (again in descending order) in MT-ND1 and ND5 of complex I, MT-CO3 of complex IV and MT-CYB of complex III, respectively. MT-ATP6 also contained a couple of variations, while MT-CO1, MT-CO2, MT-ND2 and MT-ND6 each contained just one variation; MT-ND3, ND4, ND4L, and MT-ATP8 (not shown), however, remained variation free. Among the non-synonymous variations, 48% (11 of 23) were conserved between the human sequences and the bovine models (Figure S1). Despite complex I containing most of the non-synonymous variations, just 30% (3 of 10) were conserved, followed by 40% in complex III (2 of 5), 60% in complex IV (4 of 6) and 100% in complex V (2 of 2)(Figure 4, Figure S1 and Table S3).

3.6 Three functional candidates identified

Owing to the 61-91% identity that exists between the human sequences and the homologs ([19] and Table S3), we were able to perform, for the first time, detailed 3D structural analysis on all complexes of the OXPHOS system containing mtDNA-encoded subunits. Using this approach, we were able to predict just 3 out of the 23 non-synonymous variations were likely to have a functional impact at the level of protein structural changes, and could be put into previously defined structural/functional classes [19,20]. One occurred in a complex III ligand/inhibitor binding pocket region (class 3) and two occurred in protein-protein interaction regions (class 4), one in complex IV and one in complex V. The remaining variations, including all 10 in complex I, are predicted to be non-functional (Table 2).

The absence, or very low frequency, of all 3 of the functional candidates in the bloods of normal healthy individuals further hints that the variations could be important. According to the HmtDB: the complex IV variation (E153G) has previously been documented in diabetes, while *T14819insTTCTATA* and S99P have not been previously been documented were not reported (Table

S4 and S5). Intriguingly, S99P has been reported as a somatic variant in the TCGAs-BRCA project of primary breast invasive carcinoma patients.

3.7 Mechanistic insights

The structural consequences of the 3 variations predicted to be functional are illustrated in Figures 5 and 7.

Figure 5A to D depicts the structural consequence of the *14819insTCTATA* in frame insertion on the only mtDNA-encoded subunit of Complex III: MT-CYB. Complex III is at the centre of the OXPHOS system, and catalyses the transfer of electrons from ubiquinol to cytochrome c, which is coupled with the translocations of protons across the inner mitochondrial membrane. Complex III is also a major site for ROS production. MT-CYB lies at the centre of dimeric complex III, and forms intimate interactions with 20 to 22 additional subunits (10 to 11 per MT-CYB monomer) that are nuclear DNA-encoded. The subunit contains two hemes and two inhibitor binding sites, the Qo and Qi sites. Although the codons remain in register, the introduction of 6 new nucleotides at position 14819 of the mitochondrial genome eventually causes a serine to phenylalanine substitution at position 25 before the addition of a further 2 new amino acids, a tyrosine and a threonine at position 26 and 27, respectively, of the MT-CYB polypeptide (Figure S2). Although it is difficult to model the precise effects of these changes, such a substantial change is unlikely to be easily accommodated by the wild type protein (Figure 5A, C) and we predict local mis-folding of the mutant MT-CYB in a region proximal to the ubiquinol/ubiquinone binding site (the Qi-site; Figure 5B). The mis-folding could also affect the surface of MT-CYB (Figure 5C, D), disrupting its interactions with the nuclear subunits, and therefore the stability of complex III. Taken together, *14819insTCTATA* is likely have an effect on complex III activity. As inhibitors can also bind to the Qi-site, this variation could also influence the efficacy of complex III inhibitors eliciting mitochondrially-mediated apoptosis.

A9664G causes the amino acid substitution E153G in MT-CO3. Complex IV, which is the terminal enzyme of the MRC that catalyses the electron transfer from cytochrome c to oxygen, which (like in complex III) is also coupled to proton translocation across the inner mitochondrial membrane. MT-CO3 forms a homodimer at the centre of complex IV. Although MT-CO3 does not house any active sites/binding pockets, through its interactions with MT-CO1 and various nuclear encoded subunits e.g. COX6A, it is thought to play a key role in the formation and therefore stabilisation of complex IV dimer [41]. In wild type MT-CO3, E153 occupies a position on the surface of MT-CO3 and interacts with alanine at position 13 of the nuclear-encoded subunit COX6A (Figure 6A). The E153G substitution is predicted to eliminate a hydrogen bond interaction (Figure 6B) that usually occurs between the large, negatively charged wild type glutamic acid residue of MT-CO3 and the smaller hydrophobic alanine of nuclear COX6A, potentially undermining the interaction between E153G and COX6A (Figure 6C, D). In summary, E153G could affect complex IV stability and therefore activity. Consistent with this view, knockdown of the COX6A subunit (although a more severe scenario than E153G) has been shown to reduce activity of complex IV [42–44].

T8821C causes an S to P substitution at position 99 of the MT-ATP6 polypeptide (Figure 7), which forms part of the F0 region of complex V. Complex V is the primary producer of ATP in eukaryotic cells, and in addition to the hydrophobic F0 region that traverses the inner mitochondrial membrane, it has a hydrophilic F1-ATPase region that protrudes like a 'lollipop' head into the mitochondrial matrix. In addition to MT-ATP6 (also known as the a-subunit) the F0 region is formed of an additional mtDNA-encoded subunit (MT-ATP8, also referred to as A6L-subunit), several nuclear encoded subunits (e, f, g, DAPIT, 2 hydrophobic alpha-helices of b), a proteolipid, and the c8-ring ('rotor'). The F0 and F1 regions are linked via the subunits OSCP, d, F6 and the hydrophilic portions of subunit b, referred to as the peripheral stalk. It is proposed that complex V drives the protons that arise in the inter membrane space (IMS) as a result of MRC activity through two half-channels formed by MT-ATP6 and the 'rotor'. In one half channel, conserved Glu58 in one of the c-subunits receives a proton from the IMS, which induces the c-ring to rotate, with the protonated Glu residues moving away from the static portions of the F0 region (the 'stator' which includes the two mtDNA-encoded subunits as well as components of the peripheral stalk). Once the protonated Glu residues have

moved almost a full circle around the 'rotor', returning to the static portion of the F0 region, the other half channel provides an exit for the protons into the matrix. The resulting in deprotonation of the Glu residue, makes it ready to receive another round of protonation via the IMS. The mechanical energy created by the rotation, is transmitted via subunits gamma, sigma and epsilon (the 'shaft'), causing conformational changes in the alpha and beta subunits (the 'blades') of the F1 region, and ultimately ATP hydrolysis (reviewed in [29]). Complex V is often referred to as a Brownian ratchet as rotation of the c-ring occurs through Brownian motion, but it is also referred to as a turbine as the enzyme has all the parts required to make a simple turbine (a rotor, stator, shaft and blades). Wild type MT-ATP6 is a 6-alpha helical bundle, whose tertiary folds are likely to be stabilised by the 12 proline residues (Figure S1). These proline-mediated kinks are likely to be indispensable for the positioning of MT-ATP6 close to the nuclear-encoded subunits of the c-ring rotor, as well as subunit b that forms part of the stator, and MT-ATP8, whose function of which is of yet unknown. The introduction of an additional proline residue, downstream of a conserved proline at position 94 of MT-ATP6 (Figure S1, Figure 7), is likely to affect thermodynamics of the whole subunit, and as a consequence, the numerous inter-helical interactions of MT-ATP6 mentioned above [45], this in turn could reduce complex V activity.

3.8 Prevalence of functional candidates in patient bloods

The detailed structural modelling suggests that altogether just 3 out of the 13 (23%) BBM patients carried a single pathogenic variation: BBM4 carries *T14819insTTCTATA* in complex III, BBM12 carries E153G in complex IV, and BBM2 carries S99P in complex V. While the possibility that D-loop, rRNA, tRNA and synonymous variations could also affect the OXPHOS function through a variety of mechanisms, the current lack of *in silico* tools available to accurately predict the structural/functional effect of such variations on OXPHOS proteins limits informative interpretation.

3.9 Summary and future prospects

Although more patients are surviving primary breast cancer, more and more are developing secondary metastases, including to the brain. Although there are existing markers that relate to the likelihood of developing BBM (e.g. triple negative or Her2+, or ER-), more needs to be done to fully understand the process of BBM and develop additional markers that predict whether a patient will develop BBM. MtDNA variations are known to play a role in various diseases, including some cancers. With ever improving and more cost-effective sequencing technologies, an increasing number of mtDNA variations are being detected that are associated with disease. However, proving their role in processes like BBM remains difficult, one reason being that it is impossible to introduce single mtDNA variations into human mtDNA and observe direct phenotypic effects.

Our focus here was two-fold: to screen the bloods of BBM patients using an established and sensitive deep sequencing approach for mtDNA variation detection (and to predict their effect on mitochondrial function using an equally established 3D structural modelling approach [19]). We reveal that the vast majority of mtDNA protein-coding variations detected in the majority of patients tested are not predicted to have a major impact on mitochondrial function. This said, 3 of the mtDNA variations detected in three of the patients were predicted to have an impact at the level of OXPHOS protein structure. The potential multiple origins of the mtDNA variations (e.g. neutrophil, monocyte, myeloid dendritic, natural killer, T and B [49], as well tumour circulating cells [4]) detected in the bloods, as well as the different selective pressures which may act upon them, however, make it difficult to fully predict their significance on the process of BBM (even if they are predicted to evoke a change on the OXPHOS system), and requires further investigation using matched primary tumour-blood-secondary tumour samples to confirm their origin. Other future work could include determining how the 3 mtDNA variations correlate with 1) the predicted alterations in mitochondrial function using *in vitro* studies and 2) with increased risk of developing breast brain metastasis using a much larger cohort of patient samples. This study also serves as a resource by highlighting a list of several mtDNA variations which are predicted to be non-functional and therefore not worthy of further investigation, which should help scientists and clinicians avoid wasting resources. Such a

systematic approach of analyzing mtDNA variations could help improve our understanding of their role in the process of BBM as well as aid the development of markers that predict whether patients develop BBM in the future.

Acknowledgments: Running costs were provided by the Institute of Biomedical and Biomolecular Science (IBBS-fund) and REM's salary was supported by the charity Brain Tumour Research. We would also like to gratefully thank John E. McGeehan for his discussion during the design of the project and comments and guidance on the structural section of this work; Geoffrey J. Pilkington and Helen L. Fillmore for the use of the Brain Tumour Research Centre facilities; Sajid Rafiq, and Will Tapper for their discussions during the design of the project, and Nikki J. Graham and Lorraine T. Duncan for providing the patient blood DNA samples and associated clinical details.

Declaration of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Berman, A.T.; Thukral, A.D.; Hwang, W.T.; Solin, L.J.; Vapiwala, N. Incidence and patterns of distant metastases for patients with early-stage breast cancer after breast conservation treatment. Clin. Breast Cancer. 2013, 13, 88-94.
2. Gil-Gil, M.J.; Martinez-Garcia, M.; Sierra, A.; Conesa, G.; Del Barco, S.; González-Jimenez, S.; Villà, S. Breast cancer brain metastases: A review of the literature and a current multidisciplinary management guideline. Clin Transl Oncol. 2014, 16, 436-46.
3. Engel, J.; Eckel, R.; Aydemir, Ü.; Aydemir, S.; Kerr, J.; Schlesinger-Raab, A; Dirschedl, P.; Hölzel, D. Determinants and prognoses of locoregional and distant progression in breast cancer. Int. J. Radiat. Oncol. Biol. Phys. 2003, 55, 1186-95.
4. Fidler, IJ. The Biology of Brain Metastasis: Challenges for Therapy. Cancer J. 2015, 21, 284-93.
5. Khan, S.; Fagerholm, R.; Rafiq, S.; Tapper, W.; Aittomäki, K.; Liu, J.; Blomqvist, C.; Eccles, D.; Nevanlinna, H. Polymorphism at 19q13.41 predicts breast cancer survival specifically after endocrine therapy. Clin. Cancer Res. 2015, 21, 4086-96.
6. Fagerholm, R.; Schmidt, M.K.; Khan, S.; Rafiq, S.; Tapper, W.; Aittomäki, K.; Groco, D.; Heikkinen, T.; Muranen, T.A.; Fasching, P.A.; et al. The SNP rs6500843 in 16p13.3 is associated with survival specifically among chemotherapy-treated breast cancer patients. Oncotarget. 2015, 6, 7390-407.
7. Rizzuto, R.; De Stefani, D.; Raffaello, A.; Mammucari, C. Mitochondria as sensors and regulators of calcium signalling. Nat. Rev. Mol. Cell Biol. 2012, 13, 566-78.
8. Sena, L.A.; Chandel, N.S. Physiological roles of mitochondrial reactive oxygen species. Mol. Cell. 2012, 48, 158-66.
9. Jeong, S.Y.; Seol, D.W. The role of mitochondria in apoptosis. BMB Rep. 2008, 41, 11-22.
10. Wang, C.; Youle, R.J. The role of mitochondria in apoptosis*. Annu. Rev. Genet. 2009, 43, 95-118.
11. Schon, E.A.; DiMauro, S.; Hirano, M. Human mitochondrial DNA: roles of inherited and somatic mutations. Nat. Rev. Genet. 2012, 13, 878-90.
12. Rossignol, R.; Faustin, B.; Rocher, C.; Malgat, M.; Mazat, J.P.; Letellier, T. Mitochondrial threshold effects. Biochem. J. 2003, 370, 751-62.
13. Chen, E.I. Mitochondrial dysfunction and cancer metastasis. J. Bioenerg. Biomembr. 2012, 44, 619-22.
14. Ishikawa, K.; Takenaga, K.; Akimoto, M.; Koshikawa, N.; Yamaguchi, A.; Imanishi, H.; Nahada, K.; Honma, Y.; Hayashi, J. ROS-generating mitochondrial DNA mutations can regulate tumor cell metastasis. Science. 2008, 320, 661-4.
15. Wang, Y.; Liu, V.W.S.; Tsang, P.C.K.; Chiu, P.M.; Cheung, A.N.Y.; Khoo, U.S.; Nagley, P.; Ngan, H.Y.S. Microsatellite instability in mitochondrial genome of common female cancers. Int. J. Gynecol. Cancer. 2006, 16, 259-66.
16. Bai, R.K.; Leal, S.M.; Covarrubias, D.; Liu, A.; Wong, L.J.C. Mitochondrial genetic background modifies breast cancer risk. Cancer Res. 2007, 67, 4687-94.
17. Czarnecka, A.M.; Krawczyk, T.; Plak, K.; Klemba, A.; Zdrozny, M.; Arnold, R.S.; Kofler, B.; Golik, P.; Szybinskia, A.; Lubinski, J.; Mossakowska, M.; Bartnik, E.; Petros, J.A. Mitochondrial genotype and breast cancer predisposition. Oncol. Rep. 2010, 24, 1521-34.

18. Canter, J.A.; Kallianpur, A.R.; Parl, F.F.; Millikan, R.C. Mitochondrial DNA G10398A polymorphism and invasive breast cancer in African-American women. Cancer Res. 2005, 65, 8028-33.
19. Lloyd, R.E.; Keatley, K.; Littlewood, D.T.J.; Meunier, B.; Holt, W.V.; An, Q.; Higgins, S.C.; Polyzoidis, S.; Stephenson, K.F.; Ashkan, K.; Fillmore, H.L.; Pilkington, G.J.; McGeehan, J.E. Identification and functional prediction of mitochondrial complex III and IV mutations associated with glioblastoma. Neuro-oncology. 2015, 17, 942–52.
20. Lloyd, R.E.; McGeehan, J.E. Structural Analysis of Mitochondrial Mutations Reveals a Role for Bigenomic Protein Interactions in Human Disease. PLoS One. 2013, 8, e69003.
21. Eccles, D.; Gerty, S.; Simmonds, P.; Hammond, V.; Ennis, S.; Altman, D.G. Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH): study protocol. BMC Cancer. 2007, 7, 160.
22. Rafiq, S.; Tapper, W.; Collins, A.; Khan, S.; Politopoulos, I.; Gerty, S.; Blomqvist, C.; Couch, F.J.; Nevanlinna, H.; Liu, J.; Eccles, D. Identification of inherited genetic variations influencing prognosis in early-onset breast cancer. Cancer Res. 2013, 73, 1883-91.
23. Andrews, R.M.; Kubacka, I.; Chinnery, P.F.; Lightowlers, R.N.; Turnbull, D.M.; Howell, N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. 1999, 23, 147.
24. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009, 25, 1754-60.
25. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; DePristo, M.A. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010, 20, 1297 303.
26. DePristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V; Maguire, J.R.; Hartl, C.; Philippakis, A.A.; del Angel, G.; Rivas, M.A.; Hanna, M.; et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 2011, 43, 491-8.
27. Ju, Y.S.; Alexandrov, L.B.; Gerstung, M.; Martincorena, I.; Nik-Zainal, S.; Ramakrishna, M.; Davies, H.R.; Papaemmanuil, E.; Gundem, G.; Shlien, A.; et al. Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. Elife. 2014, 3, e02935.
28. Zhu, J.; Vinothkumar, K.R.; Hirst, J. Structure of mammalian respiratory complex I. Nature. 2016, 536, 354-8.
29. Zhou, A.; Rohou, A.; Schep, D.G.; Bason, J.V.; Montgomery, M.G.; Walker, J.E.; Grigorieffniko, N.; Rubinstein, J.L. Structure and conformational states of the bovine mitochondrial ATP synthase by cryo-EM. Elife. 2015, 4, e10180.
30. Gao, X.; Wen, X.; Esser, L.; Quinn, B.; Yu, L.; Yu, C.A.; Xia, D. Structural Basis for the Quinone Reduction in the bc1 Complex: A Comparative Analysis of Crystal Structures of Mitochondrial Cytochrome bc1 with Bound Substrate and Inhibitors at the Qi Site. Biochemistry. 2003, 42, 9067-80.
31. Iwata, S.; Lee, J.W.; Okada, K.; Lee, J.K.; Iwata, M.; Rasmussen, B.; Link, T.A.; Ramaswamy, S.; Jap, B.K. Complete structure of the 11-subunit bovine mitochondrial cytochrome bc1 complex. Science. 1998, 281, 64-71.
32. Muramoto, K.; Hirata, K.; Shinzawa-Itoh, K.; Yoko-o, S.; Yamashita, E.; Aoyama, H.; Tsukihara, T.; Yoshikawa, S. A histidine residue acting as a controlling site for dioxygen reduction and proton pumping by cytochrome c oxidase. Proc. Natl. Acad. Sci. U.S.A. 2007, 104, 7881-6.
33. Larkin, M.; Blackshields, G.; Brown, N.; Chenna, R.; McGettigan, P.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; Thompson, J.D.; Gibson, T.J.; Higgins, D.G. ClustalW and ClustalX version 2. Bioinformatics. 2007, 23, 2947-8.
34. Gouet, P.; Courcelle, E.; Stuart, D.I.; Metoz, F. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics. 1999, 15, 305-8.
35. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of Coot. Acta. Crystallogr. Sect. D Biol. Crystallogr. 2010, 66, 486-501.
36. Rostami, R.; Mittal, S.; Rostami, P.; Tavassoli, F.; Jabbari, B. Brain metastasis in breast cancer: a comprehensive literature review. J. Neurooncol. 2016, 127, 407-14.
37. Wiseman, H.; Halliwell, B. Damage to DNA by reactive oxygen and nitrogen species: role in inflammatory disease and progression to cancer. Biochem. J. 1996, 313, 17–29.
38. Akouchekian, M.; Houshmand, M.; Hemati, S.; Ansaripour, M.; Shafa, M. High rate of mutation in mitochondrial DNA displacement loop region in human colorectal cancer. Dis. Colon Rectum. 2009, 52, 526-30.
39. Fliss, M.S.; Usadel, H.; Caballero, O.L.; Wu, L.; Buta, M.R.; Eleff, S.M.; Jen, J.; Sidransky, D. Facile detection of mitochondrial DNA mutations in tumors and bodily fluids. Science. 2000, 287, 2017-9.

40. Parrella, P.; Xiao, Y.; Fliss, M.; Sanchez-Cespedes, M.; Mazzarelli, P.; Rinaldi, M.; Nicol, T.; Gabrielson, E.; Cuomo, D.; Cohen, D.; et al. Detection of mitochondrial DNA mutations in primary breast cancer and fine-needle aspirates. Cancer Res. 2001, 61, 7623-6.
41. Sedlák, E.; Robinson, N.C. Sequential dissociation of subunits from bovine heart cytochrome c oxidase by urea. Biochemistry. 2009, 48, 8143–50.
42. Quintens, R.; Singh, S.; Lemaire, K.; de Bock, K.; Granvik, M.; Schraenen, A.; Vroegrijk, I.O.C.M; Costa, V.; van Noten, P.; Lambrechts, D.; et al. Mice Deficient in the Respiratory Chain Gene Cox6a2 Are Protected against High-Fat Diet-Induced Obesity and Insulin Resistance. PLoS One. 2013, 8, e56719.
43. Tamiya, G.; Makino, S.; Hayashi, M.; Abe, A.; Numakura, C.; Ueki, M.; Tanaka, A.; Ito, C.; Tohimori, K.; Ogawa, N.; et al. A mutation of COX6A1 causes a recessive axonal or mixed form of Charcot-Marie-Tooth disease. Am. J. Hum. Genet. 2014, 95, 294-300.
44. Fornuskova, D.; Stiburek, L.; Wenchich, L.; Vinsova, K.; Hansikova, H.; Zeman, J. Novel insights into the assembly and function of human nuclear-encoded cytochrome c oxidase subunits 4, 5a, 6a, 7a and 7b. Biochem. J. 2010, 428, 363-74.
45. Schmidt, T.; Situ, A.J.; Ulmer, T.S. Structural and thermodynamic basis of proline-induced transmembrane complex stabilization. Sci Rep. 2016, 6, 29809.
46. Kenny TC, Hart P, Ragazzi M, Sersinghe M, Chipuk J, Sagar MA, Eliceiri KW, LaFramboise T, Grandhi S, Santos J, Riar AK, Papa L, D'Aurello M, Manfredi G, Bonini MG, Germain D. Selected mitochondrial DNA landscapes activate the SIRT3 axis of the UPR(mt) to promote metastasis. Oncogene. 2017 Apr 3. doi: 10.1038/onc.2017.52. [Epub ahead of print] PubMed PMID: 28368421.
47. Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD, Marshall D. Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform. 2013 Mar;14(2):193-202.
48. Song Z, Laleve A, Vallières C, McGeehan JE, Lloyd RE, Meunier B. Human Mitochondrial Cytochrome b Variants Studied in Yeast: Not All Are Silent Polymorphisms. Hum Mutat. 2016 Sep;37(9):933-41.
49. Zhang P, Samuels DC, Wang J, Zhao S, Shyr Y, Guo Y. Mitochondria single nucleotide variation across six blood cell types. Mitochondrion. 2016 May;28:16-22.

Supplementary Materials: Figure S1. Location of non-synonymous amino acid substitution sites following the alignments of human (rCRS) against the bovine homologous mtDNA-encoded polypeptide sequences. Figure S2. Comparison of the wild type *mt-cyb* and MT-CYB sequences against the variant sequences which contain a 6-nuclotide insertion at position 14819. Table S1: Details of the breast-to-brain metastasis patients from which whole-blood mitochondrial genomes were derived. Table S2. Long PCR Primer details. Table S3. Similarity of human sequences vs the homologous structural models used to analyse the effect mtDNA-encoded OXPHOS variations identified in breast-to-brain metastasis bloods. Table S4. Details of each mtDNA variation identified in breast-brain metastasis blood samples including location, average heteroplasmy and prevalence. Table S5. Reported disease associations of each mtDNA variation identified from breast-to-brain metastasis patient bloods.

Tables with captions

Table 1. Summary of sequencing read statistics used to generate new complete breast-to-brain metastasis blood mtDNAs.

Sample ID	Total number of reads	Number of reads mapped	Mean read length	Reads Mapped (%)	Mean DOC/nt	Accession No
BBM1	1286809	1244758	150	97	11255	MF317865
BBM2	1652912	1598870	150	97	14471	MF317866
BBM3	1398435	1354576	150	97	12236	MF317867
BBM4	1733369	1675403	150	97	15133	MF317868
BBM5	1606805	1505339	150	94	13587	MF317869
BBM6	2080624	2044788	150	98	18515	MF317870
BBM7	1616133	1587251	150	98	14358	MF317871
BBM8	1540641	1482632	150	96	13398	MF317872
BBM9	1709391	1649532	150	96	14911	MF317873
BBM10	1796414	1746692	150	97	15804	MF317874
BBM11	1517582	1475338	150	97	13331	MF317875
BBM12	1709260	1623266	150	95	14655	MF317876
BBM13	1500821	1466356	150	98	13252	MF317877
Average	1626861	1573446	150	97	14224	
Standard Deviation	196610	196129	0.3	1.2	1781	

Abbreviation: nt - nucleotide, DOC - depth of coverage

Table 2. Heteroplasmy, prevalence of non-synonymous variations in the mtDNA complex genes identified within the mtDNA landscapes of breast-to-brain blood samples.

Locus	Specific Nucleotide Variant	Amino Acid change	Average heteroplasmy (% ± SD)	% Hmt DB	% BBM	Specific Sample IDs (BBM)
MT-ND1	*A3796G*	T164A	99.6*	0.5	7.7	6
MT-ND1	*C3992T*	T229M	99.2*	0.7	7.7	10
MT-ND1	*A4024G*	T240A	99.2*	0.6	7.7	10
MT-ND1	*T4216C*	Y304H	99.4±0.3*	9.9	15.4	1 and 12
MT-ND2	*A4917G*	N150D	98.6±0.9*	4.7	15.4	1 and 12
MT-ND5	*A12397G*	T21A	99.6*	0.5	7.7	5
MT-ND5	*C12557T*	T74I	98.4*	0.3	7.7	4
MT-ND5	*G13813A*	V493I	99.6*	<0.1	7.7	5
MT-ND5	*G13889A*	C518Y	99.6*	0.1	7.7	10
MT-ND6	*A14582G*	V31A	99.6*	0.6	7.7	10
MT-CYB	*C14766T*	T7I	98.9±0.8*	77.3	46.2	1/ 4/ 8/ 9/ 12/ 13
MT-CYB	*A14793G*	H16R	99.6±0.0*	2.2	15.4	9 and 13
MT-CYB	*T14819insTTCT ATA*	S25	2.4	-	7.7	12
MT-CYB	*A15218G*	T158A	100±0.0*	1.9	7.7	9
MT-CYB	*C15452A*	L236I	99.4±0.9*	9.1	15.4	1 and 12
MT-CO1	*G6267A*	A122T	98.8*	0.2	7.7	2
MT-CO2	*T8265C*	L227P	24.8	<0.1	7.7	4
MT-CO3	*C9469T*	T88I	100*	0.1	7.7	13
MT-CO3	*G9477A*	V91I	99.8±0.3*	4.0	15.4	9 and 13
MT-CO3	*A9664G*	E153G	98.8*	0.1	7.7	4

MT-CO3	*A9667G*	N154S	99.2*	0.7	7.7	9
MT-ATP6	*T8821C*	S99P	83.2	<0.1	7.7	2
MT-ATP6	*G8989A*	A155T	97.2	<0.1	7.7	12

Grey shading indicates variation classed as functional at the level of protein structural changes. The unshaded variations are classified as non-functional. %hmtDB and %BBM indicate prevalence in the human mitochondrial database and BBM cohort (this study), respectively. *Signifies homoplasmic (pure) mutant (see materials and methods).

Figures with captions

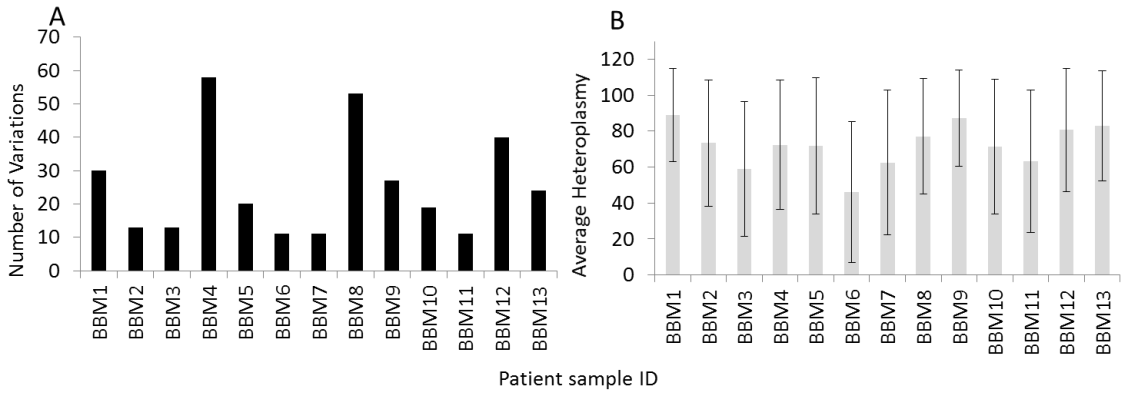

Figure 1. Total number (A) and average heteroplasmy (B) of mtDNA variations in breast-to-brain metastasis patient blood mtDNAs. Error bars indicate standard deviation.

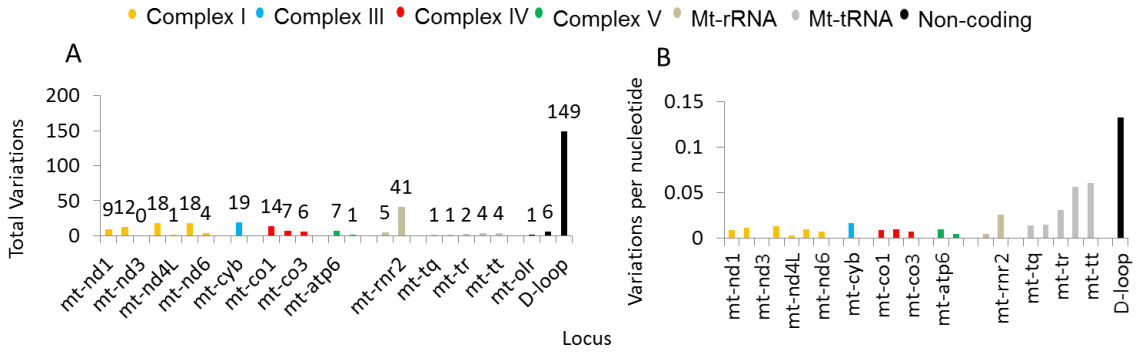

Figure 2. Presence of variations in breast-to-brain metastasis patient blood across mtDNA loci. (A) Total number of mtDNA variations identified in both coding and non-coding loci. (B) Total number of variations expressed per nucleotide in each of the mtDNA loci.

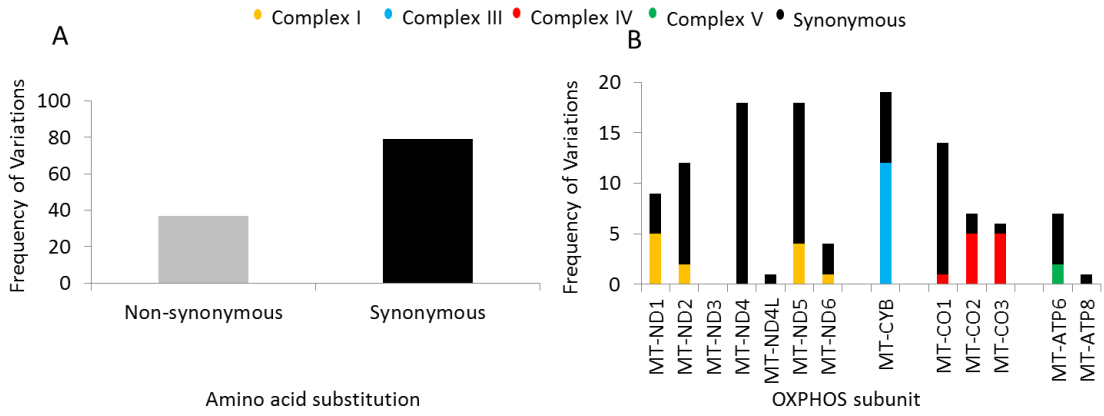

Figure 3. Pattern of synonymous and non-synonymous mtDNA variations in breast-to-brain blood mtDNAs. (A) Total and (B) across loci.

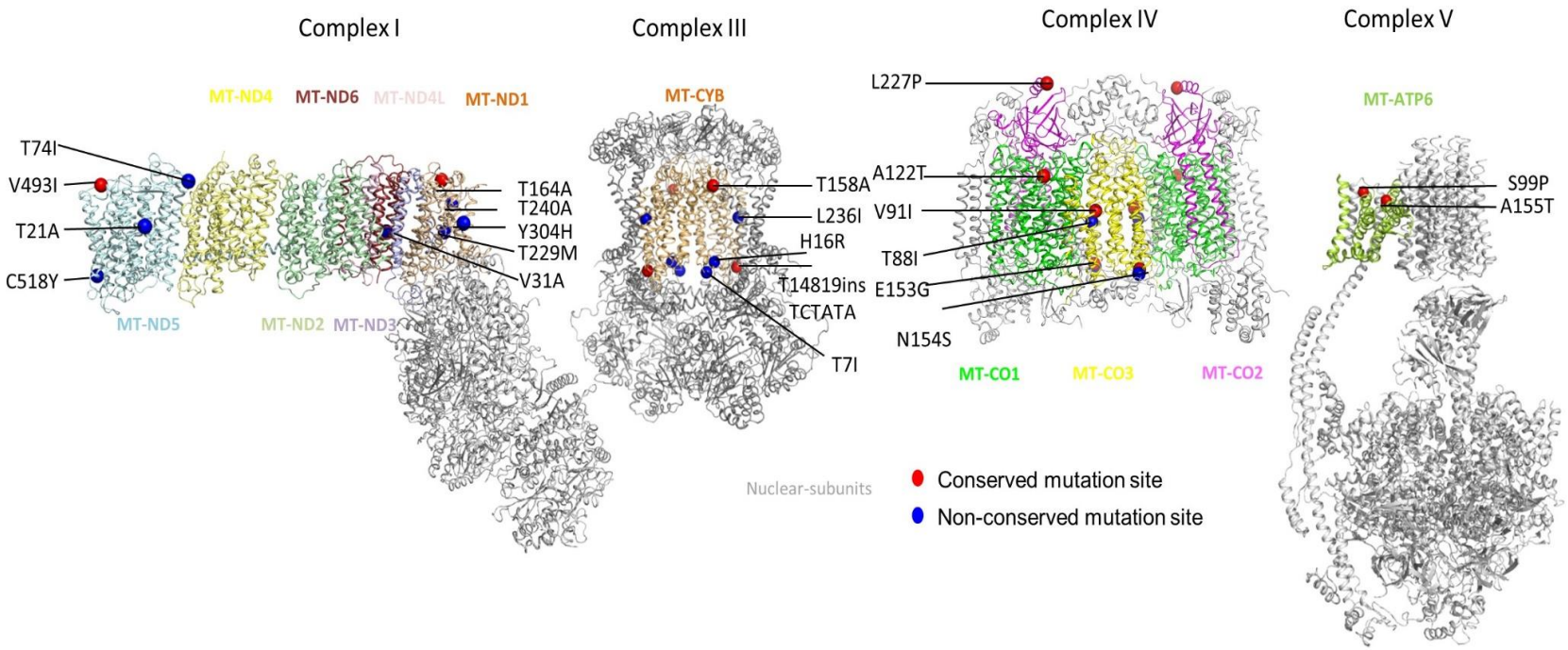

Figure 4. Global map of non-synonymous variations identified in the mitochondrial OXPHOS proteins of breast-to-brain metastasis patient blood mtDNAs. Variations mapped to T. thermophilus complex I homologues [PDB:4HEA], B. taurus complex III homologue [PDB:1BE3], B. taurus complex IV homologue [PDB:2EIJ] and B. taurus complex V homologoue [PDB: 5ARA]. The carbon alphas of residues that are conserved and non-conserved between the human sequences and the bovine models are highlighted as red and blue spheres respectively, including for complex I eventhough the T. thermophilus structure is shown. The conserved residue N150D is not shown as there is no equivalent residue present in 4HEA

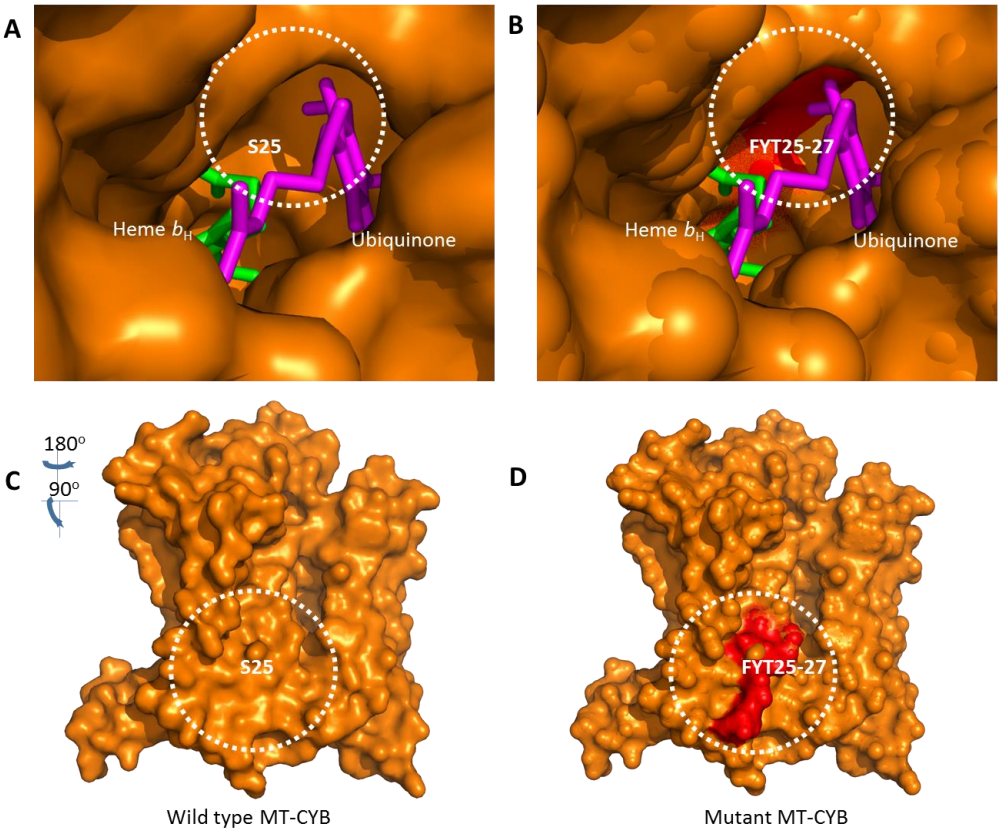

Figure 5. Structural consequences of the class 3 binding pocket variation *14819insTCTATA*. The variation causes a serine to phenylalanine substitution and the addition of two amino acids at positions 25-27 of the complex III protein MT-CYB. MT-CYB is rendered as a space filling model (orange), with (A) the wild type Qi-site shown with the deeply buried heme (green) and bound ubiquinone (magenta; PDB 1NTZ). (B) The amino acid changes induced by the variation occur in the vicinity of the Qi-site (in red). (C) The surface of the wild type MT-CYB monomer. (D) The variant is likely to cause local mis-folding of MT-CYB (in red), disrupting its interaction with surrounding nuclear encoded subunits. Such changes to the MT-CYB Qi-site and surface are likely to interfere with ubiquinone/ubiquinol binding, and ultimately complex III activity.

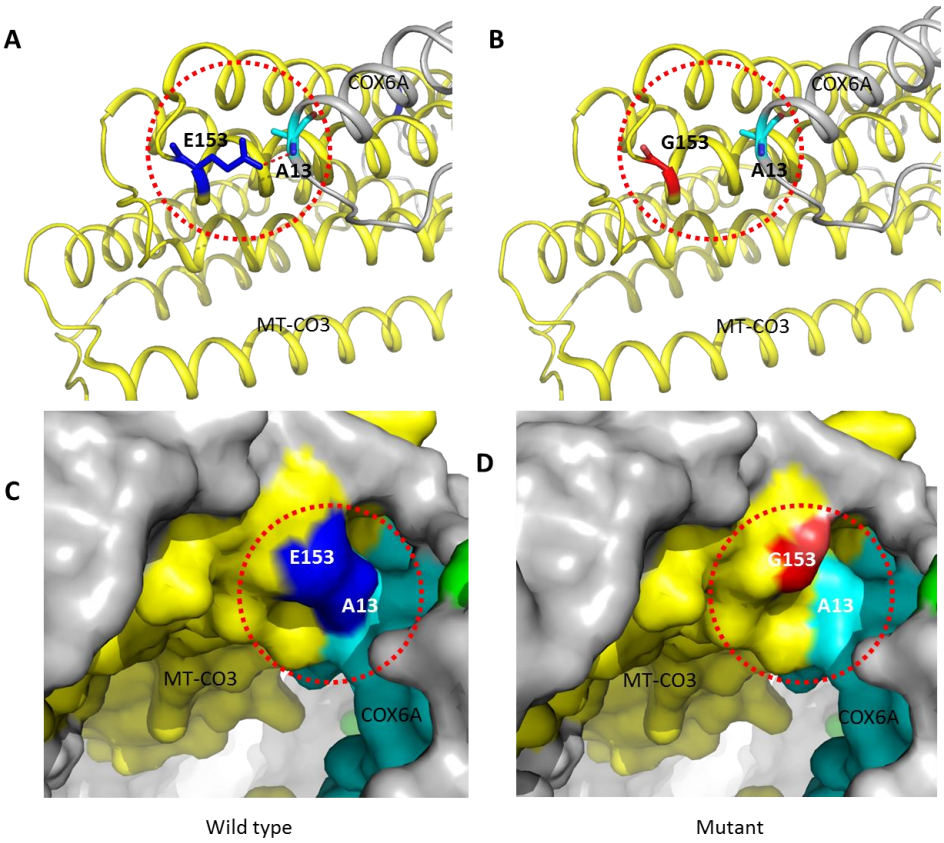

Figure 6. Structural consequences of the class 4 interaction variation *A9664G*. (A) Wild type E153 (blue sticks) in MT-CO3 (yellow ribbons) forms an interaction with A13 (cyan sticks) in COX6A (grey ribbon). (B) The variant G153 (red sticks) results in the loss of a hydrogen bond with A13. (C) The wild type MT-CO3 is shown as a surface model, showing E153 (blue) forms a tight interaction with A13 of COX6A (cyan). (D) Surface illustration of variant G153 (red), which no longer interacts with COX6A, which could affect the stability of complex IV.

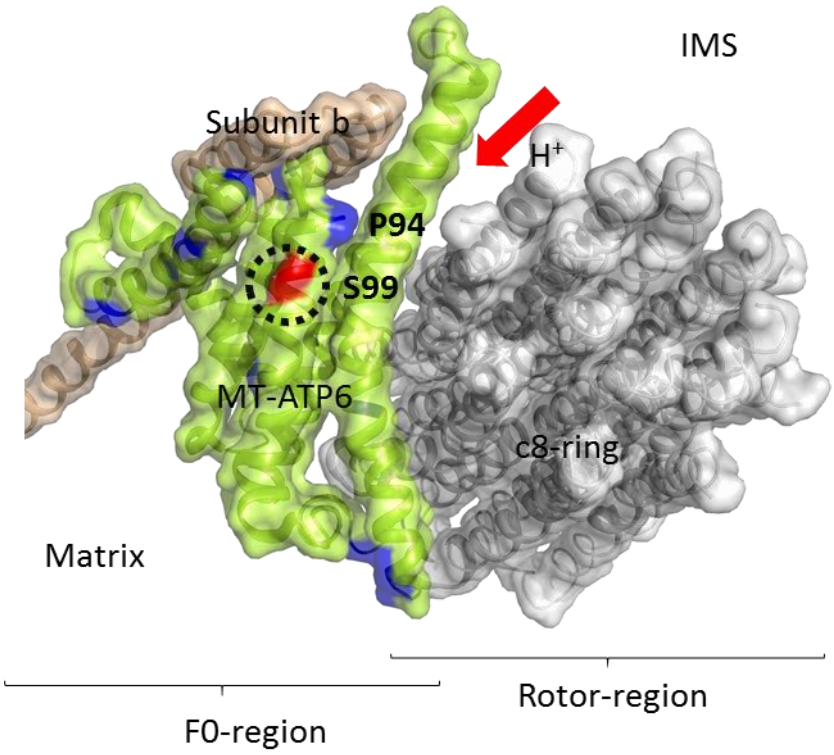

Figure 7. Structural consequences of the class 4 variation *T8821C*. Wild type MT-ATP6 (light green) is shown as a surface model and is formed of six kinked alpha-helices that stabilise the subunit and allow it to form intricate inter-helical interactions with subunit b (wheat), MT-ATP8 (not shown) and the c-ring (grey). Together with the hydrophobic portions of subunit b and MT-ATP8, MT-ATP6 form the static part of the F0 region, with the c-ring forming the rotary portion. It is likely that several proline residues (blue), conserved between the human sequence and the bovine model, are responsible for the kinks observed in MT-ATP6. The substitution of a serine for a proline (red) at position 99 in one of the helices, is likely to interfere with tertiary fold and destabilise the protein and disrupt its interactions, influencing complex V activity.