

ORIGINAL ARTICLE

The pattern of natural selection in somatic cancer mutations of human mtDNA

Pete Stafford^{1,3} and Estella B Chen-Quin²

Tumors frequently contain somatic mutations in the mitochondrial DNA (mtDNA). Whether these mutations have a causal function or are merely an effect is still unclear. As tumor formation is a type of somatic evolution, we examine the cancer mutation pattern for consistency with random forces or selection. We also compare the tumor mutation pattern to that observed in the population to gain insight on the mutation process in cancer. Among germline mtDNAs, all genes show strong deficiency in missense changes, reflecting negative selection during human history. In somatic cancer sequences, mtDNA genes show relaxed negative selection relative to germline, or mutation consistent with neutrality. NADH dehydrogenase subunit 3, cytochrome *c* oxidase subunit 3 and NADH dehydrogenase subunit 4 L in particular show cancer missense mutation rates 9–18 times that of germline. Bootstrap analysis shows cytochrome *B* to have cancer changes in positions of unusually high conservation, suggesting that tumors select for mutations in residues of high functionality. Strong negative selection was detected in mitochondrially encoded cytochrome *c* oxidase 1 (MTCO1), suggesting that tumor cells are dependent upon MTCO1 function. Common population polymorphisms were also frequently reported among somatic tumor mutations. The implication of these ‘somatic polymorphisms’ in tumor growth is discussed.

Journal of Human Genetics (2010) 55, 605–612; doi:10.1038/jhg.2010.76; published online 8 July 2010

Keywords: cancer; germline mtDNA; mitochondria; selection; somatic mtDNA mutation

INTRODUCTION

The discovery of mitochondrial DNA (mtDNA) mutations in most cancer types re-opened the possibility that defects in mitochondria have a causal function in tumor formation. This supported historical observations that tumors have abnormal mitochondria and elevated levels of glycolysis in the presence of oxygen¹ (the ‘Warburg Effect’). Normal mitochondria are also necessary for apoptosis, the loss of which can be a step in oncogenesis.^{2,3} mtDNA mutations associated with cancer were rediscovered in 1998⁴ and received intense investigation for several years following.^{5–10} However, the significance of these mutations in cancer development remains unclear.

Salient arguments are made for aerobic glycolysis being a driving force in tumorigenesis rather than mitochondrial dysfunction.¹¹ In early stages, neoplastic cells encounter a hypoxic environment as cellular oxygen becomes deficient in the center of the growing, unvascularized tumor. Normal oxygen levels may be reestablished upon growth of new blood vessels into the tumor. Tumor cells continue to be subject to hypoxic or anoxic microenvironments as they metastasize and establish new growth in other tissue areas. These periods of normoxia-hypoxia-anoxia may select for constitutive upregulation of glycolysis. Prolonged glycolysis may give cancer cells an advantage, creating an acidic microenvironment, inducing apoptosis in normal cells and selecting for acid-resistant cancer cells.

Dichloroacetate treatment of cancer cells can restore respiratory activity and inhibit cancerous growth, suggesting that mitochondria are not ‘broken’ in cancerous tissue, but only quiescent in response to aerobic glycolysis.¹²

Equally convincing evidence supports an active function for mitochondrial dysfunction in tumor formation, as shown by hereditary cancers of the head and neck caused by inherited mutations in mitochondrial complexes succinate dehydrogenase and fumarate hydratase.² In addition, mtDNA mutations of known pathology (for example, Leigh’s Syndrome) can be transformed into cancer cell lines and result in greatly enhanced tumor growth of the recipient cells *in vivo*.^{13–15} Particularly convincing are cell lines (cybrids) in which mitochondria from tumors of high or low metastatic potential were transformed into recipient tumor cell lines.¹⁶ The recipient tumor cells took on the metastatic potential (aggressive or non-aggressive) of the transferred mtDNA, despite differing nuclear backgrounds, showing that specific mtDNA alleles cause more aggressive tumor behavior.

Tumor formation is a type of somatic evolution; precancerous cells acquire mutations conferring enhanced growth or invasion, leading to a selective advantage in the tissue and rising to clonal frequency within primary tumors. Cancerous cells are also a highly mutagenic environment, containing elevated levels of reactive oxygen species (ROS) and defects in DNA repair and cell cycle checkpoints. Thus, it is unclear

¹Department of Mathematics and Statistics, Kennesaw State University, Kennesaw, GA, USA and ²Department of Biology and Physics, Kennesaw State University, Kennesaw, GA, USA

³Current address: Department of Statistics, Florida State University, Tallahassee, FL 32306, USA.

Correspondence: Dr EB Chen-Quin, Department of Biology and Physics, Kennesaw State University, 1000 Chastain Road, Kennesaw, GA 30144, USA.

E-mail: echen1@kennesaw.edu

Received 30 December 2009; revised 24 April 2010; accepted 31 May 2010; published online 8 July 2010

whether mtDNA mutations found in tumors have a causal function or are merely an effect. If mtDNA mutations are only a result of the cancer environment, then homoplasmic somatic mutations detected in a mature tumor are primarily due to drift, under random or neutral forces. If tumor mtDNA mutations confer a cell proliferation advantage, then the observed DNA variation should be because of positive selection.

The many mtDNA mutations and cancer types render association between mutation and phenotype at too low of power to determine statistical significance. However, the growing use of microarrays on cancer-associated mtDNAs greatly increases the sequence information available, making it possible to run neutrality tests of significance. In natural populations, the pattern of DNA variation may indicate whether a locus has undergone recent selection or is evolving under neutrality. An excess of amino acid changes relative to the rate of silent changes indicates positive selection; fewer missense mutations indicate negative selection. Such tests were used to identify fast-evolving areas of the Trim 5 α -restriction enzyme in primates, which combats the Simian immunodeficiency virus and is heavily selected for a fast mutation rate.¹⁷ Here we apply tests of neutrality to the tumor mutation pattern in the 13 genes of the mitochondrial chromosome. We also compare the somatic tumor mutation pattern to that observed at the organismal or population level to gain insight on the mutation process in cancer cells.

MATERIALS AND METHODS

Mutation information

Whole genome mtDNA mutations that were reported for tumors of the thyroid, head and neck, breast, lung, bladder, esophagus, prostate, colon, ovary, pancreas were analyzed (see Supplementary Table S1 for publication sources of tumor mutation data). Renal tumors, oral tumors and glioma were also studied. The mutations occurring in the 13 coding regions of the mitochondrial chromosome—*MTATP6* [MIM +516060]; *MTAPT8* [MIM +516070]; *MTCO1* [MIM *516030]; *MTCO2* [MIM *516040]; *MTCO3* [MIM *516050]; *MTCYB* [MIM *516020]; *MTND1* [MIM *516000]; *MTND2* [MIM *516001]; *MTND3* [MIM *516002]; *MTND4* [MIM *516003]; *MTND4L* [MIM *516004]; *MTND5* [MIM *516005]; *MTND6* [MIM *516006]—were analyzed for patterns of nonneutral selection or preferential mutation in amino acid positions of high conservation. Each somatic mutation in an independent tumor was counted as a new mutation event. For analyzing mutations at the germline level, 200 sequences of the general population were obtained from the public database mtDB at <http://www.genpat.uu.se/mtDB>.¹⁸ The continental origins of the germline sequences were broadly matched in proportion to the continental origins of the cancer set, which is estimated to be 75% European, 18% Asian and 8% African. The counts of missense and silent changes in germline sequences were calculated according to the method of Ingman and Gyllensten.¹⁹

Public germline databases checked against somatic cancer mutations

Details of database mtDB are available at <http://www.genpat.uu.se/mtDB>,¹⁸ Mitowheel at <http://www.mitowheel.org/mitowheel.html>²⁰ and Mitomap at <http://www.mitomap.org>.²¹ Database frequencies are based on mtDB.

Statistical tests

Missense to silent ratio. Departure from neutrality was determined by deviation of the ratio of missense to silent mutations from that expected by chance. It was calculated according to a modified method of Greenman *et al.*²² and is based on the binomial. The software program DNAsp was used to determine the total possible nonsynonymous and synonymous sites.

Neutrality index. The magnitude of amino acid variation in tumors relative to germline amino acid variation was calculated as (A-cancer/S-cancer)/(A-germline/S-germline) and is based on the neutrality index (NI) for determining excess amino acid variation within species.²³

Calculation of conservation indices. The conservation index (CI) is the percentage of species with amino acids identical to that of healthy humans at each amino acid position.²⁴ Interspecific protein alignments were created using Clustal X 2.05 (freeware available from the Conway Institute, Dublin, Ireland, at www.clustal.org)²⁵ and T-Coffee Vers. 5.72 (freeware opensource package from the Center for Genomic Regulation (CRG), Barcelona, Spain; available at www.tcoffee.org)²⁶ Alignment scores as evaluated by T-Coffee ranged between 50 and 91; both methods gave very similar alignments with identical scores. In the rare instance that scores between Clustal X and T-Coffee alignments differed (*MTATP6*), then we used the higher scoring alignment. Representative amino acid sequences for each of the mtDNA coding genes were obtained from RefSeq. The curated sequences were chosen if more than one sequence was available.

Bootstrap analysis. Bootstrap simulations for the cancer-specific mutations were performed on each of the mtDNA coding genes, 1 000 000 000 iterations. The sample size for each iteration matched the observed sample size of cancer-specific amino acid mutations and allowed for replacement. In an effort to separate neutral positions from the potentially cancer-causing positions, polymorphic germline positions were excluded. The average CI was calculated for each simulation, and *P*-values were derived from the bootstrap distribution of mean conservation indices.

RESULTS

Summary of data

This data set represents the querying of 13 cancer types and a minimum of 543 tumors, 356 of which were reported with somatic mutations different from the patient's germline (see Table 1, for summary of tumor types and Supplementary Table S1 for publication list of tumor mutation data). The ethnic composition of the cancer data set is estimated to be approximately 75% European descent, 18% Asian descent and 8% of African descent. Germline sequences for comparison between cancer mutation processes and organismal were obtained from the public database, *mtDB—Human Mitochondrial Genome Database*¹⁸ at <http://www.genpat.uu.se/mtDB>.

Spectrum of mutations

A total of 485 point mutations were reported in the coding regions of the mtDNA, 245 of which are missense. Overall, 59 of the 245 tumor

Table 1 Samples reported^a with mtDNA tumor mutations, by cancer type

Cancer type	No. of samples sequenced	No. of samples with somatic mutation	Proportion of total samples with somatic mutation (%)
Bladder	14	12	3.4
Breast	72	57	16.0
Colon	10	7	2.0
Esophageal	20	11	3.1
Glioma	16	7	2.0
Head-neck	96–180 ^{b,c}	72	20.2
Lung	71 ^d	40	11.2
Oral	18	14	3.9
Ovarian	10	6	1.7
Pancreatic	5	4	1.1
Prostate	33	8	2.2
Renal	32	19	5.3
Thyroid	146 ^b	99	27.7
Total	543 ^e	356	100

^aList of publication sources is in Supplementary Information Table S1.

^bTotal number of samples sequenced cannot be absolutely determined from publication.

^cTwenty five of these samples were selected for presence of known somatic mutation.

^dTwo of these samples were selected for presence of known somatic mutation.

^eTotal represents the minimum possible number.

Table 2 (a) Somatic cancer missense mutations also found as germline polymorphisms and (b) somatic missense mutations unique to cancer set or at very low frequency in public germline databases

No. of independent tumors	Polymorphism	Gene	a.a. change ^a	mtDB frequency (%)	Average CI ^b
(a)					
12	m.8701A>G	<i>MTATP6</i>	p.Thr59Ala	34	0.15
10	m.10398A>G	<i>MTND3</i>	p.Thr114Ala	45	0.31
4	m.9477G>A	<i>MTCO3</i>	p.Val91Ile	3.6	0.92
2	15 different mutations	Multiple	Multiple	0.074–2.0	0.45
1	45 different mutations	Multiple	Multiple	0.037–9.02	0.92
(b)					
3	m.4831G>A, m.10695G>A	<i>MTND2, MTND4L</i>	p.Gly121Asp, p.Ala76Thr	NA	1.0, 0.67
2	15 different mutations	Multiple	Multiple	NA	0.74
1	106 different mutations	Multiple	Multiple	NA	0.74

Abbreviations: a.a., amino acid; CI, conservation index; NA, not applicable.

^aAmino acid changes are theoretically deduced.

^bConservation index, a measure of interspecific conservation based on the percentage of species that match healthy humans at this amino acid position. CI=1.0 denotes 100% conservation (Ruiz-Pesini *et al.*, 2004²⁴).

missense mutations represent the same mutation event in independent tumor samples; thus, 186 different amino acid mutations are represented. Out of 186 somatic missense mutations, 63 coincide with polymorphisms found in the public germline databases, mtDB (<http://www.genpat.uu.se/mtDB>),¹⁸ or at high frequency in MitoWheel²¹ (<http://www.mitowheel.org/mitowheel.html>) or Mitomap²⁰ (<http://www.mitomap.org>; Table 2a). The remaining 123 missense mutations were not reported in any of the databases or at extremely low frequency (Table 2b).

Common haplogroup lineages found among reported somatic mtDNA mutations

Common germline polymorphisms such as m.8701A>G or m.10398A>G were found at high frequency among the somatic tumor mutations (Table 2a). This may be due to inherent instability at these sites, detected at both the germline and somatic level, or due to contamination by germline mtDNAs during sequencing of tumor mtDNAs. To distinguish whether these mutations are truly *de novo* somatic mutations or germline contaminants, we checked for other germline variants associated with the m.8701A >G or m.10398A >G polymorphisms and which comprise well-known mtDNA haplogroups. Among the 12 tumor samples reported to have a somatic mutation at m.8701A>G, eight of these samples also contain somatic mutations at m.10398A >G. Other linked polymorphisms include m.9540T>C, m.12705C>T, m.15301G>A and m.16223C>T. Four tumor samples were reported with all six of these somatic mutations. The public germline database mtDB¹⁸ (<http://www.genpat.uu.se/mtDB>) contains 623 sequences (23% of the database) with all six of these polymorphisms. These sequences are from African, European, Asian, Southeast Asian, Amerindian, Australian and Melanesian mtDNAs. Haplogroups D, G, M7 (Japan), E (Taiwan), L (Africa) and M (India) are reported to have these polymorphisms as well. Given the improbability that all six of these mutations would arise through somatic mutation on one mtDNA chromosome, it is most likely that these mutations, and perhaps others, are due to contamination by germline mtDNAs during sequencing.

Is the pattern of mutation in tumor mtDNA mutations more consistent with neutrality or selection?

Under pure neutrality, the rate of amino acid changes should be approximately equal to the rate of silent changes, and the ratio of missense changes per such site to the number of silent changes per

such site should be equal to one.²⁷ Under negative selection, the ratio is < 1; under positive selection the ratio > 1. As a test of selection, we compared the numbers of observed changes relative to the number expected in germline and cancer mutations (see Table 3). For determining the mutation count among inherited germline sequences, we used the method of Ingman and Gyllensten,¹⁹ which uses a majority rule consensus sequence by which to compare observed sequences. This avoids the use of pairwise comparisons between samples, which at the intraspecific level frequently results in zero synonymous differences and prohibits tests of neutrality based on the missense to silent ratio. Furthermore, tumor mutations represent a somatic event arising from the patient's inherited mtDNA, and thus it is inappropriate to count the number of missense and silent mutations by phylogenetic or pairwise comparisons. Each sequence difference represents an intratissue *de novo* mutation event.

Among the germline sequences, all genes show a strong deficiency of amino acid changes, indicating negative selection during human history (Table 3, left). We also observe a high rate of amino acid change for *MTATP6* relative to other mtDNA genes in the germline (Figure 1). The high population variation in *MTATP6* has been reported before and was posited as the result of positive or balancing selection or relaxed negative selection.^{19,28–30} By contrast, selection is greatly relaxed in tumor mutations relative to germline (Table 3, right). In particular, the mutation patterns of *MTATP6*, *MTCO3*, *MTND3* and *MTND4L* in tumor cells are consistent with neutrality, whereas in the remainder of the genes the deficiency of missense mutations is less severe than that seen in germline. Furthermore, comparing the ratio of amino acid to silent changes (A/S) between tumor and germline mutations reveals a greatly enhanced rate of missense changes in tumors (Figure 1). The tumor variation in *MTND3*, *MTCO3* and *MTND4L*, in particular, shows amino acid changes 9 to 18 times greater than their counterpart genes at the germline level (Figure 1 and Supplementary Table S2).

Somatic missense mutations in tumors occur more frequently among core domain residues in comparison with germline mutations

At the population level, mtDNA-encoded polypeptides exhibit amino acid variation that is inversely proportional to the number of residues comprising the core domain(s) of the protein.¹⁹ Proteins composed primarily of core domain residues have correspondingly low amino acid polymorphism in the population, and the relative lack of core

Table 3 Analysis of amino-acid changing and silent mutations^a

Gene	Germline				Cancer ^b			
	a.a. Change	Silent change	P-value ^c	Selection pattern	a.a. Change	Silent change	P-value ^c	Selection pattern
<i>MTATP6</i>	32	24	0.00242	Negative selection	27	9	0.15096	Neutral
<i>MTATP8</i>	8	11	0.00096	Negative selection	1	3	0.03836	Relaxed negative selection
<i>MTCO1</i>	17	65	0.00000	Negative selection	7	20	0.00000	Negative selection
<i>MTCO2</i>	10	35	0.00000	Negative selection	10	10	0.00872	Negative selection
<i>MTCO3</i>	12	41	0.00000	Negative selection	27	8	0.15102	Neutral
<i>MTCYB</i>	41	53	0.00000	Negative selection	28	15	0.04570	Relaxed negative selection
<i>MTD1</i>	21	48	0.00000	Negative selection	21	13	0.04841	Relaxed negative selection
<i>MTND2</i>	18	46	0.00000	Negative selection	29	25	0.00046	Negative selection
<i>MTND3</i>	5	14	0.00001	Negative selection	13	4	0.22238	Neutral
<i>MTND4</i>	15	58	0.00000	Negative selection	20	25	0.00002	Negative selection
<i>MTND4L</i>	2	14	0.00000	Negative selection	8	3	0.26177	Neutral
<i>MTND5</i>	42	82	0.00000	Negative selection	33	30	0.00004	Negative selection
<i>MTND6</i>	9	27	0.00000	Negative selection	2	6	0.00425	Negative selection

Abbreviation: a.a., amino acid.

^aMissense and silent changes are theoretically deduced.

^bExcluding putative germline contaminants in the cancer sequences.

^cDeviation from neutrality was calculated according to a modified version of Greenman *et al.*²²

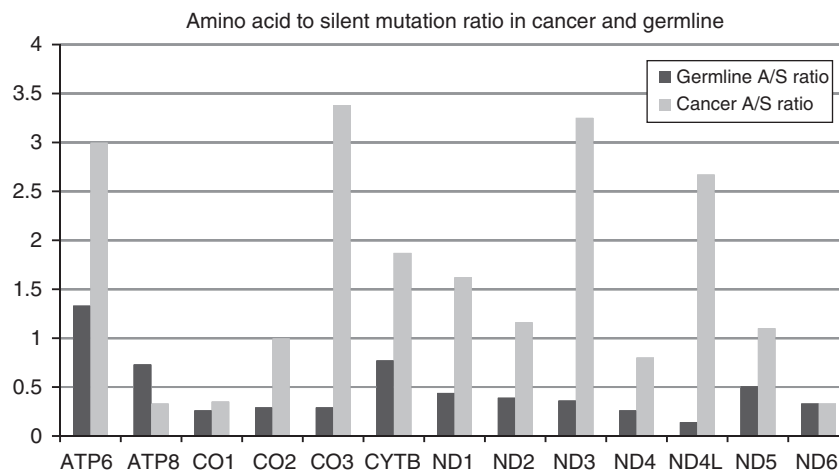


Figure 1 Amino acid to silent ratio (A/S) in cancer and germline show excess amino acid variation in cancer.

domain residues in *MTATP6* was proposed as the reason for the high population variation associated with this gene in comparison with other mtDNA genes.¹⁹ Of all mtDNA-encoded proteins, *MTATP6* has the lowest proportion of core domain residues and thus may best tolerate changes in protein sequence. We observed a similar correlation among the germline sequences analyzed in this set. For a given protein, the A/S ratio drops as the proportion of core residues increases relative to the entire polypeptide (Figure 2a; $P=0.02$). However, the correlation does not hold true at the somatic tumor level (see Figure 2b). *MTCO3* and *MTND4L*, in particular, exhibit high levels of amino acid variation, despite the composition of these proteins being nearly 100% core residues.

Are essential amino acid positions preferentially mutated in tumors?

In vitro and *in vivo* assays of mitochondrial loss-of-function mutations suggest that mitochondrial dysfunction contributes to tumor growth or causes more aggressive metastasis.^{13–16} To see if tumors preferentially mutate conserved amino acid positions, we performed bootstrap

analysis on the cancer-specific mutations. The bootstrap resamples polypeptide positions to assess whether the positions of observed amino acid mutations show a higher interspecific conservation than expected under neutral forces. A Conservation Index (CI), defined by the frequency of species sharing the same amino acid per position as found in healthy humans,²⁴ was calculated for each position. On the basis of this index, amino acid mutations of known pathology (for example, Leber's hereditary optic neuropathy, mitochondrial myopathy, Leigh's syndrome) occurred in positions with an average CI of 0.936 or 93.6% conserved.²⁴ If tumor cells are largely glycolytic, and mitochondria are free of selective constraints throughout tumor development, the somatic mutation pattern should be random. Bootstrap of polypeptides encoded by mtDNA showed significant results for the cytochrome *B* protein and mitochondrially encoded cytochrome oxidase 1 (MTCO1) (Figures 3a and b). Tumor-specific mutations in cytochrome *B* occurred in positions with an unusually high average CI of 0.973 (Figure 3a; $P=0.008$). By contrast, cancer-specific mutations in MTCO1 show unusually low conservation

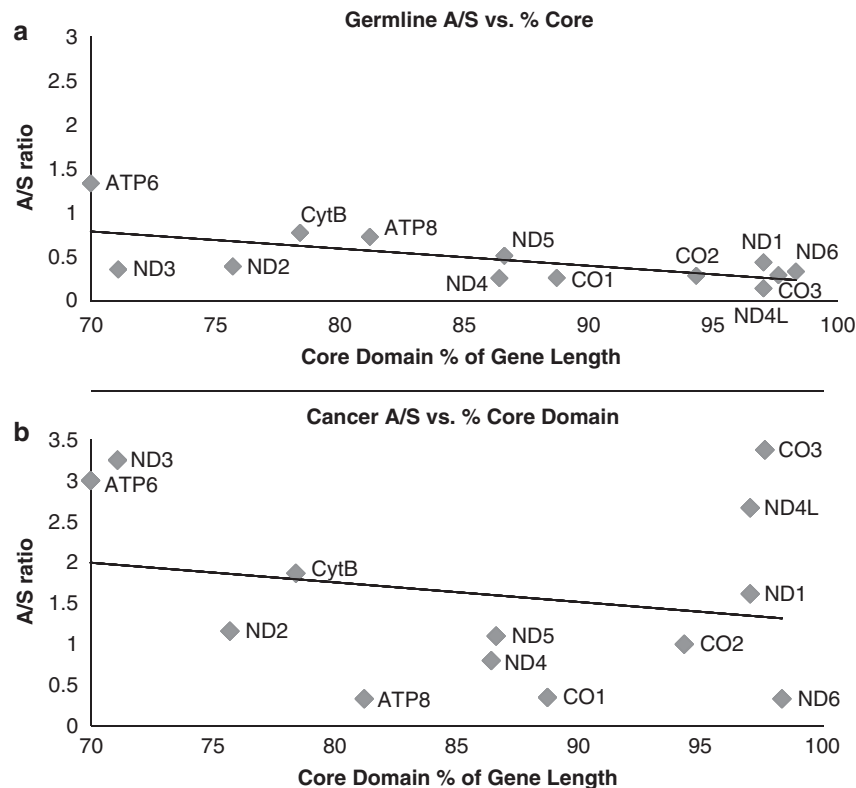


Figure 2 Regression between amino acid to silent mutation ratio (A/S) and percent gene length comprising functional protein domain. (a) Germline A/S vs % core domain. Germline polymorphisms show negative correlation. Linear regression, $R^2=0.3996$, $P=0.0204$. (b) Cancer A/S vs % core domain. Linear regression, $R^2=0.048$, $P=0.47$. Cancer mutations show a release from the constraints of protein domains. Core functional domain residues were determined as in Ingman and Gyllenstein,¹⁹ based on the Pfam annotations as reported in Ensembl v36.

indices (Figure 3b), with a mean CI of 0.565. Overall, 99.3% of the resampling trials had a higher mean CI than that observed in cancer.

DISCUSSION

mtDNA cancer mutations show relaxed negative selection and an enhanced missense mutation rate among functional domains

Although somatically acquired mutations in mtDNA are found in numerous cancer types, the function that these mutations have in cancer is still unclear. Comparing the pattern of mtDNA changes at the tumor level and in the general population may give insight on differences in the mutation and selection process. Among inherited germline sequences, all genes show a strong deficiency of amino acid changes, reflecting negative selection during human history and consistent with known constraints on mitochondria within an organism. However, in tumor cells the somatic mtDNA mutation patterns of *MTATP6*, *MTCO3*, *MTND3* and *MTND4L* are consistent with neutrality. The remainder of the genes shows negative selection, although much weaker than that observed in inherited germline sequences. Direct comparison of A/S between cancer and germline mutations reveals a greatly enhanced rate of missense changes in cancer. Thus, negative selection is greatly relaxed in cancer mutations relative to germline, suggesting that mtDNA in tumors mutate more freely than at the population level.

Ingman and Gyllenstein¹⁹ proposed that the rate of germline amino acid variation in mtDNA-encoded proteins is inversely related to how much of the polypeptide is comprised of core functional domain(s). The A/S ratio decreases as the proportion of core residues relative to the entire polypeptide increases (Figure 2a and Ingman and

Gyllenstein¹⁹). However, amino acid changes in tumors do not correlate with the functional domains of the protein. *MTCO3* and *MTND4L*, in particular, exhibit high levels of amino acid variation despite these proteins being completely composed of core residues (Figure 2b). These results suggest that functional constraints on mtDNA are relaxed within tumor cells and are consistent with the fact that metabolism in cancer is primarily glycolytic. Cell function and tumor growth can occur independently of adenosine triphosphate (ATP) production from aerobic respiration, allowing the accumulation of mitochondrial defects that would normally not be tolerated at the organismal level.

Bootstrap analysis of cancer-mutated positions

If mitochondria are free of selective constraints during tumor development, the somatic mutation pattern should be random. To test this hypothesis, we performed bootstrap analysis on the conservation indices associated with each amino acid position in mtDNA-encoded proteins. Cancer-specific amino acid mutations in cytochrome *B* have an unusually high average CI of 0.973 (Figure 3a), implying that these mutations are not under neutral forces and that late-stage tumors select for cells with impaired cytochrome *B* function. These results are consistent with *in vivo* assays in which deletion versions of cytochrome *B* were expressed in tumor cells, with a resulting increase in tumorigenicity, ROS production and expression of mitogenic factors NF- κ B, cyclin D and matrix metalloproteinase-2.¹³ Cytochrome *B*, as a subunit of Complex III in the electron transport chain, is involved in the reduction and oxidation of ubiquinone (Q). The oxidation-reduction cycle of ubiquinone (Q-cycle) is implicated in stabilization of mitogenic factor hypoxia-inducible factor1- α as a normal response to hypoxic

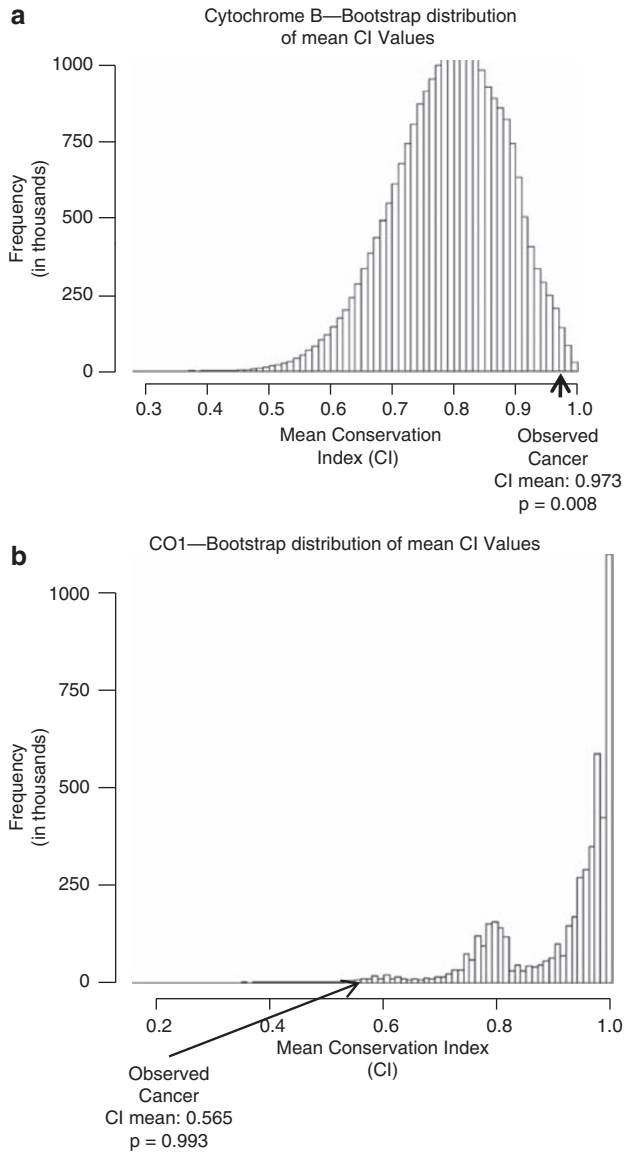


Figure 3 Bootstrap distribution of conservation indices in mtDNA-encoded polypeptides. (a) Tumor-specific mutations in cytochrome *B* occur in positions with an unusually high average conservation index (CI) of 0.973; (b) cancer-specific mutations in MTCO1 have unusually low conservation indices, with a mean CI of 0.565. Bootstrap of MTATP8, MTND3, and MTND6 could not be performed due to small sample sizes.

conditions.³¹ Missense mutations in cytochrome *B* can interfere with the Q-cycle, possibly generating higher levels of ROS and stabilizing hypoxia-inducible factor1- α . Cytochrome *B* is also implicated in the FAS-induced apoptotic pathway, in which cytochrome *B* is cleaved in half and its C-terminal portion transported into the cytoplasm to mediate cell death.^{31,32} It is noteworthy that a seven amino acid deletion in cytochrome *B* originally discovered in bladder cancer⁵ also falls within the 42 amino acid tag that interferes with the ability of endogenous cytochrome *B* to mediate FAS-induced apoptosis.³² When expressed in bladder cancer cell lines, this deletion increased tumorigenic behavior.¹³ A majority of the cancer-specific missense mutations examined in this paper also fall within the C-terminal portion of the protein.

In contrast to cytochrome *B*, MTCO1 had tumor mutations occurring in positions with unusually low conservation indices

(Figure 3b; mean CI of 0.565), despite the fact that MTCO1 is the most highly conserved mtDNA protein with an overall CI of 0.91. Greater than 99% of the resampling trials had a higher mean CI than that observed for cancer positions, implying that tumors select or tolerate MTCO1 mutations occurring only in poorly conserved residues. It may be that tumors are not wholly independent of Complex IV activity, or that MTCO1 function is necessary for tumor progression. To date, an increase in tumorigenic properties has not been found in functional assays for pathogenic MTCO1 mutations, including a stop codon mutation (G6930A) that was tested for increased ROS production and oxidative damage.³³ MTCO1 transcripts also show an increased level of expression in rat hepatoma relative to normal cells.³⁴

The role of somatic mtDNA mutations in cancer

Our data on somatic mtDNA cancer mutations show relaxed negative selection, release of missense mutations from protein domain constraints and nonrandom occurrence of amino acid changes in cytochrome *B* and cytochrome oxidase 1. As tumor cells synthesize ATP primarily through glycolysis, they are less dependent upon the electron transport chain (ETC), creating a selectively neutral environment for mtDNA mutations. Mutations impairing ETC function should frequently arise, as these can occur at multiple points along the chain or in the diverse subunits of a single complex. Relaxed negative selection on mtDNA may allow mutations not seen at the population level to accumulate within cancerous or precancerous cells. It is possible that key mutations impairing mitochondrial function may then contribute to increased mitosis or metastasis within the tumor.

It remains a puzzle as to why deleterious mtDNA mutations can cause mitochondrial pathology such as MERRE, MELAS, Leigh's syndrome and LHON when maternally inherited, but these same mutations do not cause heritable predisposition to cancer. The lack of such cancer susceptibility presumably points to somatic cancer mtDNA mutations being only a result of the mutagenic environment in tumor tissue rather than having a causal function. However, mtDNA mutations of known pathology can be transformed into cancer cell lines and result in greatly enhanced (for example, >7x) tumor growth of the recipient cells *in vivo*.^{14,15,35} Similarly, over-expression of a 21-bp deletion in cytochrome *B* discovered in bladder cancer induced greater tumor growth *in vivo* and upregulated cell cycle progression genes *NF- κ B2*, *cyclin D1* and *matrix metalloproteinase-2*.¹³ Transformation of tumor-associated mutations into cancer cell lines also resulted in enhanced foci formation and generation of ROS in culture.^{13,15} And when mitochondria from tumors of high or low metastatic potential were transformed into tumor cell lines, the recipient cells took on the metastatic behavior (aggressive or nonaggressive) of the transferred mitochondria despite a differing nuclear background.¹⁶ This showed that specific cancer mutations in mtDNA contribute to aggressive tumor behavior. Furthermore, nuclear genes encoding ETC Complex II—SDHB, SDHC and SDHD—and matrix enzyme fumarate hydratase are dominantly inherited tumor suppressor genes.² Heterozygous germline mutations in these genes cause hereditary paragangliomas and pheochromocytomas, and leiomyotomas and renal clear cell carcinomas, respectively. Surprisingly, mutations in the first subunit of Complex II, SDHA, do not predispose to cancer when heterozygous but only cause Leigh's syndrome in homozygotes. Thus, there is a clear role for mitochondrial dysfunction in tumor formation, but the basis for why some mutations cause mitochondrial myopathy while other mutations lead to cancer remains a mystery.

One possibility is that some mutations in the ETC cause dysfunction in the citric acid cycle without causing a cancer-promoting increase in ROS, whereas the converse applies for other ETC mutations. Metabolic dysfunction in the citric acid cycle or ATP production leads to neuromuscular degeneration and other symptoms associated with the mitochondrial encephalomyopathy syndromes, whereas intermediate increases in ROS lead to apoptosis resistance and stabilization of mitogenic/angiogenic factors. The *in vitro* suppression of SDHA and SDHB has shown just such a dichotomy between different subunits of the same ETC complex, mirroring the myopathy and cancer phenotypes associated with these genes, respectively.³⁶ SDHA and SDHB were suppressed using small interfering RNA and pharmacological agents, showing that loss of SDHB—but not SDHA—causes an increase of ROS levels, tumor growth and stabilization of mitogenic/angiogenic factor such as hypoxia-inducible factor1- α .³⁶ Few other studies provide a model so clearly distinguishing between molecular mechanisms leading to mitochondrial encephalomyopathy vs tumor promotion.

It is likely that the diversity of phenotypes observed with mutations in different subunits of Complex II—SDH extends to other ETC complexes, and may account for the contrasting mutation patterns observed in this study. Genes of Complex IV showed such diversity—the somatic cancer pattern of MTCO3 showed a high amino acid mutation rate, whereas the low cancer missense mutation rate of MTCO1 and bootstrap resampling implies that tumors do not tolerate deleterious mutations in MTCO1. Likewise, two of the subunits of Complex I—MTND3 and MTND4L—show high levels of cancer missense mutations, whereas the remaining Complex I subunits show negative selection.

Common haplogroup lineages found among reported somatic mtDNA mutations

Common germline polymorphisms m.8701A>G, m.9540T>C, m.10398A>G, m.12705C>T, m.15301G>A and m.16223C>T were found at high frequency in the somatic tumor mutations and comprised somatic mutation haplotypes among sequences from individual tumors. Because of the prevalence of this haplotype in the population, these tumor samples were assumed to be contaminated from an external source and excluded from the analysis. It has been posited that the occurrence of common population haplotypes—including the mutations listed above—in somatic cancer mutations and at nodal (haplogroup-defining) positions in the human phylogeny must be due to positive selection, indicative of similar selective forces at both the tumor and population level.³⁷ We feel that this scenario should take additional factors into account. The metabolism of cancer is primarily glycolytic, contrary to a cell environment that would select for normal or more efficient mitochondrial function. Mitochondria are dysfunctional in a large majority of tumors, prompting the theory that dysfunction in mitochondria gives a proliferative advantage to tumor cells. This is supported by more aggressive tumor behavior upon transfection of tumor cell lines with pathogenic mtDNA mutants that are at low frequency in the population. Selection for mtDNA alleles at the population level would be under the opposite constraints, where humans require functional mitochondria for ATP production. Haplogroup-defining mutations that are selected for generation after generation must be at least neutral or confer an ATP production advantage. Furthermore, the theory that mutations in nodal positions of the human phylogeny also have a selective function when occurring as a somatic mutation is based on only two large surveys of mtDNA cancer mutations from the same facility. It has been reported that these somatic common polymorphisms are detected in

tumors by multiple technicians (reviewed by Brandon *et al.*³⁸). It may be that two classes of mtDNA mutations—common population polymorphisms and low frequency deleterious mutations—are involved in tumor formation as suggested by Brandon *et al.*³⁸ However, it remains to be clearly proven that haplogroup-defining polymorphisms detected as linked somatic cancer mutations are found by multiple investigators and are not due to an external source.

Overall, results support relaxed negative selection or neutrality in the somatic mtDNA mutations found in cancer. The nonrandom occurrence of somatic cancer mutations in highly conserved residues of cytochrome *B* also suggests that mutations impairing mitochondrial function may confer a mitogenic advantage in the changing cancer microenvironment. This scenario is consistent with the ability of pathological mtDNA cybrids to increase the tumorigenicity of recipient tumor lines, as well as the occurrence of tumors with no obvious mitochondrial dysfunction which still metabolize under aerobic glycolysis. This data set still represents a small number of tumors and tumor-associated mutations. In the very near future, microarray analysis may provide the high volume of sequence data necessary for distinguishing the function(s) of mtDNA mutation in carcinogenesis.

ACKNOWLEDGEMENTS

We thank K Callahan for additional bootstrap analysis and T McElroy for helpful comments about the paper.

- Warburg, O. (ed). Ueber den Stoffwechsel der Tumoren (The metabolism of tumours; investigations from the Kaiser Wilhelm institute for biology, Berlin-Dahlem). *J. Am. Med. Assoc.* **96**, 1982–2309 (1931).
- Eng, C., Kiuru, M., Fernandez, M. J. & Aaltonen, L. A. A role for mitochondrial enzymes in inherited neoplasia and beyond. *Nat. Rev. Cancer* **3**, 193–202 (2003).
- Weinberg, R. A. *The Biology of Cancer* (Garland Science, Taylor and Francis Group LLC, New York, NY, 2007).
- Polyak, K., Li, Y., Zhu, H., Lengauer, C., Willson, J. K., Markowitz, S. D. *et al.* Somatic mutations of the mitochondrial genome in human colorectal tumours. *Nat. Genet.* **20**, 291–293 (1998).
- Fliss, M. S., Usadel, H., Caballero, O. L., Wu, L., Buta, M. R., Eleff, S. M. *et al.* Facile detection of mitochondrial DNA mutations in tumors and bodily fluids. *Science* **287**, 2017–2019 (2000).
- Jeronimo, C., Nomoto, S., Caballero, O. L., Usadel, H., Henrique, R., Varzim, G. *et al.* Mitochondrial mutations in early stage prostate cancer and bodily fluids. *Oncogene* **20**, 5195–5198 (2001).
- Jones, J. B., Song, J. J., Hempen, P. M., Parmigiani, G., Hruban, R. H., Kern, S. E. *et al.* Detection of mitochondrial DNA mutations in pancreatic cancer offers a 'mass'-ive advantage over detection of nuclear DNA mutations. *Cancer Res.* **61**, 1299–1304 (2001).
- Liu, V. W., Shi, H. H., Cheung, A. N., Chiu, P. M., Leung, T. W., Nagley, P. *et al.* High incidence of somatic mitochondrial DNA mutations in human ovarian carcinomas. *Cancer Res.* **61**, 5998–6001 (2001).
- Maximo, V., Soares, P., Lima, J., Cameselle-Teijeiro, J. & Sobrinho-Simoes, M. Mitochondrial DNA somatic mutations (point mutations and large deletions) and mitochondrial DNA variants in human thyroid pathology: a study with emphasis on Hurthle cell tumors. *Am. J. Pathol.* **160**, 1857–1865 (2002).
- Parrella, P., Xiao, Y., Fliss, M., Sanchez-Cespedes, M., Mazzarelli, P., Rinaldi, M. *et al.* Detection of mitochondrial DNA mutations in primary breast cancer and fine-needle aspirates. *Cancer Res.* **61**, 7623–7626 (2001).
- Gatenby, R. A. & Gillies, R. J. Why do cancers have high aerobic glycolysis? *Nat. Rev. Cancer* **4**, 891–899 (2004).
- Bonnet, S., Archer, S. L., Allalunis-Turner, J., Haromy, A., Beaulieu, C., Thompson, R. *et al.* A mitochondria-K⁺ channel axis is suppressed in cancer and its normalization promotes apoptosis and inhibits cancer growth. *Cancer Cell* **11**, 37–51 (2007).
- Dasgupta, S., Hoque, M. O., Upadhyay, S. & Sidransky, D. Mitochondrial cytochrome B gene mutation promotes tumor growth in bladder cancer. *Cancer Res.* **68**, 700–706 (2008).
- Petros, J. A., Baumann, A. K., Ruiz-Pesini, E., Amin, M. B., Sun, C. Q., Hall, J. *et al.* mtDNA mutations increase tumorigenicity in prostate cancer. *Proc. Natl Acad. Sci. USA* **102**, 719–724 (2005).
- Zhou, S., Kachhap, S., Sun, W., Wu, G., Chuang, A., Poeta, L. *et al.* Frequency and phenotypic implications of mitochondrial DNA mutations in human squamous cell cancers of the head and neck. *Proc. Natl Acad. Sci. USA* **104**, 7540–7545 (2007).

- 16 Ishikawa, K., Takenaga, K., Akimoto, M., Koshikawa, N., Yamaguchi, A., Imanishi, H. *et al.* ROS-generating mitochondrial DNA mutations can regulate tumor cell metastasis. *Science* **320**, 661–664 (2008).
- 17 Sawyer, S. L., Wu, L. I., Emerman, M. & Malik, H. S. Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proc. Natl Acad. Sci. USA* **102**, 2832–2837 (2005).
- 18 Ingman, M. & Gyllensten, U. mtDB: human mitochondrial genome database, a resource for population genetics and medical sciences. *Nucleic Acids Res.* **34**, D749–751 (2006).
- 19 Ingman, M. & Gyllensten, U. Rate variation between mitochondrial domains and adaptive evolution in humans. *Hum. Mol. Genet.* **16**, 2281–2287 (2007).
- 20 Kogelnik, A. M., Lott, M. T., Brown, M. D., Navathe, S. B. & Wallace, D. C. MITOMAP: a human mitochondrial genome database. *Nucleic Acids Res.* **24**, 177–179 (1996).
- 21 Zsurka, G. & Csordás, A. MitoWheel, visualizing the human mitochondrial genome (MitoWheel 1.2). in *Nature Preceedings*. [dx.doi.org/10.1038/npre.2009.3167.1](https://doi.org/10.1038/npre.2009.3167.1). (Nature Publishing Group, Cambridge, 2009).
- 22 Greenman, C., Wooster, R., Futreal, P. A., Stratton, M. R. & Easton, D. F. Statistical analysis of pathogenicity of somatic mutations in cancer. *Genetics* **173**, 2187–2198 (2006).
- 23 Rand, D. M. & Kann, L. M. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol. Biol. Evol.* **13**, 735–748 (1996).
- 24 Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V. & Wallace, D. C. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* **303**, 223–226 (2004).
- 25 Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
- 26 Poirot, O., O'Toole, E. & Notredame, C. tcoffee@igs: a web server for computing, evaluating and combining multiple sequence alignments. *Nucleic Acids Res.* **31**, 3503–3506 (2003).
- 27 Hughes, A. L. & Nei, M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**, 167–170 (1988).
- 28 Elson, J. L., Turnbull, D. M. & Howell, N. Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. *Am. J. Hum. Genet.* **74**, 229–238 (2004).
- 29 Mishmar, D., Ruiz-Pesini, E., Brandon, M. & Wallace, D. C. Mitochondrial DNA-like sequences in the nucleus (NUMTs): insights into our African origins and the mechanism of foreign DNA integration. *Hum. Mutat.* **23**, 125–133 (2004).
- 30 Sun, C., Kong, Q. P. & Zhang, Y. P. The role of climate in human mitochondrial DNA evolution: a reappraisal. *Genomics* **89**, 338–342 (2007).
- 31 Klimova, T. & Chandel, N. S. Mitochondrial complex III regulates hypoxic activation of HIF. *Cell Death Differ.* **15**, 660–666 (2008).
- 32 Komarov, A. P., Rokhlin, O. W., Yu, C. A. & Gudkov, A. V. Functional genetic screening reveals the role of mitochondrial cytochrome *b* as a mediator of FAS-induced apoptosis. *Proc. Natl Acad. Sci. USA* **105**, 14453–14458 (2008).
- 33 Vives-Bauza, C., Gonzalo, R., Manfredi, G., Garcia-Arumi, E. & Andreu, A. L. Enhanced ROS production and antioxidant defenses in cybrids harbouring mutations in mtDNA. *Neurosci. Lett.* **391**, 136–141 (2006).
- 34 Luciakova, K. & Kuzela, S. Increased steady-state levels of several mitochondrial and nuclear gene transcripts in rat hepatoma with a low content of mitochondria. *Eur. J. Biochem.* **205**, 1187–1193 (1992).
- 35 Shidara, Y., Yamagata, K., Kanamori, T., Nakano, K., Kwong, J. Q., Manfredi, G. *et al.* Positive contribution of pathogenic mutations in the mitochondrial genome to the promotion of cancer by prevention from apoptosis. *Cancer Res.* **65**, 1655–1663 (2005).
- 36 Guzy, R. D., Sharma, B., Bell, E., Chandel, N. S. & Schumacker, P. T. Loss of the SdhB, but not the SdhA, subunit of Complex II triggers reactive oxygen species-dependent hypoxia-inducible factor activation and tumorigenesis. *Mol. Cell Biol.* **28**, 781–731 (2008).
- 37 Zhidkov, I., Livneh, E. A., Rubin, E. & Mishmar, D. MtDNA mutation pattern in tumors and human evolution are shaped by similar selective constraints. *Genome Res.* **19**, 576–580 (2009).
- 38 Brandon, M., Baldi, P. & Wallace, D. C. Mitochondrial mutations in cancer. *Oncogene* **25**, 4647–4662 (2006).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)