Oncogene (2006) 25, 4647–4662

www.nature.com/onc

REVIEW

Mitochondrial mutations in cancer

M Brandon[1,2], P Baldi[1,2] and DC Wallace[1,3]

[1]Center for Molecular and Mitochondrial Medicine and Genetics (MAMMAG) and Institute for Genomics and Bioinformatics, University of California at Irvine, Irvine, CA, USA; [2]Department of Computer Science, University of California at Irvine, Irvine, CA, USA and [3]Departments of Biological Chemistry, Ecology and Evolutionary Biology, and Pediatrics, University of California at Irvine, Irvine, CA, USA

The metabolism of solid tumors is associated with high lactate production while growing in oxygen (aerobic glycolysis) suggesting that tumors may have defects in mitochondrial function. The mitochondria produce cellular energy by oxidative phosphorylation (OXPHOS), generate reactive oxygen species (ROS) as a by-product, and regulate apoptosis via the mitochondrial permeability transition pore (mtPTP). The mitochondria are assembled from both nuclear DNA (nDNA) and mitochondrial DNA (mtDNA) genes. The mtDNA codes for 37 genes essential of OXPHOS, is present in thousands of copies per cell, and has a very high mutations rate. In humans, severe mtDNA mutations result in multisystem disease, while some functional population-specific polymorphisms appear to have permitted humans to adapt to new environments. Mutations in the nDNA-encoded mitochondrial genes for fumarate hydratase and succinate dehydrogenase have been linked to uterine leiomyomas and paragangliomas, and cancer cells have been shown to induce hexokinase II which harnesses OXPHOS adenosine triphosphate (ATP) production to drive glycolysis. Germline mtDNA mutations at nucleotides 10398 and 16189 have been associated with breast cancer and endometrial cancer. Tumor mtDNA somatic mutations range from severe insertion–deletion and chain termination mutations to mild missense mutations. Surprisingly, of the 190 tumor-specific somatic mtDNA mutations reported, 72% are also mtDNA sequence variants found in the general population. These include 52% of the tumor somatic mRNA missense mutations, 83% of the tRNA mutations, 38% of the rRNA mutations, and 85% of the control region mutations. Some associations might reflect mtDNA sequencing errors, but analysis of several of the tumor-specific somatic missense mutations with population counterparts appear legitimate. Therefore, mtDNA mutations in tumors may fall into two main classes: (1) severe mutations that inhibit OXPHOS, increase ROS production and promote tumor cell proliferation and (2) milder mutations that may permit tumors to adapt to new environments. The former may be lost during subsequent tumor oxygenation while the latter may become fixed. Hence, mitochondrial dysfunction does appear to be a factor in cancer etiology, an insight that may suggest new approaches for diagnosis and treatment.
Oncogene (2006) **25,** 4647–4662. doi:10.1038/sj.onc.1209607

Keywords: mtDNA; cancer; OXPHOS; adaptation; Warburg

Correspondence: Dr DC Wallace, Center for Molecular and Mitochondrial Medicine and Genetics (MAMMAG), Hewitt Hall, Room 2014, University of California, Irvine, CA 92697-3940, USA.
E-mail: dwallace@uci.edu

Cancer cell 'aerobic glycolysis'

Over 70 years ago Otto Warburg observed that cancer cells actively metabolize glucose and produce excessive lactic acid while at the same time consuming oxygen via mitochondrial oxidative phosphorylation (OXPHOS), a phenomenon that he called 'aerobic glycolysis' (Warburg, 1931, 1956). Since aerobic glycolysis was contrary to the normal 'Pasteur effect' Warburg proposed that cancer could be the result of mitochondrial defects.

The mitochondria produce most of the cellular energy, generate much of the endogenous reactive oxygen species (ROS), and regulate programmed cell death (apoptosis) via the mitochondrial permeability transition pore (mtPTP). The mtPTP is assembled from the pro- and anti-apoptotic members of the Bax and Bcl2 gene family; the mitochondrial outer membrane voltage-dependent anion ion channel (VDAC); the mitochondrial inner membrane adenine nucleotide (nt) translocator (ANT); cyclophilin D and the benzodiazepine receptor (Wallace, 2005).

Warburg's observation stimulated many investigators to analyse mitochondrial function in tumor cells. This revealed that mitochondrial number and OXPHOS activities were frequently downregulated in many cancers (Pedersen, 1978) and that mRNA levels for certain mitochondrial DNA (mtDNA)-encoded genes were upregulated (Glaichenhaus *et al*., 1986; LaBiche *et al*., 1988, 1992). However, the first significant insight into the cause of 'aerobic glycolysis' came from the discovery that hepatic tumors (hepatomas) express hexokinase II while normal hepatic cells express hexokinase IV (glucokinase). Hexokinase IV, also known as glucokinase, is a 50 kDa enzyme with one catalytic site and a high K_m ($K_m = 5$–8 mM). Hexokinases I–III, by contrast, are 100 kilodalton (kDa) proteins that duplicate the basic hexokinase IV unit and have much lower K_ms ($K_m = 0.02$–0.03 mM).

Moreover, hexokinase II, unlike hexokinases I and III, has both catalytic sites active (Wallace, 2005).

Hexokinase II can directly bind to the mitochondrial outer membrane protein VDAC (Bustamante and Pedersen, 1977). Voltage-dependent anion ion channel is the pore in the mitochondrial outer membrane through which all solutes including the adenosine triphosphate (ATP) pass. Mitochondrial ATP is exported through the mitochondrial inner membrane by the ANT in exchange for spent cytosolic ADP. Hence, VDAC-bound hexokinase II can directly bind and utilize the mitochondrial ATP to rapidly phosphorylate glucose to glucose-6-phosphate. Thus cancer cells harness mitochondrial OXPHOS to drive glycolysis to make ATP under low-oxygen conditions and as a result generate excessive lactate (Bustamante and Pedersen, 1977; Wallace, 2005). Consequently, aerobic glycolysis is an adaptive strategy that permits the cancer cell to grow in hypoxic environments.

The binding of hexokinase II to VDAC also stabilizes the mtPTP and inhibits induction of apoptosis (Pedersen *et al*., 2002; Gatenby and Gillies, 2004) by antagonizing the pro-apoptotic action of Bax and Bak (Gottlob *et al*., 2001; Birnbaum, 2004; Majewski *et al*., 2004a, b). The interaction of hexokinase II and VDAC is promoted by phosphoryation of hexokinase II by the *Akt* kinases, also known as protein kinase D (PKB). The *Akt* kinases, in turn, mediate the signals of a variety of oncogenes (Aoki *et al*., 1998; Testa and Bellacosa, 2001; Brazil *et al*., 2004; Yanagihara *et al*., 2005). Therefore, the induction of hexokinase II and its binding to VDAC facilitates growth under low-oxygen tension by switching the cellular metabolism from oxidative to glycolytic and by inhibiting apoptosis through blocking Bax and Bak.

Mitochondrial bioenergetics

Mitochondrial OXPHOS encompasses respiratory complexes I–IV of the electron transport chain (ETC) plus complex V, the ATP synthase. Energy generated by OXPHOS is used to synthesize ATP to drive cellular work, to produce heat to maintain body temperature, etc. Electrons from the carbohydrates and fats are transferred via nicotinamide adenine dinucleotide (reduced form) (NADH) + H^+ to complex I (NADH dehydrogenase or NADH:ubiquinone oxidoreductase) or from succinate to complex II (succinate dehydrogenase or succinate:ubiquinone oxidoreductase). Each pair of electrons is then used to reduce ubiquinone to ubisemiquinone and then to ubiquinol. Ubiquinol then transfers the electrons to complex III (the bc_1 complex or ubiquinone:cytocrome *c* oxidoreductase), after which the electrons flow through cytochrome *c* (cyt*c*), to complex IV (cyt*c* oxidase, cyclooxygenase or reduced cyt*c*:oxygen oxidoreductase) and finally to $\frac{1}{2}$ O_2 to generate H_2O. The energy that is released by the ETC is used to pump protons out cross the mitochondrial inner membrane through complexes I, III and IV. This creates an electrochemical gradient ($\Delta P = \Delta\Psi + \Delta pH$) that is positive and acid on the outside and negative and alkaline on the inside. These high capacitance protons can then flow through the ATP synthase proton channel, driving the condensation of bound ADP and Pi to generate ATP. The mitochondrial matrix ATP is then exchanged for cytosolic ADP by the ANTs.

The efficiency by which the ETC pumps protons out of the mitochondrial matrix and by which the ATP synthase converts this proton gradient into ATP is known as the coupling efficiency. Tightly coupled mitochondria make more ATP and less heat while loosely coupled mitochondria make more heat and less ATP.

Mitochondrial OXPHOS also produces most of the cellular ROS at complexes I and III, by the transfer of an unpaired electron to O_2 to generate superoxide ($O_2^{\bullet-}$). Superoxide, a potent oxidizing agent, is detoxified by the mitochondrial matrix manganese superoxide dismutase (MnSOD) to generate hydrogen peroxide (H_2O_2). The relatively stable H_2O_2 persists in the mitochondrial matrix and can diffuse into the cytosol and the nucleus. However, in the presence of reduced transition metals, H_2O_2 can acquire an additional electron, generating the highly reactive hydroxyl radical ($^{\bullet}OH$). In the mitochondria, H_2O_2 is slowly reduced to water by glutathione peroxidase while in the cytosol H_2O_2 can be removed by the peroxisomal enzyme catalase.

Inhibition of electron flux through the ETC causes electrons to be retained on complex I–III electron carriers. This makes the electrons available to be transferred directly to O_2 to generate $O_2^{\bullet-}$. Mitochondrial ROS can then damage mitochondrial enzymes, lipids and mutagenize the mtDNA.

Chronically elevated oxidative stress, reduced ΔP, reduced adenine nts and increased matrix Ca^{++} can all activate the mtPTP. This opens a channel through the mitochondrial inner membrane that leads to the collapse of ΔP, swelling of the mitochondria, and release of apopototic proteins stored in the mitochondrial intermembrane space, specifically cyt*c*, procaspase-9, apoptosis initiating factor, and endonuclease G. These degrade the proteins and nucleic acids of the cell resulting in apoptosis (Wallace, 2005).

Nuclear encoded mitochondrial gene mutations cause cancer

While the expression of hexokinase II explains why cancer cells make excess lactate in the presence of oxygen, it does not demonstrate that mitochondrial dysfunction is causally related to neoplastic transformation. However, genetic mapping and cloning of chromosomal 'oncogenes' revealed that some cancers are caused by mutations in nuclear-encoded mitochondrial enzyme genes. Uterine leiomyomas and renal cell carcinomas have been linked to mutations in fumarate hydratase (Lehtonen *et al*., 2004). Paraganglions have been linked to mutations in three of the four subunits of

succinate dehydrogenase (SDH) (Baysal *et al*., 2000; Niemann and Muller, 2000; Astuti *et al*., 2001; Vanharanta *et al*., 2004).

SDH (complex II) is assembled from four nuclear DNA (nDNA)-encoded subunits, SDH A, B, C and D. SDHA which contains a flavin adenine dinucleotide collects electrons from succinate in the tricarboxylic acid cycle and passes the electrons to the iron-sulfur components in SDHB. Succinate dehydrogenase B then transfers the electrons to the cytochrome *b* (cyt*b*) and ubiquinone associated components SDH C and SDH D. Mutations in SDH B (Astuti *et al*., 2001; Vanharanta *et al*., 2004), SDH C (Baysal *et al*., 2000), and SDH D (Niemann and Muller, 2000) all cause the cancer paraganglioma. However, mutations in SDH A cause the lethal pediatric neurodegenerative disease Leigh's syndrome (Bourgeron *et al*., 1995), not paraganglioma.

In *Caenorhabditis elegans,* inactivation of the cyt*b* subunit of complex II (the *mev-1* mutant) markedly increases mitochondrial ROS production in association with a reduction in life span (Ishii *et al*., 1998; Senoo-Matsuda *et al*., 2001). It follows then that mutations in the SDH B, C and D subunits would also result in increased mitochondrial ROS production by retarding the electron flow out of complex II. Mutations in SDH A, on the other hand, would block the entry of electrons into complex II, thus eiminating ROS production. However, this would also diminish mitochondrial ATP production resulting in neurodegenerative disease (Wallace, 2005). Thus, mitochondrial ROS production appears to be the factor that links mitocondrial defects to neoplastic transformation.

Mitochondrial DNA variation in disease and climatic adaptation

The mitochondrial genome consists of approximately 1500 genes, 37 encoded by the 16 569 nt mtDNA and the remainder located in the nDNA. The mtDNA genes include a 12S and 16S rRNA, 22 tRNAs, and 13 essential OXPHOS polypeptide subunits (Wallace and Lott, 2002). The 13 mtDNA polypeptides include seven (ND1, 2, 3, 4, 4L, 5, 6) of the 46 subunits of complex I, none of the four subunits of complex II, one (cyt*b*) of the 11 subunits of complex III, three (COI, II, III) of the 13 subunits of complex IV, and two (ATP6 and 8) of the 16 subunits of complex V. The mtDNA also contains an approximately 1000 nt control region which encompasses the promoter for transcribing the G-rich heavy (H) strand (P_H) and the adjacent promoter to transcribe the C-rich light (L) strand (P_L), the intervening mitochondrial transcription factor (mtTFA)-binding sites, three conserved sequence boxes (CSB I–III), the origin of H-strand replication, and the termination-associated sequence (TAS). The replication of the mtDNA H-strand has been proposed to start at O_H using a cleaved transcript from P_H as the primer and pausing at TAS. This creates the 7S DNA which forms a triple-stranded displacement (D)-loop. The origin of L-strand origin of replication (O_H) is located 2/3 around the mtDNA circle.

The nDNA encodes most of the mitochondrial polypeptides including the mtDNA polymerase γ, the mtRNA polymerase, the mitochondrial ribosomal proteins, the enzymes of mitochondrial metabolism, etc. The genes for these proteins are replicated and transcribed in the nucleus, the mRNAs are translated into protein on cytosolic ribosomes, and the mitochondrial proteins selectively imported into the mitochondrion using specific import peptide sequences (Wallace and Lott, 2002).

Each cell contains hundreds of mitochondria and thousands of mtDNAs. The mitochondria and mtDNAs are transmitted through the oocyte and thus are exclusively maternally inherited. A mutation in one of the mtDNAs of the cell results in an intracellular mixture of mutant and normal mtDNAs, a state known as heteroplasmy. With repeated cell divisions the percentage of mutant and normal mtDNAs can drift (replicative segregation) until it reaches either pure mutant or normal homoplasmy.

Mutations in the mtDNA can be either deleterious (pathogenic), neutral or beneficial (adaptive). In the germ line, mtDNA variants generally must reach a relatively high percentage of heteroplasmy before they affect the cellular phenotype. Within post-mitotic cells somatic mtDNA mutations accumulate over time, eroding cellular function (Trifunovic *et al*., 2004; Kujoth *et al*., 2005; Schriner *et al*., 2005). Ultimately, this activates the mtPTP and the cell dies by apoptosis. As a result, the accumulation of mtDNA mutations acts as the aging clock (Wallace, 2005). It is the interaction between the partial defects resulting from inherited mitochondrial defects plus the accumulation of somatic mtDNA mutations which exacerbates the defect that account for the delayed-onset and progressive course of mitochondria-associated diseases (Wallace, 2005).

Since the mtDNA is exclusively maternally inherited, mtDNAs between different lineages cannot mix or recombine. Therefore, the mtDNA sequence can only change by the sequential accumulation of mtDNA mutations along radiating female lineages. As a result, the human mtDNA tree is in a sequential mutational tree that coalesces to a single mtDNA in Africa about 200 000 years before present (Johnson *et al*., 1983; Cann *et al*., 1987; Merriwether *et al*., 1991). Moreover, the branches of the mtDNA tree show striking regional specificity. It is now believed that this is because about 25% of the mtDNA protein sequence variation (Mishmar *et al*., 2003; Wallace *et al*., 2003; Ruiz-Pesini *et al*., 2004), 10 to 20% of the tRNA variation, and at least some of the rRNA variation (Ruiz-Pesini and Wallace, 2006) altered the mitochondrial coupling efficiency. This increases heat production at the expense of ATP production and permitted humans to adapt to the colder climates of Eurasia. When an adaptive mutation arises in the mtDNA, that mtDNA becomes enriched in the population by positive selection. Since the mtDNA does not recombine, all of the sequence variants linked to the adaptive mutation are also enriched, a

phenomenon known as 'hitchhiking.' The collection of linked variants on an individual mtDNA molecule is called a mtDNA haplotype, and the descendants of an original mutant mtDNA will acquire additional variants to generate a group of related haplotypes, referred to as a haplogroup. These haplogroups frequently constitute region-specific branches of the mtDNA tree.

The greatest mtDNA diversity is found in Africa (Johnson *et al*., 1983; Cann *et al*., 1987; Merriwether *et al*., 1991; Mishmar *et al*., 2003) indicating that Africa was the origin of humans. The most ancient African branches of the mtDNA tree (haplogroups L0, L1 and L2) are defined by a T at nt 3594. This nt becomes a C when African L3 diverges from L2. Many L0–L2 haplotypes also harbor a $tRNA^{Asp}$ G to A variant at nt 7521. African L3 gave rise to haplogroups M and N, the only two lineages that left Africa to colonize all of Eurasia about 65 000 YBP. Macro-haplogroup N was derived from haplogroup L3 and differs by two missense mutations, one in ND3 at nt 10398 which is G in African L3 and A in macro-haplogroup N causing an A114T amino acid substitution and the other in ATP6 at nt 8701 which is a G in African L3 that changes to an A in macro-haplogroup N resulting in an A59T amino-acid change. In Europe, macro-haplogroup N gave rise to the European-specific haplogroups H, I, J, Uk, T, U, V, W and X, each associated with distinctive polymorphisms. In haplogroups J1, J2 and Uk the ND3 nt 10 398A of macro-haplogroup N reverted back to a G and codon 114 returned to an A. Haplogroup U radiated from macro-haplogroup N by the acquisition of a $tRNA^{Leu(CUN)}$ nt 12308G variant and the haplogroup subsequently was subdivided by the appearance of a 16S rRNA variant at nt 1811 A to G on one sub-branch and a COIII missense mutation at nt 9477 G to A on another sub-branch. Similarly, haplogroup J was founded by the appearance of a ND5 polymorphism at nt 13708A (A458T). Asia was colonized by derivatives of both macro-haplogroups M and N which gave rise to a plethora of Asian haplogroups including A, B, F, etc. from N and C, D, G, etc. from M. Sublineages of Asian haplogroup A and D also acquired the 13708 G to A variant in ND5, an example of the repeated appearance of the same mutation on different mtDNA lineages indicated convergent evolution by adaptive selection. Of all of the mtDNA lineages in Central Asia, only three lineages (A, C and D) became enriched in Northeastern Siberia by 20 000–30 000 years ago and these were in a position to colonize the Americas when the Bering Land Bridge became exposed. Similarly, the northern European haplogroup X arrived in north-central North America about 15 000 YBP (Brown *et al*., 1998).

Thus, human mtDNA variation exhibits two striking geographic discontinuities, one between the enormous African mtDNA diversity and the colonization of Eurasia by only two mtDNAs, M and N; and the second between the extensive radiation of M and N in Eurasia and the colonization of the Americas through the arctic by only four mtDNAs: A, C, D and X. The best explanation for these geographic discontinuities in mtDNA diversity is that selection limited the mtDNAs that could move into the more northern and colder latitudes (Wallace *et al*., 1999, 2003; Mishmar *et al*., 2003; Ruiz-Pesini *et al*., 2004; Wallace, 2005).

Today, these ancient adaptive mtDNA variants influence individual predisposition to disease. For example, the reduced mitochondrial ATP production of the partially uncoupled mitochondria of mtDNA haplogorup J increases susceptibility to blindness of individuals which also acquire one of the milder Leber's hereditary optic neuropathy (LHON) pathogenic mutations (Brown *et al*., 1997, 2001; Torroni *et al*., 1997; Wallace, 2005). However, these same uncoupling mutations also keep the ETC more oxidized in the presence of excess calories, thus reducing ROS production and increasing longevity and protecting against age-related degenerative diseases (Wallace, 2005).

Mitochondrial DNA mutations in cancer

The linkage of mutations in nDNA-encoded mitochondrial gene mutations to cancer (Baysal *et al*., 2000; Niemann and Muller, 2000; Astuti *et al*., 2001; Lehtonen *et al*., 2004; Vanharanta *et al*., 2004) raises the possibility that mtDNA mutations might also contribute to cancer etiology. Cancer mtDNA mutations could either arise in the female germ line and predispose to cancer (oncogenic germline mutations) or arise in the mtDNAs of the tissues (tumor-specific somatic mutations) and participate in the tumor progression process.

Germline mitochondrial DNA mutations in cancer

Two mtDNA population polymorphisms have been associated with increased cancer risk suggesting that germline mtDNA mutations might be important in the etiology of certain cancers. The first of these occurs in the ND3 gene at nt 10398 (nt G10398A, codon A114T) variant in which the 10398A allele was linked to increased risk for invasive breast cancer in both pre- and post-menopausal African-American women, compared to African-American women with the 10398G allele. However, the 10398A allele did not increase breast cancer risk for Caucasian women (Canter *et al*., 2005). The second polymorphism occurs at nt 16189 (T > C) and has been associated with endometrial cancer (Liu *et al*., 2003).

Further evidence for a role of germline mtDNA variants in cancer risk came from a study of mtDNA COI gene variants in European-American prostate cancer patient specimens. The COI gene was studied because proteomic studies had reported an increase in the ratio of nDNA/mtDNA encoded complex IV polypeptides in prostate cancer (Herrmann *et al*., 2003; Krieg *et al*., 2004) and because European mtDNAs lack common COI polymorphisms, thus minimizing possible confusion between ancient variants and cancer-related mutations (Mishmar *et al*., 2003; Ruiz-Pesini *et al*., 2004). The mtDNA COI genes were sequenced from 260 prostectomy specimens from North American Eurasian

patients and from blood samples of 54 prostate cancer-negative men over age 50 years whose prostate-specific antigen was <1 and in which a prostate needle biopsy was found to be free of cancer cells. This analysis revealed that about 11% of the prostate cancer specimens harbored COI mutations while none of the prostate cancer negative controls had COI mutations. Four prostate cancer COI mutations were found to be germline mutations since they were homoplasmic and found in multiple independent patient tumors, often on different mtDNA backgrounds. The first mutation, nt T6253C (M117T, conservation index (CI) = 69%), was found in three independent prostate cancer specimens, all haplogroup H. The second mutation, nt C6340T (A120T, CI = 97%) was found in six patients on haplogroups J, T, L1, and N. The third mutation, nt G6261A (T146I, CI = 79%) was found in two cases on haplogroups of H and N; and the final mutation, nt A6663G (I125V, CI = 95%), was found in five cases on haplogroups LO and L2. Thus, these data prove that germline mtDNA COI mutations can cause prostate cancer and thus that mutant mtDNA genes can be oncogenes (Petros *et al*., 2005).

Surprisingly, a survey of 898 random European mtDNA sequences revealed that 5.5% of the general population also harbored COI variants. One reason why COI mutations that cause prostate cancer could persist in the population may be because only men develop prostate cancer but only women transmit the mtDNA. Hence, the mutant mtDNAs are insulated against selection (Petros *et al*., 2005).

Somatic mitochondrial DNA mutations in cancer

While studies linking germline mtDNA mutations to cancer can be confounded by the high background frequency of functional mtDNA polymorphisms, studies of somatic mtDNA mutations in cancer can be more definitive since the cancer cell should have the neoplastic mtDNA mutation while the normal tissue should not. Numerous mtDNA mutations have now been reported in cancer (Chinnery *et al*., 2002; Copeland *et al*., 2002). Early studies of cancer cell mtDNAs reported dimeric mtDNAs in human leukemia leukocytes (Clayton *et al*., 1970), partial deletions in the mtDNA $tRNA^{Tyr}$ and $tRNA^{Trp}$ genes in rat liver tumor mtDNAs (Taira *et al*., 1983), and heteroplasmy based on restriction digests of tumor mtDNAs (Bianchi *et al*., 1995). However, the first clear demonstration that a mtDNA mutation in cancer cells could be functionally significant came from the report of a renal adenocarcinoma in which 50% of the mtDNAs contained a 294 nt in-frame deletion in the mtDNA ND1 gene resulting in a truncated mRNA (Horton *et al*., 1996). A patient with such a mtDNA mutation would manifest severe mitochondrial disease. Hence, this somatic tumor-specific mutation had to be functionally significant.

Analysis of 10 colon cancer cell lines revealed that seven (70%) harbored functional tumor-specific mtDNA coding region mutations including missense mutations in six lines and a chain termination and a frame shift mutation in the seventh. All of the mutations were homoplasmic except for missense mutations in two of the cell lines (Alonso *et al*., 1997; Polyak *et al*., 1998; Habano *et al*., 1999). A survey of the mtDNAs of bladder, head and neck, and lung cancers revealed that 51% (21/41) contained tumor-specific mtDNA mutations (Fliss *et al*., 2000). Surveys of mtDNA control region variation revealed that 48% (15/31) of gastric tumors harbored tumor-specific mtDNA mutations (Wu *et al*., 2005), 39% (24/61) of hepatocellular carcinomas harbored mtDNA control region mutations in a Taiwan study (Lee *et al*., 2004) and 68% (13/19) of hepatocellular carcinomas had control region mutations in a Japanese study (Nomoto *et al*., 2002). Predominantly mtDNA control region mutations were found in glioblastomas but also two coding region mtDNA variants (Kirches *et al*., 2001) while neurofibromas from neurofibromatosis patients harbored only control region mutations (Kurtz *et al*., 2004). Analysis of mtDNA variation of thyroid tumors revealed that 49% (32/66) of tumors had control region somatic mutations and 52% (34/66) of tumors harbored coding region somatic mutations in one study (Maximo *et al*., 2002) and 23% of 13 papillary thyroid carcinomas had coding region mtDNA mutants in an independent study (Yeh *et al*., 2000). However, thyroid tumors had a relatively low frequency of control region mutants, particularly in the hypervariable homopolymeric C region between nt 303 and 315 (Tong *et al*., 2003). Ovarian tumors harbored both control region (3/15) and coding region variants in 60% (6/10) of cases (Liu *et al*., 2001), breast cancer samples contained control region and coding region mutations in 61% (11/18) of samples (Parrella *et al*., 2001) and prostate cancer tumors had somatic control region mutations in 88% of cases (14/16) (Chen *et al*., 2002, 2003) and control region and coding region mutations in 19% (3/16) of cases (Jeronimo *et al*., 2001). The functional relevance of tumor-specific polypeptide gene missense mutations was demonstrated by three heteroplasmic COI mutations reported in prostate tumors: G5949 (G16X, Stop), T6124C (M74T, CI = 95%) and C6924T (A341S, CI = 100%) (Petros *et al*., 2005). The CI of the two somatic COI missense mutations exceeds the average of all COI variants found in the general European population (CI = 71 ± 35%) as well as the average of all prostate cancer COI mutations (CI = 83 ± 25%). Hence, these mutations must be functionally relevant.

Some somatic mitochondrial DNA mutations are the same as adaptive mutations

To learn more about the nature of the somatic mtDNA mutations in tumors, we assembled all of the data on somatic mtDNA mutations found in ovarian (Liu *et al*., 2001); bladder, head and neck (Fliss *et al*., 2000); colon (Polyak *et al*., 1998); thyroid (Yeh *et al*., 2000) and prostate (Jeronimo *et al*., 2001) tumors and compared these variants to those that have been found in the general population. Since chain termination and insertion–deletion mutations are commonly found in tumor

mtDNA, but would be lethal in humans, these cancer mutations were excluded from the cancer mutation-population polymorphism comparisons.

We then compared the remaining 'tumor-specific somatic mtDNA mutations' with all of the mtDNA variant sites contained in our global human mtDNA database of 1920 complete mtDNA sequences plus 532 coding region mtDNA sequences. These sequences encompass 3512 total population variants including 453 non-coding, 2563 mRNA, 216 tRNA and 280 rRNA variant sites (Brandon *et al*., 2005). Surprisingly, of the 190 tumor-specific somatic mtDNA mutations, 72% (137) were also sequence variants in our population database. Subdividing the tumor-specific somatic mutations by the type of genetic function that was affected revealed that of the 42 tumor-specific somatic mtDNA mRNA gene polypeptide missense mutations 52% (22) were also in the population database, of the six tumor-specific somatic tRNA mutations 83% (five) were present in the population database, of the 13 tumor-specific somatic rRNA mutations 38% (five) were in the population database, and of the 84 tumor-specific somatic control region mutations 85% (71) were found in the population database (Table 1). The probability that 72% of tumor-specific somatic mutation positions would also coincide with population variant positions by chance is much less than 0.001 and the probability that any of the gene-specific classes of tumor somatic mtDNA mutation positions would contain the observed proportion of population variant positions are all 0.01 or less (Table 1).

There are two possible explanations for the discovery that a high proportion of tumor-specific somatic mtDNA mutations are also population variants. Either the majority of the 'tumor-specific somatic mtDNA mutations' are sequencing errors in which population variants were overlooked in either the tumor tissue or the normal tissue sequence, or that cancer cells are prone to acquire some of the same functional mtDNA mutations as they migrate into new environments as did humans when they migrated out of Africa into new environments. The latter possibility would imply that similar selective influences acted on our ancient human ancestors and are active on cells undergoing neoplastic transformation.

One reason to suspect that sequencing errors may be a cause of the high percentage of 'tumor-specific somatic mtDNA mutations' is that for mtDNA mRNA gene synonymous site mutations, 76% (34/45) of the tumor mRNA mutations were also present among the population variants (Table 1). It is generally assumed that the 'synonymous' mRNA variants seen in human mtDNAs are functionally neutral, and that the common synonymous variants have increased in frequency through their linkage due to hitchhiking with adaptive variants located elsewhere on the same mtDNA. If the synonymous variants are neutral, then there would be no reason for them to be selected for in the cancer cell. Hence, they could not be new tumor-specific somatic mtDNA mutations. Rather they must be pre-existing polymorphisms. The only reason why this argument does not completely negate the observed association between tumor mutations and population polymorphisms is that our knowledge of the functions of the mtDNA mRNA nts are rudimentary. Therefore, it is possible that nt changes in the mRNA that do not change an amino acid might change some as yet unidentified function in the mRNA. For example, the mtDNA mRNAs do not have 5′ or 3′ non-translated regions. Hence, internal mRNA sequences must be required to initiate mRNA translation and these could be rendered non-functional by a nt change that did not affect the amino-acid sequence. Thus, though it seems possible that at least some of the 'tumor-specific' variants are sequencing errors, the association between 'tumor-specific' variants and population polymorphisms is too strong to be completely spurious.

The biological relevance of the apparent association between 'tumor-specific somatic mutations' and population variants is supported by a review of selected individual tumor mutations that alter the amino-acid sequence of mtDNA mRNA genes. The most compelling evidence that at least one of the tumor-specific somatic mtDNA mutations is also a population variant is seen for the nt 13708 G to A missesnse mutation in ND5 (A458 T) reported in a breast cancer tumor (Parrella *et al*., 2001). The original publication provides the actual sequencing gel for this tumor-specific somatic mutation. This gel clearly shows that the 13708 nt change is present in the tumor and in a positive lymph

Table 1 Summary of somatic variants

Genomic category	*Nature*	*N*[a]	#[b]	*Percent*	*Probability*[c]
mRNA	NS	42	22	52.38	<0.001[d]
mRNA	S	45	34	75.56	
tRNA	NC	6	5	83.33	<0.001
rRNA	NC	13	5	38.46	0.01
Control	NC	84	71	84.52	<0.001
Total	All	190	137	72.11	<0.001

Abbreviations: NC = nucleotide change; NS = non-synonymous; S = synonymous; All = all types of variants. G9949A and C6924T were excluded from these calculations because both the somatic and tumor alleles were absent from the population database. [a]Number of variants extracted from the literature. [b]Number of these variants also found in a population database of 2452 human mitochondrial DNA sequences. [c]*P*-value corresponding to the probability of overlap between somatic and population variants equal or greater than observed, under the null binomial model in which the probability of a somatic mutation is set to its empirical frequency. [d]Synonymous and non-synonymous variants were treated as one group for calculating the probability.

node but is absent in the adjacent normal tissue and in peripheral blood leucocytes (Parrella *et al*., 2001). Hence, we can conclude that a somatic mutation did arise at nt 13708 in a breast tumor. In African L0–L3, Asian M, and at the base of Eurasian N mtDNAs, the ND5 nt 13708 nt is a G (458A). However, at the root of European haplogroup J and in important sublineages of the European X and Asian B haplogroups the 13708 nt is an A (458T). Therefore, the tumor-specific mutation converts the mtDNA sequence to the more Eurasian-like allele (Ruiz-Pesini *et al*., 2004). The second instructive case involves nt 4977 in the COIII gene. Variants at this nt have been reported in three independent thyroid tumors in one paper (Maximo *et al*., 2002) (Table 2). Two of the three tumors were reported to have a G to A transition (V91I), which is the same base change that defines a major sublineage of haplogroup U (U5a, U5b, U5ab, etc) (Montiel-Sosa *et al*., 2005) (Table 2). However, the third tumor is reported to have a G to T transversion (V91F) which appears to be a more significant amino-acid substitution and which does not appear in our population database. Since both nt changes affect the same codon, but only one is a population variant, this would suggest that changes in codon 91 of COIII may be important in neoplastic transformation. Additional support for a congruence between tumor-specific somatic mutations and population polymorphisms comes from observations at the nt 10 398 ND3 (G to A, A114T) and nt 8701 ATP6 (G to A, A59 T) nts. The 10398 A to G variant was reported as a tumor-specific somatic mtDNA mutation in a thyroid tumor (Yeh *et al*., 2000) (Table 2). The 10 398G allele (114A) is the primary allele for African haplogroups L0–L3 and Asian macro-haplogorup M. The nt 10 398A allele has been linked to breast cancer in African-American women who would more commonly harbor the 10 398G allele (Canter *et al*., 2005). Similarly, the nt 8701 (A to G or T59A) mutation has been reported as a tumor-specific somatic mtDNA mutation in two independent thyroid tumors (Maximo *et al*., 2002) (Table 2). Since the nts 10 398G and 8701G base changes occur together at the base of macro-haplogroup M, it is possible that alterations in these nts could have related functional consequences for both population adaptation and neoplastic transformation.

In addition to these examples of correlations between tumor somatic mtDNA mutations and high abundance population polymorphisms, some tumor-specific somatic mutations correspond with low abundance population variants. Tumor-specific somatic mutations at nts 10 320 and 10 321 in ND3 were reported in two different papers yet change the same valine codon to either an isoleucine or an alanine, the former being seen in one haplotype of haplogroup B and the later in a sublineage of L1 (Table 2). The nt 3308 tumor-specific somatic mutation in ND1 is also seen in subhaplogroups of African L1 and L2, the nt 3505 tumor-specific somatic mutation in ND1 is also found in European haplogroup W, and the tumor somatic nt 3992 mutation in ND1 is also associated with European haplogroup H4 (Table 2). Since each of these 'tumor-specific mutants' is relatively rare in the general population, it would seem unlikely that all these population-specific variant nts would be observed in tumors through random sampling of tumor patient populations and then also be missed when sequencing the normal mtDNA. It seems more likely that these are important functional variants that have been selected for in both human populations and in tumors for functionally related reasons.

Analysis of the interspecific CI for the amino-acid changes associated with the tumor-specific somatic polypeptide mutations revealed another interesting association. Those tumor-specific somatic mutations that do not correspond to population polymorphisms generally have a higher CI than those that do correspond to population variants (Table 2). This suggests that there may be two different classes of tumor mutations, those that are found only in tumors and have a relatively severe deleterious effect on mitochondrial metabolism and ROS production and those that have a milder impact on mitochondrial OXPHOS, alter the same nts as population variants, and help the tumors to adapt to adverse environments (Table 2).

Returning to the 'synonymous' tumor-specific somatic mtDNA mutations the available data indicate that the frequency of synonymous mutations is roughly equal to the frequency of non-synonymous mutations, 42 versus 45 (Table 1). This is often viewed as evidence of positive selection. Also, the synonymous tumor-specific somatic mRNA gene mutations encompass five transversions and 40 transitions (Table 3), consistent with the predominance of transition mutations found in inter-specific and intra-specific population studies (Brown *et al*., 1982; Ingman *et al*., 2000; Mishmar *et al*., 2003). However, in three of the 'tumor-specific somatic mutations' (nt 3357, 3480 and 12372) the tumor-specific somatic mutation is the common population allele while the normal tissue allele is very rare. This could suggest sequencing errors in the normal tissue, which would seem to be the case for the nt 3357 variant in which the 'normal tissue allele' was only found once in our population database while the 'tumor-specific somatic mutation' was present in every other sequence. Moreover, the rarity of these three 'tumor-specific somatic mutations' in the general population would make it unlikely that three individuals with these rare alleles would all be sampled by chance in these tumor studies. On the other hand, the 'normal' alleles of the nts 3480 and 12372 variants are present in 100 and 307 mtDNA sequences in our database, respectively. Hence, it is possible that these might be functionally relevant population variants that have reverted in the tumor. Also, the COII nt 7873, ATP6 nt 8697, ND4 nt 10793 and ND4 nt 10822 have all been observed in two different tumors. In the case of the nt 10793 variant the two reports occur in independent publications and another tumor was reported with a synonymous mutation in the adjacent nt, nt 10792, all of which suggest biological relevance. Thus for the moment, there seems to be no definitive way to assess the validity of the

Table 2 Non-synonymous variants

Position	*Gene*	*Loc*[a]	*Norm*[b]	#[c]	*Tum*[d]	#[e]	*#Obs*[f]	*Norm*[g]	*Tum*[h]	*CI*[i]	*Haplogroups*[j]	*References*
3308	ND1	C	T	2434	C	17	1	M	T	0.79	L1, L2a	Polyak *et al.* (1998)
3434	ND1	P	G	5	A	2447	1	C	Y	0.72	A2, A*, B1119, B4a, B4b, B4d, B9950, B*, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, F1, F2, F*, G1, G2, G*, H1, H10, H13, H2, H3, H4, H5, H6, H6776, H7, HV1, HV*, H*, I1, I2, I*, J1, J1*, J1b, J1c, J2, L0, L1, L2a, L2b, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N1b, N9a, N9*, N*, P, P*, Q, R1, R30, R31, R5, R6, R7, R8, R*, T, T1, T2, U1, U1811*, U2, U3, U4, U5a, U5a1, U5ab*, U5b, U6, U7, U8, Uk1, Uk2, V1, V2, V*, W1a, W3, W*, X, Y, Z, preHV1(B4, B4bd*, N1a, N1d, N5, R9*, T*, W1c)	Jeronimo *et al.* (2001)
3505	ND1	P	A	2399	G	53	1	T	A	0.28	W1a, W3, W*(H1, W1c)	Jeronimo *et al.* (2001)
3526	ND1	T	G	2452	A	0	1	A	T	0.90		Maximo *et al.* (2002)
3910	ND1	T	G	2452	A	0	1	E	K	1.00		Maximo *et al.* (2002)
3992	ND1	T	C	2426	T	26	1	T	M	0.26	H4(H3)	Maximo *et al.* (2002)
6124	CO1	P	T	2452	C	0	1	M	T	0.95		Petros *et al.* (2005)
6924	CO1	P	C	0	T	0	1	P	S	1.00		Petros *et al.* (2005)
7775	CO2	T	G	2452	A	0	1	V	I	0.36		Maximo *et al.* (2002)
7785	CO2	T	T	2452	C	0	1	I	T	0.97		Maximo *et al.* (2002)
8009	CO2	C	G	2452	A	0	1	V	M	0.79		Polyak *et al.* (1998)
8701	ATP6	T	A	1690	G	762	2	T	A	0.44	C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, G1, G2, G*, L0, L1, L2a, L2b, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, Q, Z(A*, B*, H1, L2bc*, N9a, T1, U1)	Maximo *et al.* (2002)
8716	ATP6	T	A	2452	G	0	1	K	E	0.82		Maximo *et al.* (2002)
9137	ATP6	T	T	2450	C	2	1	I	T	0.18	(F2, HV*)	Maximo *et al.* (2002)
9477	CO3	T	G	2366	A	86	2	V	I	0.90	D4*, U5a, U5a1, U5ab*, U5b(B4b)	Maximo *et al.* (2002)
9477	CO3	T	G	2366	T	0	1	V	F	0.90		Maximo *et al.* (2002)
9655	CO3	T	G	2452	A	0	1	S	N	0.97		Maximo *et al.* (2002)
9691	CO3	T	C	2452	T	0	1	A	V	0.87		Maximo *et al.* (2002)
9949	CO3	C	G	0	A	0	1	G	D	0.92		Polyak *et al.* (1998
10197	ND3	T	G	2450	C	0	1	A	P	0.95		(Maximo *et al.* (2002)
10269	ND3	T	C	2452	T	0	1	L	F	0.95		Maximo *et al.* (2002)
10272	ND3	T	C	2452	T	0	1	L	F	0.97		Maximo *et al.* (2002)
10320	ND3	T	G	2450	A	2	1	V	I	0.10	(B4b, F*)	Maximo *et al.* (2002)
10321	ND3	D	T	2437	C	15	1	V	A	0.10	L1(T*)	Fliss *et al.* (2000)
10398	ND3	T	A	1379	G	1072	1	T	A	0.54	B1119, B9950, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, G1, G2, G*, I1, I2, I*, J1, J1*, J1b, J1c, J2, L0, L1, L2a, L2b, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N*, P*, Q, R*, Uk1, Y, Z(L2bc*, N1a, U1811*, U2)	Yeh *et al.* (2000)
10563	ND4L	C	T	2452	C	0	1	C	R	0.97		Polyak *et al.* (1998)
10639	ND4L	T	A	2452	G	0	1	N	S	0.36		Maximo *et al.* (2002)
11016	ND4	T	G	2428	A	24	1	S	N	0.08	H*, N9*, P*(M*, N9a)	Maximo *et al.* (2002)
11150	ND4	H	G	2449	A	3	1	A	T	0.08	(H*, M*, U1811*)	Fliss *et al.* (2000)
11900	ND4	S	G	2452	A	0	1	V	M	0.62		Parrella *et al.* (2001)
12344	ND5	S	T	2452	A	0	1	M	K	0.44		Parrella *et al.* (2001)
12967	ND5	T	A	2451	C	1	1	T	P	0.10	(L0)	Maximo *et al.* (2002)
13708	ND5	S	G	2243	A	209	1	A	T	0.33	A2, B4b, D4-8020, D5, F2, H2, H*, J1, J1*, J1b, J1c, J2, L2a, L2d, N*, X(B9950, D4*, H1, N9*, P*, R5, T2, U1811*, U4, Uk2)	Parrella *et al.* (2001
13943	ND5	T	C	2450	T	2	2	T	M	0.10	(H1, Uk1)	Maximo *et al.* (2002)
14053	ND5	P	G	6	A	2446	1	A	T	0.10	A2, A*, B1119, B4a, B4b, B4d, B9950, B*, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, F1, F2, F*, G1, G2, G*, H1, H10, H13, H2, H3, H4, H5, H6, H6776, H7, HV1, HV*, H*, I1, I2, I*, J1, J1*, J1b, J1c, J2, L0, L1, L2a, L2b, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N1b, N9a, N9*, N*, P, P*, Q, R1, R30, R31, R5, R6, R7, R8, R*, T, T1, T2, U1, U1811*, U2, U3, U4, U5a, U5a1, U5ab*, U5b, U6, U7, U8, Uk1, Uk2, V1, V2, V*, W1a, W3, W*, X, Y, Z, preHV1(B4, B4bd*, L2bc*, N1a, N1d, N5, R9*, T*, W1c)	Jeronimo *et al.* (2001)
14498	ND6	T	T	2452	A	0	1	Y	F	0.97		Maximo *et al.* (2002)
14985	CYB	C	G	2451	A	1	1	R	H	1.00	(H1)	Polyak *et al.* (1998)
15179	CYB	T	G	2452	A	0	1	V	M	1.00		Yeh *et al.* (2000)
15182	CYB	T	A	2452	G	0	1	I	V	1.00		Maximo *et al.* (2002)
15312	CYB	T	T	2452	G	0	1	I	S	0.82		Maximo *et al.* (2002)
15572	CYB	C	T	2452	C	0	1	F	L	0.97		Polyak *et al.* (1998)

[a]Location of the tumor, D = bladder, N = brain, S = breast, C = colorectal, H = head or neck, L = lung, O = ovarian, P = prostate, T = thyroid. [b]Normal germline nucleotide found in the patient. [c]Number of sequences in the population database having the normal allele. [d]Tumor nucleotide. [e]Number of sequences in the population database having the tumor allele. [f]Number of times the somatic variant was observed. [g]Amino acid encoded by the patient's germline DNA. [h]Amino acid encoded by the tumor. [i]Conservation index of the amino-acid position. [j]Haplogroups from the population database with one or more sequences having the tumor allele, haplogroups with only one sequence are listed in parentheses at the end of the list.

Table 3 Synonymous variants

Position	*Gene*	*Loc*[a]	*Norm*[b]	#[c]	*Tum*[d]	#[e]	*#Obs*[f]	*Norm*[g]	*Tum*[h]	*Haplogroups*[i]	*Ref*
3357	ND1	P	A	1	G	2451	1	AUA	AUG	A2, A*, B1119, B4a, B4b, B4d, B9950, B*, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, F1, F2, F*, G1, G2, G*, H1, H10, H13, H2, H3, H4, H5, H6, H6776, H7, HV1, HV*, H*, I1, I2, I*, J1, J1*, J1b, J1c, J2, L0, L1, L2a, L2b, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N1b, N9a, N9*, N*, P, P*, Q, R1, R30, R31, R5, R6, R7, R8, R*, T, T1, T2, U1, U1811*, U2, U3, U4, U5a, U5a1, U5ab*, U5b, U6, U7, U8, Uk1, Uk2, V1, V2, V*, W1a, W3, W*, X, Y, Z, preHV1(B4, B4bd*, L2bc*, N1a, N1d, N5, R9*, T*, W1c)	Jeronimo *et al*. (2001)
3480	ND1	P	G	100	A	2352	1	AAG	AAA	A2, A*, B1119, B4a, B4b, B4d, B9950, B*, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, F1, F2, F*, G1, G2, G*, H1, H10, H13, H2, H3, H4, H5, H6, H6776, H7, HV1, HV*, H*, I1, I2, I*, J1, J1*, J1b, J1c, J2, L0, L1, L2a, L2b, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N1b, N9a, N9*, N*, P, P*, Q, R1, R30, R31, R5, R6, R7, R8, R*, T, T1, T2, U1, U1811*, U2, U3, U4, U5a, U5a1, U5ab*, U5b, U6, U7, U8, Uk1, V1, V2, V*, W1a, W3, W*, X, Y, Z, preHV1(B4, B4bd*, L2bc*, N1a, N1d, N5, R9*, T*, W1c)	Jeronimo *et al*. (2001)
3594	ND1	T	C	2335	T	117	1	GUC	GUU	L0, L1, L2a, L2b, L2c, L2d(L2bc*)	Maximo *et al*. (2002)
3918	ND1	S	G	2431	A	21	1	GAG	GAA	B4b, L2a(EM9, H10, T2, U5a1, Uk1, V*)	Parrella *et al*. (2001)
4613	ND2	T	A	2452	G	0	1	AUA	AUG		Maximo *et al*. (2002)
4940	ND2	T	C	2452	T	0	1	CUC	CUU		Maximo *et al*. (2002)
4985	ND2	T	A	2450	G	2	1	CAA	CAG	(H1, U1)	Maximo *et al*. (2002)
6473	CO1	T	C	2435	T	17	1	AUC	AUU	B4b	Maximo *et al*. (2002)
6650	CO1	T	A	2452	C	0	1	CCA	CCC		Maximo *et al*. (2002)
7103	CO1	T	C	2452	T	0	1	UUC	UUU		Maximo *et al*. (2002)
7819	CO2	T	C	2447	A	5	1	CUC	CUA	L3*	Maximo *et al*. (2002)
7873	CO2	T	C	2450	T	2	2	ACC	ACU	T2	Maximo *et al*. (2002)
8697	ATP6	T	G	2320	A	132	2	AUG	AUA	T, T1, T2, Uk2(I*, T*)	Maximo *et al*. (2002)
8706	ATP6	T	A	2452	G	0	1	AUA	AUG		Maximo *et al*. (2002)
9030	ATP6	T	C	2452	T	0	1	CAC	CAU		Maximo *et al*. (2002)
9296	CO3	T	C	2411	T	41	1	GGC	GGU	D4-8020(R5)	Maximo *et al*. (2002)
9575	CO3	T	G	2414	C	0	1	CCG	CCC		Maximo *et al*. (2002)
9746	CO3	T	G	2452	A	0	1	GAG	GAA		Maximo *et al*. (2002)
10071	ND3	D	T	2451	C	1	1	UUA	CUA	(L1)	Fliss *et al*. (2000)
10181	ND3	T	C	2431	T	21	1	UUC	UUU	D4b	Maximo *et al*. (2002)
10691	ND4L	T	C	2452	G	0	1	GGC	GGG		Maximo *et al*. (2002)
10792	ND4	D	A	2446	G	6	1	UUA	UUG	L1(B4b)	Fliss *et al*. (2000)
10793	ND4	D	C	2447	T	5	1,1	CUA	UUA	L1	Fliss *et al*., (2000) and Maximo *et al*. (2002)
10822	ND4	H	C	2449	T	3	2	CAC	CAU	T2	Fliss *et al*. (2000)
10978	ND4	D	A	2443	G	9	1	CUA	CUG	Uk1(R*)	Fliss *et al*. (2000)
11065	ND4	D	A	2448	G	4	1	CUA	CUG	I2(F*)	Fliss *et al*. (2000)
11332	ND4	T	C	2442	T	10	1	GCC	GCU	U4(R6)	Maximo *et al*. (2002)
11518	ND4	D	G	2451	A	1	1	CUG	CUA	(B4b)	Fliss *et al*. (2000)
11674	ND4	P	C	2399	T	53	1	ACC	ACU	W1a, W3, W*(V*, W1c)	(Jeronimo *et al*. (2001)
11840	ND4	T	C	2442	T	10	1	CUA	UUA	Uk1	Maximo *et al*. (2002)
11947	ND4	P	A	2399	G	53	1	ACA	ACG	W1a, W3, W*(W1c)	Jeronimo *et al*. (2001)
12049	ND4	D	C	2447	T	5	1	UUC	UUU	L1	Fliss *et al*. (2000)
12345	ND5	L	G	2452	A	0	1	AUG	AUA		Fliss *et al*. (2000)
12372	ND5	P	A	307	G	2145	1	CUA	CUG	A2, A*, B1119, B4a, B4b, B4d, B9950, B*, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, F1, F2, F*, G1, G2, G*, H1, H10, H13, H2, H3, H4, H5, H6, H6776, H7, HV1, HV*, H*, I1, I2, I*, J1, J1*, J1b, J1c, J2, L0, L1, L2a, L2b, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N1b, N9*, N*, P, P*, Q, R1, R30, R31, R5, R6, R7, R8, R*, T, T1, T2, V1, V2, V*, W1a, W3, W*, X, Y, Z, preHV1(B4, B4bd*, L2bc*, N1a, N1d, N5, R9*, T*, U1, W1c)	(Jeronimo *et al*. (2001)
12414	ND5	P	T	2391	C	61	1	CCU	CCC	W1a, W3, W*, X(C, L3a, P, R*, W1c)	Jeronimo *et al*. (2001)
12519	ND5	D	T	2436	C	16	1	GUU	GUC	L1(L3*)	Fliss *et al*. (2000)
12705	ND5	P	C	1407	T	1044	1	AUC	AUU	A2, A*, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, G1, G2, G*, I1, I2, I*, L0, L1, L2a, L2b, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N1b, N9a, N9*, N*, Q, W1a, W3, W*, X, Y, Z(B9950, L2bc*, N1a, N1d, N5, W1c)	Jeronimo *et al*. (2001)
12918	ND5	T	C	2452	T	0	1	AAC	AAU		Maximo *et al*. (2002)
14560	ND6	T	G	2407	T	0	1	GUC	GUA		Maximo *et al*. (2002)
14869	CYB	S	G	2442	A	9	1	CUG	CUA	H1, L3a(H6, H*, M11, U7, Uk2)	Parrella *et al*. (2001)
15280	CYB	T	C	2452	T	0	1	UUC	UUU		Maximo *et al*. (2002)

[a]Location of the tumor, D = bladder, N = brain, S = breast, C = colorectal, H = head or neck, L = lung, O = ovarian, P = prostate, T = thyroid. [b]Normal germline nucleotide found in the patient. [c]Number of sequences in the population database having the normal allele. [d]Tumor nucleotide. [e]Number of sequences in the population database having the tumor allele. [f]Number of times the somatic variant was observed. [g]Codon encoded by the patient's germline DNA. [h]Codon encoded by the tumor. [i]Haplogroups from the population database with one or more sequences having the tumor allele, haplogroups with only one sequence are listed in parentheses at the end of the list.

various sequence variants in the published data (Table 3).

Of the six tumor-specific somatic tRNA mutations, 83% (five) were also population variants (Table 4). The $tRNA^{Ile}$ gene mutation at nt 4312 (C to T) is found in African haplogroup L0. The $tRNA^{Ala}$ gene mutation at nt 5633 (C to T) is associated with European subhaplogroup J2, and the $tRNA^{Asp}$ mutation at nt 7521 (G to A) is found in African haplogroups L0–L2. The $tRNA^{Ser(CUN)}$ gene mutation at nt 12 236 (G to A) is found in African L2 and the nt 12308 (A to G) variant defines European haplogroup U (Table 4).

Of the 13 tumor-specific 12S and 16S rRNA mutations, 38% (five) were also found among the population variants and four are common variants (Table 4). For the 12S rRNA, while the tumor-specific somatic nt 664 mutation is not associated with a haplogroup, a nt 663 polymorphism defines the Asian and Native American haplogroup A. Moreover, the tumor-specific somatic 12S rRNA mutation at nt 710 is found in African haplogroup L1 and Eurasia N*. For the 16S rRNA gene, the nt 1811 A to G variant (Fliss *et al*., 2000) subdivides haplogroup U (Montiel-Sosa *et al*., 2005). The tumor-specific somatic mutation nt 1738 (T to C) is found in African L1, and the nt 2056 (G to A) mutation is found in haplogroup R30 (Table 4).

Finally, a stunning 85% (71/84) of the tumor-specific control region mutations have also been observed as population variants (Table 1). Many of the tumor variants are seen in a large number of population-specific haplogroups. This is extreme in nine cases in which less than 300 individuals in the population database have the allele reported for the 'normal tissue' while over 2000 population samples harbor the 'tumor-specific somatic mutation.' On the other hand, 13 of the tumor-specific control region mutations have been reported to occur more than once and most of these reports are from independent publications (Table 5). Hence, many of the observed associations between 'tumor-specific somatic mutations' and population variants for the control region may be functionally relevant. The tumor-specific somatic mutations at nt 73, 16189, 16311 and 16519 (Yoneyama *et al*., 2005) alter nts associated with important population-specific control region mutations. Moreover, a number of the tumor control region mutants (Table 5) have been observed in aging and in age-related diseases. The nt 189 variant accumulates with age in skeletal muscle (Wang *et al*., 2001) and the nt 150 variant accumulates in leukocytes (Zhang *et al*., 2003). Other tumor-specific somatic mutations at nt 73, 146, 152, 195 have also been observed to accumulate in Alzheimer's disease brains (Coskun *et al*., 2004). Interestingly, mtDNA origins of heavy strand replication occur at nt 191, 151 and 57 (Fish *et al*., 2004) and tumor-specific somatic mutations have been reported at 195, 189, 152, 150, 73 and 75 (Table 5). Moreover, multiple tumor-specific somatic mtDNA control region mutations have been observed between nt 16 058 and 16 188, the so-called pre-TAS region (Table 5). This same region has been found to harbor about half of the mtDNA variants that are

Table 4 tRNA and rRNA variants

Position	*Gene*	*Loc*[a]	*Norm*[b]	#[c]	*Tum*[d]	#[e]	*#Obs*[f]	*Haplogroups*[g]	*Ref*
4312	tRNA isoleucine	T	C	2424	T	28	1	L0	Maximo *et al*. (2002)
5521	tRNA tryptophan	L	G	2452	A	0	1		Fliss *et al*. (2000)
5633	tRNA alanine	T	C	2436	T	16	1	J2	Maximo *et al*. (2002)
7521	tRNA aspartic acid	T	G	2331	A	121	1	G*, L0, L1, L2a, L2b, L2c, L2d(L2bc*, M7a, M7c, N*)	Yeh *et al*. (2000)
12236	tRNA serine2	T	G	2434	A	18	1	L2b, L2c(B4d, H3, H*, L3*)	Maximo *et al*. (2002)
12308	tRNA leucine2	P	A	2190	G	262	1	U1, U1811*, U2, U3, U4, U5a, U5a1, U5ab*, U5b, U6, U7, U8, Uk1, Uk2	Jeronimo *et al*. (2001)
664	12S ribosomal RNA	O	G	2452	A	0	1		Liu *et al*. (2001)
710	12S ribosomal RNA	C	T	2433	C	19	1	H2, L1, N*(C, U3)	Polyak *et al*. (1998)
1401	12S ribosomal RNA	O	G	2452	A	0	1		Liu *et al*. (2001)
1738	16S ribosomal RNA	C	T	2436	C	16	1	L1(F1)	Polyak *et al*. (1998)
1811	16S ribosomal RNA	H	A	2289	G	163	1	J1c, U1811*, U2, U3, U4, U7, U8, Uk1, Uk2(N*, R31)	Fliss *et al*. (2000)
1952	16S ribosomal RNA	O	T	2452	C	0	1		Liu *et al*. (2001)
1967	16S ribosomal RNA	C	T	2452	C	0	1		Polyak *et al*. (1998)
2056	16S ribosomal RNA	D	G	2447	A	5	1	R30(B4b, P*, Z)	Fliss *et al*. (2000)
2299	16S ribosomal RNA	C	T	2452	A	0	1		Polyak *et al*. (1998)
2445	16S ribosomal RNA	D	T	2451	C	1	1	(U5a)	Fliss *et al*. (2000)
2664	16S ribosomal RNA	L	T	2452	C	0	1		Fliss *et al*. (2000)
2923	16S ribosomal RNA	P	G	2452	A	0	1		Jeronimo *et al*. (2001)
3054	16S ribosomal RNA	D	G	2452	A	0	1		Fliss *et al*. (2000)

[a]Location of the tumor, D = bladder, N = brain, S = breast, C = colorectal, H = head or neck, L = lung, O = ovarian, P = prostate, T = thyroid. [b]Normal germline nucleotide found in the patient. [c]Number of sequences in the population database having the normal allele. [d]Tumor nucleotide. [e]Number of sequences in the population database having the tumor allele. [f]Number of times the somatic variant was observed. [g]Haplogroups from the population database with one or more sequences having the tumor allele, haplogroups with only one sequence are listed in parentheses at the end of the list.

Table 5 Control region variants

Position	*Loc*[a]	*Norm*[b]	*#*[c]	*Tum*[d]	*#*[e]	*#Obs*[f]	*Haplogroups*[g]	*Ref*
1	P	G	2452	C	0	1		Chen *et al*. (2002)
73	P	A	889	G	1563	3	A2, A*, B1119, B4a, B4b, B4d, B9950, B*, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, F1, F2, F*, G1, G2, G*, H1, H2, H3, H4, HV*, H*, I1, I2, J1, J1*, J1b, J1c, J2, L0, L1, L2a, L2b, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N1b, N9a, N9*, N*, P, P*, Q, R1, R30, R31, R5, R6, R7, R8, R*, T, T1, T2, U1, U1811*, U2, U3, U4, U5a, U5a1, U5ab*, U5b, U6, U7, U8, Uk1, Uk2, W1a, W3, W*, X, Y, Z(B4, B4bd*, I*, L2bc*, N1d, N5, R9*, T*, W1c)	Chen *et al*. (2002)
75	H	G	2450	A	2	1	(B4a, D4-8020)	Fliss *et al*. (2000)
94	P	G	2438	A	14	1	D2, N9*(F1)	Chen *et al*. (2002)
106	P	G	2451	A	0	1		Chen *et al*. (2002)
114	D	T	8	C	2444	1	A2, A*, B1119, B4a, B4b, B4d, B9950, B*, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, F1, F2, F*, G1, G2, G*, H1, H10, H13, H2, H3, H4, H5, H6, H6776, H7, HV1, HV*, H*, I1, I2, I*, J1, J1*, J1b, J1c, J2, L0, L1, L2a, L2b, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N1b, N9a, N9*, N*, P, P*, Q, R1, R30, R31, R5, R6, R7, R8, R*, T, T1, T2, U1, U1811*, U2, U3, U4, U5a, U5a1, U5ab*, U5b, U6, U7, U8, Uk1, Uk2, V1, V2, V*, W1a, W3, W*, X, Y, Z, preHV1(B4, B4bd*, L2bc*, N1a, N1d, N5, R9*, T*, W1c)	Fliss *et al*. (2000)
115	T	T	2452	C	0	1		Maximo *et al*. (2002)
146	P	T	2234	C	217	1,1	A2, A*, B1119, B4a, B4b, B9950, C, D4-8020, D4a, D4*, D5, F1, F*, G2, H1, H*, L0, L1, L2a, L2b, L2c, L2d, L3*, M11, M7a, M7c, Q, U1811*, Uk1, Uk2, Y(B4, B4bd*, D4b, D*, F2, HV*, L2bc*, M2, M8a, M*, N1b, N9a, N9*, R31, R6, R7, T1, T2, U2, U4, U5b, W3, X, Z)	Jeronimo *et al*. (2001) and Liu *et al*. (2001)
150	P	C	2210	T	240	1,1	B1119, B4a, D4-8020, D5, G1, HV*, J1c, L0, L2b, L2c, L2d, L3a, L3*, M7b, N9a, N*, T2, U3, U5a, U5ab*, U5b, U6(B4b, B*, D4*, G*, H3, H7, HV1, H*, J2, L2a, L2bc*, M7a, N1b, R*, U5a1)	Chen *et al*., (2002) and Maximo *et al*. (2002)
152	O	T	2035	C	389	1	A*, B4a, B4b, D2, D4a, D4*, D4b, D5, F1, F*, G2, G*, H1, H2, H3, HV*, H*, I2, L0, L1, L2a, L2b, L2c, L3*, M11, M2, M7a, M8a, M*, N1b, N*, P*, R30, R5, T1, T2, U1811*, U2, U5a1, U6, U7, Uk1, Uk2, Z(B9950, D*, H10, H13, H4, J1*, J2, L2d, M1, N9*, P, Q, R7, T*, U4, V2, Y, preHV1)	Liu *et al*. (2001)
174	P	C	2452	T	0	1		Chen *et al*. (2002)
185	N	A	81	G	2355	1	A2, A*, B1119, B4a, B4b, B4d, B9950, B*, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, F1, F2, F*, G1, G2, G*, H1, H10, H13, H2, H3, H4, H5, H6, H6776, H7, HV1, HV*, H*, I1, I2, I*, J1*, J1b, J1c, J2, L0, L1, L2a, L2b, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N1b, N9a, N9*, N*, P, P*, Q, R1, R30, R31, R5, R6, R7, R8, R*, T, T1, T2, U1, U1811*, U2, U3, U4, U5a, U5a1, U5ab*, U5b, U6, U7, U8, Uk1, Uk2, V1, V2, V*, W1a, W3, W*, X, Y, Z, preHV1(B4, L2bc*, N1a, N1d, N5, R9*, T*, W1c)	Kirches *et al*. (2001)
185	T	G	2355	A	81	1	J1, J1c, L0, L3a, L3*, R*(B4bd*, F1, G2, H7, J1b, R31, Z)	Maximo *et al*. (2002)
189	P	A	2341	G	93	1	B9950, D4*, J1c, J2, L0, L1, L3a, L3*, R*, W1a, W3, W*(B4a, B4b, B4bd*, D5, F1, N9*, U5ab*, V*, W1c)	Jeronimo *et al*. (2001)
195	P	T	2134	C	309	2,1,1	B1119, B9950, C, D4*, D5, EM9, F1, G1, H2, H4, H6776, HV*, H*, J2, L0, L1, L2a, L2b, L2c, L3a, L3*, M1, M7b, N9a, N*, R6, R8, T1, T2, U4, U6, Uk1, W1a, W3, W*, X(B4a, B4bd*, D2, D4b, F*, G2, G*, H1, H3, J1c, L2bc*, M7a, P, R30, R31, R*, T*, U1, U1811*, U5a1, V*, W1c)	Chen *et al*. (2002), Kirches *et al*. (2001) and Maximo *et al*. (2002)
199	O	T	2338	C	114	1	B4b, B9950, D4-8020, D4*, I1, I2, L0, M7b, M7c(L2c, M7a, M*, N1d, Q, T2, U1, U1811*)	Liu *et al*. (2001)
204	N	C	129	T	2322	1	A2, A*, B1119, B4a, B4b, B4d, B9950, B*, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, EM9, F1, F2, F*, G1, G2, G*, H1, H10, H13, H2, H3, H4, H5, H6, H6776, H7, HV1, HV*, H*, I1, I2, I*, J1, J1*, J1b, J1c, J2, L0, L1, L2a, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N1b, N9a, N9*, N*, P, P*, Q, R1, R30, R31, R5, R6, R7, R8, R*, T, T1, T2, U1, U1811*, U2, U3, U4, U5a, U5a1, U5ab*, U5b, U6, U7, U8, Uk1, Uk2, V1, V2, V*, W3, X, Y, Z, preHV1(B4, B4bd*, D*, L2bc*, N1a, N5, R9*, T*, W*)	Kirches *et al*. (2001)
204	P	T	2322	C	129	1	B9950, D4*, I1, I2, L0, L1, L2b, M7a, M7b, N9*, W1a, W3, W*(B4a, D*, F1, H*, J1c, M2, M7c, N1d, P*, R31, T1, T2, W1c, X, Z)	Jeronimo *et al*., 2001)
207	T	G	2327	A	125	1,1,1	B4b, B9950, D4-8020, D4*, F*, G2, I2, L0, L2b, M7a, N9*, R31, W1a, W3, W*(A*, F1, G1, I1, M7b, P*, U5b, Uk2, V*, W1c, X, Z)	Chen *et al*. (2002), Jeronimo *et al*. (2001) and Maximo *et al*. (2002)
235	P	A	2391	G	61	1	A2, A*(D5, N9a)	Jeronimo *et al*. (2001)
235	P	G	61	C	0	1		Chen *et al*. (2002)
295	N	T	102	C	2349	1	A2, A*, B1119, B4a, B4b, B4d, B9950, B*, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, F1, F2, F*, G1, G2, G*, H1, H10, H13, H2, H3, H4, H5, H6, H6776, H7, HV1, HV*, H*, I1, I2, I*, J1*, J1b, J1c, J2, L0, L1, L2a, L2b, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N1b, N9a, N9*, N*, P, P*, Q, R30, R31, R5, R6, R7, R8, R*, T, T1, T2, U1, U1811*, U2, U3, U4, U5a, U5a1, U5ab*, U5b, U6, U7, U8, Uk1, Uk2, V1, V2, V*, W1a, W3, W*, X, Y, Z, preHV1(B4, B4bd*, L2bc*, N1a, N1d, N5, R1, R9*, T*, W1c)	Kirches *et al*. (2001)
313	P	C	2452	T	0	1		Chen *et al*. (2002)
315	P	C	2452	T	0	1		Chen *et al*. (2002)
325	T	C	2442	T	10	2	D4*, L2c	Maximo *et al*. (2002)

Table 5 (*continued*)

Position	*Loc*[a]	*Norm*[b]	#[c]	*Tum*[d]	#[e]	*#Obs*[f]	*Haplogroups*[g]	*Ref*
386	D	C	2452	A	0	1		Fliss *et al.* (2000)
456	T	C	2403	T	47	1	D5, H5, H*, L0(L2bc*, L3a, R8)	Maximo *et al.* (2002)
460	T	T	2450	C	2	1	(H1, Uk1)	Maximo *et al.* (2002)
462	T	C	2363	T	89	1	J1, J1*, J1b, J1c(V1)	Maximo *et al.* (2002)
481	T	C	2449	T	3	1	B9950(D4a)	Maximo *et al.* (2002)
489	P	T	1762	C	690	1,1	C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, G1, G2, G*, J1, J1*, J1b, J1c, J2, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, Q, Z(N9a, U2)	Chen *et al.* (2002) and Liu *et al.* (2001)
497	T	C	2429	T	23	1	H2, Uk1	Maximo *et al.* (2002)
499	T	G	2422	A	30	1,1	B4b, U4(L2c)	Chen *et al.* (2002) and Maximo *et al.* (2002)
549	T	C	2452	T	0	1		Maximo *et al.* (2002)
16093	S	C	90	T	2361	1	A2, A*, B1119, B4a, B4b, B4d, B9950, B*, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, F1, F2, F*, G1, G2, G*, H1, H10, H13, H2, H3, H4, H5, H6, H6776, H7, HV1, HV*, H*, I1, I2, I*, J1, J1*, J1b, J1c, J2, L0, L1, L2a, L2b, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N1b, N9a, N9*, N*, P, P*, Q, R1, R30, R5, R6, R7, R8, R*, T, T1, T2, U1, U1811*, U2, U3, U4, U5a, U5a1, U5ab*, U5b, U6, U7, Uk1, Uk2, V1, V2, V*, W1a, W3, W*, X, Y, Z, preHV1(B4, B4bd*, N1a, N1d, N5, R31, R9*, T*, U8, W1c)	Parrella *et al.* (2001)
16093	P	T	2361	C	90	1	B1119, C, D2, D4a, D4*, G2, H1, H10, H*, L0, L3*, M10, M7a, M*, N9a, R31, U8, Uk1(A*, B4a, B*, D4b, D5, G*, H5, H6, J1c, L1, L2a, L2bc*, L2c, L3a, Q, R6, R8, R*, T2, U3, U5a1, U5b, U7, X)	Chen *et al.* (2002)
16111	P	C	2409	T	40	1	A2, B1119, B9950, D2, N9a(B4b, D4-8020, D4*, F1, H*, J1c, N5)	Chen *et al.* (2002)
16126	N	C	213	T	2239	1	A2, A*, B1119, B4a, B4b, B4d, B9950, B*, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, F1, F2, F*, G1, G2, G*, H1, H10, H13, H2, H3, H4, H5, H6, H6776, H7, HV1, HV*, H*, I1, I2, I*, J1*, J1b, J1c, J2, L0, L1, L2a, L2b, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N1b, N9a, N9*, N*, P, P*, Q, R1, R30, R31, R5, R6, R7, R8, R*, T1, T2, U1, U1811*, U2, U3, U4, U5a, U5a1, U5ab*, U5b, U6, U7, U8, Uk1, Uk2, V1, V2, V*, W1a, W3, W*, X, Z(B4, B4bd*, L2bc*, N1a, N1d, N5, R9*, W1c, preHV1)	Kirches *et al.* (2001)
16134	N	C	2452	T	0	1		(Kirches *et al.* (2001)
16172	H	C	134	T	2315	1	A2, A*, B1119, B4a, B4b, B4d, B9950, B*, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, F1, F2, F*, G1, G2, G*, H1, H10, H13, H2, H3, H4, H5, H6, H6776, H7, HV1, HV*, H*, I1, I2, I*, J1, J1*, J1b, J1c, J2, L0, L1, L2a, L2b, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N1b, N9a, N9*, N*, P, P*, Q, R1, R30, R31, R5, R6, R7, R8, R*, T1, T2, U1, U1811*, U2, U3, U4, U5a, U5a1, U5ab*, U5b, U6, U7, U8, Uk1, Uk2, V1, V2, V*, W1a, W3, W*, X, Y, Z, preHV1(B4, B4bd*, L2bc*, N1a, N1d, N5, R9*, T, T*, W1c)	Fliss *et al.* (2000)
16182	P	A	2328	C	120	2	B1119, B4a, B4b, B9950, B*, D5, F1, N9*, R30, R31, R*, U1, U2(A*, B4bd*, B4d, H4, L3a, M1, M7b, T1, T2, Uk2, X)	Chen *et al.* (2002)
16183	P	A	2197	C	237	2	B1119, B4a, B4b, B4d, B9950, B*, D4-8020, D5, D*, F1, G2, L2a, L3a, M7a, M7b, N9*, R30, R31, R*, T1, U1, U2, X(A*, B4bd*, G*, H1, H4, J1c, M1, N*, P, T2, Uk2)	Chen *et al.* (2002)
16183	P	A	2197	G	14	1	D4*, U2, V1(U1)	Jeronimo *et al.* (2001)
16183	L	C	237	A	2197	1	A2, A*, B1119, B4a, B4b, B9950, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, EM9, F1, F2, F*, G1, G2, G*, H1, H10, H13, H2, H3, H4, H5, H6, H6776, H7, HV1, HV*, H*, I1, I2, I*, J1, J1*, J1b, J1c, J2, L0, L1, L2a, L2b, L2c, L2d, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N1b, N9a, N*, P, P*, Q, R1, R30, R5, R6, R7, R8, R*, T, T1, T2, U1, U1811*, U2, U3, U4, U5a, U5a1, U5ab*, U5b, U6, U7, U8, Uk1, Uk2, V1, V2, V*, W1a, W3, W*, X, Y, Z, preHV1(B4, B4d, L2bc*, N1a, N1d, N5, N9*, R31, R9*, T*, W1c)	Fliss *et al.* (2000)
16187	L	C	2391	T	61	1	A*, D4-8020, G1, L0(D4*, L1, L2c, Q, U7)	Fliss *et al.* (2000)
16189	P	T	1926	C	422	2,1	A*, B1119, B4a, B4b, B4d, B9950, B*, C, D4-8020, D4*, D5, D*, F1, G1, G2, G*, H1, H*, I1, J1c, L0, L1, L2a, L3a, M7a, M7b, N9a, N9*, R30, R31, R*, T1, T2, U1, U2, U5b, U6, X(B4, B4bd*, H4, H6, H7, J1, L2c, M1, M11, N*, P, R5, T, U1811*, Uk2, Y)	Chen *et al.* (2002) and Jeronimo *et al.* (2001)
16217	P	T	2361	C	88	1	B1119, B4a, B4b, B4d, B*(B4, B4bd*, C, HV*, R*)	Chen *et al.* (2002)
16218	P	C	2445	T	7	1	F*, R31(F1, L3*, N*)	Chen *et al.* (2002)
16232	P	C	2430	A	18	1	F1	Chen *et al.* (2002)
16249	P	T	2412	C	40	1	B*, D4a, F1, L1, M1, U1(A*, B4b, C, H6, P*)	Chen *et al.* (2002)
16265	D	A	2423	C	15	1	G*, L1, Q(F1, J2)	Fliss *et al.* (2000)
16274	P	G	2412	A	40	1	B1119, D4*, G*, H2, J1b, L1, P*, R6(D2, D*, G2, J1c, M7a, N1d, N*)	Chen *et al.* (2002)
16278	O	T	219	C	2233	1	A2, A*, B1119, B4a, B4b, B4d, B9950, B*, C, D1, D2, D4-8020, D4a, D4*, D4b, D5, D*, EM9, F1, F2, F*, G1, G*, H1, H10, H13, H2, H3, H4, H5, H6, H6776, H7, HV1, HV*, H*, I1, I2, I*, J1, J1*, J1b, J1c, J2, L0, L3a, L3*, M1, M10, M11, M2, M7a, M7b, M7c, M8a, M*, N1b, N9a, N9*, N*, P, P*, Q, R1, R30, R31, R5, R6, R7, R8, R*, T, T1, T2, U1, U1811*, U2, U3, U4, U5a, U5a1, U5ab*, U5b, U6, U7, U8, Uk1, Uk2, V1, V2, V*, W1a, W3, W*, Y, Z, preHV1(B4, B4bd*, G2, N1a, N1d, N5, R9*, T*, W1c)	Liu *et al.* (2001)
16292	H	C	2384	T	66	1,1	D4*, L2d, L3*, R31, T2, W1a, W3, W*(B*, D5, L1, N*, P, R8)	Fliss *et al.* (2000), Parrella *et al.* (2001)
16293	N	A	2420	G	31	1	H*, L0, L1(HV*, M8a, R7, Uk1)	Kirches *et al.* (2001)
16298	P	T	2292	C	160	1	B4b, C, HV*, M7b, M8a, R30, V1, V2, V*, Z(D4b, F*, H*, L2a, U2, X)	Chen *et al.* (2002)
16300	H	A	2445	G	7	1	D5, EM9(H*, J1*, L2bc*)	Fliss *et al.* (2000)
16304	P	T	2320	C	132	1	F1, F2, F*, H5, H*, N9a, R5, T2(HV*, N9*, R1, R31, R7, R9*, R*, T, U5a1, U5ab*, U7)	Chen *et al.* (2002)
16311	P	T	2150	C	301	1	A*, B1119, B4a, D2, D4-8020, D4a, D4*, D5, F1, F*, H1, H13, H2, H6776, HV*, H*, I1, J1c, L0, L1, L3a, L3*, M1, M10, M11,	Chen *et al.* (2002)

Table 5 (*continued*)

Position	*Loc*[a]	*Norm*[b]	*#*[c]	*Tum*[d]	*#*[e]	*#Obs*[f]	*Haplogroups*[g]	*Ref*
							M8a, M*, N*, P*, Q, R5, R*, T1, U1811*, U5a, U6, Uk1, Uk2, Y(C, D*, G2, H3, H5, J1, L2a, L2b, L2bc*, L2d, M2, M7b, N5, N9a, N9*, R1, R30, R7, R8, U2, U5a1, V2, W3, Z)	
16356	N	T	2427	C	25	1	C, H1, R5, U4(A*, D4*, HV*, M*, N1d, N*, U3)	Kirches *et al.* (2001)
16365	O	C	2452	T	0	1		Liu *et al.* (2001)
16403	P	C	2452	T	0	1		Chen *et al.* (2002)
16459	P	C	2452	T	0	1		Chen *et al.* (2002)
16474	P	G	2452	C	0	1		Chen *et al.* (2002)
16519	P	T	1529	C	923	1,1	A*, B1119, B4a, B4b, B4d, B9950, B*, C, D2, D4-8020, D4a, D4*, D5, F1, F2, G1, G2, H1, H10, H13, H2, H3, H6, H6776, H7, HV*, H*, I1, J1b, J1c, L0, L1, L2a, L2c, L2d, L3a, L3*, M1, M10, M2, M7a, M7c, M8a, M*, N1b, N9*, N*, P*, R1, R30, R5, R6, R8, R*, T, T1, T2, U1, U1811*, U2, U3, U4, U7, Uk1, Uk2, V1, V2, V*, W3, W*, X, Y(A2, B4, B4bd*, D1, D4b, H5, HV1, I*, J1*, L2bc*, M7b, N9a, P, R31, R9*, T*, U5a, U5ab*, U5b, W1a, W1c, Z, preHV1)	Chen *et al.* (2002) and
16532	D	A	2452	T	0	1		Fliss *et al.* (2000)

[a]Location of the tumor, D = bladder, N = brain, S = breast, C = colorectal, H = head or neck, L = lung, O = ovarian, P = prostate, T = thyroid. [b]Normal germline nucleotide found in the patient. [c]Number of sequences in the population database having the normal allele. [d]Tumor nucleotide. [e]Number of sequences in the population database having the tumor allele. [f]Number of times the somatic variant was observed. [g]Haplogroups from the population database with one or more sequences having the tumor allele, haplogroups with only one sequence are listed in parentheses at the end of the list.

associated with maternally inherited cyclic vomiting syndrome and migraine headaches without aura (Wang *et al*., 2004).

The 16 189 T to C variant is also intriguing since it converts a T that disrupts a string of Cs (CCCCCT^{16189}CCCC) to a C (CCCCCCCCCC) thus encouraging slip mispairing and the destabilization of this region on the mtDNA. This same variant has been associated with increased risk for type 2 diabetes mellitus (Poulton, 1998, 1998), dilated cardiomyopathy (Khogali *et al*., 2001) and endometrial cancer (Liu *et al*., 2003). Similarly in one Taiwanese case of gastric cancer, the patient's tumor was found to be heteroplasmic for a 270 nt insertion between nt 309 and 568. This was associated with instability in the number of Cs in a homopolymeric string starting at nt 568 (Wu *et al*., 2005). This same duplication was found to be a characteristic of European haplogroup I (Torroni *et al*., 1994) and has been implicated in predisposition to mtDNA rearrangements (Brockington *et al*., 1993).

One of the most common mtDNA sequence variants seen in the control region of tumor cells is variation in the number of Cs in the homopolymer string of Cs in the nt region 303–315 (Nomoto *et al*., 2002; Kurtz *et al*., 2004; Lee *et al*., 2004; Wu *et al*., 2005; Yoneyama *et al*., 2005). This is also a common variant in normal tissues (Torroni *et al*., 1994; Trounce *et al*., 1995).

Assuming that the tumor-specific somatic control region mutations are real, which from the sheer mass of data seems irrefutable, then we can conclude that there is a very strong correlation between the somatic tumor mutations and the population variants. This association then shows that most of the common mtDNA control region sequence variants found in tumors, aging tissues and populations are not neutral, as was previously assumed, but instead are functional and in certain contexts may be either tumorigenic or adaptive.

Oncogenicity of mitochondrial DNA mutations in cancer

That mtDNA mutations can contribute to the oncogenicity of tumors was demonstrated by introducing a known pathogenic mtDNA mutation into the prostate cancer cell line, PC3, and demonstrating increased tumorigenicity. The mutant mtDNA was derived from a mitochondrial disease patient who was heteroplasmic for the pathogenic mtDNA T8993G mutation (Holt *et al*., 1990). Cell lines were derived from the patient's cells that were homoplasmic for either the mutant (T8993G) or wild-type (T8993T) mtDNAs. The T8993G mutation causes a L156R amino acid substitution that results in a 70% reduction in ATP synthase (complex V) activity (Trounce *et al*., 1994) and a significant increase in mitochondrial ROS production (Mattiazzi *et al*., 2004). The T8993T and T8993G cell lines were then enucleated and the cytoplasts fused to PC3 cells that had been cured of their resident mtDNAs by treatment with the mitochondrial toxin rhodamine-6G (Trounce *et al*., 1996). This created PC3 (mtDNA T8993T) and PC3

(mtDNA T8993G) cytoplasmic hybrids or cybrids (Trounce *et al*., 1994, 1996).

The PC3 (mtDNA T8993T) and PC3 (mtDNA T8993G) cybrids were then injected into nude mice. The PC3 cells with the wild-type (T8993T) mtDNA barely grew in the nude mice. By contrast, the PC3 cells harboring the mutant (T8993G) mtDNA grew rapidly and required that the mice be euthanized. Therefore, mtDNA mutations can functionally contribute to the tumorigenesis of cancer cells (Petros *et al*., 2005).

Consistent with the prediction from paraganglioma genetics that mitochondrial ROS may be important in oncogenicity, the cells from the PC3 (mtDNA T8993G) tumors stained much more strongly for ROS production by dihydroethidium than did the cells from the wild-type PC3 (mtDNA T8993T) nodules (Petros *et al*., 2005). Thus, mitochondrial ROS production would appear to be an important component in the etiology of cancer.

A mitochondrial DNA genetics of cancer

These observations suggest that cancer mtDNA mutations may fall into two main classes: tumorigenic and adaptive. The tumorigenic mutants are the more severe mutations and include heteroplasmic insertion–deletion mutations, chain termination mutations, and missense mutations that change highly conserved amino acids. These mutations are envisioned as inhibiting the ETC resulting in a marked increase in mitochondrial ROS production. The increased mitochondrial ROS may then facilitate cellular transformation by acting as both a tumor initiator, mutagenizing proto-oncogenes into oncogenes, and as a tumor promoter, stimulating the cell to start replicating.

The adaptive mtDNA mutations are milder mutations that are also seen in different human populations and/or during aging and disease. The polypeptide variants are missense mutations that alter functionally important amino acids and the control region mutations modulate replication and transcription. These adaptive mutations can be envisioned as facilitating the tumors survival under the adverse environments that it encounters as it metastasizes. Environmental challenges that might be encountered include altered and reduced energetic substrates, reduced oxygen tension, variation in environmental temperature, tolerance of increased ROS toxicity and exposure to conditions that induce apoptosis. Hence, the tumorigenic mutations are analogous to human pathogenic mutations while the tumor adaptive mutations are analogous to the human regional adaptive variants.

The elevated ROS production of the tumorigenic mutations would result in increased H_2O_2 diffusion out of the mitochondrion. In the nucleus, the H_2O_2 can interact with reduced transition metals to generate hydroxyl radical that can act as a potent nDNA damaging agent and mutagen. In the cytosol, the H_2O_2 may interact with mitogenic signal transduction pathways including nuclear factor (NF)-κB, apurinic/apyrimidinic endonuclease 1 (APE-1), *Fos*, *Jun* and tyrosine kinases (Src kinase, protein kinase C, mitogen-activated protein kinases and receptor tyrosine kinases) (McCord, 2000). APE-1 is a dual functional protein which is also redox sensitive and involved in the redox regulation of *Fos*, *Jun*, NF-κB, PAX, hypoxia-inducible factor-1α and p53 (Evans *et al*., 2000; Kelley and Parsons, 2001). Thus mitochondrial H_2O_2 could drive cell proliferation.

The activation of the tryrosine kinases and/or Ras would activate the phosphatidylinositol-3 kinase, which in turn would activate the *Akt*/PKB kinases. The oncogene-induced replication of the pre-transformed cells in solid unvascularized tissue would also result in hypoxia. This would induce hexokinase II transcription and result in its phosphorylation by *Akt*/PKB kinases thus promoting its binding to mitochondrial VDAC. This would couple mitochondrial ATP production to driving glycolysis and inhibit apoptosis by stabilizing the mtPTP.

The activated *Akt*/PKB kinases would phosphorylate the FOXO transcription factors. Phosphorylation inactivates the FOXOs resulting in the downregulation of nDNA genes which have insulin response elements in their promoters. This class of genes includes antioxidant genes such as mitochondrial MnSOD gene and the master mitochondrial biogenesis gene regulator peroxisome-proliferation-activated receptor γ-coactivator 1α (Wallace, 2005). The downregulation of OXPHOS via the *Akt*/PKB and FOXO pathway could than explain the general reduction in mitochondrial number and OXPHOS enzymes in cancer cells (Pedersen, 1978).

Since the two different classes of somatic tumor mtDNA mutations have different functions in the cancer cell, they might also be expected to arise and be lost from the tumor cells at different times during the neoplastic transformation process. The severly deleterious tumorigenic mutations that inhibit the ETC would be advantageous in the intial phases of tumor growth when the tumor requires mitochondrial H_2O_2 to drive cell proliferation. In this early phase, the tumor is hypoxic and thus can tolerate OXPHOS deficiency since it is exploiting hexokinase II to generate ATP by glycolysis. However, when the tumor becomes vascularized and/or metastasizes and the cells return to a high oxygen tension environment, then it may be more advantageous for the established transformed cells to revert to a more oxidative metabolism. In this instance, the severely deleterious heteroplasmic tumorigenic mtDNA mutations can be lost by replicative segregation and the cell revert back toward a more oxidative mtDNA genotype favored in the metastases. This conjecture could explain why the heteroplasmic 294 nt ND1 deletion mtDNA observed to be in 50% of the mtDNAs in the primary renal cell carcinoma was absent in the subsequent metastatic tumors from the same patient (Horton *et al*., 1996).

By contrast, the adaptive mtDNA mutations may permit the tumor cells to flourish in new environments as they metastasize. Since the migratory tumor cells could be exposed to similar envrionmental challenges as were the humans who migrated out of Africa, the same mtDNA mutations might be adaptive in both tumors and people. Since these mutations would continue to be

adaptive once the tumor is actively growing in its new metastatic site, the mutations would shift from the initially heteroplasmic to homoplasmic mutations. This could explain why many of the tumor somatic mutations are homoplasmic.

Thus, the considerations of this review support the conclusion that mtDNA mutations play an important role in the etiology of solid tumors, consistent with Warburg's hypothesis. The more deletious mutations are tumorigenic, resulting in increased mitochondrial ROS production which is mutagenic without significantly imparing cellular ATP production under the intial anaerobic conditions. As the tumor vascularizes and/or metastasizes and moves into a more oxygen rich environment, the severely deleterious mtDNA mutations become disadvantageous and are lost, but less deleterious adaptive mtDNA mutations are acquired. Since these become increasingly beneficial they segregate to homoplasy. Thus, the central role of the mitochondria in energy production, ROS generation and apoptosis regulation combines with the novel stochastic genetics of the 'thousand copy' mtDNA to provide rational explanations for many of the metabolic and stochastic features of solid tumor biology. Therefore, understanding mtDNA variation in cancer cells promises to provide major new insights into the etiology of solid tumors and be of great potential benefit for prognosis and possible treatment of cancer.

Acknowledgements

We thank Ms MT Lott for her assistance with this manuscript. This work has been supported by NIH Biomedical Informatics Training Grant LM07443 supporting MB and PB awarded to PB and NIH RO1 Grants NS21328, AG13154, NS41650, AG24373 and HL64017 awarded to DCW.

References

Alonso A, Martin P, Albarran C, Aquilera B, Garcia O, Guzman A *et al*. (1997). *Electrophoresis* **18**: 682–685.
Aoki M, Batista O, Bellacosa A, Tsichlis P, Vogt PK. (1998). *Proc Natl Acad Sci USA* **95**: 14950–14955.
Astuti D, Latif F, Dallol A, Dahia PL, Douglas F, George E *et al*. (2001). *Am J Hum Genet* **69**: 49–54.
Baysal BE, Ferrell RE, Willett-Brozick JE, Lawrence EC, Myssiorek D, Bosch A *et al*. (2000). *Science* **287**: 848–851.
Bianchi MS, Bianchi NO, Bailliet G. (1995). *Cytogenet Cell Genet* **71**: 99–103.
Birnbaum MJ. (2004). *Dev Cell* **7**: 781–782.
Bourgeron T, Rustin P, Chretien D, Birch-Machin M, Bourgeois M, Viegas-Pequignot E *et al*. (1995). *Nat Genet* **11**: 144–149.
Brandon MC, Lott MT, Nguyen KC, Spolim S, Navathe SB, Baldi P *et al*. (2005). *Nucl Acids Res* **1**: 33 (database issue): D611–D613.
Brazil DP, Yang ZZ, Hemmings BA. (2004). *Trends Biochem Sci* **29**: 233–242.
Brockington M, Sweeney MG, Hammans SR, Morgan-Hughes JA, Harding AE. (1993). *Nat Genet* **4**: 67–71.
Brown MD, Hosseini SH, Torroni A, Bandelt HJ, Allen JC, Schurr TG *et al*. (1998). *Am J Hum Genet* **63**: 1852–1861.
Brown MD, Sun F, Wallace DC. (1997). *Am J Hum Genet* **60**: 381–387.
Brown MD, Zhadanov S, Allen JC, Hosseini S, Newman NJ, Atamonov VV *et al*. (2001). *Hum Genet* **109**: 33–39.
Brown WM, Prager EM, Wan A, Wilson AC. (1982). *J Mol Evol* **18**: 225–239.
Bustamante E, Pedersen PL. (1977). *Proc Natl Acad Sci USA* **74**: 3735–3739.
Cann RL, Stoneking M, Wilson AC. (1987). *Nature* **325**: 31–36.
Canter JA, Kallianpur AR, Parl FF, Millikan RC. (2005). *Cancer Res* **65**: 8028–8033.
Chen JZ, Gokden N, Greene GF, Green B, Kadlubar FF. (2003). *Carcinogenesis* **24**: 1481–1487.
Chen JZ, Gokden N, Greene GF, Mukunyadzi P, Kadlubar FF. (2002). *Cancer Res* **62**: 6470–6474.
Chinnery PF, Samuels DC, Elson J, Turnbull DM. (2002). *Lancet* **360**: 1323–1325.
Clayton DA, Davis RW, Vinograd J. (1970). *J Mol Biol* **47**: 137–153.
Copeland WC, Wachsman JT, Johnson FM, Penta JS. (2002). *Cancer Invest* **20**: 557–569.
Coskun PE, Beal MF, Wallace DC. (2004). *Proc Natl Acad Sci USA* **101**: 10726–10731.
Evans AR, Limp-Foster M, Kelley MR. (2000). *Mutat Res* **461**: 83–108.
Fish J, Raule N, Attardi G. (2004). *Science* **306**: 2098–2101.
Fliss MS, Usadel H, Caballero OL, Wu L, Buta MR, Eleff SM *et al*. (2000). *Science* **287**: 2017–2019.
Gatenby RA, Gillies RJ. (2004). *Nat Rev Cancer* **4**: 891–899.
Glaichenhaus N, Leopold P, Cuzin F. (1986). *EMBO J* **5**: 1261–1265.
Gottlob K, Majewski N, Kennedy S, Kandel E, Robey RB, Hay N. (2001). *Genes Dev* **15**: 1406–1418.
Habano W, Sugai T, Yoshida T, Nakamura S. (1999). *Int J Cancer* **83**: 625–629.
Herrmann PC, Gillespie JW, Charboneau L, Bichsel VE, Paweletz CP, Calvert VS *et al*. (2003). *Proteomics* **3**: 1801–1810.
Holt IJ, Harding AE, Petty RK, Morgan-Hughes JA. (1990). *Am J Hum Genet* **46**: 428–433.
Horton TM, Petros JA, Heddi A, Shoffner J, Kaufman AE, Graham Jr SD *et al*. (1996). *Genes Chromosomes Cancer* **15**: 95–101.
Ingman M, Kaessmann H, Paabo S, Gyllensten U. (2000). *Nature* **408**: 708–713.
Ishii N, Fujii M, Hartman PS, Tsuda M, Yasuda K, Senoo-Matsuda N *et al*. (1998). *Nature* **394**: 694–697.
Jeronimo C, Nomoto S, Caballero OL, Usadel H, Henrique R, Varzim G *et al*. (2001). *Oncogene* **20**: 5195–5198.
Johnson MJ, Wallace DC, Ferris SD, Rattazzi MC, Cavalli-Sforza LL. (1983). *J Mol Evol* **19**: 255–271.
Kelley MR, Parsons SH. (2001). *Anti Redox Signaling* **3**: 671–683.
Khogali SS, Mayosi BM, Beattie JM, McKenna WJ, Watkins H, Poulton J. (2001). *Lancet* **357**: 1265–1267.
Kirches E, Krause G, Warich-Kirches M, Weis S, Schneider T, Meyer-Puttlitz B *et al*. (2001). *Int J Cancer* **93**: 534–538.
Krieg RC, Knuechel R, Schiffmann E, Liotta LA, Petricoin III EF, Herrmann PC. (2004). *Proteomics* **4**: 2789–2795.

Kujoth GC, Hiona A, Pugh TD, Someya S, Panzer K, Wohlgemuth SE *et al*. (2005). *Science* **309**: 481–484.
Kurtz A, Lueth M, Kluwe L, Zhang T, Foster R, Mautner VF *et al*. (2004). *Mol Cancer Res* **2**: 433–441.
LaBiche RA, Demars M, Nicolson GL. (1992). *In Vivo* **6**: 317–324.
LaBiche RA, Yoshida M, Gallick GE, Irimura T, Robberson DL, Klostergaard J *et al*. (1988). *J Cellular Biochem* **36**: 393–403.
Lee HC, Li SH, Lin JC, Wu CC, Yeh DC, Wei YH. (2004). *Mutat Res* **547**: 71–78.
Lehtonen R, Kiuru M, Vanharanta S, Sjoberg J, Aaltonen LM, Aittomaki K *et al*. (2004). *Am J Pathol* **164**: 17–22.
Liu VW, Shi HH, Cheung AN, Chiu PM, Leung TW, Nagley P *et al*. (2001). *Cancer Res* **61**: 5998–6001.
Liu VW, Wang Y, Yang HJ, Tsang PC, Ng TY, Wong LC *et al*. (2003). *Hum Mutat* **22**: 173–174.
Majewski N, Nogueira V, Bhaskar P, Coy PE, Skeen JE, Gottlob K *et al*. (2004a). *Mol Cell* **16**: 819–830.
Majewski N, Nogueira V, Robey RB, Hay N. (2004b). *Mol Cell Biol* **24**: 730–740.
Mattiazzi M, Vijayvergiya C, Gajewski CD, DeVivo DC, Lenaz G, Wiedmann M *et al*. (2004). *Hum Mol Genet* **13**: 869–879.
Maximo V, Soares P, Lima J, Cameselle-Teijeiro J, Sobrinho-Simoes M. (2002). *Am J Pathol* **160**: 1857–1865.
McCord JM. (2000). *Am J Med Genet* **108**: 652–659.
Merriwether DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, Jenkins T *et al*. (1991). *J Mol Evol* **33**: 543–555.
Mishmar D, Ruiz-Pesini EE, Golik P, Macaulay V, Clark AG, Hosseini S *et al*. (2003). *Proc Natl Acad Sci USA* **100**: 171–176.
Montiel-Sosa F, Ruiz-Pesini E, Enríquez JA, Marcuello A, Díez-Sánchez C, Montoya J *et al*. (2005). *Gene* (E-Pub ahead of print) doi, 10.1016/j.gene.2005.09.015.
Niemann S, Muller U. (2000). *Nat Genet* **26**: 268–270.
Nomoto S, Yamashita K, Koshikawa K, Nakao A, Sidransky D. (2002). *Clin Cancer Res* **8**: 481–487.
Parrella P, Xiao Y, Fliss M, Sanchez-Cespedes M, Mazzarelli P, Rinaldi M *et al*. (2001). *Cancer Res* **61**: 7623–7626.
Pedersen PL. (1978). *Prog Exp Tumor Res* **22**: 190–274.
Pedersen PL, Mathupala S, Rempel A, Geschwind JF, Ko YH. (2002). *Biochim Biophys Acta* **1555**: 14–20.
Petros JA, Baumann AK, Ruiz-Pesini E, Amin MB, Sun CQ, Hall J *et al*. (2005). *Proc Natl Acad Sci USA* **102**: 719–724.
Polyak K, Li Y, Zhu H, Lengauer C, Willson JK, Markowitz SD *et al*. (1998). *Nat Genet* **20**: 291–293.
Poulton J. (1998). *Trends Genet* **14**: 387–389.
Poulton J, Brown MS, Cooper A, Marchington DR, Phillips DI. (1998). *Diabetologia* **41**: 54–58.
Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC. (2004). *Science* **303**: 223–226.
Ruiz-Pesini E, Wallace DC. (2006). *Hum Mut* (in press).
Schriner SE, Linford NJ, Martin GM, Treuting P, Ogburn CE, Emond M *et al*. (2005). *Science* **308**: 1909–1911; published online 5 May 2005 (10.1126/science.1106653).
Senoo-Matsuda N, Yasuda K, Tsuda M, Ohkubo T, Yoshimura S, Nakazawa H *et al*. (2001). *J Biol Chem* **276**: 41553–41558.
Taira M, Yoshida E, Kobayashi M, Yaginuma K, Koike K. (1983). *Nucl Acids Res* **11**: 1635–1643.
Testa JR, Bellacosa A. (2001). *Proc Natl Acad Sci USA* **98**: 10983–10985.
Tong BC, Ha PK, Dhir K, Xing M, Westra WH, Sidransky D *et al*. (2003). *J Surg Oncol* **82**: 170–173.
Torroni A, Lott MT, Cabell MF, Chen Y, Laverge L, Wallace DC. (1994). *Am J Hum Genet* **55**: 760–776.
Torroni A, Petrozzi M, D'Urbano L, Sellitto D, Zeviani M, Carrara F *et al*. (1997). *Am J Hum Genet* **60**: 1107–11021.
Trifunovic A, Wredenberg A, Falkenberg M, Spelbrink JN, Rovio AT, Bruder CE *et al*. (2004). *Nature* **429**: 417–423.
Trounce I, Neill S, Wallace DC. (1994). *Proc Natl Acad Sci USA* **91**: 8334–8338.
Trounce I, Neill S, Wallace DC. (1995). *Am J Hum Genet* **57**: A252.
Trounce IA, Kim YL, Jun AS, Wallace DC. (1996). *Methods Enzymol* **264**: 484–509.
Vanharanta S, Buchta M, McWhinney SR, Virta SK, Peczkowska M, Morrison CD *et al*. (2004). *Am J Hum Genet* **74**: 153–159.
Wallace DC. (2005). *Annu Rev Genet* **39**: 359–407.
Wallace DC, Brown MD, Lott MT. (1999). *Gene* **238**: 211–230.
Wallace DC, Lott MT. (2002). *Emery and Rimoin's Principles and Practice of Medical Genetics*. In: Rimoin DL, Connor JM, Pyeritz RE, Korf BR (eds). Churchill Livingstone: London, pp 299–409.
Wallace DC, Ruiz-Pesini E, Mishmar D. (2003). *Cold Spring Harbor Symposia Quant Biol* **68**: 479–486.
Wang Q, Ito M, Adams K, Li BU, Klopstock T, Maslim A *et al*. (2004). *Am J Med Genet* **131A**: 50–58.
Wang Y, Michikawa Y, Mallidis C, Bai Y, Woodhouse L, Yarasheski KE *et al*. (2001). *Proc Natl Acad Sci USA* **98**: 4022–4027.
Warburg O. (1931). *The Metabolism of Tumors*. R.R. Smith: New York.
Warburg O. (1956). *Science* **123**: 309–314.
Wu CW, Yin PH, Hung WY, Li AF, Li SH, Chi CW *et al*. (2005). *Genes Chromosomes Cancer* **44**: 19–28.
Yanagihara M, Katano M, Takahashi-Sasaki N, Kimata K, Taira K, Andoh T. (2005). *Cancer Sci* **96**: 620–626.
Yeh JJ, Lunetta KL, van Orsouw NJ, Moore Jr FD, Mutter GL, Vijg J *et al*. (2000). *Oncogene* **19**: 2060–2066.
Yoneyama H, Hara T, Kato Y, Yamori T, Matsuura ET, Koike K. (2005). *Mol Cancer Res* **3**: 14–20.
Zhang J, Asin-Cayuela J, Fish J, Michikawa Y, Bonafe M, Olivieri F *et al*. (2003). *Proc Natl Acad Sci USA* **100**: 1116–1121.