

# mtDNA mutation pattern in tumors and human evolution are shaped by similar selective constraints

Ilia Zhidkov,<sup>1,2</sup> Erez A. Livneh,<sup>2</sup> Eitan Rubin,<sup>2,3,4</sup> and Dan Mishmar<sup>1,2,4</sup>

<sup>1</sup>Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel; <sup>2</sup>National Institute of Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel; <sup>3</sup>Department of Microbiology and Immunology, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

Multiple human mutational landscapes of normal and cancer conditions are currently available. However, while the unique mutational patterns of tumors have been extensively studied, little attention has been paid to similarities between malignant and normal conditions. Here we compared the pattern of mutations in the mitochondrial genomes (mtDNAs) of cancer (98 sequences) and natural populations (2400 sequences). De novo mtDNA mutations in cancer preferentially colocalized with ancient variants in human phylogeny. A significant portion of the cancer mutations was organized in recurrent combinations (COMs), reaching a length of seven mutations, which also colocalized with ancient variants. Thus, by analyzing similarities rather than differences in patterns of mtDNA mutations in tumor and human evolution, we discovered evidence for similar selective constraints, suggesting a functional potential for these mutations.

[Supplemental material is available online at [www.genome.org](http://www.genome.org). The scripts, input files, intermediate files and raw output files are available at <http://bioinfo.bgu.ac.il/rubin/Zhidkov2009/>.]

Natural selection leaves a characteristic signature in the human genome in the form of changes in allele frequency in the population (Bamshad and Wooding 2003; Sabeti et al. 2006). This principle equally applies to natural human populations and to populations of tumor cells, thought to experience selective constraints (Anderson et al. 2006). Tumor cells are thought to present special features, as they are under unique selective constraints designed to favor proliferation and avoid cell death. Moreover, tumors have unique population dynamics, characterized by clonal inheritance, as well as extreme bottlenecks, followed by exponential growth under severe selective conditions (Beerenwinkel et al. 2007). It is possible, therefore, that the selective constraints acting on tumors and inherited variants could differ. Hence, comparing patterns and signatures of selection in the genomes of tumor cells and natural human populations may provide insight into the functional constraints acting on both systems.

Recent advances in sequencing technology and whole genome analysis of polymorphic sites have enabled the identification of a repertoire of genetic landscapes both in global populations and in several types of cancer (International HapMap Consortium 2005; Chimpanzee Sequencing and Analysis Consortium, 2005; Nielsen et al. 2005; Sabeti et al. 2006; Sjoblom et al. 2006; Wang et al. 2006; Weir et al. 2007; Wood et al. 2007). Although most of the mutations comprising natural variation are thought to be neutral, signatures of selection have been identified (Wang et al. 2006). Similarly, tumor genomes mostly harbor neutral (passenger) mutations but also contain a smaller fraction of positively selected mutations that promote tumor progression (Beerenwinkel et al. 2007). Hence, like those genetic variants that define genetic backgrounds in the population, part of the genetic landscapes in tumors could also harbor signatures of selection.

Cancer constitutes a unique system for studying evolution. Comparing the mutational patterns in genomes of tumors to genomes of the corresponding healthy tissues (i.e., the “ancestral genomes” of the tumor) allows for investigation of the mutational outcome of severe selective constraints within the lifetime of humans. Indeed, mutational patterns in tumors have recently attracted attention. In their pioneering work aimed at characterizing the genomic attributes of cancer, Wood and colleagues resequenced most of the human transcriptome in 22 different breast and colorectal cancer tissues, recording de novo mutations (Sjoblom et al. 2006; Wood et al. 2007). Along with meeting their original goal, these sequencing efforts also provided a growing body of data documenting the genetic landscape of tumors at a similar resolution to the reported genetic variation in natural human populations, allowing for identification of signatures of natural selection in tumor genomes.

The characterization of genome-wide signatures of selection in both natural human populations and cancer makes it possible, in principle, to compare selective patterns. Such a comparison would allow for the identification of commonalities in the selective constraints underlying the genetic landscape of both systems. Nevertheless, such comparison is difficult to perform with the nuclear genome, due to it presenting a mosaic of multiple haploblocks, each having evolved independently in different lineages. Moreover, the tremendous efforts required for resequencing the nuclear genome make it difficult to test whether selection acts preferentially on specific mutations or on combinations of mutations, i.e., on de novo genetic backgrounds. Major efforts, such as the 1000 genomes project (<http://www.1000genomes.org/>), may, in the future, assist in overcoming some of these obstacles. However, there are currently not enough data to compare the nuclear genomic landscape of cancer cells to that of corresponding normal individuals.

With this in mind, we now present a new approach that focuses on the mitochondrial genome (mtDNA) and compares the selective signatures formed during rapid (cancer) and slow (human mitochondrial phylogeny) evolutionary processes at the genome scale level. Addressing mtDNA offers several advantages in such

#### <sup>4</sup>Corresponding authors.

E-mail [dmishmar@bgu.ac.il](mailto:dmishmar@bgu.ac.il); fax 972-8-6461356.

E-mail [erubin@bgu.ac.il](mailto:erubin@bgu.ac.il); fax 972-8-6461356.

Article published online before print. Article and publication date are <http://www.genome.org/cgi/doi/10.1101/gr.086462.108>.

studies. First, many years of extensive mtDNA usage in human population genetic studies have generated a tremendous amount of data of the mtDNA genetic variation in many global populations, information far more extensive than current knowledge of nuclear genomic variation. Second, its relative short length (~16.5 kb) makes mtDNA highly amenable for whole-genome sequencing. As it follows a simple uniparental inheritance pattern, mtDNA genomic variation is dominated by mutations rather than by recombination. In addition, mtDNA has been subjected to various types of selection over the course of human evolution (Mishmar et al. 2003; Meiklejohn et al. 2007). Finally, mtDNA shows high sequence variability in tumors (for review, see Brandon et al. 2006; Verma and Kumar 2007). With these attributes in mind, we set out to compare mutational patterns of the mtDNA in cancer and normal conditions.

## Results and Discussion

The mutational landscapes of mtDNA in cancer and natural populations were compared using two mutational compendia, one including mutations that occurred during human evolution and the other containing de novo cancer mutations. The human evolution compendium was created by assembling 2400 publicly available nonidentical mtDNA sequences, containing the entire coding region, into a phylogenetic tree and subsequently testing for the most parsimonious pattern of mutations that fits the tree (Supplemental Table 1). To assemble the de novo cancer mutations compendium, we scanned the literature for appropriate screens in which such mutations were documented, comparing whole mtDNA sequences from tumors and corresponding healthy tissues. Two mtDNA mutation surveys—one involving Head and Neck Squamous Cell Carcinoma (HNSCC) (Zhou et al. 2007) and another involving pancreatic cancer (PANC) (Kassaoui et al. 2006)—were identified as the only studies that had directly compared numerous tumor and adjacent normal tissues over the entire mitochondrial genome, reporting all de novo mutations, including those occurring in known polymorphic sites. Interestingly, both studies described a mutational landscape generally similar to that reported for the nuclear genome, dominated by “hills” of nonrecurring de novo mutations. This view is supported by other, less comprehensive surveys (for review, see Ohta 2006). We combined these data into a single cancer compendium of de novo mtDNA mutations, including 299 de novo point mutations identified in 83 tumor–normal tissue pairs from the HNSCC and 15 pairs from the PANC surveys (Supplemental Table 2).

Remarkable similarities were observed between the cancer and human mitochondrial evolution mutational landscapes. When considering individual positions (i.e., a one-dimensional analysis), de novo mutations in the cancer compendium preferentially colocalized with mutations in the inner nodes rather than the external branches (tips) of the human phylogenetic tree ( $P < 1.4 \times 10^{-25}$ ,  $\chi^2$  test). Hence, cancer mutations preferentially colocalize with variants that were fixed during ancient times (inner nodes) as opposed to recent changes (tips), as previously suggested (Brandon et al. 2006). Moreover, positions mutated in cancer preferentially coincide with ancient variants, as reflected in their maximal depth of mutation fixation (MDF) values. MDF at a given mtDNA nucleotide position was defined as the maximal depth of any branch in which it was mutated (based on a maximum parsimony model), with nonvariable positions and positions mutated in the tree tips having a MDF value of zero (for a complete description of the MDF calculation, see

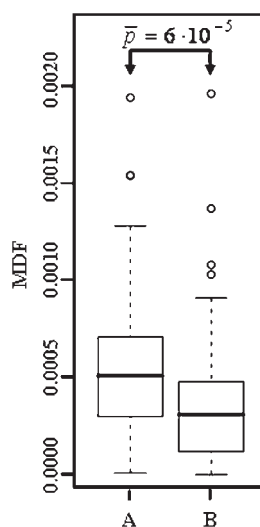
Methods). Comparing the MDF values of all positions harboring de novo cancer mutations in the assembled cancer compendium with those of positions not included in the compendium revealed that de novo cancer mutations preferentially occur at positions involving deeper branches ( $P \leq 1.7 \times 10^{-31}$ , Mann-Whitney test). Moreover, when noncancer positions were sampled, such that they presented precisely the same distribution of natural variability as at those nucleotide positions mutated in our cancer compendium, a significant bias in MDF values toward deeper branches in the cancer positions was observed (Fig. 1). This observation, in agreement with previous predictions (Brandon et al. 2006), suggests that de novo somatic mtDNA mutations in the cancer compendium correlate with ancient mutations in human phylogeny.

It was previously suggested that cancer is caused by combinations of, rather than by single nuclear DNA mutations (Knudson 1971). Therefore, the cancer compendium was scanned for combinations of de novo mtDNA mutations (COMs), defined as two or more mutations that recurred in two or more tumors. A complex pattern of combinations was revealed, with 15 partially overlapping COMs, involving 25 mtDNA positions and 14 individual samples, being identified (Fig. 2A). These COMs could be clustered into three defined patterns (i.e.,  $\alpha$ ,  $\beta$ , and  $\gamma$ ) (Fig. 2B) by combining any two COMs that overlap in at least one position. Most strikingly, two COMs,  $\alpha_1$  and  $\beta_1$ , involved seven different homoplasmic mtDNA mutations, with COM  $\alpha_1$  containing two HNSCC tumor mutations and  $\beta_1$  including mutations seen in HNSCC and pancreatic cancer samples. Six of the seven mutations comprising COM  $\alpha_1$  were shared by an additional HNSCC tumor, thus defining a pattern of six de novo mutations that recurred in three individuals. These patterns represent “ridges” in a multidimensional mutational landscape that are highly unlikely to have occurred by chance. Indeed, reshuffling tests suggest that randomly obtaining even one COM reaching a length of seven mutations is highly unlikely ( $P = 2.9 \times 10^{-4}$ ). In addition, the number of positions involved in the COMs ( $n = 25$ ) was significantly larger than the expected value ( $P = 4 \times 10^{-3}$ ). Hence, rejecting the null hypothesis (i.e., that the observed COMs pattern is a coincidence, potentially resulting from elevated mutation rate at COM positions) suggests that the observed correlation is due to selection. It is noteworthy that the reshuffling test was performed based on the observed mutation rates in cancer and may, therefore, overestimate the length of COMs that are expected to occur by chance. We thus conclude, that the bias in the mutational landscape of mtDNA toward nonrandom recurrence of COMs is best explained by an elevated fitness of cells harboring particular combinations of mutations during the development and/or progression of cancer (Mithani et al. 2007).

Previously, it was proposed that the spectrum of single mtDNA mutations observed in some cancers could be explained by random fixation events (Coller et al. 2001). Our results suggest that when a multidimensional mutational landscape is considered, a significant fraction of the reported mutations are probably under selection and hence bear functional potential. In other studies, the conservation index (CI) was used to assess the functionality of amino acid changes (Ruiz-Pesini et al. 2004) and lineage-defining mutations in mitochondrial RNA genes (Ruiz-Pesini and Wallace 2006). However, only five of 25 and six of 25 of the COM mutations replaced amino acids or occurred in RNA genes, respectively. Intriguingly, nine of the 25 COM mutations are synonymous, and an additional five involve noncoding D-loop mutations. Of these positions, synonymous and D-loop mutations also occurred in the seven-position-long  $\alpha_1$  and  $\beta_1$  COMs. Hence,

in evolutionary terms, a large proportion of the functional potential of COMs could affect fitness without altering protein sequence. This observation supports previous suggestions that synonymous mutations should not be regarded as nonfunctional. Our results thus call for a re-evaluation of the functional potential of common genetic variants, at least in the case of the mtDNA (Chamary et al. 2006).

Having established that the mtDNA multidimensional mutational landscape in cancer is shaped by selection, we aimed to compare this pattern with the mutational landscape created during natural evolution. Our hypothesis was that similarities in the selective forces acting on the two genetic landscapes (i.e., in cancerous and natural populations) would be reflected in a fixation of variants at the same nucleotide positions in both systems. When the compendium of mutations that occurred during human evolution (see above) was screened for the presence of mutations comprising the cancer mtDNA COMs, it was revealed that most (24/25) of the cancer mtDNA mutations corresponding to mtDNA COMs matched parts of natural mtDNA haplogroups (Supplemental Tables 3, 4). Perhaps the most striking example is COM  $\alpha_1$ , involving seven de novo mutations (8701G, 9540C, 10398G, 10873C, 12705T, 15301A, and 16223T). This COM is the exact mirror image (i.e., it presents mutations in the opposite direction) of the combination of variants defining macro-haplogroup R, a lineage encompassing all western Eurasian mtDNA haplogroups.



**Figure 1.** mtDNA positions harboring de novo cancer mutations preferentially occur in deep branches of human phylogenetic tree. Using maximum parsimony to assign mutations to branches of the human mtDNA phylogenetic tree, a maximal depth of fixation (MDF) value was calculated for 3328 mtDNA positions that vary in a compendium of coding region sequences (for description of the compendium, see text). The distributions of MDF values were compared to reported mutations in cancer data sets (A) and to noncancer positions, i.e., positions that were mutated only in the human compendium (B). To control for difference in the variability levels of the two types of positions, a sample of noncancer positions was used, sampled such that their variability levels precisely matched the variability level of the cancer positions. For estimating the statistical significance of the bias toward higher MDF values in cancer positions compared with noncancer positions ( $\bar{p}$ ), the  $P$ -value estimate from a one-sided Mann-Whitney  $U$ -test was averaged over 1000 independently generated variability matched samples. To ensure the repeated sampling is meaningful, highly variable positions (assigned to 13 or more inner nodes) were ignored, since these could not be matched between the cancer and noncancer sets.

Moreover, it appears that other COMs also recapitulate major events in human mitochondrial phylogeny. COM  $\beta_1$  perfectly recapitulates those mutations defining haplogroup U4, while COM  $\alpha_6$  reconstructs the exact path from haplogroup H to haplogroup V, skipping the intermediate stage observed in the HV cluster. Interestingly, some of the haplogroups that overlap with COMs were associated with phenotypes. Cybrids carrying haplogroup N (overlapping with COM  $\alpha_1$ ) differed from haplogroup M cybrids in calcium uptake, while haplogroups H (COM  $\alpha_6$ ) and U4 (COM  $\beta_1$ ) are associated with high and low sperm motility, respectively (Ruiz-Pesini et al. 2000; Montiel-Sosa et al. 2006). These results support our working hypothesis, suggesting that similar selective forces act on cancer and during human mitochondrial evolution. Moreover, the observation that several COMs occur at positions defining haplogroups associated with pathological phenotypes further suggests that COMs possess functional potential. While it is tempting to propose that some human haplogroups carry mutations that alter their propensity to cancer, the functional consequences of COMs are most likely affected by the genetic backgrounds upon which they occurred, a hypothesis that will be tested upon the creation of the relevant data sets.

Finally, we noticed that tumors containing COMs comprise only 24% of tumors for which any de novo mtDNA mutations were reported in our tested compendium and only 14% of all tumors tested. What could be the reason for this apparent paucity? One possible explanation is that some tumors have followed adaptation routes that do not involve mtDNA changes at all. Accordingly, 41 of the 83 tumors analyzed from the HNSCC group completely lacked de novo mtDNA mutations. Alternatively, some of the COM-free tumors may have originated from individuals that inherited cancer-compatible mtDNA variants to begin with. Indeed, inherited mtDNA variants have been proposed to play a role in cancer susceptibility (Canter et al. 2005; Wang et al. 2006; Darvishi et al. 2007), although some of these associations have been recently questioned (Setiawan et al. 2008). The possibility that inherited COMs occur in cancer patients will be tested by examining the complete mtDNA sequences of the normal tissues analyzed in Zhou et al. (2007) and Kassaoui et al. (2006), to be deposited in public databases in the near future (J.A. Califano, pers. comm.). A third possible explanation for the paucity of COMs is that the mutations reported for some of the COM-free tumors represent “half COMs,” harboring de novo mtDNA mutations which were cofixed with unreported mutations, such as mutations in the nuclear genome or inherited mtDNA variants. This possibility further underlines the need to include mitochondrial genomic sequences in genome-wide mutational landscape surveys.

In summary, our comparisons of de novo cancer mtDNA mutations in two types of tumors to mutations that have been fixed during human evolution have revealed notable similarity in terms of their mutational landscapes. We present a novel approach for the analysis of mutational landscapes, i.e., the search for COMs, and present complex and nonrandom patterns in the de novo cancer mutational pattern. This mutational pattern is more likely to reflect a response to positive selection pressures in cancer, rather than random nonadaptive processes. Many of the COMs recapitulate combinations of mutations that occurred during human mitochondrial evolution, leading to the establishment of mtDNA haplogroups, suggesting a similarity in the patterns of mutations in normal human evolution and cancer. Thus, our findings pave the path toward the development of new approaches for assessing the functional potential of naturally occurring changes and for investigating basic principles in cancer and natural genomic evolution.

## A.

Position	94	630	16311	7424	2352	8101	9540	10398	10873	12705	15301	16223	2706	7028	15904	1580	4639	16334	12308	12372	1811	4646	14620	15693	16356	
HNSCC #2051			T*	G*																						
HNSCC #1736			C																							
HNSCC #2828				G		G	C	G	C	T	A	T	G	T												
HNSCC #2714					G	C	G			T	A	T														
HNSCC #1680					C*	G*		G*						G*	T*											
HNSCC #1691								G*						G	T	T*	A*									
HNSCC #2455														G	T	T	A	C								
HNSCC #2555															C											
HNSCC #2043																			G	A						
Pancreas #14																		T	G	A	G	C	T	C	C	
HNSCC #1836																		G	A	G	C	T	C	C		
Pancreas #11																	T*				G			C*		
Pancreas #9	A*	T*																								
Pancreas #12	A*	T*																								

## B.

COM	Number of sequences in control data set	# Positions in mtDNA and corresponding bases							
$\alpha_1$	654	8*01G	9540C	10398G	10873C	12705T	15301A	16223T	
$\alpha_2$	656	8*01G	9540C	10398G	12705T	15301A	16223T		
$\alpha_3$	34	2352C	8*01G	10398G					
$\alpha_4$	9	7424G	12705T						
$\alpha_5$	766	12705T	16311C						
	203	12705C	16311T						
$\alpha_6$	57	2706G	4580A	7028T	15904T				
$\alpha_7$	754	2706G	7028T	8*01G	10398G				
$\alpha_8$	1070	2706G	7028T	10398G					
$\alpha_9$	1907	2706G	7028T						
$\alpha_{10}$	467	4639T	7028C						
	14	4639C	7028T						
$\beta_1$	6	1811G	4646C	12308G	12372A	14620T	15693C	16356C	
$\beta_2$	12	1811G	15693C						
$\beta_3$	287	12308G	12372A						
$\beta_4$	0	1811G	15693C	16134C					
$\gamma_1$	0	94A	630T						

**Figure 2.** Recurrent tumor-related mutation combinations of HNSCC and pancreatic tumors. Tumor-related mutations reported in HNSCC (Zhou et al. 2007) and pancreatic cancer patients (Kassauei et al. 2006) were scanned for COMs, defined as recurrent mutation of at least two positions in two different patients. (A) Mutations involved in the COMs. For each tumor (left column), the cancer type and patient identifier are provided as shown in the above-mentioned references. Nucleotide positions of mutations are according to the revised Cambridge reference sequence (GenBank AC\_000021.2). Heteroplasmic mutations are marked with an asterisk (\*). (B) A proposed naming convention for COMs. Overlapping COMs are considered to be part of a larger COM group, designated using Greek letters ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ). In particular, COMs that belong to a certain group are named using a Greek letter in combination with a number.

## Methods

## Sequences and haplotype assignment

A data set of 2655 human mtDNA coding-region sequences (~15.5 kb, nucleotide positions 577–16023) was compiled from Mitomaster (Brandon et al. 2009) and from unpublished sequences generated in our lab. For sequences generated in our lab, haplogroup assignment was performed as described previously (Feder et al. 2008). Upon removing redundant (i.e., 100% identical) sequences ( $n = 245$ ), 2400 sequences were retained (Supplemental Table 1). The D-loop region was clipped from sequences for which it was available to ensure equivalent coverage of the mtDNA.

## Multiple sequence alignment and phylogenetic reconstruction

A multiple sequence alignment was generated, after removal of insertions and deletions, followed by the creation of a neighbor-joining (NJ) phylogenetic tree using the *Pan troglodytes* (NC\_001643) mtDNA sequence as an outgroup, utilizing the distance matrix program (PHYLIP package; <http://evolution.genetics.washington.edu/phylip.html>). Manual inspection of the resulting tree showed no significant deviations from accepted human mtDNA phylogeny at the haplogroups level, and only minor deviations from other published trees (Ruiz-Pesini et al. 2007).

## Assigning sequence changes to specific tree branches

Dnapars (<http://evolution.genetics.washington.edu/phylip/doc/dnapars.html>) was used to infer mtDNA sequences at the inner nodes of the NJ phylogenetic tree described above. The Dnapars program was chosen as it provides a model that minimizes the number of mutational events. The results of this analysis were analyzed using a Perl/BioPerl script developed for this purpose in our group. We designed a script, MDFCALC, that parses the Dnapars output, identifies mutational events, assigns them to specific branches, and calculates their MDF values (see below). This script first resolves underdetermined mutational events, i.e., events in which multiple patterns equally explain the underlying sequences. In these cases, Dnapars uses IUPAC codes for multiple nucleotides to represent all possible solutions (i.e., ambiguous bases). These ambiguities are resolved by randomly assigning one of the possible bases to the first mutational event, and adjusting the tree to replace all other IUPAC codes, accordingly.

## Calculation of the maximal depth of fixation

We define maximal depth of fixation (MDF) as the depth of the deepest inner node in which a nucleotide position was mutated. The depth of all the nodes in which a tested position mutated was calculated as the median distance from the node to its terminal descendants. The MDF was taken to be the maximal depth for any node in which the position was mutated. These calculations were carried out by the MDFCALC script described above.

## Identification of mutational combinations and their statistical analysis

COM identification was performed using ComsFinder, a PERL script written for this purpose. ComsFinder takes a table of mutation-individual pairs as input, finds COMs (for definition, see

text), and reports all the mutations found at COM positions (Fig. 1). Statistical testing is also carried out by ComsFinder, applying the following procedure: The input file was converted to a binary matrix,  $D$ , comprising 299 columns over 56 rows, in which the columns represent mtDNA positions that were de novo mutated in cancer and the rows represent the samples constituting our tumor compendium. The value of each cell, either 1 or 0, reflects the presence or absence, respectively, of a de novo mutation in the tumor sample. From this matrix,  $l_{\max}(D)$  and  $k_{\max}(D)$  were calculated, representing the maximal number of positions and the maximal number of samples involved in any COM within  $D$ , respectively. To test the null hypothesis that  $l_{\max}(D)$  and  $k_{\max}(D)$  are not larger than expected by chance, assuming independence of fixation events, a collection of  $10^7$  randomized matrices ( $D'$ ) was created, in which each matrix  $d$  is derived from  $D$  by column reshuffling. For each of the reshuffled matrices,  $d_i$ , the  $l_{\max}(d_i)$  and  $k_{\max}(d_i)$  were calculated as for the un-shuffled matrix. The probability of the null hypothesis was estimated from the fraction of reshuffled matrices,  $d$ , for which  $l_{\max}(D) \leq l_{\max}(d)$  or  $k_{\max}(D) \leq k_{\max}(d)$ . As the background fixation rate for each position is directly derived from the data, cofixation events will increase the estimated rate of individual fixation, and thus possibly the expected rate of cofixation.

## Acknowledgments

We thank the National Institute of Biotechnology in the Negev (NIBN) for funding (D.M., E.R.) and the Center of Complexity Science for a scholarship given to I.Z. We also thank Prof. Jerry Eichler (BGU) for critical reading of the manuscript. We also thank Dr. Tal Pupko, Tel Aviv University, for critical discussions of the manuscript. This work was also funded by grants from the Israeli Cancer Association (D.M., E.R.) and from the Israeli Science Foundation (D.M.).

## References

Anderson, A.R., Weaver, A.M., Cummings, P.T., and Quaranta, V. 2006. Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. *Cell* **127**: 905–915.

Bamshad, M. and Wooding, S.P. 2003. Signatures of natural selection in the human genome. *Nat. Rev. Genet.* **4**: 99–111.

Beerenwinkel, N., Antal, T., Dingli, D., Traulsen, A., Kinzler, K.W., Velculescu, V.E., Vogelstein, B., and Nowak, M.A. 2007. Genetic progression and the waiting time to cancer. *PLoS Comput. Biol.* **3**: e225. doi: 10.1371/journal.pcbi.0030225.

Brandon, M., Baldi, P., and Wallace, D.C. 2006. Mitochondrial mutations in cancer. *Oncogene* **25**: 4647–4662.

Brandon, M.C., Ruiz-Pesini, E., Mishmar, D., Procaccio, V., Lott, M.T., Nguyen, K.C., Spolim, S., Patil, U., Baldi, P., and Wallace, D.C. 2009. MITOMASTER: A bioinformatics tool for the analysis of mitochondrial DNA sequences. *Hum. Mutat.* **30**: 1–6.

Canter, J.A., Kallianpur, A.R., Parl, F.F., and Millikan, R.C. 2005. Mitochondrial DNA G10398A polymorphism and invasive breast cancer in African-American women. *Cancer Res.* **65**: 8028–8033.

Chamary, J.V., Parmley, J.L., and Hurst, L.D. 2006. Hearing silence: Non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.* **7**: 98–108.

Chimpanzee Sequencing and Analysis Consortium, C.S.a.A. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**: 69–87.

Coller, H.A., Khrapko, K., Bodyak, N.D., Nekhaeva, E., Herrero-Jimenez, P., and Thilly, W.G. 2001. High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection. *Nat. Genet.* **28**: 147–150.

Darvishi, K., Sharma, S., Bhat, A.K., Rai, E., and Bamezai, R.N. 2007. Mitochondrial DNA G10398A polymorphism imparts maternal Haplogroup N a risk for breast and esophageal cancer. *Cancer Lett.* **249**: 249–255.

Feder, J., Blech, I., Ovadia, O., Amar, S., Wainstein, J., Raz, I., Dadon, S., Arking, D.E., Glaser, B., and Mishmar, D. 2008. Differences in mtDNA

haplogroup distribution among 3 Jewish populations alter susceptibility to T2DM complications. *BMC Genomics* **9**: 198. doi: 10.1186/1471-2164-9-198.

International HapMap Consortium 2005. A haplotype map of the human genome. *Nature* **437**: 1299–1320.

Kassaei, K., Habbe, N., Mullendore, M.E., Karikari, C.A., Maitra, A., and Feldmann, G. 2006. Mitochondrial DNA mutations in pancreatic cancer. *Int. J. Gastrointest. Cancer* **37**: 57–64.

Knudson Jr., A.G. 1971. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci.* **68**: 820–823.

Meiklejohn, C.D., Montooth, K.L., and Rand, D.M. 2007. Positive and negative selection on the mitochondrial genome. *Trends Genet.* **23**: 259–263.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D., et al. 2003. Natural selection shaped regional mtDNA variation in humans. *Proc. Natl. Acad. Sci.* **100**: 171–176.

Mithani, S.K., Taube, J.M., Zhou, S., Smith, I.M., Koch, W.M., Westra, W.H., and Califano, J.A. 2007. Mitochondrial mutations are a late event in the progression of head and neck squamous cell cancer. *Clin. Cancer Res.* **13**: 4331–4335.

Montiel-Sosa, F., Ruiz-Pesini, E., Enriquez, J.A., Marcuello, A., Diez-Sanchez, C., Montoya, J., Wallace, D.C., and Lopez-Perez, M.J. 2006. Differences of sperm motility in mitochondrial DNA haplogroup U sublineages. *Gene* **368**: 21–27.

Nielsen, R., Bustamante, C., Clark, A.G., Glanowski, S., Sackton, T.B., Hubisz, M.J., Fedel-Alon, A., Tanenbaum, D.M., Civallo, D., White, T.J., et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**: e170. doi: 10.1371/journal.pbio.0030170.

Ohta, S. 2006. Contribution of somatic mutations in the mitochondrial genome to the development of cancer and tolerance against anticancer drugs. *Oncogene* **25**: 4768–4776.

Ruiz-Pesini, E. and Wallace, D.C. 2006. Evidence for adaptive selection acting on the tRNA and rRNA genes of human mitochondrial DNA. *Hum. Mutat.* **27**: 1072–1081.

Ruiz-Pesini, E., Lapena, A.C., Diez-Sanchez, C., Perez-Martos, A., Montoya, J., Alvarez, E., Diaz, M., Urries, A., Montoro, L., Lopez-Perez, M.J., et al. 2000. Human mtDNA haplogroups associated with high or reduced spermatozoa motility. *Am. J. Hum. Genet.* **67**: 682–696.

Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V., and Wallace, D.C. 2004. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* **303**: 223–226.

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

Sabeti, P.C., Schaffner, S.E., Fry, B., Lohmueller, J., Varily, P., Shamovsky, O., Palma, A., Mikkelsen, T.S., Altshuler, D., and Lander, E.S. 2006. Positive natural selection in the human lineage. *Science* **312**: 1614–1620.

Setiawan, V.W., Chu, L.H., John, E.M., Ding, Y.C., Ingles, S.A., Bernstein, L., Press, M.F., Ursin, G., Haiman, C.A., and Neuhausen, S.L. 2008. Mitochondrial DNA G10398A variant is not associated with breast cancer in African-American women. *Cancer Genet. Cytogenet.* **181**: 16–19.

Sjblom, T., Jones, S., Wood, L.D., Parsons, D.W., Lin, J., Barber, T.D., Mandelker, D., Leary, R.J., Ptak, J., Silliman, N., et al. 2006. The consensus coding sequences of human breast and colorectal cancers. *Science* **314**: 268–274.

Verma, M. and Kumar, D. 2007. Application of mitochondrial genome information in cancer epidemiology. *Clin. Chim. Acta* **383**: 41–50.

Wang, E.T., Kodama, G., Baldi, P., and Moyzis, R.K. 2006. Global landscape of recent inferred Darwinian selection for *Homo sapiens*. *Proc. Natl. Acad. Sci.* **103**: 135–140.

Weir, B.A., Woo, M.S., Getz, G., Perner, S., Ding, L., Beroukham, R., Lin, W.M., Province, M.A., Kraja, A., Johnson, L.A., et al. 2007. Characterizing the cancer genome in lung adenocarcinoma. *Nature* **450**: 893–898.

Wood, L.D., Parsons, D.W., Jones, S., Lin, J., Sjblom, T., Leary, R.J., Shen, D., Boca, S.M., Barber, T., Ptak, J., et al. 2007. The genomic landscapes of human breast and colorectal cancers. *Science* **318**: 1108–1113.

Zhou, S., Kachhap, S., Sun, W., Wu, G., Chuang, A., Poeta, L., Grumbine, L., Mithani, S.K., Chatterjee, A., Koch, W., et al. 2007. Frequency and phenotypic implications of mitochondrial DNA mutations in human squamous cell cancers of the head and neck. *Proc. Natl. Acad. Sci.* **104**: 7540–7545.

Received September 11, 2008; accepted in revised form February 3, 2009.



## mtDNA mutation pattern in tumors and human evolution are shaped by similar selective constraints

Iliia Zhidkov, Erez A. Livneh, Eitan Rubin, et al.

*Genome Res.* 2009 19: 576-580 originally published online February 10, 2009

Access the most recent version at doi:[10.1101/gr.086462.108](https://doi.org/10.1101/gr.086462.108)

---

**Supplemental Material** <http://genome.cshlp.org/content/suppl/2009/03/09/gr.086462.108.DC1>

**References** This article cites 32 articles, 10 of which can be accessed free at:  
<http://genome.cshlp.org/content/19/4/576.full.html#ref-list-1>

### License

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

Affordable, Accurate  
Sequencing.



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>