Mutation Research 709–710 (2011) 15–20

Contents lists available at ScienceDirect

Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis

journal homepage: www.elsevier.com/locate/molmut
Community address: www.elsevier.com/locate/mutres

Rapid identification of mtDNA somatic mutations in gastric cancer tissues based on the mtDNA phylogeny

Rui Bi[a,c,1], Wen-Liang Li[b,1], Ming-Qing Chen[b], Zhu Zhu[b], Yong-Gang Yao[a,*]

[a] Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, 32 Jiaochang Donglu, Kunming, Yunnan 650223, China
[b] Department of Oncology, the First Affiliated Hospital of Kunming Medical College, Kunming, Yunnan 650032, China
[c] Graduate School of the Chinese Academy of Sciences, Beijing 100039, China

ARTICLE INFO

Article history:
Received 4 January 2011
Received in revised form 15 February 2011
Accepted 16 February 2011
Available online 16 March 2011

Keywords:
mtDNA
Somatic mutations
Cancer
Phylogeny

ABSTRACT

Mitochondrial DNA (mtDNA) somatic mutations have been identified in nearly all kinds of cancer during the past decade. Normally one need to determine the complete mtDNA sequences from both cancerous and normal tissues of the same patient to score the somatic mutation in cancer. In this study, we intended to explore a strategy to quickly identify somatic mutations in the entire mtDNA genome based on its phylogeny. The principal assumption for this strategy is that somatic mutations, as recently accumulated in cancerous tissue, have younger age and will be located in the terminal branches of mtDNA phylogenetic tree. In contrast, the haplogroup-specific variants, which appear as germline variants and have ancient age, will be located in the basal or intermediate-node branches of the tree, depending on their relative age. When the complete mtDNA sequence of the cancerous tissue is determined and is classified relative to the available mtDNA phylogeny, we only need to screen the variants that are located in the terminal branch in the paracancerous tissue or other normal tissue from the same patient to identify somatic mutations in cancer. We validated this strategy by using paired gastric cancer tissue and paracancerous tissue or blood from 10 Chinese patients (including one with gastric stromal tumor). A total of seven somatic mutations were identified in the cancerous tissues from four patients. Our result suggests that employing mtDNA phylogenetic knowledge facilitates rapid identification of mitochondrial genome somatic mutations in cancer.

1. Introduction

Since the Warburg effect was proposed over half a century ago [1], mitochondrial dysfunction in cancer has drawn a lot of concerns. Mitochondrial dysfunction may contribute to tumorigenesis due to its important roles in cell metabolism, energy production and apoptosis [2]. The properties of limited DNA repair mechanism, high oxidative stress, and lacking histone protein protection make mtDNA to be susceptible to mutations [3,4]. For these reasons, the relationship between mtDNA mutations and cancer became a hot topic during the past decade [5–7]. Previous studies paid more attention to the somatic mutations in mtDNA control region and did not analyze the mutations in coding region [8,9]. In addition, problems in data caused by sequencing errors, contamination or sample mix-up are almost ubiquitous, which lead to overestimation of mtDNA somatic mutations in cancer [10]. All these problems can be attributed to the deficiency of mtDNA phylogenetic knowledge when analyzing sequencing data [11,12].

Because of high mutation rate and lack of recombination, mtDNA accumulates mutations in a time order within a relatively short time and form different haplotypes [13]. In the process of human evolution and migration, descendants of an original mtDNA haplotype would sequentially accumulate additional mutations to generate a group of related haplotypes that was termed as a haplogroup [14–16]. Current global mtDNA phylogenetic tree is constituted by different haplogroups, which present strikingly region-specific distribution patterns [15–18].

In principle, every newly sequenced mtDNA can be positioned into one of the branches of the mtDNA phylogenetic tree according to the variants it contains [11]. The nucleotide substitutions that define a specific haplogroup are called haplogroup-specific (or haplogroup-diagnostic, or haplogroup-defining) variants and these variants are of relatively ancient age, while mutations that are located in the terminal branch of mtDNA tree are called private mutations and are recently accumulated in the molecule. Using the phylogenetic approach, one can reconstruct the time order of the mutations that were accumulated on mtDNA sequence and distin-

* Corresponding author. Tel.: +86 871 5180085; fax: +86 871 5180085.
E-mail addresses: ygyaozh@gmail.com, ygyaozh@yahoo.com (Y.-G. Yao).
[1] These authors contributed equally to this work.

0027-5107/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.mrfmmm.2011.02.016

Table 1
Clinical information for the patients with gastric cancer.

Patient	Age	Sex[a]	Cancer staging [stage][b]	Clinical expression
P1	67	M	PT_{4a} (G_{2-3}) N_2 (3/16) M_0 [IIIb]	Hiccup
P2	61	M	PT_3 (G_{2-3}) N_2 (3/16) M_0 [IIIa]	Upper gastrointestinal bleeding, abdominal pain
P3	73	M	PT_3 (G_3) N_1 (1/18) M_0 [IIb]	Gastrointestinal bleeding, anemia
P4	57	F	PT_3 (G_{1-2}) N_0 (0/12) M_0 [IIa]	Fundus of stomach occupying lesion
P5	68	M	PT_{4a} (G_2) N_2 (16/38) M_0 [IIIb]	Abdominal pain
P6	71	M	PT_{4a} (G_{2-3}) N_2 (5/16) M_0 [IIIb]	Hiccup
P7	40	M	PT_3 (G_3) N_{3a} (12/14) M_0 [IIIb]	Upper gastrointestinal bleeding
P8	60	M	gastric stromal tumor	Upper gastrointestinal bleeding
P9	52	M	PT_{4a} (G_{2-3}) N_{3a} (12/20) M_0 [IIIc]	Abdominal distention, vomit, upper gastrointestinal bleeding
P10	44	M	PT_3 (G_{2-3}) N_2 (4/37) M_0 [IIIa]	Abdominal pain and distension

[a] "M" indicates male, "F" indicates female.
[b] Stage of cancer was classified following the 7th edition of AJCC Cancer Staging Handbook [21]. Patient P8 was diagnosed as gastric stromal tumor, which was not a malignant tumor.

guish young mutations from the ancestral variants. Based on a fact that the accumulated variants in mtDNA sequences are subjected to selection and this effect will ultimately eliminate the deleterious mutations from the population, potential pathogenic mtDNA mutations are most likely to be acquired rather recently and appear as private mutation(s) in the lineage [19]. Therefore, phylogenetic approach is useful in assessing the function of mtDNA mutations in disease [15,16,20].

Here, we propose a strategy on the basis of mtDNA phylogenetic knowledge that somatic mutations are located in the terminal branch(es) of mtDNA phylogenetic tree, which can facilitate the identification of cancerous mtDNA somatic mutations in the entire mtDNA genome efficiently. When the entire mtDNA sequence and phylogenetic status of tumor tissue have been determined, it is only necessary to detect the following mutations/variants that are located in the terminal braches of mtDNA tree in the paracancerous tissue under strict quality control to distill somatic mutations: (1) private mutations; (2) haplogroup-specific variants which are back mutations or heteroplasmic. By using this strategy, we investigated mtDNA somatic mutations in the cancerous tissue samples and corresponding paracancerous tissues or blood tissues from 10 Chinese patients with gastric cancer (including one with gastric stromal tumor). As a result, we found seven heteroplasmic somatic mutations in gastric cancer tissues from four patients. Our result suggests that mtDNA phylogenetic knowledge is very useful when investigating somatic mutations in cancer.

2. Materials and methods

2.1. Samples

Nine gastric cancer tissues (GC1–GC7, GC9–GC10) and the corresponding adjacent normal tissues (GN1–GN3, GN5–GN7, GN9–GN10) or whole blood tissues (GB4, GB6) from nine patients (P1–P7, P9–P10), which were diagnosed as gastric cancer at the First Affiliated Hospital of Kunming Medical College, were collected with patients' informed consent. The tumorous tissue (GC8) and adjacent normal tissue (GN8) of a patient with gastric stromal tumor (P8) were also analyzed in this study (Table 1). The stage of cancer was classified on the basis of the 7th edition of AJCC Cancer Staging Handbook [21]. Whole genomic DNA was extracted by the TIANamp Genomic DNA Kit (Tiangen Bio Inc., Beijing, China). The institutional review boards of the Kunming Institute of Zoology and the First Affiliated Hospital of Kunming Medical College approved this study.

2.2. Entire mtDNA sequence analysis of cancerous tissue

The complete mtDNA sequence of each cancerous tissue was amplified by 9 pairs of primers and sequenced as described in our previous studies [22,23]. Sequence variants in each mtDNA were scored relative to the revised Cambridge Reference Sequence (rCRS) [24]. Haplogroup status of each mtDNA was determined according to the available phylogenetic tree of global human mtDNA (http://phylotree.org/tree/main.htm; mtDNA tree Build 11, 7 Feb 2011) [17] and was checked by MitoTool (www.mitotool.org) [25]. The phylogenetic relationship between the mtDNAs from different tissues was demonstrated by a schematic tree, following the same approach in our recent studies [20,26,27].

2.3. Somatic mutation analysis

The status of private mutations (which are located in the terminal branches of the phylogenetic tree) and haplogroup-specific variants, whether presented as either heteroplasmic or back mutations in cancerous tissue, was detected in corresponding paracancerous tissue or whole blood sample from the same donor. Variants that were not detected in the (putatively) normal tissues were regarded as somatic mutations in tumor tissue. We employed the sequencing primers reported in our recent study [22] to direct sequence the region containing certain somatic mutation(s). We defined the uniqueness of mtDNA variants/mutations in certain mtDNA by an exhaustive database search following the available guidelines [28]. The frequency of each mutation was calculated as the number of occurrences of the mutation in 2196 complete mtDNA sequences that were summarized by Soares et al. [29]. Evolutionary conservation analysis for certain mtDNA variant was performed using the MitoTool [25].

2.4. Quality control

We employed the following strategies to avoid sequencing errors and potential contamination: (1) genomic DNA from tissues of each patient was isolated independently. Each sample was amplified and sequenced independently; (2) the mtDNA control region (regions 16024–16569 and 1–576) of each sample was sequenced and analyzed following our previously reported method [27] to discern potential sample contamination and to confirm the paired samples from the same patient. Sample contamination can be recognized by the presence of heteroplasmy at nearly all the sites bearing germline variations of the patient. Unpaired samples can be distinguished by their different haplogroup status [30]; (3) we examine the original sequencing chromatograms site by site to identify heteroplasmic mutations; (4) we followed the available guideline to eliminate potential pseudo mtDNA gene [31]; (5) suspicious mutations were double-strand sequenced or re-sequenced in both cancerous and paracancerous tissues.

3. Results

The paired samples from the same donor were confirmed to have no problem for sample crossover by sequencing the mtDNA control region of cancerous tissues and corresponding paracancerous tissues or whole blood (Supplementary Table 1). Two variants, *m.16092T>C* and *m.204T>C*, identified in samples GC6, GN6 and GB6 from donor 6, were all located in the hypervariable segments of the mtDNA control region and showed different level of heteroplasmy in different tissues, suggested being inherited or occurred as a *de novo* somatic mutation (Supplementary Fig. 1). The observation of fewer somatic mutations in mtDNA control region was consistent with our previous report for a low level of mtDNA control region somatic mutations in early stage breast cancer [26].

Fig. 1 shows a phylogenetic tree of the 10 mtDNAs from the cancerous tissues (sequences have been deposited in GenBank under Accession No. HQ713443–HQ713452). According to the haplogroup-specific variants in each mtDNA and the available mtDNA phylogenetic tree [17], these mtDNAs could be classified into haplogroups B4a1, B4g, F1a1, F1a1a1, F1b1, F2a, F3a, A, A4 and D5a2a1, respectively. All the haplogroup-specific variants in cancerous tissues are homoplasmic. We screened the mutations/variants in the terminal branches and haplogroup-specific

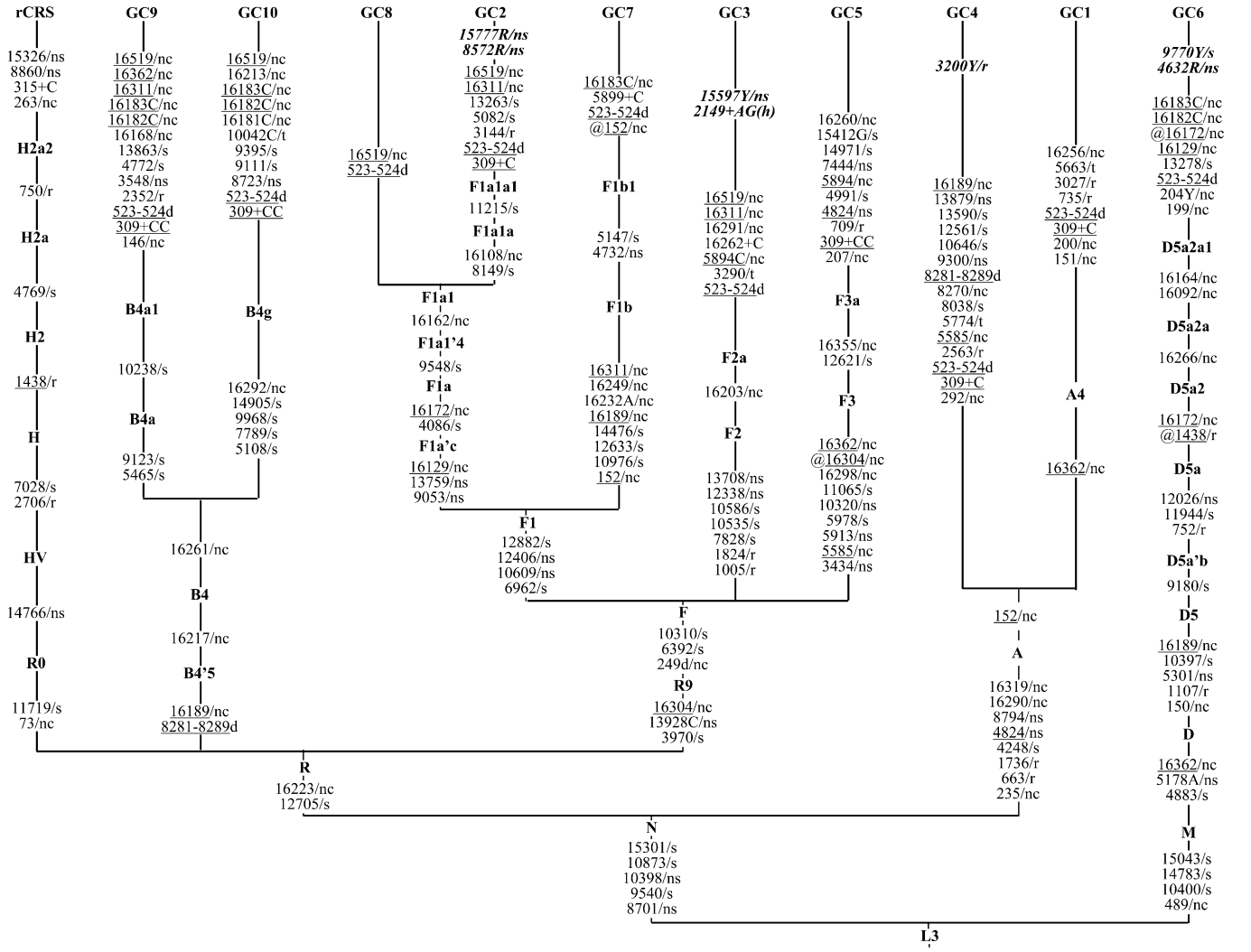


Fig. 1. Haplogroup tree of the complete mtDNA sequences of 10 gastric cancer tissues (including one gastric stromal tumor tissue (GC8)), plus the revised Cambridge reference sequence (rCRS) [24]. Somatic mutations are italicized and in bold face. "d" indicates a deletion; "+" indicates an insertion; "s" indicates a synonymous mutation; "ns" indicates a non-synonymous mutation; "r" indicates the mutation/variant occurs in the rRNA genes; "t" indicates the mutation/variant occurs in the tRNA genes; "nc" indicates the mutation/variant occurs in the non-coding region; suffixes A, C, G mean transversions; suffix "Y" indicates a heteroplasmy of T and C; suffix "R" indicates a heteroplasmy of A and G; recurrent mutations are underlined; back mutations are underlined and marked "@". Sample GC3 presents a heteroplasmic status of insertion AG at site 2149, and is marked as "2149+AG(h)".

Table 2
Somatic mtDNA mutations in gastric cancer tissues of 4 Chinese patients.

Sample	Haplogroup	Mutation (amino acid change)	Gene	Reported (population context)[a]	Reported (disease context)[a]	Haplogroup specific variant[b]	Variant frequency[c]	Conservation Index (CI)[d]
GC2	F1a1a1	*m.8572G>A* (synonymous)	*MT-ATP8*	Yes	No	Yes (H1s, M33a1a)	6/2196	1
		m.8572 G>A (p.G16S)	*MT-ATP6*	Yes	No			0.744
		m.15777G>A (p.S344N)	*MT-CYB*	Yes	Yes	Yes (L1c1a1b)	4/2196	0.302
GC3	F2a	*m.2148_2149dupAG*	*MT-RNR2*	Yes	No	No	0/2196	1
		m.15597T>C (p.V284A)	*MT-CYB*	No	No	No	0/2196	0.512
GC4	A	*m.3200T>C*	*MT-RNR2*	Yes	Yes	Yes (M20)	3/2196	0.116
GC6	D5a2a1	*m.4632G>A* (p.A55T)	*MT-ND2*	No	No	No	0/2196	1
		m.9770T>C (synonymous)	*MT-CO3*	Yes	No	Yes (U3b2)	2/2196	0.977

[a] The search was performed on 11 February 2011 according to the described strategy [28] (e.g. both "G8572A mtDNA" and "8572G>A mtDNA" were queried).
[b] The presence of the mutations as a haplogroup-specific variant were scored relative to the available global mtDNA phylogenetic tree at the Phylotree (http://phylotree.org/tree/main.htm; mtDNA tree Build 11, 7 Feb 2011) [17]. In round brackets we indicated the haplogroup status as it defined in that tree.
[c] The mutation frequency was calculated as the number of occurrences of each mutation in 2196 complete mtDNA sequences that were summarized by Soares et al. [29].
[d] The conservation index was defined by Ruiz-Pesini et al. [32] and was computed by using MitoTool (http://www.mitotool.org/) [25]. A total of 43 primate species were considered and a CI value of 0.744 meant 74.4% of 43 primate species have the wide type allele with human sequence (GenBank Accession number NC_012920).

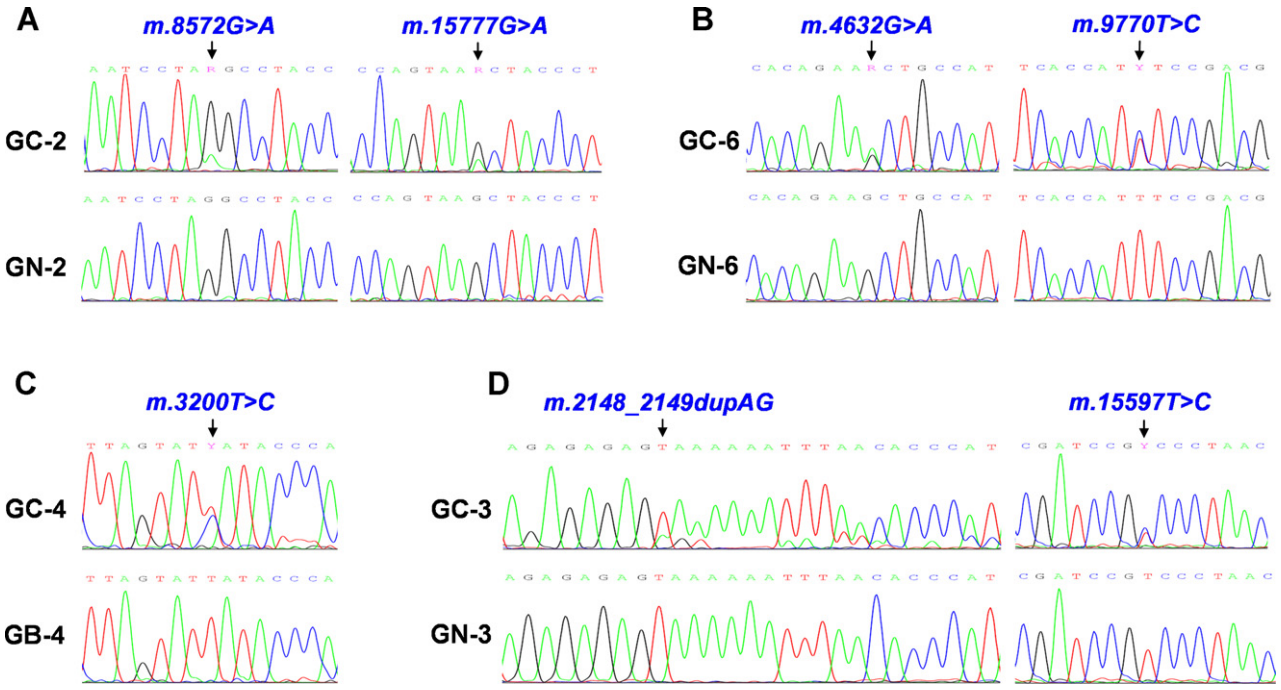


Fig. 2. Somatic mutation status in different tissues of 4 gastric cancer patients. "GC" means gastric cancer tissue; "GN" means adjacent normal tissue; "GB" means blood tissue. (A) The cancerous tissue of patient P2 contains two heteroplasmic mtDNA somatic mutations: *m.8572G>A* (left) and *m.15777G>A* (right). (B) The cancerous tissue of patient P6 contains two heteroplasmic mtDNA somatic mutations: *m.4632G>A* (left) and *m.9770T>C* (right). (C) The cancerous tissue of patient P4 contains one heteroplasmic mtDNA somatic mutation *m.3200T>C*. (D) The cancerous tissue of patient P3 contains two heteroplasmic mtDNA somatic mutations: *m.2148_2149dupAG* (left) and *m.15597T>C* (right).

variants which undergone back mutations in corresponding paracancerous tissue or blood sample from the same patient. In total, seven somatic mutations were identified in 4 gastric cancer tissues (Table 2 and Fig. 2). All these somatic mutations are in coding region and are heteroplasmic. Four mutations (*m.8572G>A* in the *MT-ATP6* gene, *m.15777G>A* and *m.15597T>C* in the *MT-CYB* gene, and *m.4632G>A* in the *MT-ND2* gene) are non-synonymous, one mutation (*m.9770T>C*) is synonymous, and two mutations (*m.2148_2149dupAG*, *m.3200T>C*) are in the *MT-RNR2* gene. In particular, mutations *m.15597T>C* and *m.4632G>A* have not been found in available data set according to a web-based search [28] (Table 2). Note that position 4632 in the *MT-ND2* gene had a conservation index [32] up to 1, which suggested that mutation *m.4632G>A* (p.A55T) might be deleterious.

4. Discussion

Comparing the entire mtDNA sequences between the cancerous and normal tissues from the same donor is usually required for identifying somatic mutations. In this study, we proposed a strategy to rapidly identify mtDNA somatic mutations based on mtDNA phylogeney. This strategy is established on a fact that somatic mutations (which were newly acquired in the mtDNA) are mainly located at the terminal branches of the mtDNA phylogenetic tree. By sequencing the complete mtDNA sequence of a tumor tissue and plotting all the sequence mutations/variants in the tree using the latest version of mtDNA phylogeny, the young private mutations/variants in the lineage can be distinguished from the ancient haplogroup-diagnostic variants. Note that the haplogroup-defining variants, which were accumulated during the evolution and were the product of natural selection, are unlikely to be a hot spot for back mutation [32–34]. On the contrary, the private mutations/variants in certain mtDNA, which occurred very recently, are relatively younger and are located in the terminal branch of a phylogenetic tree. Therefore, to identify the recently accumulated somatic mutation(s) in the tumor tissue, we only need to check the variants/mutations in the terminal branch in the paracancerous tissue or other normal tissues. In our recent study, we analyzed the entire mtDNA sequences of cancerous tissues and the corresponding adjacent and distant normal tissues of 10 Chinese patients with early stage breast cancer under strict data quality control [26]. A re-analysis of these complete mtDNA sequence data of the paired cancerous tissues and normal tissues according to the latest version of the global mtDNA phylogenic tree [17] did show that all the identified somatic mutations were located in the terminal branches of the tree (data not shown), suggested that our assumption is valid.

In this study, we validated the efficiency of our strategy for rapid identification of mtDNA somatic mutations in cancerous tissues by analyzing 10 gastric cancer tissue samples (including one with gastric stromal tumor) and the corresponding paracancerous or blood samples. We identified seven somatic mutations in cancerous tissues from four patients. Compared with the routine approach to sequence the entire mtDNA genomes in both cancerous and normal tissues to identify somatic mutation, the labor and cost could be efficiently reduced by employing our strategy, as we only need to sequence a few mtDNA fragments in the normal tissue.

The function of mtDNA somatic mutations in cancer has been under debate for decades [35–39]. Purifying selection was considered to play important roles in shaping human mitochondrial genome and the transmission of mammalian mtDNA [33,40,41]. But this selection against deleterious non-synonymous mutations was recently found to be less-intense in mtDNA genes in cancer [42]. As cancer cells mainly produce energy through aerobic glycolysis, somatic mutations in mtDNA can be tolerated and accumulated [42]. Note that some of the somatic mutations identified in this study are haplogroup-specific variants for young haplogroups to which the patients do not belong. For instance, mutation *m. 8572G>A* in sample GC2 is a characteristic variant

for haplogroup H1s and M33a1a, mutation *m.15777G>A* in sample GC2, *m.3200T>C* in sample GC4 and *m.9770T>C* in sample GC6 are haplogroup-specific variants for haplogroups L1c1a1b, M20 and U3b2, respectively (Table 2). The phenomenon of a same variant arising independently on different lineages is not rare in mtDNA phylogenetic tree [17]. Such variants may not be the direct target of selection, therefore can be randomly acquired by different lineages. Intriguingly, the somatic mutations identified in the four patients seemed to be clustered in the *MT-RNR2* gene and the *MT-CYB* gene (Table 2) despite a fact that the sample size was too small to make a firm conclusion. Whether these mutations enacted a biological function during the carcinogenesis remains an open question.

In short, we proposed and validated an efficient strategy that facilitates quick identification of mtDNA somatic mutations in cancerous tissues on the basis of mtDNA phylogenetic knowledge. We hoped that our strategy will be widely employed in the field and be helpful in the investigation of the role of somatic mutations in cancer.

Conflict of interest

The authors declare that there are no conflicts of interest.

Acknowledgements

We are grateful to the subjects who donated their samples. We thank the members of Dr. Yao's laboratory for helpful discussions. This study was supported by Yunnan Province (云南省高端人才计划 2009CI119), the National Natural Science Foundation of China (30925021), and the MOST (Ministry of Science and Technology of China; 2011CB910900).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.mrfmmm.2011.02.016.

References

[1] P.P. Hsu, D.M. Sabatini, Cancer cell metabolism: Warburg and beyond, Cell 134 (2008) 703–707.
[2] D.C. Wallace, A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine, Annu. Rev. Genet. 39 (2005) 359–407.
[3] J.F. Harrison, S.B. Hollensworth, D.R. Spitz, W.C. Copeland, G.L. Wilson, S.P. LeDoux, Oxidative stress-induced apoptosis in neurons correlates with mitochondrial DNA base excision repair pathway imbalance, Nucleic Acids Res. 33 (2005) 4660–4671.
[4] N.M. Druzhyna, G.L. Wilson, S.P. LeDoux, Mitochondrial DNA repair in aging and disease, Mech. Ageing Dev. 129 (2008) 383–390.
[5] A. Chatterjee, E. Mambo, D. Sidransky, Mitochondrial DNA mutations in human cancer, Oncogene 25 (2006) 4663–4674.
[6] M. Brandon, P. Baldi, D.C. Wallace, Mitochondrial mutations in cancer, Oncogene 25 (2006) 4647–4662.
[7] W.-Y. Hung, C.-W. Wu, P.-H. Yin, C.-J. Chang, A.F. Li, C.-W. Chi, Y.-H. Wei, H.-C. Lee, Somatic mutations in mitochondrial genome and their potential roles in the progression of human gastric cancer, Biochim. Biophys. Acta 1800 (2010) 264–270.
[8] C. Ye, X.O. Shu, L. Pierce, W. Wen, R. Courtney, Y.-T. Gao, W. Zheng, Q. Cai, Mutations in the mitochondrial DNA D-loop region and breast cancer risk, Breast Cancer Res. Treat. 119 (2010) 431–436.
[9] C.-W. Wu, P.-H. Yin, W.-Y. Hung, A.F. Li, S.-H. Li, C.-W. Chi, Y.-H. Wei, H.-C. Lee, Mitochondrial DNA mutations and mitochondrial DNA depletion in gastric cancer, Genes Chromosomes Cancer 44 (2005) 19–28.
[10] A. Salas, Y.-G. Yao, V. Macaulay, A. Vega, A. Carracedo, H.-J. Bandelt, A critical reassessment of the role of mitochondria in tumorigenesis, PLoS Med. 2 (2005) e296.
[11] H.-J. Bandelt, A. Achilli, Q.-P. Kong, A. Salas, S. Lutz-Bonengel, C. Sun, Y.-P. Zhang, A. Torroni, Y.-G. Yao, Low "penetrance" of phylogenetic knowledge in mitochondrial disease studies, Biochem. Biophys. Res. Commun. 333 (2005) 122–130.
[12] H.-J. Bandelt, A. Salas, Contamination and sample mix-up can best explain some patterns of mtDNA instabilities in buccal cells and oral squamous cell carcinoma, BMC Cancer 9 (2009) 113.
[13] B. Pakendorf, M. Stoneking, Mitochondrial DNA and human evolution, Annu. Rev. Genomics Hum. Genet. 6 (2005) 165–183.
[14] A. Torroni, M. Richards, V. Macaulay, P. Forster, R. Villems, S. Nørby, M.-L. Savontaus, K. Huoponen, R. Scozzari, H.-J. Bandelt, mtDNA haplogroups and frequency patterns in Europe, Am. J. Hum. Genet. 66 (2000) 1173–1177.
[15] A. Achilli, U.A. Perego, C.M. Bravi, M.D. Coble, Q.-P. Kong, S.R. Woodward, A. Salas, A. Torroni, H.-J. Bandelt, The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies, PLoS One 3 (2008) e1764.
[16] Q.-P. Kong, H.-J. Bandelt, C. Sun, Y.-G. Yao, A. Salas, A. Achilli, C.-Y. Wang, L. Zhong, C.-L. Zhu, S.-F. Wu, A. Torroni, Y.-P. Zhang, Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations, Hum. Mol. Genet. 15 (2006) 2076–2086.
[17] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, Hum. Mutat. 30 (2009) E386–394.
[18] Q.-P. Kong, H.-J. Bandelt, M. Zhao, Y.-P. Zhang, Y.-G. Yao, Reply to van Oven: suggestions and caveats for naming mtDNA haplogroup, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) E40–E41.
[19] D. Yu, X. Jia, A.-M. Zhang, S. Li, Y. Zou, Q. Zhang, Y.-G. Yao, Mitochondrial DNA sequence variation and haplogroup distribution in Chinese patients with LHON and m.14484T>C, PLoS One 5 (2010) e13426.
[20] Y.-G. Yao, A. Salas, C.M. Bravi, H.-J. Bandelt, A reappraisal of complete mtDNA variation in East Asian families with hearing impairment, Hum. Genet. 119 (2006) 505–515.
[21] S.B. Edge, D.R. Byrd, C.C. Compton, A.G. Fritz, F.L. Greene, A. Trotti, AJCC Cancer Staging Handbook, American Joint Committee on Cancer, Chicago, 2010.
[22] H.-W. Wang, X. Jia, Y. Ji, Q.-P. Kong, Q. Zhang, Y.-G. Yao, Y.-P. Zhang, Strikingly different penetrance of LHON in two Chinese families with primary mutation G11778A is independent of mtDNA haplogroup background and secondary mutation G13708A, Mutat. Res. 643 (2008) 48–53.
[23] R. Bi, A.-M. Zhang, W. Zhang, Q.-P. Kong, B.-L. Wu, X.-H. Yang, D. Wang, Y. Zou, Y.-P. Zhang, Y.-G. Yao, The acquisition of an inheritable 50-bp deletion in the human mtDNA control region does not affect the mtDNA copy number in peripheral blood cells, Hum. Mutat. 31 (2010) 538–543.
[24] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nat. Genet. 23 (1999) 147.
[25] L. Fan, Y.-G. Yao, MitoTool: a web server for the analysis and retrieval of human mitochondrial DNA sequence variations, Mitochondrion 11 (2011) 351–356.
[26] C.-Y. Wang, H.-W. Wang, Y.-G. Yao, Q.-P. Kong, Y.-P. Zhang, Somatic mutations of mitochondrial genome in early stage breast cancer, Int. J. Cancer 121 (2007) 1253–1256.
[27] Y. Ji, A.-M. Zhang, X. Jia, Y.-P. Zhang, X. Xiao, S. Li, X. Guo, H.-J. Bandelt, Q. Zhang, Y.-G. Yao, Mitochondrial DNA haplogroups M7b1'2 and M8a affect clinical expression of leber hereditary optic neuropathy in Chinese families with the m.11778G→A mutation, Am. J. Hum. Genet. 83 (2008) 760–768.
[28] H.-J. Bandelt, A. Salas, R.W. Taylor, Y.-G. Yao, Exaggerated status of "novel" and "pathogenic" mtDNA sequence variants due to inadequate database searches, Hum. Mutat. 30 (2009) 191–196.
[29] P. Soares, L. Ermini, N. Thomson, M. Mormina, T. Rito, A. Röhl, A. Salas, S. Oppenheimer, V. Macaulay, M.B. Richards, Correcting for purifying selection: an improved human mitochondrial molecular clock, Am. J. Hum. Genet. 84 (2009) 740–759.
[30] Y.-G. Yao, H.-J. Bandelt, N.S. Young, External contamination in single cell mtDNA analysis, PLoS One 2 (2007) e681.
[31] Y.-G. Yao, Q.-P. Kong, A. Salas, H.-J. Bandelt, Pseudomitochondrial genome haunts disease studies, J. Med. Genet. 45 (2008) 769–772.
[32] E. Ruiz-Pesini, D. Mishmar, M. Brandon, V. Procaccio, D.C. Wallace, Effects of purifying and adaptive selection on regional variation in human mtDNA, Science 303 (2004) 223–226.
[33] J.L. Elson, D.M. Turnbull, N. Howell, Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection, Am. J. Hum. Genet. 74 (2004) 229–238.
[34] T. Kivisild, P. Shen, D.P. Wall, B. Do, R. Sung, K. Davis, G. Passarino, P.A. Underhill, C. Scharfe, A. Torroni, R. Scozzari, D. Modiano, A. Coppa, P. de Knijff, M. Feldman, L.L. Cavalli-Sforza, P.J. Oefner, The role of selection in the evolution of human mitochondrial genomes, Genetics 172 (2006) 373–387.
[35] H.A. Coller, K. Khrapko, N.D. Bodyak, E. Nekhaeva, P. Herrero-Jimenez, W.G. Thilly, High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection, Nat. Genet. 28 (2001) 147–150.
[36] P.F. Chinnery, D.C. Samuels, J. Elson, D.M. Turnbull, Accumulation of mitochondrial DNA mutations in ageing, cancer, and mitochondrial disease: is there a common mechanism? Lancet 360 (2002) 1323–1325.
[37] I. Zhidkov, E.A. Livneh, E. Rubin, D. Mishmar, MtDNA mutation pattern in tumors and human evolution are shaped by similar selective constraints, Genome Res. 19 (2009) 576–580.
[38] Y.-G. Yao, Y. Ogasawara, S. Kajigaya, J.J. Molldrem, R.P. Falcão, M.-C. Pintão, J.P. McCoy Jr., E.G. Rizzatti, N.S. Young, Mitochondrial DNA sequence variation in single cells from leukemia patients, Blood 109 (2007) 756–762.
[39] Y. He, J. Wu, D.C. Dressman, C. Iacobuzio-Donahue, S.D. Markowitz, V.E. Velculescu, L.A. Diaz Jr., K.W. Kinzler, B. Vogelstein, N. Papadopoulos, Hetero-

plasmic mitochondrial DNA mutations in normal and tumour cells, Nature 464 (2010) 610–614.
[40] W. Fan, K.G. Waymire, N. Narula, P. Li, C. Rocher, P.E. Coskun, M.A. Vannan, J. Narula, G.R. Macgregor, D.C. Wallace, A mouse model of mitochondrial disease reveals germline selection against severe mtDNA mutations, Science 319 (2008) 958–962.
[41] J.B. Stewart, C. Freyer, J.L. Elson, A. Wredenberg, Z. Cansu, A. Trifunovic, N.-G. Larsson, Strong purifying selection in transmission of mammalian mitochondrial DNA, PLoS Biol. 6 (2008) e10.
[42] P. Stafford, E.B. Chen-Quin, The pattern of natural selection in somatic cancer mutations of human mtDNA, J. Hum. Genet. 55 (2010) 605–612.

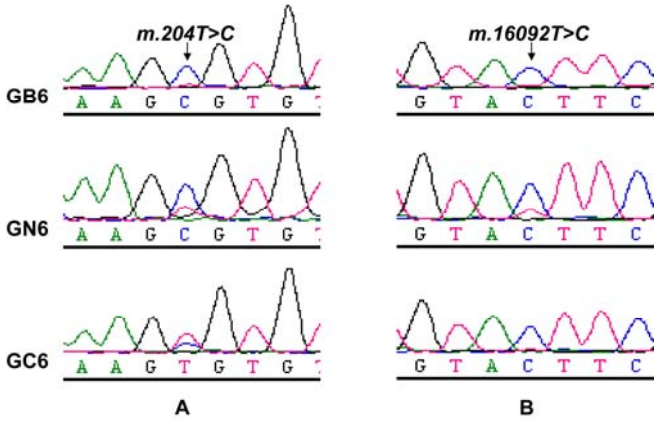
Supplementary Material

Supplementary Table 1. mtDNA control region sequence variants in different tissues from 10 Chinese patients with gastric cancer (including one gastric stromal tumor patient (P8))

Patients	Haplogroup	Tissue [a]	Variations [b]
P1	A4	GC1	73 151 152 200 235 263 309+C 315+C 523-524d 663 735 750 16223 16256 16290 16319 16362
		GN1	73 151 152 200 235 263 309+C 315+C 523-524d 663 735 750 16223 16256 16290 16319 16362
P2	F1a1a1	GC2	73 249d 263 309+C 315+C 523-524d 750 16108 16129 16162 16172 16304 16311 16519
		GN2	73 249d 263 309+C 315+C 523-524d 750 16108 16129 16162 16172 16304 16311 16519
P3	F2a	GC3	73 249d 263 315+C 523-524d 750 16203 16262+C 16291 16304 16311 16519
		GN3	73 249d 263 315+C 523-524d 750 16203 16262+C 16291 16304 16311 16519
P4	A	GC4	73 152 235 263 292 309+C 315+C 523-524d 663 750 16189 16223 16290 16319
		GB4	73 152 235 263 292 309+C 315+C 523-524d 663 750 16189 16223 16290 16319
P5	F3a	GC5	73 207 249d 263 309+CC 315+C 709 750 16260 16298 16355 16362
		GN5	73 207 249d 263 309+CC 315+C 709 750 16260 16298 16355 16362
P6	D5a2a1	GC6	73 150 199 204Y 263 315+C 489 523-524d 750 752 16092 16129 16164 16182C 16183C 16189 16223 16266 16362
		GN6	73 150 199 204Y 263 315+C 489 523-524d 750 752 16092Y 16129 16164 16182C 16183C 16189 16223 16266 16362
		GB6	73 150 199 204 263 315+C 489 523-524d 750 752 16092 16129 16164 16182C 16183C 16189 16223 16266 16362
P7	F1b1	GC7	73 249d 263 315+C 523-524d 750 16183C 16189 16232A 16249 16304 16311
		GN7	73 249d 263 315+C 523-524d 750 16183C 16189 16232A 16249 16304 16311
P8	F1a1	GC8	73 249d 263 315+C 523-524d 750 16129 16162 16172 16304 16519
		GN8	73 249d 263 315+C 523-524d 750 16129 16162 16172 16304 16519
P9	B4a1	GC9	73 146 263 309+CC 315+C 523-524d 750 16168 16182C 16183C 16189 16217 16261 16311 16362 16519
		GN9	73 146 263 309+CC 315+C 523-524d 750 16168 16182C 16183C 16189 16217 16261 16311 16362 16519
P10	B4g	GC10	73 263 309+CC 315+C 523-524d 750 16181C 16182C 16183C 16189 16213 16217 16261 16292 16519
		GN10	73 263 309+CC 315+C 523-524d 750 16181C 16182C 16183C 16189 16213 16217 16261 16292 16519

[a] "GC" means gastric cancer tissue or gastric stromal tumor tissue; "GN" means adjacent normal tissue; "GB" means blood tissue.

[b] Sequence variations were scored relative to the revised Cambridge reference sequence (rCRS) [1]. "d" indicates deletion; "+" indicates insertion; Suffixes A, C indicate transversions; Suffix "Y" indicates the mutation is heteroplasmic with T and C.



Supplementary Figure 1. Variants *m.204T>C* and *m.16092T>C* show different heteroplasmic levels in different tissues of patient P6. "GC" means gastric cancer tissue; "GN" means adjacent normal tissue; "GB" means blood tissue. (A): Variant *m.204T>C* is homoplasmic in blood tissue (GB6), but is heteroplasmic in paracancerous tissue (GN6) and cancerous tissue (GC6). (B): Variant *m.16092T>C* is homoplasmic in blood tissue (GB6) and cancerous tissue (GC6), but is heteroplasmic in paracancerous tissue (GN6).

Supplemental Reference

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