

## Correlational study on mitochondrial DNA mutations as potential risk factors in breast cancer

Linhai Li<sup>1,\*</sup>, Lidan Chen<sup>1,\*</sup>, Jun Li<sup>2,\*</sup>, Weiyun Zhang<sup>1</sup>, Yang Liao<sup>1</sup>, Jianyun Chen<sup>1</sup>, Zhaohui Sun<sup>1</sup>

<sup>1</sup>Department of Laboratory Medicine, Guangzhou General Hospital of Guangzhou Military Command of PLA, Guangzhou, Guangdong 510010, P.R. China

<sup>2</sup>Department of Information, No.4 Hospital of PLA, XiNing, Qinghai 810007, P.R. China

\*These authors have contributed equally to this work

Correspondence to: Linhai Li, e-mail: mature303@126.com

Keywords: mitochondrial DNA, mutation, high-throughput sequencing, breast cancer

Received: January 01, 2016

Accepted: April 02, 2016

Published: April 21, 2016

### ABSTRACT

The presented study performed an mtDNA genome-wide association analysis to screen the peripheral blood of breast cancer patients for high-risk germline mutations. Unlike previous studies, which have used breast tissue in analyzing somatic mutations, we looked for germline mutations in our study, since they are better predictors of breast cancer in high-risk groups, facilitate early, non-invasive diagnoses of breast cancer and may provide a broader spectrum of therapeutic options. The data comprised 22 samples of healthy group and 83 samples from breast cancer patients. The sequencing data showed 170 mtDNA mutations in the healthy group and 393 mtDNA mutations in the disease group. Of these, 283 mtDNA mutations (88 in the healthy group and 232 in the disease group) had never been reported in the literature. Moreover, correlation analysis indicated there was a significant difference in 32 mtDNA mutations. According to our relative risk analysis of these 32 mtDNA mutations, 27 of the total had odds ratio values (ORs) of less than 1, meaning that these mutations have a potentially protective role to play in breast cancer. The remaining 5 mtDNA mutations, RNR2-2463 indelA, COX1-6296 C>A, COX1-6298 indelT, ATP6-8860 A>G, and ND5-13327 indelA, whose ORs were 8.050, 4.464, 4.464, 5.254 and 4.853, respectively, were regarded as risk factors of increased breast cancer. The five mutations identified here may serve as novel indicators of breast cancer and may have future therapeutic applications. In addition, the use of peripheral blood samples was procedurally simple and could be applied as a non-invasive diagnostic technique.

### INTRODUCTION

Breast cancer is the most commonly diagnosed malignant tumor among females. Most breast cancers originate in the epithelial cells of lobules and in the non-epithelial tissue of the breast, although occasionally mixed carcinosarcomas can be found [1]. Breast cancers account for more than 1.2 million deaths each year in China alone [2]. Among women in China, breast cancer now has the highest incidence, while in the overall population it is the sixth most deadly cancer [3]. Breast cancer accounts for 12.2% of newly diagnosed cancer cases within China and

9.6% of global cancer related deaths [4]. In China, the mortality rate from cancer is much higher in urban areas than in the rural locales [5], and rising. Thus, there has been considerable effort applied to the development of campaigns for early detection, diagnosis and treatment of breast cancer in urban areas. In addition, investigations into the types and incidence rate of cancer, risk factors and an analysis of the high-risk populations are being supported [6–11].

Breast cancer is a systemic disease and can be widespread in the early stages. While doctors may overlook tumors less than 1 cm in size, studies report

circulating breast cancer cells present in blood samples in cases with these small tumors [12]. The majority of metastases are hematogenous metastases, which can generate tumors in all tissues and organs. The aggressive nature of breast cancer metastasis often leads to treatment failure due to cancer spread and recurrence [13]. The ability of breast cancer to metastasize in its early stages, the difficulty of detecting small tumors, and the relatively low survival rates of patients with recurrences make breast cancer a high-priority in the search for early-stage markers.

Recent developments in the search for genetic links to breast cancer have included the discovery of the BRCA genes, many of whose mutations have been mapped [14]. While encouraging, many factors influence the course of the disease and the search for gene mutations that correlate with cancer must continue. The methods used to locate these mutations have included high-frequency mutation screening, exon 11 screening and sequencing monitoring of all exons and adjacent introns [15]. However, a much more comprehensive analysis is required to identify molecular mechanisms in breast cancer and assess high-risk populations further. This would require large sample sizes and a great deal of expensive high-throughput screening.

Using mitochondrial DNA (mtDNA), though, could obtain meaningful results with considerably less resources spent on testing. Human mitochondrial DNA is a double-loop chain of 16.6 kb, with light and heavy chains. MtDNA is the only genetic material outside the nucleus in eukaryotic cells, and is correlated with oxidative phosphorylation. MtDNA contains 13 subunit coding genes [16–18] (Supplementary Table S1) for the oxidative phosphorylation system (OXPHOS), including seven subunits in complex I (ND1, ND2, ND3, ND4, ND4L, ND5, ND6), a subunit in complex III (CytB), three subunits in complex IV (COX1, COX2, COX3) and two subunits in complex V (ATPase6, ATPase8). MtDNA is also capable of encoding 22 ATP tRNAs and 2 rRNAs for protein and ATP synthesis (Supplementary Table S2). MtDNA has no introns. The only non-coding region is the D-loop region (displacement-region), which has been identified as a regulatory region of mtDNA replication and transcription, containing a replication origin and promoter region [19].

Over the past ten years, researchers have found that the mtDNA mutation rate is several times that of nuclear DNA. These mutations include gene deletions, missense mutations, frame-shift mutations and insertions [20]. Both structural and functional features lie behind mtDNA's greater susceptibility to carcinogen-induced mutation. MtDNA mutations can impair normal respiratory function and release large amounts of ROS, increasing the risk of tumorigenesis. Study data suggests that mtDNA has a strong correlation with tumorigenesis [17].

The objective of this study was to investigate the mitochondrial genome association with breast cancer.

Many scholars have confirmed that detection of mtDNA mutations in tumor cells is simpler and more reliable than of nDNA, owing to the fact that nuclear genetic damage is often missed in detection schemes, and that the number of copies of mtDNA is 200 times higher than those of the p53 gene [13, 16]. This can reduce sequencing labor and costs significantly. The study screened peripheral blood of breast cancer patients and researched the relationship between germline mutations and breast cancer. It is much more convenient to obtain samples of peripheral blood than breast cancer tissue samples, and it can be directly applied to non-invasive diagnostic techniques. Additionally, various factors of breast cancer occurrence are maternally inherited, possibly via the mitochondria. Using mtDNA markers of inheritable germline mutations for screening in high risk populations will play an important guiding role for the prevention and early diagnosis of breast cancer.

## RESULTS

### Purification and amplification of mitochondrial extracts

After mtDNA was extracted and purified, the Qubit® 2.0 results indicated that the concentrations of mtDNA samples were  $\geq 10$  ng/ $\mu$ l, which is in line with the requirements of PCR template concentration. When tested with 1 % agarose gel electrophoresis at 120 V, 25 min, the 16 kb bands were specific and clear, and without degradation or nuclear genomic contamination, which indicates high-purity mtDNA was extracted.

After amplification and gel extraction purification of the mtDNA, their concentrations, as measured by Qubit® 2.0 measurement of the 1-20 coded mtDNA samples indicated a concentration of 50 ng/ $\mu$ l and a volume of 50  $\mu$ l, which meet the requirements of library construction. The results from the 1 % agarose gel electrophoresis test are shown in Figure 1. As seen in the figure, there are bright bands at 7287 bp (Figure 1A) and 9391 bp (Figure 1B). These bands agreed with expectations, indicating that the acquisition of mtDNA was successful.

### Quality control of the library

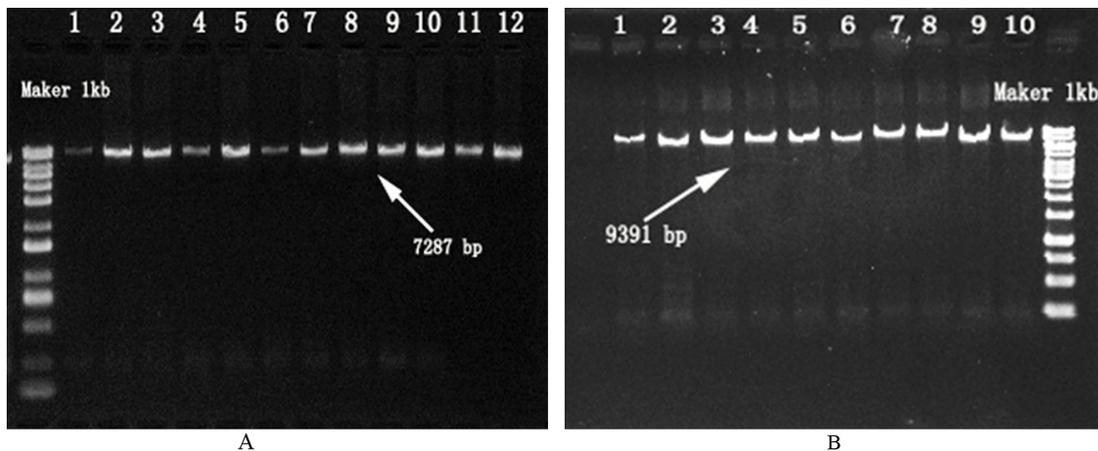
After the A and P1 ends of mtDNA were connected, the library was formed. Precise quality control of the library is the key to the quality of library sequencing data. The mtDNA library concentrations were measured by Qubit® 2.0 and each library met the requirements (1–10 ng/ $\mu$ l). This sequence had a read length of 200 reactions, interrupted by peak connectors at both ends of 200–250 bp plus 80 bp, which meant the target band was between 280 and 330 bp. Thus we used 2 % low-range detection agarose (120 V, 90 min) and gel extraction was performed to acquire the target band of 280–330bp. Figure 2 shows the correct band.

The sizes and concentrations of the mitochondrial fragment libraries were confirmed and analyzed by Agilent 2100. The results are shown in Figure 3. Figure 3A indicates that the sizes of the mitochondrial fragment libraries were between 280–330 bp (from mt1 to mt9), which is the expected band range and the same as the agarose gel results.

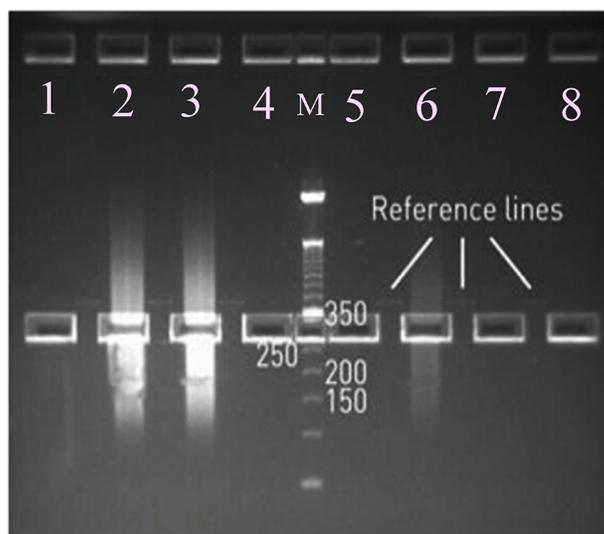
### Ion Torrent sequencing preliminary results

After using Fast QC-3.4.1.1 software to perform the analysis, the sequencing quality control report was as follows: the average read length of the mitochondrial

fragments was 250 bp; and the sequencing quality control between 50 and 250 bp reached the Q20 value. Also, the total base number generated from the original 318 chip sequence data was 1080 Mbp, and the Q20 base number was 920 Mbp. Using 250 bp as the read length of the mtDNA sequence fragments, the coverage of the chip sequencing was 88 %. Stripping out the polyclonal, low-quality data, 61 percent of the data's total read numbers came to 5,096,640. The sequencing quality control between 50 and 250 bp met or exceeded Q20's standards, which indicated that the quality of sequencing data was reliable. This in turn meant the mtDNA sequencing



**Figure 1: Confirmation of Amplification Products from the Mitochondrial Genome.** A. F7065 / R16455 Primers; Maker: molecular mass marker; Lanes 1-12: samples amplified bands using F7065 / R16455 primers. B. F16394 / R7111 Primers; Maker: molecular mass marker; Lanes 1-10: samples amplified bands using F16394 / R7111 primers.



**Figure 2: Selections of Mitochondrial Fragment Libraries: The length of sequencing bands: 280-350 bp.** M: molecular mass marker; Lanes 1, 4, 5, 6, 7, 8: Reference line bands (which were used as the 250-350bp cue lines); Lanes 2, 3: Amplified sample bands.

had been successful. The libraries' adapter primer was designed and the library construction kits were successfully assembled, so that the large-scale sequencing model of mtDNA was established.

### Data assembly and comparison

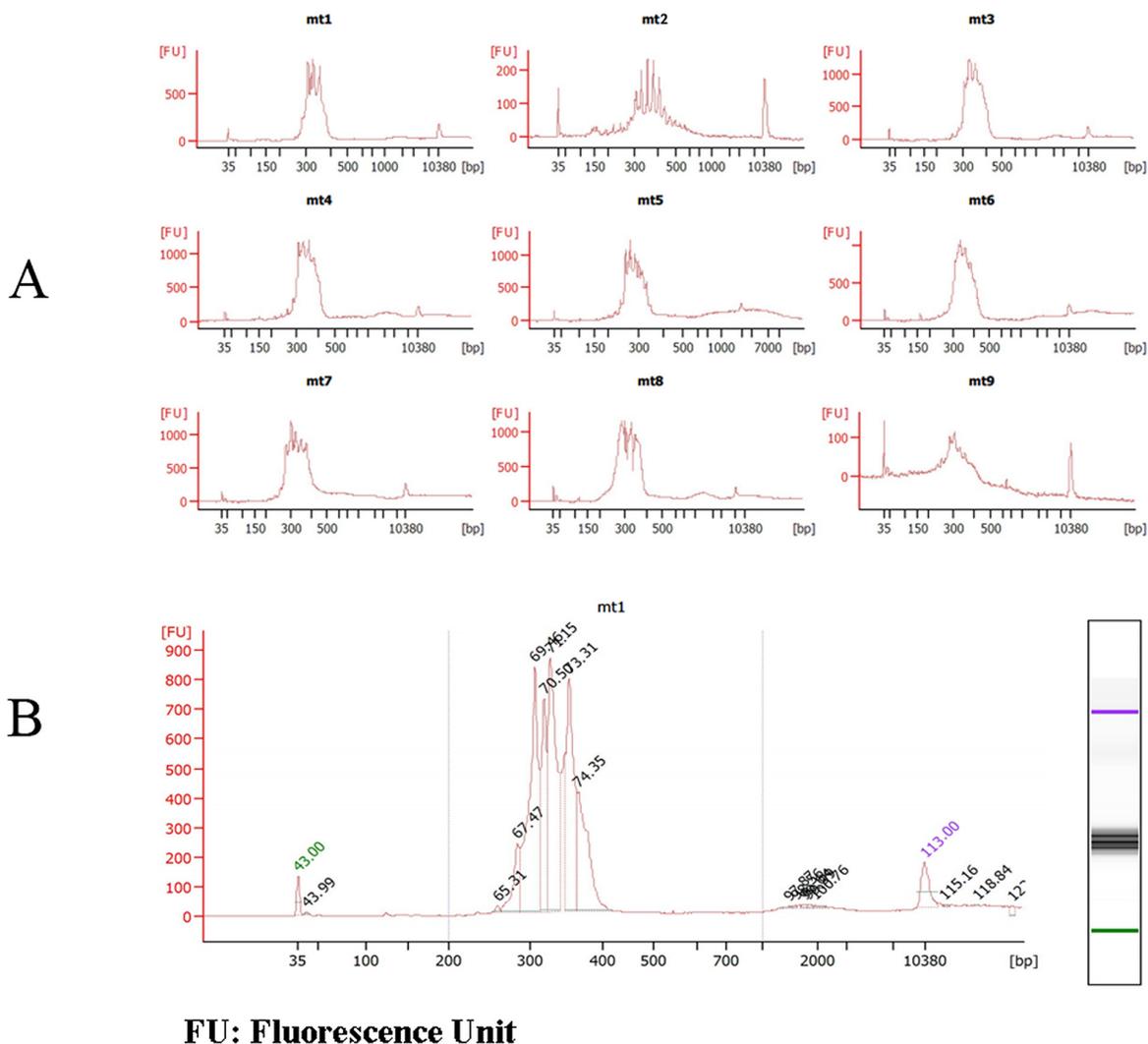
We chose the following software as the analysis tools: Assembler-3.4.2.0, Alignment 2.0, coverage Analysis 3.0, and sequence alignment software based on the Fasn2 algorithm. We assembled and aligned the data using hg38chrM data as the reference genome, and allowed 10 base mismatches. After pre-filtering and aligning the original sequence, we excluded sample data with a filtration rate of less than 40 %, a mapping rate of less than 80%, a depth of coverage less than 100X or where there was uneven coverage. Eventually, 22 cases with valid data from the control group and 83 cases from

the disease group were successfully screened. All of these data were used to do the correlation analysis between mutation sites and breast cancer.

### Search of mutation site

In order to perform an SNP loci search to obtain a list of mutations in the samples, we used hg38chrM as our reference genome (NC\_012920) and used the Ion Torrent variant caller for MtDNA-3.0 software in combination with the online analysis system (<https://ionreporter.LifeTechnologies.com/ir/>). Table 1 lists the mutation sites from sample No. 031 of the Control Group. The mutation sites for sample no. 137 from the Disease Group are listed in Table 2.

We then calculated statistics for each mutation site, using the number of samples for each locus mutation. The mutation frequency was calculated according to the following formula: mutation



**Figure 3: Quality Control Chart of mtDNA Fragment Library Size on Agilent 2100 bioanalyzer. A.** the sizes of some mtDNA fragments **B.** the concentrations of mtDNA1.

**Table 1: List of the Mutation Sites in Sample No. 031 of the Control Group**

Site	Mutation	Proportion	Site	Mutation	Proportion
46	insA	0.45	7271	A>G	1.00
146	T>C	1.00	8032	C>A	0.37
709	G>A	1.00	8033	delA	0.35
750	A>G	0.96	9123	G>A	1.00
1438	A>G	1.00	9412	G>C	0.29
3109	delT	0.98	9414	delC	0.29
3718	C>A	0.44	9907	G>C	0.29
3721	delA	0.38	9908	delC	0.29
5618	T>C	1.00	10476	C>T	0.26
6305	G>A	0.27	11719	G>A	1.00
6307	delA	0.26	12630	G>T	0.41
6958	G>C	0.68	12631	delT	0.41
6960	delC	0.66	15661	C>T	0.99
7028	C>T	0.99	16518	G>C	0.37
7052	A>G	1.00	16519	T>C	0.39
7078	G>C	0.58	16520	delC	0.27
7079	delC	0.58			

**Table 2: List of the Mutation Sites in Sample No. 137 of the Disease Group**

Site	Mutation	Proportion	Site	Mutation	Proportion
73	A>G	1.00	8032	C>A	0.27
146	T>C	0.99	8033	delA	0.26
199	T>C	1.00	8701	A>G	0.99
263	A>G	1.00	8860	A>G	1.00
489	T>C	1.00	9824	T>C	1.00
523	A>C	0.91	10398	A>G	1.00
524	delC	0.91	10400	C>T	1.00
525	C>G	0.89	11516	C>T	0.28
526	delG	0.88	11517	delT	0.26
750	A>G	0.98	11636	C>A	0.27
1438	A>G	1.00	11637	delA	0.26
2463	delA	0.35	11665	C>T	1.00
3109	delT	0.99	11719	G>A	1.00
3392	G>C	0.36	12091	T>C	1.00
3393	delC	0.36	12630	G>T	0.39
3882	G>A	1.00	12631	delT	0.39
4071	C>T	1.00	12705	C>T	1.00

*(Continued)*

Site	Mutation	Proportion	Site	Mutation	Proportion
4611	delA	0.40	13237	delA	0.32
4769	A>G	1.00	15043	G>A	1.00
4850	C>T	1.00	15301	G>A	1.00
5255	C>A	0.39	15326	A>G	1.00
5258	delA	0.39	16092	T>C	0.97
5442	T>C	0.99	16295	C>T	0.98
6455	C>T	1.00	16319	G>A	0.98
6698	delA	0.33	16518	G>C	0.36
6958	G>C	0.51	16519	T>C	0.88
6960	delC	0.49	16520	delC	0.36
7028	C>T	1.00	16528	C>T	0.27
7079	C>T	0.28	16529	delT	0.27

frequency = the number of sites with mutation/total number of cases. The results for each site are shown in Supplementary Table S3. These data include 170 mutations detected from the 22 samples in the control group, and 393 mutations from the 83 samples in the disease group. The occurrence of SNP of the disease group is much larger than that of the group control. In addition, the main types of mutations shown in Supplementary Table S3 are point mutations and base deletions, as well as some of the insertion mutations in the base fraction. In addition, a large fragment of 8283 – 8292 bp missing also found in two breast cancer patients' samples. A search on the NCBI database revealed 1321 reported mitochondrial DNA mutations, of which 182 were clearly confirmed to be associated with breast cancer. Our comparison and analysis found 283 mutation sites which had never been reported (indicated as sites with "\*" in Supplementary Table S3). These included 232 from the disease group and 88 from the control group.

### Analysis of the correlation between mtSNP and breast cancer

The principle of SNP correlation analysis is to use SNP molecular genetic markers in the overall correlation analysis. Genetic variations in the genome-wide data were selected to do the genotyping and the differences between each genetic variation and frequency variation were compared. Afterwards, we performed a statistical analysis to arrive at the correlation between each variation and target trait, and screened the most relevant genetic variation to validate the results. We then finalized its correlation with the target traits based on the verified results.

### Correlation analysis of breast cancer and mtSNP

SPSS19.0 statistical software was employed to run the  $\chi^2$  test on  $2 \times 2$  tables for each of the corresponding locus mutation frequencies in the two groups. Calculated using Fisher's exact test, we arrived at the test value from the double-sided exact test (p value); if  $p < \alpha$  (significance level:  $\alpha = 0.05$ ), it can be considered that a certain site demonstrates a significant difference between the two groups. Based on this principle, the results shown in Table 3 and Figure 4 indicate that there are 32 mutations with significant differences between these two groups. Previously, 25 of these 32 mutations have not been reported in the literatures.

These 32 SNP loci with significant differences were assessed for relative risk, and the OR and 95 % CI values were calculated. The analysis results showed there to be 27 SNPs with OR values of less than 1 among the 32 associated SNPs, and they that they had a negative correlation with the incidence of breast cancer. This indicated that these SNPs are protective factors which can reduce the risk of breast cancer occurrence. Meanwhile, the other five SNPs: 2463 (delA), 6296 (C>T), 6298 (delT), 8860 (A>G), 13237 (delA), whose OR values were 8.050, 4.464, 4.464, 5.254, and 4.853 respectively, were positively related to the development of breast cancer, and were thus considered high-risk factors. The incidence of these SNP may thus increase the chance of breast cancer. These five mutations are all located in the coding region of the mtDNA genome: the deletion in the 2463A site of RNR2, C6296T site mutation and 6298T site deletion in COX1, A8860G site mutation on ATP6, and the 13237A site deletion on ND5. Excepting A8860G (rs2001031), the four others were all discovered for the first time in this study.

**Table 3: Association Analysis between the 32 Significantly Different SNP Mutations and Breast Cancer**

Site	SNP	Gene	Mutation frequency		P	OR	95%CI
			Disease group	Control group			
16518*	G>C*		0.12	0.36	0.021	0.24	0.080~0.714
16520*	delC*	D-loop	0.12	0.32	0.046	0.294	0.096~0.895
146	T>C		0.02	0.18	0.017	0.111	0.019~0.654
150	C>T		0.07	0.23	0.05	0.265	0.072~0.970
2463* HR	delA* HR	RNR2	0.28	0.05	0.022	8.05	1.023~63.342
6296*HR	C>T* HR		0.17	0	0.038	4.464	0.555~35.900
6298* HR	delT* HR		0.17	0	0.038	4.464	0.555~35.900
6958*	G>C*		0.3	0.64	0.006	0.246	0.092~0.661
6960*	delC*	COX1	0.3	0.68	0.002	0.201	0.073~0.554
7078*	G>C*		0.08	0.64	0	0.053	0.016~0.168
7079*	delC*		0.08	0.64	0	0.053	0.016~0.168
7079*	C>T*		0.11	0.32	0.039	0.261	0.084~0.809
7081*	delT*		0.02	0.23	0.004	0.084	0.015~0.469
8032*	C>A*		0.06	0.68	0	0.03	0.008~0.107
8033*	delA*	COX2	0.08	0.73	0	0.035	0.010~0.107
8254*	C>G*		0.01	0.14	0.028	0.077	0.008~0.784
8860 HR	A>G HR	ATP6	0.19	0	0.021	5.254	0.658~41.923
9412*	G>C*		0	0.55	0	0.01	0.001~0.086
9414*	delC*	COX3	0	0.55	0	0.01	0.001~0.086
9907*	G>C*		0.06	0.68	0	0.03	0.008~0.107
9908*	delC*		0.06	0.68	0	0.03	0.008~0.107
10476*	C>T*	TRNR	0	0.23	0	0.041	0.004~0.373
10477*	delT*		0	0.09	0.042	0.12	0.010~1.396
11478*	G>C*		0	0.09	0.042	0.12	0.010~1.396
11479*	delC*	ND4	0	0.09	0.042	0.12	0.010~1.396
11719	G>A		0.45	0.73	0.03	0.302	0.107~0.848
12630	G>T		0.2	0.77	0	0.076	0.024~0.235
12631*	delT*	ND5	0.2	0.73	0	0.097	0.033~0.284
12705*	C>T*		0.28	0.64	0.003	0.219	0.081~0.591
13237* HR	delA* HR		0.18	0	0.037	4.853	0.606~38.866
15301	G>A	CYTB	0.13	0.36	0.025	0.267	0.091~0.784
15326	A>G		0.2	0.45	0.027	0.309	0.114~0.835

\*means the mtSNP is newly discovered; HR means a high risk factor.



**Table 4: Heterogeneity of the 32 Significantly Different Mutations**

Site	SNP	Mutation frequency	
		Disease Group	Control Group
146	T>C	0.9946	0.997
150	C>T	0.9932	0.9972
2463HR	DelAHR	0.331	0.33
6296HR	C>THR	0.467	-
6296HR	DelTHR	0.4618	-
6958	G>C	0.6194	0.6599
6960	delC	0.6021	0.6468
7078	G>C	0.3763	0.6076
7079	delC	0.3688	0.6045
7079	C>T	0.5943	0.6794
7081	delT	0.3082	0.3052
8032	C>A	0.3156	0.3444
8033	delA	0.3128	0.3225
8254	C>G	0.2781	0.2927
8860HR	A>GHR	0.9999	-
9412	G>C	-	0.3249
9414	delC	-	0.3253
9907	G>C	0.2814	0.2824
9908	delC	0.2819	0.2816
10476	C>T	-	0.2672
10477	delT	-	0.2674
10478	G>C	-	0.2637
10479	delC	-	0.2603
11719	G>A	0.9982	0.9982
12630	G>T	0.4032	0.462
12631	delT	0.402	0.4258
12705	C>T	0.9969	0.9989
13237HR	delAHR	0.315	-
15301	G>A	0.9987	0.9998
15326	A>G	0.9987	0.9991
16518	G>C	0.4551	0.4962
16520	delC	0.4575	0.4804

of researchers have subsequently reported this kind of mutation in other tumors, including 60% of breast cancers. Many somatic mtDNA mutations have been found in varying degrees, different sites, and different types [23–25]. Tan et al. found that many mtDNA missense mutations

of breast cancer have resulted in amino acid substitutions [26], such as T14110C and G14207A gene mutations in ND5 and ND6, respectively. Gallardo et al found that the G6267A mutation, which is involved in encoding COX in the mitochondria, is correlated with breast cancer [27].

In this study, genome-wide association analysis was performed on 22 samples from a control group and 83 samples from breast cancer patients in order to locate germline mutations of mtDNA. We found 32 site mutations associated with breast cancer, including 25 newly discovered mutations. After relative risk analysis based on OR values, 27 of the 32 mutations were found to have a negative correlation probability with breast cancer, and thus may have a protective effect that reduces the risk of breast cancer, while the remaining five mutations were positively related to the development of breast cancer. Compared with the normal group, these five mutations were associated with an increased risk of breast cancer by a factor of between 4 and 9. In addition, these five mutations are all located in the coding region of the mtDNA genome. They are: a deletion in the 2463A site of RNR2, C6296T site mutation and 6298T site deletion at COX1, A8860G site mutation at ATP6, and a 13237A site deletion at ND5. Excepting A8860G (rs2001031), the four others are all previously undiscovered. Mitochondrial DNA codes for important respiratory redox enzymes and RNA. Mutations in this coding region may result in frameshift mutations or missense mutations which lead to the encoded proteins losing their physiological functions; they may also produce a chain reaction, thereby causing disorders of physiological function, and these processes may eventually cause cancer.

The 5 mtDNA mutations identified in this study are likely to affect the synthesis of the respiratory chain. They may also result in increases in nuclear gene mutation and induce disorders of apoptosis by nDNA and mtDNA gene regulation a process which plays a very important role in the occurrence and development of breast cancer [28–30].

The present study differs from previous studies in that it focuses on the relationship between mtDNA mutations and breast cancer. Previous studies often used breast cancer tissues and analyzed the relationship between somatic mutations and cancer. However, the samples collected were often middle or late term, which yields limited guidance on early diagnosis. Additionally, many studies focused on nDNA, requiring the collection and screening of thousands of cases for a genome-wide association analysis. By investigating mtDNA mutations, this study focuses on germline mutation association with breast cancer, reduces the amount of screening required, and identifies biomarkers with potential for use in early detection in high-risk populations.

Due to time constraints, this study collected only 255 samples, of which 22 cases in the control group and 83 cases in disease group were used in the association analysis. While we acquired promising and beneficial results, a larger-scale research project must be performed to confirm the scientific relevance and reliability of our analyses. The five mutations identified in the presented study will be further validated in large-scale future studies.

In conclusion, our study differed from much previous work in investigating the relationship between

mtDNA germline mutations and breast cancer. The five mutations we found as a result may be indicators in the prevention of breast cancer, with the added possibility that they may be used in treating breast cancer. Moreover, the use of peripheral blood samples was procedurally much easier to conduct and less invasive, and may be applied to non-invasive diagnostic techniques in the future. Although only 255 samples were collected in our study, our use of high-throughput sequencing of mtDNA had never before been reported in China, and we believe the five key mutations identified in this study will be validated by future large-scale studies.

## MATERIALS AND METHODS

### Samples

Patients (the case group): from January 2013 to June 2014, the department of breast surgery in the General Hospital of Guangzhou Military Command collected 118 peripheral blood samples from patients who had just received their first pathological diagnosis of breast cancer and had never before undergone treatment for the condition. All were female, and between 25 and 84 years old (average:  $49.92 \pm 11.82$  years old). The pathological types included 70 cases of invasive ductal carcinoma, 14 cases of intraductal carcinoma, 16 cases of invasive and intraductal carcinoma, 3 cases of intraductal carcinoma with focal infiltration, 7 cases of mucinous carcinoma, 3 cases of ductal carcinoma in situ, 2 cases of invasive lobular carcinoma, 2 cases of atypical medullary carcinoma, and 1 case of invasive papillary carcinoma. Control group: Samples were collected from 137 individuals at the same hospital and the donors received comprehensive physical examination and imaging to exclude the presence of a variety of tumors. All were women, aged 26-80 years of age (average:  $55.17 \pm 13.28$  years old).

Peripheral blood samples were collected after obtaining the approval of the Ethics Committee of Guangzhou General Hospital of Guangzhou Military Command of PLA and the informed consents of donors. The samples, each 5ml of fresh blood stored with EDTA anticoagulant, were collected from the patients before treatment and from the control group at the same time, and were stored at  $-70^{\circ}\text{C}$ .

### Primer design

Two pairs of PCR amplification primers were designed in order to perform a complete enrichment of mitochondrial DNA and in vitro amplification for the extracted total genomic DNA. The lengths of the amplified products were 9391 bps and 7287 bps, both capable of covering the entirety of the amplification products' complete mtDNA sequences. These primer sequences were:

The first site: 7065/16455; the length of amplification product is 9391 bps.

F7065: 5'-GCCATCATAGGAGGCTTCATTAC-3';  
R16455: 5'-CGGAGCGAGGAGAGTAGCACTC  
TT-3'

The second site: 16394/7111; the length of amplification product is 7287 bps.

F16394: 5'-CCTTGACCACCATCCTCCGTGAA-3'  
R7111: 5'-TAGCCTGAGAATAGGGGAAATCA  
GT-3'

## Acquisition of mitochondrial DNA

Total DNA was extracted and purified from blood cells using the Gene JET Genomic DNA purification Kit (Thermo, USA), according to the manufacturer's instructions. The concentration of total DNA was detected using the Qubit® dsDNA HS Assay Kit (Invitrogen, USA), according to the manufacturer's protocol. Agarose gel (0.8%) electrophoresis was employed to confirm the purification and detect contamination of the total DNA.

Because the nuclear genome contains many sequences that share some homology with mtDNA, the mitochondrial genome from each sample was first amplified as a series of 14 overlapping 1.5–3 kb PCR products to avoid amplification of the nuclear mitochondrial-like sequences. Each of the reaction mixtures contained 10–50 ng total DNA, 2 U/μl Phusion DNA polymerase (0.5 μl), 5 × Phusion HF Buffer (10 μl), 2.5 mM of each dNTP (8 μl), 10 mM each of the forward and reverse primers (2.5 μl), and ddH<sub>2</sub>O to a final volume of 50 μl. PCR was performed using a thermocycler (4375305, Life Technologies) with the following conditions: 10 s at 95 °C, 6 min at 72 °C, followed by 30 cycles, and a final extension step of 10 min at 72 °C. The resulting products were stored at 4 °C. Gel electrophoresis was used to visualize the PCR products and AMPure XP Beads (Beckman Coulter Inc., Brea, CA, USA) were used to remove any contaminants. The Qubit® dsDNA HS Assay Kit (Q32851, Life Technologies, USA) and agarose gel electrophoresis were used to determine the concentration and purity of the 7287 bp and 9391 bp DNA products.

## Construction of mitochondrial DNA library

The 7287 bp and 9391 bp PCR products were mixed equally according to the DNA concentrations in order to create the sample which would be used to construct the mtDNA library. The flow chart of the DNA library construction is shown in Figure 5.

The genomic DNA was sheared using Bioruptor system reagents (Life Technologies, USA). A 2% agarose gel was used to visualize the sample fragment and the main site of cutting. DNA (100 ng in 79 μL) was combined with end repair enzymes (1 μl) in 20 μl 5X end repair buffer and incubated at room temperature for 0.5 hr. Agencourt® AM Pure® reagents were utilized

to purify the end-repair DNA. Finally, in order to join the ends and fill the gaps, DNA (25 μl) was mixed with Ion P1 Adapters (2 μl), Ion Xpress Barcode X (2 μl), 10×dNTP;Ligase Buffer (10 μl), DNA ligase (2 μl), dNTP Mix (2 μl) and Nick Repair Polymerase (8 μl), and nuclease-free water (49 μl), to a total volume of 100 μl. Incubation and thermal cycling were 25 min at 25 °C, followed by 5 min at 72 °C, then the temperature was held at 4 °C. Agencourt® AM Pure® reagents were used, as before, to purify the linked DNA.

The library was amplified via PCR using Platinum PCR Super Mix High Fidelity (100 μl), Library Amplification Primer Mix (5 μl) and 25 μl of the linked DNA. PCR was performed on a thermocycler with the following conditions: 5 min at 95°C, 8–10 cycles at 95°C (15 s), 15 s at 62°C, 1 min at 70°C, an additional 5 min at 79°C and finally hold at 4°C.

After amplification, Agencourt® AM Pure® reagents were used to purify the library. A Qubit dsDNA HS Assay Kit was used to determine DNA concentration and Agilent® High Sensitivity DNA Kit was used to determine the fragment size.

## Template preparation and sequencing

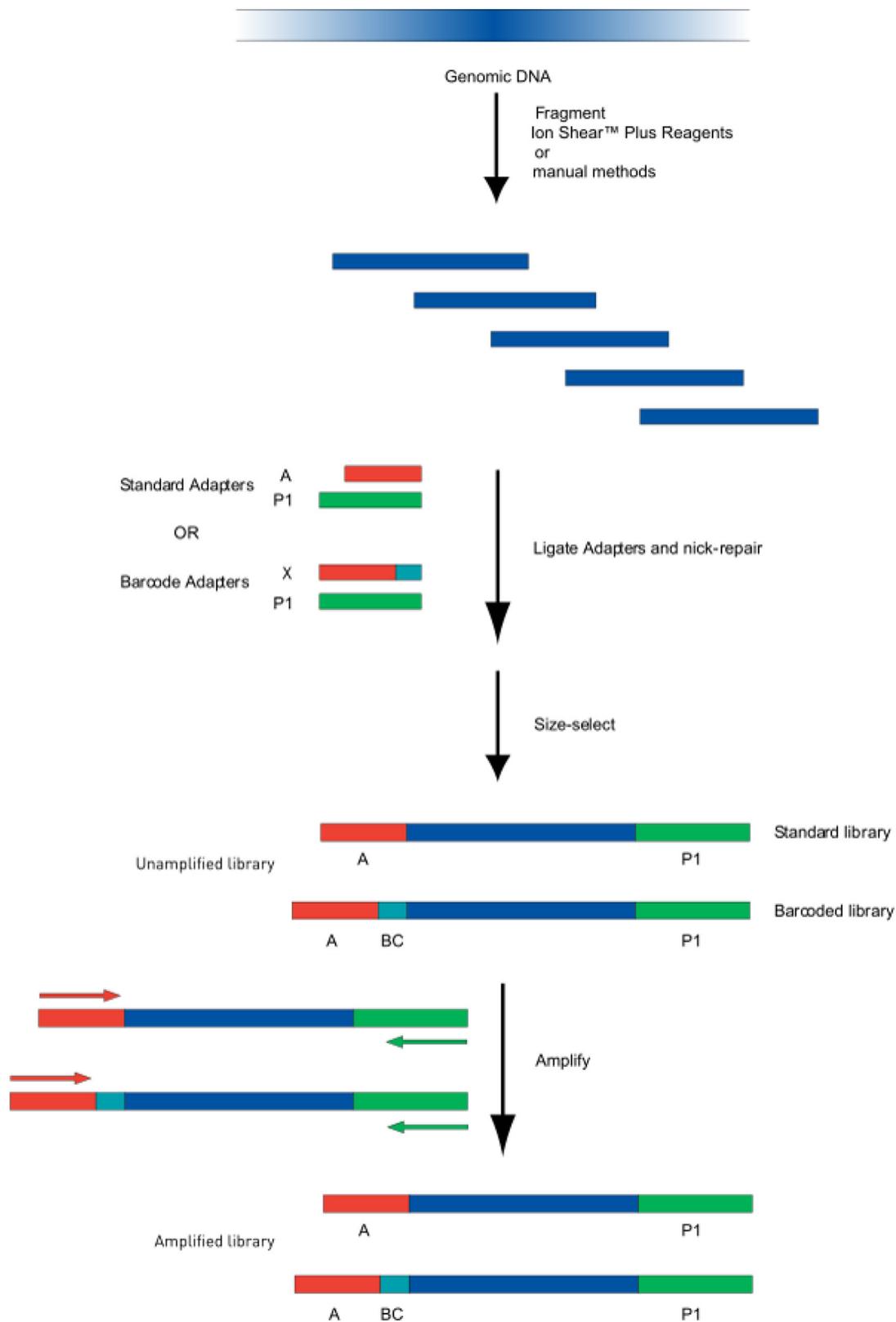
Template preparation was performed with the Ion PGM OneTouch OT2 200 Template Kit (Life Technologies, Carlsbad, CA, USA), according to the manufacturer's instructions. The library was diluted to a final concentration of 26 pM. 20 μL of the 20 code libraries above were mixed with equal volumes of amplification solution and emulsion PCR was performed using the Ion One Touch 2 instrument (Life Technologies, Carlsbad, CA, USA). An Ion OneTouch Enrichment System (Life Technologies, Carlsbad, CA, USA) was utilized to isolate template-positive Ion Sphere Particles, producing an enriched library with 3' end beads.

The library was sequenced using the Personal Genome Machine (Ion Torrent PGM, Life Technologies, Carlsbad, CA, USA). The results were analyzed with the Ion Torrent PGM server system (Torrent Suite 4.2) which exported raw data along with a sequencing report.

## Correlation analysis and statistical analysis

Based on the Fanse 2 algorithm, we made a comparison using the Homosapiens GRCh38 genome (NC\_012920) sequence as the reference and performed mutation site analysis and gene annotation. The mutation frequency was calculated based on the following formula: mutation frequency = number of samples with mutations / total number of samples.

SPSS19.0 (SPSS Inc., Chicago, IL) statistical analysis software was used to analyze the differences between the control group and the patient group, based on the chi-square test, and calculate the mutation frequency.



**Figure 5: Flow chart of DNA Library Construction.** Genomic DNA was sheared using a Bioruptor. The sheared fragments underwent end-repair and purified with Agencourt AM Pure reagents. The ends were joined and gaps filled using Ion P1 adapters, IonXpress Barcode adapters, DNA ligase and Nick Repair Polymerase. After additional purification, PCR amplification was performed to create amplified libraries.

When the number of samples was not less than 40 ( $n > 40$ ) and the theoretical frequency was not less than 5 ( $T > 5$ ), the  $\chi^2$  formula without correction was applied; the Pearson method to calculate the  $\chi^2$  is as follows:

$$\chi^2 = \sum \frac{(A - T)^2}{T}$$

When the number of samples was not less than 40 ( $n > 40$ ), and the theoretical frequency was between 1 and 5, the  $\chi^2$  calculation was performed with correction. Continuous correction of the chi-square value was calculated using the formula:

$$\chi^2 = \sum \frac{(|A - T| - 0.5)^2}{T}$$

When the number of samples was less than 40 ( $n < 40$ ), or the theoretical frequency was less than 1, Fisher's exact test was employed. Simultaneously, the relative risk assessment was characterized by the odds ratio (OR) and 95 % confident interval (95 % CI). Using the combined p value of the chi-square test, the correlation between SNP and breast cancer was analyzed so as to screen according to the high risk evaluation index:

When OR = 1, it indicates that the SNP is not associated with breast cancer;

When OR > 1, it indicates that the SNP is a high-risk factor for the occurrence of breast cancer;

When OR < 1, it indicates that the SNP is a protective factor, and could reduce the risk of breast cancer.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

## FINANCIAL SUPPORT

This study was supported by grants from Guangdong Province, the China Science and Technology Development Project of Guangdong Province (2010B011300018-7) and the Natural Science Foundation of Guangdong Province (8451051501000491).

## Author contributions

LHL, JL LDC and JL designed the research and wrote the paper. LHL, LDC, JL, WYZ, YL, JYC and ZHS performed the research. All authors discussed the results and commented on the manuscript.

## REFERENCES

1. Abdolrasoul Talei, Majid Akrami, Maral Mokhtari and Sedigheh Tahmasebi (2012). Surgical and Clinical Pathology of Breast Diseases, Histopathology - Reviews and Recent Advances, Dr. Enrique Poblet (Ed.), ISBN: 978-953-51-0866-5, InTech, doi: 10.5772/52105. Available from: <http://www.intechopen.com/books/histopathology-reviews-and-recent-advances/surgical-and-clinical-pathology-of-breast-diseases>.
2. Fan L, Strasser-Weippl K, Li JJ, St Louis J, Finkelstein DM, Yu KD et al. Breast cancer in China. *Lancet Oncol* 2014; 15: 79-89.
3. Zheng Y, Wu C, Zhang M. Epidemic status and disease characteristics of breast cancer in China. *Chin J Cancer* 2013; 23: 561-538.
4. Zhang M, Huang Z, Zheng Y. Estimation and prediction of the incidence, mortality and prevalence of female breast cancer of China in 2008. *Chin J Epidemiol* 2012; 33: 1049-1051.
5. Boyle P, Levin B. *World Cancer Report, 2008*. IARC Press: Lyon, France, 2008.
6. Lin Y, Yin W, Yan T, Zhou L, Di G, Wu J et al. Site-specific relapse pattern of the triple negative tumors in chinese breast cancer patients. *BMC Cancer* 2009; 9: 342-348.
7. Liu ZB, Liu GY, Yang WT, Di GH, Lu JS, Shen KW et al. Triple-negative breast cancer types exhibit a distinct poor clinical characteristic in lymph node-negative Chinese patients. *Oncol Rep* 1994; 20: 987-994.
8. Xing P, Li J, Jin F. A case-control study of reproductive factors associated with subtypes of breast cancer in Northeast China. *Med Oncol* 2010; 27: 926-931.
9. Zhao J, Liu H, Wang M, Gu L, Guo X, Gu F et al. Characteristics and prognosis for molecular breast cancer subtypes in Chinese women. *J Surg Oncol* 2009; 100: 89-94.
10. Su Y, Zheng Y, Zheng W, Gu K, Chen Z, Li G et al. Distinct distribution and prognostic significance of molecular subtypes of breast cancer in Chinese women: a population-based cohort study. *BMC Cancer* 2011; 11: 292-302.
11. Xu YL, Sun Q, Shan GL, Zhang J, Liao HB, Li SY et al. Risk Factors of Breast Cancer in China: A Case-Control Study. *Arch Med Sci* 2012; 8: 303-9.
12. Vachon CM, van Gils CH, Sellers TA, Ghosh K, Pruthi S, Brandt KR et al. Mammographic density breast cancer risk and risk prediction. *Breast Cancer Res* 2007; 9: 217-225.
13. Euhus DM. Understanding mathematical models for breast cancer risk assessment and counseling. *Breast J* 2001; 7: 224-232.
14. Joly Y, Tonin PN. Social, ethical and legal considerations raised by the discovery and patenting of the BRCA1 and BRCA2 genes. *New Genetics & Society* 2014; 33: 160-180.
15. Wang JJ, Lee CL, Pan TM. Modified mutation method for screening low citrinin-producing strains of *Monascus purpureus* on rice culture. *J Agric Food Chem* 2004; Nov 17; 52: 6977-6982.

16. Howell N. Leber hereditary optic neuropathy: How do mitochondrial DNA mutation cause degeneration of the optic nerve? *J Bioenerg Biomemb* 1997; 29: 165-173.
17. Brown MD, Wallace DC. Spectrum of mitochondrial DNA mutations in Leber's hereditary optic neuropathy. *Clin Neurosci* 1994; 2: 138-145.
18. Harding AE, Sweeney MG, Govan GG, Riordan-Eva P. Pedigree analysis in Leber hereditary optic neuropathy families with a pathogenic mtDNA mutation. *Am J Hum Genet* 1995; 57:77-86.
19. Taanman JW. The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta* 1999; 1410:103.
20. Panta JS, Johnson FM, Wachsman JT, et al. Mitochondrial DNA in human malignancy. *Mutat Res* 2001;488:119-133.
21. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J et al. Sequence and organization of the human mitochondrial genome. *Nature* 1981; 290: 457-465.
22. Polyak K, Li Y, Zhu H, Lengauer C, Willson JK, Markowitz SD et al. Somatic mutations of the mitochondrial genome in human colorectal tumours. *Nat Genet* 1998; 20: 291-293.
23. Tan DJ, Chang J, Chen WL, Agress LJ, Yeh KT, Wang B et al. Novel heteroplasmic frameshift and missense somatic mitochondrial DNA mutations in oral cancer of betel quid chewers. *Genes Chromosomes Cancer*, 2003, 37: 186-194.
24. Fliss MS, Usadel H, Caballero OL, Wu L, Buta MR, Eleff SM et al. Facile detection of mitochondrial DNA mutations in tumors and bodily fluids. *Science* 2000; 287: 2017-2019.
25. Canter JA, Kallianpur AR, Parl FF, Millikan RC. Mitochondrial DNA G10398A polymorphism and invasive breast cancer in African-American women. *Cancer Res* 2005; 65: 8028-8033.
26. Tan DJ, Bai RK, Wong LJ. Comprehensive scanning of somatic mitochondrial DNA mutations in breast cancer. *Cancer Res* 2002; 62: 972-976.
27. Gallardo ME, Moreno-Loshuertos R, López C, Casqueiro M, Silva J, Bonilla F et al. m.6267G>A: a recurrent mutation in the human mitochondrial DNA that reduces cytochrome c oxidase activity and is associated with tumors. *Hum Mutat* 2006; 27: 575-582.
28. Scarpulla RC. Transcriptional activators and coactivators in the nuclear control of mitochondrial function in mammalian cells. *Gene* 2002; 286: 81-89.
29. Richard CS. Nuclear control of respiratory chain expression in mammal cells. *J Bioenerg Biomemb* 1997; 29: 109-119.
30. Morases CT. What regulates mitochondrial DNA copy number in animal cells. *Trends Genet* 2001; 17: 199-205.