

Mitochondrial Genetic Background Modifies Breast Cancer Risk

Ren-Kui Bai,¹ Suzanne M. Leal,¹ Daniel Covarrubias,^{1,2} Aiyi Liu,³ and Lee-Jun C. Wong¹

¹Department of Molecular and Human Genetics, Baylor College of Medicine; ²Department of Statistics, Rice University, Houston, Texas and ³Biometry and Mathematical Statistics Branch, National Institute of Child Health and Human Development, NIH, Rockville, Maryland

Abstract

Inefficient mitochondrial electron transport chain (ETC) function has been implicated in the vicious cycle of reactive oxygen species (ROS) production that may predispose an individual to late onset diseases, such as diabetes, hypertension, and cancer. Mitochondrial DNA (mtDNA) variations may affect the efficiency of ETC and ROS production, thus contributing to cancer risk. To test this hypothesis, we genotyped 69 mtDNA variations in 156 unrelated European-American females with familial breast cancer and 260 age-matched European-American female controls. Fisher's exact test was done for each single-nucleotide polymorphism (SNP)/haplogroup and the *P* values were adjusted for multiple testing using permutation. Odds ratio (OR) and its 95% confidence interval (95% CI) were calculated using the Sheeher correction. Among the 69 variations, 29 were detected in the study subjects. Three SNPs, G9055A (OR, 3.03; 95% CI, 1.63–5.63; *P* = 0.0004, adjusted *P* = 0.0057), A10398G (OR, 1.79; 95% CI, 1.14–2.81; *P* = 0.01, adjusted *P* = 0.19), and T16519C (OR, 1.98; 95% CI, 1.25–3.12; *P* = 0.0030, adjusted *P* = 0.0366), were found to increase breast cancer risk; whereas T3197C (OR, 0.31; 95% CI, 0.13–0.75; *P* = 0.0043, adjusted *P* = 0.0526) and G13708A (OR, 0.47; 95% CI, 0.24–0.92; *P* = 0.022, adjusted *P* = 0.267) were found to decrease breast cancer risk. Overall, individuals classified as haplogroup K show a significant increase in the risk of developing breast cancer (OR, 3.03; 95% CI, 1.63–5.63; *P* = 0.0004, adjusted *P* = 0.0057), whereas individuals bearing haplogroup U have a significant decrease in breast cancer risk (OR, 0.37; 95% CI, 0.19–0.73; *P* = 0.0023, adjusted *P* = 0.03). Our results suggest that mitochondrial genetic background plays a role in modifying an individual's risk to breast cancer. [Cancer Res 2007;67(10):4687–94]

Introduction

Breast cancer, a multifactorial disease, is the most commonly diagnosed cancer of women and the second leading cause of cancer deaths among women. In addition to the high-risk genes, such as *BRCA1* and *BRCA2*, there is a mixture of genetic effects, environmental exposures, and gene-environment interactions that contribute to breast cancer risk. Studies have suggested that reactive oxygen species (ROS) play a role in the development of cancer (1–3). ROS are byproducts of normal mitochondrial electron transport chain (ETC) function, which oxidizes molecular oxygen to water and generates energy molecule, ATP. Oxidative stress may

occur if there is an imbalance between the ROS production and the antioxidant capacity due to impaired mitochondrial function, high dietary fat content, or insufficient dietary and endogenous defense activities against ROS (4).

The mitochondrial genome is highly polymorphic among individuals. *In vitro* studies showed that mitochondrial DNA (mtDNA) variations have subtle effect on mitochondrial respiratory chain activity (5). However, the sum of many subtle changes may cause significant consequences. Some mtDNA variations may be synergistic in combination with other mtDNA mutations (6, 7). Pathogenic expression of homoplasmic mtDNA mutations may also need a complex nuclear-mitochondrial interaction (8). Studies have suggested that mtDNA polymorphisms are associated with a variety of disorders, including Alzheimer's disease (9), acute myocardial infarction (10), bipolar disorder (11), Leber's hereditary optic neuropathy (6, 12), sensorineural hearing impairment (13, 14), Parkinson's disease (15–18), stroke (19, 20), and Wolfram (diabetes insipidus, diabetes mellitus, optic atrophy, and deafness) syndrome (21, 22). Most of these studies focused on one or limited number of mtDNA single-nucleotide polymorphisms (SNP). Studies using trans-mitochondrial cybrid system have shown that cells containing mitochondria bearing certain polymorphisms or haplotype may exert synergistic effect on excessive production of ROS, leading to oxidative stress (23–25).

The essential roles of mitochondria in energy metabolism, the generation of ROS, the initiation of apoptosis, and other aspects of tumor biology have implicated the importance of mitochondrial function in the neoplastic process. Low levels of ROS regulate cellular signaling and are essential in normal cell proliferation (26). ROS production is increased in cancer cells causing oxidative stress and DNA damage, leading to genetic instability (26–28). Thus, ROS are thought to play multiple roles in tumor initiation, progression, and maintenance (1). Somatic mtDNA mutations have been identified in a variety of malignant tumors, including breast cancer, colorectal cancer, ovarian cancer, gastric carcinoma, hepatocellular cancer, pancreatic cancer, prostate cancer, lung cancer, thyroid cancer, brain tumors, and esophageal carcinomas (2, 29). It is not clear if these somatic mtDNA mutations are the causes or results of the neoplastic process. However, based on the vicious cycle of ROS, it is conceivable that if the variant mitochondria function at a reduced efficiency, even with subtle changes, the accumulation of ROS effect may increase an individual's cancer risk over time. Thus, we hypothesize that mtDNA variations could act individually or in combination with other mtDNA variations or through interaction with nuclear gene or environmental factors to modify cancer risk. Studies of the single mtDNA SNP have shown that I0398A increases the risk of developing invasive breast cancer in African-American women (30) and increases the risk of developing prostate cancer in African-American males (31). In addition, mtDNA haplogroups also play an important role in disease expression (6, 17). We genotyped 69 mtDNA SNPs on 156 unrelated non-Jewish European-American female with breast cancer and 260 age-matched healthy non-Jewish

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Requests for reprints: Lee-Jun C. Wong, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, NAB 2015, Houston, TX 77030. Phone: 713-798-1940; Fax: 713-798-8937; E-mail: ljwong@bcm.edu.

©2007 American Association for Cancer Research.
doi:10.1158/0008-5472.CAN-06-3554

Table 1. Haplogroup classification

Haplogroups	SNPs															
	1719	4216	4580	7028	8251	8994	9055	10034	10398	12308	13368	13708	14470	14766	15607	16069
H	C	A	C
HV	G	T	A	C
I	A	T	A	C	G
J	...	C	G	A	T
K	A	G
T	...	C	...	T	A	A	A	G	...
U	G	...	A	G
V	A	T	A	C
W	T	A	A	A
X	A	T	A	C

European-American female controls. Our results showed that individual SNPs, 9055G>A, 10398A>G, and 16519T>C, increase a woman's risk of developing breast cancer or are in linkage disequilibrium with functional variants that increase a woman's risk, whereas 3197T>C and 13708G>A have a protective effect or are in linkage disequilibrium with protective variants. Belonging to haplogroup K increases a woman's risk of developing breast cancer, whereas membership in haplogroup U decreases a woman's risk of developing breast cancer.

Materials and Methods

Samples. A total of 156 unrelated European-American breast cancer patients with family history of breast cancer and 260 unrelated age/gender/ethnic matched control subjects were included in the present study. The patients were referred to the Molecular Genetics Laboratory at Georgetown University during 1997 to 2002 for mutational analysis of the common mutations in *BRCA1* and *BRCA2* genes because of family history of breast cancer. All patients, ages ≥ 26 years, had family history (at least one first-degree or two second-degree relatives also have breast cancer) of breast cancer. Blood DNA was extracted from peripheral blood lymphocytes. The mutations analyzed were 185delAG, 5382insC, T300G, 1294del40, 4184del4, C4446T, 1136insA, T>Gins59bp, 2800delAA, and 2798del4 for *BRCA1* and 6174delT, 3034del4, 6503delTT, and 982del4 for *BRCA2*. The control DNA samples were banked DNA specimens from age- and gender-matched anonymous individuals referred for genetic testing unrelated to cancer or mitochondrial disease during the same period. The specimens were stripped of personal identifiers, except ethnic background, age, and gender, according to approved protocol. Patients and controls with Jewish background were not included because the frequency of haplogroup K is much higher within the Jewish population compared with non-Jewish European-Americans (32). The mean age of the patient group was 50.7 ± 11.6 years (range, 26–86). The control group had a mean age of 50.2 ± 11.8 years (range, 26–88).

mtDNA genotyping. Four sets of multiplex PCR (Supplementary Table S1), one 4-plex and three 5-plex, were designed to include the mtDNA regions containing 69 reported mtDNA variations. These mtDNA variations were selected from Mitomap database.⁴ They are distributed along the rRNA, mRNA, tRNA, and displacement loop regions of the mitochondrial genome (Supplementary Table S2). Fifty-five of the 69 mtDNA variations studied had been reported in patients with various diseases, including Parkinson's disease, Alzheimer's disease, Leber's hereditary optic neuropathy, deafness,

chronic progressive external ophthalmoplegia, lethal infantile mitochondrial myopathy, diabetes, colon cancer, and other diseases (Supplementary Table S2). Sixteen of the SNPs studied have been reported to determine mtDNA haplogroups (Table 1). The primer sequences of the multiplexes are listed in Supplementary Table S1. Each 100 μ L of PCR mixture contained $1 \times$ PCR Buffer II (Applied Biosystems), 3 mmol/L MgCl₂, 0.2 mmol/L of each deoxynucleotide triphosphate, 0.5 μ mol/L each of the primers, 1 unit of Taq DNA polymerase, and 20 ng of total genomic DNA. The reaction mixture was denatured at 94°C for 5 min followed by 36 cycles of 1 min of denaturation at 94°C, 1 min of reannealing at 55°C, and 2 min of extension at 72°C. The PCR was completed by a final extension at 72°C for 5 min. Two microliters of PCR products were spotted onto Zeta-Probe Blotting Membrane (Bio-Rad). Dot blot preparation and hybridization conditions were those of published procedures (33, 34). The sequences of the allele-specific oligonucleotide (ASO) probes are listed in Table S3. For each ASO blot, both variant and wild-type controls are included as quality controls. This method has been routinely used in DNA diagnosis with an accuracy rate of >99.9%. About 10% of samples were randomly chosen for repeat with either ASO or sequencing. The results were 100% consistent with the first analysis. Among ~60,000 ASO genotypings, 582 failed, with a successful rate of ~99%.

Determination of mtDNA haplogroup. European mtDNA haplogroups H, Super HV (HV), I, J, K, T, U, V, W, and X were classified according to Table 1, which was based on published references (35–37) and the Mitomap database.⁴ Individuals with possible African admixture were excluded by the presence of C3594T SNP for the African haplogroup L. Similarly, individuals with possible Asian admixture were excluded by the presence of C10400T, determinant of haplogroup M.

Statistical analysis. Initial analysis was carried out using Fisher's exact test for each individual SNP and haplogroup. To adjust the *P* values for multiple testing and control for family wise error rate (FWER), 10,000 replicates were generated where the case/control status of the 416 study subjects was permuted and reanalyzed. Additionally, for each SNP and haplogroup, the odds ratio (OR) and its 95% confidence interval (95% CI) were calculated using the Sheeche correction (38).

Several of the SNPs had variant frequencies of ≤ 0.05 or were correlated with other SNPs. The method of Carlson et al. (39) was used to tagSNPs, which have pairwise values $r^2 \geq 0.8$ and implemented using Haploview (40). Those SNPs with a variant frequency of ≤ 0.05 or were tagged by other SNPs within the data set were not included in the stepwise logistic regression analysis. Haplogroup data and a subset of SNPs that met the r^2 and variant frequency criteria were analyzed with stepwise logistic regression using Akaike's information criteria (AIC) selection method (Akaike, 1974). To control the FWER for the stepwise logistic regression analysis, 10,000 replicates were generated and permutation was used to estimate the empirical *P* values. All statistical analyses with the exception of tagSNP selection were implemented using the R statistical package (R Core Development Team, 2006).

⁴ <http://www.mitomap.org>

Table 2. The results of the Fisher's exact test for individual SNP loci and haplogroups

SNPs/ haplogroup	Cases (n = 156)		Controls (n = 260)		OR	95% CI		Fisher's exact test P value	Adjusted P value
	Positive/ haplogroup	%	Positive/ haplogroup	%		Lower	Upper		
G709A	20	12.82	25	9.62	1.39	0.74	2.57	0.33	0.99
T710C	2	1.28	0	0.00	8.43	0.40	176.7	0.14	0.89
A1555G	0	0.00	2	0.77	0.33	0.02	6.93	0.53	0.99
G1719A	17	10.90	18	6.92	1.64	0.83	3.27	0.20	0.97
T3197C	6	3.85	31	11.92	0.31	0.13	0.75	0.0043*	0.0526 [†]
T3394C	3	1.92	4	1.54	1.30	0.32	5.32	1	1
T4216C	26	16.67	52	20.00	0.81	0.48	1.35	0.44	0.99
A4295G	3	1.92	2	0.77	2.36	0.45	12.09	0.36	0.99
T4336C	4	2.56	4	1.54	1.68	0.44	6.31	0.48	0.99
A4529T	10	6.41	8	3.08	2.13	0.84	5.37	0.13	0.86
G4580A	4	2.56	5	1.92	1.37	0.38	4.84	0.73	1
A4917G	17	10.90	20	7.69	1.47	0.75	2.87	0.28	0.99
G5460A	4	2.56	18	6.92	0.38	0.13	1.10	0.07	0.67
C7028T	91	58.33	149	57.31	1.04	0.69	1.55	0.91	1
G8251A	16	10.26	16	6.15	1.74	0.85	3.55	0.13	0.85
G8994A	2	1.28	3	1.15	1.19	0.23	6.11	1	1
G9055A	29	18.59	18	6.92	3.03	1.63	5.63	0.0004*	0.0057*
T10034C	10	6.41	10	3.85	1.71	0.71	4.12	0.24	0.98
A10398G	50	32.05	54	20.77	1.79	1.14	2.81	0.01 [†]	0.19
A12308G	41	26.28	66	25.38	1.05	0.66	1.65	0.91	1
G13368A	15	9.62	21	8.08	1.22	0.61	2.42	0.59	1
G13708A	12	7.69	40	15.38	0.47	0.24	0.92	0.0220 [†]	0.267
T14470C	2	1.28	3	1.15	1.19	0.23	6.11	1	1
C14766T	84	53.85	137	52.69	1.04	0.70	1.56	0.84	1
A15607G	15	9.62	19	7.31	1.35	0.67	2.73	0.46	0.99
A15924G	15	9.62	16	6.15	1.62	0.78	3.34	0.24	0.98
C16069T	10	6.41	30	11.54	0.54	0.26	1.12	0.12	0.83
T16189C	21	13.46	35	13.46	1.01	0.56	1.79	1	1
T16519C	122	78.21	167	64.23	1.98	1.25	3.12	0.0030*	0.0366 [†]
H	65	41.67	109	41.92	0.99	0.66	1.47	1	1
HV	4	2.56	9	3.46	0.78	0.25	2.44	0.77	1
I	9	5.77	9	3.46	1.71	0.71	4.12	0.24	0.98
J	10	6.41	29	11.15	0.54	0.26	1.12	0.12	0.82
K	29	18.6	18	6.92	3.03	1.63	5.63	0.0004*	0.0057*
T	13	8.33	19	7.31	1.19	0.59	2.41	0.71	1
U	12	7.69	48	18.46	0.38	0.20	0.73	0.0023*	0.03 [†]
V	3	1.92	5	1.92	1.05	0.27	4.11	1	1
W	2	1.28	3	1.15	1.19	0.23	6.11	1	1
X	2	1.28	3	1.15	1.19	0.23	6.11	1	1
UC	7	4.49	8	3.08	—	—	—	—	—

Abbreviation: UC, unclassified.

* $P \leq 0.01$.[†] $P \leq 0.05$.

Hierarchical clustering. To illustrate the relationships of the 29 SNPs that were detected, hierarchical clustering dendrograms were constructed for cases and controls separately using Euclidean distances, where the Euclidean distance is defined by $d_{rs}^2 = \sum_{i=1}^p (x_{ri} - x_{si})^2$. The Euclidean distance is calculated for each pairwise SNP combination of r and s such that $r \neq s$. At each SNP, the i^{th} individual can either be 0 for wild-type or 1 if a variant is present. The squared difference of the two observations at SNPs r and s is taken for the i^{th} individual and summed across all p individuals.

Drawing of phylogenetic network. Phylogenetic network (median-joining network; ref. 41) was analyzed and drawn by using software Network

4.112 available at the Web site of Fluxus technology.⁵ All the individuals from 156 breast cancer patients and 260 controls were analyzed together, and the phylogenetic network was generated automatically by the software. Each circle represents a haplotype harboring a set of specific SNPs. The area of the circle is roughly proportional to the number of sampled individuals that belong to that haplotype. The total number of individuals classified

⁵ <http://www.fluxus-technology.com/sharenet.htm>

into that haplotype is written beside the circle. Nucleotide transitions are indicated as the base changes at the nucleotide position number according to the revised Cambridge reference sequence.⁴ The phylogenetic network output was manually modified such that the number beside each circle represents the total number of women belonging to that haplotype, the number inside the black pie area indicates the normalized proportion (percentage) of women with breast cancer belonging to a given haplotype, whereas the remaining white pie area represents the control women belonging to that haplotype. The nucleotide abbreviation next to the nucleotide position represents the nucleotide found in the haplotype node closest to it. The normalized proportion (percentage) of women with breast cancer belonging to a given haplotype was calculated by the following:

$$P = \frac{\frac{n_{b\alpha}}{N_B}}{\frac{n_{b\alpha}}{N_B} + \frac{n_{c\alpha}}{N_C}} \times 100$$

where $n_{b\alpha}$ is the number of breast cancer patients belonging to the α haplotype, N_B is the total number of breast cancer patients, $n_{c\alpha}$ is the number of controls belonging to the α haplotype, and N_C is the total number of controls.

Results

Of the 69 mtDNA variants tested, only 29 were polymorphic in our study population (Table 2). All variants seem to be homoplasmic as detected using sensitive multiplex PCR/radioactive ASO method.

When each SNP is tested individually using the Fisher's exact test (Table 2), the frequencies of five SNPs (T3197C, G9055A, A10398G, G13708A, and T16519C) were found to be significantly different ($P \leq 0.05$) between the breast cancer patient group and the control group. After the adjustment for FWER using permutation, SNP G9055A remained highly significant ($P = 0.0057$), whereas SNPs T3197C and T16519C have P values of 0.05 and 0.04, respectively (Table 2). The ORs for G9055A (OR, 3.0; 95% CI, 1.6–5.6) and T16519C (OR, 2.0; 95% CI, 1.3–3.1) suggest that these SNPs increase a woman's risk of developing breast cancer or are in linkage disequilibrium with a functional SNP that increases a woman's risk, whereas T3197C has a protective effect (OR, 0.31; 95% CI, 0.13–0.75) or is in linkage disequilibrium with a protective variant.

Of the 29 SNPs that were genotyped, 9 SNPs (T710C, A1555G, T3394C, A4295C, T4336C, A4529T, G4580A, T10034C, and T14470C) were not included in the stepwise logistic regression due to variant frequencies ≤ 0.05 . Additionally, A4197G and A15607G were not

Table 3. Results of the stepwise logistic regression using AIC model selection for individual SNP loci

Variations	Coefficient estimate	SE	Z value	P value	Adjusted P value
Intercept	−0.46	0.12	−3.6	0.000283*	—
T3197C	−1.17	0.46	−2.53	0.0113*	0.03*
G5460A	−0.91	0.56	−1.60	0.1092	0.66
A10398G	0.92	0.27	3.40	0.00067 [†]	0.0007 [†]
G13708A	−1.36	0.39	−3.45	0.00055 [†]	0.0006 [†]

* $P < 0.05$.

[†] $P < 0.01$.

Table 4. Results of the stepwise logistic regression using AIC model selection for the haplogroups

Haplogroup	Coefficient estimate	SE	Z value	P value	Adjusted P value
Intercept	−0.45	0.12	−3.62	0.0003	—
J	−0.061	0.38	−1.58	0.1137	0.33
K	0.92	0.32	2.86	0.0043*	0.005*
U	−0.93	0.34	−2.70	0.0069*	0.008*

* $P < 0.01$.

included in the stepwise logistic regression analysis because they are both highly correlated with G13368A [A4917G ($r^2 = 0.81$) and A15607G ($r^2 = 0.88$)].

For the AIC stepwise logistic regression (Table 3), of the original 18 SNPs that were included in the analysis, SNPs T3197C ($P = 0.01$), G5460A ($P = 0.11$), A10398G ($P = 0.00067$), and G13708A ($P = 0.00055$) remained in the model. After adjusting the P values for multiple testing, SNPs A10398G ($P = 0.0007$) and G13708A ($P = 0.0006$) remained significant. For SNP A10398G, having the G variant increased a Caucasian woman's risk of developing breast cancer (OR, 2.5; 95% CI, 1.48–4.28), whereas for SNP G13708A the A variant is protective (OR, 0.26; 95% CI, 0.12–0.26). SNP T3197C is also protective but less significant ($P = 0.03$).

All 10 common European haplogroups (H, I, J, K, HV, T, U, V, W, and X) were observed. The haplogroup distribution in the control group is very similar to that in non-Jewish Europeans (32, 42). When the haplogroup data were analyzed using the Fisher's exact test, haplogroups K ($P = 0.0004$) and U ($P = 0.0023$) were shown to influence a woman's risk of developing breast cancer. The resulting P values remained significant after adjusting for multiple testing (Table 2). Whereas belonging to haplogroup K increases a woman's risk of developing breast cancer (OR, 3.03; 95% CI, 1.63–5.63), membership in haplogroup U decreases a woman's risk of developing breast cancer (OR, 0.38; 95% CI, 0.20–0.73). These results are consistent with those obtained from the AIC stepwise logistic regression (Table 4).

The hierarchical clustering dendrograms among cases display several distinct clusters (Fig. 1A), including a cluster that defines the H haplogroup (determined by nucleotide positions 7028, 14766, and 10398), K haplogroup (determined by nucleotide positions 9055 and 12308), and U haplogroup (determined by nucleotide positions 10398, 9055, and 12308). However, in controls (Fig. 1B), none of these variants are in the same cluster, except nucleotide positions 7028 and 14766, which partially define haplogroup H. There are several variants that form well-defined clusters in both cases and controls, including a cluster that contains nucleotide positions 709, 4917, 13368, and 15607 defining haplogroup T and another cluster containing nucleotide positions 13708 and 16069, which partially define haplogroup J. Only in controls do all of the nucleotide positions that define haplogroup J cluster (13708, 16069, 10398, and 4216). For both cases and controls, the Euclidean distance between nucleotide positions 16519 and other nucleotide positions is substantial (Fig. 1A and B).

Median-joining network analysis (Fig. 2) shows that there is much higher proportion of breast cancer cases compared with

controls in haplogroup K, whereas the converse is observed for haplogroup U. Both the Fisher's exact test and stepwise logistic regression showed that there was a statistical difference between cases and controls for these two haplogroups. The most prevalent haplogroup is H, which contains approximately equal proportions (42%) of both cases and controls.

Discussion

Oxidative stress is one of the major risk factors for cancer. Mitochondrial function plays a crucial role in ROS production. mtDNA variations can cause inefficient oxidative phosphorylation

leading to the accumulation of ROS, DNA damage, and increased cancer risk. In this study, we provide evidence that specific mitochondrial haplogroup background and mtDNA variations can be protective or risk contributors to breast cancer.

Two mtDNA variations, T3197C and G13708A, are found to have protective effect against breast cancer. T3197C is located at a stem region of 16S rRNA that may involve in the stability of the 16S rRNA structure. The T3197C SNP is right next to the G3196A variant, which was found to be associated with increased risk of Alzheimer's and Parkinson's diseases (43). G13708A SNP, changing an alanine residue to threonine in ND5, is a secondary mutation for Leber's hereditary optic neuropathy. It is conceivable that these

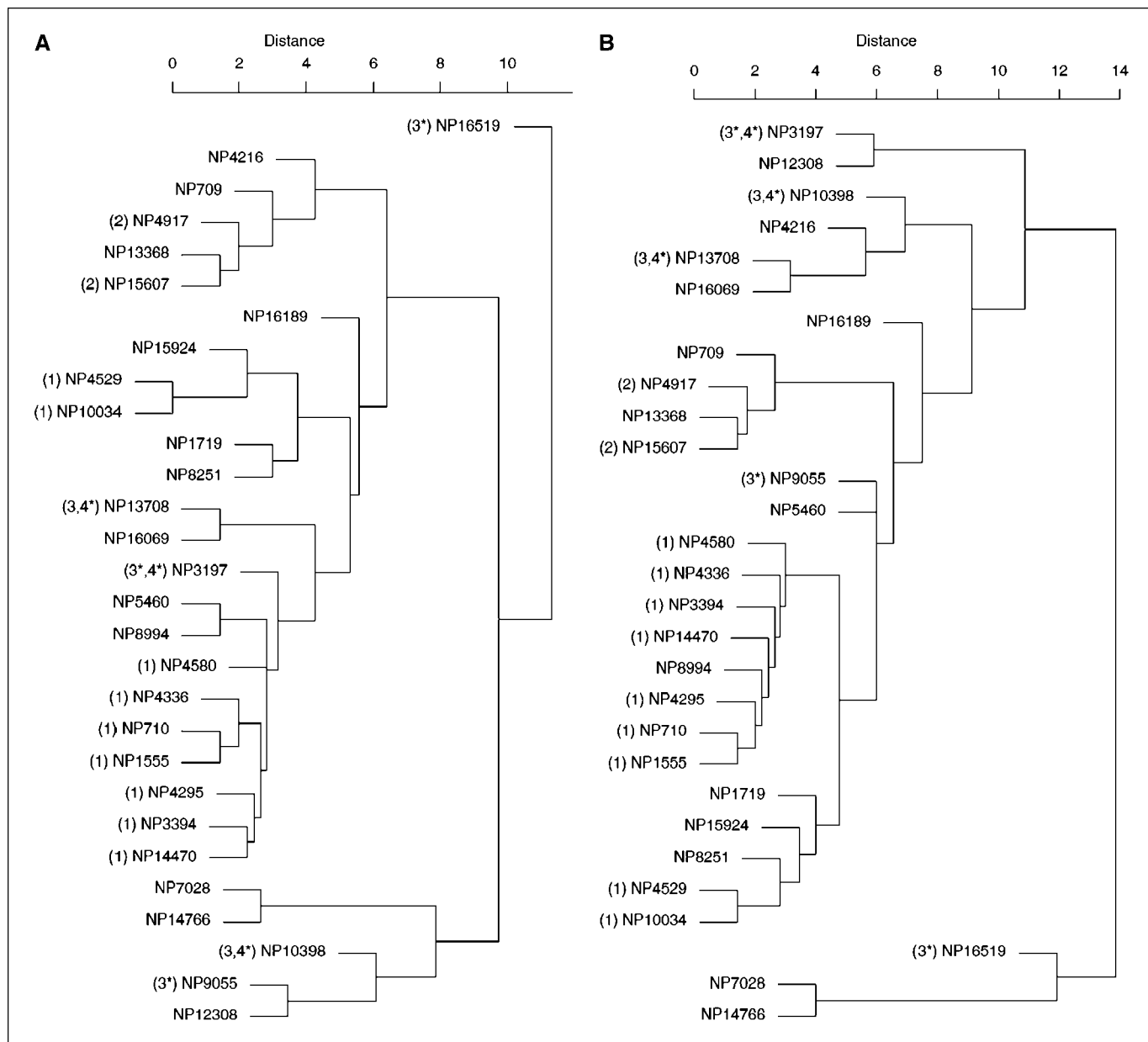


Figure 1. Hierarchical cluster dendrogram for cases and controls. X axis, clustered groups of SNP loci; Y axis, Euclidean distance between clusters. SNPs that were not included in the stepwise logistic regression model due to a low variant frequency are denoted by a 1 and those that were not included due to being highly correlated with SNP G13368A are denoted by a 2. SNPs that have a nominal significance of $P \leq 0.05$ for the Fisher's exact test are denoted by a 3 and a 3* if they remained significant at a $P \leq 0.05$ after adjusting for multiple testing using permutation. SNPs that have a nominal significance of $P \leq 0.05$ for the AIC stepwise logistic regression analysis are denoted by a 4 and a 4* if they remained significant at a $P < 0.05$ after adjusting for multiple testing using permutation. A, hierarchical cluster dendrogram for cases. B, hierarchical cluster dendrogram for controls. NP, nucleotide position.

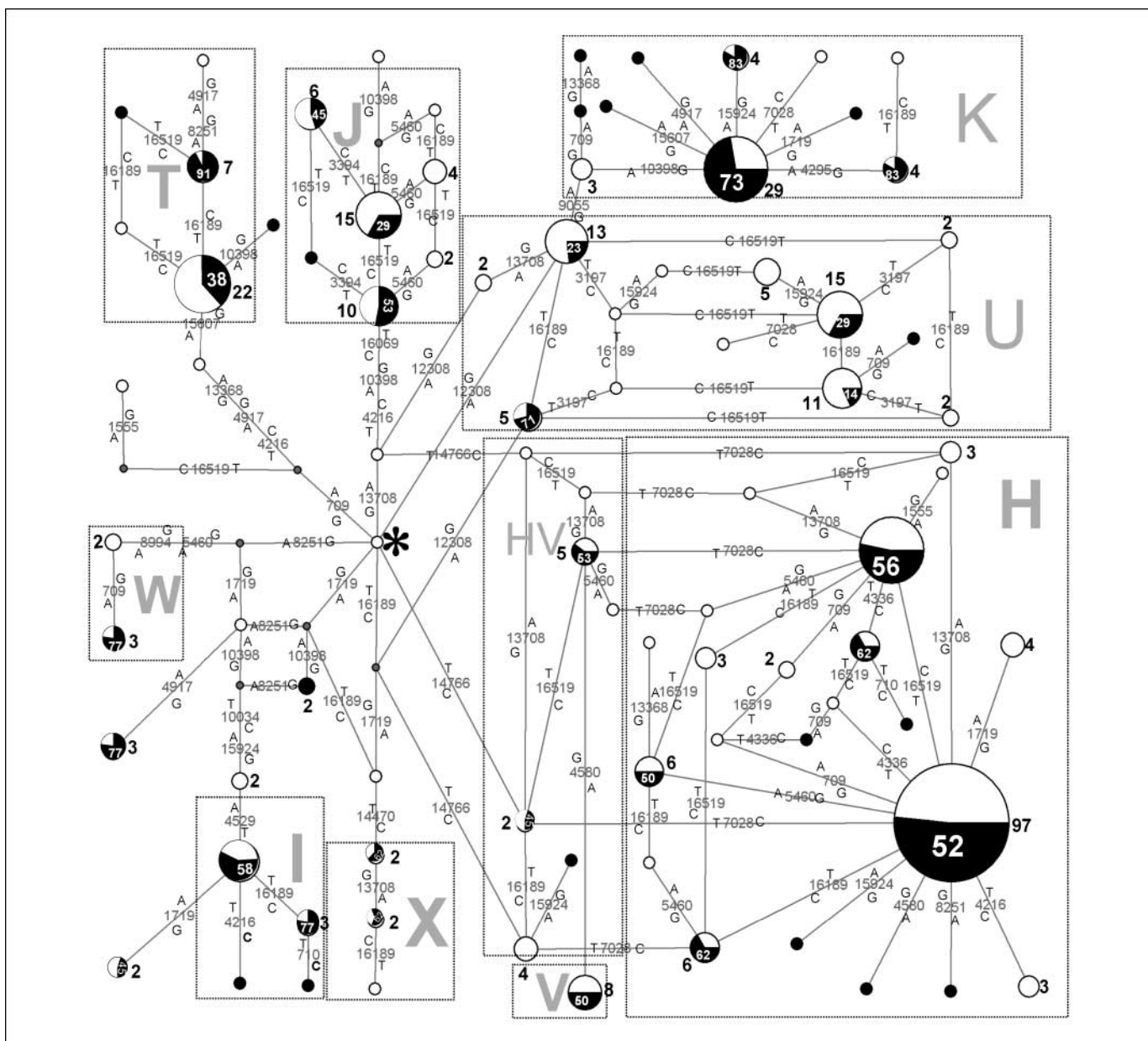


Figure 2. Median-joining network of mtDNA from breast cancer patients and normal controls. *Black and white circles*, haplotypes. *Black pie*, normalized proportion (percentage) of breast cancer patients; *white pie*, normalized proportion (percentage) of controls. *Small circles in gray*, median vector. *Number beside each circle*, total number of women belonging to that haplotype (the smallest white circles or black nodes without any number beside them have only one control or patient belonging to that haplotype). *Number in the black pie*, normalized percentage of breast cancer patients belonging to that haplotype. The size of the pie is proportional to the total number of individuals belonging to that haplotype. ***, center of the clusters, which is suggested by the branching pattern of the network, from which the other nodes (haplotypes) were formed after evolution. *Number between two adjacent haplotype nodes*, nucleotide position of mitochondrial genome. Clusters H, HV, I, J, K, T, U, V, W, and X stand for 10 different European mtDNA haplogroups. Those nodes that do not belong to any of the above haplogroups are unclassified.

variations, T3197C and G13708A, with detrimental effects in neurodegenerative diseases caused by cell death have protective effect in abnormal growth of cancer cells.

There is statistical evidence that the mtDNA variants G9055A, A10398G, and T16519C increase the risk of breast cancer or are in linkage disequilibrium with a functional variant that increases a Caucasian woman's risk of developing breast cancer. G9055A, a missense variant in *ATP6* gene of mtDNA, changing alanine to threonine, was reported to be a protective SNP for Caucasian women against Parkinson's disease (18). This nucleotide substitution, G9055A, defining haplogroup K, also occurs more frequently

in individuals with longevity (44). In our study, we found that G9055A is present at higher frequency in breast cancer patients than that in control. These results suggest that detrimental mtDNA variations for neurodegenerative disorders are actually protective factors for cancer and vice versa.

A10398G changes a nonconserved threonine residue to alanine in ND3 subunit of complex I. The 10398A variant is found in ~74% of white individuals, whereas it is present at a much lower frequencies in Asians (34%) and African-Americans (~5%; refs. 45, 46). The 10398G SNP was found to be strongly associated with the protective effect in white population against Parkinson's

disease (18). The protective effect seemed to be stronger in women than in men (18). A recent report from the Carolina Breast Cancer Study showed that the 10398A variant had a significant increased risk of invasive breast cancer in African-American women (30). They found that the 10398A is not a breast cancer risk polymorphism among white women. Our results also show that the 10398A variant is not a breast cancer risk SNP in Caucasian women. In fact, the frequency of 10398A variant is lower in white women with breast cancer. It is the 10398G variant that increases the risk of breast cancer in European-American women. Canter et al. (30) interpreted that the increase in risk for breast cancer in African-American women for carrying the 10398A reference allele is due to the interaction with other unidentified genetic and environmental risk factors. We studied 29 mtDNA SNPs and found that individuals belonging to haplogroup U, which contains 10398A, actually had lower risk of breast cancer. Because the 10398A>G occurred at a nonconserved amino acid, it is possible that 10398G is in linkage disequilibrium with other causative polymorphisms.

Individuals bearing haplogroup K have increased risk for breast cancer. This is opposite to the recent report that mtDNA haplogroups K and J have apparent protective effect on Parkinson's disease (16, 18). In our study, haplogroup J has neither protective nor detrimental effect on breast cancer risk. As a matter of fact, contradictory results on haplogroup J and disease association have been reported. An association between haplogroup J and longevity in Northern Italian men was found (47, 48). Conversely, haplogroup J has also been found to increase the risk for disease expression of Leber's hereditary optic neuropathy (6). A recent study by Booker et al. (49) showed that North American white individuals carrying mitochondrial haplogroup U has an ~2-fold increased risk of prostate cancer. Interestingly, our results show that European-American women with haplogroup U have significantly decreased risk of breast cancer. The study of Parkinson's disease clearly showed the differential gender effect of mitochondrial genetic background on disease risk (18).

Fisher's exact test and the AIC stepwise logistic regression may seem to be inconsistent. However, the hierarchical cluster dendrograms make the results clearer by elucidating the relationships between SNPs. For example, SNPs G9055A, A12308G, and A10398G are within the same branch in cases (Fig. 1A). SNP A10398G is retained in the AIC stepwise logistic regression model and this helps to explain why G9055A was removed from the AIC stepwise logistic regression model. Although T16519C was not included in the AIC stepwise logistic regression model, the *P* value for this SNP after adjusting for multiple testing was borderline significant at *P* = 0.04. The three SNPs (T3197C, A10398G, and G13708A) that were retained in the AIC stepwise logistic regression model were all significant for the Fisher's exact test.

The median-joining network suggests that additional SNPs to the haplogroup backbone may result in different observed proportions between cases and controls belonging to a given haplotype. For example, in haplogroup T, the proportion of breast cancer patients in the major node is 38%, although the addition of T16819C changes the proportion of breast cancer patients to 91%. Noteworthy is that none of these differences are statistically significant (data not shown). It should be noted that, for each haplotype, there are only a few observations; therefore, the power to detect whether these haplotypes play a role in breast cancer risk is low.

Despite a very thorough study of a large number of mitochondrial variants as well as extensive statistical and phylogenetic analysis, this study has some limitations. This study focuses on familial cases where nuclear gene alterations are most likely the primary cause. Mitochondrial genetic background may exacerbate the aberrant effect of nuclear genes. However, ~90% of breast cancers are sporadic cases with later onset, where environmental factors may play important roles by interacting with genetic factors. Inefficient mitochondrial function can accelerate a detrimental environmental effect to cause oxidative DNA damage leading to cancer. Whether the same group of mtDNA SNPs or haplogroups affects both familial and sporadic breast cancer risk similarly or not requires further investigation. In addition, due to limited information, we are unable to adjust the data for estrogen-related factors, such as menopausal status, body mass index, years of menstruation, age at menarche, hormone replacement therapy use, and treatment.

In conclusion, our results, for the first time, show the importance of mtDNA haplogroups and variations in modifying an individual's risk of breast cancer. It is possible that the effect of mitochondrial genetic background is influenced by physiologic conditions, such as hormonal state, of an individual. Thus, to understand the etiology of the protective or detrimental effect of mtDNA haplogroups or polymorphisms, more studies of all types of cancers of both genders with stratification of the data set by sex are necessary. Because mitochondria play an important role in modulating oxidative stress, the identification of significant mtDNA SNPs and haplogroups associated with breast cancer suggests that mitochondria may be involved in gene-gene and gene-environment interactions that may affect the pathogenetic mechanism of disease and cancer.

Acknowledgments

Received 10/4/2006; revised 2/2/2007; accepted 3/6/2007.

Grant support: NIH grants CA87327 and CA10023 and Department of Defense U.S. Army Breast Cancer Research Program grant DAMD17-01-1-0258.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Song-Ping Wang for technical assistance.

References

1. Benhar M, Engelberg D, Levitzki A. ROS, stress-activated kinases and stress signaling in cancer. *EMBO Rep* 2002;3:420-5.
2. Brandon M, Baldi P, Wallace DC. Mitochondrial mutations in cancer. *Oncogene* 2006;25:4647-62.
3. Sander CS, Chang H, Hamm F, Elsner P, Thiele JJ. Role of oxidative stress and the antioxidant network in cutaneous carcinogenesis. *Int J Dermatol* 2004;43:326-35.
4. Ambrosone CB, Freudenheim JL, Thompson PA, et al. Manganese superoxide dismutase (MnSOD) genetic polymorphisms, dietary antioxidants, and risk of breast cancer. *Cancer Res* 1999;59:602-6.
5. Polyak K, Li Y, Zhu H, et al. Somatic mutations of the mitochondrial genome in human colorectal tumours. *Nat Genet* 1998;20:291-3.
6. Brown MD, Starikovskaya E, Derbeneva O, et al. The role of mtDNA background in disease expression: a new primary LHON mutation associated with Western Eurasian haplogroup J. *Hum Genet* 2002;110:130-8.
7. Torroni A, Petrozzi M, D'Urbano L, et al. Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations 11778 and 14484. *Am J Hum Genet* 1997;60:1107-21.
8. Carelli V, Giordano C, d'Amati G. Pathogenic expression of homoplasmic mtDNA mutations needs a complex nuclear-mitochondrial interaction. *Trends Genet* 2003;19:257-62.
9. Carrieri G, Bonafe M, De Luca M, et al. Mitochondrial DNA haplogroups and APOE4 allele are non-independent variables in sporadic Alzheimer's disease. *Hum Genet* 2001;108:194-8.
10. Mukae S, Aoki S, Itoh S, et al. Mitochondrial 5178A/C

- genotype is associated with acute myocardial infarction. *Circ J* 2003;67:16–20.
11. Kato T, Kunugi H, Nanko S, Kato N. Mitochondrial DNA polymorphisms in bipolar disorder. *J Affect Disord* 2001;62:151–64.
 12. Isashiki Y, Sonoda S, Izumo S, Sakamoto T, Tachikui H, Inoue I. Phylogenetic assessment of the mitochondrial DNA displacement loop haplotype in Japanese patients with Leber's hereditary optic neuropathy harboring the mitochondrial DNA G11778A mutation. *Ophthalmic Res* 2003;35:224–31.
 13. Finnila S, Autere J, Lehtovirta M, et al. Increased risk of sensorineural hearing loss and migraine in patients with a rare mitochondrial DNA variant 4336A>G in tRNAGln. *J Med Genet* 2001;38:400–5.
 14. Lehtonen MS, Moilanen JS, Majamaa K. Increased variation in mtDNA in patients with familial sensorineural hearing impairment. *Hum Genet* 2003;113:220–7.
 15. Autere J, Moilanen JS, Finnila S, et al. Mitochondrial DNA polymorphisms as risk factors for Parkinson's disease and Parkinson's disease dementia. *Hum Genet* 2004;115:29–35.
 16. Ghezzi D, Marelli C, Achilli A, et al. Mitochondrial DNA haplogroup K is associated with a lower risk of Parkinson's disease in Italians. *Eur J Hum Genet* 2005;13:748–52.
 17. Ross OA, McCormack R, Maxwell LD, et al. mt4216C variant in linkage with the mtDNA T1 cluster may confer a susceptibility to mitochondrial dysfunction resulting in an increased risk of Parkinson's disease in the Irish. *Exp Gerontol* 2003;38:397–405.
 18. van der Walt JM, Nicodemus KK, Martin ER, et al. Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease. *Am J Hum Genet* 2003;72:804–11.
 19. Majamaa K, Finnila S, Turkka J, Hassinen IE. Mitochondrial DNA haplogroup U as a risk factor for occipital stroke in migraine. *Lancet* 1998;352:455–6.
 20. Pulkes T, Sweeney MG, Hanna MG. Increased risk of stroke in patients with the A12308G polymorphism in mitochondria. *Lancet* 2000;356:2068–9.
 21. Hofmann S, Bezold R, Jaksch M, Kaufhold P, Obermaier-Kusser B, Gerbitz KD. Analysis of the mitochondrial DNA from patients with Wolfram (DID-MOAD) syndrome. *Mol Cell Biochem* 1997;174:209–13.
 22. Hofmann S, Bezold R, Jaksch M, et al. Wolfram (DIDMOAD) syndrome and Leber hereditary optic neuropathy (LHON) are associated with distinct mitochondrial DNA haplotypes. *Genomics* 1997;39:8–18.
 23. Cardoso SM, Santana I, Swerdlow RH, Oliveira CR. Mitochondria dysfunction of Alzheimer's disease cybrids enhances A β toxicity. *J Neurochem* 2004;89:1417–26.
 24. Cassarino DS, Fall CP, Swerdlow RH, et al. Elevated reactive oxygen species and antioxidant enzyme activities in animal and cellular models of Parkinson's disease. *Biochim Biophys Acta* 1997;1362:77–86.
 25. Swerdlow RH, Parks JK, Miller SW, et al. Origin and functional consequences of the complex I defect in Parkinson's disease. *Ann Neurol* 1996;40:663–71.
 26. Burdon RH. Superoxide and hydrogen peroxide in relation to mammalian cell proliferation. *Free Radic Biol Med* 1995;18:775–94.
 27. Jackson AL, Loeb LA. The contribution of endogenous sources of DNA damage to the multiple mutations in cancer. *Mutat Res* 2001;477:7–21.
 28. Kang DH. Oxidative stress, DNA damage, and breast cancer. *AACN Clin Issues* 2002;13:540–9.
 29. Chatterjee A, Mambo E, Sidransky D. Mitochondrial DNA mutations in human cancer. *Oncogene* 2006;25:4663–74.
 30. Canter JA, Kallianpur AR, Parl FF, Millikan RC. Mitochondrial DNA G10398A polymorphism and invasive breast cancer in African-American women. *Cancer Res* 2005;65:8028–33.
 31. Mims MP, Hayes TG, Zheng S, et al. Mitochondrial DNA G10398A polymorphism and invasive breast cancer in African-American women. *Cancer Res* 2006;66:1880; author reply 1.
 32. Behar DM, Hammer MF, Garrigan D, et al. MtDNA evidence for a genetic bottleneck in the early history of the Ashkenazi Jewish population. *Eur J Hum Genet* 2004;12:355–64.
 33. Liang MH, Wong LJ. Yield of mtDNA mutation analysis in 2,000 patients. *Am J Med Genet* 1998;77:395–400.
 34. Wong LJ, Senadheera D. Direct detection of multiple point mutations in mitochondrial DNA. *Clin Chem* 1997;43:1857–61.
 35. Herrnstadt C, Elson JL, Fahy E, et al. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet* 2002;70:1152–71.
 36. Macaulay V, Richards M, Hickey E, et al. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 1999;64:232–49.
 37. Torroni A, Huoponen K, Francalacci P, et al. Classification of European mtDNAs from an analysis of three European populations. *Genetics* 1996;144:1835–50.
 38. Sheehy PR. Combination of log relative risk in retrospective studies of disease. *Am J Public Health Nations Health* 1966;56:1745–50.
 39. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004;74:106–20.
 40. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–5.
 41. Bandelt HJ, Forster P, Rohl A, Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 1999;16:37–48.
 42. Richards M, Macaulay V, Hickey E, et al. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 2000;67:1251–76.
 43. Shoffner JM, Brown MD, Torroni A, et al. Mitochondrial DNA variants observed in Alzheimer disease and Parkinson disease patients. *Genomics* 1993;17:171–84.
 44. Niemi AK, Hervonen A, Hurme M, Karhunen PJ, Jylha M, Majamaa K. Mitochondrial DNA polymorphisms associated with longevity in a Finnish population. *Hum Genet* 2003;112:29–33.
 45. Torroni A, Lott MT, Cabell MF, Chen YS, Lavergne L, Wallace DC. mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am J Hum Genet* 1994;55:760–76.
 46. Wallace DC. Mitochondrial diseases in man and mouse. *Science* 1999;283:1482–8.
 47. de Benedictis G, Carrieri G, Varcasia O, Bonafè M, Franceschi C. Inherited variability of the mitochondrial genome and successful aging in humans. *Ann N Y Acad Sci* 2000;908:208–18.
 48. De Benedictis G, Rose G, Carrieri G, et al. Mitochondrial DNA inherited variants are associated with successful aging and longevity in humans. *FASEB J* 1999;13:1532–6.
 49. Booker LM, Habermacher GM, Jessie BC, et al. North American white mitochondrial haplogroups in prostate and renal cancer. *J Urol* 2006;175:468–72; discussion 72–3.

Mitochondrial Genetic Background Modifies Breast Cancer Risk

Ren-Kui Bai, Suzanne M. Leal, Daniel Covarrubias, et al.

Cancer Res 2007;67:4687-4694.

Updated version Access the most recent version of this article at:
<http://cancerres.aacrjournals.org/content/67/10/4687>

Supplementary Material Access the most recent supplemental material at:
<http://cancerres.aacrjournals.org/content/suppl/2007/05/10/67.10.4687.DC1>

Cited articles This article cites 49 articles, 8 of which you can access for free at:
<http://cancerres.aacrjournals.org/content/67/10/4687.full#ref-list-1>

Citing articles This article has been cited by 15 HighWire-hosted articles. Access the articles at:
<http://cancerres.aacrjournals.org/content/67/10/4687.full#related-urls>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
<http://cancerres.aacrjournals.org/content/67/10/4687>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.