Polymorphisms in Mitochondrial Genes and Prostate Cancer Risk

Liang Wang, 1 Shannon K. McDonnell, 2 Scott J. Hebrinong, 1 Julie M. Cunningham, 1 Jennifer St Sauver, 2 James R. Cerhan, 2 Grazia Isaya, 3 Daniel J. Schaid, 2 and Stephen N. Thibodeau 1

Departments of 1Laboratory Medicine and Pathology, 2Health Sciences Research, and 3Pediatrics, Mayo Clinic College of Medicine, Rochester, Minnesota

Abstract

The mitochondrion, conventionally thought to be an organelle specific to energy metabolism, is in fact multifunctional and implicated in many diseases, including cancer. To evaluate whether mitochondria-related genes are associated with increased risk for prostate cancer, we genotyped 24 single-nucleotide polymorphisms (SNP) within the mitochondrial genome and 376 tagSNPs localized to 78 nuclear-encoded mitochondrial genes. The tagSNPs were selected to achieve ≥80% coverage based on linkage disequilibrium. We compared allele and haplotype frequencies in ~1,000 prostate cancer cases with ~500 population controls. An association with prostate cancer was not detected for any of the SNPs within 10 mitochondrial common haplotypes when evaluated using a global score statistic. For the nuclear-encoded genes, none of the tagSNPs were significantly associated with prostate cancer after adjusting for multiple testing. Nonetheless, we evaluated unadjusted P values by comparing our results with those from the Cancer Genetic Markers of Susceptibility (CGEMS) phase I data set. Seven tagSNPs had unadjusted P ≤ 0.05 in both our data and in CGEMS (two SNPs were identical and five were in strong linkage disequilibrium with CGEMS SNPs). These seven SNPs (rs17184211, rs4147684, rs4233367, rs2070902, rs3829037, rs7830235, and rs1203213) are located in genes MTRR, NDUFA9, NDUFS2, NDUFB9, and COX7A2, respectively. Five of the seven SNPs were further included in the CGEMS phase II study; however, none of the findings for these were replicated. Overall, these results suggest that polymorphisms in the mitochondrial genome and those in the nuclear-encoded mitochondrial genes evaluated are not substantial risk factors for prostate cancer. (Cancer Epidemiol Biomarkers Prev 2008;17(12):3558–66)

Introduction

In 2007, ~ 218,890 men in the United States will be diagnosed with prostate cancer, and >27,000 deaths will be attributed to the disease (1). Several etiologic factors for prostate cancer have been suggested, including genetic and environmental factors. However, only age, race or ethnicity and family history are established risk factors (2). Age is the strongest known risk factor, and the incidence of prostate cancer increases more steeply with age than for any other cancer (3). There is a large variation in its incidence among men in different countries, with the highest rates in the United States, Canada, Sweden, Australia, and France and Asian countries having the lowest rates (4). Although the causes of the variation of prostate cancer incidence are likely to be related to differences in screening methods, diet and health-related behaviors, clinical practice patterns, and environmental risk factors, there is a large body of literature that also strongly implicates a genetic etiology (5). This evidence comes from a variety of study designs, including case-control, cohort, twin, and family-based studies (6).

The mitochondrion, an organelle central to energy metabolism, also has multiple additional roles, including cell signaling, apoptosis, and cellular homeostasis. Mitochondria can generate reactive oxygen species and activate apoptosis. These reactive oxygen species function as crucial pro-apoptotic factors but may also be involved in the initiation and promotion of carcinogenesis. Interestingly, mitochondrial dysfunction has been found to be a common feature of cancer cells. Somatic mutations of mitochondrial DNA have been reported in a variety of cancers, including prostate cancer (7-13). Somatic alterations include intragenic deletions (14), missense and chain-terminating point mutations (7), and alterations of homopolymeric sequences (15), and these have been identified in nearly every type of tumor studied.

In addition to somatic alterations, several reports have shown that alterations in mitochondrial enzymes are also implicated in hereditary cancer syndromes (16). For example, the complex II of mitochondrial respiratory chain is composed of four nuclear-encoded subunits and is localized in the mitochondrial inner membrane.
Germline heterozygous mutations in three of the four subunits (SDHB, SDHC, and SDHD) cause the inherited syndromes that feature pheochromocytoma and paraganglioma (17). Another example is the nuclear-encoded mitochondrial enzyme fumarase (FH), an enzymatic component of the tricarboxylic acid cycle that catalyzes the formation of l-malate from fumarate. Mutations in the FH gene cause a predisposition to cutaneous and uterine leiomyomas, as well as kidney cancers (18). Finally, more recent studies have suggested that single-nucleotide polymorphisms (SNP) in the mitochondrial genome (mtSNPs) are associated with increased risk for several types of cancers, including invasive breast cancer in African-American women (19) and prostate cancer (10, 13, 20).

The findings in previous studies suggest that genetic variations in mitochondria might play an important role in developing cancers, a hypothesis recently emphasized in a conference report (21). To systematically test the role of mitochondria in prostate cancer risk, we genotyped 24 SNPs from coding and regulatory regions in the mitochondrial genome and 376 tagSNPs in 78 nuclear-encoded mitochondrial genes among groups of cases and controls. The nuclear-encoded mitochondrial genes included 138 tagSNPs in 30 genes associated with mitochondria-related cancer pathways, 161 tagSNPs in 36 genes involved in mitochondrial respiratory chain, and 77 tagSNPs in 12 genes serving as zinc transporters. Subjects included ~1,000 prostate cancer cases, derived from men with familial prostate cancer, men with sporadic prostate cancer, men with aggressive prostate cancer, and 495 population controls.

Materials and Methods

Familial Prostate Cancer. Ascertainment of families with prostate cancer has been described elsewhere (22). In brief, ~200 high-risk families were identified following a survey of 12,675 men enrolled in the Mayo Clinic radical prostatectomy database; families having a minimum of three men with prostate cancer were enrolled for further study. Blood was collected from as many family members as possible, resulting in a total of 498 affected men from 189 families. Of these, 490 affected men from 187 families are of non-Hispanic Caucasian ancestry. All men with prostate cancer who contributed a blood specimen had their cancers verified by review of medical records and pathologic confirmation. For the mitochondrial genome study, 435 affected men from 177 families were utilized for the analysis. For nuclear-encoded mitochondrial gene analysis, we were able to genotype 395 men from 187 families, of which 355 were also included in the mitochondrial genome study. The research protocol and informed consent forms were approved by the Mayo Clinic institutional review board.

Sporadic Prostate Cancer. Patients with sporadic prostate cancer were selected from respondents to our family history survey who reported no family history of prostate cancer (23). To ensure that the sporadic prostate cancer group was similar to the familial prostate cancer group, except for family history, eligible patients with sporadic prostate cancer were selected by frequency matching them to the familial prostate cancer index patients according to year of diagnosis, age at diagnosis, and number of brothers. Blood samples were available for 491 men of non-Hispanic Caucasian ancestry. All but 10 of these men were treated surgically for their prostate cancer.

Aggressive Prostate Cancer. Patients with aggressive prostate cancer (Gleason grade, ≥8) were also identified through the Mayo Clinic radical prostatectomy database. All men not previously contacted for our family history survey and who were diagnosed with high-grade prostate cancer were invited to participate. Of the 315 men eligible, 211 contributed a blood specimen and 204 were of non-Hispanic Caucasian ancestry. All of these men were treated surgically for their prostate cancer.

Population Controls. From a sampling frame of the local population provided by the Rochester Epidemiology Project (24), men were randomly selected for a clinical urological examination (25). This examination included digital rectal examination and transrectal ultrasound of the prostate, abdominal ultrasound for postvoid residual urine volume, measurement of serum levels of prostate-specific antigen and creatinine, focused urological physical examination, and cryopreservation of serum for subsequent sex hormone assays. Any patient with an abnormal digital rectal examination, elevated serum prostate-specific antigen level, or suspicious lesion on transrectal ultrasound was evaluated for prostatic malignancy. If the digital rectal examination and transrectal ultrasound were unremarkable but the serum prostate-specific antigen level was elevated (>4.0 ng/mL), then a sextant biopsy (three cores from each side) of the prostate was done. An abnormal digital rectal examination or transrectal ultrasound result, regardless of the serum prostate-specific antigen level, prompted a biopsy of the area in question. In addition, a sextant biopsy of the remaining prostate was done. These men have been followed with biennial examinations. All men without prostate cancer on the basis of this work up and any follow-up exams were used in the control sample. The mitochondrial genome study included 490 population controls, whereas the nuclear-encoded mitochondrial gene analysis included 495 population controls, 355 of who were in both control sets. All men are of non-Hispanic Caucasian ancestry.

Cancer Genetic Markers of Susceptibility (CGEMS) Data. After receiving approval to use the individual genotype data from the CGEMS phase I data (26), the genomewide association data for 1,172 prostate cancer cases and 1,157 controls were downloaded from the website (http://cgems.cancer.gov/data).

SNP Selection. Although thousands of mtSNPs have been reported (www.mitomap.org), most of these are rare (allele frequency, <1%). Based on allele frequency and the presence of common haplotypes in Caucasians (27), we selected 24 mtSNPs distributed across the mitochondrial genome for genotyping. Among these, 17 are within protein coding regions and the remaining 7 are in the regulatory region (displacement loop). Ten of the 24 variants are SNP that define common haplotypes in the Caucasian population (27).

There are >700 known nuclear-encoded mitochondrial proteins (www.mitoproteome.org). In this study, however,
we focused our efforts on the three categories of proteins: (a) respiratory chain proteins, (b) mitochondria-related cancer proteins, and (c) zinc transporter proteins. For these nuclear-encoded SNPs, our SNP selection relied on tagSNPs selected on the basis of linkage disequilibrium as implemented in IdSelect (28). To identify tagSNPs for each of the nuclear-encoded genes selected for study, we used the publicly available genotype data from the HapMap Consortium based on NCBI build 35 assembly and dbSNP build 125. SNPs were binned using an $r^2$ threshold of 0.8. A set of tagSNPs were identified such that each exceeded this $r^2$ threshold with all other SNPs in the same bin. To choose between multiple tagSNPs within a bin, we implemented hierarchical selection criteria based on larger design scores provided by Illumina, greater minor allele frequency, and preference for coding over noncoding SNPs.

In this study, we selected only those genes that had ≥80% coverage by the linkage disequilibrium bins. For example, if a gene extends from position 1 to 100 and contains one linkage disequilibrium bin with markers ranging from position 1 to position 80, then that gene is described as having a coverage of 80%. Additional “singleton” SNPs within this gene, that is, SNPs that are not in linkage disequilibrium with any surrounding typed SNPs, do not increase the computed coverage because they are not sufficiently correlated with any neighboring SNPs, although we recognize the possibility that these singleton SNPs could be in linkage disequilibrium with SNPs that are not in HapMap. We elected not to include singleton SNPs in our set of selected tagSNPs, focusing instead on those tagSNPs that are known to be in linkage disequilibrium with at least one HapMap SNP. Of the 393 selected SNPs, 379 were successfully genotyped and 376 were of sufficient quality for further analysis. Two SNPs were excluded because the genotype frequencies in controls deviated from Hardy-Weinberg equilibrium ($P < 0.001$), and one SNP was removed because of poor clustering.

**Genotyping for mtSNP.** We used the Beckman SNPstream system for mtSNP genotyping. Two 12plex panels of primer sets were designed using the Web-based Autoprimer. For each 12plex panel, 2 ng DNA isolated from peripheral blood lymphocytes was amplified with the pooled primer sets (50 nmol/L each) under universal PCR conditions [5 mmol/L MgCl$_2$, 75 µmol/L deoxy-nucleotide triphosphates, 0.1 unit AmpliTaq Gold (Applied Biosystems)] in a final volume of 5 µL. After initial denaturation at 94°C for 1 min, 34 cycles were done at 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min. The amplified products were then cleaned by incubation with SBE Clean-up (shrimp alkaline phosphatase and exonuclease I) at 37°C for 30 min and 96°C for 10 min. After 9 µL of extension mixes for each panel (C/T or G/A) was added to 7 µL of cleaned PCR product, the plates were cycled at 96°C for 3 min, then 45 cycles at 94°C for 20 s and 40°C for 11 s. SNPware array plates were prepared (washing with buffers 1 and 2). Eight microliters of hybridization solution was added to each well of the plate following primer extension reaction, and 10 µL of this was added to the corresponding well in the SNPware tag plate, incubated at 42°C and 100% humidity for 2 h. The arrays were then washed, vacuum dried, and imaged on the scanner. The SNPstream software was used for image data analysis and genotype calls. For each 384 well plate, quality control samples included 8 no-DNA template and 8 genomic controls. Final analysis revealed no signs of contamination or other technical problems associated with the genotype calls.

**Genotyping for Nuclear Gene SNP.** We designed a custom GoldenGate oligonucleotide pool and used the Illumina platform to do the genotyping (29, 30). Most SNPs assayed had SNP design scores ≥0.6. Automated genotype clustering and calling was made using BeadStudio II. Summary and report files were generated within BeadStudio and transferred electronically to a server for analysis. Samples with GenCall scores below 0.25 and/or call rates below 95% were removed, as were SNPs with GenCall scores below 0.4 and/or call rates <95%. To confirm and refine clusters, we used eight replicates of a CEPH family of nuclear trion. In addition, replicates of two DNA samples were included on each plate and 18 SNPs with known baseline genotype data on all cases and controls were included for quality control purposes. The error rate measured for the replicate samples and the 18 previously genotyped SNPs was 0.03%.

**Statistical Analyses**

**Mitochondrial SNPs. Single-SNP Analysis.** Although there are multiple mitochondrial DNA copies per cell, there is generally only a single allele for any given individual. In addition, mitochondria are transmitted maternally so that all familial prostate cancer cases in the same pedigree with a common maternal ancestor will have identical alleles. We therefore identified independent familial prostate cancer cases in the pedigrees by clustering subjects into groups defined by maternal ancestry and selecting a single subject per cluster. In this manner, 213 independent familial prostate cancer cases were selected for analysis. The mtSNP allele frequency was used to assess the difference between the cases and controls using standard contingency table methods. Unconditional logistic regression models, which treated case-control status as the outcome, were used to test the association between prostate cancer risk and SNP carrier status. Odds ratios and 95% confidence intervals (95% CI) were computed for carriers of the minor allele versus subjects homozygous for the major allele.

**Haplotype Analysis.** Genotypes for 10 core mtSNPs that define common haplotypes in Caucasians (31) were combined to construct mitochondrial haplotypes. Haplotypes are specific combinations of nucleotides on the same mitochondrial genome. To test for an association between the mitochondrial haplotype and case-control status, we calculated a score statistic using a modified version of the haplo-stats program (32), which implements an expectation-maximization algorithm to infer missing alleles for a haplotype. The analyses were based on global score statistics that compare all haplotypes between cases and controls.

**Nuclear-Coding Mitochondrial SNPs**

**Single-SNP Analysis.** For nuclear-encoded mitochondrial genes, single-SNP genotype frequencies were compared between prostate cancer cases and controls using...
Armitage test for trend for the number of minor alleles. This coding assumes an additive model, with heterozygote risk intermediate between the two homozygotes. Permutation $P$ values based on 10,000 random simulations are reported, wherein case-control status was randomly permuted for each simulation. To account for the relatedness among familial prostate cancer cases, a single familial prostate cancer case was randomly selected from each pedigree.

Because testing was done on a large number of SNPs, as well as multiple groups, both unadjusted $P$ values and those adjusted for multiple testing were computed. Adjustments were made separately for mitochondrial versus nuclear-encoded SNPs. Because some of these comparisons are dependent because of overlapping control group and SNPs in linkage disequilibrium within a gene, the usual Bonferroni Correction is too conservative. Hence, adjusted $P$ values were computed by 10,000 simulations. For each simulation, case-control status was randomly permuted and a new $P$ value was computed. The adjusted $P$ value was computed from the number of times out of 10,000 simulations that the minimum simulated $P$ value (over all SNPs and all group comparisons) was less than the observed $P$ value. These corrections accounted for the total number of SNPs evaluated, as well as the number of group comparisons (33).

**Haplotype Analyses.** The number of SNPs studied per gene ranged from 1 to 28. Haplotype analyses were conducted using all of the tagSNPs within each gene when more than a single SNP was studied. Rare haplotypes (frequencies, <1%) were collapsed into a single haplotype group, and the most frequent haplotype was considered the reference in the analyses. Global tests were conducted to assess the significance of all haplotypes simultaneously. In addition, the maximum of all haplotype-specific tests, comparing each haplotype to the pool of all other haplotypes, was identified. Simulated $P$ values were used to avoid problems that may result from sparse data. To account for the relatedness among familial prostate cancer cases, we randomly selected a single case from each pedigree. This randomization was repeated 100 times and the average $P$ values are reported. Analyses were conducted using the haplo-stats package in Splus Version 8.0.1.

**Principal Components.** The haplotype analyses described above might not be optimal in the sense that analyses of genes with many SNPs compare many different haplotypes between cases and controls, resulting in a global test with many degrees of freedom. As an alternative, we used principal components, a variable reduction procedure that typically results in a small number of components that account for most of the variance in a set of observed variables, in this case, the observed SNP genotypes within a candidate gene. The first N principal components that explained at least 90% of the variance in the observed SNPs were used to test for associations with prostate cancer. For each principal component, scores for each subject were calculated and these scores were used in a logistic regression model comparing all prostate cancer cases to controls. A global test for association was obtained for each gene with degrees of freedom equal to the number of principal component scores fit in the model. To account for the relatedness of the familial prostate cancer cases, a single affected man was randomly selected from each pedigree to test for association. The process of randomly selecting familial prostate cancer cases, computing the principal component scores, and fitting logistic regression models was repeated 100 times. The average global $P$ value is reported.

**Population Stratification.** To investigate whether stratification exists among our patient groups due to differences in ethnic ancestry, we used all 376 nuclear-encoded SNPs to create principal components. Plots of the first two principal components revealed random scatter for all four patient groups, suggesting that population stratification is not likely to influence our results.

**Results**

The characteristics of the four study groups are shown in Table 1. The sets of familial prostate cancer cases and controls used for each of the different studies are reported separately. Although the distribution of the age at diagnosis and body mass index levels are similar between the familial prostate cancer and sporadic prostate cancer cases, the cases with aggressive prostate cancer and the controls tended to be younger and have higher body mass index than the familial prostate cancer and sporadic prostate cancer cases. However, analyses adjusting for age and body mass index did not alter our findings. Therefore, for simplicity, we present only the unadjusted results.

We first evaluated mtSNPs and mitochondrial haplotypes for their association with prostate cancer utilizing 908 prostate cancer cases (213 cases from 177 familial prostate cancer pedigrees, representing independent maternal clusters, 491 sporadic prostate cancer and 204 aggressive prostate cancer) and 490 population-based controls. All 24 mtSNPs were genotyped in the familial prostate cancer and sporadic prostate cancer cases, as well as in the controls. Only 19 of the 24 mtSNPs were genotyped in the aggressive prostate cancer cases. The frequencies of these 24 mtSNPs along with odds ratios for the carriers of minor alleles, and their 95% CIs are presented in Fig. 1 (Supplementary Table 1). Overall, none of our findings were statistically significant after correcting for multiple comparisons, with all corrected $P \geq 0.95$. Given the power limitations of this study, we used the CGEMS data as an independent follow-up data set to further evaluate our findings. Of the 24 mtSNPs tested, we found 12 identical SNPs in the CGEMS project data set. None of these 12 SNPs showed an association with prostate cancer in the CGEMS study.

To analyze mitochondrial haplotypes, we used 10 core mtSNPs that define common haplotypes in Caucasians: SNPs 1719, 4580, 7028, 8251, 9055, 10398, 12308, 13368, 13708, and 16391. To reduce computations, subjects who were missing ≥7 loci were excluded (2 controls, 22 sporadic prostate cancer, and 7 aggressive prostate cancer). Eleven haplotypes with frequencies at least 1% were included in the analysis (Table 2). Based on a global scorestatistical that compares all 11 haplotypes between cases and controls, we did not identify any significant associations with familial prostate cancer, sporadic prostate cancer, aggressive prostate cancer, or all cases combined (global $P = 0.091, 0.416, 0.604, and 0.499$
Table 1. Characteristics of prostate cancer cases and population controls

<table>
<thead>
<tr>
<th></th>
<th>SPC (n = 491)</th>
<th>APC (n = 204)</th>
<th>FPC* (n = 435)</th>
<th>Nuclear-coding mtDNA SNPs (n = 395)</th>
<th>Cont (n = 491)</th>
<th>Nuclear-coding mtDNA SNPs (n = 495)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong> (median [range])</td>
<td>65 (46-79)</td>
<td>64 (40-86)</td>
<td>66 (45-84)</td>
<td>65 (42-80)</td>
<td>63 (43-91)</td>
<td>63 (44-91)</td>
</tr>
<tr>
<td>Age (quartiles)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤59</td>
<td>90 (18)</td>
<td>59 (29)</td>
<td>66 (15)</td>
<td>86 (22)</td>
<td>205 (42)</td>
<td>186 (38)</td>
</tr>
<tr>
<td>60-65</td>
<td>175 (36)</td>
<td>53 (26)</td>
<td>128 (29)</td>
<td>131 (33)</td>
<td>84 (17)</td>
<td>102 (21)</td>
</tr>
<tr>
<td>66-70</td>
<td>167 (34)</td>
<td>47 (23)</td>
<td>131 (30)</td>
<td>117 (30)</td>
<td>68 (14)</td>
<td>83 (17)</td>
</tr>
<tr>
<td>&gt;70</td>
<td>59 (12)</td>
<td>44 (22)</td>
<td>110 (25)</td>
<td>61 (15)</td>
<td>133 (27)</td>
<td>124 (25)</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Body mass index (kg/m² [%])</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;28</td>
<td>308 (63)</td>
<td>96 (49)</td>
<td>288 (67)</td>
<td>254 (64)</td>
<td>213 (43)</td>
<td>211 (43)</td>
</tr>
<tr>
<td>≥28</td>
<td>183 (37)</td>
<td>99 (51)</td>
<td>145 (33)</td>
<td>140 (36)</td>
<td>277 (57)</td>
<td>284 (57)</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>PSA (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤4</td>
<td>90 (23)</td>
<td>26 (17)</td>
<td>40 (12)</td>
<td>40 (13)</td>
<td>444 (91)</td>
<td>457 (92)</td>
</tr>
<tr>
<td>4.9-9.9</td>
<td>173 (44)</td>
<td>74 (50)</td>
<td>153 (46)</td>
<td>138 (46)</td>
<td>42 (9)</td>
<td>35 (7)</td>
</tr>
<tr>
<td>10-19.9</td>
<td>72 (18)</td>
<td>28 (19)</td>
<td>68 (20)</td>
<td>62 (20)</td>
<td>3 (1)</td>
<td>3 (1)</td>
</tr>
<tr>
<td>≥20</td>
<td>61 (15)</td>
<td>21 (14)</td>
<td>72 (22)</td>
<td>62 (20)</td>
<td>1 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Unknown</td>
<td>95</td>
<td>55</td>
<td>102</td>
<td>93</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Pathologic characteristics (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nodal status*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>428 (88)</td>
<td>152 (85)</td>
<td>299 (86)</td>
<td>283 (87)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>57 (12)</td>
<td>27 (15)</td>
<td>50 (14)</td>
<td>43 (13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>6</td>
<td>25</td>
<td>86</td>
<td>69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/2</td>
<td>287 (67)</td>
<td>62 (40)</td>
<td>268 (71)</td>
<td>229 (69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3/4</td>
<td>139 (33)</td>
<td>92 (60)</td>
<td>108 (29)</td>
<td>102 (31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>8</td>
<td>23</td>
<td>9</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gleason score**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤7</td>
<td>263 (67)</td>
<td>49 (70)</td>
<td>249 (70)</td>
<td>215 (68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥7</td>
<td>127 (33)</td>
<td>108 (30)</td>
<td>102 (32)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>101</td>
<td>78</td>
<td>78</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: Cont, control; FPC, familial prostate cancer; SPC, sporadic prostate cancer; APC, aggressive prostate cancer; PSA, prostate-specific antigen; mtDNA, mitochondrial DNA.

*The 435 familial prostate cancer cases included in mtDNA SNP analysis are from 187 families. The 395 men used for nuclear-coding mitochondrial DNA SNP association are from 177 families. The two familial cases groups have 355 men in common.

† The 2 control groups have 380 men in common.

‡ These 435 men represent 213 independent subjects defined by clustering men based on maternal lineage.

§ Defined as age at diagnosis for prostate cancer patients and age at last follow-up for controls.

‖ Subjects whose nodal status was positive are excluded.

**All aggressive prostate cancer have Gleason score ≥ 8 (99 grade 8, 95 grade 9, 10 grade 10).
the CGEMS SNPs with the lowest \( P \) values (Supplementary Table 4). Of the 104 tagSNPs described above, 7 also had a \( P < 0.05 \) in the initial phase of the CGEMS study. These seven SNPs are located in five genes (MTRR, NDUFA9, NDUFS2, NDUFB9, and COX7A2). A summary of the association results for the 7 CGEMS SNPs is shown in Table 4. Five of the seven SNPs were also included in the CGEMS phase II study, two of which had \( P < 0.10 \) in that follow-up analysis. However, for both of these SNPs, the risk estimates were in opposite directions in the two studies.

Haplotype analysis was then done for each of the nuclear-encoded mitochondrial genes. For the gene NDUFV2, we observed a significant haplotype association with sporadic prostate cancer (global \( P = 8.89 \times 10^{-10} \), simulated global \( P < 1/20,000 \), as well as marginal association with aggressive prostate cancer (simulated global \( P = 0.08 \)). However, no association was observed with familial prostate cancer (global \( P = 0.82 \); Supplementary Table 5). Two haplotypes were associated with an increased risk of sporadic prostate cancer; however, these same haplotypes were associated with a decreased risk of aggressive prostate cancer. The association of NDUFV2 haplotypes with aggressive and nonaggressive prostate cancer was further explored using the CGEMS data. Our study includes five SNPs in the NDUFV2 gene, only one of which was also genotyped in the CGEMS study. We inferred the four remaining SNPs in the CGEMS subjects using Mach, a Markov Chain–based haplotyping method to infer missing genotypes in unrelated individuals (http://www.sph.umich.edu/csg/abecasis/MACH). This approach combined the sparser SNPs from the CGEMS data with the high-density SNPs of the HapMap CEU data to use the linkage disequilibrium in the CEU data to impute the unmeasured SNPs in the CGEMS data. The imputation quality for the four SNPs was excellent, with individual SNP quality scores >0.98 (on a scale of 0-1). Haplotype analyses were done with CGEMS aggressive prostate cancer and nonaggressive prostate cancer cases, resulting

![Figure 1](http://www.sph.umich.edu/csg/abecasis/MACH)

**Figure 1.** Statistical analysis of mtSNPs and prostate cancer risk. Minor allele frequencies of the 24 mtSNPs in four different populations are on the left. Odds ratios for the carriers of minor alleles and their 95% CIs are displayed. SPC, sporadic prostate cancer; FPC, familial prostate cancer; APC, aggressive prostate cancer; AILPC, all prostate cancer cases; CONT, control. Five mtSNPs were not genotyped in the cases with aggressive prostate cancer.
in global haplotype score statistical $P$ values of 0.14 and 0.06, respectively. The maximum haplotype-specific score tests were also not statistically significant ($P = 0.30$ and 0.62, respectively; Supplementary Table 5). Interestingly, the haplotype that had the strongest evidence for association with sporadic prostate cancer and aggressive prostate cancer in the Mayo data was not observed in the CGEMS data set, suggesting that the Mayo association is a chance finding.

Discussion

In this study, we tested the hypothesis that common variants in mitochondrial related genes are associated with an increased risk for prostate cancer. To accomplish this, we examined 24 SNPs in the mitochondrial genome, 11 mitochondrial haplotypes, and 376 tagSNPs for nuclear-encoded mitochondrial genes for potential associations. There are >700 known mitochondrial proteins (www.mitoproteome.org). In this study, however, we focused on the three categories of protein function: (a) respiratory chain proteins, (b) mitochondria-related cancer proteins, and (c) zinc transporter proteins. The respiratory chain proteins were chosen because some of these have been reported to cause hereditary cancer syndromes (34, 35). Mitochondria-related cancer genes were identified through comparing lists of all nuclear-encoded mitochondrial genes (www.mitoproteome.org) to those of all cancer-related genes from a variety of databases. The zinc transporters were chosen because of the potential role of these proteins in the development and progression of prostate malignancy (36). Importantly, the SNPs chosen for these genes were selected to achieve >80% coverage based on linkage disequilibrium. Overall, no statistically significant associations were detected for any of the SNPs individually or for the mtSNP haplotypes after adjusting for multiple comparisons. Our conclusions, however, are limited to relatively large effects of SNPs on prostate cancer risk. With ~490 cases and 490 controls and correcting for ~1,600 statistical tests (across all SNPs and the four group comparisons), we had 85% power to detect an odds ratio of 2.5 if the risk allele has a population frequency of 10% and 77% power to detect an odds ratio of 2.0 if the risk allele has a population frequency of 20%.

The CGEMS project, a collaborative whole genome association study initiated by National Cancer Institute, has the goal of identifying common genetic variations associated with risk for prostate and breast cancer (http://cgems.cancer.gov). The CGEMS project is projected to genotype ~8,000 prostate cancer cases over a 3 year period. The initial phase (phase I) of the CGEMS study scanned ~550,000 tagSNPs in 1,172 prostate cancer cases and 1,157 controls. The replication phase (phase II) of this study genotyped 26,958 selected SNPs in 3,941 cases and 3,964 controls. These association results, now publicly available, provide valuable information for further candidate gene selection, gene evaluation, and replication of association results.

By taking advantage of this public database, 12 of the 24 mtSNPs from the current study were also found to be present in the CGEMS data set. As with our results, a statistically significant association was not detected for any of these 12 mtSNPs in the CGEMS analysis. In
addition to these 12 SNPs, 86 additional mtSNP were also tested in the CGEMS project. Of these, five showed an incidence density adjusted score test P < 0.05. Given that ~550,000 SNPs were tested, these signals are not extreme enough to warrant statistical significance. Overall, the combined results suggest that common variants within the mitochondrial genome do not play a significant role in prostate cancer risk.

Our findings differ from those previously reported. Booker et al. (20) compared mitochondrial haplotypes in a total of 121 prostate cancer cases, 221 renal cancer cases, and 246 controls. They found that mitochondrial haplotype U was a highly significant risk factor for prostate and renal cancer versus controls (16.74% and 29.66% versus 9.35%; P = 0.019 and 0.005, respectively). Because multiple haplotypes were analyzed and the sample size was relatively small, these findings most likely represent false-positive associations. Of interest, the frequency of haplotype U among the controls in the current study is ~17%, a value closer to that reported for the cases in the Brooker study, suggesting that their control group is artificially low.

Of the 376 tagSNPs for the nuclear-encoded genes, 104 had at least 1 of 11 tests had unadjusted P < 0.05. Recognizing that these findings are most likely due to chance (marginal P values along with large number of statistical tests and modest sample size), we nevertheless further explored the significance of these findings by examining the data available through the CGEMS project. For these 104 tagSNPs, we identified 39 identical, 30 equivalent (based on linkage disequilibrium, r² = 1) and 32 associated (1 > r² ≥ 0.7) SNPs in the CGEMS phase I database (3 tagSNPs had no corresponding CGEMS linkage disequilibrium SNP). Among these 104 SNPs, 7 CGEMS SNPs had an incidence density adjusted score test P < 0.05. These seven CGEMS SNPs are located within five genes, four involved in the mitochondrial respiratory chain (NDUFS2, COX7A2, NDUFB9, and NDUFA9) and one in the mitochondria-related cancer pathway (MTRR). The three genes, NDUFS2, NDUFB9, and NDUFA9, encode components of mitochondrial complex I. NDUFB9 and NDUFA9 have NADH dehydrogenase and oxidoreductase activities. COX7A2 is a subunit of cytochrome C oxidase, the terminal component of the mitochondrial respiratory chain that catalyzes the electron transfer from reduced cytochrome C to oxygen. The gene MTRR encodes a protein that regenerates a functional methionine synthase via reductive methylation. Because methionine is an essential amino acid required for protein synthesis and one carbon

### Table 3. Summary of gene categories and tagSNP selection

<table>
<thead>
<tr>
<th>No. of genes</th>
<th>No. of selected genes</th>
<th>No. of successfully genotyped tagSNPs (genes)*</th>
<th>No. of tagSNPs (genes) that passed QC filters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mitochondrial genome</td>
<td>—</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Nuclear-encoded genes</td>
<td>—</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Respiratory chain</td>
<td>85</td>
<td>46</td>
<td>176</td>
</tr>
<tr>
<td>Cancer-related</td>
<td>94</td>
<td>32</td>
<td>161</td>
</tr>
<tr>
<td>Zinc transporters</td>
<td>24</td>
<td>16</td>
<td>81</td>
</tr>
<tr>
<td>Total</td>
<td>193</td>
<td>94</td>
<td>393</td>
</tr>
</tbody>
</table>

Abbreviation: QC, quality control.

* Nineteen SNPs were not successfully genotyped; 11 SNPs were dropped and 8 were excluded due to atypical clustering.

### Table 4. CGEMS phase I and II study analysis of Mayo SNPs of interest

<table>
<thead>
<tr>
<th>Gene</th>
<th>Mayo</th>
<th>CGEMS</th>
<th>P value</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNPs</td>
<td>Minor allele (MAF)</td>
<td>OR (95% CI)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mitochondrial respiratory chain</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDUFS2</td>
<td>rs2433367</td>
<td>A (0.41)</td>
<td>0.90 (0.77, 1.05)</td>
<td></td>
</tr>
<tr>
<td>rs2079092</td>
<td>0.24</td>
<td>1.01 (0.84, 1.22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COX7A2</td>
<td>rs2103213</td>
<td>A (0.05)</td>
<td>1.41 (0.98, 2.04)</td>
<td></td>
</tr>
<tr>
<td>rs638590</td>
<td>1.0</td>
<td>C (0.04)</td>
<td>0.004 (1.25, 2.40)</td>
<td></td>
</tr>
<tr>
<td>NDUFB9</td>
<td>rs3829037</td>
<td>G (0.23)</td>
<td>1.26 (1.05, 1.52)</td>
<td></td>
</tr>
<tr>
<td>rs730235</td>
<td>A (0.13)</td>
<td>1.07 (0.84, 1.36)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDUFA9</td>
<td>rs417684</td>
<td>A (0.27)</td>
<td>0.94 (0.79, 1.12)</td>
<td></td>
</tr>
<tr>
<td>Mitochondria-related cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTRR</td>
<td>rs1718421</td>
<td>A (0.20)</td>
<td>0.95 (0.78, 1.15)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: OR, odds ratio; MAF, minor allele frequency.

* Odds ratios and their 95% CIs are estimated using a logistic regression model assuming an ordinal (log additive) genotypic effect on prostate cancer risk. For the Mayo analyses, the reported odds ratio (95% CI) are for all PC cases versus controls.

P value from CGEMS phase II replication data set. The minimum adjusted P value over four genetic models tested (dominant, recessive, additive, and genotype trend) and four phenotype definitions (dichotomous phenotype, nonaggressive prostate cancer versus controls, aggressive prostate cancer versus controls, and all prostate cancer cases versus controls; and trichotomous phenotype) is reported.
metabolism, polymorphisms in the gene have shown an association with various diseases, including cancer (37–42). However, the seven CGEMS SNPs in the five genes are not located in coding regions or any other regulatory regions. Their functional consequences are not clear. In addition, the five of these seven SNPs that were evaluated in the CGEMS phase II replication study did not have $P < 0.05$.

In summary, we did not find a significant role of mitochondrial SNPs in prostate cancer risk. For the 78 nuclear-encoded mitochondrial genes tested, none of the tagSNPs were significant after correcting for multiple comparisons and none of the findings were replicated using the CGEMS data set.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Acknowledgments**

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.

**References**

23. Berry R, Schaid DJ, Smith JR, et al. Linkage analyses at the chromosome 1 loci 1q24-25 (HP1C1), 1q42.43 (PCAP), and 1p36 (CAPB) in families with hereditary prostate cancer. Am J Hum Genet 2000;66:539–46.
Polymorphisms in Mitochondrial Genes and Prostate Cancer Risk


Cancer Epidemiol Biomarkers Prev 2008;17:3558-3566.

Updated version
Access the most recent version of this article at:
http://cebp.aacrjournals.org/content/17/12/3558

Supplementary Material
Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2009/01/19/17.12.3558.DC1

Cited articles
This article cites 40 articles, 8 of which you can access for free at:
http://cebp.aacrjournals.org/content/17/12/3558.full#ref-list-1

Citing articles
This article has been cited by 3 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/17/12/3558.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, contact the AACR Publications Department at permissions@aacr.org.