Association of PLUNC gene polymorphisms with susceptibility to nasopharyngeal carcinoma in a Chinese population

Y He, G Zhou, Y Zhai, X Dong, L Lv, F He, K Yao

Nasopharyngeal carcinoma (NPC) is a serious health problem in the southern Chinese population, with an incidence rate ranging from 15 to 50 per 100 000. NPC is an epithelial malignancy with a striking racial and geographic distribution differences. High incidence rates are observed in the southeast Chinese population, and similar rates have been reported in these people wherever they have migrated, including Singapore, Taiwan, and Hong Kong. These incidence rates are almost 100 fold higher than in white populations. The marked racial and geographic differences in NPC susceptibility are considered a multifactorial and polygenic event with environmental and genetic components and correlation with Epstein-Barr virus infection.

Genetic mapping of variants conferring a small disease risk can identify pathways in complex disorders. Thus, the hope of resolving the genetic aetiology of NPC and the mechanisms of racial and geographic differences in NPC susceptibility has prompted a search for genetic variants in gene candidates thought to play a role in the pathogenesis of the NPC. A genetic linkage study based on affected sibling pairs collected from different Chinese populations in southeast Asia, which located the susceptibility locus to within the 6p22 region, supported the involvement of the human leukocyte antigens (HLA) in the pathogenesis of NPC. Recently, two genome-wide searches, carried out in 20 Cantonese speaking families from Guangdong province and 18 families from Hunan province in southern China, provided evidences of susceptibility loci for NPC on chromosome 4p15.1-q12 and 3p21.31-21.2, respectively. Case-control studies have established associations between specific HLA molecules and NPC in several populations including Asians, whites, and North Africans. In addition to the HLA, some non-HLA loci, including the heat shock protein 70-2 (HSP70-2), cyclin D1 (CCND1), glutathione S-transferase M1 (GSTM1), and CYP2E1 genes have also shown associations with susceptibility to NPC. Based on the fact that the disease susceptibility to NPC is determined at different functional levels, such as metabolism of carcinogenic constituents, tumour antigen presentation, cell cycle regulation, and antigenic peptide chaperones, we hypothesised that an unknown number of other unidentified genes are likely to modify the susceptibility to NPC.

PLUNC (palate, lung, and nasal epithelial clone, also designated YH1, LUNX, NASG, SPURT, and SPLUNC1) is a newly identified human homologue of the murine plunc gene, and, like the mouse gene, is specifically expressed in the upper airways and nasopharyngeal regions. The abnormal expression of PLUNC may be an important molecular event in NPC development; it was found to be downregulated in 34/48 NPC biopsies. It has been reported that CNE-2Z cells transfected with PLUNC had a significant decrease in cell proliferation. More interestingly, this gene is located on chromosome segment 20q11.2, and the loss or gain of chromosome arm 20q has been previously described in NPC.

On the basis of the potential functional relevance of PLUNC in the pathogenesis of NPC in vivo and in vitro, we hypothesised that PLUNC might be an excellent biological candidate susceptibility gene for NPC. It is expected that genetic polymorphisms within PLUNC could result in genotype dependent differences in susceptibility to NPC. In the present study, we therefore systematically screened single nucleotide polymorphisms (SNPs) in PLUNC, and investigated association of the polymorphisms in PLUNC with susceptibility to NPC in the Chinese population.
MATERIALS AND METHODS

Patients and controls

The case–control population contained 525 adult unrelated Chinese who were selected from the same population living in southeastern China between February 2000 and May 2003 (table 1). A total of 239 NPC patients were recruited from Nanfang Hospital and Jianmen Center Hospital in Guangzhou City. The only selection criterion for patients was that their NPC diagnosis had been pathologically confirmed. In 90.4% patients, the diagnosis was poorly differentiated squamous cell carcinoma. The patients (163 men; 76 women) had a mean (SD) age of 46.9 (11.2) years. The control group consisted of 286 unrelated blood donors, and laboratory and medical staff. Selection criteria for controls were no evidence of any personal or family history of cancer or other serious illness. The mean age of the control group (151 males and 135 females) was 47.2 (12.8) years. There was no significant difference between patients and control subjects in terms of mean age distribution. However, although an effort was made to obtain a frequency match on gender distribution between cases and controls, there were men in the case group than in the control group (68.2% vs 52.8%, χ² = 12.2, p<0.01). Written informed consent was obtained from all the subjects, and the study was performed with the approval of the ethics committee of Chinese Human Genome.

SNP discovery

SNP screening of all exons, relevant intron–exon boundaries and the approximately 2 kb promoter region of PLUNC (GenBank accession no. gi:17458490) was performed by PCR direct sequencing as described previously. The exact sequence regions investigated are indicated in fig 1A by first and last nucleotide position. The screening panel included 27 unrelated individuals randomly selected from the 525 individuals, regardless of disease status. The sample size gave us 95% probability of detecting alleles with a minimal frequency of 5.4%. Briefly, the primers for the target regions were designed using the web based software Primer 3.0 (http://www-genome.wi.mit.edu/). DNA samples from 27 Chinese individuals were amplified and purified, then the PCR products were sequenced using an ABI Prism Dye Terminator Sequencing kit with AmpliTaq DNA polymerase (ABI, Foster City, CA, USA) and loaded onto an ABI 3700 sequencer. SNP candidates were identified by the PolyPhred (http://www-genome.wi.mit.edu/clayton/software/) and AutoSNP (https://www.gene.cmir.cam.ac.uk/clayton/software/) programs. SNP positions and individual genotypes were confirmed by reamplifying and resequencing the SNP sites from the opposite strand. Primers are available on request.

### Table 1: Distributions of selected characteristics by case–control status

<table>
<thead>
<tr>
<th>Variable</th>
<th>NPC patients (n = 239)</th>
<th>Controls (n = 286)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>163 (68.2)*</td>
<td>151 (52.8)</td>
</tr>
<tr>
<td>Female</td>
<td>76 (31.8)</td>
<td>135 (47.2)</td>
</tr>
<tr>
<td>Mean (SD) age (years)</td>
<td>46.9 (11.2)</td>
<td>47.2 (12.8)</td>
</tr>
<tr>
<td>Histological type (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorly differentiated squamous cell carcinoma</td>
<td>216 (90.4)</td>
<td>23 (9.6)</td>
</tr>
<tr>
<td>Others*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<0.01, compared with controls; †Others include poorly differentiated adenocarcinoma (n=8), higher differentiated squamous cell carcinoma (n=6), and undifferentiated cancer (n=9).

Haplotype construction and htSNP determination

PLUNC haplotypes from the unrelated Chinese samples were assigned by the PHASE program, which uses a Bayesian approach incorporating a priori expectations of haplotypic structure from population genetic and coalescent theory. The program htSNP2 (http://www-genome.wi.mit.edu/). was used to determine the htSNPs as described previously.

SNP genotyping

We extracted genomic DNA from peripheral blood leukocytes of 5 ml whole blood using standard phenol/chloroform protocols. DNA samples were diluted to 8 ng/µl and placed in 96 well plates; each 96 well plate contained 94 samples and two controls containing water with no DNA. Three htSNPs (A-3348C, C-2128T, and C-1888T) were then selected for genotyping in the case–control population using PCR direct sequencing or PCR-RFLP analysis.

The polymorphism A-3348C was genotyped by PCR direct sequencing. The primers 5’TGCCAGGTGCCTCCTACATT-3’ and 5’TTCATTAACGTGCACAA-3’ were used for amplifying and sequencing the target region. PCR conditions were identical to those for SNP discovery, except for an annealing temperature of 56.5°C.

For the C-1888T polymorphism, a amplification using forward primer 5’AGGTGAGGACATTTAGCCTAT-3’ and reverse primer 5’AGGGCAAGAGATGAGACT-3’ was performed. An FbII recognition site was introduced by a one base mismatch (underlined) in the forward primer. PCR conditions were identical to those for SNPs discovery except for an annealing temperature of 54.8°C and a total reaction volume of 25 µl. The reaction yielded a 192 bp amplicon. An aliquot (5 µl) of PCR product was digested with 8 U of FbII (TaKaRa; Otsu, Shiga, Japan) and separated on a 3% agarose gel. The presence of the 1888T allele creates an FbII restriction site; digested amplicons from 1888T homozygotes appear as a 170 bp and a 22 bp band, homozygotes for the 1888C allele appear as a 192 bp band, and heterozygotes have all three of these bands.

For the C-2128T polymorphism, an amplification of a 164 bp fragment using forward primer 5’AGGTGAGGACATTTAGCCTAT-3’ and reverse primer 5’AGGGCAAGAGATGAGACT-3’ was performed. An FbII recognition site was introduced by a one base mismatch (underlined) in the reverse primer. PCR conditions were identical to those for C-1888T, except for an annealing temperature of 55°C. The 164 bp amplicon was digested with 8 U of FbII (TaKaRa) and separated on a 3% agarose gel. Homozygotes for the 2128T allele yield two restriction fragments of 145 bp and 19 bp after FbII digestion, homozygotes for 2128C remained uncut (164 bp band), and heterozygotes yielded all three of these bands.

Genotyping was performed by staff blinded to the subjects’ case–control status. The accuracy of genotyping data for each SNP obtained from PCR-RFLP analyses was validated by direct sequencing of a 15% masked, random sample of cases and controls.

Statistical analysis

Genotype and allele frequencies for each SNP were determined by gene counting, and the significance of deviations from Hardy-Weinberg equilibrium was tested using the random permutation procedure implemented in the Arlequin package (http://www.ubine.ch/arlequin/). Logistic regression analysis was performed to evaluate whether there was association with susceptibility to NPC for each SNP after adjustment for age and gender, and p value, odds ratios (OR), and 95% confidence intervals (CI) were calculated. Haplotype
frequencies between NPC patients and controls were compared using the $\chi^2$ test. In view of the multiple comparisons in the case-control study, the correction factor $n(m–1)$ ($m$ loci with $n$ alleles each) was applied to correct the significance level. These analyses were performed using SPSS software (version 9.0; SPSS Inc, Chicago, IL, USA).

The pairwise linkage disequilibrium (LD) measurement (Lewontin's $D^*$ and $r^2$) calculation was performed using the Arlequin package. MatInspector (http://www.gsf.de/biodv/matinjector.html) was used to search for candidate transcription factors that might bind to regions surrounding the SNPs we observed.

**RESULTS**

Resequencing of 5743 bp of *PLUNC* genomic regions in the 27 samples revealed eight SNPs (table 2, fig 1A). Haplotypes were constructed on the basis of the genotype data from these eight SNPs using PHASE software. A 98% PHASE assignment was made with more than 96% certainty. Six haplotypes were identified (fig 1B). Three common SNPs (1, 4, and 5, table 2), which captured more than 95% of the haplotype diversity observed within gene regions, were determined as htSNPs. In addition, the SNPs A-3348C, C-3098A, C-2971G, C5438T, and C5591A were in absolute LD in the Chinese sample ($(D^*)^2 = 1$ and $r^2 = 1$, fig 1C), suggesting the polymorphism A-3348C can represent the other four SNPs as marker. We therefore selected polymorphisms A-3348C, C-2128T, and C-1888T as markers for subsequent genotyping analysis.

On the basis of logistic regression analysis with adjustment for age and gender, significant associations with the susceptibility to NPC were observed with the polymorphisms C-2128T and C-1888T (table 3). Subjects carrying the 2128C/C genotype had an increased susceptibility to NPC compared with those carrying the 2128T/T genotype (OR 2.8; 95% CI 1.7 to 4.9; $p = 0.0006$). Similarly, the 1888C/C genotype was associated with an increased susceptibility to NPC compared with the 1888T/T genotype (3.3; 1.8 to 6.1; $p < 0.0001$). The associations remained significant even after correction for multiple comparisons. No significant association was found between A-3348C and susceptibility to NPC. The genotype distributions for the three SNPs were in Hardy-Weinberg equilibrium (table 3).

Furthermore, we performed the haplotype analysis for evaluating the haplotype frequencies of SNPs located nearby at the same chromosome regions, trying to derive haplotypes specifically correlated with NPC. Haplotypes based on the polymorphisms C-2128T and C-1888T were constructed. Four haplotypes, T-T, C-T, T-C and C-C, were observed. The estimated haplotype distribution was significantly different between the NPC patients and controls, indicating that individuals with haplotype C-C had significantly increased susceptibility to NPC (OR 1.86; 95% CI 1.34 to 2.56; $p = 0.0016$), even after correction for multiple comparisons (table 4).

**DISCUSSION**

In this study, we systematically screened SNPs in all exons, relevant intron-exon boundaries, and the approximately 2 kb promoter region of *PLUNC*, and from this, we identified eight SNPs. Two promoter SNPs, C-2128T and C-1888T, showed significant association with susceptibility to NPC. When compared with the statistical significance level required after correction for multiple testing, the associations remained significant. To our knowledge, this is the first report of the genetic association between *PLUNC* and susceptibility to NPC, indicating that *PLUNC* may play a role in the pathogenesis of this disorder.

![Figure 1](http://jmg.bmj.com/)  
**Figure 1** SNPs distribution, haplotypes, and LD pattern of *PLUNC*. (A) Gene structure and location of SNPs of *PLUNC*. The positions of the eight SNPs are shown above the intron-exon structure of *PLUNC*. Coding exons are marked by black boxes, and 5$'$ and 3$'$ UTRs by grey boxes. The exact sequence regions investigated are indicated by first and last nucleotide position (GenBank accession no: GI:17458490). (B) *PLUNC* haplotypes: all eight SNPs were used to construct haplotypes (SNPs are numbered as in table 2). Three htSNPs (underlined) describe 95% of the haplotype diversity observed within the *PLUNC* region. (C) The pairwise LD measurement at *PLUNC* (SNPs are numbered as in table 2). The bottom left of the table indicates the values of $r^2$, and the top right absolute values of $D^*$.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Positions and frequencies of SNPs within <em>PLUNC</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>SNP*</td>
</tr>
<tr>
<td>1</td>
<td>A-3348C</td>
</tr>
<tr>
<td>2</td>
<td>C-3098A</td>
</tr>
<tr>
<td>3</td>
<td>C-2971G</td>
</tr>
<tr>
<td>4</td>
<td>C-2128T</td>
</tr>
<tr>
<td>5</td>
<td>C-1888T</td>
</tr>
<tr>
<td>6</td>
<td>A3536G</td>
</tr>
<tr>
<td>7</td>
<td>C5438T</td>
</tr>
<tr>
<td>8</td>
<td>C5591A</td>
</tr>
</tbody>
</table>

*The position of the SNPs is relative to the first nucleotide of the open reading frame of *PLUNC*, referred to as GenBank accession number GI:17458490; †minor allele frequency.
Although a number of SNPs within PLUNC have been identified, no study has so far addressed the functional consequences of these mutations. Computer analysis indicates that sequences surrounding the polymorphism C-2128T might bind transcription factor winged helix protein (WHN)\(^23\) or Epstein-Barr virus (EBV) transcription factor R (EBVR)\(^36\) and the sequences surrounding C-1888T might bind transcription factor winged helix protein forkhead domain factor 3 (FKHD3)\(^35\), respectively. One mechanism by which this could occur is if the risky 2128C/888C alleles and the C-C haplotype may affect the susceptibility of NPC. This study, however, does not exclude the possibility that the effect observed here could be attributable to linkage disequilibrium with functional polymorphisms in nearby genes. Indeed, the loss\(^27\)–\(^28\) or gain\(^29\) of chromosome arm 20q has been previously described in NPC. Additional analyses of the nearby genes may help to verify the possibility.

In reviewing the results of this study, one potential limitation must be kept in mind. A number of association studies have addressed identification of the genes that may relate to the susceptibility to NPC.\(^14\)–\(^19\) Most of the results, however, could not be replicated in subsequent studies in other populations. Although the highly significant association between PLUNC and susceptibility of NPC derived from a biologically based a priori hypothesis, our initial findings should be independently verified in other populations with high incidence rates of NPC, such as other southern Chinese, Singaporeans, and Taiwanese.

In conclusion, we have shown that PLUNC may be a genetic risk factor for NPC in Cantonese speaking Chinese patients. Knowledge of genetic factor contributing to the pathogenesis of the NPC as presented here could lead to improved treatment and prevention of this disorder.

ACKNOWLEDGEMENTS

This study was supported by grants from the Chinese High-tech Research Program, the Medicine and Health Research Program, the Science Fund for Creative Research Groups, and the Major Program of Guangdong Province.

Table 3: The genotype and allele distribution of three SNPs in NPC patients and controls

<table>
<thead>
<tr>
<th>Polymorphism</th>
<th>NPC patients (n = 239)</th>
<th>Controls (n = 286)</th>
<th>OR (95% CI)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-3348C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genotype</td>
<td>A/A</td>
<td>9 (3.9%)</td>
<td>15 (5.4%)</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>A/C</td>
<td>90 (38.6%)</td>
<td>116 (41.6%)</td>
<td>1.3 (0.6–3.1)</td>
</tr>
<tr>
<td></td>
<td>C/C</td>
<td>134 (57.5%)</td>
<td>148 (53.0%)</td>
<td>1.5 (0.6–3.6)</td>
</tr>
<tr>
<td>Allele</td>
<td>A</td>
<td>108 (23.2%)</td>
<td>146 (26.2%)</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>358 (76.8%)</td>
<td>412 (73.8%)</td>
<td>1.2 (0.9–1.6)</td>
</tr>
<tr>
<td>C-2128T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genotype</td>
<td>T/T</td>
<td>120 (50.2%)</td>
<td>160 (56.9%)</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>T/C</td>
<td>50 (20.9%)</td>
<td>81 (28.8%)</td>
<td>0.9 (0.6–1.4)</td>
</tr>
<tr>
<td></td>
<td>C/C</td>
<td>69 (28.9%)</td>
<td>40 (14.2%)</td>
<td>2.8 (1.7–4.9)</td>
</tr>
<tr>
<td>Allele</td>
<td>T</td>
<td>290 (60.7%)</td>
<td>401 (71.4%)</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>188 (39.3%)</td>
<td>161 (28.6%)</td>
<td>1.7 (1.3–2.2)</td>
</tr>
<tr>
<td>C-1888T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genotype</td>
<td>T/T</td>
<td>64 (27.6%)</td>
<td>99 (35.1%)</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>T/C</td>
<td>121 (52.2%)</td>
<td>163 (57.1%)</td>
<td>1.2 (0.8–1.7)</td>
</tr>
<tr>
<td></td>
<td>C/C</td>
<td>47 (20.2%)</td>
<td>20 (7.1%)</td>
<td>3.3 (1.8–6.1)</td>
</tr>
<tr>
<td>Allele</td>
<td>T</td>
<td>249 (53.7%)</td>
<td>361 (64.0%)</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>215 (46.3%)</td>
<td>203 (36.0%)</td>
<td>1.6 (1.2–2.0)</td>
</tr>
</tbody>
</table>

The frequencies of genotypes are indicated in absolute values and percentages in parentheses. The sums of genotypes in patients and controls do not add up to the total number of subjects, owing to failure of genotyping for some individuals. All ORs are adjusted for age and gender.

Table 4: Haplotype distribution in the patients with NPC and in controls

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>NPC patients (2n = 478)</th>
<th>Controls (2n = 572)</th>
<th>OR (95% CI)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-T</td>
<td>187 (41.2%)</td>
<td>292 (53.7%)</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>C-T</td>
<td>57 (12.6%)</td>
<td>56 (10.3%)</td>
<td>1.59 (1.05–2.40)</td>
<td>0.027</td>
</tr>
<tr>
<td>C-C</td>
<td>90 (19.8%)</td>
<td>95 (17.5%)</td>
<td>1.48 (1.05–2.08)</td>
<td>0.024</td>
</tr>
</tbody>
</table>

The frequencies of haplotypes are indicated in absolute values and percentages in parentheses. Haplotypes are based on C-2128T and C-1888T. The sums of genotypes in patients and controls do not add up to the total number of subjects, owing to failure of genotyping for some individuals. No correction was made for testing multiple alleles.

PLUNC may directly affect the susceptibility of NPC. This study, however, does not exclude the possibility that the effect observed here could be attributable to linkage disequilibrium with functional polymorphisms in nearby genes. Indeed, the loss\(^27\)–\(^28\) or gain\(^29\) of chromosome arm 20q has been previously described in NPC. Additional analyses of the nearby genes may help to verify the possibility.

In reviewing the results of this study, one potential limitation must be kept in mind. A number of association studies have addressed identification of the genes that may relate to the susceptibility to NPC.\(^14\)–\(^19\) Most of the results, however, could not be replicated in subsequent studies in other populations. Although the highly significant association between PLUNC and susceptibility of NPC derived from a biologically based a priori hypothesis, our initial findings should be independently verified in other populations with high incidence rates of NPC, such as other southern Chinese, Singaporeans, and Taiwanese.

In conclusion, we have shown that PLUNC may be a genetic risk factor for NPC in Cantonese speaking Chinese patients. Knowledge of genetic factor contributing to the pathogenesis of the NPC as presented here could lead to improved treatment and prevention of this disorder.
REFERENCES


