Genetic variants in \textit{RUNX3} and risk of bladder cancer: a haplotype-based analysis

Zhizhong Zhang$^{1,2}$, Shizhi Wang$^2$, Melin Wang$^{1,2}$, Na Tong$^2$, Guangbo Fu$^4$ and Zhengdong Zhang$^{1,2,*}$

$^1$Cancer Center, $^2$Department of Molecular and Genetic Toxicology and Department of Epidemiology and Biostatistics, Nanjing Medical University, Nanjing 210029, China and $^3$Department of Urology, The Huai-An First Affiliated Hospital, Nanjing Medical University, Huai-An 223300, China

To whom correspondence should be addressed. Department of Molecular and Genetic Toxicology, School of Public Health, Nanjing Medical University, 140 Hanzhong Road, Nanjing 210029, China. Tel: +86 25 86862937; Fax: +86 25 86527613; Email: zdzhang@njmu.edu.cn

Transforming growth factor-\(\beta\) (TGF-\(\beta\)) is a multifunctional growth factor that plays important roles in many biological processes, whereas \textit{RUNX3} is a target of TGF-\(\beta\)-mediated tumor suppressor pathway. In humans, \textit{RUNX3} inactivation may lead to the cancer development, including bladder cancer. To determine whether the \textit{RUNX3} polymorphisms are associated with risk of bladder cancer, we conducted a case–control study of 368 bladder cancer patients and 368 cancer-free controls to assess the associations between the \textit{RUNX3} tagging single-nucleotide polymorphisms (tSNPs) and bladder cancer risk. In the single-locus analysis, we found a significantly increased risk of bladder cancer associated with the SNP7 rs760805 A allele (adjusted odds ratio = 1.97, 95% confidence interval = 1.44–2.69), compared with the AT/TT genotype. Haplotype-based association analysis revealed that the increased risk of bladder cancer was significantly associated with two haplotypes TATGCCAAAA (2.37, 1.16–4.83) and AGCTTGAGAG (2.70, 1.08–6.72) that included the rs760805 A allele. Multifactor dimensionality reduction (MDR) analysis identified a significant more than multiplicative interaction between the SNP7 rs760805 AA and smoking and an additive interaction between the SNP3 rs11249206 TT and smoking on bladder cancer risk. The SNP3 rs11249206, SNP5 rs1395621, SNP7 rs760805, SNP8 rs2236852 and the trichotomized cumulative smoking were the five factors best predicted by the MDR models. When the variables were combined and dichotomized and fitted into the MDR model, the subjects carrying the combined risk stratum had a significantly increased risk for bladder cancer (6.37, 4.57–8.87, \(P = 7.03 \times 10^{-28}\)). These results suggest that the genetic variants in \textit{RUNX3} may modulate the risk of bladder cancer.

Introduction

Bladder cancer, the ninth common frequent cancer, is an important health problem worldwide, and it was reported that 357 000 cases of bladder cancer occurred in 2002 and 145 000 patients die of this disease annually (1). In China, bladder cancer is the 10th most common cancer, accounting for 17 365 deaths in 2005, and the mortality has increased steadily between 1991 and 2005 (2). Among urinary system tumors, the incidence or mortality of bladder cancer is the 17th most common cancer, accounting for 17 365 deaths in 2005, and the mortality has increased steadily between 1991 and 2005 (1). Notably, bladder cancer is the 10th most common cancer in men (3). It has been estimated that tobacco smoking and occupational exposure to chemical carcinogens are the strongest risk factors for bladder cancer (4). Although many people are exposed to these risk factors, only a fraction of the exposed develop bladder cancer in their lifetime, suggesting a variation in individual susceptibility to bladder carcinogenesis in the general population.

Abbreviations: CI, confidence interval; CVC, cross-validation consistency; OR, odds ratio; SNP, single-nucleotide polymorphism; TGF-\(\beta\), transforming growth factor-\(\beta\); tSNP, tagging single-nucleotide polymorphism.

Materials and methods

Study subjects

The methods of recruiting study subjects of Han Chinese have been described previously (15). The present study population included 368 eligible patients with newly diagnosed transitional cell carcinoma of bladder and 368 cancer-free control subjects recruited between January 2003 and April 2008. Those cases who had previous cancer, metastasized cancer from other or unknown origin and previous radiotherapy or chemotherapy were excluded. Approximately 95% of the eligible patients contacted chose to participate. The cancer-free control subjects were blood unrelated to the cases, had no individual history of cancer and were recruited from those who accompanied the patients to the hospital and who were seeking health care. Controls were frequency-matched to the cases by age (±5 years) and sex. Among the willing respondents contacted for recruitment, the response rate was >85%. Each eligible subject was interviewed to obtain demographic information as well as data on smoking status and alcohol use. Those subjects who smoked daily for >1 year were defined as ever smokers. Ever smokers who had quit smoking for >1 year were defined as former smokers and the other smokers as current smokers. Pack-years [cigarettes per day/20 \times years smoked] were calculated to indicate the cumulative smoking dose and the smokers were further categorized into...
two groups; light smokers (pack-years < 22) and heavy smokers (pack-years > 22) according to the median pack-year value of the ever smokers. Those subjects who consumed three or more alcohol drinks per week for at least 1 year were considered ever drinkers, and the rest were defined as never drinkers. After having provided their written informed consent, each subject donated 5 ml of blood used for DNA extraction. The research protocol was approved by the institutional review board of Nanjing Medical University.

SNP selection, tSNPs identification and genotyping

We selected the tSNPs by using genotype data obtained from unrelated Han Chinese in Beijing individuals in the HapMap (HapMap Data Rel 21a/Phase II, Jan07, on NCBI B35 assembly, dbsNP b125). For RUNX3, we selected the tSNPs that had a minor allele frequency >0.10 in Han Chinese in Beijing within a 67 kb region spanning the RUNX3 gene (including 2 kb upstream and 2 kb downstream) using the pairwise option of the Haploview 4.0 software (16). The LD plot based on D’ values was made of the region (supplementary Figure 1 is available at Carcinogenesis Online). An r2 of 0.8 was selected as a threshold for the analyses, and performance was defined as the number of SNPs in the evaluated population that had an r2 of >0.8 with the tSNPs among the total number of SNPs. Because the reported mutations that may abolish the tumor-suppressive effect of RUNX3 were very rare (7,12), we did not include them in this study. As a result, 10 tSNPs were selected at a resolution of one SNP per 6.7 kb that captured all variant alleles with a mean r2 of 0.965. The ‘‘y’’ number and relative position of the 10 tSNPs are shown in Figure 1A and their LD was calculated and visually presented by the Haploview 4.0 software in Figure 1B.

The selected SNPs were genotyped in all 736 subjects by polymerase chain reaction–restriction fragment length polymorphism method. The tSNPs information, primers and restriction enzymes of polymorphisms are all listed in supplementary Table I (available at Carcinogenesis Online). The polymorphism analysis was performed independently by two persons in a blind fashion. About 1% of polymerase chain reaction products were randomly selected and confirmed by sequencing (data not shown), and >10% of the samples were randomly selected for repeated genotyping. The results were 100% concordant.

Statistical analysis

Chi-square test was used to evaluate differences in frequency distributions of selected demographic variables and known risk factors, such as tobacco smoking and alcohol use as well as each allele and genotype of the RUNX3 polymorphisms, between the cases and controls. Unconditional univariate and multivariate logistic regression analyses were performed to obtain the crude and adjusted odds ratios (ORs) for risk of bladder cancer and their 95% confidence intervals (CIs). Hardy–Weinberg equilibrium of the genotype distribution of the controls was tested by a goodness-of-fit chi-square test. To reduce the potential of spurious findings due to multiple testing, we applied 10 000 permutations to the empirically derived P-values.

The computation of LD between SNPs was estimated using the normalized measure of allelic association D’ and r2, and the characterization of these patterns was determined by Haploview 4.0 software (16). Haplotypes were generated by the Haploview software using the approach given by Gabriel et al. (17). Also, the permutation test in Haploview was used to obtain a measure of significance corrected for multiple testing bias. This method takes account of multiple testing by comparing the maximum statistic obtained from the individual tests with its permutation distribution obtained by 10 000 random relabelings of the cases and controls. A significant result was obtained if the observed maximum r2 statistic fell in the top 5% of its permutation distribution. Haplotypes were inferred by the Haplo.stats program (18), which is a score test based on generalized linear models. This program also provides several different global and haplotype-specific tests for association and allows the possibility to include several covariates. Empirical P-values, based on 10 000 simulations, were computed for the global score test and each of the haplotype-specific score tests.

Potential locus–locus and gene–environment interactions were performed by using the non-parametric multifactor dimensionality reduction (MDR) software (version 0.5.1) (19) with 10 tSNPs and the trichotomized cumulative smoking dose because smoking is an established risk factor for bladder cancer. The MDR analysis redefined a variable of high and low risk by combining information on several loci and environmental factors that may interact in the cancer etiology. The new one-dimensional variable can be evaluated for its ability to classify and predict disease status using cross-validation and permutation testing. A 10-fold cross-validation test was conducted such that a model was constructed based on 9/10 of the data (training data) and then evaluated using the remaining 1/10 of the data (testing data). The fitness of an MDR model was assessed by estimating the testing accuracy. Models that were true positive would have estimated testing accuracy of >0.5. The cross-validation consistency was a measure of how many times of 10 divisions of the data that MDR found in the same best model. The sign test counted the number of cases, k, where the testing accuracy was >0.5 of 10 cross-validation cases. Among these models, the one with the highest cross-validation consistency and the highest testing accuracy was selected and further evaluated using permutation testing.

Then, interaction dendrogram was employed. The attributes (i.e. SNPs) that were strongly interacting appeared close together at the leaves of the tree, whereas those that do not interact appeared far from one another. The colors used in the dendrogram comprised a spectrum of colors representing a continuum from synergy to redundancy. The colors range from red representing a high degree of synergy, orange a lesser degree and gold the midway point between synergy and redundancy. On the redundancy end of the spectrum, the highest degree was represented by the blue color with a lesser degree represented by green.

For each variable generated in the best model, gene–environment interaction was also evaluated by logistic regression analysis and the Stata software. We tested the null hypotheses of multiplicative gene–environment interaction and evaluated departure from multiplicative interaction model by including main effect variables and their product terms in the logistic regression model (20). When the test for multiplicative interaction was not rejected, the additive interaction was further assessed by a bootstrapping test by using the Stata software (version 8.2; StataCorp LP, College Station, TX). Finally, all the variables in the best model were combined and dichotomized according to the MDR software, and their ORs and 95% CIs for bladder cancer risk were calculated accordingly. All the statistical analyses were performed with SAS software (version 9.1.3), unless indicated otherwise.

Results

Characteristics of the study subjects

The frequency distributions of selected characteristics of the cases and controls are presented in supplementary Table II (available at Carcinogenesis Online). Briefly, there was no significant difference in the distribution of age (P = 0.476) and sex (P = 0.510) between the cases and controls. However, there were more ever smokers (55.7%) and ever alcohol users (48.1%) among the cases than among the controls (40.8 and 37.2%, respectively), and these differences were statistically significant (P < 0.001 for smoking; P = 0.003.
for alcohol use). Specifically, light smokers (<22 pack-years) had a 1.74-fold (95% CI = 1.21–2.50) and heavy smokers (>22 pack-years) had a 1.92-fold (95% CI = 1.34–2.74) increased risk, compared with non-smokers (P < 0.001 for the trend test). Thus, these variables were further adjusted for in the multivariate logistic regression analysis.

**Individual SNP association analysis**

The SNP IDs, locations and allele frequencies are shown in supplementary Table III (available at Carcinogenesis Online). All RUNX3 genotype distributions in the controls were consistent with those expected from the Hardy–Weinberg equilibrium model. As shown in supplementary Table III (available at Carcinogenesis Online), only the allele frequency of rs760805 was significantly different between the cases and controls (P = 0.002), and the association remained significant after 10 000 permutations (P = 0.011). The genotype distributions of the RUNX3 polymorphisms in cases and controls are summarized in supplementary Table IV (available at Carcinogenesis Online). The single-SNP analysis indicated that the genotype frequency of SNP7 rs760805 was significantly different between the cases and controls (P < 0.001). Multivariate logistic regression analyses revealed that a significantly decreased risk was associated with the AT and TT genotype (adjusted OR = 0.46, 95% CI = 0.33–0.64 for AT; OR = 0.63, 95% CI = 0.42–0.97 for TT), compared with the AA genotype.

In further stratification analysis for the SNP7 rs760805 (Table I), when we used the combined genotype AT/TT as the reference, we found that the AA genotype was associated with an increased risk of bladder cancer (adjusted OR = 1.97, 95% CI = 1.44–2.69). This increased risk was also more pronounced among subgroups of aged >65 years (2.83, 1.77–4.53), male (2.16, 1.52–3.07), smokers (3.46, 2.16–5.54) and drinkers (2.60, 1.58–4.27). Furthermore, we found that the risk of bladder cancer associated with the AA genotype was increased in a dose–response manner as the pack-years increased (1.15, 0.74–1.78 for non-smokers; 3.05, 1.56–5.98 for light smokers and 4.04, 2.05–8.00 for heavy smokers).

**Association between haplotypes and risk of bladder cancer**

Haplotype analysis including all 10 ISNPs was performed, and there were >100 possible haplotypes derived from the known genotypes. Haplotypes with a frequency <0.01 in both the cases and controls were pooled into a single group, and the remaining 20 haplotypes were analyzed with Haplo.stats. As shown in Table II, after adjustment for age, sex, pack-years of smoking and alcohol use, the risk of bladder cancer was statistically significantly increased among individuals carrying the haplotype TATCCAAA (adjusted OR = 2.37, 95% CI = 1.16–4.83) and haplotype AGCTTGTGAG (adjusted OR = 2.70, 95% CI = 1.08–6.72), compared with those carrying the most common haplotype AGCTTGTGAG. Interestingly, the only difference between the risk haplotype AGCTTGTGAG and the most common haplotype AGCTTGTGAG was the SNP7 rs760805 A allele, suggesting that the SNP7 rs760805 A allele may be the risk allele for bladder cancer, which is consistent with the individual SNP association analysis. For the association between total haplotypes and bladder cancer risk, the global score statistic was 26.79 (degree of freedom = 20) with a P-value of 0.141 and a simulation-based P-value of 0.138 when no covariates were included; and when adjusted for age, sex, pack-years of smoking and alcohol use, the global score statistic was 28.90 with a P-value of 0.090 and a simulation-based P-value of 0.086 (data not shown).

**Locus–locus and gene–environment–interactions**

All 10 ISNPs in RUNX3 and the dichotomized cumulative smoking dose were included in the MDR analysis. Figure 2A summarizes the best interaction models obtained from the MDR analysis. In the one-factor model, SNP7 rs760805 was the best attribute for predicting bladder cancer risk (testing accuracy = 55.48%; cross-validation consistency (CVC) = 9/10). But the best interaction model was the five-factor model (i.e. SNP3 rs11249206, SNP5 rs1395621, SNP7 rs760805, SNP8 rs2236852 and pack-years of smoking), with an improved testing accuracy to 61.22% and a perfect CVC of 10 that was statistically significant (P = 0.001), as determined empirically by the permutation testing. All models including six or more factors had a decrease in both testing accuracy and CVC.

---

**Table II. Associations between risk of bladder cancer and frequencies of inferred haplotypes on the basis of the observed genotypes in bladder cancer cases and cancer-free controls**

<table>
<thead>
<tr>
<th>Haplotypes</th>
<th>Haplotype frequencies</th>
<th>P</th>
<th>P-sim</th>
<th>Adjusted OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGCTTGTGAG</td>
<td>0.0924 0.0929</td>
<td></td>
<td></td>
<td>0.2156 0.2189 1.00</td>
</tr>
<tr>
<td>TATCCGAGA</td>
<td>0.0622 0.1120</td>
<td></td>
<td></td>
<td>0.0550 0.0546 0.80</td>
</tr>
<tr>
<td>TATTTGTGAG</td>
<td>0.0731 0.0765</td>
<td></td>
<td></td>
<td>0.7474 0.7438 1.12</td>
</tr>
<tr>
<td>TATCCGTGAG</td>
<td>0.0608 0.0764</td>
<td></td>
<td></td>
<td>0.1954 0.1960 0.87</td>
</tr>
<tr>
<td>TATCCCAAAG</td>
<td>0.0386 0.0464</td>
<td></td>
<td></td>
<td>0.9017 0.8997 1.01</td>
</tr>
<tr>
<td>AGCTTGAAGA</td>
<td>0.0381 0.0316</td>
<td></td>
<td></td>
<td>0.7318 0.7289 0.99</td>
</tr>
<tr>
<td>TATTTGAGA</td>
<td>0.0201 0.0308</td>
<td></td>
<td></td>
<td>0.8902 0.8907 1.19</td>
</tr>
<tr>
<td>TATCCAAAAA</td>
<td>0.0653 0.0286</td>
<td></td>
<td></td>
<td>0.0319 0.0293 2.37</td>
</tr>
<tr>
<td>TATCCGAAGA</td>
<td>0.0180 0.0283</td>
<td></td>
<td></td>
<td>0.4147 0.4254 0.76</td>
</tr>
<tr>
<td>TATCCCTGAG</td>
<td>0.0360 0.0260</td>
<td></td>
<td></td>
<td>0.1792 0.1782 2.68</td>
</tr>
<tr>
<td>TATTTCTGAG</td>
<td>0.0071 0.0256</td>
<td></td>
<td></td>
<td>0.9480 0.9488 1.21</td>
</tr>
<tr>
<td>AGCTTGGAGA</td>
<td>0.0365 0.0224</td>
<td></td>
<td></td>
<td>0.0413 0.0429 2.70</td>
</tr>
<tr>
<td>TATCCCAAAG</td>
<td>0.0267 0.0178</td>
<td></td>
<td></td>
<td>0.2289 0.2318 2.14</td>
</tr>
<tr>
<td>TATTTGGAGA</td>
<td>0.0065 0.0146</td>
<td></td>
<td></td>
<td>0.0616 0.0587 0.36</td>
</tr>
<tr>
<td>TATCCGAGA</td>
<td>0.0183 0.0106</td>
<td></td>
<td></td>
<td>0.7352 0.7302 1.32</td>
</tr>
<tr>
<td>AGCTTGAAGA</td>
<td>0.0171 0.0100</td>
<td></td>
<td></td>
<td>0.5345 0.5348 2.56</td>
</tr>
<tr>
<td>TATCCCAAAG</td>
<td>0.0133 0.0079</td>
<td></td>
<td></td>
<td>0.6659 0.6635 2.50</td>
</tr>
<tr>
<td>TATTTGAGA</td>
<td>0.0227 0.0074</td>
<td></td>
<td></td>
<td>0.0398 0.0328 2.10</td>
</tr>
<tr>
<td>TATCCGAAGA</td>
<td>0.0210 0.0027</td>
<td></td>
<td></td>
<td>0.0663 0.0655 3.95</td>
</tr>
<tr>
<td>TATCCGAGA</td>
<td>0.0139 0.0019</td>
<td></td>
<td></td>
<td>0.2098 0.2137 2.80</td>
</tr>
</tbody>
</table>

*aAdjusted for age, sex, pack-years of smoking and alcohol use.

*bTwo-sided χ² test.

*cSimulation-based P-value.

*dAdjusted for age, sex, pack-years of smoking and alcohol use.
Then, we applied the interaction dendrogram and graph to determine whether there was a synergistic relationship among these five factors in the best model (Figure 2B). The SNP4 and SNP10 were placed on the same branch, whereas the SNP3, SNP5, SNP7, SNP8 and pack-years of smoking were placed on the other branch. The closer position in the diagram clearly showed that the five-factor model might have a synergistic interaction effect on modulating risk of bladder cancer. An interaction effect might also exist between the SNP4 rs7551188 and SNP10 rs2282718 polymorphisms.

For each variable generated in the best five-factor model, we first evaluated greater than multiplicative gene–smoking interaction by logistic regression analyses between each dichotomized genotypes and smoking status (smokers and non-smokers) (Table III). The results indicated a supermultiplicative interaction between SNP7 and smoking ($P = 0.001$). For SNP7 rs760805, compared with non-smokers who carried the AT/TT genotype, non-smokers with the AA genotype had a 1.15-fold (95% CI = 0.74–1.78) increased risk of bladder cancer and smokers with the AT/TT had a 1.14-fold (95% CI = 0.77–1.70) increased risk of bladder cancer, whereas smokers with AA genotype had the highest risk, with the OR being 3.91 (95% CI = 2.43–6.29), which is 3-fold greater than the product of the OR for non-smokers with AA genotype and the OR for smokers with the

### Table III. Risk of bladder cancer associated with selected tSNPs in MDR by smoking

<table>
<thead>
<tr>
<th>tSNPs</th>
<th>Cases/controls</th>
<th>OR (95% CI)$^a$</th>
<th>Cases/controls</th>
<th>OR (95% CI)$^a$</th>
<th>$P$ for interaction (multiplicative)</th>
<th>$P$ for interaction (additive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP3 rs11249206</td>
<td>CC/TC</td>
<td>1.00</td>
<td>TT</td>
<td>1.00</td>
<td>0.04 (0.68–1.58)</td>
<td>0.0169</td>
</tr>
<tr>
<td>Non-smokers</td>
<td>96/130</td>
<td>1.44 (0.95–2.18)</td>
<td>67/87</td>
<td>2.14 (1.46–3.61)</td>
<td>0.487</td>
<td>0.210</td>
</tr>
<tr>
<td>Smokers</td>
<td>106/94</td>
<td>2.29 (1.4–3.61)</td>
<td>94/54</td>
<td>2.39 (1.6–3.91)</td>
<td>0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SNP5 rs1395621</td>
<td>TT/CT</td>
<td>1.00</td>
<td>CC</td>
<td>1.00</td>
<td>0.07 (0.7–1.65)</td>
<td>0.047</td>
</tr>
<tr>
<td>Non-smokers</td>
<td>104/143</td>
<td>1.57 (1.06–2.33)</td>
<td>59/75</td>
<td>2.09 (1.33–3.31)</td>
<td>0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Smokers</td>
<td>118/99</td>
<td>2.09 (1.33–3.31)</td>
<td>79/50</td>
<td>2.09 (1.33–3.31)</td>
<td>0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SNP7 rs760805</td>
<td>AT/TT</td>
<td>1.00</td>
<td>AA</td>
<td>1.15 (0.74–1.78)</td>
<td>0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Non-smokers</td>
<td>109/151</td>
<td>1.14 (0.77–1.70)</td>
<td>54/67</td>
<td>2.43 (2.43–6.29)</td>
<td>0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Smokers</td>
<td>96/113</td>
<td>3.91 (2.43–6.29)</td>
<td>105/36</td>
<td>1.40 (0.84–2.32)</td>
<td>0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SNP8 rs2236852</td>
<td>GA/AA</td>
<td>1.00</td>
<td>GG</td>
<td>0.86 (0.55–1.37)</td>
<td>0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Non-smokers</td>
<td>118/150</td>
<td>1.00</td>
<td>45/65</td>
<td>1.37 (1.05–1.77)</td>
<td>0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Smokers</td>
<td>151/108</td>
<td>1.40 (0.84–2.32)</td>
<td>48/41</td>
<td>1.40 (0.84–2.32)</td>
<td>0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>MDR best model 5</td>
<td>0$^b$</td>
<td>1.00</td>
<td>1$^b$</td>
<td>3.91 (2.43–6.29)</td>
<td>0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>0$^b$</td>
<td>102/259</td>
<td>6.37 (4.57–8.87)</td>
<td>254/102</td>
<td>6.37 (4.57–8.87)</td>
<td>0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

$^a$Adjusted for age, sex and alcohol use.

$^b$New class level in MDR: ‘1’ represents that the combinations exceed the ratio threshold, whereas ‘0’ represents that the combinations do not exceed the ratio threshold.

$^cP = 7.03 \times 10^{-28}$. 

---

[Fig. 2.](#) The MDR models, interaction dendrogram for locus–locus and gene–environment interactions on bladder cancer risk. (A) Summary of the MDR interaction models. (B) Interaction dendrogram. The colors indicated the type of interactions. Red denoted a high degree of interaction, orange a lesser degree and green weak interaction. The hierarchical cluster analysis placed SNP4 and SNP10 on the same branch but SNP3, SNP5, SNP7, SNP8 and pack-years of smoking on the other branch.]
AT/TT genotype. In the later additive interaction analyses, SNP3 rs11249206 appeared to have an additive interaction with smoking \((P = 0.032)\). However, similar interaction was not found between the other tSNPs and smoking.

Finally, all the variables in the best five-factor model were combined and dichotomized according to the MDR software. Subjects carrying the combined risk stratum of the five factors had a significant 6.37-fold increased risk \((95\% \text{ CI } 4.57–8.87)\) for bladder cancer with a \(P\)-value as small as \(7.03 \times 10^{-28}\).

**Discussion**

In this study, we selected 10 tSNPs in the RUNX3 gene and investigated their associations with the risk of bladder cancer in a Chinese population. We found that the SNP7 rs760805 AA genotype was associated with a statistically significantly increased risk of bladder cancer, compared with the AT/TT genotype; and this increased risk was more pronounced among subgroups of aged \(>65\) years, male smokers and drinkers. Statistical evidence was also observed for a more than multiplicative interaction between the SNP7 rs760805 polymorphism and tobacco smoking and an additive interaction between the SNP3 rs11249206 polymorphism and tobacco smoking.

Furthermore, we found that the haplotypes TATCCCCAAA and AGCTTGGAGAG were associated with a significantly increased risk of bladder cancer. The MDR analysis identified a significant five-factor interaction model including SNP3, SNP5, SNP7, SNP8 and the trichotomized cumulative smoking dose, which suggested that \(>6\)-fold significantly increased risk of bladder cancer was associated with the combined genetic variations in RUNX3 and smoking. To the best of our knowledge, this is the first study that investigates the association between the RUNX3 tSNPs and risk of bladder cancer.

Given RUNX3 is an integral component of TGF-\(\beta\)-induced signaling pathway that plays an important role in carcinogenesis (8), it is possible that genetic variations in RUNX3 may affect TGF-\(\beta\) signaling and contribute to the development of cancers. It is known that missense mutations occurred in the Runt domain can abolish the DNA-binding ability of RUNX3 (12), which could result in inactivation of RUNX3 and ultimately the development of bladder cancer. Hu et al. (21) investigated the association between the RUNX3 C364T polymorphism and risk of gastric cancer in a Chinese population. However, no significant association was observed. To date, there is no report on the association between the RUNX3 polymorphism and risk of bladder cancer.

In the present study, the most significant finding was the association between SNP7 rs760805 T > A and bladder cancer risk. In the single-locus analysis, SNP7 AA genotype was associated with a significantly increased risk of bladder cancer. In the haplotype analysis, both of the two risk haplotypes contained the SNP7 rs760805 A allele. Furthermore, SNP7 was the best one-factor model for predicting bladder cancer risk in MDR analyses. This significant association appears to be consistent across all analyses, suggesting that SNP7 rs760805 may be involved in the development of bladder cancer in this study population. SNP7 rs760805 is an intronic polymorphism whose functional consequences is less intuitive, but intronic polymorphisms have been reported to be associated with a variety of chronic diseases including breast cancer (22), essential hypertension (23) and type II diabetes (24). Recent data from immunoprecipitation and gene expression experiments suggest that up to 40% of transcription factor-binding sites are located within introns (25). When we used the Genomax program (http://www.genomax.de/) to analyze the putative transcription factors for RUNX3, we found that the SNP7 rs760805 T > A locus is centered in a 13-nucleotide sequence that is nearly identical to the myelin transcription factor 1 consensus binding site, which is known to play a role in the development of neurons and oligodendroglia in the mammalian central nervous system (26) and is reported to be overexpressed in human brain tumors (27).

Consistently, RUNX3 is proven to be necessary for neurogenesis of the dorsal root ganglia (28). Although the exact molecular mechanisms of how the SNP7 rs760805 variant works to have an effect on risk of bladder cancer are unknown, it is possible that this tSNP may have an effect on the gene expression and that it may be in LD with other functional variants. However, these hypotheses require further investigation.

In this study, we also found that the increased risk associated with the SNP7 rs760805 AA genotype was more pronounced among the older male, smokers and drinkers. The possible explanation is that individuals in those subgroups may be more probably to have been exposed to some risk factors involved in the etiology of bladder cancer, such as tobacco smoking, alcohol use or occupational chemical carcinogen (29,30). However, this hypothesis needs to be confirmed in further studies.

Accumulating evidence has shown that the effect of single genetic variation may be dependent on other genetic variations (gene–gene interaction) and environmental factors (gene–environment interaction) (31). It is conceivable that bladder cancer is probably the result of interactions between multiple genetic variations and environmental factors. However, such complex multifactor interactions are often difficult to detect by using traditional parametric statistical methods such as logistic regression because of the sparseness of data in high dimensions (32). Apart from type I errors due to multiple tests, parametric statistical methods can also be problematic, leading to an increase in type II errors and a decrease in power (19). To address this problem, we adopted both parametric statistical methods and non-parametric MDR approach to detect and characterize the gene–environment interaction between the RUNX3 gene and smoking. With non-parametric and genetic model-free MDR, multifaceted genotypes and environmental factors are pooled into high-risk and low-risk groups, effectively reducing the dimensionality from \(n\) dimensions to one dimension (33). In order to interpret the locus–locus and gene–environment interaction, we further applied the interaction denogram to determine the nature of their joint effects. As the results shown, the five-factor model (i.e. SNP3, SNP5, SNP7, SNP8 and pack-years of smoking) is the best model for predicting bladder cancer risk. Notably, although all the SNPs except for SNP7 were not associated with bladder cancer risk in the single-locus analysis, they might act in concert to modulate the risk of bladder cancer. In addition, we tested the gene–environment interaction with MDR-selected variables using parametric statistical methods, and the results suggested a multiplicative joint effect between the SNP7 and smoking in identifying bladder cancer risk. Besides, SNP3 also appeared to have an additive joint effect between with smoking. Because smoking is an established risk factor for bladder cancer (34) and has a destructive effect on immune responses (35), such an interaction may be expected.

One strength of this study is the adoption of a haplotype-based analysis to assess the association between the RUNX3 genetic variations and risk of bladder cancer. Haplotype-based tSNP analysis assumes that unassayed and risk-related SNPs may be linked with one or more assayed SNPs. As cancer is a disease involving multiple SNPs in multiple genes, a haplotype-based analysis may be more powerful than the approach of analyzing a single locus. In our present study, we found that two haplotypes and best five-factor model were significantly associated with the risk of bladder cancer, suggesting that these variants and environment factors may play a joint role in the development of bladder cancer. Our results further support the notion that a single polymorphism may only contribute a modest effect, whereas the combined variants of a gene may provide a more comprehensive evaluation of genetic susceptibility in candidate gene with low penetration.

Several limitations in our study need to be addressed. First, our study was a hospital-based case–control study. Thus, selection bias may occur. However, the allele frequencies of the genotyped SNPs in our controls were similar to those obtained from the HapMap Project. Moreover, we recruited subjects from two different large hospitals in Jiangsu province and the response rates in the cases and controls were \(>95\) and \(85\%\), respectively. We believe that the selection bias, if any, is unlikely to be substantial. Second, the sample size (368 cases and 368 controls) in this study may not be large enough to detect a small effect from very low-penetrance SNPs and evaluate
gene–environment interaction adequately, though we have 80% power at a 0.05 significance level to detect an OR of 1.55 or greater with an exposure frequency of 30% under the current sample size. Third, in addition to tobacco smoking and alcohol use, we did not obtain enough information on occupational exposure that may interact with the RUNX3 gene. Finally, because the tSNPs were chosen to maximize SNP tagging for genetic variation rather than for the functionality of the RUNX3 gene, further mechanistic studies are needed to determine how these tSNPs influence the protein function.

In conclusion, our study shows that some representative genetic variants in RUNX3 may modulate the risk of bladder cancer. Besides, these variants along with some environment factors could play a joint role in the development of bladder cancer. Studies with ethnically diverse populations and functional evaluation are warranted to confirms our findings.

Supplementary material

Supplementary Figure 1 and Tables I–IV can be found at http://carcin.oxfordjournals.org/

Funding

National Natural Science Foundation of China (30571583 and 30271105); PhD Programs Foundation of Ministry of Education of China (20060312002); Natural Science Foundation of Jiangsu Province (BK2006231); Post-doctoral Science Foundation of China (20060390293); Post-doctoral Science Foundation of Jiangsu Province (06010149); ‘Qinglan Project’ Foundation for the Young Academic Leader of Jiangsu Province (2006).

Acknowledgements

We thank Qinbo Yuan, Lin Yuan and Jianfeng Shao for their assistance in recruiting the subjects; Dongmei Wu, Haixia Zhu, Yuanjuan Tian and Miaomiao Wang for their technical support; Chunping Li and Aiping Li for their laboratory assistance; Guangfu Jin and Jianling Bo for their assistance in data analysis and Qingyi Wei (Department of Epidemiology, The University of Texas MD Anderson Cancer Center) for critical comments and scientific editing.

Conflict of Interest Statement: None declared.

References
