Contribution of ZBTB20 Polymorphisms to Esophageal Cancer Risk Among the Chinese Han Population

Shuyong Yu¹*, Guihong Yuan²*, Feixiang Hu¹*, Yongyu Li², Zhuang Chen², Ronglin Zhang³, Ping Li³, Zhaowei Chen², Jian Song⁴

¹Department of Gastrointestinal Surgery, Hainan Cancer Hospital, Haikou, 570100, People’s Republic of China; ²Department of Gastroenterology, Hainan Cancer Hospital, Haikou, 570100, People’s Republic of China; ³Digestive Endoscopy Center, Hainan Cancer Hospital, Haikou, 570100, People’s Republic of China; ⁴Department of Gastroenterology, Southern University of Science and Technology Hospital, Shenzhen, 518000, People’s Republic of China

*These authors contributed equally to this work

Background: ZBTB20 was overexpressed in esophageal cancer (EC). The study aimed to identify genotypes of ZBTB20 polymorphisms and their correlation with EC occurrence in a Chinese Han population.

Methods: Four single nucleotide polymorphisms (SNPs) in ZBTB20 were randomly selected for genotyping through Agena MassARRAY system among 525 EC patients and 522 healthy controls. Multiple genetic models were applied to assess the association of ZBTB20 polymorphisms with EC susceptibility by calculating odds ratios (ORs) with 95% confidence intervals (CIs).

Results: Rs10934270 was associated with lower EC susceptibility (OR = 0.64, p = 0.004) with statistical power >90% in overall analysis. Specifically, the correlation of rs10934270 with EC susceptibility was found in subgroups including patients with esophageal squamous cell carcinoma (ESCC), males, subjects aged ≤65 years, subjects with BMI ≤ 24 kg/m², and smokers. Rs9841504 might be a risk-increasing factor for ESCC. Moreover, rs9288999 in subjects aged ≤65 years and rs73230162 in females were related to lower EC risk.

Conclusion: Our research is the first to report that ZBTB20 rs10934270 is associated with reduced EC susceptibility in the Chinese Han population. These data provide a scientific basis for understanding the influence of the ZBTB20 gene on EC occurrence.

Keywords: esophageal cancer, ZBTB20, genetic polymorphisms, genotype–phenotype analyses, FPRP analysis

Introduction

Esophageal cancer (EC) is a highly aggressive cancer of the digestive system with an increasing incidence and a 5-year survival rate of about 15–25%.¹ EC is the eighth most common cause of cancer worldwide (604,100 new cases) overall, and the mortality of EC ranks sixth in malignancy-related mortality (544,076 new deaths).² Esophageal squamous cell carcinoma (ESCC) accounts for 88.84% of all EC cases in China and is the main pathological type worldwide.³ There are obvious gender differences in EC, EC is 2 to 8 times more common in men than in women in most areas of the world because of the use of tobacco and alcohol, especially in the developed countries. However, the incidence of EC in male and female can be very close in some regions where smoking and drinking play only a minor role in EC development (eg, Huai’an in Jiangsu,⁴,⁵ and Taihang mountain region)⁶,⁷. EC is considered to be a complex disease caused by the interaction of multiple factors such as genetics and environmental factors. It is now generally accepted that unfavorable habits (tobacco and alcohol), poor nutritional status, caloric intake and obesity are the main risk factors for EC.⁸ Previous studies on the attributable risk of EC in China showed 46% of EC (51% in men and 33% in women) were attributable to tobacco smoking, alcohol drinking, and low vegetable and fruit intake. Tobacco smoking and alcohol use are major risk
factors for squamous cell carcinoma. A negative correlation between overall obesity, as measured by body mass index (BMI) and risk of ESCC has been reported. Nevertheless, not everyone exposed to risk factors will eventually develop EC. Recently, the role of genetic polymorphisms in the occurrence of EC has been reported.

Zinc finger and BTB domain containing 20 (ZBTB20, also named ZNF288), a dendritic cell-derived BTB/POZ zinc finger (DPZF), belongs to a family of transcription factors with BTB/POZ domain (N-terminal) and zinc finger domain (C-terminal). ZBTB20 is considered as a key transcriptional repressor, and its deficiency may lead to high expression levels of alpha-fetoprotein. ZBTB20 plays a role in many processes, including glucose homeostasis, and tumor progression. Studies have reported that the high expression of ZBTB20 is closely related to the prognosis of hepatocellular carcinoma. In addition, studies have also clarified that ZBTB20 promotes the migration and invasion of gastric cancer cells. ZBTB20 was overexpressed in glioblastoma, and ZBTB20 knockdown inhibited glioblastoma progression. These findings indicated that ZBTB20 acted as a tumor progression gene in tumor progression. Based on the GEPIA database (http://gepia.cancer-pku.cn/), ZBTB20 was overexpressed in EC. A study has evaluated the relation between ZBTB20 rs9841504 and ESCC susceptibility, but no association has been found. Furthermore, the contribution of other polymorphisms in ZBTB20 to EC has not been investigated.

In this study, SNPs (rs10934270, rs9288999, rs9841504, and rs73230612) in the intronic region of ZBTB20 were randomly selected for genotyping based on the following criteria: 1) minor allele frequency (MAF) >0.05 and $r^2$ of ≥0.8 in the Chinese Han population from the Chinese Han population in Beijing of the 1000 Genomes Project and the dbSNP database; 2) Hardy-Weinberg equilibrium (HWE) >0.05, and the call rate for genotyping >99.5%; 3) previous literatures. HaploReg v4.1 (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php) was used to determine the frequencies of these SNPs in other populations.
populations. HaploReg v4.1, RegulomeDB (https://regulome.stanford.edu/regulome-search/) and GTEx Portal (https://gtexportal.org/home/) are databases used to predict the potential functions of selected SNPs. Genotyping was determined through Agena MassARRAY system (Agena, San Diego, CA, USA) with built-in software. The MassARRAY platform is based on matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS).26–28 The general principle of the MassARRAY platform is based on the difference in primer masses caused by sequence changes. The primer sequences were shown in Suppl_Table 1. All samples were genotyped using a double-blind model. For quality control, approximately 10% of randomly chosen samples were run in duplicate with 100% consistency.

Statistical Analysis
Student’s t-test or χ² test was used to identify differences in baseline data between EC patients and healthy controls, as appropriate. A goodness-of-fit χ² test was used for HWE analysis in controls. Multiple genetic models were applied to assess the association of ZBTB20 polymorphisms with EC susceptibility. Logistic regression analysis adjusted for age and gender was employed to calculate odds ratios (ORs) with 95% confidence intervals (CIs) by PLINK software. Furthermore, subgroup analyses stratified by histological, demographic, and behavioral data were also estimated. False-positive report probability (FPRP) analysis was applied to assess noteworthy associations by setting a threshold of 0.2 and a prior probability of 0.1. The SNP–SNP interactions were analyzed through multifactor dimensionality reduction (MDR) (version 3.0.2) software. SPSS 18.0 software (SPSS Inc., Chicago, IL, USA) was used for data analysis, and a p value < 0.05 indicated statistically different, whereas an adjusted p value < 0.05/4 was considered significant after Bonferroni correction.

Results
Participants’ Characteristics
A total of 1047 subjects were recruited, including 525 EC patients (63.92 ± 9.18 years) and 522 healthy controls (63.70 ± 7.07 years). The demographic and clinical characteristics of all subjects were shown in Table 1. The ratio of male to female in the two groups was 3:1. The distributions of age (p = 0.657) and sex (p = 0.956) were not statistically different. However, there were significant differences in BMI (p < 0.001), smoking (p = 0.040), and drinking (p < 0.001) between the two groups. Among 525 EC patients, 73.7% (N = 387) were confirmed to have ESCC. There were 177 cases of lymphatic metastasis and 189 cases in stage III/IV.

The Association Between ZBTB20 SNPs and EC Risk
Table 2 summarized the information on ZBTB20 four polymorphisms. The genotype distribution of these polymorphisms in the control group was in accordance with HWE (p > 0.05), indicating that the selected population had a good representativeness. The detection rate of genotyping was >99.5%. The MAFs of these SNPs in healthy controls were similar to those in Asians. The potential functions of these polymorphisms explored by HaploReg v4.1 and RegulomeDB were displayed in Table 2. The results from HaploRegv4.1 indicated that these selected SNPs were associated with the regulation of enhancer histones, changed motifs, DNase, and selected eQTL hits. Besides, rs10934270 might be related to transcription factor (TF) binding, any motif and DNase peak. Based on the GTEx Portal database, the genotypes of rs10934270 (p = 1.4e-12, Suppl_Figure 1A) and rs73230612 (p = 3.4e-6, Suppl_Figure 1B) were related to ZBTB20mRNA expression in whole blood.

The allele and genotype frequencies of these polymorphisms are listed in Table 3. The T allele frequency of rs10934270 in EC patients was lower than that in controls. Based on multiple genetic models, rs10934270 was related to lower EC susceptibility under the allele (T vs C, OR = 0.64, 95% CI: 0.47–0.87, p = 0.004), genotype (CT vs CC, OR = 0.68, 95% CI: 0.49–0.96, p = 0.029), dominant (CT-TT vs CC, OR = 0.65, 95% CI: 0.46–0.91, p = 0.011), and log-additive (OR = 0.64, 95% CI: 0.47–0.88, p = 0.005) models. No effect of other SNPs on EC risk was found overall. After Bonferroni correction, the significant association of rs10934270 with EC risk in the allele, dominant and log-additive models still existed.
Table 1 The Information of the Patients with Esophageal Cancer and Healthy Controls

| Characteristics                  | Case     | Control   | p    |
|----------------------------------|----------|-----------|------|
| Number                           | 525      | 522       |      |
| Age (mean ± SD, years)           | 63.92 ± 9.18 | 63.70 ± 7.07 | 0.657 |
| > 65                             | 236 (45.0%) | 191 (36.6%)  |      |
| ≤ 65                             | 289 (55.0%) | 331 (63.4%)  |      |
| Gender                           |          |           | 0.956|
| Male                             | 390 (74.3%) | 387 (74.1%)  |      |
| Female                           | 135 (25.7%) | 135 (25.9%)  | <0.001|
| BMI (kg/m²)                      |          |           |      |
| ≤ 24                             | 435 (82.9%) | 162 (31.0%)  |      |
| > 24                             | 74 (14.1%)  | 193 (37.0%)  |      |
| Missing                          | 16 (3.0%)   | 167 (32.0%)  |      |
| Smoking                          |          |           | 0.040|
| Yes                              | 245 (46.7%) | 117 (22.4%)  |      |
| No                               | 274 (52.2%) | 179 (34.3%)  |      |
| Missing                          | 6 (1.1%)    | 226 (43.3%)  |      |
| Drinking                         |          |           | <0.001|
| Yes                              | 125 (23.8%) | 116 (22.2%)  |      |
| No                               | 354 (67.4%) | 155 (29.7%)  |      |
| Missing                          | 46 (8.8%)   | 251 (48.1%)  |      |
| Pathological type                |          |           |      |
| Squamous cell carcinoma          | 387 (73.7%) |            |      |
| Other                            | 138 (26.3%) |            |      |
| Lymph node metastasis            |          |           |      |
| Yes                              | 177 (33.7%) |            |      |
| No                               | 162 (30.9%) |            |      |
| Missing                          | 186 (35.4%) |            |      |
| Clinical stages                  |          |           |      |
| I+II                             | 159 (30.3%) |            |      |
| III+IV                           | 189 (36.0%) |            |      |
| Missing                          | 177 (33.7%) |            |      |

Note: p values were calculated by χ² test or the Student’s t-test.

Stratified Analysis by Histological Data
The stratified analysis by subtype, lymph node metastasis, and stage was performed to assess the association of candidate SNPs with histological data of EC patients. We found that rs10934270 was a protective factor against ESCC (allele: T vs C, OR = 0.66, p = 0.017; dominant: CT-TT vs CC, OR = 0.67, p = 0.031; and log-additive: OR = 0.67, p = 0.020), while rs9841504 was related to an elevated risk of ESCC (genotype: GG vs CC, OR = 2.59, p = 0.045; and recessive: GG vs CC-CG, OR = 2.56, p = 0.047, Table 3). However, no significant associations were detected between ZBTB20 SNPs and lymph nodes metastasis and staging in EC patients (data no shown).

Stratified Analysis by Demographic and Behavioral Data
Tables 4 and 5 summarized the results of subgroup analyses to explore the interaction of ZBTB20 variants and demographic data with EC risk. According to gender-stratified analysis (Table 4), rs10934270 was related to reduced EC risk in males under the allele (T vs C, OR = 0.64, p = 0.017), genotype (CT vs CC, OR = 0.66, p = 0.042), dominant (CT-TT vs CC, OR = 0.64, p = 0.025), and log-additive (OR = 0.65, p = 0.020) models. Among females, rs73230612 was a protective factor against EC in the allele (C vs T, OR = 0.70, p = 0.043), genotype (CC vs TT, OR = 0.42, p = 0.035), and log-additive (OR = 0.67, p = 0.034) models.
Table 2 The Information of Four Polymorphisms on the ZBTB20 Gene

| SNP ID  | Chr: Position | Alleles (Ref/Alt) | MAF                     | Call rate | HWE       | Frequency* | RegulomeDB                                                                 |
|---------|---------------|------------------|-------------------------|-----------|-----------|------------|---------------------------------------------------------------------------|
|         |               |                  |                         |           |           |            | Case:O(HET):E(HET):p:AFR:AMR:ASN:EUR:Pair of SNPs with \( r^2 \geq 0.8^a \) |
| rs10934270 | 3:114384900  | C/T              | 0.069:0.103             | 100.0%    | 0.176:0.186:0.239:0.05:0.39:0.11:0.43 | DNAse, Motifs Changed, Selected eQTL hits |
| rs9288999 | 3:114429080  | A/G              | 0.416:0.430             | 99.9%     | 0.488:0.490:0.929:0.67:0.41:0.43:0.24 | Enhancer histone marks, Motifs changed |
| rs9841504 | 3:114643917  | C/G              | 0.147:0.139             | 100.0%    | 0.251:0.239:0.359:0.34:0.17:0.16:0.08 | Enhancer histone marks, Motifs changed |
| rs73230612 | 3:115131989  | T/C              | 0.404:0.416             | 100.0%    | 0.475:0.486:0.652:0.81:0.83:0.40:0.89 | Motifs changed |

Notes: *Data from Haploreg (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php). RegulomeDB (https://regulome.stanford.edu/regulome-search/).

Abbreviations: SNP, Single nucleotide polymorphism; MAF, Minor allele frequency; HWE, Hardy-Weinberg equilibrium; O(HET), Observed heterozygosity frequency; E(HET), Expected heterozygosity frequency; AFR, African; AMR, American; ASN, Asian; EUR, European.
In age stratification, the median age (65-year) was set as the cut-off value for all subjects. In order to explore the effect of age on EC risk, we divided all subjects into two groups as ≤65 years and >65 years. The contributions of rs10934270 (allele: T vs C, OR = 0.40, \( p = 3.25 \times 10^{-5} \); genotype: CT vs CC, OR = 0.40, \( p = 3.29 \times 10^{-4} \); dominant: CT-TT vs CC, OR = 0.38, \( p = 1.07 \times 10^{-4} \); and log-additive: OR = 0.41, \( p = 1.15 \times 10^{-4} \) and rs9288999 (genotype: GG vs
Table 4 Stratified Analysis by Gender and Age for the Associations Between ZBTB20 Polymorphisms and the Risk of Esophageal Cancer

| SNP ID  | Model   | Genotype | Control | Case | OR (95% CI) | p-value | Control | Case | OR (95% CI) | p-value |
|---------|---------|----------|---------|------|-------------|---------|---------|------|-------------|---------|
|         |         |          | Males   | Females |            |         | Males   | Females |            |         |
| rs10934270 | Allele  | C        | 699     | 730   | 1           |         | 237     | 248   | 1           |         |
| Genotype |         | T        | 75      | 50    | 0.64 (0.44–0.93) | 0.017   | 33      | 22    | 0.64 (0.36–1.12) | 0.118   |
|         |         | CC       | 317     | 342   | 1           |         | 105     | 113   | 1           |         |
|         |         | CT       | 65      | 46    | 0.66 (0.44–0.99) | 0.042   | 27      | 22    | 0.76 (0.41–1.41) | 0.381   |
|         |         | TT       | 5       | 2     | 0.37 (0.07–1.92) | 0.237   | 3       | 0     | /           | /       |
|         | Dominant| CC/ CT-TT | 317/70 | 342/48 | 0.64 (0.43–0.95) | 0.025   | 105/30  | 113/22 | 0.68 (0.37–1.26) | 0.219   |
|         | Recessive| CC-CT/ TT | 382/5  | 388/2 | 0.39 (0.08–2.04) | 0.267   | 132/3   | 135/0 | /           | /       |
|         | Log-additive |        |         |       | 0.65 (0.45–0.93) | 0.020   |         |       | 0.64 (0.36–1.13) | 0.122   |
|         |         |          |         |       |             |         |         |       |             |         |
| rs73230612 | Allele  | T        | 459     | 452   | 1           |         | 151     | 174   | 0.70 (0.50–0.99) | 0.043   |
| Genotype |         | C        | 315     | 328   | 1.06 (0.86–1.29) | 0.588   | 119     | 96    | 0.72 (0.42–1.22) | 0.220   |
|         |         | TT       | 142     | 132   | 1           |         | 39      | 52    | 1           |         |
|         |         | TC       | 175     | 188   | 1.16 (0.84–1.58) | 0.364   | 73      | 70    | 0.72 (0.42–1.22) | 0.220   |
|         |         | CC       | 70      | 70    | 1.07 (0.71–1.61) | 0.739   | 23      | 13    | 0.42 (0.19–0.94) | 0.035   |
|         | Dominant| TT/ TC-CC | 142/245 | 132/258 | 1.13 (0.84–1.52) | 0.409   | 39/96   | 52/83 | 0.65 (0.39–1.08) | 0.097   |
|         | Recessive| TT-TC/ CC | 317/70 | 320/70 | 0.99 (0.68–1.42) | 0.943   | 112/23  | 122/13 | 0.52 (0.25–1.08) | 0.078   |
|         | Log-additive |        |         |       | 1.05 (0.86–1.29) | 0.605   |         |       | 0.67 (0.46–0.97) | 0.034   |
|         |         |          |         |       |             |         |         |       |             |         |
| rs9288999 | Allele  | A        | 437     | 467   | 1           |         | 157     | 146   | 1.20 (0.85–1.69) | 0.292   |
| Genotype |         | G        | 337     | 313   | 0.87 (0.71–1.06) | 0.173   | 111     | 124   | 1.38 (0.80–2.38) | 0.251   |
|         |         | AA       | 124     | 143   | 1           |         | 46      | 37    | 1.41 (0.69–2.86) | 0.347   |
|         |         | AG       | 189     | 181   | 0.83 (0.61–1.14) | 0.251   | 65      | 72    | 1.39 (0.82–2.33) | 0.219   |
|         |         | GG       | 74      | 66    | 0.78 (0.51–1.17) | 0.225   | 23      | 26    | 1.15 (0.62–2.14) | 0.657   |
|         | Dominant| AA/ AG-GG | 124/263 | 143/247 | 0.82 (0.61–1.10) | 0.179   | 46/88   | 37/98 | 1.21 (0.85–1.71) | 0.284   |
|         | Recessive| AA-AG/ GG | 313/74 | 324/66 | 0.86 (0.60–1.25) | 0.434   | 111/23  | 109/26 | 1.21 (0.85–1.71) | 0.284   |
|         | Log-additive |        |         |       | 0.87 (0.71–1.07) | 0.183   |         |       | 1.21 (0.85–1.71) | 0.284   |
|         |         |          |         |       |             |         |         |       |             |         |
| rs9841504 | Allele  | C        | 667     | 672   | 1           |         | 232     | 224   | 1.25 (0.79–2.00) | 0.342   |
| Genotype |         | G        | 107     | 108   | 1.00 (0.75–1.34) | 0.990   | 38      | 46    | 1.25 (0.79–2.00) | 0.342   |
|         |         | CC       | 286     | 291   | 1           |         | 98      | 95    | 1           |         |
|         |         | CG       | 95      | 90    | 0.93 (0.67–1.29) | 0.660   | 36      | 34    | 0.97 (0.56–1.68) | 0.916   |
|         |         | GG       | 6       | 5     | 1.46 (0.51–4.15) | 0.483   | 1       | 6     | 6.20 (0.73–52.44) | 0.094   |
|         | Dominant| CC/ CG-GG | 286/101 | 291/99 | 0.96 (0.70–1.32) | 0.801   | 98/37   | 95/40 | 1.11 (0.66–1.89) | 0.693   |
|         | Recessive| CC-CG/ GG | 381/6  | 381/9 | 1.48 (0.52–4.21) | 0.460   | 134/1   | 129/6 | 6.24 (0.74–52.59) | 0.092   |
|         | Log-additive |        |         |       | 1.00 (0.75–1.33) | 0.987   |         |       | 1.25 (0.78–1.99) | 0.350   |

(Continued)
Table 4 (Continued).

| SNP ID     | Model          | Genotype | Control | Case | OR (95% CI)     | p-value | Control | Case | OR (95% CI)     | p-value |
|------------|----------------|----------|---------|------|-----------------|---------|---------|------|-----------------|---------|
| rs10934270 | Allele         |          |         |      |                 |         |         |      |                 |         |
|            |                | C        | 351     | 429  | 1.14 (0.70–1.84) | 0.607   | 585     | 549  | 0.40 (0.26–0.62) | 3.25×10⁻⁵* |
|            |                | T        | 31      | 43   | 1.19 (0.70–2.01) | 0.519   | 77      | 29   | 0.40 (0.25–0.66) | 3.29×10⁻⁴** |
|            |                | CT       | 29      | 41   | 0.85 (0.05–13.81) | 0.909   | 63      | 27   | 0.19 (0.02–1.54) | 0.119   |
|            | Genotype       | CC       | 161     | 194  |                 |         | 261     | 261  |                 |         |
|            |                | CT       | 29      | 41   |                 |         | 63      | 27   |                 |         |
|            |                | TT       | 1       | 1    |                 |         | 7       | 1    |                 |         |
|            | Dominant       | CC/ CT   | 161/30  | 194/42| 1.18 (0.70–1.98) | 0.538   | 261/70  | 261/28| 0.38 (0.24–0.62) | 1.07×10⁻⁴** |
|            | Recessive      | CC/CT/ TT| 190/1  | 235/1| 0.83 (0.05–13.40) | 0.893   | 324/7   | 288/1| 0.21 (0.03–1.73) | 0.147   |
|            | Log-additive   |          |         |      | 1.15 (0.70–1.90) | 0.571   |         |      | 0.41 (0.26–0.64) | 1.15×10⁻⁴** |
| rs9288999  | Allele         | A        | 225     | 266  | 1.11 (0.84–1.46) | 0.455   | 369     | 347  | 0.84 (0.67–1.06) | 0.143   |
|            |                | G        | 157     | 206  |                 |         | 291     | 231  |                 |         |
|            | Genotype       | AA       | 65      | 78   | 1.09 (0.70–1.71) | 0.697   | 243     | 220  | 0.89 (0.61–1.30) | 0.546   |
|            |                | AG       | 95      | 110  | 2.56 (0.67–9.76) | 0.168   | 105     | 102  | 1.85 (0.50–6.81) | 0.358   |
|            |                | GG       | 31      | 48   |                 |         | 66      | 44   | 0.93 (0.64–1.35) | 0.707   |
|            | Dominant       | AA/AG    | 65/126 | 78/158| 1.00 (0.66–1.51) | 0.995   | 105/225 | 102/187| 0.77 (0.55–1.09) | 0.146   |
|            | Recessive      | AA/AG/ GG| 160/31| 188/48| 1.26 (0.76–2.09) | 0.363   | 264/66  | 245/44| 0.61 (0.39–0.94) | 0.025   |
|            | Log-additive   |          |         |      | 1.07 (0.82–1.41) | 0.613   |         |      | 0.77 (0.60–0.97) | 0.027   |
| rs9841504  | Allele         | C        | 329     | 393  |                 |         | 570     | 503  | 0.92 (0.67–1.28) | 0.635   |
|            |                | G        | 53      | 79   | 1.25 (0.86–1.82) | 0.250   | 42      | 35   | 0.89 (0.61–1.30) | 0.546   |
|            | Genotype       | CC       | 141     | 166  |                 |         | 243     | 220  | 1.85 (0.50–6.81) | 0.358   |
|            |                | CG       | 47      | 61   | 1.09 (0.70–1.71) | 0.697   | 84      | 63   | 0.93 (0.64–1.35) | 0.707   |
|            |                | GG       | 3       | 9    | 2.56 (0.67–9.76) | 0.168   | 4       | 6    | 1.85 (0.50–6.81) | 0.358   |
|            | Dominant       | CC/CG    | 141/50 | 166/70| 1.18 (0.77–1.82) | 0.451   | 243/88  | 220/69| 0.93 (0.64–1.35) | 0.707   |
|            | Recessive      | CC/CG/ GG| 188/3 | 227/9| 2.5 (0.66–9.49)  | 0.177   | 327/4   | 283/6| 1.90 (0.32–6.99) | 0.335   |
|            | Log-additive   |          |         |      | 1.23 (0.85–1.8)  | 0.275   |         |      | 0.99 (0.70–1.38) | 0.933   |
| rs73230612 | Allele         | T        | 223     | 283  |                 |         | 387     | 343  | 0.96 (0.77–1.21) | 0.753   |
|            |                | C        | 159     | 189  | 0.94 (0.71–1.23) | 0.640   | 275     | 235  | 0.96 (0.77–1.21) | 0.753   |
|            | Genotype       | TT       | 62      | 87   |                 |         | 119     | 97   |                 |         |
|            |                | TC       | 99      | 109  | 0.80 (0.52–1.23) | 0.309   | 149     | 149  | 1.19 (0.83–1.70) | 0.353   |
|            |                | CC       | 30      | 40   | 0.92 (0.51–1.65) | 0.781   | 63      | 43   | 0.83 (0.51–1.35) | 0.459   |
|            | Dominant       | TT/TC    | 62/129 | 87/149| 0.83 (0.55–1.24) | 0.365   | 119/212 | 97/192| 1.08 (0.77–1.52) | 0.650   |
|            | Recessive      | TT/TC/ CC| 161/30| 196/40| 1.05 (0.62–1.77) | 0.862   | 268/63  | 246/43| 0.75 (0.49–1.17) | 0.204   |
|            | Log-additive   |          |         |      | 0.93 (0.70–1.23) | 0.593   |         |      | 0.96 (0.76–1.21) | 0.706   |

Notes: p values were calculated by logistic regression analysis with adjustments for age and gender. Bold p < 0.05 respects the data is statistically significant. *p indicate that after Bonferroni correction (p < 0.05/4) means the data is statistically significant.

Abbreviations: SNP, single nucleotide polymorphism; OR, odds ratio; 95% CI, 95% confidence interval.
### Table 5: Stratified Analysis by BMI and Smoking for the Associations Between ZBTB20 Polymorphisms and the Risk of Esophageal Cancer

| SNP ID | Model | Genotype | Control | Case | OR (95% CI) | p-value | Control | Case | OR (95% CI) | p-value |
|--------|--------|----------|---------|------|-------------|---------|---------|------|-------------|---------|
|        |        |          | BMI, kg/m² | > 24 | ≤ 24 |          |        |       |              |         |
| rs10934270 | Allele | C | 356 | 135 | I | 1.14 (0.58–2.26) | 0.701 | 277 | 812 | 0.42 (0.28–0.63) | 2.11×10⁻⁵* |
|          | Genotype | CC | 164 | 61 | I | 1.30 (0.62–2.73) | 0.484 | 120 | 379 | 0.48 (0.30–0.77) | 0.002* |
|          | Dominant | CC/ CT-TT | 164/29 | 61/13 | I | 1.26 (0.6–2.64) | 0.534 | 120/42 | 379/56 | 0.44 (0.28–0.69) | 3.64×10⁻⁴* |
|          | Recessive | CC/ CT/ TT | 192/1 | 74/0 | I | 1.21 (0.59–2.47) | 0.605 | 157/5 | 433/2 | 1.16 (0.66–2.05) | 0.157 |
| rs9288999 | Allele | A | 212 | 83 | I | 0.95 (0.65–1.40) | 0.810 | 186 | 507 | 0.98 (0.76–1.27) | 0.874 |
|          | Genotype | AA | 58 | 24 | I | 0.96 (0.58–1.63) | 0.610 | 55 | 148 | 0.95 (0.66–1.38) | 0.850 |
|          | Dominant | AA/ AG-GG | 58/135 | 24/50 | I | 0.88 (0.49–1.59) | 0.671 | 55/106 | 148/287 | 0.82 (0.44–1.59) | 0.990 |
|          | Recessive | AA-AG/ GG | 154/39 | 59/15 | I | 0.93 (0.47–1.85) | 0.838 | 131/30 | 359/76 | 0.92 (0.57–1.47) | 0.997 |
| rs9841504 | Allele | C | 334 | 125 | I | 1.18 (0.69–2.01) | 0.538 | 281 | 741 | 0.96 (0.78–1.24) | 0.496 |
|          | Genotype | CC | 141 | 54 | I | 1.01 (0.60–1.71) | 0.957 | 120 | 318 | 1.00 (0.71–1.41) | 0.963 |
|          | Dominant | CC/ CG-GG | 141/52 | 54/20 | I | 0.96 (0.52–1.80) | 0.907 | 120/42 | 318/117 | 1.08 (0.71–1.64) | 0.713 |
|          | Recessive | CC-CG/ GG | 193/0 | 71/3 | I | 1.16 (0.66–2.05) | 0.613 | 161/1 | 426/12 | 1.17 (0.81–1.70) | 0.408 |
| rs73230612 | Allele | T | 214 | 93 | I | 0.97 (0.50–1.09) | 0.122 | 190 | 517 | 0.97 (0.75–1.26) | 0.807 |
|          | Genotype | TT | 59 | 31 | I | 0.59 (0.32–1.09) | 0.092 | 54 | 149 | 0.50 (0.30–0.84) | 0.036 |
|          | Dominant | TT/ TC-CC | 59/134 | 31/43 | I | 0.60 (0.34–1.07) | 0.082 | 54/108 | 149/286 | 0.98 (0.67–1.44) | 0.918 |
|          | Recessive | TT-TC/ CC | 155/38 | 62/12 | I | 0.85 (0.41–1.77) | 0.667 | 136/26 | 368/67 | 0.93 (0.56–1.52) | 0.761 |
|          | Log-additive | 0.75 (0.50–1.12) | 0.159 | 0.97 (0.74–1.27) | 0.816 |

(Continued)
Table 5 (Continued).

| SNP ID   | Model   | Genotype | Control | Case | OR (95% CI)       | p-value | Control | Case | OR (95% CI)       | p-value |
|----------|---------|----------|---------|------|--------------------|---------|---------|------|--------------------|---------|
|          |         | Allele   |         |      |                    |         |         |      |                    |         |
| rs10934270 |        | T        | 206     | 459  | 1                  |         |          |      |                    |         |
|          |         | C        | 28      | 31   | 0.50 (0.29–0.85)   | 0.009*  |          |      | 319               | 0.66 (0.42–1.05) | 0.077  |
|          |         | G        | 91      | 216  | 1                  |         |          |      |                    |         |
|          |         | C        | 24      | 27   | 0.43 (0.24–0.79)   | 0.006*  |          |      | 319               | 0.81 (0.49–1.36) | 0.431  |
|          |         | T        | 2       | 2    | 0.38 (0.05–2.72)   | 0.334   |          |      | 319               | /       | /     |
|          |         | CC/CT-TT | 91/26   | 216/29 |        |         |          |      | 319               | 0.72 (0.43–1.18) | 0.188  |
|          |         | CC-CT/ TT| 115/2   | 243/2 |        |         |          |      | 319               | /       | /     |
|          |         | Dominant |         |      | 0.47 (0.28–0.81)   | 0.006*  |          |      |                    |         |
|          |         | Recessive |        |      |                    |         |          |      |                    |         |
|          |         | Log-additive |     |      |                    |         |          |      |                    |         |
| rs9841504 |        | Allele   | 214     | 420  | 1                  |         |          |      |                    |         |
|          |         | C        | 20      | 70   | 1.78 (1.06–3.01)   | 0.029   |          |      | 301               | 0.96 (0.66–1.38) | 0.810  |
|          |         | G        | 97      | 181  | 1                  |         |          |      |                    |         |
|          |         | CC       | 20      | 58   | 1.51 (0.84–2.69)   | 0.166   |          |      | 124               | 0.77 (0.5–1.18) | 0.229  |
|          |         | CG       | 0       | 6    | /                  | /       |          |      | 2                 | 2.69 (0.57–12.74) | 0.211  |
|          |         | Dominant | 97/20   | 181/64 |        |         |          |      | 124               | 0.84 (0.55–1.28) | 0.413  |
|          |         | Recessive |        |      |                    |         |          |      |                    |         |
|          |         | Log-additive |     |      |                    |         |          |      |                    |         |
| rs9288999 |        | Allele   | 137     | 289  | 1                  |         |          |      |                    |         |
|          |         | A        | 97      | 201  | 0.98 (0.72–1.35)   | 0.912   |          |      | 204               | 0.99 (0.75–1.29) | 0.925  |
|          |         | G        | 44      | 86   | 1                  |         |          |      |                    |         |
|          |         | AA       | 49      | 117  | 1.10 (0.66–1.83)   | 0.711   |          |      | 62                | 1.10 (0.72–1.70) | 0.653  |
|          |         | AG       | 24      | 42   | 0.85 (0.45–1.62)   | 0.630   |          |      | 37                | 0.89 (0.52–1.54) | 0.688  |
|          |         | GG       | 44/73   | 86/159 |        |         |          |      | 62                | 1.04 (0.69–1.55) | 0.854  |
|          |         | Dominant |        |      | 1.72 (1.01–2.95)   | 0.047   |          |      |                    |         |
|          |         | Recessive |        |      |                    |         |          |      |                    |         |
|          |         | Log-additive |     |      |                    |         |          |      |                    |         |
| rs73230612 |       | Allele   | 130     | 275  | 1                  |         |          |      |                    |         |
|          |         | T        | 104     | 215  | 0.98 (0.71–1.34)   | 0.886   |          |      | 215               | 0.88 (0.67–1.16) | 0.380  |
|          |         | C        | 38      | 79   | 1                  |         |          |      |                    |         |
|          |         | TT       | 54      | 117  | 1.02 (0.61–1.71)   | 0.949   |          |      | 63                | 0.91 (0.60–1.38) | 0.658  |
|          |         | TC       | 25      | 49   | 0.93 (0.49–1.76)   | 0.831   |          |      | 27                | 0.73 (0.40–1.33) | 0.302  |
|          |         | Dominant | 38/79   | 79/166 |        |         |          |      | 63                | 0.87 (0.58–1.29) | 0.484  |
|          |         | Recessive |        |      |                    |         |          |      |                    |         |
|          |         | Log-additive |     |      |                    |         |          |      |                    |         |

Notes: p values were calculated by logistic regression analysis with adjustments for age and gender. Bold p < 0.05 respects the data is statistically significant. *p indicate that after Bonferroni correction (p < 0.05/4) means the data is statistically significant.

Abbreviations: SNP, single nucleotide polymorphism; OR, odds ratio; 95% CI, 95% confidence interval.
AA, OR = 0.56, \( p = 0.019 \); recessive: GG vs AA-AG, OR = 0.61, \( p = 0.025 \); and log-additive: OR = 0.77, \( p = 0.027 \) to a decreased EC risk (Table 4) were observed in subjects aged \( \leq 65 \) years. The significant association of rs10934270 with EC risk in subjects aged 65 or younger still existed after Bonferroni correction.

In the subgroup with BMI \( \leq 24 \text{ kg/m}^2 \) (Table 5), the risk-reducing association of rs10934270 with EC occurrence was found under the allele (T vs C, OR = 0.42, \( p = 2.11 \times 10^{-5} \)), genotype (CT vs CC, OR = 0.48, \( p = 0.002 \), and TT vs CC, OR = 0.13, \( p = 0.018 \)), dominant (CT-TT vs CC, OR = 0.44, \( p = 3.64 \times 10^{-4} \)), recessive (TT vs CC-CT, OR = 0.15, \( p = 0.027 \), and log-additive (OR = 0.45, \( p = 1.35 \times 10^{-4} \)) models. The significant association of rs10934270 with EC risk remained after Bonferroni correction.

In smokers (Table 5), rs10934270 (allele: T vs C, OR = 0.50, \( p = 0.009 \); genotype: CT vs CC, OR = 0.43, \( p = 0.006 \); dominant: CT-TT vs CC, OR = 0.43, \( p = 0.004 \); and log-additive: OR = 0.47, \( p = 0.006 \)) was associated with a decreased risk of EC, whereas rs9841504 (allele: G vs C, OR = 1.78, \( p = 0.029 \); and log-additive: OR = 1.72, \( p = 0.047 \)) contributed to increased EC susceptibility. After Bonferroni correction, the significant association of rs10934270 with EC risk in smokers still existed.

When stratified by drinking, there was no correlation between ZBTB20 SNPs and EC risk in drinkers and non-drinkers (data no shown).

### FPRP Analysis

FPRP analysis was performed to calculate positive findings, as shown in Table 6. At a prior probability level of 0.1, the significant association of rs10934270 remained noteworthy overall (FPRP = 0.040, and 0.055, and statistical power >90%). The correlation of rs10934270 with EC susceptibility was also positive in subgroups including ESCC patients, males, subjects aged \( \leq 65 \) years, subjects with BMI \( \leq 24 \text{ kg/m}^2 \), and smokers (FPRP < 0.2). Moreover, the association of rs9288999 with EC risk in subjects aged \( \leq 65 \) years was still noteworthy (FPRP = 0.193). The low statistical power of subgroups may be related to the small sample size after stratification.

| Group/SNPs ID | Model                  | OR (95% CI)     |  \( p \)   | Statistical Power | Prior Probability |
|---------------|------------------------|-----------------|------------|-------------------|-------------------|
|               |                        |                 |            | 0.25              | 0.01              | 0.001             | 0.0001             | 0.00001                       |
| **Overall**   |                        |                 |            |                   |                   |                   |                   |                   |
| rs10934270    | Allele                 | 0.64 (0.47–0.87)| 0.004      | 0.942             | 0.014             | 0.040             | 0.315             | 0.823              | 0.979              |
|               | Genotype               | 0.68 (0.49–0.96)| 0.029      | 0.960             | 0.081             | 0.210             | 0.745             | 0.967              | 0.997              |
|               | Dominant               | 0.65 (0.46–0.91)| 0.011      | 0.906             | 0.091             | 0.232             | 0.768             | 0.971              | 0.997              |
|               | Log-additive           | 0.64 (0.47–0.88)| 0.005      | 0.936             | 0.019             | 0.055             | 0.389             | 0.865              | 0.985              |
| **Esophageal squamous cell carcinoma** | | | | | |
| rs10934270    | Allele                 | 0.66 (0.47–0.93)| 0.017      | 0.944             | 0.053             | 0.143             | 0.648             | 0.949              | 0.995              |
|               | Dominant               | 0.67 (0.47–0.96)| 0.031      | 0.945             | 0.085             | 0.217             | 0.753             | 0.969              | 0.997              |
|               | Log-additive           | 0.67 (0.48–0.94)| 0.020      | 0.955             | 0.060             | 0.162             | 0.679             | 0.955              | 0.995              |
|               | Genotype               | 2.59 (1.02–6.59)| 0.045      | 0.294             | 0.319             | 0.584             | 0.939             | 0.994              | 0.999              |
|               | Recessive              | 2.56 (1.01–6.47)| 0.047      | 0.301             | 0.319             | 0.584             | 0.939             | 0.994              | 0.999              |
| rs9841504     | Allele                 | 2.59 (1.02–6.59)| 0.045      | 0.294             | 0.319             | 0.584             | 0.939             | 0.994              | 0.999              |
|               | Dominant               | 2.56 (1.01–6.47)| 0.047      | 0.301             | 0.319             | 0.584             | 0.939             | 0.994              | 0.999              |
|               | Log-additive           | 2.56 (1.01–6.47)| 0.047      | 0.301             | 0.319             | 0.584             | 0.939             | 0.994              | 0.999              |
| **Males**     |                        |                 |            |                   |                   |                   |                   |                   |                   |
| rs10934270    | Allele                 | 0.64 (0.44–0.93)| 0.017      | 0.902             | 0.060             | 0.161             | 0.679             | 0.955              | 0.995              |
|               | Genotype               | 0.66 (0.44–0.99)| 0.042      | 0.910             | 0.128             | 0.306             | 0.829             | 0.980              | 0.998              |
|               | Dominant               | 0.64 (0.43–0.95)| 0.025      | 0.890             | 0.083             | 0.213             | 0.749             | 0.968              | 0.997              |
|               | Log-additive           | 0.65 (0.45–0.93)| 0.020      | 0.924             | 0.056             | 0.152             | 0.664             | 0.952              | 0.995              |

(Continued)
MDR Analysis

MDR analysis was used to detect the relationship between higher order interactions and EC risk (Table 7 and Figure 1). Rs10934270 was the most influential attribution factor for EC risk in the single-locus model (testing balanced accuracy of 0.5297, and cross-validation consistency of 10/10), which was consistent with the logistic analysis results. The combination of rs10934270, rs9288999 and rs73230612 (testing balanced accuracy of 0.5211, and cross-validation consistency of 9/10) was also significant.

### Table 6 (Continued)

| Group/SNPs ID       | Model       | OR (95% CI) | p       | Statistical Power | Prior Probability |
|---------------------|-------------|-------------|---------|-------------------|------------------|
|                     |             |             |         |                   | 0.25  | 0.1   | 0.01  | 0.001 | 0.0001 |
| **Females**         |             |             |         |                   |       |       |       |       |        |
| rs73230612          | Allele      | 0.70 (0.50–0.99) | 0.043   | 0.971             | 0.119  | 0.288 | 0.817  | 0.978  | 0.998   |
|                     | Genotype    | 0.42 (0.19–0.94) | 0.035   | 0.336             | 0.237  | 0.483 | 0.911  | 0.990  | 0.999   |
|                     | Log-additive| 0.67 (0.46–0.97) | 0.034   | 0.939             | 0.098  | 0.245 | 0.781  | 0.973  | 0.997   |
| **Age ≤ 65 years**  |             |             |         |                   |       |       |       |       |        |
| rs10934270          | Allele      | 0.40 (0.26–0.62) | 3.25×10⁻⁵ | 0.159             | 0.001  | 0.002 | 0.025  | 0.207  | 0.724   |
|                     | Genotype    | 0.40 (0.25–0.66) | 3.29×10⁻⁴ | 0.191             | 0.005  | 0.016 | 0.148  | 0.637  | 0.946   |
|                     | Dominant    | 0.38 (0.24–0.62) | 1.07×10⁻⁴ | 0.136             | 0.002  | 0.007 | 0.072  | 0.440  | 0.887   |
|                     | Log-additive| 0.41 (0.26–0.64) | 1.15×10⁻⁴ | 0.191             | 0.001  | 0.004 | 0.043  | 0.312  | 0.820   |
|                     | Genotype    | 0.56 (0.34–0.91) | 0.019   | 0.676             | 0.079  | 0.204 | 0.738  | 0.966  | 0.996   |
|                     | Recessive   | 0.61 (0.39–0.94) | 0.025   | 0.816             | 0.084  | 0.216 | 0.752  | 0.968  | 0.997   |
|                     | Log-additive| 0.77 (0.60–0.97) | 0.027   | 0.889             | 0.074  | 0.193 | 0.724  | 0.964  | 0.996   |
| **BMI ≤ 24 kg/m²**  |             |             |         |                   |       |       |       |       |        |
| rs10934270          | Allele      | 0.42 (0.28–0.63) | 2.11×10⁻⁵ | 0.200             | <0.001 | 0.001 | 0.013  | 0.121  | 0.579   |
|                     | Genotype    | 0.48 (0.30–0.77) | 0.002   | 0.433             | 0.016  | 0.046 | 0.348  | 0.844  | 0.982   |
|                     | Dominant    | 0.13 (0.03–0.71) | 0.018   | 0.060             | 0.481  | 0.735 | 0.968  | 0.997  | 1.000   |
|                     | Log-additive| 0.44 (0.28–0.69) | 3.64×10⁻⁴ | 0.289             | 0.004  | 0.011 | 0.107  | 0.546  | 0.923   |
|                     | Recessive   | 0.15 (0.03–0.81) | 0.027   | 0.081             | 0.505  | 0.753 | 0.971  | 0.997  | 1.000   |
|                     | Log-additive| 0.45 (0.30–0.68) | 1.35×10⁻⁴ | 0.308             | 0.001  | 0.004 | 0.046  | 0.327  | 0.829   |
| **Smokers**         |             |             |         |                   |       |       |       |       |        |
| rs10934270          | Allele      | 0.50 (0.29–0.85) | 0.009   | 0.500             | 0.059  | 0.158 | 0.674  | 0.954  | 0.995   |
|                     | Genotype    | 0.43 (0.24–0.79) | 0.006   | 0.313             | 0.059  | 0.158 | 0.674  | 0.954  | 0.995   |
|                     | Dominant    | 0.43 (0.24–0.77) | 0.004   | 0.306             | 0.042  | 0.117 | 0.594  | 0.937  | 0.993   |
|                     | Log-additive| 0.47 (0.28–0.81) | 0.006   | 0.412             | 0.046  | 0.125 | 0.612  | 0.941  | 0.994   |
|                     | Allele      | 1.78 (1.06–3.01) | 0.029   | 0.668             | 0.124  | 0.298 | 0.823  | 0.979  | 0.998   |
|                     | Log-additive| 1.72 (1.01–2.95) | 0.047   | 0.708             | 0.171  | 0.383 | 0.872  | 0.986  | 0.999   |

Notes: p values were calculated by logistic regression analysis with adjustments for age. Statistical power was calculated using the number of observations in the subgroup and the OR and p values in this table. Bold prior probability < 0.2 (false-positive report probability threshold) respects the data is statistically significant.

Abbreviations: SNP, single nucleotide polymorphism; OR, odds ratio; 95% CI, 95% confidence interval.

---

### Table 7 SNP–SNP Interaction Models of Candidate SNPs Analyzed by the MDR Method

| Model                     | Training Bal. Acc | Testing Bal. Acc | CVC | p       |
|---------------------------|-------------------|------------------|-----|---------|
| rs10934270                | 0.5287            | 0.5297           | 10/10 | 0.0119  |
| rs9288999, rs73230612     | 0.5412            | 0.4933           | 4/10 | 0.0187  |
| rs10934270, rs9288999, rs73230612 | 0.5629       | 0.5211           | 9/10 | <0.0001 |
| rs10934270, rs9288999, rs9841504, rs73230612 | 0.5810       | 0.5086           | 10/10 | <0.0001 |

Notes: p values were calculated using χ² tests. Bold indicate that p < 0.05 indicates statistical significance.

Abbreviations: MDR, multifactor dimensionality reduction; Bal. Acc., balanced accuracy; CVC, cross-validation consistency; OR, odds ratio; CI, confidence interval.
consistency of 9/10) was the best multi-locus model. The dendrogram (Figure 1A) presented that rs10934270 and rs9841504 exhibited strong redundancy effects on EC susceptibility. The Fruchterman-Reingold (Figure 1B) revealed that rs9288999 and rs73230612 had synergistic interaction with the positive information gain (0.21%) of EC.

Discussion

The ZBTB20 gene, located on chromosome 3q13.31, is reported to be involved in the proliferation, migration and invasion of cancer.18,19 It has been found that ZBTB20 expression is increased in EC by silico analyses. Previous studies have revealed that ZBTB20 polymorphisms are related to many diseases, such as cognitive aging,20 systemic lupus erythematosus,21 autism spectrum disorders,22 and gastric cancer.23 Here, a hospital-based study of 525 EC patients and 522 healthy controls explored the relationship between four SNPs (rs10934270, rs9288999, rs9841504, and rs73230612) in ZBTB20 and EC occurrence among the Chinese Han population. The results demonstrated for the first time that rs10934270-T was associated with lower EC susceptibility with statistical power >90% in overall analysis. However, there are no reports on rs10934270. Moreover, we also found that the rs9841504 GG genotype might be a risk-increasing factor for ESCC. Nevertheless, a previous study has shown no significant relationship between rs9841504-GG and ESCC risk;24 such inconsistencies in these studies might be due to different behaviors or sample sizes. Based on the GTEx Portal database, genotypes of rs10934270 were related to ZBTB20 mRNA expression in whole blood. These results suggest that rs10934270 may be involved in EC carcinogenic by affecting the expression or function of ZBTB20. This may be new biological findings in the development of EC; however, further experimental confirmation is still required.

It is well known that genetic, environmental, and behavioral risk factors may affect EC development.33 According to reports, the risk of EC increases with age, and the incidence of EC is higher in men than in women.34 In age stratification, the associations of rs10934270 T allele and rs9288999 GG genotype with decreased EC risk were observed in subjects aged ≤65 years, but not in subjects aged >65 years. According to the gender-stratified analysis, rs10934270-T was associated with reduced EC risk in males, and rs73230612-C was a protective factor against EC in females. These results indicated that the association between ZBTB20 polymorphisms and EC susceptibility appeared to be age- and gender-dependent. Obesity, cigarette-smoking, and alcohol-drinking are known risk factors for EC.35 Higher BMI levels also increase the risk of EC.36 Additionally, the risk-reducing association of rs10934270-T with EC occurrence was found in
the subgroup with BMI ≤ 24 kg/m². Smokers have a 2.21–3.73 fold increased risk of EC compared with non-smokers.37 In smokers, rs10934270-T was related to a decreased EC risk, whereas rs9841504-G contributed to increased EC susceptibility. These results need to be verified in a larger population. Alcohol consumption is associated with increased EC occurrence.38 When stratified by drinking, no association between ZBTB20 SNPs and EC risk was found. These findings suggested that gene-behavioral habit interactions might play a certain role in the carcinogenesis of EC. However, the results should be interpreted with caution due to the relatively small sample size in stratified analyses.

Furthermore, exploring intragenic SNP–SNP interactions can also help us discover potential risk factors for the onset of EC.39 The results of MDR showed that rs10934270 was the most influential attribution factor for EC risk in the single-locus model and the combination of rs10934270, rs9288999 and rs73230612 was the best multi-locus model.

Inevitably, this study has some limitations. First, the hospital-based research has selection bias, and all participants are Chinese Han population, so these findings may not be applicable to other populations. Second, only four SNPs were chosen to explore the effect of ZBTB20 variants on EC occurrence, and other loci in ZBTB20 were not investigated. The association between other SNPs in ZBTB20 and the risk of EC requires further evaluation in the future. Third, research on the functions of these SNPs and their association with the expression level of ZBTB20 should be conducted, which will further confirm the results of our study. The potential mechanisms and functions of these SNPs hidden behind the association need to be further explored in detailed experiments. Fourth, there was a limited sample size in the stratification analysis. Hence, a larger sample size is needed to verify our findings. Fifth, given that EC is a complex multifactorial disease that may be influenced by genetic and environmental factors, the role of environmental factors in the association of ZBTB20 variants with EC risk should be considered. In the future we will enlarge the cohort of subjects to explore the interaction between ZBTB20 variants and environmental factors on EC risk.

**Conclusion**

Taken together, our study is the first to report that ZBTB20 rs10934270-T is associated with lower EC susceptibility in the Chinese Han population. These data provide scientific evidence for understanding the influence of ZBTB20 on the occurrence of EC. However, it is still necessary to conduct functional studies to clarify the molecular mechanisms of EC behind these associations.

**Data Sharing Statement**

The datasets generated and/or analysed during the current study are available in the zenodo repository ([https://zenodo.org/record/6318712#.Yh22wnbE_J0](https://zenodo.org/record/6318712#.Yh22wnbE_J0)).

**Ethics Approval and Consent to Participate**

This study was conducted under the approval of the Ethics Committee of Hainan Cancer Hospital (No.: ZDKJ202005) according to the Helsinki Declaration. Written informed consent to participate in the study was obtained.

**Informed Consent**

All individuals provided written informed consent prior to sample collection.

**Acknowledgment**

The authors thank all participants and volunteers in this study. Shuyong Yu, Guihong Yuan, and Feixiang Hu are co-first authors to this study.

**Author Contributions**

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.
**Funding**
This work was supported by 2020 Hainan Province Major Science and Technology Plan Project (ZDKJ2020005) and Supported by the specific research fund of The Innovation Platform for Academicians of Hainan Province.

**Disclosure**
The authors declare that they have no conflicts of interest in this work.

**References**
1. Rustgi AK, El-Serag HB, Ingelfinger JR. Esophageal carcinoma. *N Engl J Med.* 2014;371(26):2499–2509. doi:10.1056/NEJMra1314530
2. Ferlay JEM, Lam F, Soerjomataram I, et al. Global cancer observatory: cancer today. *Int J Cancer.* 2021;149(4):778–789. doi:10.1002/ijc.33588
3. Yang J, Liu X, Cao S, Dong X, Rao S, Cai K. Understanding esophageal cancer: the challenges and opportunities for the next decade. *Front Oncol.* 2020;10:1727. doi:10.3389/fonc.2020.01727
4. Wang S, Pan D, Chen Z, et al. Trends in incidence and mortality of esophageal cancer in Huai’an District, a high-risk area in Northern Jiangsu Province, China. *Cancer Control.* 2022;29:10732748221076824. doi:10.1177/10732748221076824
5. Pan D, Su M, Zhang T, et al. A distinct epidemiologic pattern of precancerous lesions of esophageal squamous cell carcinoma in a high-risk area of Huai’an, Jiangsu Province, China, *Cancer Prev Res.* 2019;12(7):449–462. doi:10.1158/1940-6207.CAPR-18-0462
6. Tran GD, Sun XD, Abnet CC, et al. Prospective study of risk factors for esophageal and gastric cancers in the Linxian general population trial cohort in China. *Int J Cancer.* 2005;113(3):456–463. doi:10.1002/ijc.20616
7. Gao Y, Hu N, Han XY, et al. Risk factors for esophageal and gastric cancers in Shanxi Province, China: a case-control study. *Cancer Epidemiol.* 2011;35(6):e91–99. doi:10.1016/j.canep.2011.06.006
8. Reichenbach ZW, Murray MG, Saxena R, et al. Clinical and translational advances in esophageal squamous cell carcinoma. *Adv Cancer Res.* 2019;144:95–135.
9. Li S, Chen H, Man J, et al. Changing trends in the disease burden of esophageal cancer in China from 1990 to 2017 and its predicted level in 25 years. *Cancer Med.* 2021;10:1889–1899. doi:10.1002/cam4.3775
10. Lindkvist B, Johansen D, Stocks T, et al. Metabolic risk factors for esophageal squamous cell carcinoma and adenocarcinoma: a prospective study of 580,000 subjects within the Me-Can project. *Cancer Biomark.* 2014;14:103. doi:10.1186/1471-2407-14-103
11. Klimczak-Bitner AA, Bitner J, Hiruta K, Szemraj J. Exploring a possible association between the occurrence of the SERPINE1-675 4G/5G polymorphism and the increased risk of esophageal cancer in the Caucasian population. *Biochem Biophys Res. B. 2021;28:101147. doi:10.3389/fbrep.2021.101147
12. Yang PW, Lin MC, Huang PM, et al. Risk factors and genetic biomarkers of multiple primary cancers in esophageal cancer patients. *Front Oncol.* 2020;10:585621. doi:10.3389/fonc.2020.585621
13. Dighe SG, Chen J, Yan L, et al. Germline variation in the insulin-like growth factor pathway and risk of Barrett’s esophagus and esophageal adenocarcinoma. *Carcinogenesis.* 2021;42(3):369–377. doi:10.1093/carcin/bga132
14. Liu G, Zhou L, Zhang H, et al. Regulation of hepatic lipogenesis by the zinc finger protein Zbtb20. *Nat Commun.* 2017;8:14824. doi:10.1038/ ncomms14824
15. Zhao J, Xu X, Ellwein LB, et al. Prevalence of vision impairment in older adults in Rural China in 2014 and comparisons with the 2006 China nine-province survey. *Am J Ophthalmol.* 2018;185:81–93. doi:10.1016/j.ajo.2017.10.016
16. Chevrier S, Corcoran LM. BTB-ZF transcription factors, a growing family of regulators of early and late B-cell development. *Immunol Cell Biol.* 2014;92(6):481–488. doi:10.1038/icb.2014.20
17. He Z, Zhu J, Mo J, Zhao H, Chen Q. HBV DNA integrates into upregulated ZBTB20 in patients with hepatocellular carcinoma. *Mol Med Rep.* 2020;22(1):380–386. doi:10.3892/mmr.2020.11074
18. Zhang Y, Zhou X, Zhang M, Cheng L, Zhang Y, Wang X. ZBTB20 promotes cell migration and invasion of gastric cancer by inhibiting IκBα to induce NF-κB activation. *Biosci Rep.* 2020;40(9). doi:10.1042/BSR20200102
19. Kichaev G, Bhatia G, Loh PR, et al. Leveraging polygenic functional enrichment to improve GWAS power. *Cancer Med.* 2020;9(5):1889–1899. doi:10.1002/cam4.3775
20. He Z, Zhu J, Mo J, Zhao H, Chen Q. HBV DNA integrates into upregulated ZBTB20 in patients with hepatocellular carcinoma. *Mol Med Rep.* 2020;22(1):380–386. doi:10.3892/mmr.2020.11074
21. Bai F, Xiao K. Prediction of gastric cancer risk: association between ZBTB20 genetic variance and gastric cancer risk in Chinese Han population. *Biosci Rep.* 2020;40(9). doi:10.1042/BSR20200102
22. Kichaev G, Bhatia G, Loh PR, et al. Leveraging polygenic functional enrichment to improve GWAS power. *Am J Hum Genet.* 2019;104(1):65–75. doi:10.1016/j.ajhg.2018.11.008
23. Liu M, Jiang Y, Wedow R, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet.* 2019;51(2):237–244. doi:10.1038/s41435-018-0307-5
24. Yan C, Zhu M, Ding Y, et al. Meta-analysis of genome-wide association studies and functional assays decipher susceptibility genes for gastric cancer in Chinese populations. *Gut.* 2020;69(4):641–651. doi:10.1136/gutjnl-2019-318760
25. Imamura M, Takahashi A, Yamashita T, et al. Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes. *Nat Commun.* 2016;7:10531. doi:10.1038/ncomms10531
26. Thomas RK, Baker AC, Debiasi RM, et al. High-throughput oncogene mutation profiling in human cancer. *Nat Genet.* 2007;39(3):347–351. doi:10.1038/ng1975
27. Gabriel S, Ziaugra L, Tabbaa D. SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Curr Protocols Human Genet.* 2009; Chapter 2:Unit12.12.
28. Guo T, Hao H, Zhou L, Zhou F, Yu D. Association of SNPs in the TIMP-2 gene and large artery atherosclerotic stroke in southern Chinese Han population. *Oncotarget*. 2018;9(4):4698–4706. doi:10.18632/oncotarget.23473

29. Lin E, Kuo PH, Lin WY, Liu YL, Yang AC, Tsai SJ. An association study in the Taiwan Biobank elicits three novel candidates for cognitive aging in old adults: NCAM1, TTC12 and ZBTB20. *Aging*. 2021;13(14):18769–18788. doi:10.18632/aging.203321

30. Lee HS, Kim T, Bang SY, et al. Ethnic specificity of lupus-associated loci identified in a genome-wide association study in Korean women. *Ann Rheum Dis*. 2014;73(6):1240–1245. doi:10.1136/annrheumdis-2012-202675

31. Wiśniowiecka-Kowalnik B, Kastory-Bronowska M, Bartnik M, et al. Application of custom-designed oligonucleotide array CGH in 145 patients with autistic spectrum disorders. *Eur J Human Genet*. 2013;21(6):620–625. doi:10.1038/ejhg.2012.219

32. Dai N, Zheng M, Wang C, et al. Genetic variants at 8q24 are associated with risk of esophageal squamous cell carcinoma in a Chinese population. *Cancer Sci*. 2014;105(6):731–735. doi:10.1111/cas.12399

33. Musa IH, Musa TH, Musa HH, Ahmed ME. Esophageal cancer epidemiology, diagnosis, and management in Sudan - a review. *Med J Malaysia*. 2021;76(5):691–697.

34. Huang FL, Yu SJ. Esophageal cancer: risk factors, genetic association, and treatment. *Asian J Surg*. 2018;41(3):210–215. doi:10.1016/j.asjsur.2016.10.005

35. Short MW, Burgers KG, Fry VT. Esophageal cancer. *Am Fam Physician*. 2017;95(1):22–28.

36. Schlottmann F, Dreifuss NH, Patti MG. Obesity and esophageal cancer: GERD, Barrett’s esophagus, and molecular carcinogenic pathways. *Expert Rev Gastroenterol Hepatol*. 2020;14(6):425–433. doi:10.1080/17474124.2020.1764348

37. Oze I, Matsuoka K, Ito H, et al. Cigarette smoking and esophageal cancer risk: an evaluation based on a systematic review of epidemiologic evidence among the Japanese population. *Jpn J Clin Oncol*. 2012;42(1):63–73. doi:10.1093/jjco/hyr170

38. He F, Sha Y, Wang B. Relationship between alcohol consumption and the risks of liver cancer, esophageal cancer, and gastric cancer in China: metanalysis based on case-control studies. *Medicine*. 2021;100(33):e26982. doi:10.1097/MD.00000000000026982

39. Zhang H, Zhang Z, Zhang J, et al. Fine-mapping of ABO gene identifies two novel SNPs associated with large artery atherosclerotic stroke in a Chinese Han population. *Mol Neurobiol*. 2017;54(3):2107–2113. doi:10.1007/s12035-016-9794-5