Risk prediction of esophageal squamous-cell carcinoma with common genetic variants and lifestyle factors in Chinese population

Jiang Chang[1], Ying Huang[1], Lixuan Wei[1], Baoshan Ma[2], Xiaoping Miao[3], Yun Li[4], Zhibin Hu, Dianke Yu[1], Weihua Jia[5], Yu Liu[1], Wen Tan[1], Zhonghu He[6,7], Yang Ke[7], Tangchun Wu[3], Hongbing Shen[6], Yixin Zeng[5], Chen Wu[1] and Dongxin Lin[1,*]

[1]State Key Laboratory of Molecular Oncology, Cancer Institute and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China, [2]College of Information Science and Technology, Dalian Maritime University, Dalian, Liaoning, China, [3]Key Laboratory for Environment and Health (Ministry of Education), School of Public Health, Huazhong University of Sciences and Technology, Wuhan, Hubei, China, [4]Department of Genetics, University of North Carolina, Chapel Hill, NC, [5]State Key Laboratory of Oncology in Southern China, Sun Yat-Sen University Cancer Center, Guangzhou, Guangdong, China, [6]Department of Epidemiology and Biostatistics, Cancer Center, Nanjing Medical University, Nanjing, Jiangsu, China and [7]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University School of Oncology, Beijing Cancer Hospital and Institute, Beijing, China

*To whom correspondence should be addressed. Tel: +86-10-87788491; Fax: +86-10-67722460;
Email: lindx72@cicams.ac.cn
Correspondence may also be addressed to Chen Wu. Tel: +86-10-87788349; Fax: +86-10-67722460;
Email: cindywuchen@gmail.com

Genome-wide association studies have identified multiple genetic variants associated with risk of esophageal squamous-cell carcinoma (ESCC) in Chinese populations. We examined whether these genetic factors, along with non-genetic factors, can contribute to ESCC risk prediction. We examined 25 single nucleotide polymorphisms (SNPs) and 4 non-genetic factors (sex, age, smoking and drinking) associated with ESCC risk in 9805 cases and 10 493 controls from Chinese populations. Weighted genetic risk score (wGRS) was calculated and logistic regression was used to analyze the association between wGRS and ESCC risk. We calculated the area under the curve (AUC) using receiver operating characteristic curve analysis to measure the discrimination after adding genetic variants to the model with only non-genetic factors. Net reclassification improvement (NRI) was used to quantify the degree of correct reclassification using different models. wGRS of the combined 17 SNPs with significant marginal effect (G SNPs) increased ~4-fold ESCC risk ($P = 1.49 \times 10^{-164}$) and the associations were significant in both drinkers and non-drinkers. However, wGRS of the eight SNPs with significant effect in gene × drinking interaction (GE SNPs) increased ~4-fold ESCC risk only in drinkers ($P_{interaction} = 8.76 \times 10^{-41}$). The AUC for a risk model with 4 non-genetic factors, 17 G SNPs, 8 GE SNPs and their interactions with drinking was 70.1%, with the significant improvement of 7.0% compared with the model with only non-genetic factors ($P < 0.0001$). Our results indicate that incorporating genetic variants, lifestyle factors and their interactions in ESCC risk models can be useful for identifying patients with ESCC.

Introduction

In China, esophageal squamous-cell carcinoma (ESCC) ranks the fourth cause of cancer-related death and kills ~250 000 people per year. ESCC

Abbreviations: AIC, Akaike's information criterion; AUC, area under curve; CI, confidence interval; ESCC, esophageal squamous-cell carcinoma; GRS, genetic risk score; GWAS, genome-wide association studies; NRI, net reclassification improvement; ROC, receiver operating characteristic; SNPs, single nucleotide polymorphisms; wGRS, weighted genetic risk score.

is hard to detect in its early stage and most patients are in advanced disease stage when they are diagnosed and, thus, the long-term outcome of this malignancy is dismal, with 5-year survival rates being around 30% (1,2). Therefore, the establishment of risk prediction models capable of identifying individuals at high-risk for ESCC development may be of benefit for intervention, early detection and better clinical care of this disease if they are demonstrated to have clinical validity and utility.

Alcohol drinking is an established risk factor for ESCC development (3–5). In China, there has been a drinking culture for at least 7000 years and alcohol consumption has an important place in many cultural celebrations and even in daily life. It is estimated that there are currently >500 million Chinese who regularly consume alcohol and the overall heavy drinking rate of adults also increases rapidly (6–8). Convincing evidence based on international studies and studies in China has shown that high alcohol consumption is associated with an increased risk of developing ESCC (9–11).

However, although the lifestyle risk factors such as alcohol drinking mentioned above may play important roles in the etiology of ESCC, not all exposed individuals develop this disease, indicating that genetic factors, alone or in interaction with the environmental factors, may also be important in the development of the cancer. In the last decades, efforts have been made to identify such genetic risk factors and a few loci or allelic variants have been shown to be associated with susceptibility to ESCC (12–15). Recently, we have conducted genome-wide association studies (GWAS) of ESCC in Chinese population and identified 25 risk loci, among which 17 showing significant genetic effect and 8 showing significant gene–alcohol drinking interaction effect (16,17). These findings provide an unique opportunity for us to explore whether this panel of risk loci and their interactions with drinking can contribute to the prediction of ESCC risk.

Here, we report an evaluation of the contribution of the 25 risk loci and their interactions with drinking to the risk prediction of ESCC compared with the commonly used non-genetic factors in 9805 patients and 10 493 controls from the Chinese population.

Patients and methods

Subjects

This study included 10 123 ESCC patients and 10 664 controls. Detailed information on their recruitment and characteristics were described in our previous reports (16,17). Briefly, all patients were Han Chinese recruited in Beijing city and Jiangsu, Guangzhou, Henan and Hubei Provinces in China. Demographic characteristics including sex, age, smoking and drinking were obtained from patients' medical records. Controls were selected on the basis of physical examinations and frequency matched for age and sex to ESCC patients in different geographical regions (Supplementary Table 1, available at *Carcinogenesis* Online). At recruitment, informed consent was obtained from each subject and the study was approved by the Institutional Review Boards of the Chinese Academy of Medical Sciences Cancer Institute.

Genetic variants

We genotyped 26 SNPs associated with ESCC risk at genome-wide significance in our previously GWAS in 10 123 ESCC cases and 10 664 controls (16,17). Due to genotyping failure of some DNA samples, only 9805 cases and 10 493 controls with complete genotyping data were available for further analysis. To obtain a set of independent markers for this analysis, we removed rs11066280 on chromosome 12q24 because of its high linkage disequilibrium with rs11066015 ($r^2 > 0.8$), the strongest association marker at this genetic region and therefore 25 SNPs were remained for final analysis. We estimated linkage disequilibrium among the loci at the same chromosome using Haploview.

Statistical analysis

Individual locus analysis. Marginal effects between genotypes and ESCC risk were analyzed by additive model in a logistic regression with age, sex, smoking and drinking as covariates. For gene–drinking interaction analysis, we conducted

a one degree-of-freedom Wald test of a single interaction parameter (SNP × drinking) as implemented in an unconditional logistic regression based on the equation $Y = \beta_0 + \beta_1 \times \text{SNP} + \beta_2 \times \text{drinking} + \beta_3 \times (\text{SNP} \times \text{drinking})$ as described in ref. (17). Sex, age and smoking served as covariates in both gene × drinking interaction and stratified analyses. All the odds ratios were calculated for the minor allele of each variant. Significant SNPs with $P < 0.002$ (0.05/25) after Bonferroni correction were separated into G or GE group for further genetic risk score (GRS) computation. The 17 SNPs showing only significant marginal effects on ESCC risk were classified as the G group, whereas the 8 SNPs, among them 6 showed both significant marginal effects and gene–drinking interactions and 2 showed sole gene–drinking interactions, were classified as the GE group.

Genetic risk score computation. Two methods were used to create GRS: a simple count method (GRS) and a weighted method (wGRS). Both methods assumed each locus independently associated with ESCC risk. We assumed an additive genetic model for each locus, and applied a linear weight of 0, 1 and 2 to genotypes containing 0, 1 or 2 risk alleles, respectively. The simple GRS assumed that each locus in the panel contributes equally to ESCC risk and was calculated by summing the risk alleles of each locus. For wGRS, each locus was weighted by β-coefficients obtained from individual locus analysis shown in Supplementary Table 2, available at *Carcinogenesis* Online. We calculated GRS and wGRS for 17 loci in the G group and 8 loci in the GE group and named them as GRS17SNP, GRS8SNP, wGRS17SNP and wGRS8SNP, respectively. To simplify interpretation and facilitate comparison with the simple GRS for two groups, all wGRS values were divided by 8.26 or 6.32 (twice the sum of the β-coefficients) and multiplied by 17 or 8, respectively. The GRS and wGRS were also categorized into quintiles based on the distribution in controls and named as GRS17SNPQ5, GRS8SNPQ5, wGRS17S-NPQ5 and wGRS8SNPQ5, respectively. Locus–locus interaction term was not included in our analysis. The marginal and interaction effects for the association between GRS or wGRS and ESCC risk were calculated by unconditional logistic regression as used for individual locus analysis.

Discrimination ability and absolute risk estimation. To measure the discriminative improvement after adding genetic factors and their gene–drinking interactions to the non-genetic model, we plotted receiver operating characteristic (ROC) curves and calculated the corresponding areas under the curve (AUCs) obtained from a logistic regression model. The non-genetic model included sex, age, smoking and drinking. The joint effects of the 25 risk loci were incorporated in five different ways: (i) twenty-five individual loci in an additive main effect, (ii) GRS group included GRS17SNP and GRS8SNP, (iii) wGRS group included wGRS17SNP and wGRS8SNP, (iv) quintile categories of GRS group included GRS17SNPQ5 and GRS8SNPQ5 and (v) quintile categories of wGRS group included wGRS17SNPQ5 and wGRS8SNPQ5, respectively. We also investigated whether the interactions between drinking and the risk loci can improve the performance of the relevant models by including product terms of drinking multiplied by the eight individual loci that have interaction effects or their corresponding score group GRS8SNP, wGRS8SNP, GRS8SNPQ5 or wGRS8SNPQ5, respectively. We then compared the goodness-of-fit of various unconditional logistic regression models using Akaike's information criterion (AIC).

Comparison of performance of risk models. To compare the performance of different models, we calculated the cross-classification of the assignment of cases into predicted risk bands with similar risk. Cases and controls were randomly divided into 10 approximately equal subgroups and cross-validation was performed by setting aside one subgroup for the purpose of testing (testing set), while using the remaining 9 subgroups (training set) to construct a model to test in the testing subset. This procedure was repeated nine additional times with a different subgroup set aside for testing. Estimates of AUC were obtained for training and testing sets from each of the 10 cross-validation sets and a final estimated AUC was averaged over the 10 cross-validation sets. We evaluated the standard error of AUC and its significance by fitting each model 1000 times based on bootstrap re-sampling with replacement of cases and controls.
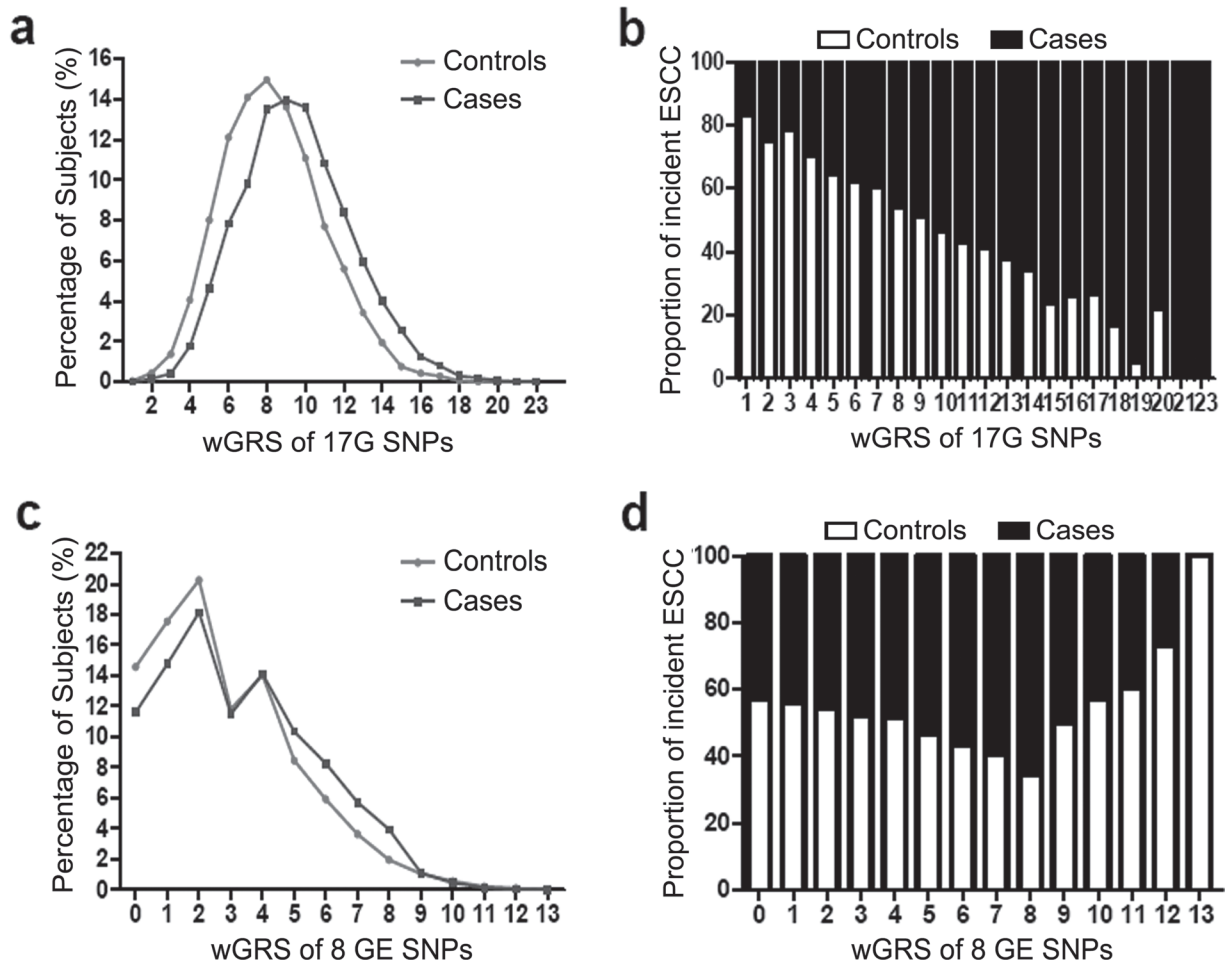

Fig. 1. Distributions of the wGRS among ESCC cases and controls. (**a**) percentage of wGRS of 17 SNPs displaying significant marginal effect (G SNPs) among cases and controls, (**b**) distributions of wGRS of 17 G SNPs among cases and controls, (**c**) percentage of wGRS of eight SNPs displaying the effect of significant gene × drinking interaction (GE SNPs) among cases and controls and (**d**) distributions of wGRS of eight GE SNPs among cases and controls.

Reclassification analysis. We used the category-less NRI to quantify the degree of correct reclassification when using the different models with genetic factors compared with the model without them. A confidence interval (CI) for NRI was constructed with the bootstrap method. All the statistical analyses were performed in R version 2.15.1.

Results

Cumulative association of 25 loci and risk of ESCC

Twenty-five loci previously identified as risk variants for ESCC are shown in Supplementary Table 2, available at *Carcinogenesis* Online. After correction for multiple comparisons, 17 loci showed significant marginal effects ($P_G = 1.96 \times 10^{-10}$ to 7.13×10^{-36}), 6 loci showed both significant marginal effects and interaction with drinking ($P_{GE} = 3.44 \times 10^{-6}$ to 1.02×10^{-46}) and 2 loci showed only significant interaction with drinking ($P_{GE} = 1.29 \times 10^{-7}$ to 2.25×10^{-10}).

To evaluate the joint effect of these variants, we calculated GRS and wGRS for the 17 G-loci or 8 GE-loci. The median of wGRS17SNP and wGRS8SNP was 9 and 3, respectively, and their distributions among cases or controls are shown in Figure 1. The incidence of ESCC increased significantly along with the increase in number of risk alleles (Figure 1a and c) and cases had more risk alleles than controls (Figure 1b and d) in both 17 G-loci group and 8 GE-loci group (all $P < 0.001$). Furthermore, we computed the associations of ESCC risk with quintiles of wGRS17SNP or wGRS8SNP and found increased odds ratios across quintiles of wGRS ($P_{trend} = 1.49 \times 10^{-164}$ and 9.38×10^{-50}). Compared with the lowest wGRS quintile, the highest wGRS quintile of wGRS17SNP had 3.74-fold increased odds ratio for ESCC (95% CI, 3.33–4.20), but risk was only 1.92-fold increase in wGRS8SNP (95% CI, 1.74–4.20). The analyses of stratification and interaction with drinking were performed to determine whether genetic factors varied across subgroups with different drinking status. wGRS17SNP, which includes 17 G-loci showing only significant marginal effects, was significantly associated with the risk in both non-drinkers and drinkers and did not significantly interact with drinking ($P_{interaction} = 0.1695$). However, wGRS8SNP, which contained 8 GE-loci showing gene–drinking interaction, had much stronger genetic effect among drinkers than among non-drinkers, with the odds ratio of 3.61 (95% CI, 3.11–4.19) and 1.17 (95% CI, 1.02–1.34) for drinkers and non-drinkers, respectively. The P value for interaction between wGRS8SNP and drinking was 8.76×10^{-41} (Table I). The risk estimates by using GRS17SNP and GRS8SNP were not much different (Table I).

Discriminative ability of genetic risk models

Five formats of genetic risk factors were incorporated into the non-genetic model including sex, age, smoking and drinking: individual locus (models 2–4), GRS (models 5–7), wGRS (models 8–10), quintiles of GRS (models 11–13) and quintiles of wGRS (models 14–16) (Supplementary Table 3, available at *Carcinogenesis* Online). We then calculated their goodness-of-fit and chose the best-fit models with (i) only genetic factors, (ii) both genetic and non-genetic factors and (iii) genetic factors, non-genetic factors and gene–drinking interactions. Using AIC value, we found that the model group including individual loci as genetic factors were the best and chosen for further predictor performance. The AUC of the models including non-genetic or genetic parameters alone was 0.639 (95% CI, 0.632–0.647) or 0.632 (95% CI, 0.625–0.640), respectively, and they were not significantly different. Compared with non-genetic model, the addition of 25 genetic risk locus produced a significant 5.8% increase in AUC value (95% CI, 5.3–6.4%) (Table II and Figure 2). Because 8 of the 25 loci used in this analysis showed significant interaction with drinking, we therefore added SNP × drinking interaction term in the model and the AUC for this model was 0.709 (95% CI, 0.70–0.716), which also significantly improved the performance of the genetic risk prediction model compared with the non-genetic model (AUC improvement = 7%; 95% CI, 6.4–7.6%). However, only 1.2% (95% CI, 0.9–1.4%) of AUC improvement was seen between the model with or without interaction terms. The corresponding cross-validated values for models including non-genetic and genetic parameters with or without gene–drinking interaction were 0.707 (95% CI, 0.682–0.731) and 0.695 (95% CI, 0.675–0.716), respectively, and the performance improvements of these two models after cross-validations were similar to those before cross-validations (Table II).

Table I. Associations of quintile groups of GRS with risk of ESCC

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	P_{trend}	$P_{interaction}$
wGRS17SNP (median = 9)	≤5	6, 7	8, 9	10, 11	≥12		
No. of cases/controls	686/1459	1731/2750	2686/3000	2396/1972	2306/1312		
Marginal	1.00 (reference)	1.33 (1.18–1.48)	1.88 (1.69–2.10)	2.58 (2.30–2.89)	3.74 (3.33–4.20)	1.49×10^{-164}	0.1695
Non-drinker	1.00 (reference)	1.43 (1.22–1.67)	1.91 (1.64–2.22)	2.59 (2.22–3.03)	3.73 (3.17–4.38)	9.64×10^{-81}	
Drinker	1.00 (reference)	1.24 (1.05–1.46)	1.91 (1.63–2.24)	2.61 (2.21–3.07)	3.85 (3.24–4.57)	1.43×10^{-84}	
GRS17SNP (median = 9)	≤5	6, 7	8, 9	10, 11	≥12		
No. of cases/controls	1048/2118	2066/2778	1295/1577	2455/2203	2941/1817		
Marginal	1.00 (reference)	1.49 (1.35–1.64)	1.65 (1.48–1.83)	2.25 (2.05–2.48)	3.26 (2.96–3.59)	1.78×10^{-147}	0.1297
Non-drinker	1.00 (reference)	1.57 (1.38–1.80)	1.60 (1.38–1.86)	2.24 (1.96–2.56)	3.19 (2.79–3.65)	8.34×10^{-72}	
Drinker	1.00 (reference)	1.41 (1.22–1.63)	1.72 (1.47–2.02)	2.30 (2.00–2.66)	3.39 (2.94–3.92)	8.08×10^{-77}	
wGRS8SNP (median = 3)	0	1	2	3, 4	≥5		
No. of cases/controls	1136/1527	1444/1838	1778/2123	2507/2710	2940/2295		
Marginal	1.00 (reference)	1.07 (0.96–1.19)	1.16 (1.05–1.29)	1.31 (1.19–1.45)	1.92 (1.74–2.12)	9.38×10^{-50}	8.76×10^{-41}
Non-drinker	1.00 (reference)	0.98 (0.84–1.14)	1.06 (0.92–1.22)	1.02 (0.89–1.16)	1.17 (1.02–1.34)	0.0122	
Drinker	1.00 (reference)	1.17 (1.00–1.37)	1.27 (1.09–1.47)	1.78 (1.54–2.06)	3.61 (3.11–4.19)	2.05×10^{-81}	
GRS8SNP (median = 3)	0	1	2	3, 4	≥5		
No. of cases/controls	1251/1672	1194/1437	2558/3092	1339/1442	3463/2850		
Marginal	1.00 (reference)	1.15 (1.03–1.28)	1.16 (1.05–1.27)	1.30 (1.17–1.45)	1.73 (1.58–1.90)	9.89×10^{-39}	1.35×10^{-50}
Non-drinker	1.00 (reference)	1.03 (0.89–1.20)	1.01 (0.89–1.15)	1.08 (0.93–1.26)	1.15 (1.01–1.31)	0.0095	
Drinker	1.00 (reference)	1.28 (1.09–1.50)	1.35 (1.18–1.56)	1.58 (1.35–1.86)	2.75 (2.41–3.16)	7.32×10^{-58}	

Table II. AUC values of various implemented models and their improvement values compared with the non-genetic model

Base model	Original AUC (95% CI)	AUC (95% CI) after 10-fold cross-validation	AUC improvement (95% CI)	
			Original	After 10-fold cross-validation
Non-genetic	0.639 (0.632–0.647)	0.639 (0.614–0.663)	Reference	Reference
Genetic	0.632 (0.625–0.640)	0.629 (0.610–0.649)	−0.007 (−0.018 to 0.004)	−0.009 (−0.044 to 0.025)
Genetic + non-genetic	0.698 (0.691–0.705)	0.695 (0.675–0.716)	0.058 (0.053–0.064)	0.057 (0.039–0.075)
Genetic + non-genetic + interaction	0.709 (0.702–0.716)	0.707 (0.682–0.731)	0.070 (0.064–0.076)	0.068 (0.042–0.094)

Non-genetic model includes sex, age, smoking and drinking as the variables; genetic model includes 17 G SNPs and 8 GE SNPs without interaction with drinking; genetic + non-genetic model includes sex, age, smoking, drinking as well as 17 G SNPs and 8 GE SNPs without interactions with drinking; genetic + non-genetic + interaction model includes sex, age, smoking, drinking as well as 17 G SNPs and 8 GE SNPs with interaction with drinking.
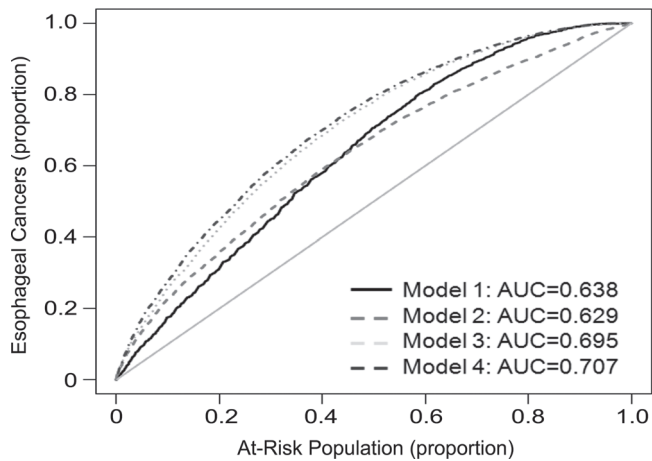

Fig. 2. ROC curves of four models for ESCC risk. The straight line indicates random classification. The area under the ROC curves (AUCs) are based on logistic regression models incorporating non-genetic risk factors (sex, age, smoking and drinking) only (model 1), genetic factors only (model 2), both non-genetic and genetic factors without their interactions (model 3) and non-genetic, genetic factors and their interaction with drinking (model 4).

Reclassification analysis

We used NRI to account for the correct movement of subjects in categories (higher risk subjects who developed ESCC and lower risk subjects who did not develop ESCC). The addition of genetic factors or their gene–drinking interactions to the non-genetic model produced significant improvements in net reclassification, with the NRI being 0.070 (95% CI, 0.061–0.079) and 0.086 (95% CI, 0.077–0.096), respectively (all $P < 0.001$; Table III). However, the NRI was 0.017 (95% CI, 0.010–0.024) between the models containing the non-genetic factors and genetic loci with or without gene–drinking interactions and this improvement was still significant.

Discussion

In this study, we confirmed 25 SNPs associated with ESCC risk. Among them, 17 variants showed unique marginal effect and 8 variants showed gene–drinking interaction manner. We demonstrated that GRS or wGRS, which combines information of multiple genetic loci, had substantial impact on ESCC risk and that the model with combination of both non-genetic and genetic factors performed significantly better than model with either non-genetic or genetic parameters alone in predicting ESCC risk.

We observed that combination of multiple loci had substantial influence on ESCC risk prediction, despite of the modest effect of each risk locus. In the 17 G-loci group, individuals with the highest quintile of GRS or wGRS had nearly 4-fold increased ESCC risk compared with those with the lowest quintile and no difference was seen between drinkers and non-drinkers. In eight GE-loci group, the highest quintile of GRS or wGRS was also associated with 4-fold increased risk in drinkers but not non-drinkers compared with the lowest quintile. The

interactions with drinking were only seen for GRS or wGRS of 8 loci but not for 25 loci. We assigned GRS and wGRS to the 17 loci in G group and 8 loci in GE group, respectively, so the effect on risk may be estimated more accurately considering the cumulative effect. However, the results using wGRS, which accounts for different magnitudes of effect of each locus, were similar to those using simple GRS, suggesting that the effect range of individual locus was narrow.

The first strength of this study is that different genetic parameters, including individual SNP, GRS, wGRS or their quintiles, were used to evaluate ESCC risk in prediction models and AIC values were used to compare the goodness-of-fit of these models. The genetic factor including individual locus was chosen due to its lowest AIC although the magnitude of the AUC improvement was similar to other genetic factors. Our data showed that the genetic-only model might not be superior to the non-genetic model, suggesting that the identification of novel or uncovered non-genetic risk factors for ESCC is needed for better prediction of this cancer in the future. However, combination of genetic factors with lifestyle should increase the ability of the models to predict ESCC. Indeed, the AUC for the model incorporating both genetic and non-genetic parameters increased by 5.78% compared with that for the model with non-genetic factors merely. This moderate improvement in the AUC with incorporating genetic loci identified in our previous ESCC GWAS seems to be better than those in prediction models reported for other diseases (18–21). This might reflect the fact that the AUC of non-genetic models for other diseases, for example, of type 2 diabetes (0.699 for the model including clinical predictors only), are relatively higher than that for ESCC and may be insensitive to the addition of any other factors. Another reason could be that the 25 SNPs identified in our population may provide more vigorous genetic susceptibility information, which contributes more improvement to risk prediction assessment.

The second strength is that we added the term of gene–drinking interaction in the risk prediction model because eight loci showed association with ESCC risk in a gene–drinking interaction fashion (17). Incorporating these loci and their interactions with drinking in the model improved prediction ability although the increment of AUC was only 1% compared with the model including simple combination of genetic and non-genetic parameters. This result is consistent with the suggestion that inclusion of gene–environment interactions is unlikely to dramatically improve risk prediction for complex disease (22).

The third strength is that the addition of DNA typing data improved not only the discriminatory power as assessed by ROC curves but also the reclassification of subjects into different risk strategies, with the use of NRI approach. NRI is a relatively new statistic and has gained increasing acceptance as an important part of new biomarkers evaluation and risk stratification (23,24). We used reclassification tables to calculate NRI and found that adding genetic risk loci to the model led to nearly 2% net gain to move the risk estimates toward the correct direction. Although this improvement is less than that for other diseases mentioned above, it remains possible that the addition of more genetic loci might prove their usefulness in risk discrimination.

Several limitations of this study deserve mention. Although we chose all ESCC-associated SNPs identified by GWAS published up to now (except for *C20orf54*, a variant reported only in one but not other studies (16,17,25,26)), the particular variants at each selected locus might just highly correlate to but not real causal

Table III. Comparison of models using NRI

Base model	Model comparison, NRI (95% CI)		
	Genetic only	Genetic + non-genetic	Genetic + non-genetic + interaction
Non-genetic only	−0.097 (−0.107 to −0.087)	0.070 (0.061–0.079)	0.086 (0.077–0.096)
Genetic only	–	0.157 (0.147–0.168)	0.174 (0.163–0.185)
Genetic + non-genetic	–	–	0.017 (0.010–0.024)
Genetic + non-genetic + interaction	–	–	–

Non-genetic model includes sex, age, smoking and drinking as the variables; genetic models includes 17 G SNPs and 8 GE SNPs without interaction with drinking; genetic + non-genetic model includes sex, age, smoking, drinking, as well as 17 G SNPs and 8 GE SNPs without interaction with drinking; genetic + non-genetic + interaction model includes sex, age, smoking, drinking, as well as 17 G SNPs and 8 GE SNPs with interaction with drinking.

variants. Because it is expected that additional GWAS, meta- or joint-analysis with other different studies and next-generation sequencing will discover additional risk variants including rare risk variants, risk prediction models as presented here require regular updates. Furthermore, our analysis was based on retrospective case–control studies. Although our sample size is large and recruited from multiple regions, whether our findings can be extended to general ESCC subjects remains to be determined and independent prospective cohort studies are needed.

In conclusion, our results indicate that incorporating genetic variants, lifestyle factors and their interactions in risk prediction model can be useful for identifying patients with ESCC. The model including genetic risk variants predicted ESCC risk better than the model including non-genetic risk factors merely. Although this effect might be too small to allow for individual risk prediction, it could be useful in reducing the number of subjects who would need to be included in the intervention studies aimed at the prevention against ESCC.

Supplementary material

Supplementary Tables 1–3 can be found at http://carcin.oxfordjournals.org/

References

1. Wang,L.S. *et al.* (1999) Prognosis of esophageal squamous cell carcinoma: analysis of clinicopathological and biological factors. *Am. J. Gastroenterol.*, **94**, 1933–1940.
2. Liu,J. *et al.* (2012) Which factors are associated with actual 5-year survival of oesophageal squamous cell carcinoma? *Eur. J. Cardiothorac. Surg.*, **41**, e7–11.
3. Gao,Y.T. *et al.* (1994) Risk factors for esophageal cancer in Shanghai, China. I. Role of cigarette smoking and alcohol drinking. *Int. J. Cancer*, **58**, 192–196.
4. Morita,M. *et al.* (2010) Alcohol drinking, cigarette smoking, and the development of squamous cell carcinoma of the esophagus: epidemiology, clinical findings, and prevention. *Int. J. Clin. Oncol.*, **15**, 126–134.
5. Toh,Y. *et al.* (2010) Alcohol drinking, cigarette smoking, and the development of squamous cell carcinoma of the esophagus: molecular mechanisms of carcinogenesis. *Int. J. Clin. Oncol.*, **15**, 135–144.
6. Newman,I. (2002) Cultural aspects of drinking patterns and alcohol controls in China. *The Globe (Global Alcohol Policy Alliance, London)*, **1**, 18–21.
7. Cochrane,J. *et al.* (2003) Alcohol use in China. *Alcohol Alcohol*, **38**, 537–542.
8. World Health Organization: Management of Substance Abuse Team. (2011) *Global Status Report on Alcohol and Health*. World Health Organization: Geneva, Switzerland.
9. Bagnardi,V. *et al.* (2001) A meta-analysis of alcohol drinking and cancer risk. *Br. J. Cancer*, **85**, 1700–1705.
10. Islami,F. *et al.* (2011) Alcohol drinking and esophageal squamous cell carcinoma with focus on light-drinkers and never-smokers: a systematic review and meta-analysis. *Int. J. Cancer*, **129**, 2473–2484.
11. Zhang,G.H. *et al.* (2010) Meta-analysis of ADH1B and ALDH2 polymorphisms and esophageal cancer risk in China. *World J. Gastroenterol.*, **16**, 6020–6025.
12. Cui,R. *et al.* (2009) Functional variants in ADH1B and ALDH2 coupled with alcohol and smoking synergistically enhance esophageal cancer risk. *Gastroenterology*, **137**, 1768–1775.
13. Sun,T. *et al.* (2007) A six-nucleotide insertion-deletion polymorphism in the CASP8 promoter is associated with susceptibility to multiple cancers. *Nat. Genet.*, **39**, 605–613.
14. Zhang,X. *et al.* (2005) Identification of functional genetic variants in cyclooxygenase-2 and their association with risk of esophageal cancer. *Gastroenterology*, **129**, 565–576.
15. Sun,T. *et al.* (2004) Polymorphisms of death pathway genes FAS and FASL in esophageal squamous-cell carcinoma. *J. Natl. Cancer Inst.*, **96**, 1030–1036.
16. Wu,C. *et al.* (2011) Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in Chinese populations. *Nat. Genet.*, **43**, 679–684.
17. Wu,C. *et al.* (2012) Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. *Nat. Genet.*, **44**, 1090–1097.
18. Wacholder,S. *et al.* (2010) Performance of common genetic variants in breast-cancer risk models. *N. Engl. J. Med.*, **362**, 986–993.
19. Cornelis,M.C. *et al.* (2009) Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. *Ann. Intern. Med.*, **150**, 541–550.
20. Janipalli,C.S. *et al.* (2012) Analysis of 32 common susceptibility genetic variants and their combined effect in predicting risk of Type 2 diabetes and related traits in Indians. *Diabet. Med.*, **29**, 121–127.
21. Lindström,S. *et al.* (2012) Common genetic variants in prostate cancer risk prediction–results from the NCI Breast and Prostate Cancer Cohort Consortium (BPC3). *Cancer Epidemiol. Biomarkers Prev.*, **21**, 437–444.
22. Aschard,H. *et al.* (2012) Inclusion of gene-gene and gene-environment interactions unlikely to dramatically improve risk prediction for complex diseases. *Am. J. Hum. Genet.*, **90**, 962–972.
23. Cook,N.R. *et al.* (2011) Performance of reclassification statistics in comparing risk prediction models. *Biom. J.*, **53**, 237–258.
24. Pencina,M.J. *et al.* (2011) Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat. Med.*, **30**, 11–21.
25. Wang,L.D. *et al.* (2010) Genome-wide association study of esophageal squamous cell carcinoma in Chinese subjects identifies susceptibility loci at PLCE1 and C20orf54. *Nat. Genet.*, **42**, 759–763.
26. Abnet,C.C. *et al.* (2010) A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. *Nat. Genet.*, **42**, 764–767.