



Original Article

Discrepancy between Genetically Predicted and Observed Alcohol Intake and Its Impact on Gastric Cancer Susceptibility

Ga-Eun Yie¹, Cheol Min Shin², Kyungtaek Park³, Jinyeon Jo⁴, Ah Ra Do¹, Sungkyoung Choi⁵, Jung Hun Ohn², Sejoon Lee⁶, Jeongseon Kim⁷, Sun Ha Jee⁸, Seung Joo Kang⁹, Nayoung Kim^{2,9}, Sungho Won^{1,4,10}

¹Interdisciplinary Program of Bioinformatics, Seoul National University, Seoul, ²Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, ³Institute of Health and Environment, Seoul National University, Seoul, ⁴Department of Public Health Sciences, Seoul National University, Seoul, ⁵Department of Mathematical Data Science, Hanyang University (ERICA), Ansan, ⁶Precision Medicine Center, Seoul National University Bundang Hospital, Seongnam, ⁷Center for Gastric Cancer, National Cancer Center Hospital, National Cancer Center, Goyang, ⁸Department of Epidemiology and Health Promotion, Institute for Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, ⁹Department of Internal Medicine and Liver Research Institute, Seoul National University College of Medicine, Seoul, ¹⁰RexSoft Inc., Seoul, Korea

Purpose We aimed to investigate how genetic predisposition to drinking and gastric cancer (GC) modifies the association between alcohol consumption and GC risk in the Korean population.

Materials and Methods Polygenic risk scores for GC (PRS-GC) and alcohol consumption (PRS-Alcohol) were formulated using genome-wide association results from BioBank Japan. Validation was performed using Korean cohorts (SNUBH-GENIE cohort), incorporating 8,846 controls and 531 patients with GC. Subsequently, these PRSs were applied to an independent Korean cohort of 67,771 participants, including 313 patients with GC during the follow-up for 14 years (KoGES cohort).

Results In KoGES cohort, the influence of alcohol consumption on GC risk was significantly altered by the PRS-GC and exhibited a synergistic interaction effect. PRS-Alcohol itself shows a negative correlation with GC risk. However, when actual alcohol consumption significantly exceeded genetically predicted levels, the risk of alcohol-related GC was notably increased (adjusted hazard ratio, 1.32; 95% confidence interval, 1.01 to 1.72). Heavy drinkers in the high-PRS-GC/low-PRS-Alcohol group had a 2.16 times higher risk of GC than non-to-light drinkers, which was prominent in males.

Conclusion Korean drinkers with higher PRS-GC who consume alcohol more than genetically predicted levels are susceptible to GC. PRS-GC and PRS-Alcohol may be beneficial for assessing the impact of alcohol consumption on GC risk in Koreans.

Key words Stomach neoplasms, Alcohol drinking, Genetic risk score, Sex, Lauren classification

Introduction

Gastric cancer (GC) is the sixth most diagnosed cancer and the third leading cause of cancer-related mortality worldwide [1]. Besides *Helicobacter pylori*, a well-known risk factor of GC, environmental and lifestyle factors have been studied for their contribution to the risk of GC, often in tandem with genetic factors. Alcohol consumption has been identified as a potential risk factor for GC in several meta-analyses [2]. However, heterogeneity in the findings across studies [3] makes it challenging to establish a clear and conclusive relationship.

Previous studies have demonstrated that single nucleotide polymorphisms (SNPs) can modify the effect of alcohol consumption on the risk of GC. In a Chinese population, SNPs such as rs2976392 in the *PSCA* gene and rs13042395 in the

SLC52A3 gene, both known GC susceptibility loci, showed significant interactions with alcohol consumption to GC risk [4,5]. SNPs in the *CUX2* region on chromosome 12q24.11-12 were identified as GC-associated variants in a large Japanese cohort [6] and were also linked to alcohol consumption in Korean men [7], suggesting a genetic overlap. Genes involved in alcohol metabolism also play crucial roles in alcohol consumption and modifying GC risk. A meta-analysis found that individuals with inactive *ALDH2* were more susceptible to the effects of alcohol on GC than those with the active form of the enzyme [8]. Similarly, polymorphisms in *ADH1C*, another alcohol-metabolizing gene, exhibited significant interactions with drinking in relation to GC risk [9].

In this context, the impact of alcohol consumption on GC risk may not depend solely on the absolute amount of drinking but rather on an individual's genetically determined

Correspondence: Sungho Won
Department of Public Health Science, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
Tel: 82-2-880-2714 E-mail: won1@snu.ac.kr

Co-correspondence: Nayoung Kim
Department of Internal Medicine, Seoul National University Bundang Hospital, 82 Gumi-ro, 173 beon-gil, Bundang-gu, Seongnam 13620, Korea
Tel: 82-31-787-7008 E-mail: nakim49@snu.ac.kr

Received January 24, 2025 Accepted May 29, 2025
Published Online May 30, 2025

*Ga-Eun Yie and Cheol Min Shin contributed equally to this work.

alcohol consumption, which can reflect their predisposition to alcohol metabolism. Individuals whose observed alcohol intake exceeds their genetically predicted consumption may be at particularly high risk, as their body may be less adapted to processing excessive alcohol. A similar concept has been observed in studies on body mass index (BMI) and type 2 diabetes (T2D), where individuals with a higher BMI than their genetically predicted value had an increased risk of T2D, regardless of their actual BMI [10]. This suggests that the discrepancy between genetically predicted and observed alcohol consumption could serve as a key factor in modifying GC risk.

To quantify genetically determined alcohol consumption and GC susceptibility, we utilized polygenic risk scores (PRSs) [11]. PRSs, which capture the cumulative effects of multiple genetic variants, each with a small effect size, are well-suited for studying complex traits such as alcohol consumption and non-hereditary GC. Using these scores, we examined whether discrepancies between genetically predicted and observed alcohol consumption, as well as genetic risk for GC, modify the association between alcohol intake and GC risk.

Therefore, in this study, we aimed to investigate how genetic predisposition, measured through PRS, modifies the association between alcohol intake and GC risk in the Korean population. Specifically, we examined whether discrepancies between PRS for alcohol consumption (PRS-Alcohol) and observed alcohol consumption, as well as the interactions of alcohol consumption with PRS for GC (PRS-GC) and PRS-Alcohol, have a significant effect on GC risk.

Materials and Methods

1. Study populations

Three independent datasets were used to generate the PRS: discovery, validation, and test sets (S1 Fig.). The discovery set required genome-wide association study (GWAS) summary statistics for GC and drinks per week, which were downloaded from the BioBank Japan (BBJ) (<https://pheweb.jp/>) [12,13]. The validation set included the Seoul National University Bundang Hospital (SNUBH) [14] and the Gene-Environment Interaction and phenotype (GENIE) cohorts [15]. The test set was derived from the Korean Genome and Epidemiology Study (KoGES) [16]. We replicated our results using data from the Korean Cancer Prevention Study (KCPS-II) [17]. No individuals overlapped between the Korean cohorts. Details of the study population, genetic data preprocessing [18,19], and GC identification are presented in Supplementary Methods (“Study population,” “Genotyping, quality control and imputation,” and “Gastric cancer

identification”). Following the quality control of genotype data and application of exclusion criteria, 8,846 participants in the validation set (SNUBH-GENIE) and 67,771 in the test set (KoGES) were eligible for the study (S2 Fig.).

2. Assessment of alcohol consumption

Participants responded to a study-specific questionnaire regarding their alcohol consumption habits. This included questions about drinking status and quantity and frequency of alcohol consumption among current drinkers. We calculated alcohol consumption (g/day) as described in “Assessment of alcohol consumption” section of the Supplementary Methods. The participants were also classified into two categories: non-to-light drinkers (< 10 g/day for female and < 20 g/day for male participants), or heavy drinkers (\geq 10 g/day for female and \geq 20 g/day for male participants).

3. PRS evaluation

Detailed explanations of the PRS development and evaluations are provided in “Polygenic risk score calculation and evaluation” section of the Supplementary Methods. The construction of the PRS required summary statistics from the GWAS and those for GC and alcohol consumption were downloaded from the BBJ (<https://pheweb.jp/>). We developed PRS-GC and PRS-Alcohol using data from GWAS of GC, conducted by Ishigaki et al. in 2020 [12] and the GWAS of alcohol consumption (drinks per week), conducted by Matoba et al. in 2020 [13]. We initially estimated the genetic correlation between summary statistics from the BBJ and KoGES [20] to ascertain the appropriateness of employing a Japanese GWAS for PRS calculations in the Korean population. Additionally, we estimated the genetic correlation between GC and alcohol consumption using the linkage disequilibrium score regression method [21]. Four different methods, clumping and thresholding (CT) [22], lasso [23], LDpred [24], and PRS-continuous shrinkage (PRS-CS) [25], were used to develop the PRS-GC and PRS-Alcohol, following the protocols suggested by Choi et al. [11]. The best PRSs were selected based on the 5-fold cross-validated area under the curve (AUC) for PRS-GC, and r^2 and Pearson’s correlation coefficients for PRS-Alcohol using the validation set. SNPs used to calculate the best PRSs were mapped to the genome using the Variant Effect Prediction tool [26]. Finally, PRSs for the test set were computed using the weights obtained from the best PRSs in the validation set.

4. Statistical analysis

Details are provided in “Statistical analysis” section of the Supplementary Methods. In the validation set, we performed multinomial logistic regression to investigate the association between GC types and PRSs, after adjusting for

Table 1. Characteristics of the study participants

Variable	Validation set (SNUBH-GENIE)			Test set (KoGES)		
	GC case (n=531)	Control (n=8,315)	p-value	Progressor to GC (n=313)	Non-progressor (n=67,458)	p-value
Male sex	371 (69.9)	5,064 (60.9)	< 0.001	205 (65.5)	24,436 (36.2)	< 0.001
Age (yr)	59.9±11.4	48.8±9.9	< 0.001	57.6±8.0	54.0±8.3	< 0.001
BMI (kg/m²)	23.9±3.2	23.3±3.1	< 0.001	24.2±2.8	24.0±2.9	0.426
Salted food intake (g/day)	NA	NA	NA	175.4±148.3	150.0±119.4	0.003
Salt preference						
Non-salty	115 (22.7)	3,252 (41.1)	< 0.001	NA	NA	NA
Mild	225 (44.5)	3,784 (47.8)		NA	NA	
Salty	166 (32.8)	884 (11.2)		NA	NA	
Smoking status						
Never	190 (36.3)	3,946 (51.6)	< 0.001	142 (45.4)	48,475 (71.9)	< 0.001
Past	223 (42.6)	2,283 (29.8)		100 (31.9)	10,647 (15.8)	
Current	111 (21.2)	1,420 (18.6)		71 (22.7)	8,306 (12.3)	
Alcohol consumption (g/day)	10.5±14.2	8.9±12.5	0.012	13.0±25.4	7.2±19.0	< 0.001
Family history of GC	117 (22.0)	1,081 (13.0)	< 0.001	41 (13.1)	6,596 (9.8)	0.061
<i>Helicobacter pylori</i>-positive	515 (97.0)	3,063 (36.8)	< 0.001	NA	NA	NA
Lauren classification						
Intestinal	301 (58.1)	NA	NA	NA	NA	NA
Diffuse	153 (29.5)	NA		NA	NA	
NA	64 (12.4)	NA		NA	NA	

Values are presented as number (%) or mean±SD. $p < 0.05$ indicates statistical significance. BMI, body mass index; GC, gastric cancer; GENIE, Gene-Environment Interaction and phenotype; KoGES, Korean Genome and Epidemiology Study; NA, not applicable; SD, standard deviation; SNUBH, Seoul National University Bundang Hospital.

age, sex, BMI, smoking status, salt preference, family history of GC, alcohol consumption and *H. pylori* infection.

In the test set, we examined the association of PRS-GC and PRS-Alcohol with each phenotype. Logistic and linear regression models, adjusting for age, sex, and first four principal components, were used for GC and alcohol consumption, respectively. We compared the performances of models with different covariate sets: model 1 (PRS-GC and PRS-Alcohol), model 2 (conventional risk factors), model 3 (model 2, PRS-GC, and PRS-Alcohol), and model 4 (model 3, interaction terms of alcohol with PRS-GC and PRS-Alcohol). We performed a mediation analysis, assuming that alcohol consumption mediates the effect of PRS-Alcohol on GC with an interaction effect involving PRS-Alcohol and PRS-GC.

Survival analysis of the association between the incidence of GC and drinking status was performed using the Cox proportional hazards model. Subgroup analyses were conducted according to PRS-GC and PRS-Alcohol levels. We defined discrepancy between alcohol consumption and genetically predicted intake as the difference between scaled alcohol consumption and scaled PRS-Alcohol. This discrepancy was categorized into three groups: (1) less than 0 (individuals who consumed less alcohol than genetically predicted),

(2) zero to the median value of discrepancies greater than 0 (0.78), and (3) greater than median value. We then tested association between these discrepancy groups and GC incidence. All statistical analyses were performed using R v3.6.3 (R Foundation for Statistical Computing) and Rex (v3.6.1), and statistical significance was set at a p-value of 0.05.

Results

1. Characteristics of the study participants

Table 1 presents the characteristics of the study participants in the validation (531 GC cases and 8,315 controls) and test sets (313 GC cases and 67,458 controls). Cases in both sets were older, had a higher proportion of males and ever-smokers, and consumed more salty food and alcohol than the controls ($p < 0.05$). Additionally, a higher proportion of cases tested positive for *H. pylori* in the validation set ($p < 0.001$).

2. SNP heritability and genetic correlations

The SNP heritability estimates were 0.02 (standard error [SE], 0.005; $p=0.001$) for GC, and 0.07 (SE, 0.04; $p=0.081$) for

alcohol consumption, using BBJ GWAS summary statistics. The genetic correlation between summary statistics from the Japanese and Korean populations for GC was 0.84 ($p=0.017$), and for alcohol consumption, it was 0.82 ($p=0.004$).

3. PRS development and evaluation in the validation set

In the SNUBH-GENIE cohort, the lassosum method exhibited the best performance for PRS-GC and PRS-Alcohol (S3 and S4 Tables). For PRS-GC, SNPs in *MUC1*, *THBS3*, *MTX1*, *GBAP1*, *RPL3P2*, *PSCA*, *JRK* and *ABO* had high weights (S5A and S5B Fig.). In the case of PRS-Alcohol, the SNPs on chromosomes 4 and 12 had higher weights including those in the *ADH* family, *ALDH2*, *MAPKAPK5*, *TMEM116*, *CCD63*, *MYL2*, *CUX2*, *RPH3A*, and *OAS2* (S5C and S5D Fig.). The SNPs involved in the development of PRS-Alcohol included variants associated with *ALDH2* expression levels according to GTEx [27] (<https://gtexportal.org/home/>) (S6 Table).

PRS-GC showed a weak correlation with PRS-Alcohol for the overall population (Pearson's correlation coefficient $r=-0.03$, $p=0.003$) (S5 Fig.), while the correlation became higher when evaluated in GC cases ($r=-0.11$, $p=0.009$), but these correlations were not significant in the test set (all samples: $r=-0.004$, $p=0.237$; GC cases: $r=-0.001$, $p=0.984$).

In the validation set, PRS-GC showed a significant association with GC even after adjusting for *H. pylori* infection and other factors (odds ratio [OR], 1.40; 95% confidence interval [CI], 1.26 to 1.57; $p=2.54\times 10^{-9}$) (S7 Table), while PRS-Alcohol was not significantly associated with GC. These results were consistent across both the intestinal and diffuse types of GC. Alcohol consumption was significantly associated with GC of the intestinal type only (OR, 1.15; 95% CI, 1.00 to 1.31; $p=0.046$).

The integration of PRSs into the risk factor model significantly increased the AUC for diffuse-type GC prediction ($\Delta\text{AUC}=0.01$, DeLong's $p=0.012$) (S8 Fig.), but it did not have a significant impact on intestinal-type GC prediction.

4. Association between PRS-GC and GC in the test set

In the KoGES cohort, we observed a significant association between PRS-GC and GC (OR for PRS-GC as a continuous variable, 1.34; 95% CI, 1.24 to 1.45; $p=9.8\times 10^{-14}$) (S9A Fig.). Individuals in the highest quartile (Q4) of PRS-GC exhibited an OR for GC of 2.09 (95% CI, 1.68 to 2.62; $p=5.9\times 10^{-11}$), compared to those in the lowest quartile (Q1). Males in Q4 of PRS-GC exhibited an OR for GC of 2.46 (95% CI, 1.84 to 3.32; $p=2.6\times 10^{-9}$) when compared to those in Q1, while the corresponding OR for females was 1.70 (95% CI, 1.23 to 2.38; $p=1.6\times 10^{-3}$). To create binary variables for PRS-GC in further analyses, we set cutoffs based on observed trends in the data. We divided groups at the 70th percentile, as the OR for GC showed a sharp increase from Q3 to Q4 of PRS-GC (S9A Fig.).

5. Association between PRS-Alcohol and alcohol consumption in the test set

A significant association was observed between PRS-Alcohol and alcohol consumption (β for continuous PRS-Alcohol, 2.09; 95% CI, 1.97 to 2.22; $p=6.0\times 10^{-232}$) (S9B Fig.). Those in Q4 of PRS-Alcohol demonstrated a 5.35 g/day increase in alcohol consumption compared to those in Q1 (95% CI, 5.00 to 5.71) ($p=3.0\times 10^{-189}$). Males in Q4 of PRS-Alcohol displayed a 12.42 g/day increase compared to those in Q1 (95% CI, 11.48 to 13.36; $p=3.2\times 10^{-162}$), while females exhibited a more moderate increase of 1.44 g/day (95% CI, 1.27 to 1.61; $p=3.7\times 10^{-62}$). We dichotomized PRS-Alcohol at the 50th percentile, given the notable increase in alcohol consumption from Q1 to Q2 of PRS-Alcohol (S9B Fig.).

6. Interaction between PRS-GC/PRS-Alcohol and alcohol consumption on GC risk

We compared models containing different sets of risk factors and PRSs (Table 2). In model 1, PRS-GC was significantly associated with GC. In model 2, age, sex, smoking status, salted food intake and family history of GC was significantly associated with GC, while BMI and alcohol consumption were not. Model 4, which incorporated the conventional risk factors, the PRSs, and their interaction terms with alcohol consumption, showed the best model fit, with the lowest Akaike information criterion value (3,738.0) of the four models (Table 2). In model 4 (Table 2, S10 Fig.), after adjusting for conventional risk factors, the estimated β value for the interaction term between alcohol consumption (as a continuous variable) and PRS-GC was 0.09 ($p=0.008$), whereas for PRS-Alcohol, it was -0.14 ($p=0.003$). Similar results were observed when evaluating the interaction between drinking status (non-to-light and heavy drinker) and PRS-GC ($\beta=0.30$, $p=0.025$) and PRS-Alcohol ($\beta=-0.31$, $p=0.049$) (S10 Fig.). In the sex-stratified analysis (S11 Table), the direction of the interaction terms remained consistent, however statistical significance was observed only in males.

7. Mediation analysis

The mediating effect of alcohol consumption between PRS-Alcohol and PRS-GC is shown in S12A Fig. When categorized as high or low PRS-GC, a positive indirect effect was observed, with a pure indirect effect estimate of 0.12 in the high PRS-GC group ($p=0.004$) (S12B Fig.). This implies a significant mediating role of alcohol consumption, primarily when PRS-Alcohol was controlled below the 50th percentile. This group showed a significant total direct effect ($\beta=-0.29$, $p=0.040$). In the low PRS-GC group, a significant total effect of -0.26 was noted ($p=0.027$) (S12C Fig.).

Table 2 Coefficients of covariates in the four gastric cancer prediction models in the test set

Variable	Model 1		Model 2		Model 3		Model 4	
	β (95% CI)	p-value	β (95% CI)	p-value	β (95% CI)	p-value	β (95% CI)	p-value
Age (yr)	-	-	0.046 (0.032 to 0.060)	3.7×10^{-11}	0.046 (0.033 to 0.060)	3.0×10^{-11}	0.046 (0.033 to 0.060)	3.0×10^{-11}
Sex (female vs. male)	-	-	-0.715 (-1.056 to -0.366)	4.8×10^{-5}	-0.714 (-1.055 to -0.364)	5.0×10^{-5}	-0.725 (-1.067 to -0.375)	4.0×10^{-5}
BMI (kg/m ²)	-	-	-0.015 (-0.056 to 0.025)	0.457	-0.014 (-0.055 to 0.026)	0.485	-0.014 (-0.055 to 0.026)	0.482
Smoking								
Past (vs. never)	-	-	0.455 (0.106 to 0.814)	0.012	0.451 (0.102 to 0.811)	0.013	0.451 (0.101 to 0.811)	0.013
Current (vs. never)	-	-	0.546 (0.174 to 0.920)	0.004	0.545 (0.173 to 0.902)	0.004	0.544 (0.171 to 0.920)	0.004
Salted food intake (per 20 g/day)	-	-	0.021 (0.005 to 0.036)	0.008	0.022 (0.005 to 0.037)	0.007	0.022 (0.005 to 0.037)	0.007
Family history of GC	-	-	0.366 (0.023 to 0.685)	0.030	0.342 (-0.002 to 0.661)	0.043	0.342 (-0.002 to 0.661)	0.043
Alcohol consumption (per 20 g/day)	-	-	0.041 (-0.043 to 0.088)	0.188	0.045 (-0.033 to 0.089)	0.113	0.017 (-0.090 to 0.075)	0.664
PRS								
PRS-GC	0.289 (0.213 to 0.365)	8.2×10^{-14}	-	-	0.277 (0.165 to 0.390)	1.4×10^{-6}	0.224 (0.105 to 0.344)	2.3×10^{-4}
PRS-Alcohol	-0.052 (-0.127 to 0.023)	0.173	-	-	-0.103 (-0.214 to 0.009)	0.069	-0.04 (-0.159 to 0.082)	0.521
Interaction terms								
Alcohol \times PRS-GC	-	-	-	-	-	-	0.088 (0.024 to 0.161)	0.008
Alcohol \times PRS-Alcohol	-	-	-	-	-	-	-0.142 (-0.235 to -0.047)	0.003
Model AIC	7,566.9		3,766.0		3,743.3		3,738.0	

Model 1 included PRS-GC and PRS-Alcohol. Model 2 included conventional risk factors (age, sex, BMI, smoking, salted food intake, family history of GC and alcohol consumption). Model 3 included variables in the model 2, PRS-GC and PRS-Alcohol. Model 4 included variables in the model 3 and interaction term of alcohol consumption with PRS-GC and PRS-Alcohol. $p < 0.05$ indicates statistical significance. AIC, Akaike information criterion; BMI, body mass index; CI, confidence interval; GC, gastric cancer; PRS, polygenic risk score.

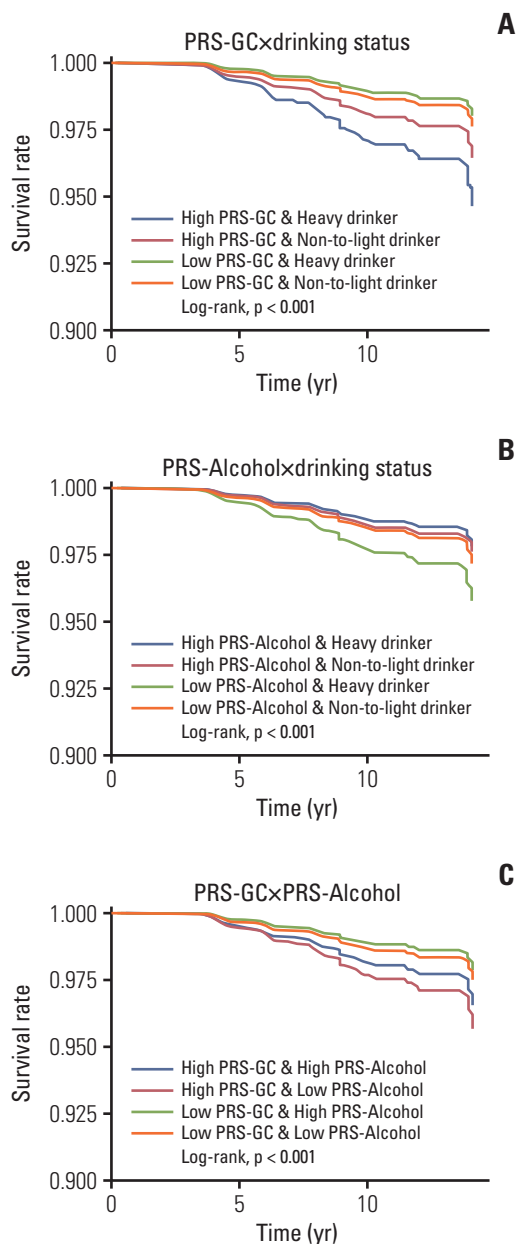


Fig. 1. Survival curves for gastric cancer (GC) incidence according to polygenic risk score for GC (PRS-GC), PRS for alcohol consumption (PRS-Alcohol), and drinking status. Kaplan-Meier curves are presented for subgroups of PRS-GC/drinking status (A), PRS-Alcohol/drinking status (B), and PRS-GC/PRS-Alcohol (C). PRS-GC and PRS-Alcohol were categorized into low and high groups based on their respective percentiles (70th for PRS-GC and 50th for PRS-Alcohol).

8. Survival analysis

Fig. 1 shows survival curves for different subgroups based on combinations of three binary factors: PRS-GC (high/low), PRS-Alcohol (high/low), and drinking status (heavy/non-to-

light drinker). Across these combinations, the high PRS-GC & heavy drinker group, the low PRS-Alcohol & heavy drinker group, and the high PRS-GC & low PRS-Alcohol group show decreased survival over time compared with other subgroups. Additionally, significant interactions were observed between alcohol consumption and PRS-GC ($p=0.025$), PRS-Alcohol ($p=0.035$) and their combination ($p=0.023$) on the incidence of GC (Fig. 2). That is, in the high PRS-GC group, heavy drinkers had a significantly higher risk of developing GC than non-to-light drinkers (HR, 1.55; 95% CI, 1.02 to 2.36; $p=0.038$). In contrast, no increased risk was observed in the low PRS-GC group. Furthermore, heavy drinkers in the high PRS-GC/low PRS-Alcohol group showed a 2.16-fold increased risk compared with non-to-light drinkers (95% CI, 1.25 to 3.72; $p=0.006$). Subgroup analysis was replicated using KCPS-II (S13 Table). When actual alcohol consumption significantly exceeded genetically predicted levels (discrepancy [actual alcohol intake–genetically predicted intake] > median value), the risk of alcohol-related GC was notably increased (HR, 1.32; 95% CI, 1.01 to 1.72; $p=0.041$) (Fig. 3).

Stratified by sex, a significant effect was found in males with high PRS-GC/low PRS-Alcohol (HR, 2.29; 95% CI, 1.28 to 4.12; $p=0.005$) (S14A Fig.) but not in females (HR, 4.29; 95% CI, 0.93 to 19.89; $p=0.063$) (S14B Fig.). The consistent direction of interaction was observed across the various PRS cut-off values (S15 Table).

Discussion

Alcohol consumption has been shown to increase the risk of GC; however, the results across studies have been heterogeneous [2], and their relationship has not been clarified. In this study, we evaluated the effects of alcohol consumption, PRS, and their interaction on GC risk in the Korean population.

In our analyses, we first found there was a significant negative interaction (β for interaction= -0.14) between alcohol consumption and PRS-Alcohol for GC risk; for non-drinkers, the OR for one standard deviation (SD) increase in PRS-Alcohol was 0.96, whereas for drinkers (consuming 20 g/day of ethanol), the OR was 0.83. These findings suggest that alcohol consumption plays a critical role in gastric carcinogenesis, with some effect modification. Consequently, the interpretation of these results should consider alcohol metabolism. PRS-Alcohol may capture not only the amount of alcohol consumed but also genetic influences on alcohol metabolism, particularly the oxidation of acetaldehyde, a group 1 carcinogen. The SNPs used to calculate PRS-Alcohol include variants linked to *ALDH2* expression levels. Notably, individuals with inactive *ALDH2* often consume less alcohol

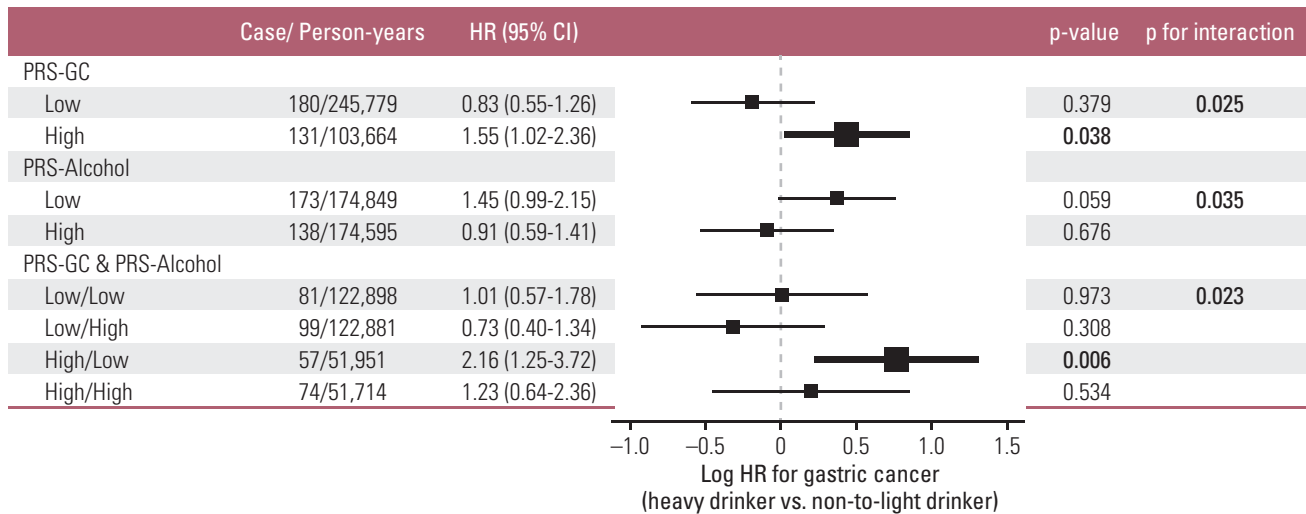


Fig. 2. Effect of drinking status on gastric cancer (GC) incidence stratified by polygenic risk score for GC (PRS-GC), PRS for alcohol consumption (PRS-Alcohol), and their combination. PRS-GC was dichotomized into low and high based on the 70th percentile, while PRS-Alcohol was dichotomized based on the 50th percentile. Drinking status was categorized into non-to-light and heavy drinkers. CI, confidence interval; HR, hazard ratio.

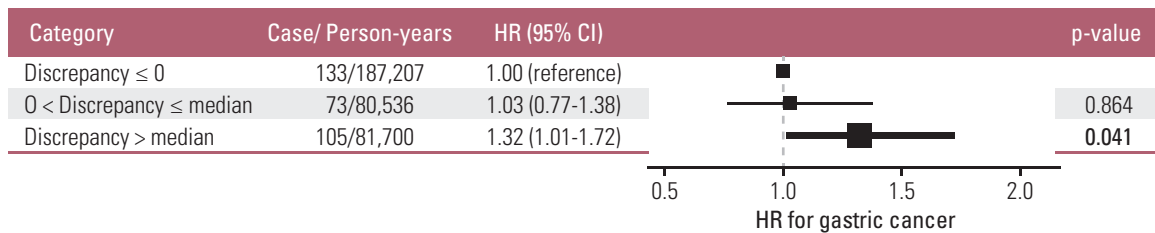


Fig. 3. Effect of discrepancy between alcohol consumption and polygenic risk score for alcohol consumption (PRS-Alcohol) on gastric cancer incidence. The discrepancy was defined as the difference between scaled alcohol consumption and scaled PRS-Alcohol. Discrepancies greater than zero were further stratified by the median value (0.78). CI, confidence interval; HR, hazard ratio.

due to adverse reactions such as flushing, tachycardia, and nausea, which are associated with severe acetaldehydemia [28]. As a result, individuals with inactive *ALDH2* tend to consume less alcohol than those with active *ALDH2* and may be less susceptible to alcohol-related GC. However, in practice, heavy drinkers with inactive *ALDH2* experience excessive acetaldehydemia, making them more susceptible to GC, underscoring the role of alcohol metabolism in gastric carcinogenesis [8]. In this context, we found that a greater discrepancy between actual alcohol consumption and genetically predicted levels was associated with an increased risk of alcohol-related GC.

We also found a significant positive interaction (β for interaction=0.09) between alcohol consumption and PRS-GC for GC risk. The OR for a 1 SD increase in PRS-GC was 1.25 for non-drinkers, compared to 1.37 for those consuming 20 g/day of ethanol. The moderating effect of PRS-GC can be

partly understood through *MUC1*, whose overexpression is associated with cancer progression [29]. Alcohol consumption has also been linked to increased *MUC1* expression [30], suggesting a potential combined effect of PRS-GC and alcohol consumption. Additionally, PRS-GC included SNPs in genes associated with alcohol dependence such as *CSMD1*, *LINGO2*, *MAG11*, and *SOX5* [31-33].

PRS-Alcohol was positively correlated with alcohol consumption, and the effect of PRS-Alcohol on GC could be interpreted as being mediated by alcohol consumption. Therefore, analyzing the effects of PRS-Alcohol on GC requires distinguishing between its direct and indirect effects through alcohol consumption. To better understand the relationship between PRS-Alcohol and alcohol consumption, we conducted a mediation analysis. Consistent with the logistic regression analysis, there was a significant interaction ($\beta=-0.08$) between PRS-Alcohol (exposure) and alcohol con-

sumption (mediator), as well as between PRS-GC and alcohol consumption ($\beta=0.09$). In the context of the PRS-Alcohol direct effect, PRS-Alcohol exhibits protective pure and total direct effects on GC, and the total direct effect (OR for a 1 SD increase=0.75) is smaller than the pure direct effect (OR, 0.81), which indicates that the protective effect of PRS-Alcohol is much stronger when individuals counterfactually consume more alcohol. Alcohol can lead to carcinogenesis through the accumulation of ethanol and acetaldehyde, which are highly reactive towards DNA. Ethanol and acetaldehyde are metabolized to acetate, which is relatively harmless [34]. Alcohol consumption was strongly linked to ethanol metabolism and had a negative genetic correlation with several diseases, such as coronary heart disease, type 2 diabetes, and BMI [35], which corresponds to our results.

Furthermore, in individuals with high PRS-GC, PRS-Alcohol showed a significant positive indirect effect on GC through alcohol consumption in the low PRS-Alcohol group (pure indirect effect; OR, 1.13), suggesting poor alcohol metabolism, whereas the indirect effect through alcohol consumption in the high PRS-Alcohol group (total indirect effect; OR, 1.04) was much smaller and not significant. From the perspective of alcohol metabolism, this result can be interpreted as indicating that alcohol consumption may have no significant effect on GC in the high PRS-Alcohol group where ethanol is well metabolized. For low PRS-GC, the total indirect effect was also smaller than the pure indirect effect, but neither effect was significant. Survival analysis also showed a significant increase in the risk of GC with higher alcohol consumption in the high-PRS-GC and low-PRS-Alcohol groups. The consistency of the interaction effects was confirmed using logistic, mediation, and survival analyses. These findings were replicated in an independent dataset of Korean individuals (KCPS-II).

Our study had several limitations. First, to develop the PRS, we used GWAS summary statistics for GC and alcohol consumption from the BBJ database. Our findings need to be confirmed using a Korean population dataset. Second, neither *H. pylori* infection status nor GC pathology data were available for the test set. In the validation set, alcohol consumption significantly increased the risk of intestinal-type GC but not diffuse-type GC. However, this was not confirmed in the test set.

In conclusion, Korean drinkers with higher PRS-GC who consume alcohol more than genetically predicted levels exhibited heightened susceptibility to the effects of alcohol consumption on the risk of incident GC. These results emphasize the importance of considering an individual's genetic profile, specifically PRS-GC and PRS-Alcohol, when assessing the impact of alcohol consumption on the risk of GC in Koreans. This personalized approach may improve

risk stratification and targeted interventions for individuals with a higher genetic risk.

Electronic Supplementary Material

Supplementary materials are available at Cancer Research and Treatment website (<https://www.e-crt.org>).

Ethical Statement

The study protocol was approved by the Ethics Committee of Seoul National University Bundang Hospital (IRB No. B-1610-366-303 and B-2008-777-301) and Seoul National University (E2308/001-020). Written informed consent was obtained from all subjects following ethical principles for medical research of the 64th World Medical Association Declaration of Helsinki.

Author Contributions

Conceived and designed the analysis: Kim N, Won S.

Collected the data: Jee SH, Kang SJ, Kim N, Won S.

Contributed data or analysis tools: Yie GE, Shin CM, Park K, Jo J, Do AR, Choi S, Ohn JH, Lee S, Kim J, Jee SH, Kang SJ, Kim N, Won S.

Performed the analysis: Yie GE, Shin CM, Park K, Jo J, Do AR, Kim N, Won S.

Wrote the paper: Yie GE, Shin CM, Kim N, Won S.

Writing - review & editing: Yie GE, Shin CM, Park K, Jo J, Do AR, Choi S, Ohn JH, Lee S, Kim J, Jee SH, Kang SJ, Kim N, Won S.

ORCID iDs

Ga-Eun Yie  : <https://orcid.org/0000-0001-8024-765X>

Cheol Min Shin  : <https://orcid.org/0000-0003-2265-9845>

Nayoung Kim  : <https://orcid.org/0000-0002-9397-0406>

Sungho Won  : <https://orcid.org/0000-0001-5751-5089>

Conflicts of Interest

Conflict of interest relevant to this article was not reported.

Funding

This study was conducted with bioresources from the National Biobank of Korea, the Korea Disease Control and Prevention Agency, Republic of Korea (KBN-2020-101). Statistical analyses were supported by the national supercomputing center with supercomputing resources including technical support (KSC-2022-CRE-0319). This work was supported by grant no. 13-2019-001, 02-2020-041, and 02-2023-0012 from the Seoul National University Bundang Hospital Research fund. This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korean government (MSIT) (Nos. RS-2024-00337453 and RS-2024-00346850).

References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71:209-49.
- Deng W, Jin L, Zhuo H, Vasiliou V, Zhang Y. Alcohol consumption and risk of stomach cancer: a meta-analysis. *Chem Biol Interact.* 2021;336:109365.
- Ferro A, Morais S, Rota M, Pelucchi C, Bertuccio P, Bonzi R, et al. Alcohol intake and gastric cancer: meta-analyses of published data versus individual participant data pooled analyses (StoP Project). *Cancer Epidemiol.* 2018;54:125-32.
- Lu Y, Chen J, Ding Y, Jin G, Wu J, Huang H, et al. Genetic variation of PSCA gene is associated with the risk of both diffuse- and intestinal-type gastric cancer in a Chinese population. *Int J Cancer.* 2010;127:2183-9.
- Cai M, Dai S, Chen W, Xia C, Lu L, Dai S, et al. Environmental factors, seven GWAS-identified susceptibility loci, and risk of gastric cancer and its precursors in a Chinese population. *Cancer Med.* 2017;6:708-20.
- Kato H, Ishikawa S. Lifestyles, genetics, and future perspectives on gastric cancer in east Asian populations. *J Hum Genet.* 2021;66:887-99.
- Baik I, Cho NH, Kim SH, Han BG, Shin C. Genome-wide association studies identify genetic loci related to alcohol consumption in Korean men. *Am J Clin Nutr.* 2011;93:809-16.
- Kang SJ, Shin CM, Sung J, Kim N. Association between ALDH2 polymorphism and gastric cancer risk in terms of alcohol consumption: a meta-analysis. *Alcohol Clin Exp Res.* 2021;45:6-14.
- Hidaka A, Sasazuki S, Matsuo K, Ito H, Sawada N, Shimazu T, et al. Genetic polymorphisms of ADH1B, ADH1C and ALDH2, alcohol consumption, and the risk of gastric cancer: the Japan Public Health Center-based prospective study. *Carcinogenesis.* 2015;36:223-31.
- Rhee TM, Choi J, Lee H, Merino J, Park JB, Kwak SH. Discrepancy between genetically predicted and observed BMI predicts incident type 2 diabetes. *Diabetes Care.* 2024;47:1826-33.
- Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc.* 2020;15:2759-72.
- Ishigaki K, Akiyama M, Kanai M, Takahashi A, Kawakami E, Sugishita H, et al. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat Genet.* 2020;52:669-79.
- Matoba N, Akiyama M, Ishigaki K, Kanai M, Takahashi A, Momozawa Y, et al. GWAS of 165,084 Japanese individuals identified nine loci associated with dietary habits. *Nat Hum Behav.* 2020;4:308-16.
- Shin CM, Park K, Kim N, Won S, Ohn JH, Lee S, et al. rs2671655 single nucleotide polymorphism modulates the risk for gastric cancer in *Helicobacter pylori*-infected individuals: a genome-wide association study in the Korean population. *Gastric Cancer.* 2022;25:573-85.
- Lee C, Choe EK, Choi JM, Hwang Y, Lee Y, Park B, et al. Health and Prevention Enhancement (H-PEACE): a retrospective, population-based cohort study conducted at the Seoul National University Hospital Gangnam Center, Korea. *BMJ Open.* 2018;8:e019327.
- Kim Y, Han BG; KoGES Group. Cohort profile: the Korean Genome and Epidemiology Study (KoGES) consortium. *Int J Epidemiol.* 2017;46:e20.
- Jee YH, Emberson J, Jung KJ, Lee SJ, Lee S, Back JH, et al. Cohort profile: the Korean Cancer Prevention Study-II (KCPS-II) biobank. *Int J Epidemiol.* 2018;47:385-6f.
- Moon S, Kim YJ, Han S, Hwang MY, Shin DM, Park MY, et al. The Korea biobank array: design and identification of coding variants associated with blood biochemical traits. *Sci Rep.* 2019;9:1382.
- Yoo SK, Kim CU, Kim HL, Kim S, Shin JY, Kim N, et al. NARD: whole-genome reference panel of 1779 Northeast Asians improves imputation accuracy of rare and low-frequency variants. *Genome Med.* 2019;11:64.
- Nam K, Kim J, Lee S. Genome-wide study on 72,298 individuals in Korean biobank data for 76 traits. *Cell Genom.* 2022;2:100189.
- Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J; Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47:291-5.
- Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics.* 2015;31:1466-8.
- Mak TS, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. *Genet Epidemiol.* 2017;41:469-80.
- Vilhjalmsson BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet.* 2015;97:576-92.
- Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun.* 2019;10:1776.
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl variant effect predictor. *Genome Biol.* 2016;17:122.
- GTEX Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45:580-5.
- Yoshida A, Huang IY, Ikawa M. Molecular abnormality of an inactive aldehyde dehydrogenase variant commonly found in Orientals. *Proc Natl Acad Sci U S A.* 1984;81:258-61.
- Apostolopoulos V, Stojanovska L, Gargosky SE. MUC1 (CD227): a multi-tasked molecule. *Cell Mol Life Sci.* 2015;72:4475-500.
- Benjamin JB, Jayanthi V, Devaraj H. MUC1 expression and its association with other aetiological factors and localization to mitochondria in preneoplastic and neoplastic gastric tissues. *Clin Chim Acta.* 2010;411:2067-72.
- Liu QR, Drgon T, Johnson C, Walther D, Hess J, Uhl GR. Addiction molecular genetics: 639,401 SNP whole genome

- association identifies many “cell adhesion” genes. *Am J Med Genet B Neuropsychiatr Genet.* 2006;141B:918-25.
32. Johnson C, Drgon T, Liu QR, Walther D, Edenberg H, Rice J, et al. Pooled association genome scanning for alcohol dependence using 104,268 SNPs: validation and use to identify alcoholism vulnerability loci in unrelated individuals from the collaborative study on the genetics of alcoholism. *Am J Med Genet B Neuropsychiatr Genet.* 2006;141B:844-53.
33. Morozova TV, Goldman D, Mackay TF, Anholt RR. The genetic basis of alcoholism: multiple phenotypes, many genes, complex networks. *Genome Biol.* 2012;13:239.
34. Rungay H, Murphy N, Ferrari P, Soerjomataram I. Alcohol and cancer: epidemiology and biological mechanisms. *Nutrients.* 2021;13:3173.
35. Sanchez-Roige S, Palmer AA, Clarke TK. Recent efforts to dissect the genetic basis of alcohol use and abuse. *Biol Psychiatry.* 2020;87:609-18.