Check for updates

Received: 9 May 2023 | Revised: 14 July 2023 | Accepted: 31 July 2023

DOI: 10.1002/ijc.34690

RESEARCH ARTICLE

Cancer Epidemiology

IJC INTERNATIONAL JOURNAL of CANCER

Genome-wide exploration of genetic interactions for bladder cancer risk

Evan Yi-Wen Yu[1,2,3] | Qiu-Yi Tang[4] | Ya-Ting Chen[1,2] | Yan-Xi Zhang[1,2] | Ya-Nan Dai[3] | Yu-Xuan Wu[1,2] | Wen-Chao Li[5] | Siamak Mehrkanoon[6] | Shi-Zhi Wang[7] | Maurice P. Zeegers[3] | Anke Wesselius[3]

[1]Key Laboratory of Environmental Medicine and Engineering of Ministry of Education, School of Public Health, Southeast University, Nanjing, China

[2]Department of Epidemiology & Biostatistics, School of Public Health, Southeast University, Nanjing, China

[3]Department of Epidemiology, CAPHRI Care and Public Health Research Institute, School of Nutrition and Translational Research in Metabolism, Maastricht University, Maastricht, the Netherlands

[4]Medical School of Southeast University, Nanjing, China

[5]Department of Urology, Affiliated Zhongda Hospital of Southeast University, Nanjing, China

[6]Information and Computing Sciences, Utrecht University, Utrecht, Netherlands

[7]Key Laboratory of Environmental Medicine Engineering, Ministry of Education, School of Public Health, Southeast University, Nanjing, China

Correspondence
Evan Yi-Wen Yu, Dingjiaqiao 87, 210009, Nanjing, China.
Email: evan.yu@maastrichtuniversity.nl

Anke Wesselius, P. Debyeplein 1 (Room A2.037), 6229 ER, Maastricht, the Netherlands.
Email: anke.wesselius@maastrichtuniversity.nl

Funding information
National Natural Science Foundation of China, Grant/Award Number: 82204033; Natural Science Foundation of Jiangsu Province, Grant/Award Number: BK20220826; Fundamental Research Funds for the Central Universities of China, Grant/Award Number: 2242022R10062/3225002202A1; Medical Foundation of Southeast University, Grant/Award Number: 4060692202/021; Dr. Evan Yi-Wen Yu is recepient of the Zhishan Young Scholar Award at the Southeast University, Grant/Award Number: 2242023R40031

Abstract

Although GWASs have been conducted to investigate genetic variation of bladder tumorigenesis, little is known about genetic interactions that may influence bladder cancer (BC) risk. By leveraging large-scale participants from UK Biobank, we established a discovery database with 4000 Caucasian participants (2000 cases vs 2000 non-cases), a database with 1648 Caucasian participants (824 cases vs 824 non-cases) and 856 non-Caucasian participants (428 cases vs 428 non-cases) as validation. We then performed a genome-wide SNP-SNP interaction investigation related to BC risk based a machine learning approach (ie, GenEpi). Moreover, we used the selected interactions to build a BC screening model with an integrated interaction-empowered polygenic risk score (iPRS) based on Cox proportional hazard model. With Bonferroni correction, we identified 10 statistically significant pairs of SNPs, which located in 17 chromosomes. Of these, four SNP-SNP interactions were found to be positively associated with BC risk among Caucasian participants (ORs 1.57-2.03), while six SNP-SNP interactions showed negatively associated with BC risk (ORs 0.54-0.65). Only four of the SNP-SNP interactions were consistently identified in non-Caucasian participants located in *ST7L-ADSS2*, *FHIT-CHDH*, *LARP4B-LHPP* and *RBFOX3-MPRIP*. In addition, the iPRS showed a

Abbreviations: AUC, area under the ROC curve comparison; BMI, body mass index; CA, Caucasian; CADD, combined annotation dependent depletion; CI, confidence interval; CV, cross-validation; EMT, epithelial-mesenchymal transition; G-E, gene-environment; G-G, gene-gene; GO, genetic ontology; GWAS, genome-wide association study; HES, hospital episode statistics; HR, hazard ratio; HRC, Haplotype Reference Consortium; ICD, international classification of diseases; iPRS, interaction-empowered polygenic risk score; LD, linkage disequilibrium; MAF, minor allele frequency; NC, non-Caucasian; OR, odds ratio; PC, principal component; PRS, polygenic risk score; PSM, propensity score matching; QC, quality control; ROC, operating characteristic curve; SNP, single nucleotide polymorphisms.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2023 The Authors. *International Journal of Cancer* published by John Wiley & Sons Ltd on behalf of UICC.

HR of 1.81 (95% CI: 1.46-2.09) compared the highest tertile to the lowest tertile, with an enhanced AUC (0.91; 95% CI:0.85-0.97) than PRS (AUC: 0.86; 95% CI:0.76-0.95; $P\text{-}_{\text{DeLong test}} = 2.2 \times 10^{-4}$). In summary, this study identified several important SNP-SNP interactions for BC risk, and developed an iPRS model for BC screening, which may help to identify the people at high-risk state of BC before early manifestation.

KEYWORDS
bladder cancer risk, GenEpi, genetic interaction, genetic screening

What's new?

Genetic variants currently associated with bladder cancer risk account for only a small proportion of familial clustering. While uncertainty remains, the gap in understanding of bladder cancer heritability is potentially explained by interactions between single nucleotide polymorphisms (SNPs). The present study describes novel SNP-SNP interactions linked to bladder cancer risk, which were identified using a machine learning approach. Based on the SNP pairs discovered, an interaction-empowered polygenic risk score (iPRS) was developed. The iPRS model successfully classified bladder cancer cases and non-bladder cancer controls, outperforming the classic PRS model and highlighting its potential to identify individuals at high-risk of malignancy.

1 | INTRODUCTION

Bladder cancer is among the most common cancers worldwide, with around 550 000 new cases and 200 000 deaths reported per year.[1] While there were few advances in clinical management of bladder cancer over the past decade, the incidence rate has been considerably increased.[2] Given its high frequency of recurrence and low health related quality of life with lifelong cystoscopy surveillance and multiple therapeutic interventions, bladder cancer is reported to be among the most expensive life-time treatments of all cancers and cause a heavy burden to the healthcare system.[3] As with many complex diseases, genetics plays an essential role in bladder cancer carcinogenesis. In the past decades, genome-wide association studies (GWASs) have identified multiple single nucleotide polymorphisms (SNPs) related to this disease.[4] Most identified variants, however, only confer relatively small increments in risk, and explain only a small proportion of familial clustering. Some of the remaining "missing heritability" might be captured by the interaction of certain SNPs (ie, genetic interaction).[5] Previous studies clearly showed that the impact of SNP interactions on the formation of diseases, including bladder cancer, is underestimated in traditional GWAS analysis,[6-9] in which only one genetic variant is considered at a time, and ignores underlying interaction of variants that might have stronger associations.[10] The evidence of previous studies have demonstrated that the existence of genetic interactions is an important factor contributing to phenotypes, especially in complex diseases such as hypertension, diabetes and cancer.[11] Particularly, genetic variants in multiple genes can synergistically lead to disease via different mechanisms, which has been described as "digenic heritance".[12,13] For both digenic heritance and genetic interactions, upon simultaneous mutation, the genetic variants either interact to produce disease or combine to produce a more complex, and usually more severe, phenotype that cannot be explained by variants in one gene alone. Nevertheless, no study investigating the genome-wide genetic interactions for bladder cancer has been constructed; therefore, the urgent need of investigation on genetic interaction in relation to bladder cancer risk has been emphasized.

As the number of SNP associated with disease risk increased exponentially, and so the number of interactions, a computational complex challenge arose in proceeding the statistical test of thousands of pairwise interactions. Despite some methods have been developed based on conventional algorithms to tackle the issue of genetic interactions (eg, FastEpistasis and BOOST),[14,15] machine learning approaches also provide an opportunity to reveal the genetic interaction, and has attracted a wide range of research interests in recent years.[16-18] Compared to the traditional methods, machine learning can effectively capture genetic interactions that characterize the biological mechanisms of disease.

The current study, therefore, aimed to construct a genomic exploration of genetic interaction in relation to bladder cancer risk based on GenEpi,[19] a computational package to uncover SNP-SNP interactions associated with phenotypes by the use of a machine learning approach that adopts two element combinatorial encoding when producing features and constructs the classification models by L1-regularized regression with stability selection. Furthermore, this study aimed to develop a bladder cancer screening model using both classic polygenetic risk score (PRS) and the detected genetic interactions for screening high-risk subpopulations.

2 | METHODS

2.1 | Study population

Study participants were originated from the UK Biobank study, which has been described in detail elsewhere.[20] In brief, the UK Biobank is a prospective study of over 0.5 million people living in the United Kingdom.

10970215, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ijc.34690 by Utrecht University , Wiley Online Library on [20/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

All people in the National Health Service registry who were aged 40 to 69 years and living <25 miles from a study center were invited to participate between 2006 and 2010. In total, 503 325 participants were recruited from over 9.2 million invitations. Extensive phenotypic data were self-reported upon baseline assessment by participants using touchscreen tests and questionnaires and at nurse-led interviews. Anthropometric assessments were conducted, and biological samples were collected at baseline. Health records were obtained from secondary care data from linked hospital episode statistics (HES).

The UK Biobank database for this project included 502 505 participants. Exclusion criteria included the withdrawal of informed consent (n = 12), lack of genetic data (n = 11 858), and missingness on either age, sex, BMI or smoking status (n = 17 886). Finally, 472 749 individuals were eligible for the current analyses (Figure 1). This study was conducted using the UK Biobank resource under Application #55889.

2.2 | Ascertainment of bladder cancer

The definitions for bladder cancer cases are presented in Table S1. Our analysis was restricted to bladder cancer patients with an International Classification of Diseases (ICD) codes of C67.0, C67.1, C67.2, C67.3, C67.4, C67.5, C67.6, C67.7, C67.8, C67.9, D09.0 (ICD10) and 1880, 1882, 1884, 1886, 1888, 1889, 2337 (ICD9); and self-report/doctor-diagnosis (code 1035 in field 20 001).

2.3 | Propensity score matching

With consideration of the imbalanced database in terms of outcomes, we used propensity score matching (PSM)[21] to select 1:1 ratio of bladder cancer cases and non-bladder cancer controls based on ethnicity, age, sex, BMI and smoking status. Accordingly, we established a discovery database with 4000 Caucasian participants (2000 cases vs 2000 non-cases), a database with 1648 Caucasian participants (824 cases vs 824 non-cases) as the first validation, and 856 non-Caucasian participants (428 cases vs 428 non-cases) as the second validation.

2.4 | Genotyping and quality control

A detailed description of the genotyping process, imputation and quality control in the UK Biobank study has been published elsewhere.[22,23] Briefly, the SNPs were genotyped using the custom UK

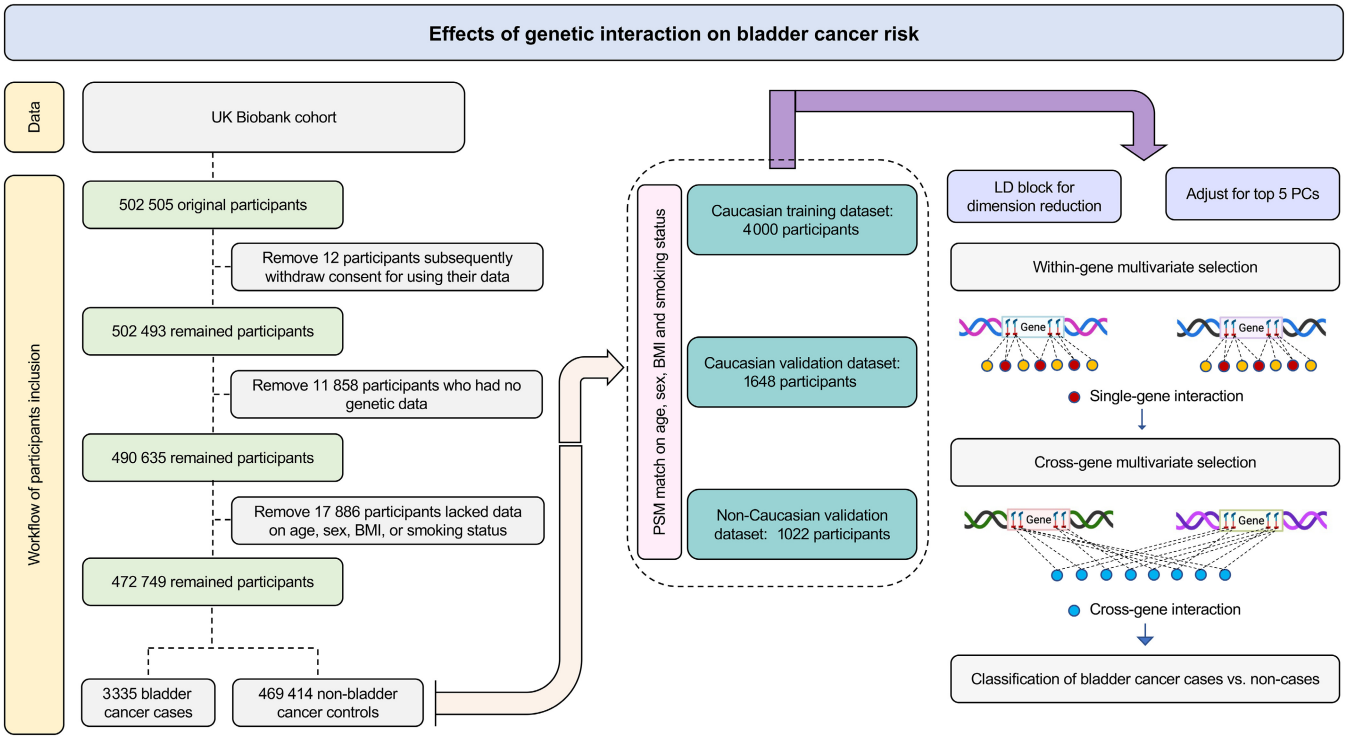


FIGURE 1 Study design, workflow and data processing based on the UK Biobank cohort. Study participants were originated from the UK Biobank study. The UK Biobank database for this project included 502 505 participants. Exclusion criteria included the withdrawal of informed consent (n = 12), lack of genetic data (n = 11 858), and missingness on either age, sex, BMI or smoking status (n = 17 886). Finally, 472 749 individuals were eligible for the current analyses. We used propensity score matching (PSM) to select 1:1 ratio of bladder cancer cases and non-bladder cancer controls based on ethnicity, age, sex, BMI and smoking status. Accordingly, we established a discovery database with 4000 Caucasian participants (2000 cases vs 2000 non-cases), a database with 1648 Caucasian participants (824 cases vs 824 non-cases) as the first validation, and 856 non-Caucasian participants (428 cases vs 428 non-cases) as the second validation. We used a machine-learning based approach, that is, GenEpi to discover genetic interactions in relation to bladder cancer risk based on the Caucasian discovery database and replicated in the Caucasian validation database and non-Caucasian validation database. BMI, body mass index; LD, linkage disequilibrium.

10970215, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ijc.34690 by Utrecht University , Wiley Online Library on [20/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Biobank Lung Exome Variant Evaluation Axiom (807 411 markers) or the UK Biobank Axiom array (825 927 markers). The two arrays shared 95% coverage resulting in >800 000 genotyped SNPs. Imputation was carried out using merged UK10K and 1000 Genomes Project Phase 3 panels as the reference panel.

Individuals who were identified by the UK Biobank as outliers based on either genotyping missingness rate or heterogeneity (968 samples), whose sex inferred from the genotypes did not match their self-reported sex (652 samples), were excluded. Population structure was captured by both principal components (PCs) analysis, along with K-means clustering on the PCs, which identified 453 964 subjects of European ancestry.

Only autosomal variants that were assayed by both aforementioned genotyping arrays employed by UK Biobank, were retained. Autosomal SNPs were pre-phased using SHAPEIT3 and imputed using IMPUTE4. In total, ~96 million SNPs were imputed. In addition, variants which had failed UK Biobank quality control procedures in any of the genotyping test, that is, batch effects (197 SNPs/batch), plate effects (284 SNPs/batch), Hardy-Weinberg equilibrium (572 SNPs/batch), sex effects (45 SNPs/batch) and array effects (5417 SNPs), were excluded (*P* value <10-12 or < 95% for all tests). Finally, related individuals were identified by estimating kinship coefficients for all pairs of samples, and only un-related participants were included for the current study.

2.5 | Assessment of genetic interactions related to bladder cancer risk

To discover genetic interactions on bladder cancer risk, a machine-learning based approach was applied, that is, GenEpi.[19] Considering the false positive rate and computational complexity, the focus of this study was only on pairwise (ie, SNP-SNP) interactions. The following main steps were conducted; at first genetic variants were grouped by a set of loci (ie, genes) in the genome using gene information available in the UCSC human genome annotation database, followed by dimensionality reduction of genetic features in each locus using linkage disequilibrium (LD).[24] This involves grouping of features into LD blocks using a given r^2 and D′ threshold of >0.8 and selection of the features with the largest minor allele frequency (MAF) to represent each block.[25] The selected genotypes of each single gene will then be independently tested for its association with bladder cancer risk by the use of a L1-regularized/logistic regression model. In the next stage, to identify cross-gene interactive features, both the individual SNPs and the previously selected within-gene interactive features were entered in the L1-regularized/logistic regression model to select the final genotype feature set. In addition, the first five PCs, that complied with the least adjustments of PCs based on "*twstat*" method,[26] were included as extra adjustments to build a final model, with a model evaluation by10-fold cross validation (CV). Given this study's focus is to identify SNP-SNP interactions associated with bladder cancer risk, the identified interactions were further analyzed by generating counts and frequencies of each two-locus genotype to understand the manner of each interaction. For the current study, GenEpi was applied to best-guess genotypes on the set of SNPs with nominal statistically significant association results (*P* value <.05) based on the filtered SNPs by LD blocks. The performance of GenEpi model was evaluated by *F1* score, which was computed by 2 × (precision × recall)/(precision + recall). More detailed information and implementation of using GenEpi on assessing the genetic interactions is described in the additional file Data S1.

2.6 | Sensitivity analyses of significant signals

Sensitivity analyses were performed by using the significant SNP-SNP interactions gained by making use of GenEpi in the additional Caucasian validation database and non-Caucasian validation database. MM, Mm and mm were used to denote the three genotypes of each SNP, that is, majority homozygous, heterozygous and minority homozygous, respectively. Then Mm and mm were attributed as SNP mutation while MM was attributed as non-SNP mutation (ie, reference). To quantify the interaction between two SNPs coded as; mm/Mm = 1, MM = 0, a standard logistic regression-based model was used. In this model, an interaction term between the two loci of identified significant SNPs was introduced with adjustments of the first five PCs (ie, outcome~$SNP_{1i} + SNP_{2i} + SNP_{1i}$*$SNP_{2i}$ + PC1 + … + PC5). The significance level for the estimation of the interaction was set at *P* value <.05 and replications should have a consistent direction of the effect.

Moreover, we presented the distribution of genotypes (ie, MM, Mm and mm) of identified SNPs and performed a chi-squared test to assess their individual and interactive differentiated distribution between bladder cancer cases and controls in the Caucasian discovery database.

2.7 | Functional annotation

An in silico approach through SNPnexus (https://www.snp-nexus.org/v4/),[27] RegulomeDB (http://regulome.stanford.edu/),[28] and HaploReg version 4.1 (http://archive.broadinstitute.org/mammals/haploreg/haploreg.php)[29] was used to predict the potential functions of the identified SNPs. Besides, to examine predicted functional impact, a combined annotation dependent depletion (CADD) method was annotated to the variants (Phred scores >5 predicted as deleterious, https://cadd.gs.washington.edu/score).

Pathway information with gene sets of all identified analytes were retrieved from the genetic ontology (GO) database (https://www.genome.jp/go/), accessed on October 18, 2022. The enrichment analyses were performed using the R package "*clusterProfiler*."[30]

2.8 | Longitudinal assessment of interaction-empowered polygenic risk score in relation to bladder cancer risk

To access the power of identified SNP-SNP interaction in predicting bladder cancer risk in longitudinal data, an enhanced cancer screening

10970215, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ijc.34690 by Utrecht University, Wiley Online Library on [20/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

model by Zhang et al[31] was applied. This method incorporated demographic factors and interaction-empowered polygenetic risk score (iPRS) to calculate the hazard ratios (HRs) of bladder cancer occurred in the follow-up data of the overall UK Biobank database based on Cox proportional hazard model, in which 469 996 study participants with 582 incident bladder cancer cases were included. The iPRS was computed based on a linear combination score of significant SNP-SNP interactions identified in the Caucasian discovery database, and reached nominal significance level (*P* value <.05) in the non-Caucasian validation database. The iPRS was calculated using the following equation, where β_{1i} and β_{2i} denoted the main effects of the two interactive SNPs and β_{3i} denoted their interaction effect. Again, the mutation of a SNP was coded with "1" as mutated while "0" as non-mutated:

$$\text{iPRS} = \sum_{i=1}^{n} (\beta 1i * \text{SNP1i} + \beta 2i * \text{SNP2i} + \beta 3i * \text{SNP1i} * \text{SNP2i})$$

The iPRS, composed of classic PRS and SNP-SNP interaction score, was used to generate an enhanced bladder cancer screening model together with age, sex, BMI and smoking status. At first the adjusted HRs (adjusted for age, sex, BMI and smoking status) and 95% confidence intervals (CIs) for bladder cancer were calculated using the continuous iPRS score. As a second step, HRs and 95% CIs were calculated using a categorized iPRS score. For this, the continuous iPRS score was categorized to a three-level categorical variable by its tertile values and the lowest group was set as the reference. Additionally, the same analysis for the classic PRS scores were repeated and compared to the performance of the iPRS score. The PRS score was calculated using the following equation (β_{1i} and β_{2i} denoted the main effects of the two interactive SNPs):

$$\text{PRS} = \sum_{i=1}^{n} (\beta 1i * \text{SNP1i} + \beta 2i * \text{SNP2i})$$

Both iPRS and PRS computed in the current study were shown to be normalized distributed (Figure S1). To assess whether the identified SNP-SNP interactions increase the prediction ability for bladder cancer risk in addition to demographic factors (ie, age, sex, BMI and smoking status), the receiver operating characteristic curve (ROC; R package "*pROC*",[32] and DeLong test[33] for area under the ROC curve comparison (AUROC) were used.

2.9 | Statistical power and multiple hypothesis testing

The following steps were used to decrease the number of false discoveries. First, high correlated SNPs were removed from the current study based on linkage disequilibrium (LD) pruning. Second, the GenEpi algorithm was used to reduce the enormous computational requirements. For this, each SNPs was independently tested for its association with bladder cancer risk. A *P* value threshold at <.05 was used as a significance level for detecting the SNP-disease associations. SNPs showing a *P* value of more than .05 were excluded. As a next step the SNP-SNP interactions that failed to reach the nominal significant level (ie, *P* value <.05) were excluded. Then a multiple correction (ie, .05/number of SNP-SNP interactions with *P* value <.05) was applied.

The statistical analyses mentioned above were performed with GenEpi (version 2.0.10) and R software (version 4.0.5).

3 | RESULTS

3.1 | Characteristics of the included participants

A detailed description of the study design, workflow and data processing based on the UK Biobank cohort is displayed in Figure 1. In total, 6504 study participants were included for GenEpi analyses. After PSM matching, no difference between bladder cancer cases and non-bladder cancer controls upon age, sex, BMI and smoking status were observed. The mean (SD) age at recruitment was 56.52 (8.09) years old, and 1610 participants (24.75%) were female. Nearly 67% of the participants were smokers, and the average BMI of all participants was 28.11 (5.06) kg/m^2 (Table 1). Figure 2A shows that the ancestry PCs differed between Caucasian and non-Caucasian participants, indicating the elicitability of *trans*-ethnic validation.

3.2 | Associations of the genetic interactions with bladder cancer risk

In the discovery phase, 3060 SNP-SNP interactions at *P* value <.05 were identified, of which 698 (22.81%) were single-gene interactions and 2362 (77.19%) were cross-gene interactions (Table S2). The median OR value obtained per SNP-SNP interaction was 1.71 (ranging from 1.44 to 2.37) for 1534 positive associations and 0.58 (ranging from 0.44 to 0.69) for 1526 negative associations. The majority of SNP-SNP interactions were shown to explain 0.5 to 1 (49.44%) and ≥1.5 (37.45%) risk estimates of interactive association with bladder cancer risk (Figure 2B). All of the interactions were distributed in 22 chromosomes, with 32 (1.01%) predicted to be with a probably deleterious effects and 126 (4.12%) showing a probably benign effect (Figure 2C). Among the 3060 SNP-SNP interactions, 72% were located in intronic regions, while 5% were up-stream located, 3% were down-stream located and only 1% were located in coding region (Figure 2D). In total the SNP-SNP interactions were attributed to 1346 genes, of which only 1% occurred within one gene, all remaining interactions (99%) were gene-gene (G-G) interactions (Figure 2E). Of the 3060 SNP-SNP interactions, 643 SNP-SNP interactions were further selected at *P* value $<1.6 \times 10^{-5}$ (0.05/3060, Bonferroni test), which were determined as the candidates for genetic interactions of bladder cancer risk (Table S2).

In the validation analysis, 10 SNP-SNP interactions were replicated with consistent directions in the discovery database at *P* value <.05. All showed to be cross-gene between 18 genes located in 17 chromosomes. Of these, four SNP-SNP interactions were found to

10970215, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ijc.34690 by Utrecht University, Wiley Online Library on [20/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Characteristics of included participants in the UK Biobank cohort[a]

Characteristics	Caucasian training database		Caucasian validation database		Non-Caucasian validation database	
	Case N = 2000	Non-case N = 2000	Case N = 824	Non-case N = 824	Case N = 428	Non-case N = 428
Year of birth[b] (year) (mean ± SD)	1945 ± 6	1945 ± 6	1945 ± 6	1945 ± 6	1947 ± 6	1947 ± 6
Sex (%)						
Male	1535 (76.75)	1535 (76.75)	618 (75.00)	618 (75.00)	294 (68.69)	294 (68.69)
Female	465 (23.25)	465 (23.25)	206 (25.00)	206 (25.00)	134 (31.31)	134 (31.31)
BMI (kg/m^2) (mean ± SD)	28.23 ± 4.74	28.14 ± 4.58	28.23 ± 4.58	28.27 ± 4.53	28.29 ± 5.03	28.04 ± 4.69
Smoking status						
Never (%)	652 (32.60)	652 (32.60)	265 (32.16)	265 (32.16)	143 (33.41)	140 (32.71)
Former (%)	325 (16.25)	325 (16.25)	113 (13.71)	113 (13.71)	75 (17.52)	75 (17.52)
Current (%)	1013 (50.65)	1013 (50.65)	441 (53.52)	441 (53.52)	208 (48.60)	208 (48.60)
Prefer not to answer	10 (0.5)	10 (0.5)	5 (0.61)	5 (0.61)	2 (0.47)	5 (1.17)
The first five PCs (mean ± SD)						
PC1	−12.34 ± 1.63	−12.38 ± 1.59	−12.31 ± 1.54	−12.36 ± 1.61	26.28 ± 81.83	44.38 ± 111.78
PC2	3.80 ± 1.51	3.77 ± 1.52	3.84 ± 1.47	3.76 ± 1.52	−9.50 ± 42.90	−9.49 ± 52.30
PC3	−1.61 ± 1.58	−1.61 ± 1.61	−1.63 ± 1.57	−1.62 ± 1.58	9.28 ± 25.89	9.28 ± 25.89
PC4	1.18 ± 2.87	1.24 ± 2.90	1.05 ± 2.86	1.34 ± 2.93	1.05 ± 2.86	1.34 ± 2.93
PC5	−1.13 ± 6.66	−0.92 ± 6.65	−1.46 ± 6.30	−0.73 ± 6.79	−1.46 ± 6.30	−0.73 ± 6.79

Abbreviations: BMI, body mass index; PC, principal component.

[a]The covariates, that is, age, sex, BMI and smoking status showed no significant different after propensity matching.

[b]Age at the time of recruitment. Descriptive statistics are presented as mean ± SD for age, BMI and PCs, and frequency (percentage, %) for sex and smoking status.

10970215, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ijc.34690 by Utrecht University, Wiley Online Library on [20/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

FIGURE 2 Characteristics of identified SNP-SNP interactions. (A) Genetic principal component of the included participants for GenEpi from UK Biobank regarding Caucasians and non-Caucasians; (B) Distribution of explained ORs that the SNP-SNP interactions contributed to the bladder cancer cases and non-bladder cancer controls; (C) Distribution of identified SNPs on chromosome locations and their predicted deleterious/benign information; (D) The proportion of predicted functional annotation classes of the identified genetic variants; (E) The proportion of single/cross gene classes of the identified SNP-SNP interactions. OR, odds ratio; PC, principal component; SNP, single nucleotide polymorphisms.

10970215, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ijc.34690 by Utrecht University , Wiley Online Library on [20/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

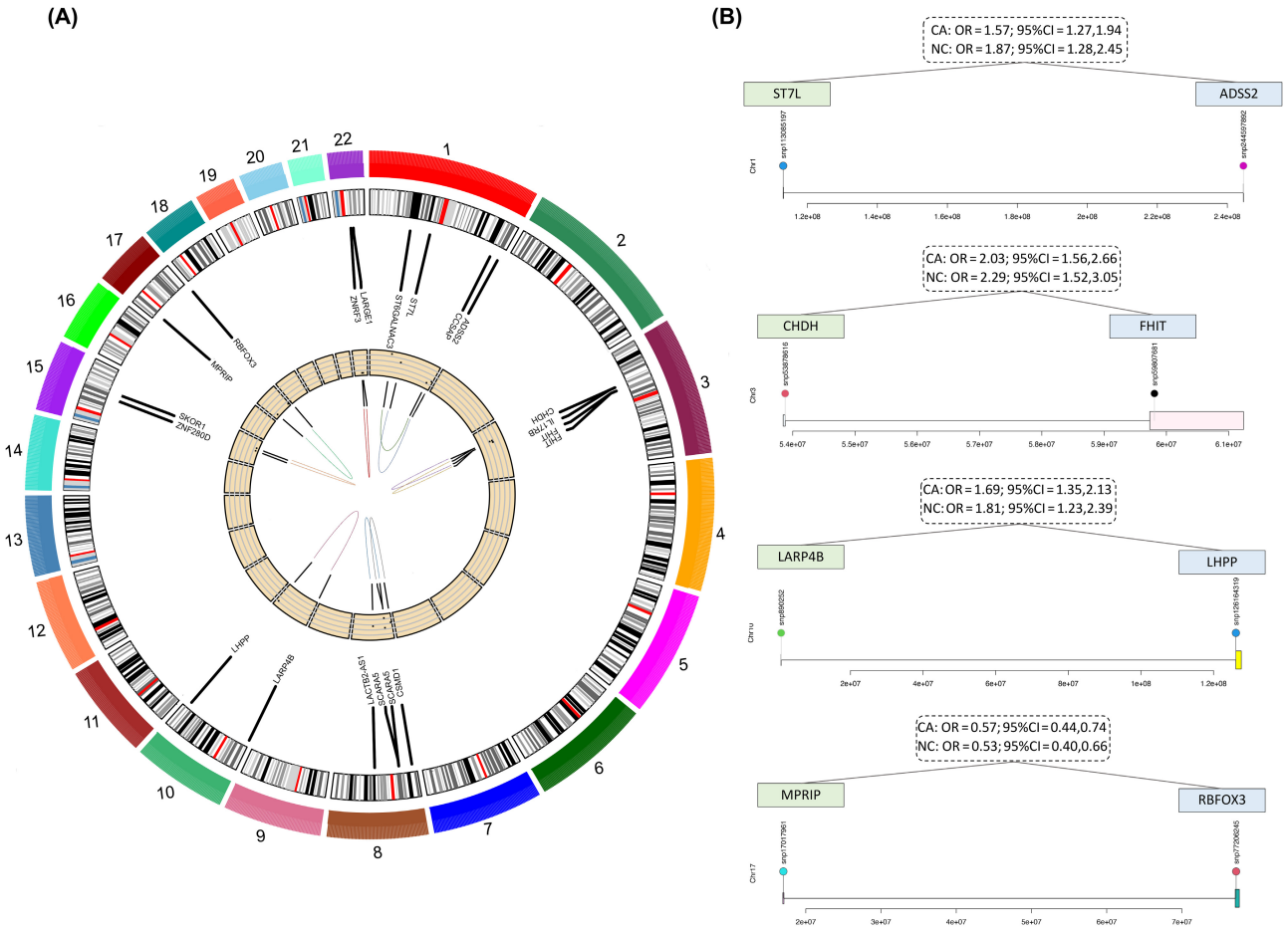


FIGURE 3 Genomic atlas of all identified SNP-SNP interactions based on Caucasian participants. (A) Overview of all identified SNPs involved in SNP-SNP interactions. Each dot represents a bladder cancer risk-associated genetic variant. The significant associations that should have (i) *P* value $<1.6 \times 10^{-5}$ (0.05/3060) in the discovery database; (ii) *P* value <.05 in the validation database; ii) consistent direction of effect between discovery and validation database; (B) The detailed description of the four SNP-SNP interactions replicated in non-Caucasian participants. CA, Caucasian; NC, non-Caucasian; OR, odds ratio; SNP, single nucleotide polymorphisms.

be positively associated with bladder cancer risk among Caucasian participants (the ORs ranging from 1.57 to 2.03), while six SNP-SNP interactions were found to be negatively associated with bladder cancer risk (the ORs ranging from 0.54 to 0.65).

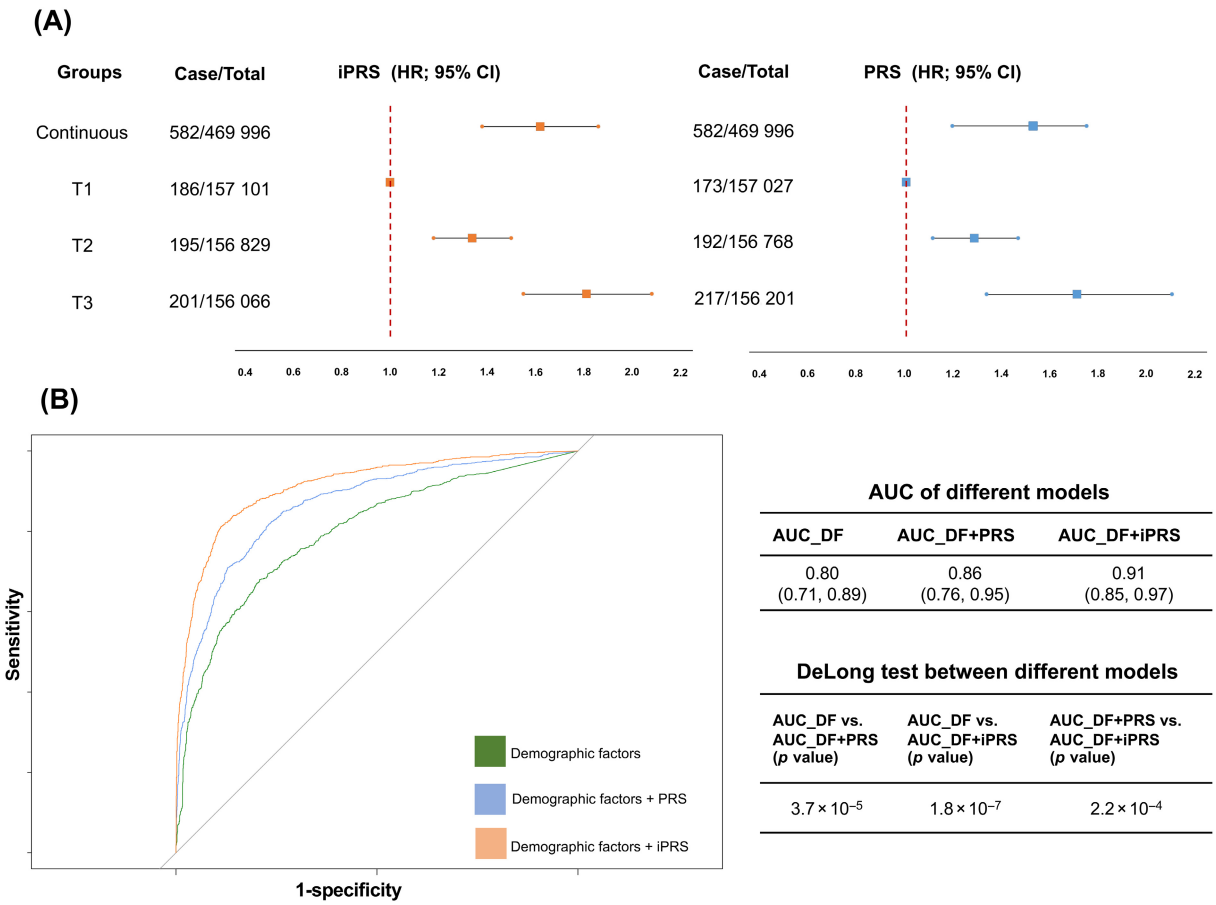
Only four of the SNP-SNP interactions above were consistently identified in non-Caucasian participants (1:113085197-1:244597892/*ST7L-ADSS2*, $OR_{Caucasian} = 1.57$, 95%CI = 1.27-1.94, $OR_{Non\text{-}Caucasian} = 1.87$, 95%CI = 1.28-2.45; 3:59807681-3:53878616/*FHIT-CHDH*, $OR_{Caucasian} = 2.03$, 95%CI = 1.56-2.66, $OR_{Non\text{-}Caucasian} = 2.29$, 95%CI = 1.52-3.05; 10:890252-10:126164319/*LARP4B-LHPP*, $OR_{Caucasian} = 1.69$, 95% CI = 1.35-2.13, $OR_{Non\text{-}Caucasian} = 1.81$, 95%CI = 1.23-2.39; 17:77206245-17:17017961/*RBFOX3-MPRIP*, $OR_{Caucasian} = 0.57$, 95% CI = 0.44-0.74, $OR_{Non\text{-}Caucasian} = 0.53$, 95%CI = 0.40-0.66; Figure 3 and Tables S3 and S4).

By evaluating the individual and interactive distribution of consistently identified SNPs between bladder cancer cases and controls in the discovery database, we observed the genotype distribution of two paired SNPs (ie, rs6537742 and rs3127462; rs75726847 and rs2289205; of rs6537742 and rs3127462; rs34598895 and rs4985733) and their genotype distribution under the interaction, which indicated the interactive distribution was more differentiated than individual distribution (Figure 5).

3.3 | Functional annotation of SNP-SNP interactions

In the RegulomeDB database, the abundant biological regulatory function was observed for 18 of the 22 identified SNPs, including eQTL, transcription factor binding site, or DNase peak. Among, three SNPs (ie, 1:229468223, 15:68115800 and 3:53879372) were predicted to be deleterious based CADD-Phred score (Table S5). Numerous enhancer histone marks and epigenetic alteration changes were observed for these SNPs (Table S6). Specifically, two SNPs (ie, 15:68115800 and 22:34208570) were observed to be located in the sites of the CpG island, which indicates a methylation of the corresponding DNA, resulting in an epigenetic alteration (Table S7). In addition, the identified SNPs were

10970215, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ijc.34690 by Utrecht University , Wiley Online Library on [20/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



AUC of different models

AUC_DF	AUC_DF+PRS	AUC_DF+iPRS
0.80 (0.71, 0.89)	0.86 (0.76, 0.95)	0.91 (0.85, 0.97)

DeLong test between different models

AUC_DF vs. AUC_DF+PRS (*p* value)	AUC_DF vs. AUC_DF+iPRS (*p* value)	AUC_DF+PRS vs. AUC_DF+iPRS (*p* value)
3.7×10^{-5}	1.8×10^{-7}	2.2×10^{-4}

FIGURE 4 iPRS and PRS in assessing the risk of bladder cancer. (A) Forest plot of association of continuous/Categorical iPRS and continuous/categorical PRS with risk of bladder cancer. The square dots denote the HRs; Horizontal lines represent the 95% confidence intervals (CIs). HR and 95% CI of each group were derived from Cox proportional hazard model adjusted for covariates (age, sex, BMI and smoking status) by setting the lowest group as reference; (B) The receiver operating characteristic curve (ROC) and prediction ability of iPRS, PRS and demographic factors in relation to bladder cancer risk. AUC, area under curve; CI, confidence interval; DF, demographic factors; HR, hazard ratio; iPRS, interaction-empowered polygenic risk score; PRS, polygenic risk score; SNP, single nucleotide polymorphisms; T, tertile.

mapped to 95 diseases that were attributed to mainly metabolic disorders (45%), cardiovascular disorders (13%) and cancers (7%) (Table S8). To biologically understand the genes mapped to interactive SNPs in the screening models, gene enrichment pathway analyses with the GO database were performed. In total, 14 pathways showed to be significant (*P* value <.05), and highly related to metabolism (Table S9).

3.4 | Integration of SNP-SNP interaction effectively distinguishes population at high risk of bladder cancer

In the longitudinal analysis of iPRS and PRS related to bladder cancer risk, 469 996 study participants from UK Biobank were included, contributed 7 153 962 person-years of follow-up (median for bladder cancer cases 3.57 years and for non-bladder cancer cases 15 years), with 582 incident bladder cancer cases (436 male, 146 female). The mean (SD) age at recruitment was 56.51 (8.09) years, and 262 677 participants (55.89%) were female (Table S10). Moreover, participants with bladder cancer were generally older (62.34 vs 56.51 years), were more often men (74.91% vs 45.07%), and had higher BMI (28.14 vs 27.41 kg/m^2; Table S10). Each subject was assigned an iPRS score, and were categorized into three groups by the tertile (ie, lowest, middle and highest) of the score. The SNP-SNP interactions were found to be maintained association with bladder cancer risk in the longitudinal analysis, both for the Caucasian participants (HRs ranged from 0.54 to 2.01 of 10 SNP-SNP interactions) and non-Caucasian participants (HRs ranged from 0.61 to 1.98 of four SNP-SNP interactions; Tables S11-S14). Subjects in the high-risk group (ie, the highest iPRS score group) had a significantly higher bladder cancer risk than those at the lowest risk group (HR: 1.81; 95% CI: 1.55-2.08; Tables S15 and S16).

Compared to the PRS score, which showed an increase of bladder cancer risk for participant in the highest tertile vs the lowest tertile;

10970215, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ijc.34690 by Utrecht University , Wiley Online Library on [20/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

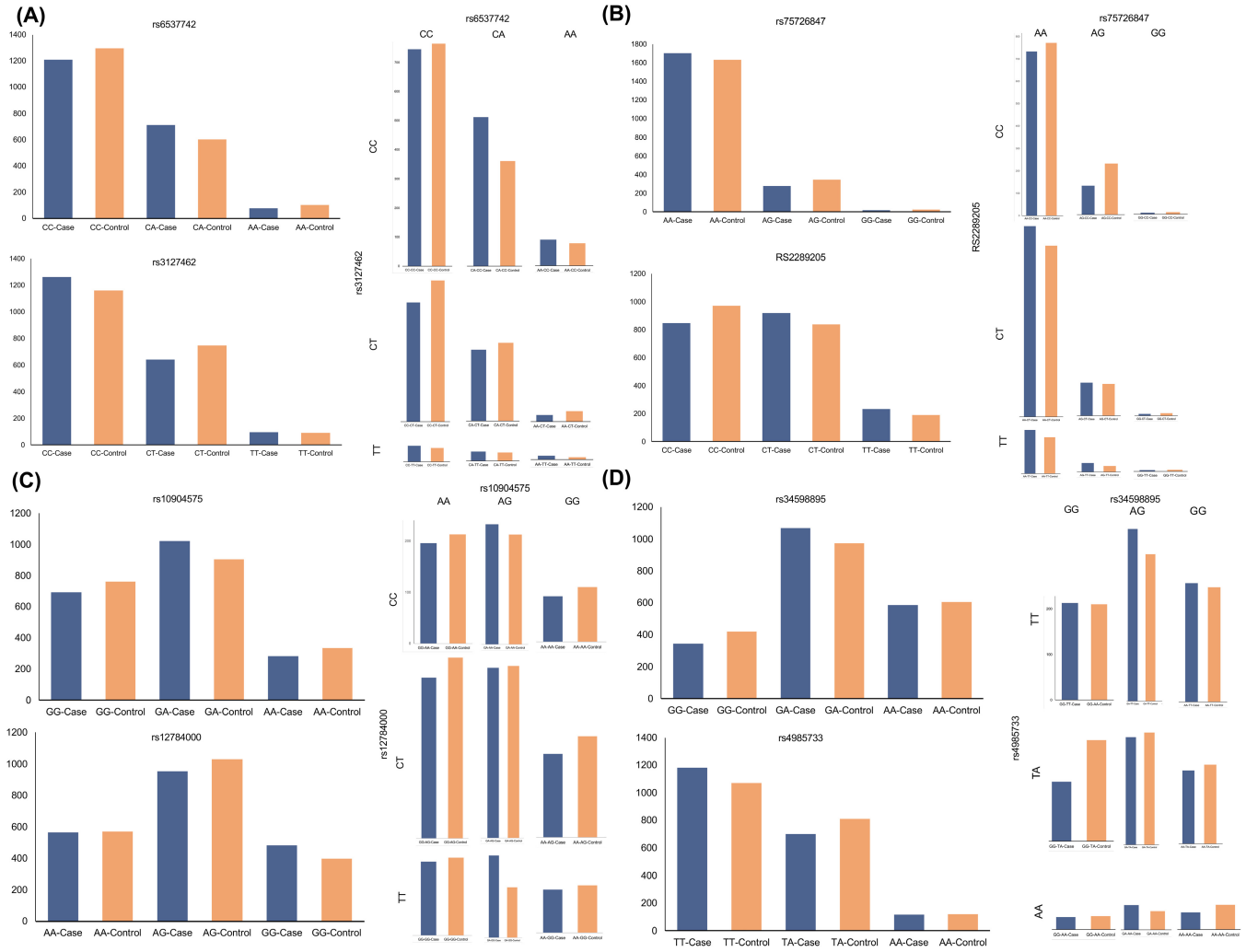


FIGURE 5 The individual and interactive distribution of consistently identified SNPs between bladder cancer cases and controls in the discovery database. (A) The distribution of cases and controls for rs6537742 (*P* value = .031) and rs3127462 (*P* value = .046); the distribution of cases and controls under the interaction of rs6537742 and rs3127462 (*P* value = .002). (B) The distribution of cases and controls for rs75726847 (*P* value = .035) and rs2289205 (*P* value = .038); the distribution of cases and controls under the interaction of rs75726847 and rs2289205 (*P* value = .003). (C) The distribution of cases and controls for rs10904575 (*P* value = .035) and rs12784000 (*P* value = .038); the distribution of cases and controls under the interaction of rs10904575 and rs12784000 (*P* value = .004). (D) The distribution of cases and controls for rs34598895 (*P* value = .030) and rs4985733 (*P* value = .034); the distribution of cases and controls under the interaction of rs34598895 and rs4985733 (*P* value = .006).

HR: 1.81 (95% CI: 1.46-2.09), the iPRS was found to have a better discrimination power with a narrower 95% CI. Meanwhile, we validated the bladder cancer screening model composed of demographic variables (age, sex, BMI and smoking status) solely, with iPRS, and with PRS, results showed the performance of classification (ie, area under ROC curve: AUC) of demographic variables (0.80; 95%CI: 0.71-0.89), and its combination with PRS (0.87; 95%CI: 0.77-0.95) and with iPRS (0.91; 95%CI: 0.85-0.97), indicating the iPRS enhanced model served as a satisfactory risk classifier (Figure 4). Furthermore, a clear pattern of high classification *F1* score separately by integration of within-chromosome identified SNP-SNP interactions, where the *F1* scores ranged from 0.72 to 0.82, which supported the evidence that the inclusion of genome-wide genetic interaction to perform a high classification accuracy between bladder cancer cases and non-bladder cancer controls (Figure S2).

4 | DISCUSSION

In the current study, a machine-learning based method was conducted to investigate genome-wide SNP-SNP interactions associated with bladder cancer risk. Here a total of 10 pairs of SNPs related to bladder cancer risk in Caucasians were identified, of which only four could be validated in non-Caucasians. To our knowledge, this is the first attempt to identify genetic interactions related to bladder cancer risk on a genome-wide scale. In addition, we developed an iPRS enhanced

10970215, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ijc.34690 by Utrecht University , Wiley Online Library on [20/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

bladder cancer screening model by incorporating the identified SNP-SNP signals. This model outperformed the classic PRS model, and therefore, has the potential to facilitate high-risk sub-populations screening.

In the past decades, large genome-wide association studies (GWASs) have identified multiple SNPs related to bladder cancer.[4] However, these SNPs only explain a small proportion of variation in the bladder cancer risk. Particularly, while certain SNPs have been identified related to bladder cancer risk in UK Biobank cohort according to Jiang et al,[34] no identified SNPs in the current study could be found at P value $<5 \times 10^{-5}$, indicating some missing heritance was not captured and may be due to genetic interactions (Table S17). Hence, recent research efforts expanded in studying genetic-environmental (G-E) interactions in relation to bladder cancer risk, including gene-smoking interactions, gene-asbestos interactions and gene-occupation interactions.[35] Studies on gene-gene (G-G) interactions in relation to bladder cancer risk, however, are still lacking, probably due to the lack of suitable algorithms and computationally intensive G-G interaction analyses on a genome-wide scale. In the current study, genome-wide genetic interaction analyses for bladder cancer susceptibility were performed based on a large-scale cohort, by which the results indicated genetic interactions or epistasis may also explain the missing heritability of bladder cancer.

All of the 10 identified SNP-SNP pairs were found to be cross-gene. Of them, 3:59807681 located in *FHIT* showed binding with *CHDH* (3:53878616) and *IL17RB* (3:53879372) in relation to bladder cancer risk, suggesting *FHIT* plays an important regulatory role in bladder cancer development. In line with this result, previous studies demonstrated that an aberrant *FHIT* expression inactivates the expression level of the FHIT protein, which is an important suppressor for tumor growth.[36-38] In addition, the *FHIT*, located at the *FRA3B* site of chromosome 3p14.2, is one of the histidine triad gene family members, which has been reported to be correlated with multiple human cancers.[39] Therefore, an aberrant *FHIT* may alter multiple biological functions in human malignancies including decreased apoptosis, increased epithelial-mesenchymal transition (EMT), increased resistance to genotoxic agents, altered production of reactive oxygen species, and ongoing genome instability.[40] The identified interaction between *FHIT* and *CHDH* was also validated in the non-Caucasian participants. Intriguingly, to our best knowledge, only one study linked to human diseases[41]; in this study it was shown that SNPs within the *CHDH* may be involved in the one-carbon metabolism and in reduction of the responses to arsenic metabolism. Arsenic contamination has shown to be an established risk factor for bladder cancer, and therefore, based on the finding of current study, it could be hypothesized that the interaction between the arsenic metabolism-related gene (ie, *CHDH*) and *FHIT* support the promotion of bladder cancer development. In addition, *FHIT* also showed to interact with the *IL17RB* in relation to bladder cancer risk; it has been demonstrated the changed patterns of the IL-17 family receptors expression, including IL17RB receptor, might be associated with infiltration of inflammatory cells and structural cells ($CD90^+$ fibroblasts and $CD31^+$ blood vessels), which may contribute to the development of bladder cancer.[42] However, the *FHIT-IL17RB* interaction could not be replicated in non-Caucasians, which suggested the generalizability of this specific interaction in trans-ethnic populations should be interpreted with caution. Further experimental studies are warranted to verify these results.

SCARA5 (8:27808579) also showed to interact with multiple genes (ie, *CSMD1* (8:4568747)-*SCARA5* and *LACTB2-AS1* (8:71542644)-*SCARA5*) in Caucasian participants. However, neither of these interactions could be replicated in non-Caucasian participants. While there is no solid evidence for the effect of *SCARA5* in the regulation of bladder cancer development, it has been reported to be involved in the growth, migration and invasion of several other cancers.[43-45]

A notable finding in the current study is the *ST7L* (1:113085197)-*ADSS2* (1:244597892) interaction, causing an increased bladder cancer risk in both Caucasian and non-Caucasian participants. So far, no previous studies showed either of these genes to be related to bladder cancer risk. As *ST7L* is a key factor of Wnt/GSK-3β/β-catenin signaling pathway,[46,47] the interaction with *ADSS2* might inhibit the function of *ST7L* that is the suppression of tumorigenicity, and thereby enhance the development of bladder cancer. Similarly, the *LARP4B* (10:890252)-*LHPP* (10:126164319) interaction was also found to increase the bladder cancer risk in both Caucasian and non-Caucasian participants. Since *LHPP* suppresses bladder cancer cell proliferation and growth via inactivating AKT/p65 signaling pathway,[48] and *LARP4B* was reported to serves as a tumor-suppressor gene, the aberrant alteration of each gene including their interaction might impede the favorable functions and suggest highly deleterious effects on bladder cancer.

In addition, the current study identified an interaction between *RBFOX3* (17:77206245) and *MPRIP* (17:17017961), showing a negative association with bladder cancer risk. According to a recent study,[49] the rs978416 G > A SNP in *RBFOX3* may be related to bladder cancer predisposition in a Chinese population and might serve as a novel biomarker for bladder cancer risk. However, the effects of all the identified genetic interactions still need to be verified in future studies.

Although the underlying mechanisms of identified SNP-SNP interactions are poorly understood, accumulative evidence has emerged suggesting that identifying high-risk individuals can enable enhanced screening and lead to better treatment options. This could directly result in a reduced cancer incidence. Therefore, the clinical use of the PRS, which integrates multiple SNPs and might distinguish individual with high or low disease risk, has been widely recognized. However, the discriminatory ability in distinguishing cancer cases from healthy controls by PRSs still needs to be enhanced.

According to Zhu et al,[50] screening of asymptomatic subjects with high cancer risk is a well-recognized way to reduce cancer morbidity and mortality by detecting very early-stage cases or those predisposed to bladder cancer. Therefore, many cancer risk screening models have been created in the last couple of decades. Whereas at first theses models only included including clinical factors, more recently also genetic factors (ie, PRS) have been included. The inclusion of the PRS showed to have a significant influence on the ability of targeting subjects at high risk for bladder cancer.[51] The current

10970215, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ijc.34690 by Utrecht University, Wiley Online Library on [20/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

study showed that the proposed iPRS, including the SNP-SNP interactions, outperformed the classically used PRS, thereby possessing an additional capability to substantially enhance the guideline- and model-based bladder cancer screening strategies. Since the individual genome-wide genetic measurement has yet to widely applied in real-word clinical practice due to high costs and technological limitation. Future development of simple and rapid methods to detect genes and SNPs will therefore enhance the clinical use of the iPRS, by developing custom-designed chips for screening usage.

The current study has several limitations. First, the focus was only on pairwise interactions, as the computational burden of high-order interactions is prohibitive and the interpretation of high-order interactions is difficult. Second, the biological mechanisms of the SNPs involved in the identified genetic interactions were not verified, which may warrant further functional studies. Third, since the current study was primarily designed for a European-ancestry population with most participants being Europeans, future G-G interaction studies on subjects with other ancestries are needed.

In summary, a machine-learning based approach was applied to construct a genetic interaction framework and identified several novel SNP-SNP interactions in association with bladder cancer risk, which might reduce the gap in genetic risk screening. In addition, an iPRS was derived from genome-wide SNPs that showed to be able to effectively classify bladder cancer cases and non-bladder cancer controls, which may help to enhance the identification of individuals at high-risk for bladder cancer.

AUTHOR CONTRIBUTIONS

The work reported in the paper has been performed by the authors, unless clearly specified in the text. Conceived and designed the study: Evan Yi-Wen Yu, Anke Wesselius; supervision: Maurice P. Zeegers, Anke Wesselius; conducted data analyses and interpretation and drafted the manuscript: Evan Yi-Wen Yu, Qiu-Yi Tang; data curation: Evan Yi-Wen Yu, Qiu-Yi Tang; Critical revision of the manuscript: Qiu-Yi Tang, Ya-Ting Chen, Yan-Xi Zhang, Ya-Nan Dai, Yu-Xuan Wu, Wen-Chao Li, Siamak Mehrkanoon, Shi-Zhi Wang, Maurice P. Zeegers, Anke Wesselius. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

The authors acknowledge Yu-Chuan Chang and Chien-Yu Chen from GenEpi team for help in solving the technical issues. The authors acknowledge certain figures (ie, graphical abstract and Figure 1) were created, adapted and exported from BioRender.com (2023). Retrieved from https://app.biorender.com/biorender-templates.

FUNDING INFORMATION

This research was funded by the National Natural Science Foundation of China (82204033), Natural Science Foundation of Jiangsu Province (BK20220826), Fundamental Research Funds for the Central Universities of China (2242022R10062/3225002202A1) and Medical Foundation of Southeast University (4060692202/021). Dr.Evan Yi-Wen Yu is recepient of the Zhishan Young Scholar Award at the Southeast University (2242023R40031).

CONFLICT OF INTEREST STATEMENT

No potential conflicts of interest relevant to this article were reported.

DATA AVAILABILITY STATEMENT

This work has been conducted using the UK Biobank Resource under Application Number 55889. The UK Biobank is an open access resource and bona fide researchers can apply to use the UK Biobank dataset by registering and applying at http://ukbiobank.ac.uk/register-apply/. Further information is available from the corresponding author upon request. Analysis code for GenEpi is available in Data S1.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by NHS National Research Ethics Service North West (11/NW/0382). The patients/participants provided their written informed consent to participate in this study.

ORCID

Evan Yi-Wen Yu https://orcid.org/0000-0001-7825-5087
Anke Wesselius https://orcid.org/0000-0003-4474-9665

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424.
2. Mitra AP, Cai J, Miranda G, et al. Management trends and outcomes of patients undergoing radical cystectomy for urothelial carcinoma of the bladder: evolution of the University of Southern California experience over 3,347 cases. *J Urol*. 2022;207(2):302-313.
3. Lenis AT, Lec PM, Chamie K, Mshs MD. Bladder cancer: a review. *Jama*. 2020;324(19):1980-1991.
4. Selinski S. Discovering urinary bladder cancer risk variants: status quo after almost ten years of genome-wide association studies. *EXCLI J*. 2017;16:1288-1296.
5. Maher B. Personal genomes: the case of the missing heritability. *Nature*. 2008;456(7218):18-21.
6. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*. 2008;9(5):356-369.
7. Eichler EE, Flint J, Gibson G, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet*. 2010;11(6):446-450.
8. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747-753.
9. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nat Rev Genet*. 2009;10(4):241-251.
10. Shriner D, Vaughan LK, Padilla MA, Tiwari HK. Problems with genome-wide association studies. *Science*. 2007;316(5833):1840-1842.
11. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet*. 2009;10(6):392-404.

10970215, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ijc.34690 by Utrecht University , Wiley Online Library on [20/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

12. Mukherjee S, Cogan JD, Newman JH, et al. Identifying digenic disease genes via machine learning in the undiagnosed diseases network. *Am J Hum Genet*. 2021;108(10):1946-1963.
13. Okazaki A, Ott J. Machine learning approaches to explore digenic inheritance. *Trends Genet*. 2022;38(10):1013-1018.
14. Schüpbach T, Xenarios I, Bergmann S, Kapur K. FastEpistasis: a high performance computing solution for quantitative trait epistasis. *Bioinformatics*. 2010;26(11):1468-1469.
15. Wan X, Yang C, Yang Q, et al. BOOST: a fast approach to detecting gene-gene interactions in genome-wide case-control studies. *Am J Hum Genet*. 2010;87(3):325-340.
16. Bureau A, Dupuis J, Falls K, et al. Identifying SNPs predictive of phenotype using random forests. *Genet Epidemiol*. 2005;28(2):171-182.
17. Schwarz DF, König IR, Ziegler A. On safari to random jungle: a fast implementation of random forests for high-dimensional data. *Bioinformatics*. 2010;26(14):1752-1758.
18. Wan X, Yang C, Yang Q, Xue H, Tang NL, Yu W. MegaSNPHunter: a learning approach to detect disease predisposition SNPs and high level interactions in genome wide association study. *BMC Bioinformatics*. 2009;10:13.
19. Chang YC, Wu JT, Hong MY, et al. GenEpi: gene-based epistasis discovery using machine learning. *BMC Bioinformatics*. 2020;21(1):68.
20. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779.
21. Liang J, Hu Z, Zhan C, Wang Q. Using propensity score matching to balance the baseline characteristics. *J Thorac Oncol*. 2021;16(6):e45-e46.
22. Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500,000 UK biobank participants. *BioRxiv*. 2017;166298.
23. Dashti HS, Jones SE, Wood AR, et al. Genome-wide association study identifies genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. *Nat Commun*. 2019;10(1):1100.
24. Rosenbloom KR, Armstrong J, Barber GP, et al. The UCSC genome browser database: 2015 update. *Nucleic Acids Res*. 2015;43(D1):D670-D681.
25. Lewontin RC. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*. 1964;49(1):49-67.
26. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2(12):e190.
27. Oscanoa J, Sivapalan L, Gadaleta E, Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). *Nucleic Acids Res*. 2020;48(W1):W185-w192.
28. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790-1797.
29. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*. 2012;40(D1):D930-D934.
30. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics*. 2012;16(5):284-287.
31. Zhang R, Shen S, Wei Y, et al. A large-scale genome-wide gene-gene interaction study of lung cancer susceptibility in Europeans with a trans-ethnic validation in Asians. *J Thorac Oncol*. 2022;17(8):974-990.
32. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
33. Demler OV, Pencina MJ, D'Agostino RB Sr. Misuse of DeLong test to compare AUCs for nested models. *Stat Med*. 2012;31(23):2577-2587.
34. Jiang L, Zheng Z, Fang H, Yang J. A generalized linear mixed model association tool for biobank-scale data. *Nat Genet*. 2021;53(11):1616-1621.
35. Horikawa Y, Gu J, Wu X. Genetic susceptibility to bladder cancer with an emphasis on gene–gene and gene–environmental interactions. *Curr Opin Urol*. 2008;18(5):493-498.
36. Kandimalla R, van Tilborg AA, Zwarthoff EC. DNA methylation-based biomarkers in bladder cancer. *Nat Rev Urol*. 2013;10(6):327-335.
37. Liu XP, Yin XH, Yan XH, Zeng XT, Wang XH. The clinical relevance of fragile histidine triad protein (FHIT) in patients with bladder cancer. *Med Sci Monit*. 2018;24:3113-3118.
38. Vecchione A, Sevignani C, Giarnieri E, et al. Inactivation of the FHIT gene favors bladder cancer development. *Clin Cancer Res*. 2004;10(22):7607-7612.
39. Ohta M, Inoue H, Cotticelli MG, et al. The FHIT gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. *Cell*. 1996;84(4):587-597.
40. Waters CE, Saldivar JC, Hosseini SA, Huebner K. The FHIT gene product: tumor suppressor and genome "caretaker". *Cell Mol Life Sci*. 2014;71:4577-4587.
41. Iliopoulos D, Guler G, Han SY, et al. Fragile genes as biomarkers: epigenetic control of WWOX and FHIT in lung, breast and bladder cancer. *Oncogene*. 2005;24(9):1625-1633.
42. Liu Y, Yang W, Zhao L, et al. Immune analysis of expression of IL-17 relative ligands and their receptors in bladder cancer: comparison with polyp and cystitis. *BMC Immunol*. 2016;17(1):1-10.
43. Liu J, Zeng ML, Shi PC, Cao YP, Zhang JL, Xie YP. SCARA5 is a novel biomarker in colorectal cancer by comprehensive analysis. *Clin Lab*. 2020;66(7):1-7.
44. Wen X, Wang N, Zhang F, Dong C. Overexpression of SCARA5 inhibits tumor proliferation and invasion in osteosarcoma via suppression of the FAK signaling pathway. *Mol Med Rep*. 2016;13(3):2885-2891.
45. Zhang H, Liu C, Wang X, Wang Y, Zheng J. SCARA5 inhibits gastric cancer progression via epithelial-mesenchymal transition suppression. *J Cancer*. 2021;12(8):2412-2421.
46. Wang H, Sun L, Jiang J, Yu S, Zhou Q. Suppression of the proliferation and invasion of breast cancer cells by ST7L occurs through inhibition of activation of Wnt/GSK-3β/β-catenin signalling. *Clin Exp Pharmacol Physiol*. 2020;47(1):119-126.
47. Zhan T, Chen X, Tian X, et al. MiR-331-3p links to drug resistance of pancreatic cancer cells by activating WNT/β-catenin signal via ST7L. *Technol Cancer Res Treat*. 2020;19:1533033820945801.
48. Li Y, Zhang X, Zhou X, Zhang X. LHPP suppresses bladder cancer cell proliferation and growth via inactivating AKT/p65 signaling pathway. *Biosci Rep*. 2019;39(7):1-11.
49. Wang X, Guo Z, Zhu H, et al. Genetic variants in splicing factor genes and susceptibility to bladder cancer. *Gene*. 2022;809:146022.
50. Zhu M, Wang T, Huang Y, et al. Genetic risk for overall cancer and the benefit of adherence to a healthy lifestyle. *Cancer Res*. 2021;81(17):4618-4627.
51. Kachuri L, Graff RE, Smith-Byrne K, et al. Pan-cancer analysis demonstrates that integrating polygenic risk scores with modifiable risk factors improves risk prediction. *Nat Commun*. 2020;11(1):6084.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Yu EY-W, Tang Q-Y, Chen Y-T, et al. Genome-wide exploration of genetic interactions for bladder cancer risk. *Int J Cancer*. 2024;154(1):81-93. doi:10.1002/ijc.34690

10970215, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ijc.34690 by Utrecht University , Wiley Online Library on [20/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

B-cell malignancies - A new knowledge hub on the latest research in therapeutic advances

EDUCATIONAL CONTENT AVAILABLE ON THE HUB:

- On-demand Webinars - earn CME credit
- Infographics
- Patient Case Studies
- Currated Research Articles

...and much more

VISIT KNOWLEDGE HUB TODAY

This educational resource has been supported by Eli Lilly.

WILEY