BMC Cancer

RESEARCH

Open Access

Novel genetic loci and functional properties of immune-related genes for colorectal cancer survival in Korea

Dabin Yun[1†], Jung-Ho Yang[2†], Soyoun Yang[1], Jin-ah Sim[3], Minjung Kim[4,5], Ji Won Park[4,5], Seung Yong Jeong[4,5], Aesun Shin[6], Sun-Seog Kweon[2†] and Nan Song[1*†]

Abstract

One major topic in colorectal cancer (CRC) research is the role of immune cells against cancer cells. The association of single-nucleotide polymorphisms (SNPs) and polygenic risk scores (PRS) with CRC was examined and their functional properties were identified using a gene-gene interaction network. 960 CRC patients at Seoul National University Hospital (SNUH, discovery) and 6,627 CRC patients at Chonnam National University Hospital (CNNUH, validation) were enrolled. SNPs were genotyped using the Korean Biobank Array. 2,729 immune-related genes were selected from the Ensembl, Gene Ontology, and Kyoto Encyclopedia of Genes and Genomes, and 37,398 SNPs were mapped. PRS were categorized into tertiles. Cox proportional hazard models were fitted for overall survival (OS) and progression-free survival (PFS). A gene-gene interaction network was analyzed. Among CRC patients from SNUH, 154 (16.0%) died, while 245 (25.5%) had progression. In CNNUH, 3,537 (53.4%) died. For OS, the most significant association was observed for rs117322760 (8q23.1, *PKHD1L1*, hazard ratio (HR) = 4.58, p-value = 1.40×10^{-6}). For PFS, it was observed in rs143531681 (7q36.1, *NOS3*, HR = 4.67, p-value = 9.72×10^{-8}). For PRS, the highest tertile group showed an increased risk for OS (HR = 59.58, p-value = 9.20×10^{-48}) and PFS (HR = 9.81, p-value = 1.69×10^{-23}). Significant interactions were observed between *PIK3R2* and *PIK3CA* for OS and *ALOX5* and *COTL1* for PFS. This study presented novel genetic variants associated with OS and PFS in CRC patients, and notable findings from the analysis of PRS and the gene-gene interaction.

Keywords Colorectal cancer, Immune-related gene, Single-nucleotide polymorphism, Polygenic risk score, Gene-gene interaction network, Survival

†Dabin Yun and Jung-Ho Yang are co-first authors and contributed equallty to this work.

†Sun-Seog Kweon and Nan Song are co-corresponding authors and contributed equally as senior authors.

*Correspondence:
Nan Song
nan.song@chungbuk.ac.kr
[1]College of Pharmacy, Chungbuk National University, Cheongju, Chungbuk, Korea
[2]Department of Preventive Medicine, Chonnam National University Medical School, Hwasun, Jeonnam, Korea
[3]Department of AI Convergence, Hallym University, Chuncheon, Gangwon, Korea
[4]Department of Surgery, College of Medicine and Hospital, Seoul National University, Seoul, Korea
[5]Cancer Research Institute, Seoul National University, Seoul, Korea
[6]Department of Preventive Medicine, College of Medicine, Seoul National University, Seoul, Korea

Introduction

In South Korea, colorectal cancer (CRC) is a leading malignancy, ranking third in incidence among men and women [1]. Moreover, it stands as the third cause of cancer-related deaths in men and women [1].

Owing to the early detection of nationwide medical screenings and the improvements in cancer treatments, the survival rate of CRC has improved from 56.2% (1993–1995) to 74.3% (2016–2020) [1]. Nonetheless, patients with a late diagnosis of CRC persistently face an elevated mortality risk [1]. This assertion is substantiated by data indicating that the five-year survival probabilities for CRC at stages 1 and 4 are 91.9% and 30.4%, respectively [2].

The role of immune cells in cancer therapy is a critical area of research and new anticancer therapies based on the immune system are being developed [3]. Immune checkpoint inhibitors (ICIs) targeting PD-L1, PD-1, CAR-T cells, and novel ICIs targeting unchallenged targets have emerged as breakthroughs in this field [4]. This trend is evident in CRC treatment, particularly in late-stage cases, where an effective immunological response can significantly improve overall survival (OS) and progression-free survival (PFS) by affecting metastatic features [4, 5]. Immunity-related biological pathways are gaining attention in CRC research. Pathways such as Interleukin-1 signaling, CLEC7A in innate immunity, and B cell receptor signaling significantly impact the survival and clinical outcomes of patients with CRC [6].

Several genetic association studies have been conducted on immune-related genes and survival outcomes of association studies. *IL23A*, *LIF*, *VGF*, *SLIT2*, and *FGF18* have a negative effect on the survival outcome of patients with CRC [7]. Additionally, rs4464148 (18q21.1, *SMAD7*) and rs6983267 (8q24.21, *CCAT2*) are associated with poorer OS outcomes in patients with CRC [8]. One of the limitations of previous candidate gene approaches and genome-wide association studies (GWAS) is that they did not apply functional analysis to provide biological interpretations and clinical implications. Therefore, functional analyses such as polygenic risk scoring and gene network analysis could be utilized to connect genetic association results with additional beneficial biological and clinical effects.

The role of polygenic risk scores (PRS) was emphasized in the genetic analysis of cancer. The importance of PRS in cancer genetics for estimating genetic risk and enabling precision medicine has been highlighted [9].

Polygenic risk score-based genetic screening of patients with CRC could yield more accurate results than the current guidelines [10]. Moreover, applying the concept of PRS to CRC revealed that the cumulative burden of CRC-associated common genetic variants is more strongly associated with early-onset CRC compared with late-onset CRC [11]. Additionally, network-based functional analysis in cancer research is being extensively studied for a detailed biological understanding of SNP analysis outcomes [12–14]. Therefore, this study aimed to identify the susceptibility of single nucleotide polymorphisms (SNPs) and genes using genome-wide immune-related gene association studies to uncover the clinical importance of the immune system for survival outcomes in patients with CRC. Additionally, we utilized PRS and gene-gene interaction networks to reveal the clinical and biological impacts of susceptible loci in CRC.

Methods

Study population and follow-up

Patients with CRC were recruited in 2002 and underwent surgical resection at the Seoul National University Hospital (SNUH), Korea. This study included patients diagnosed with CRC and who underwent surgery in 2013.Among 1,099 selected patients, we have performed genome-wide SNP arrays in 960 patients. In this study, all patients underwent surgical treatment between May 2014 and June 2017 and were prospectively followed up until April 14, 2023. Additional details of the patient selection process were presented in Supplementary Fig. 1, Additional file 1. For the validation study, patients with CRC from Chonnam National University Hospital (CNNUH), Korea, were utilized as described in detail elsewhere [15]. These patients were diagnosed with CRC between 2004 and 2014 and were prospectively followed up until December 31, 2020. They were divided into two groups (MEGA array and ONCO array), with each group having 3,465 and 3,162 patients, respectively. Further details regarding the validation study population are provided in Supplementary Table 1, Additional file 2.

Overall and progression-free survival

The definitions of OS and PFS were used to assess the clinical outcomes in patients with CRC. Overall survival was defined as the time elapsed from surgical treatment to death (event) or the end of the follow-up period if the patient was alive. In the SNUH patient group, death certificate data provided by the Korean Statistical Information Service [16] were linked to the patient with CRC data. Furthermore, the cause and date of death in the CNNUH patient group provided by the National Statistical Office were merged with the patient with CRC data. Progression-free survival was defined as the time elapsed from surgical treatment to disease progression (event, death, or relapse), or the end of the follow-up period if the patients did not have any progression events.

Gene selection

We utilized three biological databases to obtain immune-related gene data: ENSEMBL [17], Kyoto Encyclopedia

of Genes and Genomes (KEGG) [18], and Gene Ontology (GO) [19]. In ENSEMBL, we searched for immune-related genes using the keyword "immune" and obtained the resulting data in XML format. In KEGG, we selected all pathways related to the "5.1 Immune System" category. In GO, we selected the "Immune System Process" category through the AmiGO2 search browser and obtained result genes. Through integration of genes from three databases, we identified 3,523 non-redundant immune-related genes. (Supplementary Fig. 3, Additional file 1)

Genotyping, SNP selection, and quality controls

Single nucleotide polymorphism data were obtained using the Axiom Korean chip 1.1 [20] with 960 blood samples. Sample quality control and genotype calling were performed using Affymetrix Power Tools (APT) and the AxiomGT1 BRLMM-P Algorithm. As a result, we obtained 800,493 genome-wide SNPs Furthermore, we selected immune-related SNPs within ± 200 kb from the position of 3,523 immune-related genes using the bioMart [21] and SnpMatrix R package [22]. Finally, we assembled 63,459 SNPs within or near 3,017 immune-related genes.

We applied four exclusion criteria that were not mutually exclusive to the flowchart of the study design (Supplementary Fig. 3): (1) minor allele frequency (MAF) < 0.01 (N = 23,049), (2) Hardy–Weinberg equilibrium (HWE) p-value < 1×10^{-6} (N = 2,493), (3) mono-allelic (N = 10,823), and (4) call rate < 98% (N = 5,435). We used the criteria from previous quality control (QC) procedures [23] but there were no additional excluded SNPs from this procedure. In total, 37,398 genotyped SNPs were identified. Additionally, genotype imputation was performed using the Michigan Imputation Server [24]. We used MINIMAC 4 and selected 1000 genomes [25] as the reference panel from East Asia for the subpopulation. For QC, we selected an r-squared statistic (rsq) filter of 0.3, which removed over 70% of poorly imputed SNPs. Subsequently, we implemented standard settings of QC by the Michigan Imputation Server, such as minor-allele frequency < 1%, call rate > 90%, and removal of duplicate SNPs. Finally, 2,514 SNPs were excluded from the imputed dataset for two reasons (invalid alleles, 942 SNPs; allele mismatch, 1,503 SNPs), leaving 445,947 imputed SNPs. Accordingly, 450,667 genotyped and imputed SNPs within or near 3,017 immune-related genes were selected.

Single nucleotide polymorphism association analysis

We utilized the "gwasurvivr" R package [26] to identify statistically significant SNPs associated with time-to-clinical outcomes (OS and PFS) based on the Cox proportional hazards model ("plinkCoxSurv" function) with selected six covariates (sex, age, BMI, tumor location, histologic grade, and TNM stage), as described in the previous section. For the SNP association analysis, we set the cut-off values for the p-values at 1.0×10^{-4} (extended suggestive p-value), 1.0×10^{-5} (suggestive p-value), and 1.34×10^{-6} (Bonferroni-corrected p-value calculated from 0.05 divided by 37,398, total number of genotyped immune-related SNPs). To describe independent index SNPs among linkage disequilibrium (LD), we performed a pairwise LD analysis between two SNPs and calculated the r^2 values using the "LDmatrix" function in the "LDlink" web-based service [27]. We chose the SNP with the smallest p-value among the SNPs with an r^2 value equal to or greater than 0.6. To visualize the results of previous steps, we used the "CMplot" R package [28], which generated Manhattan plots of significant SNPs linked to clinical outcomes and differentiated them by distinctive colors.

Furthermore, significant SNPs from the discovery study results were used for the validation study with the same covariates as in the discovery study. The p-value threshold for the validation study was 0.05. The validation study population did not have relapse data for the PFS outcome; therefore, it was replaced with the OS outcome. We performed meta-analysis utilizing "meta" R package [29] and extracted heterogeneous p-values and I^2 values to combine the SNP effects on survival of discovery and validation studies. We selected random-effect p-values for SNPs that showed heterogeneous p-values > 0.05, and fixed-effect p-values for SNPs that did not show heterogeneous p-values > 0.05.

Polygenic risk score analysis

We followed the standard protocol outlined by Choi et al. and the clinical applications described by Lewis et al. to calculate the PRS of the study participants (Supplementary Fig. 4, Additional file 1) [30, 31]. Briefly, the effect size of each SNP and allelic dosage of the effect allele were multiplied and summed individually. For normalization, it was divided by the number of non-missing SNPs in each individual multiplied by the ploidy, which is two in humans. We used z-scores as the effect size of each SNP since we used a Cox proportional hazards model. We analyzed the PRS tertile groups stratified by the lowest, middle, and highest levels to evaluate the clinical significance of PRS in patients with CRC. Each group was designated as "Low," "High," and "Middle." We utilized the "survminer" and "survival" R packages to demonstrate the distinctive survival outcomes by PRS [32, 33]. The "coxph" function was used to perform PRS survival analysis with the selected six covariates (sex, age, BMI, tumor location, histologic grade, and TNM stage) from the previous SNP analysis. The extended suggestive p-value of 1.0×10^{-4} was used for PRS analysis. Subsequently, we performed pathway enrichment analysis using the

Reactome Knowledgebase [34] on gene sets matched to selected SNPs. Statistically significant pathways were identified using a Bonferroni-corrected p-value threshold, calculated as 0.05 divided by the total number of pathways identified in each analysis.

Gene-gene network analysis

We utilized GeneMANIA within Cytoscape software 3.10.0 to analyze and visualize the gene-gene interaction network of selected genes matched to suggestive SNPs [35, 36]. All annotation databases (co-expression, co-localization, genetic interactions, pathways, physical interactions, predicted, and shared protein domains) were selected. The tables of edges (connected nodes) and nodes (gene names) were extracted. Finally, the extended suggestive p-value of 1.0×10^{-4} was used for gene-gene interaction network analysis.

Statistical analysis

Confounding factors

We selected six variables–sex, age, body mass index (BMI), tumor location, histological grade, and Tumor Nodes Metastasis (TNM) stage–as potential confounding factors. Specifically, age, histological grade, and tumor location were statistically significant. Although sex, BMI, and age were not statistically significant within the discovery study population, extensive research has demonstrated their association with poor CRC outcomes. Therefore, we additionally selected sex, BMI and age in our analysis despite the lack of statistical significance in the study population. However, we did not select "adjuvant chemotherapy" because this factor was statistically associated with and predominantly determined by the TNM stage. This relationship between adjuvant chemotherapy and TNM stage was revealed through Pearson's correlation test with a correlation coefficient of 0.514 and p-value less than 2.2×10^{-16}, which indicates a statistically significant positive correlation. We subsequently conducted a multicollinearity test using the variance inflation factor (VIF) method (Supplementary Fig. 2, Additional file 1). The VIF values for each confounding factor ranged from 1.002695 to 1.350784, indicating that the selected confounding factors were independent [37].

Results

Study population characteristics

The characteristics of the discovery study population are described in Table 1. The non-event group in OS had a mean age of 61.9 years with a standard deviation of 10.5 years, whereas the event group had a mean age of 64.4 years with a standard deviation of 12.4 years. In PFS, the non-event group had a mean age of 62.0 years with a standard deviation of 10.7 years, and the event group had a mean age of 63.8 years with a standard deviation of 11.7 years. Regarding sex category, the non-event group in OS had 57.5% of males and 42.5% of females, while the event group had 61.2% of males and 38.8% of females. In PFS, the non-event group had 56.6% of males and 43.4% of females, while the event group had 62.1% of males and 37.9% of females. Details of the remaining clinical variables (BMI, tumor location, and so on) are represented in Table 1.

The median OS for the non-event group was 84 months, with a minimum and maximum of 3 and 96 months, respectively. For the event group, the median OS was 31 months, with a minimum and maximum of 1 and 93 months, respectively. The median PFS in the non-progression group was 71 months, with minimum and maximum values of 1 and 96 months, respectively. In the progression group, the median PFS was 14 months, with minimum and maximum PFS of 1 and 93 months, respectively.

Single nucleotide polymorphism association on overall and progression-free survival

Associations between immune-related SNPs and survival are represented as Manhattan plots in Fig. 1. The cluster of SNPs from all stages of OS highlights important positions that have risen above the suggestive p-value (1.0×10^{-5}) (Fig. 1A). One SNP (rs75663950) on chromosome 9 was in LD with rs1414944 (Fig. 1A). In this analysis, four SNPs (rs1414944 (*PIP5K1B*, 9q21.11, hazard ratio (HR) = 1.5, p-value = 3.39×10^{-6}), rs34009001 (*ADCY7*, 16q12.1, HR = 3.28, p-value = 4.38×10^{-6}), rs2967871 (*COTL1*, 16q24.1, HR = 1.58, p-value = 6.07×10^{-6}), and rs1082967 (*PRKCI*, 3q26.2, HR = 1.57, p-value = 7.86×10^{-6}) represented p-values lower than the suggestive p-value (Table 2). Although none of these SNPs reached a Bonferroni-corrected p-value of 1.34×10^{-6}, they suggested an association with OS in CRC.

Furthermore, we conducted a stratification analysis of stages 1–3 in OS. The cluster of SNPs from stages 1–3 highlights important positions that have risen above the suggestive p-value (1.00×10^{-5}) (Fig. 1B). No SNPs were observed in LD above the suggested p-value threshold in the stratification analysis by stage in contrast to the analysis of all stages based on OS. Six SNPs had p-values lower than the suggestive p-value: rs117322760 (*PKHD1L1*, 8q23.1, HR = 4.58, p-value = 1.40×10^{-6}), rs117802394 (*SEMA3C*, 7q21.11, HR = 6.29, p-value = 1.47×10^{-6}), rs79798148 (*IL19*, 1q32.1, HR = 2.5, p-value = 3.67×10^{-6}), rs142842205 (*RNF216*, 7p22.1, HR = 4.29, p-value = 5.74×10^{-6}), rs147127574 (*STAB2*, 12q23.3, HR = 3.82, p-value = 7.82×10^{-6}), and rs74996929 (*MLLT3*, 9p21.3, HR = 2.91, p-value = 7.90×10^{-6}) (Table 2). None of these SNPs reached a Bonferroni-corrected p-value of 1.34×10^{-6} from all stages.

Table 1 Characteristics of participants included in the discovery study

Characteristics	OS			PFS		
	Non-event (*N* = 718)	Event (*N* = 242)	*p*-value[a]	Non-event (*N* = 643)	Event (*N* = 317)	*p*-value[a]
Follow-up time (months) Median [range]	84 [3–96]	31 [1–93]		71 [1–96]	14 [1–93]	
Age (year)			0.006			0.02
Mean (SD)	61.9 (10.5)	64.4 (12.4)		62.0 (10.7)	63.8 (11.7)	
Sex			0.359			0.117
Male	413 (57.5%)	148 (61.2%)		364 (56.6%)	197 (62.1%)	
Female	305 (42.5%)	94 (38.8%)		279 (43.4%)	120 (37.9%)	
BMI (kg/m^2)			0.158			0.221
Underweight (BMI < 18.0)	30 (4.2%)	14 (5.8%)		25 (3.9%)	19 (6.0%)	
Normal (18.0 ≤ BMI < 23.0)	259 (36.1%)	101 (41.7%)		233 (36.2%)	127 (40.1%)	
Overweight (23.0 ≤ BMI < 25.0)	174 (24.2%)	58 (24.0%)		158 (24.6%)	74 (23.3%)	
Obese (25.0 ≤ BMI)	255 (35.5%)	69 (28.5%)		227 (35.3%)	97 (30.6%)	
Synchronous other cancer			1			0.96
Positive	703 (97.9%)	237 (97.9%)		629 (97.8%)	311 (98.1%)	
Negative	15 (2.1%)	5 (2.1%)		14 (2.2%)	6 (1.9%)	
Tumor location			0.02			0.167
Colon cancer	500 (69.6%)	154 (63.6%)		449 (69.8%)	205 (64.7%)	
Rectal cancer	199 (27.7%)	73 (30.2%)		175 (27.2%)	97 (30.6%)	
Multiple cancer	19 (2.6%)	15 (6.2%)		19 (3.0%)	15 (4.7%)	
Histologic grade			< 0.001			< 0.001
High	44 (6.1%)	37 (15.3%)		40 (6.2%)	41 (12.9%)	
Low	625 (87.0%)	195 (80.6%)		554 (86.2%)	266 (83.9%)	
Unknown	49 (6.8%)	10 (4.1%)		49 (7.6%)	10 (3.2%)	
TNM Stage			< 0.001			< 0.001
0	24 (3.3%)	1 (0.4%)		24 (3.7%)	1 (0.3%)	
1	162 (22.6%)	13 (5.4%)		154 (24.0%)	21 (6.6%)	
2	222 (30.9%)	36 (14.9%)		207 (32.2%)	51 (16.1%)	
3	267 (37.2%)	70 (28.9%)		235 (36.5%)	102 (32.2%)	
4	41 (5.7%)	121 (50.0%)		21 (3.3%)	141 (44.5%)	
Unknown	2 (0.3%)	1 (0.4%)		2 (0.3%)	1 (0.3%)	
Adjuvant chemotherapy			0.1			< 0.001
Positive	470 (65.5%)	173 (71.5%)		405 (63.0%)	238 (75.1%)	
Negative	248 (34.5%)	69 (28.5%)		238 (37.0%)	79 (24.9%)	

Abbreviations: OS (Overall survival), PFS (Progression-free survival), BMI (Body mass index), TNM Stage (Tumor Node Metastasis stage), SD (Standard deviation)

[a]Continuous variables are performed by linear-rank test and categorical variables are performed by log-rank test

Information regarding all SNPs above the suggestive *p*-value (1.00×10^{-5}) threshold is discussed in Supplementary Table 2, Additional file 2.

For the meta-analysis of all stages of OS, only SNP (rs34009001) was selected from the validation study because it achieved a *p*-value below 0.05. This SNP showed an I^2 value of 87% and a heterogeneous *p*-value of 5.17×10^{-3} and its fixed-effect *p*-value was 6.19×10^{-6}. Similarly, only one SNP (rs1111691) was selected from the validation study in the meta-analysis of stages 1–3 in OS, and it showed an I^2 value of 87%, a heterogeneous *p*-value of 5.62×10^{-3}, and a fixed-effect *p*-value of 2.82×10^{-4}. Although the validation study revealed high heterogeneity among the discovery and validation studies, the fixed effects *p*-values were < 0.05.

The cluster of SNPs from all stages in PFS highlights important positions that have risen above the suggestive *p*-value (1.00×10^{-5}) (Fig. 1C). None of the SNPs were in LD above the suggested *p*-value (Fig. 1C). In this analysis, three SNPs (rs76181827 (*ARHGEF12*, 11q23.3, HR = 2.77, *p*-value = 1.03×10^{-6}), rs896531 (*STIM1*, 11p15.4, HR = 2.03, *p*-value = 4.99×10^{-6}), and rs116332704 (*CTNNA2*, 2p12, HR = 3.83, *p*-value = 9.16×10^{-6}) had *p*-values lower than the

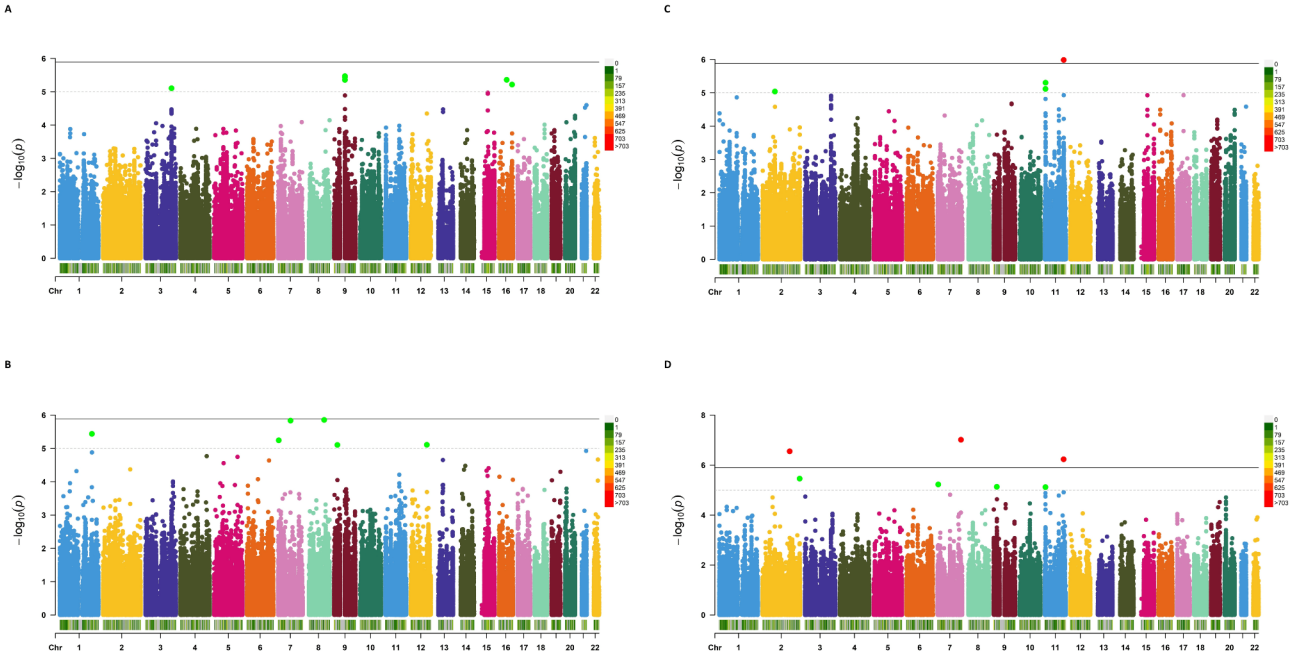

Fig. 1 Manhattan plots of immune-related SNP associations with survival. (**A**) Overall survival for all stages. (**B**) Overall survival for stages 1–3. (**C**) Progression-free survival for all stages. (**D**) Progression-free survival for stage 1–3

Each dot represents the test results for a single SNP. The x-axis represents the genomic location along each chromosome and the y-axis represents the log10 *p*-value. The solid horizontal line and the dotted horizontal line correspond to a *p*-value of 1.34×10^{-6} and 1.00×10^{-5}, respectively

suggested *p*-value (Table 2). In contrast to previous results for OS, one SNP (rs76181827) exceeded the Bonferroni-corrected *p*-value of 1.34×10^{-6}. Additionally, we conducted a stratification analysis of stage 1–3 of PFS. The cluster of SNPs highlights important positions that have risen above the suggestive *p*-value (1.00×10^{-5}) (Fig. 1D). None of the SNPs were in LD above the suggested *p*-value (Fig. 1D). Three SNPs had *p*-values lower than the suggestive *p*-value: rs143531681 (*NOS3*, 7q36.1, HR = 4.67, *p*-value = 9.72×10^{-8}), rs145049833 (*ATF2*, 2q31.1, HR = 4.04, *p*-value = 2.83×10^{-7}), and rs76181827 (*ARHGEF12*, 11q23.3, HR = 3.36, *p*-value = 5.90×10^{-7}) (Table 2). All three SNPs exceeded the Bonferroni-corrected *p*-values of 1.34×10^{-6}.

For the meta-analysis of all stages of PFS, none of the SNPs were selected from the validation study because none of the significant SNPs in the discovery study achieved a *p*-value lower than 0.05 in the validation study. However, two SNPs (rs11906029 and rs4616886) were selected from the validation study for the meta-analysis of stages 1–3 of PFS. They showed I^2 values of 91% and 90%, respectively, heterogeneous *p*-values of 8.58×10^{-4} and 1.76×10^{-3}, respectively, and their fixed-effect *p*-values were 2.58×10^{-3} and 1.32×10^{-3}, respectively. The fixed effects *p*-values were <0.05 although the validation study revealed high heterogeneity among the discovery and validation studies.

Polygenic risk score association on overall and progression-free survival

Initially, the PRS for the non-event and event groups was the same at 0.051 in the analysis of all stages of OS. The PRS distribution overlapped in the density and box plots (Fig. 2A). Second, the PRS for the non-event and event groups were 0.020 and 0.421, respectively in the analysis of stage 1–3 OS. This result was significantly different from the analysis of all stages (Fig. 2B). The PRS was not feasible as a biomarker for analyzing all stages of OS according to the survival plot and HR (Fig. 2A). This contrasted with the analysis of OS at stages 1–3, which significantly differed (Fig. 2B). The difference in the survival curve plot is consistent with the HR for both groups. The HR for high and middle-grade subgroups of all stages were 1.07 (*p*-value 0.7) and 1.03 (*p*-value, 0.9), respectively with the low-grade subgroup assigned as a reference. In contrast, the HR for subgroups of stages 1–3 was 59.58 and 9.69, respectively (*p*-values <0.001 for both) using the low-grade subgroup as the reference. Additionally, the concordance index (C-index) of all stages of OS with PRS was 0.78, which is equal to the value obtained without PRS. Moreover, the C-index of stages 1–3 of OS with PRS was 0.77, slightly higher than 0.76 observed from stages 1–3 of OS without PRS. In the analysis of all stages of PFS, the PRS for the non-event and event groups were 0.078 and 0.082, respectively. The PRS distribution overlapped in the density and box plots (Fig. 2C). The PRS for the non-event and event groups were 0.032

Table 2 Top SNPs associated with survival among CRC patients with suggestive statistical significance

OS												
Category	**SNP**	**Chr**	**Pos**	**Reference allele**	**Alternate allele**	**Band**	**Matched gene**	**Hazard ratio**	***p*-value**	***p*-value (fixed)**	***p*-value (hetero)**	**I^2 (%)**
All stages	rs1414944	9	68,841,464	G	A	q21.11	*PIP5K1B*	1.5	3.39E-06			
	rs34009001	16	50,287,266	C	T	q12.1	*ADCY7*	3.28	4.38E-06	6.19E-06	5.17E-03	87.21
	rs2967871	16	84,614,385	C	T	q24.1	*COTL1*	1.58	6.07E-06			
	rs1082967	3	170,230,623	C	T	q26.2	*PRKCI*	1.57	7.86E-06			
Stage 1–3	rs117322760	8	109,495,997	G	C, T	q23.1	*PKHD1L1*	4.58	1.40E-06			
	rs117802394	7	80,873,623	T	C	q21.11	*SEMA3C*	6.29	1.47E-06			
	rs79798148	1	206,824,216	G	A	q32.1	*IL19*	2.5	3.67E-06			
	rs142842205	7	5,620,896	C	G	p22.1	*RNF216*	4.29	5.74E-06			
	rs147127574	12	103,667,078	A	G	q23.3	*STAB2*	3.82	7.82E-06			
	rs74996929	9	20,503,822	C	A	p21.3	*MLLT3*	2.91	7.90E-06			
PFS												
Category	SNP	Chr	Pos	Reference allele	Alternate allele	Band	Matched gene	Hazard ratio	*p*-value	*p*-value (fixed)	*p*-value (hetero)	I^2 (%)
All stages	rs76181827	11	120,404,330	C	G	q23.3	*ARHGEF12*	2.77	1.03E-06			
	rs896531	11	3,902,679	A	G	p15.4	*STIM1*	2.03	4.99E-06			
	rs116332704	2	80,124,577	G	A, C,T	p12	*CTNNA2*	3.83	9.16E-06			
Stage 1–3	rs143531681	7	151,004,636	T	C	q36.1	*NOS3*	4.67	9.72E-08			
	rs145049833	2	175,145,457	T	C	q31.1	*ATF2*	4.04	2.83E-07			
	rs76181827	11	120,404,330	C	G	q23.3	*ARHGEF12*	3.36	5.90E-07			

Abbreviations: SNP, single nucleotide polymorphism; Chr, chromosome; Pos, position; fixed, fixed-effect; hetero, heterogenous; All stages, the discovery data set of 960 CRC patients with all stages in OS or PFS; Stage 1–3, the discovery data set of 795 CRC patients with stage 1–3 in OS or PFS;

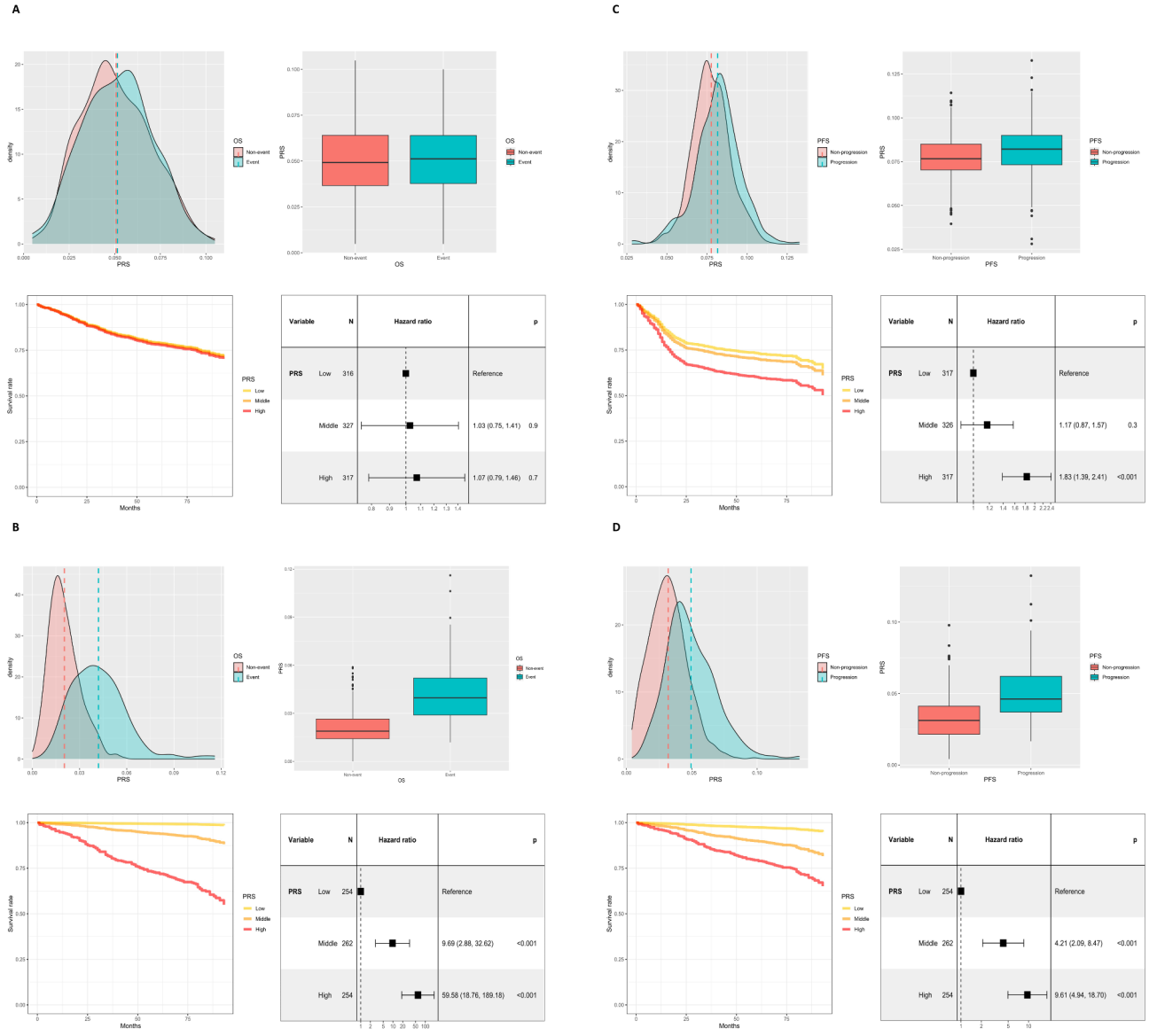

Fig. 2 Polygenic risk score (PRS) distribution and associations with survival. (**A**) Overall survival for all stages. (**B**) Overall survival for stages 1–3. (**C**) Progression-free survival for all stages. (**D**) Progression-free survival for stage 1–3

The PRS value was used as a continuous variable for distribution and box plots; its grade was used as a categorical variable for the survival curve and forest plot

and 0.050 in the analysis of stages 1–3 of PFS. The results of this analysis significantly differed from those of the analysis at all the stages (Fig. 2D). Only the high PRS differed between the middle- and low-grade groups according to the survival plot and HR (Fig. 2C). According to the results for stages 1–3, PRS is a biomarker with significantly distinct survival curves across each group. The differences in the survival curve plot were consistent with the HR results for both groups. The HR for high- and middle-grade subgroups of all stages were 1.83 and 1.17, as the low-grade subgroup was assigned to a reference. Additionally, the *p*-value for the middle-grade subgroup was 0.3 and the *p*-value for the high-grade subgroup was <0.001. The results for stages 1–3 were slightly different from the previous results for all stages. The HR for stages 1–3 were 9.61 and 4.21, as the low-grade subgroup was assigned to a reference. The *p*-values for this outcome were <0.001 for both groups. Furthermore, the C-index of all stages of PFS with PRS was 0.82, higher than the 0.70 observed from all stages of PFS without PRS. Lastly, the C-index of stages 1–3 of PFS with PRS was 0.78, higher than the corresponding value of 0.67 without PRS.

Gene-gene interaction network on overall and progression-free survival

The gene-gene interaction network from all stages of OS is depicted in Fig. 3A and the information for both the top edge and nodes is discussed in Table 3. Orange-colored nodes are query genes matched to significant SNPs and light-yellow nodes are predicted genes from GeneMANIA. Each node had 45.95% and 54.05%. For each gene, the individual scores ranged from 0.006 to 0.039 for the predicted genes, and 0.497–0.931 for the query genes. Additionally, there were four categories of edges (co-expression, genetic interactions, pathways, and physical interactions), with the proportions of edges equal to 10.10%, 57.58%, 26.26%, and 6.06%, respectively. The edges showed normalized max weights from 1.09×10^{-6} to 0.087 and "Pathway" usually showed larger weights than other types of edges (Supplementary Table 3, Additional file 2). *DTX1* (predicted gene) and *ADAM10* (query gene) had the largest weights.

The gene-gene interaction network from stages 1–3 in the OS is depicted in Fig. 3B and detailed information for both the top edges and nodes is discussed in Table 3. Query genes accounted for 59.18%, and predicted genes accounted for 40.82%. For each gene, the individual scores ranged from 0.009 to 0.041 for the predicted genes and from 0.469 to 0.991 for the query genes.

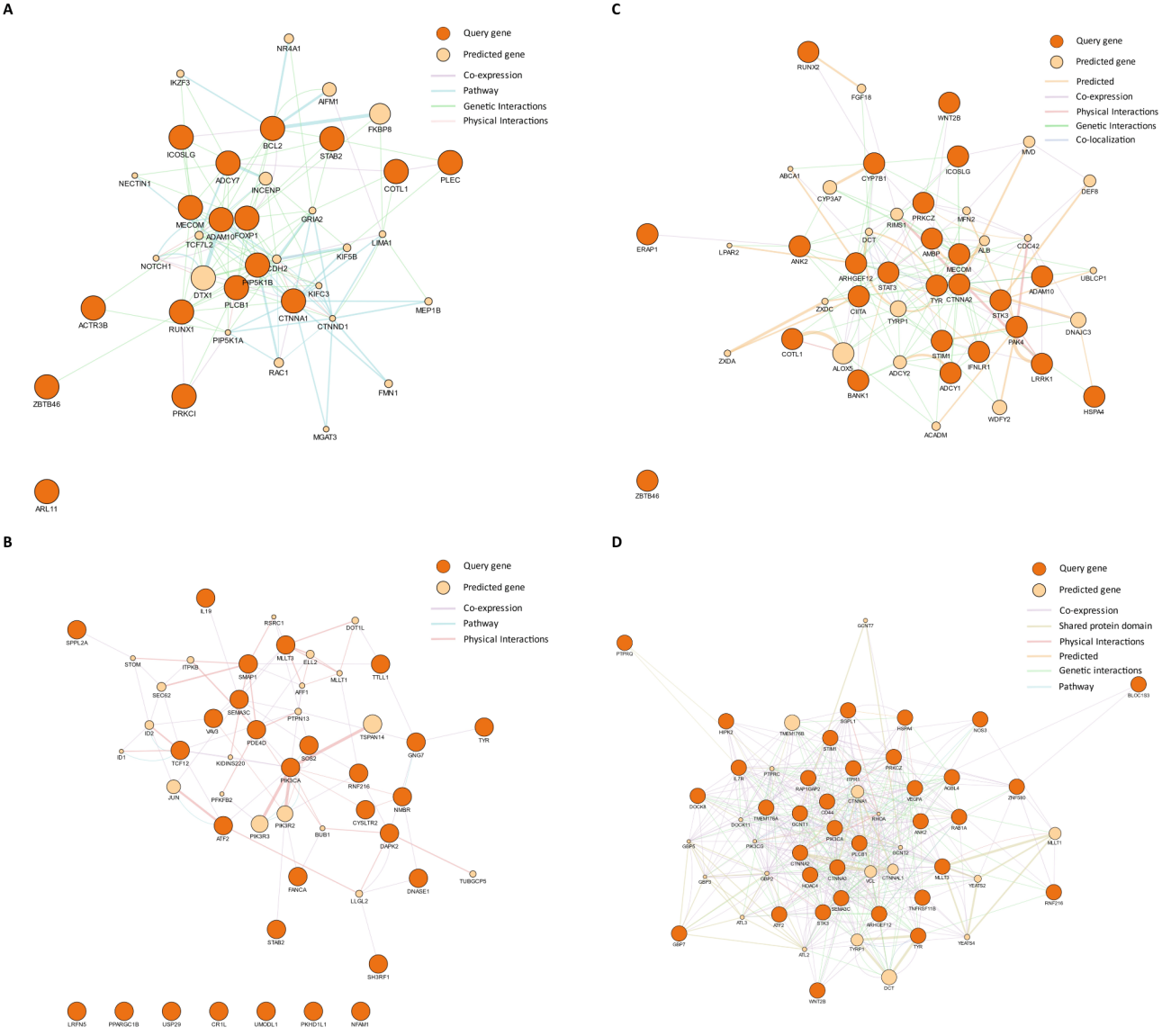

Fig. 3 Gene-gene interaction networks on survival. (**A**) Overall survival for all stages. (**B**) Overall survival for stages 1–3. (**C**) Progression-free survival for all stages. (**D**) Progression-free survival for stage 1–3

Orange nodes represent the query genes. Light-yellow nodes indicate predicted genes. Each edge has a different value for the relationship between two nodes (genes) and is distinctively colored. The thickness of each edge indicated its weight

Table 3 Top edge and node in gene-gene interaction network

OS	Top edge			Top node			
	Data type	**Name**	**Normalized max weight**	**Gene**	**Name**	**Score**	**Category**
All stages	Pathway	H__sapiens__1_-779,769| H__sapiens__1_-780,309| Pathway	0.087	*DTX1*	H__sapiens__1_-779,769	0.039	predicted
				ADAM10	H__sapiens__1_-780,309	0.749	query
Stage 1–3	Physical Interactions	H__sapiens__2_-775,897| H__sapiens__2_-777,953| Physical Interactions	0.123	*PIK3R2*	H__sapiens__2_-775,897	0.035	predicted
				PIK3CA	H__sapiens__2_-777,953	0.469	query
PFS	Top edge			Top node			
	Data type	Name	Normalized max weight	Gene	Name	Score	Category
All stages	Predicted	H__sapiens__3_-773,060| H__sapiens__3_-775,518| Predicted	0.139	*ALOX5*	H__sapiens__3_-773,060	0.089	predicted
				COTL1	H__sapiens__3_-775,518	0.761	query
Stage 1–3	Shared protein domains	H__sapiens__5_-774,226| H__sapiens__5_-774,129| Shared protein domains	0.011	*DCT*	H__sapiens__3_-774,226	0.038	predicted
				TYR	H__sapiens__3_-774,129	0.751	query

Abbreviations: All stages, the discovery data set of 960 CRC patients with all stages in OS or PFS; Stage 1–3, the discovery data set of 795 CRC patients with stage 1–3 in OS or PFS;

In addition to genes, there were three categories of edges (co-expression, pathway, and physical interactions), and the proportions of edges were 53.26%, 37.68%, and 8.70%, respectively. The edges showed normalized maximum weights from 8.00×10^{-6} to 0.123, and physical Interactions usually showed larger weights than other types of edges (Supplementary Table 3, Additional file 2). *PIK3R2* (predicted gene) and *PIK3CA* (query gene) had the largest weights.

The gene-gene interaction network from all stages in PFS is depicted in Fig. 3C and the information for both the top edge and nodes is discussed in Table 3. Query genes accounted for 55.56%, and predicted genes accounted for 44.44%. For each gene, the individual scores ranged from 0.030 to 0.089 for the predicted genes and 0.473 to 0.968 for the query genes. Additionally, there were five categories of edges (co-expression, genetic Interactions, physical interactions, predicted, and co-localization), and the proportions of edges were 25.41%, 49.18%, 4.92%, 15.57%, and 4.92%, respectively. Furthermore, the edges showed normalized max weights from 2.36×10^{-6} to 0.139, and "predicted" usually showed larger weights than other types of edges (Supplementary Table 3, Additional file 2). *ALOX5* (predicted gene) and *COTL1* (query gene) had the greatest weight.

The gene-gene interaction network from stages 1–3 in PFS is depicted in Fig. 3D and detailed information for the top edge and nodes is discussed in Table 3. Query genes and predicted genes accounted for 63.64% and 36.36%, respectively. For each gene, the individual scores ranged from 0.003 to 0.009 for the predicted genes and from 0.482 to 0.964 for the query genes. Seven categories of edges were observed (co-expression, genetic interactions, pathways, physical interactions, shared protein domains, co-localization, and predicted), and the proportions of edges were 60.40%, 20.23%, 3.70%, 5.13%, 8.55%, 1.71%, and 0.28%, respectively. The edges showed normalized max weights from 1.71×10^{-7} to 0.012, and the shared protein domains usually showed larger weights than other types of edges (Supplementary Table 3, Additional file 2). *DCT* (predicted gene) and *TYR* (query gene) had the greatest weight.

Pathway enrichment analysis of PRS gene sets on overall and progression-free survival

The pathway enrichment analysis results from all stages and stages 1–3 of OS, PFS is discussed in Supplementary Table 4. From the results for all stages in OS, we identified 'RUNX3 regulates RUNX1-mediated transcription' with *RUNX1* and 'BH3-only proteins associate with and inactivate anti-apoptotic BCL-2 members' with *BCL2*. Each pathway showed *p*-values of 1.35×10^{-5} and 1.01×10^{-4} respectively. Additionally, for stage 1–3 in OS, none of the pathway results were statistically significant. Furthermore, from the results for all stages in PFS, we identified 'RUNX2 regulates genes involved in differentiation of myeloid cells' and 'RUNX2 regulates chondrocyte maturation' with *RUNX2*, with *p*-values of 7.58×10^{-5} and 1.03×10^{-4}. Lastly, from the results for stage 1–3 in PFS, we identified 'Insulin-like Growth Factor-2 mRNA Binding Proteins (IGF2BPs/IMPs/VICKZs) bind RNA' with *CD44*, 'Signaling by VEGF' and 'VEGFA-VEGFR2 Pathway' with four genes (*PIK3CA*, *NOS3*, *ITPR1*, *PRKCZ*, *VEGFA*), and 'VEGFR2 mediated cell proliferation' with three genes (*ITPR1*, *PRKCZ*, *VEGFA*). Each pathway presented p-values of 7.04×10^{-10}, 4.93×10^{-6}, 5.18×10^{-5} and 1.31×10^{-4} respectively.

Discussion

This study investigated the association between SNPs in immune-related genes and the clinical outcomes of patients with CRC in South Korea. We identified 15 independent immune-related SNPs that were significantly associated with OS and PFS. These findings indicate the potential use of novel biomarkers for CRC prognosis. In the analysis of OS, four SNPs (rs1414944, rs34009001, rs2967871, and rs1082967) were associated with outcomes in all stages, and six SNPs (rs117322760, rs117802394, rs79798148, rs142842205, rs147127574, and rs74996929) in stage 1–3. In the PFS analysis, we observed that three SNPs (rs76181827, rs896531, and rs116332704) were associated with outcome in all stages and three SNPs were associated in stages 1–3 (rs143531681, rs145049833, and rs76181827). One of the significant differences in the OS results was that several SNPs exceeded the Bonferroni-corrected *p*-value threshold (rs76181827 in all stages and rs143531681, rs145049833, and rs76181827 in stage 1–3).

In the PRS analysis of OS, higher PRS grades were associated with poorer survival outcomes at all stages and stages 1–3. Moreover, we uncovered the potential for a new interpretation of elements that were deemed less significant in the individual SNP analysis by conducting a gene-gene interaction network analysis. The PRS analysis of PFS showed a similar trend. A higher PRS grade is associated with poorer progression outcomes at all stages and stage 1–3. However, this differed from the OS results. As more SNPs showed significance in the PFS analysis, the role of PRS as a biomarker appeared to be better in PFS than in OS. Although none of the SNPs exceeded the Bonferroni-corrected *p*-value threshold, we performed additional analysis with extended suggestive SNPs for PRS and gene-gene interaction networks.

In the gene-gene interaction network of OS at all stages, the pathway edge between *DTX1* and *ADAM10* had the largest weight. A correlation exists between both genes in the Notch signaling pathway [38], which is commonly mutated in various types of cancer. Although no direct connection between *ADAM10* and *DTX1* has been identified, individual studies on both genes have revealed their correlation with CRC survival. Several studies on *ADAM10* and CRC show that higher *ADAM10* expression is associated with higher tumor stage or poorer survival outcomes (metastasis) in CRC [39, 40]. Increased *DTX1* expression is associated with a poor prognosis for CRC [41]. Additionally, *PIK3R2* and *PIK3CA* held the highest weights for physical interactions in the OS of stages 1–3. Both genes encode the PI3K familial proteins, which have been extensively studied for anticancer therapy [42]. Both genes have been studied for their association with CRC. Differential expression analysis of the miRNA profile showed that *PIK3R2* is significantly enriched in the CRC pathway [43]. For *PIK3CA*, mutations in the *PIK3CA* gene can lead to increased *PI3K* activity, which promotes cell growth, survival, and proliferation in CRC [44].

The predicted edge between *ALOX5* and *COTL1* had the largest weight in the PFS for all stages. Both genes are involved in the association in the lipoxygenase pathway, which is significant in cancer progression and metastasis [45]. MiR-216a-3p inhibits CRC cell proliferation and targeted *ALOX5* [46]. *COTL1* has not been reported to cause CRC. However, the differential expression level of *COTL1* correlates with the metastatic features of non-small cell lung cancer (NSCLC) [47]. Further research on *COTL1* may reveal its correlation with survival outcomes considering that *ALOX5* strongly correlates with CRC. Moreover, the shared protein domain edge between *DCT* and *TYR* showed the greatest weight in the PFS of stages 1–3. Dopachrome tautomerase (DCT) correlates with CRC since it promotes COX-2 expression through β-catenin mediated mechanisms and these two factors are critical contributors to CRC [48]. Although TYR is unreported in CRC, tyrosinase suppresses vasculogenic mimicry in human melanoma cells [49]. An alteration of rs149180340 matched to TYR could be interpreted as a risk factor for CRC because this result suggests the anti-oncogenic effects of *TYR*.

Moreover, it is essential to acknowledge the strengths and limitations of this study. First, we identified novel SNPs that were not previously reported to be associated with CRC prognosis. These SNPs were associated with OS and PFS and would serve as potential biomarkers for CRC prognosis and personalized medicine. Second, we discovered a significant implication of the PRS as a biomarker of CRC prognosis. Previous studies have used the PRS in CRC case-control studies and normal populations. However, our study utilized PRS to analyze OS and PFS in CRC, providing a comprehensive approach to clinical outcomes. Third, we performed a network-based analysis of post-GWAS data that allowed us to uncover novel gene-gene interactions. This provides a deeper understanding of genetic interactions in CRC.

However, this study has several limitations. First, the most significant limitation was its small sample size. This limitation may have contributed to the failure of the validation study, the large I^2 values, and the low statistical significance. Further studies with larger populations are needed to confirm our findings and to provide a more robust understanding of the association between SNPs matched to immune-related genes and CRC prognosis. Second, we could not retrieve *p*-values exceeding the Bonferroni-corrected threshold for certain SNPs. This limitation suggests that the association of these SNPs with CRC prognosis may not be as strong as initially planned.

Conclusions

In summary, we identified novel predictive biomarker genetic variants of *PKHD1L1* and *NOS3* associated with OS and PFS, respectively, among patients with CRC. Additionally, we identified PRS as a potential biomarker for survival outcomes and suggested that a gene-gene interaction network is involved in CRC progression. Future studies including larger and more racially and ethnically diverse populations and integrating multi-omics data (such as RNA-seq and metabolomics profiling) will promote a deeper understanding of CRC survival and the development of immune-related therapeutic interventions for patients with CRC.

Abbreviations

APT	Affymetrix Power Tools
BMI	Body mass index
CRC	Colorectal cancer
DCT	Dopachrome tautomerase
GO	Gene Ontology
GWAS	Genome-wide association study
HR	Hazard ratio
HWE	Hardy–Weinberg equilibrium
ICI	Immune checkpoint inhibitor
KEGG	Kyoto Encyclopedia of Genes and Genomes
KOGES	Korean Genome and Epidemiology Study
LD	Linkage disequilibrium
MAF	Minor allele frequency
NSCLC	Non-small cell lung cancer
OS	Overall survival
PFS	Progression-free survival
PRS	Polygenic risk score
QC	Quality control
SNP	Single-nucleotide polymorphism
SNUH	Seoul National University Hospital
TNM	Tumor Nodes Metastasis
VIF	Variance inflation factor

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-13819-4.

Additional file 1 included supplementary figures. Supplementary Fig. 1. Flowchart of the study population selection. Supplementary Fig. 2. Variance inflation factor estimation to detect multicollinearity of potential confounding factors. Supplementary Fig. 3. Flowchart of gene and SNP selection. Supplementary Fig. 4. Formula for calculating polygenic risk score (PRS)

Additional file 2 included supplementary tables. Supplementary Table 1. Participant characteristics in the validation study. Supplementary Table 2. Additional information on extended suggestive SNPs. Supplementary Table 3. Additional information about the network. Supplementary Table 4. Pathway enrichment analysis results of PRS gene sets

Acknowledgements

The authors thank all individuals who participated in this study.

Author contributions

D.B.Y and N.S. wrote the main manuscript and prepared figures, tables. J.H.Y and S.S.K. provided and analyzed the replication study data. M.J.K., J.W.P., S.Y.J. and A.S.S. provided and analyzed the discovery study data. S.Y.Y., and J.A.S. provided corrections of the main manuscript. All authors reviewed the manuscript.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00358322, 2022R1C1C1009902). This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. RS-2024-00440787). This work was partly supported by grant no. 2520140010 from the Seoul National University Hospital Cohort Research Fund. This work was supported by Chonnam National University Hwasun Hospital under Grant HCRI23004 and HCRI21019. The funders of the study had no role in the design and conduct of the study and were not involved in collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

This study was approved by the institutional review board (IRB) of SNUH, and the ethics committee approved this retrospective study with a waiver of written informed consent (IRB No. 1906-116-1041, 1908-094-1055, CNUHH-2020-063).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 10 April 2024 / Accepted: 26 February 2025
Published online: 13 March 2025

References

1. Kang MJ, Jung KW, Bang SH, Choi SH, Park EH, Yun EH, Kim HJ, Kong HJ, Im JS, Seo HG. Community of Population-Based regional Cancer R. Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2020. Cancer Res Treat. 2023;55:385–99.
2. Baek SK, Lee JS, Hwang IG, Kim JG, Kim TW, Sohn SK, Kang MY, Lee SC. Clinical characteristics and survival of colorectal cancer patients in Korea stratified by age. Korean J Intern Med. 2021;36:985–91.
3. Galon J, Bruni D. Tumor immunology and tumor evolution: intertwined histories. Immunity. 2020;52:55–81.
4. Giannakis M, Mu XJ, Shukla SA, Qian ZR, Cohen O, Nishihara R, Bahl S, Cao Y, Amin-Mansour A, Yamauchi M, Sukawa Y, Stewart C, et al. Genomic correlates of Immune-Cell infiltrates in colorectal carcinoma. Cell Rep. 2016;15:857–65.
5. Liu Z, Zhang Y, Dang Q, Wu K, Jiao D, Li Z, Sun Z, Han X. Genomic alteration characterization in colorectal Cancer identifies a prognostic and metastasis biomarker: FAM83A|IDO1. Front Oncol. 2021;11:632430.
6. Sun Z, Xia W, Lyu Y, Song Y, Wang M, Zhang R, Sui G, Li Z, Song L, Wu C, Liew CC, Yu L, et al. Immune-related gene expression signatures in colorectal cancer. Oncol Lett. 2021;22:543.
7. Kang BW, Jeon HS, Chae YS, Lee SJ, Park JY, Choi JE, Park JS, Choi GS, Kim JG. Association between GWAS-identified genetic variations and disease prognosis for patients with colorectal cancer. PLoS ONE. 2015;10:e0119649.
8. Dai J, Gu J, Huang M, Eng C, Kopetz ES, Ellis LM, Hawk E, Wu X. GWAS-identified colorectal cancer susceptibility loci associated with clinical outcomes. Carcinogenesis. 2012;33:1327–31.
9. Junior HLR, Novaes LAC, Datorre JG, Moreno DA, Reis RM. Role of polygenic risk score in Cancer precision medicine of Non-European populations: A systematic review. Curr Oncol. 2022;29:5517–30.
10. Thomas M, Sakoda LC, Hoffmeister M, Rosenthal EA, Lee JK, van Duijnhoven FJB, Platz EA, Wu AH, Dampier CH, de la Chapelle A, Wolk A, Joshi AD, et al. Genome-wide modeling of polygenic risk score in colorectal Cancer risk. Am J Hum Genet. 2020;107:432–44.
11. Archambault AN, Su YR, Jeon J, Thomas M, Lin Y, Conti DV, Win AK, Sakoda LC, Lansdorp-Vogelaar I, Peterse EFP, Zauber AG, Duggan D et al. Cumulative Burden of Colorectal Cancer-Associated Genetic Variants Is More Strongly

Associated With Early-Onset vs Late-Onset Cancer. *Gastroenterology* 2020;158: 1274-86 e12.
12. Adolphe C, Xue A, Fard AT, Genovesi LA, Yang J, Wainwright BJ. Genetic and functional interaction network analysis reveals global enrichment of regulatory T cell genes influencing basal cell carcinoma susceptibility. Genome Med. 2021;13:19.
13. Climente-Gonzalez H, Lonjou C, Lesueur F, group Gs, Stoppa-Lyonnet D, Andrieu N, Azencott CA. Boosting GWAS using biological networks: A study on susceptibility to Familial breast cancer. PLoS Comput Biol. 2021;17:e1008819.
14. Yang W, Zhang T, Song X, Dong G, Xu L, Jiang F. SNP-Target genes interaction perturbing the Cancer risk in the Post-GWAS. Cancers (Basel) 2022;14.
15. Choi CK, Yang JH, Shin MH, Cho SH, Kweon SS. No association between genetically predicted C-reactive protein levels and colorectal cancer survival in Korean: two-sample Mendelian randomization analysis. Epidemiol Health. 2023;45:e2023039.
16. Cheol Seong S, Kim YY, Khang YH, Heon Park J, Kang HJ, Lee H, Do CH, Song JS, Hyon Bang J, Ha S, Lee EJ, Ae Shin S. Data resource profile: the National health information database of the National health insurance service in South Korea. Int J Epidemiol. 2017;46:799–800.
17. Martin FJ, Amode MR, Aneja A, Austine-Orimoloye O, Azov AG, Barnes I, Becker A, Bennett R, Berry A, Bhai J, Bhurji SK, Bignell A, et al. Ensembl 2023. Nucleic Acids Res. 2023;51:D933–41.
18. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44:D457–62.
19. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, Ami GOH. Web presence working G. AmiGO: online access to ontology and annotation data. Bioinformatics. 2009;25:288–9.
20. Moon S, Kim YJ, Han S, Hwang MY, Shin DM, Park MY, Lu Y, Yoon K, Jang HM, Kim YK, Park TJ, Song DS, et al. The Korea biobank array: design and identification of coding variants associated with blood biochemical traits. Sci Rep. 2019;9:1382.
21. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomart. Nat Protoc. 2009;4:1184–91.
22. snpStats DC. SnpMatrix and XSnpMatrix classes and methods, 2023:R package version 1.52.0.
23. Turner S, Armstrong LL, Bradford Y, Carlson CS, Crawford DC, Crenshaw AT, de Andrade M, Doheny KF, Haines JL, Hayes G, Jarvik G, Jiang L et al. Quality Control Procedures for Genome-Wide Association Studies. *Current Protocols in Human Genetics* 2011;68: 1.19.1-1.8.
24. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48:1284–7.
25. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. Nature. 2015;526:68–74.
26. Rizvi AA, Karaesmen E, Morgan M, Preus L, Wang J, Sovic M, Hahn T, Sucheston-Campbell LE. Gwasurvivr: an R package for genome-wide survival analysis. Bioinformatics. 2019;35:1968–70.
27. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015;31:3555–7.
28. Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X, Liu X, rMVP. A Memory-efficient, Visualization-enhanced, and Parallel-accelerated tool for Genome-wide association study. Genomics Proteom Bioinf. 2021;19:619–28.
29. Balduzzi S, Rucker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health. 2019;22:153–60.
30. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc. 2020;15:2759–72.
31. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Med. 2020;12:44.
32. Kassambara A, Kosinski M, Biecek P, Fabian S. Package 'survminer', 2022.
33. Milacic M, Beavers D, Conley P, Gong C, Gillespie M, Griss J, Haw R, Jassal B, Matthews L, May B, et al. The reactome pathway knowledgebase 2024. Nucleic Acids Res. 2024;52:D672–8.
34. Therneau TM, Grambsch PM, Therneau TM, Grambsch PM. The Cox model. Springer; 2000.
35. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, et al. The genemania prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;38:W214–20.
36. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.
37. Marcoulides KM, Raykov T. Evaluation of variance inflation factors in regression models using latent variable modeling methods. Educ Psychol Meas. 2019;79:874–82.
38. Ntziachristos P, Lim JS, Sage J, Aifantis I. From fly wings to targeted cancer therapies: a centennial for Notch signaling. Cancer Cell. 2014;25:318–34.
39. Knosel T, Emde A, Schluns K, Chen Y, Jurchott K, Krause M, Dietel M, Petersen I. Immunoprofiles of 11 biomarkers using tissue microarrays identify prognostic subgroups in colorectal cancer. Neoplasia. 2005;7:741–7.
40. Sun L, Chen B, Wu J, Jiang C, Fan Z, Feng Y, Xu Y. Epigenetic regulation of a disintegrin and metalloproteinase (ADAM) transcription in colorectal Cancer cells: involvement of beta-Catenin, BRG1, and KDM4. Front Cell Dev Biol. 2020;8:581692.
41. Jackstadt R, van Hooff SR, Leach JD, Cortes-Lavaud X, Lohuis JO, Ridgway RA, Wouters VM, Roper J, Kendall TJ, Roxburgh CS, Horgan PG, Nixon C et al. Epithelial NOTCH signaling rewires the tumor microenvironment of colorectal Cancer to drive Poor-Prognosis subtypes and metastasis. Cancer Cell 2019;36: 319– 36 e7.
42. Yang J, Nie J, Ma X, Wei Y, Peng Y, Wei X. Targeting PI3K in cancer: mechanisms and advances in clinical trials. Mol Cancer. 2019;18:26.
43. Li H, Zhang H, Lu G, Li Q, Gu J, Song Y, Gao S, Ding Y. Mechanism analysis of colorectal cancer according to the MicroRNA expression profile. Oncol Lett. 2016;12:2329–36.
44. Chong X, Chen J, Zheng N, Zhou Z, Hai Y, Chen S, Zhang Y, Yu Q, Yu S, Chen Z, Bao W, Quan M, et al. PIK3CA mutations-mediated downregulation of circLHFPL2 inhibits colorectal cancer progression via upregulating PTEN. Mol Cancer. 2022;21:118.
45. Wisastra R, Dekker FJ. Inflammation, Cancer and oxidative Lipoxygenase activity are intimately linked. Cancers (Basel). 2014;6:1500–21.
46. Wang D, Li Y, Zhang C, Li X, Yu J. MiR-216a-3p inhibits colorectal cancer cell proliferation through direct targeting COX-2 and ALOX5. J Cell Biochem. 2018;119:1755–66.
47. Sun W, Guo C, Meng X, Yu Y, Jin Y, Tong D, Geng J, Huang Q, Qi J, Liu A, Guan R, Xu L, et al. Differential expression of PAI-RBP1, C1orf142, and COTL1 in non-small cell lung cancer cell lines with different tumor metastatic potential. J Investig Med. 2012;60:689–94.
48. Xin D, Rendon BE, Zhao M, Winner M, McGhee Coleman A, Mitchell RA. The MIF homologue D-dopachrome tautomerase promotes COX-2 expression through beta-catenin-dependent and -independent mechanisms. Mol Cancer Res. 2010;8:1601–9.
49. Kamo H, Kawahara R, Simizu S. Tyrosinase suppresses vasculogenic mimicry in human melanoma cells. Oncol Lett. 2022;23:169.

Publisher's note