Identification of *MMP14* and *MKLN1* as colorectal cancer susceptibility genes and drug-repositioning candidates from a genome-wide association study

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

Abstract

Background Genome-wide association studies (GWAS) and subsequent functional interpretation have been used to identify susceptible genes and potential drug-repositioning candidates. This study aimed to identify genes associated with colorectal cancer (CRC) and potential drug-repositioning candidates.

Methods Patients with CRC at Seoul National University Hospital (SNUH, discovery study) and Chonnam National University Hospital (CNUH, replication study) were included as case groups. The Korean Genome and Epidemiology Study (KoGES) participants were included as a control group. Single-nucleotide polymorphisms (SNPs) were extracted from blood-derived DNA (N = 409,063). A SNP-based logistic regression model was applied. Furthermore, post-GWAS analysis was conducted. Drug-repositioning candidates were identified using a pre-trained deep neural network and the druggability assessment tool.

Results In the discovery study, we conducted a 1:3 age- and sex-matched case–control study that included 500 CRC cases (mean age 63.0 ± 7.15 years) and 1,500 healthy controls (mean age 62.9 ± 7.07 years), each group comprising 50% males and 50% females. The replication study enrolled 4,860 patients with CRC and 46,384 healthy controls. The two-stage GWAS revealed statistically significant associations among *MKLN1* (rs75170436, 7q32.3, beta (log odds ratio) = − 0.90, $P_{meta} = 5.90 \times 10^{-13}$), *MMP14* (rs3751489, 14q11.2, beta (log odds ratio) = − 1.91, $P_{meta} = 2.31 \times 10^{-12}$). Post-GWAS functional analysis revealed strong associations on two genes highlighting deleterious effects and increased gene expression. Drug-repositioning analysis identified GW0742 (PPARβ/δ agonist) with the highest binding score and druggability score for MMP14 with a reference allele (12.06, 0.85).

---

†Dabin Yun and Jung-Ho Yang are first co-authors who contributed equally to this study.

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

*Correspondence:
Sun-Seog Kweon
ujingogo@paran.com
Nan Song
nan.song@chungbuk.ac.kr
Full list of author information is available at the end of the article

Yun *et al. Journal of Translational Medicine*   (2025) 23:543

Page 2 of 13

Conclusions Using GWAS, *MKLN1* and *MMP14* were found to be associated with CRC development and we identified GW0742 (PPARβ/δ agonist) as a potential drug-repositioning candidate for CRC based on MKLN1 and MMP14. These findings improve the understanding of CRC development and provide insights into novel therapeutic targets and candidates for CRC treatment.

Keywords Colorectal neoplasms, Genome-wide association study, *MMP14*, *MKLN1*, Drug repositioning, PPARδ/β agonist

Background

Colorectal cancer (CRC) is a prevalent malignancy with a high incidence rate, ranking third in men and women worldwide in 2020 [1]. Globally, epidemiological factors such as obesity, smoking, alcohol consumption, and red meat consumption have been shown to increase the risk of CRC development [2]. In South Korea, CRC ranks third among all cancers in both men and women [3]. The age-standardized incidence rates of CRC per 100,000 have increased from 20.5 (1999) to 27.1 (2020) [3]. Factors such as age, sex, and obesity were found to be notably associated with CRC incidence among men and women in Korea, whereas the effects of smoking, alcohol consumption, and red meat consumption were not consistent [4–6].

Heritability of CRC has been estimated to be approximately 13–35%, necessitating rigorous genetic research to prevent CRC onset and enhance survival outcomes. Known CRC predisposition genes (*APC, MLH1, MSH2, MSH6, PMS2, STK11, MUTYH, SMAD4, BMPR1 A, PTEN, TP53, CHEK2, POLD1,* and *POLE*) are associated with CRC in 5–10% of cases, with much heritability remaining still unexplained [7–9]. In a case–control genome-wide association study (GWAS), Huyghe et al. identified 40 genetic loci linked to CRC susceptibility, including *CHD1* as a protective factor in CRC [10]. Furthermore, Jia et al. revealed three novel susceptibility loci [rs647161 (5q31.1), rs2423279 (20p12.3), and rs10774214 (12p13.32)] for CRC, validated in both East Asian and European groups [11]. Houlston et al. identified four novel susceptibility loci [rs4444235 (14q22.2), rs9929218 (16q22.1), rs10411210 (19q13.1), and rs961253 (20p12.3)] for CRC [12]. Overall, more than 250 susceptibility loci for CRC have been identified through GWAS.

As genetic research evolves, preliminary findings from case–control GWAS have enabled more complex investigations. Although identifying potential genetic loci that correlate with diseases is crucial, these findings often fail to provide information about causal genetic variants. To address this limitation, researchers have increasingly focused on post-GWAS functional analyses [13–15]. These advanced studies validate GWAS findings and integrate diverse datasets and sophisticated methodologies to uncover functional implications. Additionally, because post-GWAS functional analysis utilizes knowledge about cancer signaling pathways (e.g., LKB1/AMPK pathway, PI3 K-AKT/mTOR pathway) [16] and the functions of cancer-related proteins (e.g., Matrix Metalloproteinases) [17], it offers biological interpretations of GWAS results, particularly regarding cancer development and progression. Drug repositioning, a promising functional approach to cancer treatment, involves repositioning existing medications originally developed for other diseases to treat cancer. Using GWAS findings, medications can be identified to effectively target specific genetic variations in cancers [18].

For example, Irham et al. employed PubMed text mining to retrieve single-nucleotide polymorphisms (SNPs) associated with CRC and collected data from 40 CRC studies using GWAS [19]. They identified 170 unique SNPs linked to CRC risk, which increased to 210 CRC-associated genes upon considering linkage disequilibrium (LD). Additionally, most of these genes had missense or nonsense variants and exhibited cis-expression quantitative trait loci effects. Using the STRING database, the list of biological CRC-risk genes was reduced to 128. Additionally, 21 target genes were identified using DrugBank and the Therapeutic Target Database, resulting in the identification of 166 potential target drugs. Three of these drugs (bevacizumab, aflibercept, and tegafur uracil) are frequently used to treat CRC. However, one limitation of this approach is the lack of structural information on drug targets and candidates.

Building on previous findings of GWAS on CRC, our research aimed to identify genetic susceptibility to CRC and address limitations such as the lack of post-GWAS functional analysis and structural information. Our primary objective was to enhance the understanding of CRC through a two-stage GWAS and post-GWAS functional analysis. This approach could allow for a detailed exploration of the genetic factors associated with CRC, validating and expanding GWAS results. Additionally, we highlight the potential for drug repositioning based on functional CRC targets. To optimize this approach, we utilized a deep learning-based method to identify a novel drug target for CRC by incorporating structural information.

Yun *et al. Journal of Translational Medicine*   (2025) 23:543

Page 3 of 13

Methods

Study population

The present study recruited patients with CRC who underwent surgical resection at the Seoul National University Hospital (SNUH). Patients with CRC were enrolled starting in January 2002, and we selected 960 patients with genome-wide SNPs for the GWAS of CRC. To perform a case–control GWAS, we used the Korean Genome and Epidemiology Study (KoGES) population as a control group, with age and sex matching at a ratio of 1:3 to patients with CRC from SNUH (Table 1). Prior to matching the KoGES group (control group) with the CRC group (case group), we stratified each group into the following five sub-groups by the age of participants: 50 years ≤ age < 55 years, 55 years ≤ age < 60 years, 60 years ≤ age < 65 years, 65 years ≤ age < 70 years, and 70 years ≤ age < 75 years. After stratification, the control and case groups were frequency-matched in a ratio of 1:3 by age and sex category. The discovery GWAS included 100 and 300 individuals from the CRC case and control groups, respectively. In total, we selected 500 individuals for the CRC case group from 960 SNUH patients with CRC and 1,500 individuals for the control group from 75,639 KoGES participants.

For the replication study, data from patients with CRC from Chonnam National University Hospital (CNUH), Korea, were utilized as described in detail elsewhere [20]. Patients diagnosed with CRC between 2004 and 2014 were divided into two groups—MEGA array and ONCO array— comprising 2,545 and 2,315 patients, respectively. Further details regarding the replication study population are provided in Supplementary Table 1. To increase the statistical power and ensure extended replication beyond specific age groups, 46,384 unmatched controls from KoGES were selected for replication analysis. The ratio of the CRC case and control groups in the replication study was approximately 1:10 (Supplementary Table 1).

Data collection

Clinicopathological and epidemiological data were collected from patient surveys and electronic medical records at SNUH and CNUH. Additional epidemiological data were obtained from participant surveys and clinical measurements from the KoGES. For the GWAS, the body mass index (BMI) was included as a continuous variable and used to categorize the participants into standard groups, ranging from underweight to obese. Sex, drinking history (never or ever), and smoking history (never or ever) were included as categorical variables.

Table 1 Characteristics of participants included in the discovery genome-wide association study (GWAS) on colorectal cancer risk

Characteristics	Case (N = 500) N (%)	Control (N = 1,500) N (%)	P-value[a]
Age (year)			0.954
Mean (SD)	63.0 (7.15)	62.9 (7.07)	
Sex			> 0.999
Male	250 (50.0%)	750 (50.0%)	
Female	250 (50.0%)	750 (50.0%)	
BMI (kg/m²)			0.002
Underweight (BMI < 18.5)	23 (4.6%)	24 (1.6%)	
Normal (18.0 ≤ BMI < 23.0)	173 (34.6%)	545 (36.3%)	
Overweight (23.0 ≤ BMI < 25.0)	135 (27.0%)	411 (27.4%)	
Obese (25.0 ≤ BMI)	169 (33.8%)	519 (34.6%)	
Unknown	0 (0%)	1 (0.1%)	
Smoking history			0.226
Never	339 (67.8%)	1092 (72.8%)	
Ever	161 (31.2%)	438 (29.2%)	
Drinking history			0.394
Never	265 (53.0%)	760 (50.7%)	
Ever	235 (47.0%)	740 (49.3%)	

BMI value was calculated using the following formula: weight in kilograms divided by the square of height in meter. Participants who had ever smoked or consumed alcohol were categorized as 'Ever'

BMI body mass index, *SD* standard deviation

[a] Continuous variables were analyzed using the linear rank test and categorical variables were analyzed using the log-rank test

Genotyping and quality controls

All the samples for the discovery study were genotyped using the Korean Chip, a specialized array optimized for the Korean population. Further details regarding the KoreanChip can be found in a referenced publication [21]. All samples for the replication study were genotyped using the Infinium OncoArray-500 K BeadChip (Illumina Inc., San Diego, CA, USA) and the Infinium Multi-Ethnic Global BeadChip (MEGA, Illumina Inc.). As shown in the flowchart of the study design (Fig. 1), we applied the following three exclusion criteria for 446,300 genome-wide SNPs: (1) minor allele frequency (MAF) < 0.01, (2) Hardy–Weinberg equilibrium (HWE) *P*-value < 1.00 $\times 10^{-6}$, and (3) call rate < 98%. Finally, 409,063 genotyped SNPs were included in the discovery study.

Additionally, we imputed the SNUH GWAS data using the Michigan Imputation Server [22]. We used MINIMAC 4 software for imputation and selected 1,000 genomes for the reference panel. The subpopulation was set as East Asians because our SNUH GWAS data included only South Koreans. For quality control of the imputation results, we selected an r-squared statistic (rsq) filter value of 0.3, which removed more than 70% of the poorly imputed SNPs. Furthermore, we implemented standard quality control settings, including MAF ≥ 1%, call rate ≥ 98%, and removal of duplicate SNPs, using Michigan Imputation Server. Consequently, 2,514 SNPs were excluded from the imputed dataset for two reasons—invalid alleles (942 SNPs) and allele mismatch (1,503 SNPs). Additionally, for any SNPs duplicated between the imputed and genotyped datasets, only genotyped SNPs were included in the remaining 7,835,640 SNPs for the current analyses.

Statistical analysis and GWAS

The overall flow of the analysis is illustrated in Fig. 1. For additive SNP-based logistic regression, we utilized the Plink software [23] and selected age (continuous), sex (categorical), and the BMI (categorical) as confounding factors. Smoking and drinking histories were not included as confounding factors because these variables were not statistically significant (Table 1). These confounding factors were consistently used for both patients with CRC (case group) and the KoGES group (control group) in the discovery and replication studies. Significant SNPs identified in the discovery study were subsequently used in a replication study. We set a *P*-value threshold of 5.00×10^{-8} (GWAS nominal *P*-value) for the discovery study and 0.05 for the replication study. Beta value was presented as a log odds ratio. To combine the discovery and replication study results, we conducted the meta-analysis using the 'meta' R package [24]. From
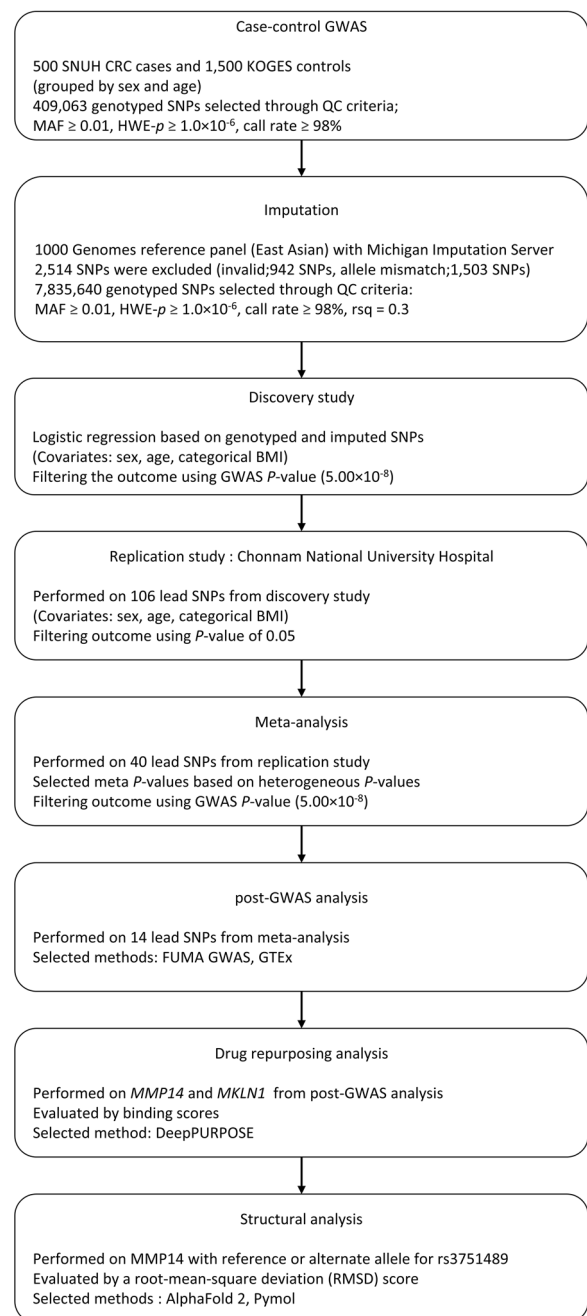
Fig. 1 Flowchart of the genome-wide association study (GWAS) and functional analyses. *SNUH* Seoul National University Hospital, *KOGES* Korean Genome and Epidemiology Study, *SNP* single-nucleotide polymorphism, *QC* quality control, *MAF* minor allele frequency, *HWE* Hardy–Weinberg equilibrium, *rsq* r-squared statistic, *FUMA GWAS* functional mapping and annotation of GWAS, *GTEx* Genotype-Tissue Expression

the meta-analysis, we extracted both the heterogeneous *P*-value and the I² value. To interpret the meta-analysis results, we used fixed-effect *P*-values, as the control

Yun *et al. Journal of Translational Medicine*   (2025) 23:543

Page 5 of 13

groups of the discovery and replication studies were from the same population cohort group.

The results of statistical analysis were visualized using the 'CMplot' R package [25]. This package generated Manhattan plots of significant SNPs linked to clinical outcomes (i.e., CRC incidence), with significant SNPs differentiated by distinctive colors based on a cutoff *P*-value. We performed a pairwise LD analysis between SNPs to distinguish independent index SNPs from those in LD located close to each chromosome. The r^2 values were calculated using the 'LDmatrix' function from the 'LDlink' web-based service [26]. To ensure precision, we specified the genome build as GRCh38 and the subpopulation as East Asians. From the SNPs with r^2 values ≥ 0.6 on each chromosome, we selected the SNP with the lowest *P*-value.

Functional mapping and annotation of GWAS

To extend the statistical analysis results, we utilized functional mapping and annotation of GWAS [27]. We uploaded the GWAS summary statistics of rsID and the meta-analysis *P*-value for each SNP. For the SNP identification step, we set a *P*-value threshold of 5.00×10^{-8} as the maximum *P*-value for lead SNPs and 0.05 as the maximum *P*-value cutoff. Additionally, thresholds of 0.6 and 0.1 were used for the first and second r^2 values, respectively, to define the independently significant SNPs. Furthermore, we selected 1,000 Genome Phase 3 East Asian patients as the reference panel population.

Subsequently, we conducted gene annotation using positional mapping, considering a maximum distance of 10 kb between the SNPs and their corresponding genes. We implemented additional SNP-filtering techniques as an extension of this approach. First, we used a combined annotation-dependent depletion (CADD) framework [28]. The CADD score provides a comprehensive framework for evaluating the impact of single-nucleotide and insertion/deletion variants in the human genome. By integrating multiple annotations, the CADD framework contrasts naturally occurring variants with simulated mutations and offers a metric that captures the evolutionary survival and potential deleterious effects of the variants. Our analysis filtered variants with scores > 12.47 (default settings). If the score was ≥ 10, the variant was expected to be among the 10% most deleterious substitutions. Furthermore, if the score was ≥ 30, the variant was expected to be among the 0.1% most deleterious substitutions. Second, we used RegulomeDB (RDB) to filter the SNPs, focusing on their established and predicted regulatory elements within the intergenic regions [29]. The RDB aggregates data from various sources, including chromatin structures, DNase hypersensitivity sites, and transcription factor binding sites. Variants were categorized on a scale of 1–7, with a low score indicating increased regulatory significance. As scores from 4 and above seemed to contain multiple supporting evidence, we primarily focused on variants with RDB scores ≤ 4.

Genotype-tissue expression analysis

In the landscape of genomics research, the Genotype-Tissue Expression (GTEx) project has facilitated the exploration of human gene expression and its relationship to genetic variation [30]. GTEx has been used to decode how genetic variants influence gene activity across various human tissues, thereby highlighting an association between genotype and gene expression. The transcripts per million (TPM) value is a scoring metric for GTEx, providing a normalization method for RNA sequencing data that adjusts for both gene length and library size. This yields a robust expression measure suitable for comparisons across samples. In addition, TPM helps discern gene expression levels, thereby distinguishing among genes expressed in various tissues. In this study, we used the GTEx portal to investigate the expression profiles of specific genes that matched our significant SNPs, visualized with violin plots. This analysis was conducted on sigmoid and transverse colon tissues. As all genes could not be used for the GTEx analysis, we used only those present in the GTEx portal. Consequently, *RP11-480I12.4* was not applicable for this analysis.

Deep learning-based drug repositioning and validation

To identify potential drug candidates, we applied Deep-PURPOSE [31], a deep learning-based tool designed for drug-target interaction predictions, drug property prediction, and drug repositioning. With a deep learning-based approach, it could offer enhanced capabilities to identify and process complex molecular data compared to traditional computational modeling methods [31]. DeepPURPOSE employs an encoder-decoder architecture. The adaptability of this structure ensures that the selection of an encoder leads to suitable decoder pairing. Implementation of a comprehensive training resulted in the establishment of a model optimized for various predictions. DeepPURPOSE efficiently processes inputs, specifically the simplified molecular input line-entry system (SMILES) structure and amino acid sequence of a compound. These elements were converted into vector representations using various molecular encoders. Transformation techniques for compounds include deploying multilayer perceptrons on established fingerprints and utilizing convolutional neural networks on SMILES sequences. The most important output component is the binding score that quantifies the affinity between a drug and its potential target, presenting a numerical measure of their interaction strength. Low binding scores

Yun *et al. Journal of Translational Medicine*   (2025) 23:543

Page 6 of 13

correspond to a strong affinity. For the present study, we used a one-liner code from DeepPURPOSE with amino acid sequences of significant target genes from post-GWAS functional analysis and selected the Broad Institute's Broad Repositioning Hub as the database [32]. For this analysis, we utilized three amino acid sequences—MKLN1, MMP14 with the reference allele, and MMP14 with the alternate allele. As rs3751489 is an exonic SNP, the amino acid sequence of MMP14 was changed from arginine (reference allele) to histidine (alternate allele) at position 431. We selected only the top six drug-repositioning targets ordered by binding affinity to provide a clear view of the results. Additionally, we validated drug repositioning results using DogSiteScorer [33]. DogSiteScorer offers a function to detect potential binding pockets for a target protein and estimate a druggability score for a small molecule. The druggability scores ranged from 0 to 1, where a higher value indicates greater druggability. We utilized druggability scores for the interaction between the potential drug repositioning candidate and the target proteins (MKLN1, MMP14 with the reference allele, and MMP14 with the alternate allele).

Structural analysis

To investigate the structural characteristics of proteins from susceptible genes, we used AlphaFold 2 [34]. We used the amino acid sequences of proteins as inputs and selected the protein structure with the highest predicted local distance difference test (pLDDT) score. The pLDDT score measures the local accuracy of the predicted protein structure, assessing how closely the predicted distances between pairs of amino acids in the protein align with the actual distances in the native structure. To investigate structural differences between proteins, we used PyMOL with an alignment function [35]. The root mean square deviation (RMSD) score was obtained using this function. An RMSD score greater than 5 is typically perceived as indicating a significant structural difference between two proteins.
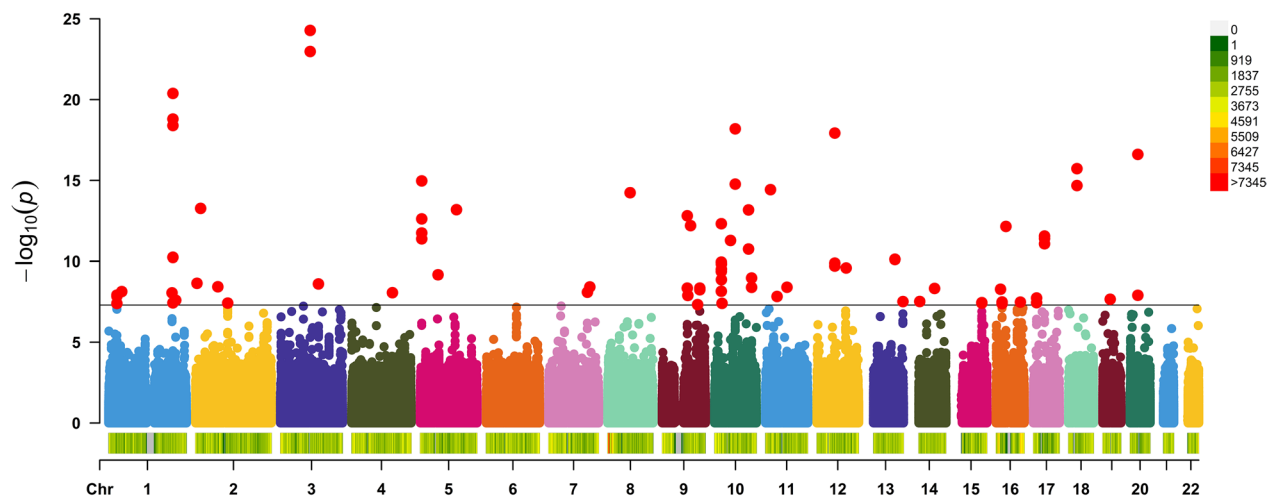
Results

Characteristics of the study population

The characteristics of the study population are summarized in Table 1. The case group had a mean age of 63.0 years with a standard deviation of 7.15 years, whereas the control group had a mean age of 62.9 years with a standard deviation of 7.07 years. Regarding sex distribution, both the case and control groups showed a balanced representation of males and females. Owing to age- and sex-matching, the distribution of age and sex between patients with CRC and controls was similar (*P*-value for age = 0.954, *P*-value for sex > 0.999). However, we noticed a statistically significant difference between the case and

control groups (*P*-value = 0.002) in the BMI categories. In the underweight category, 4.6% of individuals were in the case group, which was slightly higher than the 1.6% proportion in the control group. For the normal weight, overweight, and obese categories, the frequencies were slightly reduced in the case group (34.6%, 27.0%, and 33.8%, respectively) compared to those in the control group (36.3%, 27.4%, and 34.6%, respectively). Regarding smoking history, the case group (31.2%) had a higher proportion of smokers than the control group (29.2%). However, the difference was not statistically significant (*P*-value = 0.226). Regarding drinking history, the case group had 47.0% individuals with a history of drinking compared to 49.3% individuals in the control group, but this difference was not statistically significant (*P*-value = 0.394).

Signals and functional annotations of GWAS on CRC risk

In the discovery study, 106 SNPs showed *P*-values below the threshold of 5.00×10^{-8}, as depicted in Fig. 2. Among the SNPs utilized for the replication study, 40 showed *P*-values below the threshold of 0.05. A meta-analysis of these 40 SNPs identified 14 SNPs with meta-analysis *P*-values below the threshold of 5.00×10^{-8}. Furthermore, we identified 10 lead SNPs (rs151216315, rs79156133, rs75170436, rs3751489, rs16850506, rs116964687, rs146410474, rs141065065, rs145092508, and rs2416729) and selected four SNPs with low *P*-values in LD ($r^2 <$ 0.6) (Table 2). Additional information on the 40 significant SNPs obtained from the meta-analysis is presented in Supplementary Table 2. In the meta-analysis, we identified the strongest associations with CRC risk in rs151216315 (q41, *PTPN14*, $P_{meta} = 8.44 \times 10^{-21}$, $I^2 = 83.67\%$, $P_{het} = 1.33 \times 10^{-2}$) and rs79156133 (p13.12, *MKL2*, $P_{meta} = 1.69 \times 10^{-36}$, $I^2 = 39.90\%$, $P_{het} = 1.97 \times 10^{-1}$). These two SNPs, rs151216315 (Beta = −1.68) and rs791561330 (Beta = −0.90), were associated with protective effects against CRC development. Their effect sizes indicate a stronger protective association compared to the average effect (mean Beta = −0.22) observed for protective SNPs against CRC listed in the GWAS catalog [36]. Notably, the effect size of beta for rs151216315 (−1.68) surpasses the strongest beta value (−1.05). Other top SNPs were rs75170436 (7q32.3, *MKLN1*, $P_{meta} = 5.90 \times 10^{-36}$, $I^2 = 0.00\%$, $P_{het} = 6.68 \times 10^{-1}$) and rs3751489 (14q11.2, *MMP14*, $P_{meta} = 2.31 \times 10^{-12}$, $I^2 = 79.96\%$, $P_{het} = 2.55 \times 10^{-2}$).

These two SNPs (rs75170436 and rs3751489) exhibited excellent CADD and RDB scores. Specifically, rs75170436 had a CADD score of 15.52 (top 10% deleterious substitutions) and an RDB score of 4 (minimal binding evidence). In addition, it was identified as an intronic variant through functional annotations. Conversely,

Yun *et al. Journal of Translational Medicine*   (2025) 23:543

Page 7 of 13

Fig. 2 Manhattan plot of GWAS in the discovery study. 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 −log10 of the *P*-value. The solid horizontal line corresponds to a *P*-value of 5.00×10^{-8}. Red dots represent SNPs with *P*-value $< 5.00 \times 10^{-8}$

rs3751489 had a CADD score of 32.00 (top 0.1% deleterious substitutions) and an RDB score of 3a (less likely to affect the binding of transcription factors or DNase). Unlike rs75170436, rs3751489 was identified as an exonic variant. Therefore, rs75170436 was expected to indirectly affect MKLN1, whereas rs3751489 was expected to directly affect the structure and function of MMP14 by changing the reference allele (G, guanine) to an alternate allele (A, adenine). No other lead SNPs showed remarkable CADD scores (CADD <10) or RDB scores ≤4. Notably, two other SNPs (rs151216315 and rs79156133), despite having low meta-analysis *P*-values, exhibited less remarkable CADD and RDB scores (rs151216315: CADD =4.31, RDB =5; rs79156133: CADD =0.74, RDB =7).

Gene expression in sigmoid colon and transverse colon tissues

In the genotype-tissue expression analysis results (Fig. 3, Table 3), we found that *MMP14* was likely to have exonic and deleterious substitutions (rs3751489) from GWAS and functional annotation and exhibited increased expression levels compared to the other genes in colon tissues (TPM for sigmoid colon tissue =134.60 and TPM for transverse colon tissue =97.56). Additionally, *MKLN1*, associated with a likely indirect deleterious substitution (rs75170436), showed slightly elevated expression in colon tissues (TPM for sigmoid colon tissue =11.81 and TPM for transverse colon tissue =7.96). Additionally, *MKL2*, having the strongest signal (rs14083527) on CRC risk, exhibited clinically moderate expression levels (TPM above 10) in colon tissues (TPM for sigmoid colon tissue =12.68 and TPM for transverse colon tissue =10.05) despite

the prior functional annotation analysis of rs79156133 showing relatively low CADD (0.743) and high RDB (7) scores, making it unlikely to have causative or functional effects on CRC (Table 2).

Potential drug-repositioning candidate discovery targeting *MKLN1* and *MMP14*

Based on GWAS, functional annotation, and gene expression results, we selected two genes (*MMP14* and *MKLN1*) for potential drug repositioning. Potential drug-repositioning candidates targeting these two genes were prioritized by their binding scores, as shown in Table 4. The standout drug-repositioning candidate from our analysis was GW0742 (PPARβ/δ agonist). This compound exhibited a remarkably high binding affinity for MKLN1, with a score of 3.79. Furthermore, its binding score of 12.06 for MMP14 with the reference allele (G) was the highest among all candidates. In contrast, the binding score of GW0742 for MMP14 with an alternate allele (A) changed the most compared to that for other candidates, from 12.06 to 21.15. Druggability scores calculated using DogSiteScorer [33] showed distinct trends. We identified the lowest druggability score (0.51) for the interaction between MKLN1 and GW0742. The score for the interaction between MMP14 with an alternate allele (A) and GW0742 was 0.81. Furthermore, the interaction involving MMP14 with a reference allele (G) and GW0742 yielded the highest druggability score (0.85). As MKLN1's low druggability score was considered insufficient for validation, we selected MMP14 for further structural analysis.

Yun *et al. Journal of Translational Medicine*   (2025) 23:543

Page 8 of 13

Table 2 Signals and functional annotations from GWAS on colorectal cancer risk

SNP	Chr	Position	Ref	Alt	Band	Gene	Discovery		Replication		Meta-analysis				Functional annotation		
							Beta	P	Beta	P	Beta	P_{meta}	P_{het}	I² (%)	CADD	RDB	Annotation
rs151216315	1	214485581	A	G	q41	PTPN14	−1.23	2.50E-08	−0.65	2.58E-15	−0.72	8.44E-21	1.33E-02	83.67	4.31	5	Intronic
rs79156133	16	14083527	T	C	p13.12	MKL2	−1.43	5.23E-09	−1.8	2.10E-29	−1.69	1.69E-36	1.97E-01	39.90	0.74	7	Intronic
rs75170436	7	131235985	T	C	q32.3	MKLN1	−0.95	3.79E-09	−0.84	3.14E-05	−0.90	5.90E-13	6.68E-01	0.00	15.52	4	Intronic
rs3751489	14	22844771	G	A	q11.2	MMP14	−2.57	3.01E-08	−1.36	1.26E-06	−1.68	2.31E-12	2.55E-02	79.96	32.00	3a	Exonic
rs16850506	1	202827800	C	T	q32.1	RP11-480I12.4	−1.06	8.79E-09	−0.29	2.24E-06	−0.37	2.80E-10	7.58E-05	93.61	4.11	5	Downstream
rs116964687	10	108372632	C	T	q25.1	PTGES3P5	0.83	1.77E-11	−1	2.96E-33	−0.43	5.96E-10	9.29E-35	99.34	1.045	7	Intergenic
rs146410474	8	71529830	T	C	q13.3	RP11-1102P16.1	1.74	5.83E-15	−2.21	2.45E-04	1.26	1.50E-09	7.55E-10	97.36	3.04	6	Intronic
rs141065065	1	42453743	A	G	p34.2	ZMYND12	−1.58	7.26E-09	−0.75	1.43E-02	−1.22	2.70E-09	4.37E-02	75.43	1.90	7	Intronic
rs145092508	2	4504585	G	A	p25.2	NPM1P48	0.93	2.24E-09	−1	1.52E-25	−0.47	8.93E-09	3.91E-26	99.11	0.078	7	Intergenic
rs2416729	9	119970955	T	C	q33.2	RP11-360A18.1	−1.24	4.56E-09	−0.41	5.75E-03	−0.68	2.04E-08	1.17E-03	90.51	1.408	7	Intergenic

SNP single nucleotide polymorphism, *Chr* chromosome, *Ref* reference allele, *Alt* alteration allele, P_{fixed} fixed-effect *P*-value, P_{het} heterogeneous *P*-value, *CADD* combined annotation-dependent depletion, *RDB* RegulomeDB

Yun *et al. Journal of Translational Medicine*   (2025) 23:543

Page 9 of 13
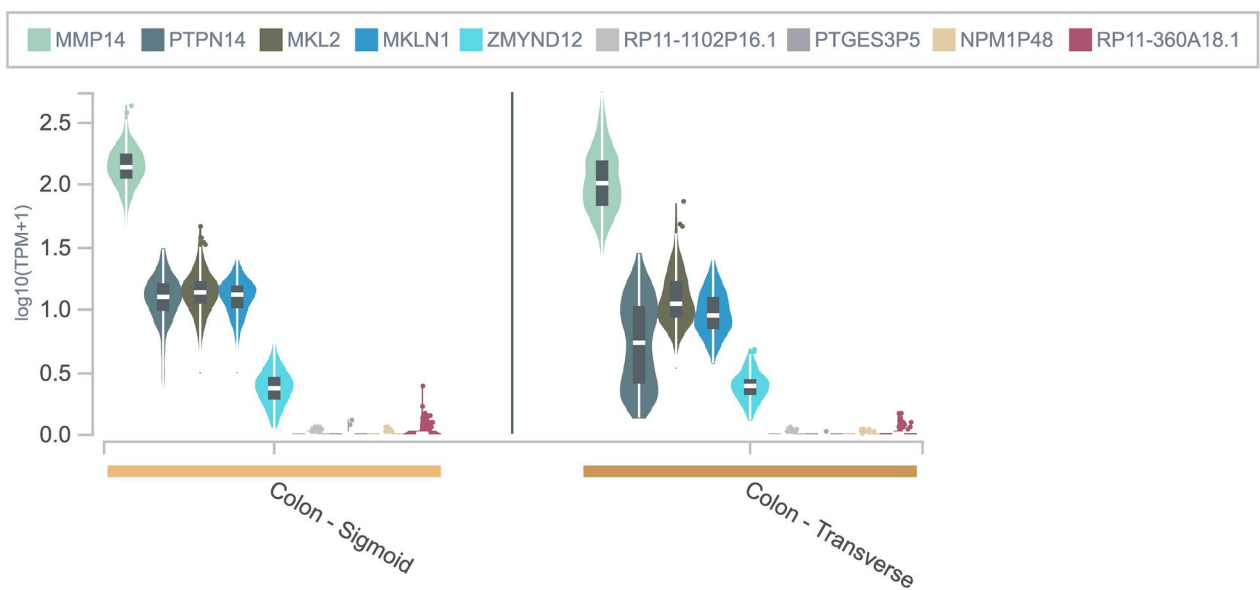
Fig. 3 Violin plot of gene expression in sigmoid colon and transverse colon tissues. The x-axis represents the category of colon tissues and the y-axis represents − log10 of (transcripts per million + 1)

Table 3 Expression of genes in sigmoid colon and transverse colon tissues

Gene	Median expression levels (TPM)[a]	
	Colon—sigmoid	Colon—transverse
MMP14	134.60	97.56
MKL2	12.68	10.05
MKLN1	11.81	7.96
PTPN14	11.5	4.27
ZMYND12	1.35	1.38
RP11-1102P16.1	0	0
PTGES3P5	0	0
NPM1P48	0	0
RP11-360 A18.1	0	0

[a] A relative expression represented by transcripts per million (TPM) values

Structure-based analysis of *MMP14*

To investigate the reasons behind the different binding scores of GW0742 based on the specific allele of MMP14, we examined the structural characteristics of MMP14 using the reference allele (G) and the alternate allele (A) at rs3751489 (Fig. 4). The RMSD score was 17.37 Å, indicating significant structural differences between the two allele-based models.

Discussion

The present study aimed to identify loci associated with CRC development in Korea using CRC patient data from SNUH and CNNUH as the case groups and individuals from KoGES as the control group. In this stage of the analysis, we focused on patients with CRC aged > 50 years. This decision was made because of an insufficient

Table 4 Targeting of MKLN1 and MMP14 and drug-repositioning candidates for colorectal cancer

MKLN1		MMP14 (ref[a])		MMP14 (alt[b])	
Drug	Binding score	Drug	Binding score	Drug	Binding score
GW0742	3.79	GW0742	12.06	Meropenem	14.72
Fulvestrant	10.18	Meropenem	13.78	Fulvestrant	15.91
Clobetasone butyrate	11.46	Fulvestrant	13.81	Halobetasol propionate	17.56
Meropenem	12.11	Halobetasol propionate	15.92	Clobetasone butyrate	19.02
Halobetasol propionate	12.27	Norethindrone	18.09	Norethindrone	21.13
Butofilolol	12.38	Clobetasone butyrate	18.69	GW0742	21.15

[a] MMP14 with the reference allele (G) for rs3751489

[b] MMP14 with the alternate allele (A) for rs3751489

Yun *et al. Journal of Translational Medicine*   (2025) 23:543

Page 10 of 13
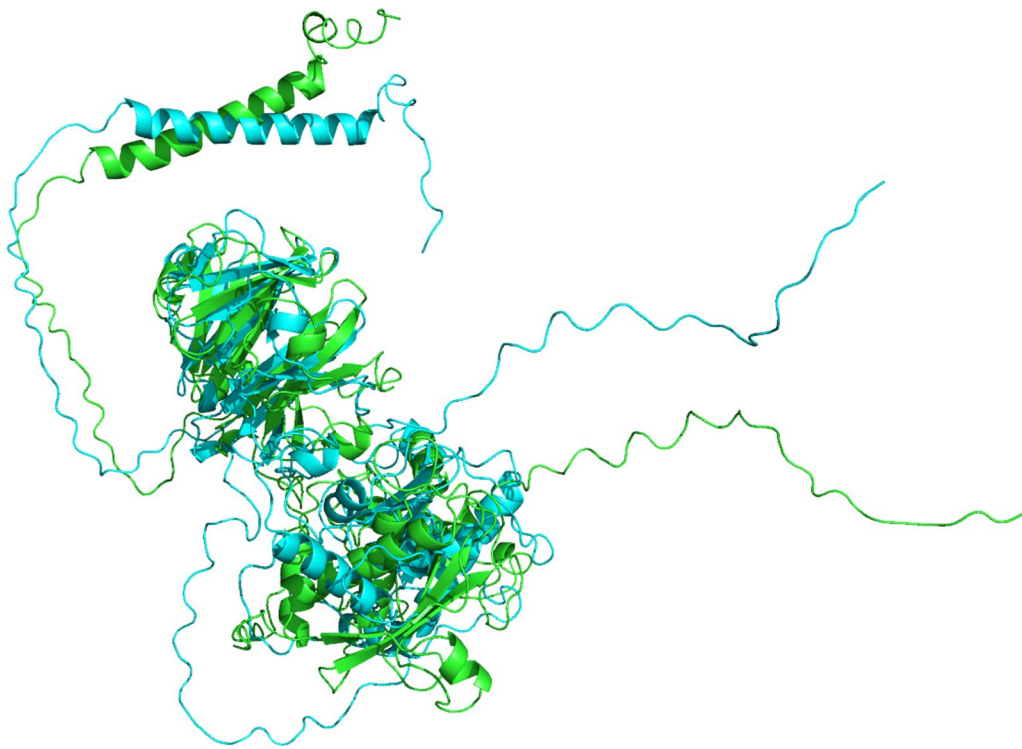
Fig. 4 Structural analysis of MMP14. The green protein is MMP14 with the reference allele (G) for rs3751489, and the blue protein is MMP14 with the alternate allele (A) for rs3751489

sample size and specific findings in South Korea, wherein the incidence rate of CRC was higher in the group aged >50 years than in that aged <50 years [4]. As national CRC screening in Korea commences at the age of 50 years, and hereditary CRC screening is recommended before this age [37], patients under 50 years of age were excluded from focusing on sporadic CRC. Our two-stage meta-GWAS identified 10 lead SNPs associated with CRC risk. We performed post-GWAS functional annotation and gene expression analyses and identified two susceptibility SNPs for CRC—rs75170436 (7q32.3, *MKLN1*) and rs3751489 (14q11.2, *MMP14*). Targeting of MKLN1 and MMP14 led to the prioritization of GW0742, a PPARβ/δ agonist, a potential drug-repositioning candidate for CRC treatment. Additionally, structural analysis of MMP14 revealed significant differences between the reference and alternate alleles supported by the differential binding scores observed in the drug repositioning analysis. An RMSD score of 17.37 Å indicates a significant structural dissimilarity [38].

The GWAS catalog [36] lists 120 publications for MKLN1 and 36 publications for MMP14. However, the specific associations we identified for rs75170436 (7q32.3, *MKLN1*) and rs3751489 (14q11.2, *MMP14*) are currently absent from the catalog. Previously reported cancer-related associations in the catalog include only lung cancer in non-smokers for *MKLN1* at chr7:131247904 and prostate cancer for *MMP14* at chr14:22836440. Therefore, our results establish novel associations between these SNPs and CRC.

Matrix Metalloproteinases (MMPs) drive tumor progression through multiple mechanisms, primarily by degrading the extracellular matrix to enable invasion and metastasis [39]. Furthermore, MMPs modulate crucial cellular processes by regulating the activity of growth factors, cytokines, and key signaling pathways involved in angiogenesis, inflammation, and epithelial-mesenchymal transition [39]. MMP14, one of the CRC susceptibility genes identified in our study, has been reported in several studies related to CRC. Cui et al. revealed that MMP14 expression levels were highly associated with CRC development and prognosis [40]. MMP14 was upregulated in CRC tissues compared to that in normal tissues (P-value < 0.05), and a high MMP14 expression was correlated with high-grade tumor stages. In addition, both disease-free survival and overall survival were significantly and negatively correlated with *MMP14* expression. Ragusa et al. revealed the mechanism whereby MMP14 contributes to CRC progression [41]. They found that MMP14 is closely related to prospero homeobox protein 1 (PROX1). In a mouse model, the knockdown of *PROX1* induced *MMP14* expression, affecting the tumor

Yun *et al. Journal of Translational Medicine*   (2025) 23:543

Page 11 of 13

microenvironment of CRC cells in various ways, including the activation of fibroblasts, dysfunctional changes in blood vessels, and loss of cytotoxic T-cell infiltration. Additionally, Ragusa et al. revealed a notable role of vascular endothelial growth factor A (VEGF-A) in the dysfunction of blood vessels and T-cell infiltration. When *MMP14* is upregulated, it activates *VEGF-A* excessively, and this change leads to the dysfunctionality of blood vessels. Consequently, the infiltration of cytotoxic T cells into the tumor deteriorates.

Regarding MKLN1, an intracellular protein involved in the cell-adhesive response and cytoskeletal complex [42], there has been limited exploration of its role in CRC development, except a study by Jung et al. [43]. Jung et al. analyzed 11,078 postmenopausal non-Hispanic white women to identify CRC susceptibility loci and found that rs117911989 (7q32.3, *MKLN1*) is associated with an increased risk of CRC development. Despite differences in ethnic characteristics between their study population and ours, the effect of *MKLN1* on CRC risk was consistent with our findings. Additionally, MKLN1 is a component of the carboxy-terminal LisH complex, which is involved in cancer cell plasticity and linked to fundamental biological processes, including proliferation, survival, programmed cell death, cell adhesion, and migration [44]. In addition, Chen et al. identified that HIF-1α directly binds to the *MKLN1-AS* (long noncoding *MKLN1* antisense RNA) promoter region, increasing its expression in pancreatic cancer [45]. *MKLN1-AS* acts as a competitive endogenous RNA, binding to miR-185-5p and preventing it from inhibiting *TEAD1* [45]. This results in upregulated TEAD1 expression, which subsequently promotes pancreatic cancer cell proliferation, migration, and tumor growth [45].

Our deep learning-based drug repositioning analysis targeting MKLN1 and MMP14, using their amino acid sequences, identified GW0742, a PPARβ/δ agonist, as the most notable drug-repositioning candidate. Several studies have provided evidence supporting GW0742 as a drug-repositioning candidate for cancer therapy. Wagner et al. investigated the effects of GW0742 treatment on lung cancer in vitro [46]. Given the high expression of PPARβ/δ in endothelial cells, it could function as a pro-angiogenic signaling molecule. Therefore, the agonist effect of GW0742 on PPARβ/δ may deteriorate cancer progression, including metastasis. However, this finding was challenged by two other studies on GW0742 as an anti-cancer treatment for CRC. Foreman et al. demonstrated that GW0742 inhibits human colon cancer cells in a dose-dependent manner [47]. Although heterogeneity among the types of human cancer cells (HT29, DLD1, and RKO) was observed, the anti-cancer effect of PPARβ/δ ligand

activation by GW0742 against human colon cancer cells was confirmed. These results, although contraindicated in previous research, have been replicated in subsequent studies [48, 49]. Similarly, Marin et al. reported results consistent with those obtained by Foreman et al. [50]. In a mouse model of colon carcinogenesis, continuous treatment with GW0742 substantially reduced the incidence, multiplicity, and size of tumors. Additionally, GW0742 demonstrated a dose-dependent effect on tumor apoptosis in the mouse model.

As the last step of our research, we conducted a structure-based analysis of MMP14 with different alleles [G (allele frequency = 0.98) and A (allele frequency = 0.02)] at rs3751489 to investigate its effect as an exonic SNP. This analysis revealed significant structural differences due to the amino acid change at position 431 from arginine to histidine. This change may have affected the different binding scores of GW0742 for MMP14 with the reference and alternate alleles. Given that the reference allele of rs3751489 was associated with an increased risk of CRC development, as indicated by the negative beta value, the high binding affinity of GW0742 for MMP14 with the reference allele underscores the potential of GW0742 as a candidate for anti-CRC therapy.

Our study has two strengths. First, we comprehensively interpreted the GWAS results, integrating post-GWAS functional analysis with structural analysis of the target proteins. This comprehensive approach allowed us to identify profound biological and clinical insights beyond the initial GWAS findings. Second, we discovered a new drug-repositioning candidate and prioritized it for CRC treatment. However, our study also has three limitations. First, the size of the study population in the discovery study was relatively small compared to that in the replication study. This limitation stemmed from the matching process based on sex and age. To address this limitation and increase statistical power, we adopted an alternative selection method for the replication study, which omitted the age-matching steps used in the discovery study and allowed for extended replication beyond specific age groups. Second, 3.1% of the individuals in the control group (46 individuals) in the discovery study reported a history of cancer diagnosis via a self-reported questionnaire. These individuals were not excluded from the current study due to the challenge of assessing clinical and recall accuracy, which may have confounded and underestimated the results of the discovery study. Finally, the efficacy of GW0742 in targeting MMP14 and MKLN1 for CRC was not validated in vivo or in vitro. Also, the results of computational drug repositioning were not validated experimentally. Further experimental studies are needed to confirm the effects and experimental validity

Yun *et al. Journal of Translational Medicine*   (2025) 23:543

Page 12 of 13

of GW0742 on CRC and its association with MMP14 and MKLN1, thereby strengthening our findings.

Conclusions

Our study identified two key CRC susceptibility genes, *MMP14* and *MKLN1*, providing a genetic basis for CRC development within the Korean population. Additionally, we discovered the potential of GW0742, a PPARβ/δ agonist, as a promising repurposed drug for CRC treatment. The high binding affinity of GW0742 for MMP14, particularly with the reference allele associated with increased CRC risk, underscores its potential efficacy. Although further validation is necessary, our research lays the foundation for the development of novel CRC treatment options.

Abbreviations

CADD	Combined annotation-dependent depletion
GWAS	Genome-wide association study
GTEx	Genotype-tissue expression
HWE	Hardy–Weinberg equilibrium
KOGES	Korean Genome and Epidemiology Study
LD	Linkage disequilibrium
MAF	Minor allele frequency
pLDDT	Predicted local distance difference test
RDB	RegulomeDB
RMSD	Root mean square deviation
rsq	R-squared statistic
SMILES	Simplified molecular input line-entry system
SNP	Single-nucleotide polymorphism
SNUH	Seoul National University Hospital
TPM	Transcript per million

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-025-06491-6.

> Supplementary material 1: Table S1. Characteristics of participants included in the replication study
>
> Supplementary material 2: Table S2. Entire signals and functional annotations from GWAS on colorectal cancer risk

the study had no role in the design and conduct of the study and were not involved in the collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Availability of data and materials
The demographic, clinicopathological, and SNP genotype data used in this study were provided by the Seoul National University Hospital, Chonnam National University Hwasun Hospital, National Biobank of Korea (NBK), and the Center for Disease Control and Prevention, Republic of Korea. To protect patients' privacy and ethical issues according to the content of our IRB, patients' clinical and genetic data are not publicly available. Controls' data can be requested via the website(https://biobank.nih.go.kr), and it is strictly prohibited to disclose them without permission. Data in this study were from the Korean Genome and Epidemiology Study (KoGES; 6635-302), National Institute of Health, Korea Disease Control and Prevention Agency, Republic of Korea.

Declarations

Ethics approval and consent to participate
The study protocol was approved by the Korea National Institute of Health and the Institutional Review Board (IRB: CBNU-202306-HR-0153) of Chungbuk National University. 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.

Author details
[1]College of Pharmacy, Chungbuk National University, Cheongju, Chungbuk, Korea. [2]Department of Preventive Medicine, Chonnam National University Medical School, Hwasun-Gun, Jeollanam-Do, 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.

References
1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024;74:12–49.
2. Hossain MS, Karuniawati H, Jairoun AA, Urbi Z, Ooi J, John A, Lim YC, Kibria KMK, Mohiuddin AKM, Ming LC, et al. Colorectal cancer: a review of carcinogenesis, global epidemiology, current challenges, risk factors, preventive and treatment strategies. Cancers (Basel). 2022;14:1732.
3. 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.
4. Jun S, Park H, Kim UJ, Choi EJ, Lee HA, Park B, Lee SY, Jee SH, Park H. Cancer risk based on alcohol consumption levels: a comprehensive systematic review and meta-analysis. Epidemiol Health. 2023;45:e2023092.
5. An S, Park S. Association of physical activity and sedentary behavior with the risk of colorectal cancer. J Korean Med Sci. 2022;37:e158.
6. Khil H, Kim SM, Hong S, Gil HM, Cheon E, Lee DH, Kim YA, Keum N. Time trends of colorectal cancer incidence and associated lifestyle factors in South Korea. Sci Rep. 2021;11:2413.

7. Mucci LA, Hjelmborg JB, Harris JR, Czene K, Havelick DJ, Scheike T, Graff RE, Holst K, Moller S, Unger RH, et al. Familial risk and heritability of cancer among twins in nordic countries. JAMA. 2016;315:68–76.

8. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish family-cancer database. Int J Cancer. 2002;99:260–6.

9. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000;343:78–85.

10. Huyghe JR, Bien SA, Harrison TA, Kang HM, Chen S, Schmit SL, Conti DV, Qu C, Jeon J, Edlund CK, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat Genet. 2019;51:76–87.

11. Jia WH, Zhang B, Matsuo K, Shin A, Xiang YB, Jee SH, Kim DH, Ren Z, Cai Q, Long J, et al. Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. Nat Genet. 2013;45:191–6.

12. Study C, Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, Chandler I, Vijayakrishnan J, Sullivan K, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet. 2008;40:1426–35.

13. Pierce SE, Booms A, Prahl J, van der Schans EJC, Tyson T, Coetzee GA. Post-GWAS knowledge gap: the how, where, and when. NPJ Parkinsons Dis. 2020;6:23.

14. Ke J, Lou J, Chen X, Li J, Liu C, Gong Y, Yang Y, Zhu Y, Zhang Y, Gong J. Identification of a potential regulatory variant for colorectal cancer risk mapping to chromosome 5q31.1: a Post-GWAS study. PLoS One. 2015;10:e0138478.

15. Wang HM, Chang TH, Lin FM, Chao TH, Huang WC, Liang C, Chu CF, Chiu CM, Wu WY, Chen MC, et al. A new method for post Genome-Wide Association Study (GWAS) analysis of colorectal cancer in Taiwan. Gene. 2013;518:107–13.

16. Rosic G. Cancer signaling, cell/gene therapy, diagnosis and role of nanobiomaterials. Adv Biol Earth Sci. 2024. https://doi.org/10.62476/abes9s11.

17. Nasibova A, Cabbarova Z, Khalilov R. The modern perspectives of nanomaterial applications in cancer treatment and drug delivery. Adv Biol Earth Sci. 2024;9:330.

18. Bell N, Uffelmann E, van Walree E, de Leeuw C, Posthuma D. Using genome-wide association results to identify drug repositioning candidates. medRxiv. 2022. https://doi.org/10.1101/2022.09.06.22279660.

19. Irham LM, Wong HS, Chou WH, Adikusuma W, Mugiyanto E, Huang WC, Chang WC. Integration of genetic variants and gene network for drug repositioning in colorectal cancer. Pharmacol Res. 2020;161:105203.

20. 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.

21. Moon S, Kim YJ, Han S, Hwang MY, Shin DM, Park MY, Lu Y, Yoon K, Jang HM, Kim YK, et al. The Korea Biobank array: design and identification of coding variants associated with blood biochemical traits. Sci Rep. 2019;9:1382.

22. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48:1284–7.

23. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.

24. 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.

25. 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.

26. 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 Proteomics Bioinform. 2021;19:619–28.

27. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 1826;2017:8.

28. Rentzsch P, Schubach M, Shendure J, Kircher M. CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. Genome Med. 2021;13:31.

29. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22:1790–7.

30. Carithers LJ, Moore HM. The genotype-tissue expression (GTEx) project. Biopreserv Biobank. 2015;13:307–8.

31. Huang K, Fu T, Glass LM, Zitnik M, Xiao C, Sun J. DeepPurpose: a deep learning library for drug-target interaction prediction. Bioinformatics. 2021;36:5545–7.

32. Corsello SM, Bittker JA, Liu Z, Gould J, McCarren P, Hirschman JE, Johnston SE, Vrcic A, Wong B, Khan M, et al. The drug repositioning hub: a next-generation drug library and information resource. Nat Med. 2017;23:405–8.

33. Volkamer A, Kuhn D, Rippmann F, Rarey M. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. Bioinformatics. 2012;28:2074–5.

34. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–9.

35. Schrodinger, LLC: The PyMOL Molecular Graphics System, Version 1.8. 2015.

36. Cerezo M, Sollis E, Ji Y, Lewis E, Abid A, Bircan KO, Hall P, Hayhurst J, John S, Mosaku A, et al. The NHGRI-EBI GWAS Catalog: standards for reusability, sustainability and diversity. Nucleic Acids Res. 2025;53:D998–1005.

37. Kim HM, Kim TI. Screening and surveillance for hereditary colorectal cancer. Intest Res. 2024;22:119–30.

38. Kufareva I, Abagyan R. Methods of protein structure comparison. Methods Mol Biol. 2012;857:231–57.

39. Hua H, Li M, Luo T, Yin Y, Jiang Y. Matrix metalloproteinases in tumorigenesis: an evolving paradigm. Cell Mol Life Sci. 2011;68:3853–68.

40. Cui G, Cai F, Ding Z, Gao L. MMP14 predicts a poor prognosis in patients with colorectal cancer. Hum Pathol. 2019;83:36–42.

41. Ragusa S, Prat-Luri B, Gonzalez-Loyola A, Nassiri S, Squadrito ML, Guichard A, Cavin S, Gjorevski N, Barras D, Marra G, et al. Antiangiogenic immunotherapy suppresses desmoplastic and chemoresistant intestinal tumors in mice. J Clin Invest. 2020;130:1199–216.

42. Adams JC, Seed B, Lawler J. Muskelin, a novel intracellular mediator of cell adhesive and cytoskeletal responses to thrombospondin-1. EMBO J. 1998;17:4964–74.

43. Jung SY, Papp JC, Sobel EM, Zhang ZF. Mendelian randomization study: the association between metabolic pathways and colorectal cancer risk. Front Oncol. 2020;10:1005.

44. Huffman N, Palmieri D, Coppola V. The CTLH complex in cancer cell plasticity. J Oncol. 2019;2019:4216750.

45. Chen J, Li L, Feng Y, Zhao Y, Sun F, Zhou X, Yiqi D, Li Z, Kong F, Kong X. MKLN1-AS promotes pancreatic cancer progression as a crucial downstream mediator of HIF-1alpha through miR-185-5p/TEAD1 pathway. Cell Biol Toxicol. 2024;40:30.

46. Wagner KD, Du S, Martin L, Leccia N, Michiels JF, Wagner N. Vascular PPARbeta/delta promotes tumor angiogenesis and progression. Cells. 2019;8:1623.

47. Foreman JE, Chang WC, Palkar PS, Zhu B, Borland MG, Williams JL, Kramer LR, Clapper ML, Gonzalez FJ, Peters JM. Functional characterization of peroxisome proliferator-activated receptor-beta/delta expression in colon cancer. Mol Carcinog. 2011;50:884–900.

48. Hollingshead HE, Borland MG, Billin AN, Willson TM, Gonzalez FJ, Peters JM. Ligand activation of peroxisome proliferator-activated receptor-beta/delta (PPARbeta/delta) and inhibition of cyclooxygenase 2 (COX2) attenuate colon carcinogenesis through independent signaling mechanisms. Carcinogenesis. 2008;29:169–76.

49. Harman FS, Nicol CJ, Marin HE, Ward JM, Gonzalez FJ, Peters JM. Peroxisome proliferator-activated receptor-delta attenuates colon carcinogenesis. Nat Med. 2004;10:481–3.

50. Marin HE, Peraza MA, Billin AN, Willson TM, Ward JM, Kennett MJ, Gonzalez FJ, Peters JM. Ligand activation of peroxisome proliferator-activated receptor beta inhibits colon carcinogenesis. Cancer Res. 2006;66:4394–401.

Publisher's Note