BMC Gastroenterology

RESEARCH Open Access

Deciphering the integrated immunogenomic landscape of colorectal cancer: insights from Mendelian randomization and immune-stratified molecular subtyping

Ke-Jie He[1*] and Guoyu Gong[2]

Abstract

Purpose This study aimed to decipher the intricate interplay between the immune landscape and CRC pathogenesis, elucidating how distinct immunophenotypes causally influence disease susceptibility and stratify patient outcomes.

Methods We obtained the immunocyte phenotypes and CRC data from their respective genome-wide association studies. The primary analysis used the inverse variance weighting (IVW) method. We also simultaneously employed MR-Egger, weighted mode, simple mode, and weighted median approaches to strengthen the findings. Consensus clustering stratified 619 TCGA CRC patients by immunome expression. Functional assays examined the tumor suppressor GPD1L.

Results The IVW MR analysis identified 17 immunocyte phenotypes positively potentially associated with increased CRC risk ($P < 0.05$, OR > 1), and 18 phenotypes negatively potentially associated with decreased CRC risk ($P < 0.05$, OR < 1). These associations were not confounded by heterogeneity or horizontal pleiotropy ($P > 0.05$). Reverse MR analysis further revealed 4 additional immunocyte phenotypes positively potentially associated with CRC ($P < 0.05$, OR > 1). Clustering resolved prognostic C1/C2 subtypes dependent on coordinated immunophenotypic programs. GPD1L knockdown promoted CRC cell proliferation.

Conclusions Genetic interrogation delineated causal immunome-CRC relationships at single-cell resolution. Immune-stratified CRC subtyping stratified patient outcomes. GPD1L exhibited tumor-suppressive functions. Our findings establish an integrated immunogenomic framework elucidating CRC pathogenesis with implications for precision immunotherapies.

Keywords Colorectal cancer, Mendelian randomization, Immunophenotypes, Immune-stratified molecular subtypes, GPD1L, Tumor suppressor

*Correspondence:
Ke-Jie He
hekejie@stu.xmu.edu.cn
[1]The Quzhou Affiliated Hospital of Wenzhou Medical University, Quzhou People's Hospital, Quzhou city, Zhejiang Province, China
[2]School of Medicine, Xiamen University, Xiamen, China

Introduction

Colorectal cancer (CRC) poses a significant global health burden and immune evasion presents a key challenge in treatment [1, 2]. However, the precise interactions between diverse immune cell populations such as T cell subsets, B cells, and others within the tumor microenvironment and their roles in CRC development remain unclear, especially at the single-cell level [3, 4].

Colorectal cancer relentlessly claims lives, posing a formidable threat. Previous studies have shown that cytotoxic T cells and memory T cells can effectively eliminate cancer cells, while regulatory T cells and myeloid-derived suppressor cells suppress anti-tumor immunity [5]. While previous studies have provided valuable insights into the general immune responses in colorectal cancer (CRC), the precise functions of distinct immune cell subpopulations in influencing disease risk and prognosis warrant further investigation. The pathogenic mechanisms underlying this prevalent malignancy remain enigmatic, with the immune system playing an obscure pivotal role. Recent studies have revealed the causal associations between 731 immune cell phenotypes and the risk of esophageal cancer [6], prostate cancer [7], and kidney cancer [8], suggesting that certain characteristic immune cell populations are potentially associated with tumor occurrence and development. However, reports on the causal associations between 731 immune cell phenotypes and colorectal cancer, as well as the exploration of key gene targets of characteristic immune cell populations, are relatively rare. This study aimed to address how variations in the immunological milieu, characterized by diverse immune cell populations, causally contribute to colorectal cancer development and progression, and to explore the corresponding gene targets of characteristic immune cell representative populations and conduct related experimental verification. Tackling these unresolved questions, this study harnesses cutting-edge analytical techniques and large-scale multi-omics data to decipher the immunogenomic landscape of colorectal cancer in unprecedented detail.

First comprehensively delineating causal immunophenotype-colorectal cancer links through bidirectional Mendelian randomization, we identified 17 phenotypes increasing and 18 decreasing risk. From risk-enhancing populations, we prioritized top 5, annotating variants to unveil 48 putative regulatory genes. Novel immune-based subtyping stratified patients into prognostically divergent subtypes.

Materials and methods

Study design

We conducted a two-sample mendelian randomization (MR) analysis to assess the causal effects of 731 immune cell signatures (across seven groups) on colorectal cancer (CRC) risk. MR leverages genetic variation as instrumental variables (IVs) for establishing causal inferences, provided that three key assumptions are met: (1) the genetic variants are robustly potentially associated with the exposure of interest; (2) they are not potentially associated with potential confounders of the exposure-outcome relationship; and (3) they influence the outcome only through the exposure and not via alternate pathways.

The underlying genome-wide association studies included in our two-sample MR analysis were approved by their respective Institutional Review Boards, and all participants provided written informed consent. For each of the 731 immune cell signatures, we identified statistically strong and independent genetic instruments from published genome-wide association studies. We then tested their associations with CRC using another independent dataset to avoid sample overlap-induced bias.

Through judiciously applying the MR design and rigorously assessing the instrument strength and exclusivity assumptions, we sought to elucidate potential causal relationships between variations in immune cell compositions and CRC susceptibility. Our findings provide novel genetic evidence for how select immune components may influence CRC etiology, with implications for immunotherapeutic and prevention strategies against this common malignancy.

Data sources

The genome-wide association study (GWAS) summary statistics for colorectal cancer (CRC) were obtained from the previous study by Sakaue et al. including 470,002 individuals of European descent [9]. Publicly available GWAS summary data for the 731 immunophenotypes were accessed from the GWAS Catalog (accession numbers GCST0001391 through GCST0002121), encompassing absolute cell counts (AC) ($n = 118$), median fluorescence intensities (MFI) reflective of surface antigen levels ($n = 389$), morphological parameters (MP) ($n = 32$) and relative cell counts (RC) ($n = 192$). Specifically, the MFI, AC and RC datasets contained B cell, conventional dendritic cell (cDC), mature T cell subsets, monocyte, myeloid cell, TBNK (T cell, B cell, natural killer cell) and regulatory T cell (Treg) characteristics, while the MP category covered cDC and TBNK parameters.

The original immune trait GWAS leveraged genotype information on ~22 million SNPs from high-density arrays in 3,757 European individuals without overlap between cohorts [10, 11]. Imputation was performed using a Sardinian reference panel, followed by association testing with adjustment for covariates including sex, age and age-squared. This ensured no confounding from hidden population substructure and provided robust instrumental variables for our causal inference analyses of CRC susceptibility.

Trait	OR(95%CI)	pvalue
IgD- CD38br AC	0.9461(0.9180 to 0.9751)	0.0003
CD27 on CD20-	0.9537(0.9251 to 0.9832)	0.0023
CD25++ CD8br %T cell	0.9348(0.8923 to 0.9794)	0.0046
SSC-A on B cell	1.0735(1.0196 to 1.1302)	0.0070
DN (CD4-CD8-) %leukocyte	0.9475(0.9092 to 0.9875)	0.0106
IgD- CD38dim %lymphocyte	1.0432(1.0094 to 1.0781)	0.0119
CD11b on basophil	1.0414(1.0086 to 1.0751)	0.0129
IgD+ CD24+ %B cell	0.9318(0.8809 to 0.9856)	0.0136
CD14 on CD14+ CD16- monocyte	0.9432(0.9002 to 0.9882)	0.0139
Activated & resting Treg %CD4+	1.0505(1.0098 to 1.0929)	0.0145
Plasmacytoid DC AC	0.9745(0.9542 to 0.9952)	0.0160
CD39+ secreting Treg AC	0.9770(0.9587 to 0.9957)	0.0161
BAFF-R on naive-mature B cell	1.0280(1.0052 to 1.0512)	0.0156
Myeloid DC %DC	1.0492(1.0092 to 1.0908)	0.0155
BAFF-R on B cell	1.0281(1.0042 to 1.0525)	0.0207
CCR2 on CD62L+ myeloid DC	1.0525(1.0071 to 1.0999)	0.0229
CD14 on CD33br HLA DR+ CD14dim	0.9592(0.9249 to 0.9949)	0.0253
BAFF-R on IgD+	1.0253(1.0031 to 1.0481)	0.0254
CD28 on secreting Treg	1.0310(1.0037 to 1.0591)	0.0260
BAFF-R on IgD+ CD24-	1.0247(1.0026 to 1.0473)	0.0283
CD39+ resting Treg AC	1.0334(1.0036 to 1.0640)	0.0277
CD39+ CD4+ AC	0.9697(0.9433 to 0.9969)	0.0294
CD25hi CD45RA- CD4 not Treg AC	0.9741(0.9509 to 0.9979)	0.0332
CD86+ plasmacytoid DC %DC	0.9685(0.9404 to 0.9975)	0.0334
CD28 on CD39+ secreting Treg	0.9591(0.9227 to 0.9969)	0.0343
CM CD8br %T cell	1.0610(1.0034 to 1.1218)	0.0374
Transitional %B cell	0.9503(0.9055 to 0.9972)	0.0382
CD24 on transitional	0.9475(0.9000 to 0.9974)	0.0396
HLA DR on CD14- CD16+ monocyte	1.0436(1.0020 to 1.0870)	0.0397
CX3CR1 on CD14- CD16+ monocyte	1.0343(1.0015 to 1.0682)	0.0403
CD8 on CD28- CD8br	1.0539(1.0019 to 1.1087)	0.0421
CD28 on CD39+ CD4+	0.9749(0.9512 to 0.9993)	0.0435
IgD+ CD38br %B cell	0.9623(0.9270 to 0.9990)	0.0443
CD25hi CD45RA- CD4 not Treg %CD4+	0.9693(0.9398 to 0.9998)	0.0484
FSC-A on HLA DR+ T cell	1.0338(1.0002 to 1.0685)	0.0487

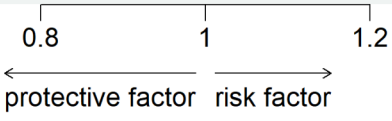

Fig. 1 Identifying Causal Immune Cell Signatures in Colorectal Cancer Susceptibility Through Two-Sample Mendelian Randomization. Forest plots were used to illustrate the causal associations between colorectal cancer (CRC) and immune cell traits. The analysis employed inverse variance weighting (IVW) and presented confidence intervals (CI) to depict the magnitude and precision of the observed relationships

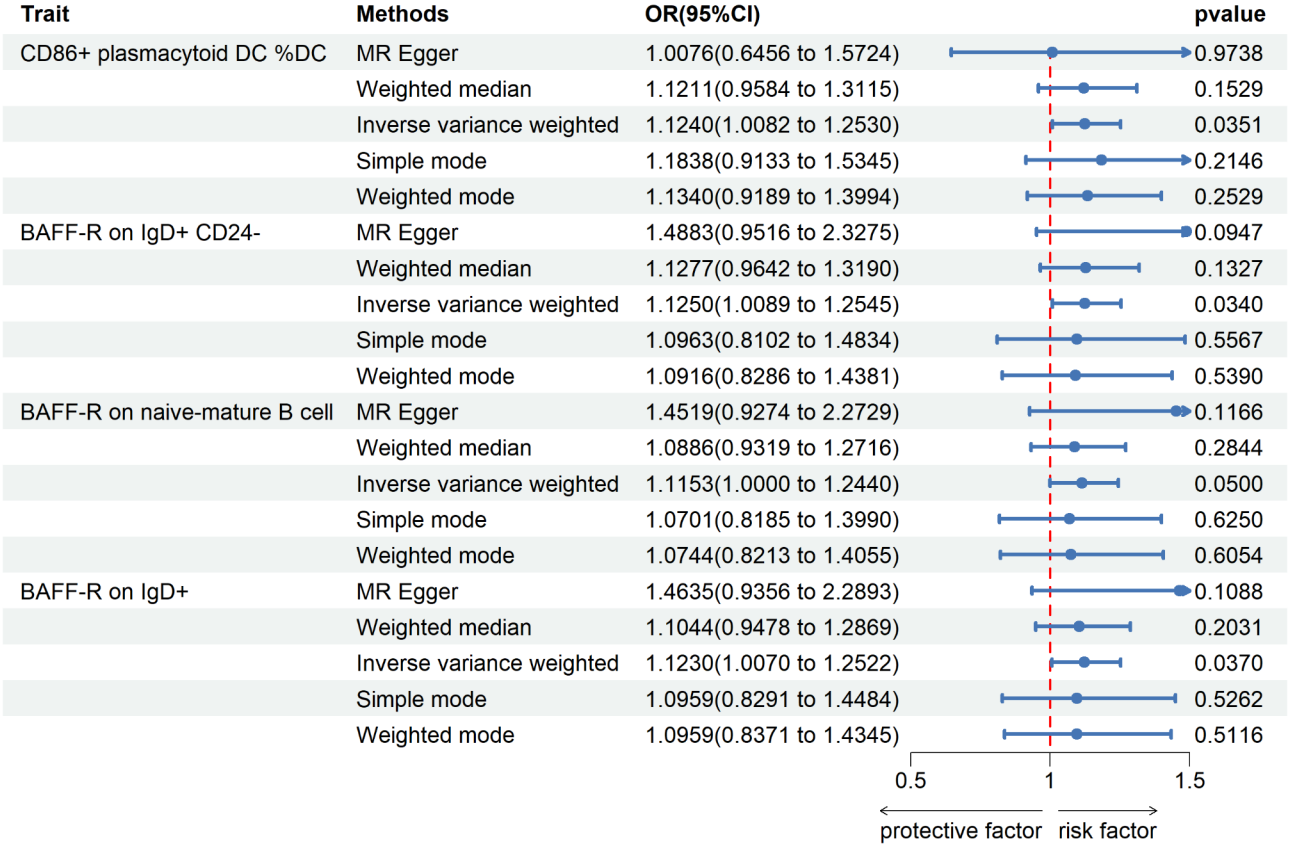

Fig. 2 Evaluating Putative Immune Modifications Potentially associated with Colorectal Cancer Onset using Two-Sample Mendelian Randomization. Forest plots were used to illustrate the causal associations between colorectal cancer (CRC) and immune cell traits. The analysis employed inverse variance weighting (IVW) and presented confidence intervals (CI) to depict the magnitude and precision of the observed relationships

Selection of instrumental variables (IVs)

We set the significance threshold for instrumental variants (IVs) of each immune trait to $P<1\times10-5$, consistent with prior work. Linkage disequilibrium (LD) pruning was performed using PLINK software (v1.90) to identify independent SNPs with an LD r2 threshold of <0.1 within 500 kb. LD r2 values were estimated based on the 1000 Genomes Project reference panel [10].

For colorectal cancer (CRC), we adopted a stringent genome-wide significance threshold of $P<5\times10^{-8}$ to ensure robust variant-outcome associations. LD was leveraged to identify independent SNPs potentially associated with CRC risk at $R2<0.005$. This filtering retained 28 genetic variants for subsequent analyses as instruments representing CRC predisposition. Once the instrumental variants were finalized, harmonization of GWAS datasets was achieved by extracting information from the CRC genome-wide study that corresponded to each retained variant. This allowed testing of CRC risk specifically mediated by the genetic instruments identified for immune cell profiles, meeting the assumptions for causal effect estimation using two-sample Mendelian randomization.

SNP annotation

Genetic variant annotation was performed using the web-based g: SNPense tool. g: SNPense maps human SNPs denoted by rs IDs to corresponding gene names by integrating chromosomal positions and predicted functional consequences from Ensembl Variation. Specifically, it retrieves relevant genic context for input variants that intersect protein-coding gene loci as defined by the reference Ensembl gene set. Leveraging g: SNPense facilitated characterization of the putative functional roles and biological pathways of instrumental variants potentially associated with immune cell traits and CRC risk. Such annotation provided molecular insight into the mechanisms whereby specific immune cell populations may exert influence on colorectal neoplasia susceptibility as inferred through our MR analyses. Overall, g:SNPense represented a powerful online resource that augmented interpretation of genetic findings by mapping variants to their cognate gene targets encoded within the human reference genome.

Construct prognostic signature of immunophenotypes

We constructed an immune signature-based prognostic model using The Cancer Genome Atlas (TCGA)

Table 1 SNP annotation

id	chr	start	end	strand	gene_names
rs62034403	16	78,334,703	78,334,703	+	WWOX
rs960502	1	55,146,605	55,146,605	+	USP24
rs138016528	16	72,077,935	72,077,935	+	TXNL4B
rs56392104	16	11,734,711	11,734,711	+	TXNDC11
rs75837226	11	68,046,204	68,046,204	+	TCIRG1
rs76882749	1	54,066,457	54,066,457	+	TCEANC2
rs62285107	4	1,729,552	1,729,552	+	TACC3
rs139258218	1	1.18E + 08	1.18E + 08	+	SPAG17
rs173262	16	12,136,304	12,136,304	+	SNX29
rs61735519	22	45,354,086	45,354,086	+	SMC1B
rs461709	19	47,163,454	47,163,454	+	SAE1
rs11121500	1	6,219,556	6,219,556	+	RNF207
rs13410035	2	1.6E + 08	1.6E + 08	+	RBMS1
rs8003606	14	68,506,533	68,506,533	+	RAD51B
rs35069110	7	1.58E + 08	1.58E + 08	+	PTPRN2
rs6139531	20	4,726,420	4,726,420	+	PRND
rs7819099	8	1.44E + 08	1.44E + 08	+	PLEC
rs2588120	8	17,607,754	17,607,754	+	PDGFRL
rs17780836	5	59,327,108	59,327,108	+	PDE4D
rs76516141	3	33,832,763	33,832,763	+	PDCD6IP
rs12573934	11	4,681,337	4,681,337	+	OR51E2
rs115044799	3	47,011,645	47,011,645	+	NRADDP
rs541367304	1	1.98E + 08	1.98E + 08	+	NEK7
rs2011807	16	83,942,959	83,942,959	+	MLYCD, OSGIN1
rs61924635	12	25,497,711	25,497,711	+	LMNTD1
rs78708224	7	26,415,621	26,415,621	+	LINC02981
rs6724848	2	38,451,509	38,451,509	+	LINC02613
rs72784338	5	1.07E + 08	1.07E + 08	+	LINC01950
rs11014701	10	25,710,311	25,710,311	+	LINC00836
rs144261817	5	40,009,547	40,009,547	+	LINC00603
rs3813658	2	30,250,656	30,250,656	+	LBH
rs6096232	20	51,021,437	51,021,437	+	KCNG1
rs2708609	7	1.12E + 08	1.12E + 08	+	IFRD1
rs13157900	5	80,359,152	80,359,152	+	HNRNPA1P12
rs9269074	6	32,473,015	32,473,015	+	HLA-DRB9
rs78689026	6	30,772,261	30,772,261	+	HCG20
rs6763382	3	32,114,791	32,114,791	+	GPD1L
rs6751481	2	38,670,668	38,670,668	+	GALM
rs6687275	1	1.62E + 08	1.62E + 08	+	FCGR3A
rs72879624	11	5,123,161	5,123,161	+	ENSG00000290651
rs61802333	1	1.62E + 08	1.62E + 08	+	ENSG00000289768
rs13006059	2	42,795,580	42,795,580	+	ENSG00000289082
rs137997033	6	1.24E + 08	1.24E + 08	+	ENSG00000285941
rs969808	8	71,687,167	71,687,167	+	ENSG00000254277
rs139325044	16	59,149,646	59,149,646	+	ENSG00000245768
rs12537451	7	49,231,213	49,231,213	+	ENSG00000234686
rs116794175	1	1.94E + 08	1.94E + 08	+	ENSG00000227240
rs6733868	2	25,276,998	25,276,998	+	DNMT3A
rs10977191	9	862,889	862,889	+	DMRT1
rs186940149	13	35,954,842	35,954,842	+	DCLK1
rs75442393	8	3,060,512	3,060,512	+	CSMD1
rs76225340	8	3,797,463	3,797,463	+	CSMD1
rs73009664	3	2,679,929	2,679,929	+	CNTN4

Table 1 (continued)

id	chr	start	end	strand	gene_names
rs113376235	16	75,480,813	75,480,813	+	CHST6
rs140690408	16	53,305,092	53,305,092	+	CHD9
rs7425430	2	86,829,357	86,829,357	+	CD8B
rs35587265	3	46,217,805	46,217,805	+	CCR3
rs10868399	9	86,235,779	86,235,779	+	C9orf153
rs139142972	18	31,651,139	31,651,139	+	B4GALT6
rs1317900	2	1.31E + 08	1.31E + 08	+	ARHGEF4
rs408686	2	38,950,179	38,950,179	+	ARHGEF33
rs11109601	12	98,780,254	98,780,254	+	ANKS1B
rs189927142	10	27,227,826	27,227,826	+	ACBD5

colorectal cancer (CRC) dataset. Clinical information including survival outcomes were obtained for 619 patients alongside RNA-seq abundance estimates (level 3). Transcript per million (TPM) values were log2-transformed after adding a pseudocount of 1 to reduce skewness. Feature selection was performed using the least absolute shrinkage and selection operator (LASSO) Cox regression with 10-fold cross-validation, implemented via the R glmnet package. This penalized approach identified a parsimonious set of immune transcripts predictive of CRC survival. Prognostic value was evaluated using Kaplan-Meier analysis and the log-rank test, with curves and 95% confidence intervals generated by the R survival package. The detailed information of TCGA-CRC is provided in Supplementary Material 9.

Signatures were dichotomized about their median to stratify patients into high- and low-risk groups for direct survival comparison. Statistical significance was set at $P < 0.05$. Our integrated computational workflow leveraged TCGA molecular profiles together with clinical follow-up to develop an immunogenomic classifier of CRC outcome. It provides a basis for future studies mechanistically linking immune landscapes to prognosis in colorectal tumorigenesis.

Identification of potential subtypes

We applied consensus clustering to elucidate robust molecular subgroups within colorectal cancer (CRC) patients. 80% of samples ($n = 495$) from The Cancer Genome Atlas were designated as the discovery cohort. ConsensusClusterPlus (v1.54.0) iteratively clustered this cohort 100 times, stabilizing consensus partitions.

Gene expression heatmaps were generated using pheatmap (v1.0.12) and focused on differentially variable genes with standard deviations over 0.1. When over 1,000 genes met this criterion, we retained the top 25% most varying. The remaining 20% of samples ($n = 124$) comprised an internal validation cohort. We evaluated whether the clusters identified in the discovery samples could correctly classify the held-out validation cases. This rigorous consensus clustering approach leveraged multiple resampling of the majority of CRC patients to reveal robust and reproducible intrinsic subgroups. Independent validation testing of the cluster solution on a separate data portion demonstrated its generalizability for stratifying CRC molecular phenotypes with functional implications.

Cell culture and transfection

The SW480 and HCT116 colorectal cancer cell line was obtained from Procell Life Science & Technology (Wuhan, China) and tested free of mycoplasma contamination. Cells were cultured in Dulbecco's Modified Eagle Medium (DMEM; Gibco) supplemented with 10% fetal bovine serum (FBS; Gibco), 100 U/mL penicillin and 100 μg/mL streptomycin (Thermo Fisher Scientific, Waltham, MA) at 37 °C in a humidified atmosphere containing 5% CO2. All human cell lines have been authenticated using STR (or SNP) profiling within the last three years has been included.

GPD1L gene silencing was achieved using lipid-based transfection of small interfering RNA (siRNA). Specifically, SW480 and HCT116 cells were transfected with either siRNA targeting GPD1L (si-GPD1L) or a non-targeting control (si-NC) using Lipofectamine 2000 reagent (Thermo Fisher Scientific) following the manufacturer's protocol. Cells were harvested 24 h post-transfection for subsequent analyses. This established an efficient in vitro system to investigate the roles of GPD1L in colorectal cancer cell biology using loss-of-function approaches with siRNA-mediated knockdown. The optimized culture conditions and transfection methodology ensured viability and reproducibility for downstream molecular and functional experiments. All of the primer sequences and small interfering RNA sequences are listed in Supplementary Material 5.

Cell proliferation

To assess the role of GPD1L in CRC cell proliferation, we performed colony formation and flow cytometry assays. For colony formation, SW480 and HCT116 cells transfected with either si-GPD1L or si-NC were seeded at 500

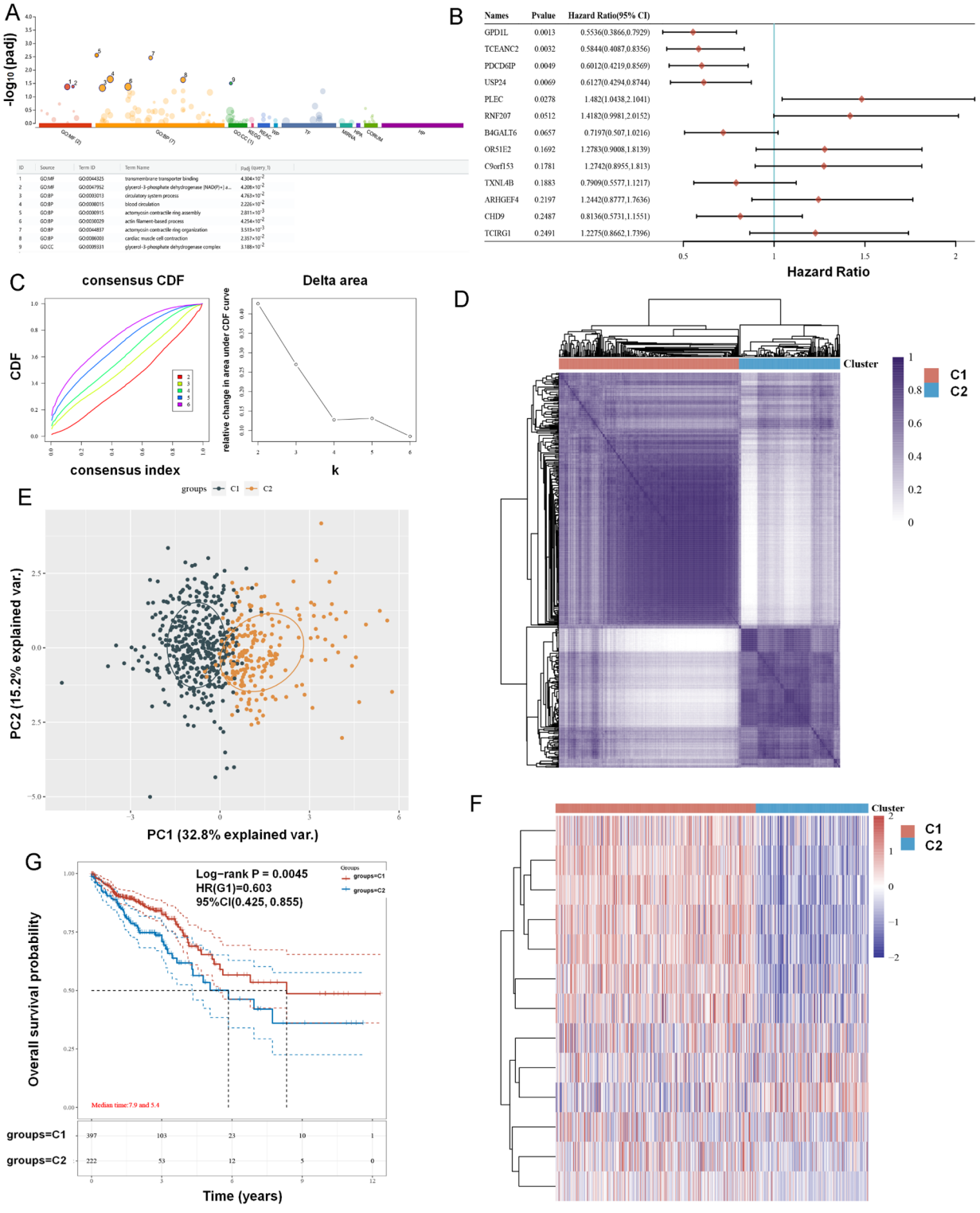

Fig. 3 Immunophenotype-Potentially associated Gene Signatures. (**a**) Pathway enrichment analysis of the 48 genes potentially potentially associated with immunophenotype risk loci; (**b**) Non-negative matrix factorization consensus clustering of the 13 gene signatures at k = 2 optimally partitioned patients per cumulative distribution function and area under the function curve; (**c**) Initial unsupervised classification separated TCGA CRC cohorts into C1 and C2 subtypes by the consensus clustering map; (**d**) Patients were divided into C1 and C2 clusters according to the consensus map; (**e**) Principal component analysis further corroborated distinct clustering; (**f**) Differential gene expression heatmaps confirmed molecular divergence between subgroups; (**g**) Kaplan-Meier survival analyses with the log-rank test revealed significantly poorer prognosis for C2 versus C1

cells per well in six-well plates and cultured for 15 days. Colonies were fixed, stained with hematoxylin, and manually counted.

Total RNA was extracted from transfected SW480 and HCT116 cells using TRIzol reagent (Thermo Fisher Scientific). Primer sequences and siRNA targets are listed in Supplementary Material 5. Colony formation assays were conducted in biological triplicate with SW480 and HCT116 cells transfected with either si-GPD1L and si-NC. Colonies were counted using ImageJ software and statistical significance between conditions was determined using two-tailed unpaired Student's t-tests. All experiments were performed in biological triplicates.

Statistical analysis

All analyses were conducted using R 3.5.3 software. The MendelianRandomization R package (version 0.4.3) was used to implement several causal estimation methods, including inverse variance weighting (IVW), MR Egger regression, weighted median and mode-based approaches [12–15]. Cochran's Q test and corresponding *p*-values quantified heterogeneity across genetic instrumental variables. In the presence of significant heterogeneity, a random effects IVW model replaced the default fixed-effects version.

To account for horizontal pleiotropy, MR Egger regression was conducted, with an intercept term different from zero suggesting pleiotropic bias. Additionally, the MR-PRESSO method robustly detected and removed outlying instrumental variables capable of unduly influencing estimates [16, 17]. Scatter plots ascertained the lack of influential outliers, while funnel plots demonstrated robust instrument strengths without asymmetry indicative of pleiotropy or bias. Overall, triangulation across multiple MR techniques and diagnostic plots provided strong evidence for causal relationships reconciled by different underlying assumptions. Rigorous methodological implementation and results appraisal in independent, well-powered immunogenetic summary data resources substantiated hypothetical links between immune profiles and schizophrenia susceptibility.

Results

Identifying causal immune cell signatures in colorectal cancer susceptibility through two-sample mendelian randomization

Two-sample Mendelian randomization was performed to investigate causal associations between immunophenotypes and colorectal cancer (CRC) risk. Inverse variance weighting (IVW) identified 18 immunophenotypes potentially associated with decreased CRC risk and 17 immunophenotypes potentially associated with increased risk (Fig. 1, Supplementary Material 1). The top five immunophenotypes demonstrating protective effects against CRC were: (1) CD24 on transitional B cells (B cell panel; OR 0.89, 95% CI 0.899–0.997, $P = 0.039$), (2) IgD- CD38br absolute counts (B cell panel; OR 0.91, 95% CI 0.918–0.975, $P = 0.0003$), (3) CD14 on CD14 + CD16- monocytes (Monocyte panel; OR 0.90, 95% CI 0.900-0.988, $P = 0.013$), (4) CD25 + + CD8br % T cells (Treg panel; OR 0.89, 95% CI 0.892–0.979, $P = 0.004$), (5) IgD + CD24+ % B cells (B cell panel; OR 0.93, 95% CI 0.880–0.985, $P = 0.013$). The top five immunophenotypes potentially associated with increased CRC risk were: (1) SSC-A on B cells (TBNK panel; OR 1.07, 95% CI 1.019–1.13, $P = 0.006$), (2) CM CD8br % T cells (Maturation stages of T cell panel; OR 1.06, CI 1.003–1.121, $P = 0.037$), (3) CD8 on CD28- CD8br (Treg panel; OR 1.05, CI 1.001–1.108, $P = 0.042$), (4) CCR2 on CD62L + myeloid dendritic cells (cDC panel; OR 1.05, CI 1.007–1.099, $P = 0.022$), (5) Activated & resting Treg %CD4+ (Treg panel; OR 1.05, CI 1.009–1.092, $P = 0.014$). The immunophenotypes demonstrated associations with CRC risk where odds ratios were closer to the null value of 1, suggesting these results require more rigorous validation before corroborating causality. The detailed immune signatures across the seven cell type groups are provided in Supplementary Material 6.

Evaluating putative immune modifications potentially associated with colorectal cancer onset using two-sample mendelian randomization

we applied two-sample Mendelian randomization (MR) to investigate the potential causal effects of colorectal cancer (CRC) onset on various immunophenotypes. Using the inverse variance weighting method, we identified several associations between CRC and specific immune cell markers: (1) Increased CD86 + expression on plasmacytoid dendritic cells (odds ratio [OR] 1.123, $P = 0.035$). (2) Increased BAFF-R expression on IgD + CD24- cells (OR 1.125, $P = 0.034$). (3) Increased BAFF-R expression on naive-mature B cells (OR 1.115, $P = 0.049$). (4) Increased BAFF-R expression on IgD + B cells (OR 1.122, $P = 0.036$). The detailed data supporting these findings are presented in Supplementary Material 2 and Fig. 2. This study adheres to the STROBE-MR (Strengthening the Reporting of Observational Studies in Epidemiology - Mendelian Randomization) reporting guidelines for observational studies using Mendelian randomization.

SNP annotation

We conducted follow-up analysis to annotate single-nucleotide polymorphisms (SNPs) meeting locus-wide significance for the top risk five immunophenotypes conferring elevated colorectal cancer (CRC) risk. Forty-eight host genes putatively linked to CRC pathogenesis were identified across the implicated loci (Table 1). We have

provided the detailed SNP annotation table including the genomic locations of the variants as a Supplementary Material 7.

Characterizing colorectal cancer molecular subtypes based on immunophenotype-potentially associated gene signatures

We conducted pathway enrichment analysis of the 48 genes potentially associated with immunophenotype risk loci, revealing significant overrepresentation of transmembrane transporter binding, glycerol-3-phosphate dehydrogenase [NAD(P)+] activity and circulatory system processes (Fig. 3A). Thirteen genes were selected through prognostic profiling to develop immune-stratified colorectal cancer (CRC) molecular subtypes (Supplementary Material 3). Non-negative matrix factorization consensus clustering of the 13 gene signatures at k = 2 optimally partitioned patients per cumulative distribution function and area under the function curve (Fig. 3B). Initial unsupervised classification separated TCGA CRC cohorts into C1 and C2 subtypes by the consensus clustering map (Fig. 3C, Supplementary Material 4). Patients were divided into C1 and C2 clusters according to the consensus map (Fig. 3D). Differential gene expression heatmaps confirmed molecular divergence between subgroups (Fig. 3F). Principal component analysis further corroborated distinct clustering (Fig. 3E). Kaplan-Meier survival analyses with the log-rank test revealed significantly poorer prognosis for C2 versus C1 ($p < 0.0045$, Fig. 3G).

Characterizing the immune landscape and prognostic signatures of colorectal cancer subtypes based on immunophenotype

We observed downregulated immune checkpoint genes in poor prognosis C2 patients (Fig. 4A). Correlation network analysis suggested that most immune phenotype-related genes are positively correlated with CD4 immune cells (Fig. 4B). Cox regression on immune traits pruned 13 genes for LASSO modeling to derive an overall survival (OS) prognostic signature (Fig. 4C, D). The risk score equation was: Risk score = (-0.2968)*TCEANC2 + (-0.5675)*PDCD6IP + (-0.2528)*USP24 + (0.3435)*PLEC + (0.1323)*RNF207 + (-0.157)*B4GALT6 + (0.1041)*OR51E2 + (0.6024)*C9orf153 + (0.6103)*TXNL4B + (0.0266)*ARHGEF4 + (0.1941)*CHD9 + (-0.1165)*TCIRG1. With λmin = 0.0031, the signature stratified patients by risk of mortality (hazard ratio 2.247, 95% confidence interval 1.56–3.236, log-rank $p = 1.36 \times 10 - 5$). Areas under receiver operating characteristic curves of 0.651, 0.657 and 0.687 validated prognostic accuracy at 1, 3 and 5 years respectively. Kaplan-Meier analysis revealed significantly greater OS in the low-risk versus high-risk group defined by this immunophenotype-derived signature ($p = 1.36 \times 10 - 5$) (Fig. 4E).

GPD1L as a potential therapeutic target in CRC

We focussed on characterizing GPD1L given its association with poor CRC prognosis. GPD1L expression was reduced in tumors versus normal tissue by boxplot and paired analysis (Fig. 5A, B). Kaplan-Meier survival curves showed elevated GPD1L correlated with improved overall survival (OS), disease-specific survival (DSS) and progression-free survival (PFI) (Fig. 5C-E). To interrogate GPD1L's involvement in colon cancer cell proliferation, SW480 and HCT116 cells were transfected with siRNA against GPD1L, confirming knockdown (Fig. 5F). To confirm knockdown at the protein level, western blot analysis in Supplementary Material 8 C demonstrates reduction of GPD1L protein expression 48 h post-transfection with si-GPD1L compared to NC group. Colony formation assays revealed si-GPD1L transfected cells formed significantly more colonies than si-NC controls, suggesting GPD1L knockdown promoted CRC cell proliferation. Colonies were counted using ImageJ software (Fig. 5G). We have since analyzed an independent cohort of 598 colorectal cancer samples from the GSE39582 dataset on GEO (Gene Expression Omnibus data base). Supplementary Material 8 A confirms reduced GPD1L expression in tumors versus normal tissues. Kaplan-Meier curves in Supplementary Material 8B recapitulate our earlier results, with elevated GPD1L correlating with improved overall survival.

Discussion

In this study, we leveraged large-scale immunogenomic and CRC datasets to elucidate causal relationships between immune cell compositions and CRC susceptibility using a two-sample Mendelian randomization approach. Our key findings provide new insights into CRC etiology and propose candidate immunological mechanisms for modulating disease risk.

Through rigorous MR analyses, we identified 17 immune cell signatures causally potentially associated with increased CRC risk and 18 signatures linked to decreased risk. Notably, our results implicate dysregulated immunophenotypic profiles localized to distinct cellular populations, including B cells, CD8 + T cells, dendritic cells, and regulatory T cells (Tregs). While previous studies have broadly linked immune activation to CRC [18], our findings provide unprecedented resolution by pinpointing discrete immune subsets and their potential contributions to tumorigenesis.

Several immunophenotypes potentially associated with elevated CRC risk merit further mechanistic investigation. Increased SSC-A expression on B cells indicates higher B cell numbers/activity. As key immune

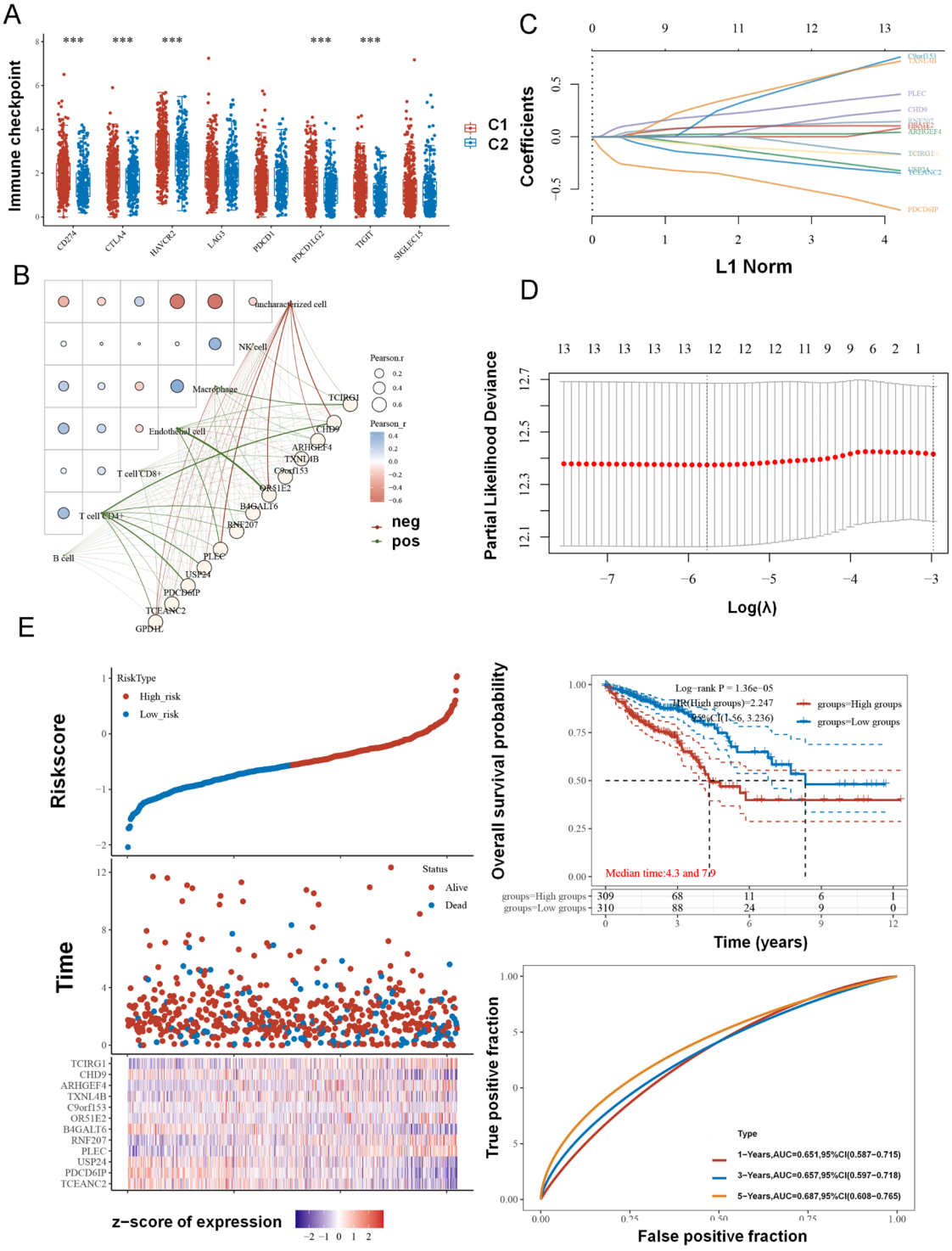

Fig. 4 Characterizing the Immune Landscape and Prognostic Signatures of Colorectal Cancer Subtypes based on Immunophenotype. (**a**) Differential expression heatmap of immune checkpoint genes between C1 and C2 subtypes. Downregulated genes are indicative of immune evasion in the poor prognosis C2 subtype; (**b**) Correlation network graph analyzing the relationship between immune phenotype-related gene expression and immune cell scores. Most genes positively correlate with CD4 + T cells; (**c**, **d**) Univariate Cox regression of immune traits on overall survival. Thirteen genes (dots) were selected for LASSO modeling based on significance; (**e**) Kaplan-Meier analysis of overall survival stratified by the immunophenotype-derived risk signature. Low-risk patients demonstrated significantly greater OS (hazard ratio 2.247, 95% CI 1.56–3.236, log-rank $p = 1.36 \times 10 - 5$), validating prognostic accuracy of the signature at 1, 3 and 5 years

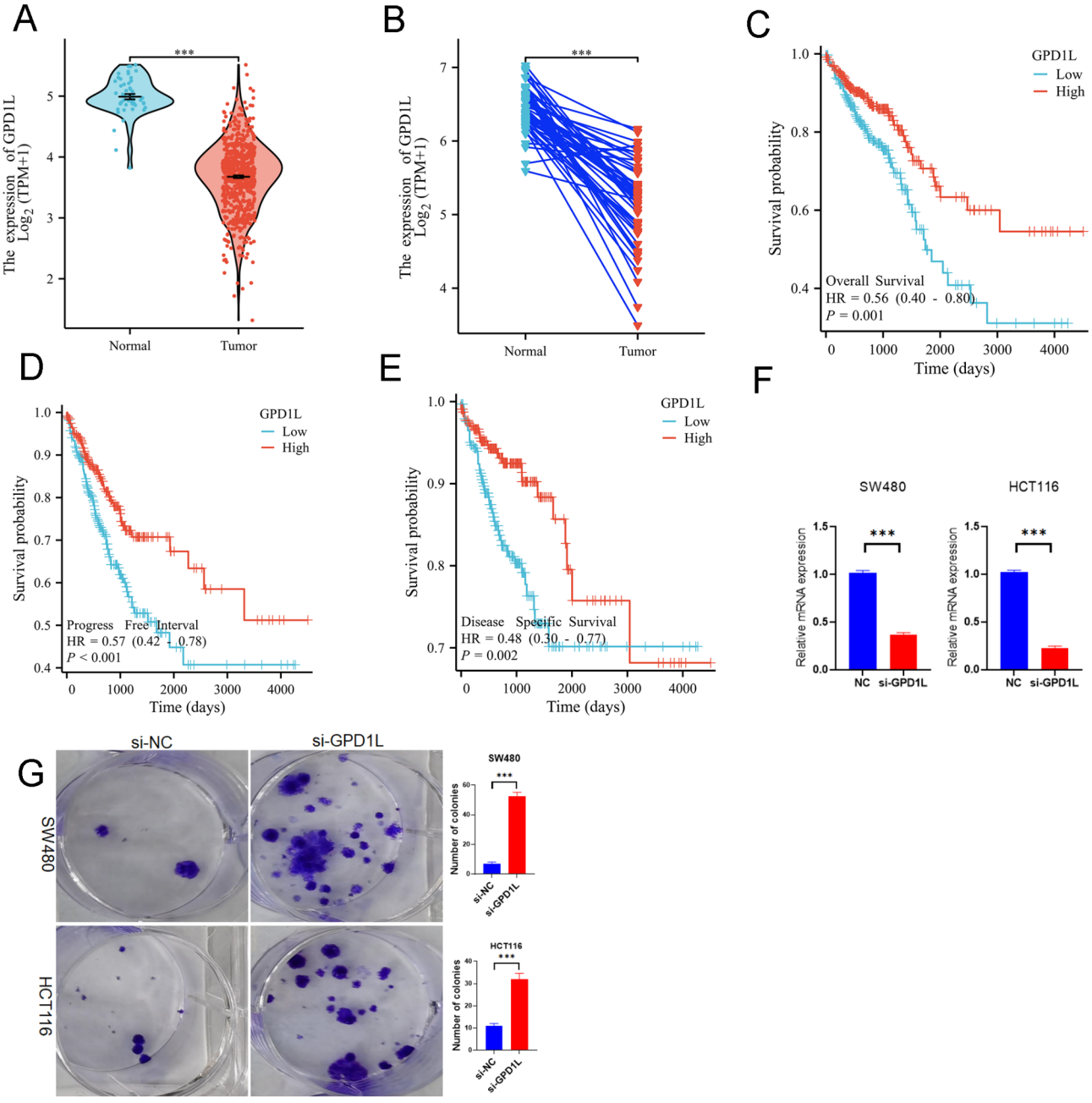

Fig. 5 GPD1L as a potential therapeutic target in CRC. (**a**) Boxplot comparing GPD1L expression between tumor and normal tissues from TCGA CRC patients; (**b**) Paired analysis showing reduced GPD1L in tumors versus matched normal tissues; (**c-e**) Kaplan-Meier survival curves showed elevated GPD1L correlated with improved overall survival (OS), disease-specific survival (DSS) and progression-free survival (PFI); (**f**) qPCR validation of GPD1L knockdown in SW480 and HCT116 cells transfected with si-GPD1L versus si-NC control; (**g**) Colony formation assays demonstrated proliferation increases with GPD1L inhibition

activators and antibody producers, elevated B cells could facilitate tumorigenesis by enhancing immune responses [19–21]. An increase in the proportion of CM differentiated CD8br T cells may suggest a potential impairment in memory CD8 + T cell responses, which are crucial for immune surveillance. This could potentially allow for cancer immune evasion [22–24]. Upregulated CD8 expression on terminally differentiated CD28- CD8br T cells may compromise their cytotoxic functions against tumors [25, 26]. Higher CCR2 levels on activated CD62L + myeloid dendritic cells suggests enhanced antigen presentation, but tumors could co-opt this to promote evasion of immune detection [27–29]. An increased proportion of activated/resting Tregs as a fraction of total CD4 + T cells implies stronger inhibitory effects of Tregs in stifling anti-tumor immunity [30–32].

We identified some immunophenotypes with relatively weak associations (ORs closer to 1) with CRC risk, suggesting that these results may require further validation. Future studies with larger sample sizes and complementary analytical approaches will be needed to ensure the robustness of these findings. Additionally, we have framed our study as an exploratory analysis to uncover as many potentially relevant immune signatures as possible, while acknowledging the need for further confirmatory investigations. Nevertheless, our overall analysis provides important insights into the role of the immune system in the pathogenesis of CRC. Some immunophenotypes identified as risk factors for colorectal cancer exhibited relatively weak effect sizes with odds ratios near 1. These marginal effects will require larger independent datasets and complementary analytic techniques to verify robustness before concluding causality. A key limitation of our study is that we did not employ false discovery rate (FDR) control or other multiple testing correction methods when performing statistical analyses involving multiple immune cell traits and genetic variants. This increases the risk of type I errors due to multiple testing. While we implemented stringent Bonferroni-corrected significance thresholds, some associations we identified require replication in independent studies to account for potential inflation of false positives without FDR control. We have highlighted several results that merit further validation. Future studies with larger sample sizes incorporating FDR are warranted to draw more robust conclusions from analyses of our vast omic datasets. Nevertheless, our exploratory analysis provides novel insights into immune-disease relationships in colorectal cancer requiring confirmation and follow-up.

Leveraging our comprehensive Mendelian randomization analyses, we prioritized the top 5 risk-enhancing immune cell populations and annotated their potentially associated genetic variants, unveiling 48 putative regulatory genes. Through an integrative risk prediction model and subsequent functional validation assays, we identified GPD1L as a key gene mediating colorectal cancer susceptibility. Integrating multi-omic profiling, we developed the first immuno-stratified taxonomy of CRC to stratify clinical behavior. Our C1/C2 subtypes depend on coordinated immunophenotypic programming, with the C2 subtype conferring poorer prognosis. Focusing on the prognostic GPD1L, its clinical significance and tumor-suppressive role in CRC were characterized both clinically and functionally through in vitro experimentation. Strikingly, GPD1L knockdown potently promoted CRC cell proliferation, highlighting its pivotal role as a tumor suppressor and representing a promising therapeutic target.

Conclusion

In this study, we employed an innovative integrated immunogenomic approach leveraging two-sample Mendelian randomization to systematically profile associations between immune cell compositions and colorectal cancer (CRC) susceptibility. An immune-stratified taxonomy of CRC was developed, where the C2 subtype showed poorer prognosis, highlighting the need to understand the molecular drivers of subtype-specific outcomes. The tumor suppressor GPD1L was characterized as a prognostic marker and a regulator of CRC pathways. This integrated immunogenomic approach is a powerful tool for dissecting precision immune-disease relationships.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12876-025-03776-4.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6
Supplementary Material 7
Supplementary Material 8
Supplementary Material 9
Supplementary Material 10

Acknowledgements
Thanks to all authors for their contributions.

Author contributions
H.K.J. designed this work. H.K.J. performed bioinformatic analyses. H.K.J. wrote the manuscript. H.K.J. and G.G.Y. revised the manuscript and performed the whole experiment. All authors have read and approved the final submitted manuscript.

Funding
This project was funded by Quzhou Municipal Science and Technology Bureau (2023K117).

Data availability
All data generated or analyzed during this study are included in this article.

Declarations

Ethics approval or consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 26 July 2024 / Accepted: 11 March 2025

Published online: 31 March 2025

References

1. O'Donnell JS, Teng MWL, Smyth MJ. Cancer immunoediting and resistance to T cell-based immunotherapy. Nat Rev Clin Oncol. 2019;16:151–67.
2. Ogino S, Nowak JA, Hamada T, Phipps AI, Peters U, Milner DA Jr., Giovannucci EL, Nishihara R, Giannakis M, Garrett WS, Song M. Integrative analysis of exogenous, endogenous, tumour and immune factors for precision medicine. Gut. 2018;67:1168–80.
3. Karki R, Kanneganti TD. Diverging inflammasome signals in tumorigenesis and potential targeting. Nat Rev Cancer. 2019;19:197–214.
4. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pagès C, Tosolini M, Camus M, Berger A, Wind P, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science. 2006;313:1960–4.
5. Osman A, Yan B, Li Y, Pavelko KD, Quandt J, Saadalla A, Singh MP, Kazemian M, Gounari F, Khazaie K. TCF-1 controls T(reg) cell functions that regulate inflammation, CD8(+) T cell cytotoxicity and severity of colon cancer. Nat Immunol. 2021;22:1152–62.
6. Yang K, Li S, Ding Y, Meng X, Zhang C, Sun X. Effect of smoking-related features and 731 immune cell phenotypes on esophageal cancer: a two-sample and mediated Mendelian randomized study. Front Immunol. 2024;15:1336817.
7. Hao X, Ren C, Zhou H, Li M, Zhang H, Liu X. Association between Circulating immune cells and the risk of prostate cancer: a Mendelian randomization study. Front Endocrinol (Lausanne). 2024;15:1358416.
8. Lu Z, Yin Y, Rao T, Xu X, Zhao K, Liu Z, Qin C, Tang M. Interaction of immune cells with renal cancer development: Mendelian randomization (MR) study. BMC Cancer. 2024;24:439.
9. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, Narita A, Konuma T, Yamamoto K, Akiyama M, et al. A cross-population atlas of genetic associations for 220 human phenotypes. Nat Genet. 2021;53:1415–24.
10. Orrù V, Steri M, Sidore C, Marongiu M, Serra V, Olla S, Sole G, Lai S, Dei M, Mulas A, et al. Complex genetic signatures in immune cells underlie autoimmunity and inform therapy. Nat Genet. 2020;52:1036–45.
11. Sidore C, Busonero F, Maschio A, Porcu E, Naitza S, Zoledziewska M, Mulas A, Pistis G, Steri M, Danjou F, et al. Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. Nat Genet. 2015;47:1272–81.
12. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. Stat Methods Med Res. 2017;26:2333–55.
13. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2016;40:304–14.
14. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal Pleiotropy assumption. Int J Epidemiol. 2017;46:1985–98.
15. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol. 2017;46:1734–9.
16. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol. 2017;32:377–89.
17. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal Pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50:693–8.
18. Schmitt M, Greten FR. The inflammatory pathogenesis of colorectal cancer. Nat Rev Immunol. 2021;21:653–67.
19. Balkwill F, Montfort A, Capasso M. B regulatory cells in cancer. Trends Immunol. 2013;34:169–73.
20. Sarvaria A, Madrigal JA, Saudemont A. B cell regulation in cancer and anti-tumor immunity. Cell Mol Immunol. 2017;14:662–74.
21. Schwartz M, Zhang Y, Rosenblatt JD. B cell regulation of the anti-tumor response and role in carcinogenesis. J Immunother Cancer. 2016;4:40.
22. Huang Q, Wu X, Wang Z, Chen X, Wang L, Lu Y, Xiong D, Liu Q, Tian Y, Lin H, et al. The primordial differentiation of tumor-specific memory CD8(+) T cells as Bona Fide responders to PD-1/PD-L1 Blockade in draining lymph nodes. Cell. 2022;185:4049–e40664025.
23. Han J, Khatwani N, Searles TG, Turk MJ, Angeles CV. Memory CD8(+) T cell responses to cancer. Semin Immunol. 2020;49:101435.
24. Zhang L, Yu X, Zheng L, Zhang Y, Li Y, Fang Q, Gao R, Kang B, Zhang Q, Huang JY, et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. Nature. 2018;564:268–72.
25. Huff WX, Kwon JH, Henriquez M, Fetcko K, Dey M. The evolving role of CD8(+) CD28(-) Immunosenescent T cells in cancer immunology. Int J Mol Sci 2019, 20.
26. Hui E, Cheung J, Zhu J, Su X, Taylor MJ, Wallweber HA, Sasmal DK, Huang J, Kim JM, Mellman I, Vale RD. T cell costimulatory receptor CD28 is a primary target for PD-1-mediated Inhibition. Science. 2017;355:1428–33.
27. MacNabb BW, Tumuluru S, Chen X, Godfrey J, Kasal DN, Yu J, Jongsma MLM, Spaapen RM, Kline DE, Kline J. Dendritic cells can prime anti-tumor CD8(+) T cell responses through major histocompatibility complex cross-dressing. Immunity. 2022;55:982–e997988.
28. Garris CS, Luke JJ. Dendritic cells, the T-cell-inflamed tumor microenvironment, and immunotherapy treatment response. Clin Cancer Res. 2020;26:3901–7.
29. Hansen M, Andersen MH. The role of dendritic cells in cancer. Semin Immunopathol. 2017;39:307–16.
30. Huppert LA, Green MD, Kim L, Chow C, Leyfman Y, Daud AI, Lee JC. Tissue-specific Tregs in cancer metastasis: opportunities for precision immunotherapy. Cell Mol Immunol. 2022;19:33–45.
31. Wang H, Franco F, Ho PC. Metabolic regulation of Tregs in cancer: opportunities for immunotherapy. Trends Cancer. 2017;3:583–92.
32. Curiel TJ. Tregs and rethinking cancer immunotherapy. J Clin Invest. 2007;117:1167–74.

Publisher's note