



The Journal of Biomedical Research

Causal genetic regulation of DNA replication on immune microenvironment in colorectal tumorigenesis: Evidenced by an integrated approach of trans-omics and GWAS

Wang Sumeng, Chen Silu, Li Huiqin, Ben Shuai, Zhao Tingyu, Zheng Rui, Wang Meilin, Gu Dongying, Liu Lingxiang

Cite this article as:

Wang Sumeng, Chen Silu, Li Huiqin, Ben Shuai, Zhao Tingyu, Zheng Rui, Wang Meilin, Gu Dongying, Liu Lingxiang. Causal genetic regulation of DNA replication on immune microenvironment in colorectal tumorigenesis: Evidenced by an integrated approach of trans-omics and GWAS[J]. *Journal of Biomedical Research*, 2024, 38(1): 37–50. doi: 10.7555/JBR.37.20230081

View online: <https://doi.org/10.7555/JBR.37.20230081>

Articles you may be interested in

[Role of inflammasome regulation on immune modulators](#)

The Journal of Biomedical Research. 2018, 32(6): 401 <https://doi.org/10.7555/JBR.32.20170120>

[Immune checkpoint inhibitors in cancer therapy](#)

The Journal of Biomedical Research. 2018, 32(5): 317 <https://doi.org/10.7555/JBR.31.20160168>

[Prostate cancer tends to metastasize in the bone-mimicking microenvironment via activating NF- \$\kappa\$ B signaling](#)

The Journal of Biomedical Research. 2018, 32(5): 343 <https://doi.org/10.7555/JBR.32.20180035>

[Comparative effects of plant oils and trans-fat on blood lipid profiles and ischemic stroke in rats](#)

The Journal of Biomedical Research. 2017, 31(2): 122 <https://doi.org/10.7555/JBR.31.20160095>

[Identification of four novel prognosis biomarkers and potential therapeutic drugs for human colorectal cancer by bioinformatics analysis](#)

The Journal of Biomedical Research. 2021, 35(1): 21 <https://doi.org/10.7555/JBR.34.20200021>

[The immune-modulating effects of viable *Weissella cibaria* JW15 on RAW 264.7 macrophage cells](#)

The Journal of Biomedical Research. 2020, 34(1): 36 <https://doi.org/10.7555/JBR.33.20190095>



Causal genetic regulation of DNA replication on immune microenvironment in colorectal tumorigenesis: Evidenced by an integrated approach of trans-omics and GWAS

Sumeng Wang^{1,Δ}, Silu Chen^{2,3,Δ}, Huiqin Li⁴, Shuai Ben^{2,3}, Tingyu Zhao¹, Rui Zheng^{2,3}, Meilin Wang^{2,3}, Dongying Gu^{5,✉}, Lingxiang Liu^{1,✉}

¹Department of Oncology, the First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu 210029, China;

²Department of Environmental Genomics, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing Medical University, Nanjing, Jiangsu 211166, China;

³Department of Genetic Toxicology, the Key Laboratory of Modern Toxicology of Ministry of Education, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu 211166, China;

⁴Department of Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu 211166, China;

⁵Department of Oncology, Nanjing First Hospital, Nanjing Medical University, Nanjing, Jiangsu 210006, China.

Abstract

The interplay between DNA replication stress and immune microenvironment alterations is known to play a crucial role in colorectal tumorigenesis, but a comprehensive understanding of their association with and relevant biomarkers involved in colorectal tumorigenesis is lacking. To address this gap, we conducted a study aiming to investigate this association and identify relevant biomarkers. We analyzed transcriptomic and proteomic profiles of 904 colorectal tumor tissues and 342 normal tissues to examine pathway enrichment, biological activity, and the immune microenvironment. Additionally, we evaluated genetic effects of single variants and genes on colorectal cancer susceptibility using data from genome-wide association studies (GWASs) involving both East Asian (7 062 cases and 195 745 controls) and European (24 476 cases and 23 073 controls) populations. We employed mediation analysis to infer the causal pathway, and applied multiplex immunofluorescence to visualize colocalized biomarkers in colorectal tumors and immune cells. Our findings revealed that both DNA replication activity and the flap structure-specific endonuclease 1 (*FEN1*) gene were significantly enriched in colorectal tumor tissues, compared with normal tissues. Moreover, a genetic variant rs4246215 G>T in *FEN1* was associated with a decreased risk of colorectal cancer (odds ratio = 0.94, 95% confidence interval: 0.90–0.97, $P_{\text{meta}} = 4.70 \times 10^{-9}$). Importantly, we identified basophils and eosinophils that both exhibited a significantly decreased infiltration in colorectal tumors, and were regulated by rs4246215 through causal pathways involving both *FEN1* and DNA replication. In conclusion, this trans-omics incorporating GWAS data provides insights into a plausible pathway connecting DNA replication and immunity, expanding biological knowledge of colorectal tumorigenesis and therapeutic targets.

^ΔThese authors contributed equally to this work.

[✉]Corresponding authors: Dongying Gu, Department of Oncology, Nanjing First Hospital, Nanjing Medical University, 68 Changle Road, Nanjing, Jiangsu 210006, China. E-mail: dygu@njmu.edu.cn; Lingxiang Liu, Department of Oncology, the First Affiliated Hospital of Nanjing Medical University, 300 Guangzhou Road, Nanjing, Jiangsu 210029, China. E-mail: llxlau@163.com.

Received: 07 April 2023; Revised: 26 May 2023; Accepted: 28 May 2023; Published online: 18 December 2023

CLC number: R735.3, Document code: A

The authors reported no conflict of interests.

This is an open access article under the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited.

Keywords: trans-omics, DNA replication, tumor immune microenvironment, causal mediation, colorectal tumorigenesis

Introduction

Colorectal cancer is among the top ten cancer types globally^[1], ranking the fourth in terms of incidence, and the second in terms of mortality in the United States as of 2022^[2]. Both sporadic and hereditary forms of colorectal cancer arise from complex factors, such as dietary and environmental hazard exposure, as well as germline mutations (accounting for up to 6%–10% of all causes)^[3–4]. A complete and accurate DNA replication is not only crucial for genomic stability but also a vulnerable cellular process that can initiate colorectal cancer and other cancers^[5–6].

The tumor immune microenvironment is a complex and dynamic entity that comprises various innate and adaptive immune cells, stromal cells, and extracellular matrix components^[7]. It is noteworthy to mention that immune cells can exert either anti-cancer or pro-tumor effects through their interactions with tumor cells, and the spatial correlation between them can be investigated by multiplex immunofluorescence^[8–10]. Although immunotherapy targeting immune cell receptors or ligands, such as programmed death-1, has revolutionized cancer treatment, the emergence of drug resistance remains an inevitable challenge^[11].

Recent advances in high-throughput technologies, including genomics, transcriptomics, epigenomics, and proteomics, have facilitated a comprehensive understanding of biological systems at multiple levels^[12]. The integration of different omics approaches is a promising strategy that may facilitate a more detailed molecular understanding of health and disease^[13–14], helping in biomarker identification to clarify disease etiology and guide therapeutic development^[15–17].

In the current study, we performed integrated analyses of genomic, transcriptomic, and proteomic profiles to explore potential causal genetic loci, susceptibility genes, and biological pathways associated with colorectal cancer risk, as well as the association between DNA replication and immunity.

Materials and methods

Ethic approval and consent to participate

Each patient who participated in the study provided a written informed consent for the biological research,

and the study was approved by the Institutional Review Board of Nanjing Medical University (approval No. 2017-515).

Availability of data and materials

Transcriptomic level data from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) are both publicly available and open accessible. The genome-wide association study (GWAS) summary statistics of East Asian populations are available from Biobank Japan (BBJ, <http://jenger.riken.jp/en/>) without any access restrictions. The GWAS summary data of European populations were obtained from the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) *via* the approval of the database of Genotypes and Phenotypes (dbGaP, accession numbers: phs001315.v1.p1, phs001078.v1.p1, and phs001415.v1.p1). Transcriptomic and proteomic data from the Nanjing Colorectal Cancer Cohort are available from the corresponding author upon reasonable request.

Transcriptomic and proteomic profiles of colorectal cancer

The Nanjing Colorectal Cancer Cohort dataset, which is a long-term follow-up clinical cohort, provides transcriptomic data for 79 pairs of colorectal tumor tissues and adjacent normal tissues, as well as proteomic data for 25 paired tumors and adjacent normal tissues. The characteristics of participants and the recruitment protocols have been described previously^[18–20]. Demographic characteristics are shown in *Supplementary Table 1* (available online).

A public dataset with information on 644 tumor tissues and 51 normal tissues was obtained from TCGA. Another three independent colorectal cancer datasets were obtained from GEO, which include GSE74602 (30 paired normal and tumor samples, Southeast Asian populations), GSE106582 (77 tumors and 117 mucosa tissues, European populations), and GSE117606 (74 tumors and 65 adjacent normal tissues, European populations).

Genetic association study of colorectal cancer

The dataset of GWAS summary statistics of East Asian populations covered 7 062 colorectal cancer cases and 195 745 controls, with 8 678 297 single nucleotide polymorphisms (SNPs) (*Supplementary*

Table 1). Details regarding genotyping and imputation were described elsewhere^[21]. The qualified GWAS summary data were selected according to the following inclusion criteria: (1) call rate > 95%; (2) P -value for Hardy-Weinberg equilibrium (HWE) > 1×10^{-6} ; (3) minor allele frequency (MAF) > 0.05; and (4) R^2 (imputation quality measure by Minimac3) > 0.80.

The GWAS summary data of European populations combined a sample size of 24 476 cases and 23 073 controls, with 6 709 910 SNPs. Detailed information about GECCO can be found in previous studies^[18,22–24]. Demographic characteristics are shown in **Supplementary Table 1**. The inclusion criteria for qualified GWAS summary data were as follows: (1) call rate > 95%; (2) MAF > 0.05; (3) $P_{\text{HWE}} > 1 \times 10^{-6}$; and (4) imputation accuracy $R^2 > 0.30$.

Pathway enrichment and activity estimation

Gene set enrichment analysis (GSEA)^[25] and gene set variation analysis (GSVA)^[26] were performed at the RNA and protein levels, based on the pathway in the Kyoto Encyclopedia of Genes and Genomes (KEGG) in Molecular Signatures Databases (MSigDB, <http://www.gsea-msigdb.org/gsea/msigdb/index.jsp>) via the R packages clusterProfiler^[27] and GSVA, respectively. We reserved enriched pathways from GSEA with a false discovery rate threshold of 0.05 for further analysis. The difference in pathway activity between tumor and normal samples calculated by GSVA was estimated by the R package limma^[28]. We selected pathways from GSVA with an adjusted P -value cutoff of 0.05 and the absolute value of $\log_2(\text{fold change, FC}) > 0.2$ ^[29]. To facilitate subsequent analysis, the resulting pathways were driven by intersecting the outcomes of GSEA and GSVA.

Gene and gene-set analysis via genetic association study

Gene and gene-set analyses were performed by Multi-marker Analysis of GenoMic Annotation (MAGMA)^[30] using genome-wide summary statistics from East Asian and European populations. In the gene-based analysis, SNPs located between transcription start and stop sites of a gene were annotated to the gene, based on dbSNP version 135 SNP locations and NCBI 37.3 gene definitions, to assess the joint association of all SNPs in the gene with the phenotype. Similarly, in the gene-set analysis, individual genes were aggregated into biological pathways or cellular functions underlying MSigDB, showing the potential genetic etiology of phenotypes. The pathways met the criteria of having significant associations, as determined by a Bonferroni-adjusted P -value of less than 0.05, were selected along with

their corresponding genes and SNPs. Comprehensive information regarding these pathways, genes, and SNPs can be found in **Supplementary Tables 2 and 3** (available online).

Deconvolution of the tumor immune microenvironment at the transcriptome level

The tumor immune microenvironment was assessed by CIBERSORT^[31] and xCell^[32] algorithms using transcriptomic data. CIBERSORT (<http://cibersort.stanford.edu/>) provides an estimation of the abundance of immune cell types associated with the LM22 gene signature matrix. xCell (<http://xCell.ucsf.edu/>), a gene signature-based method, enables cell type enrichment analysis from gene expression data for 64 kinds of cell types, including innate and adaptive immune cells, hematopoietic progenitor cells, epithelial cells, and extracellular matrix cells. Detailed immune cell types are shown in **Supplementary Table 4** (available online).

In silico analysis of genetic locus and functional annotation

Functional Annotation of Variants-Online Resource (FAVOR, <http://favor.genohub.org/>), a method providing functional annotations for 13 major categories, was used to annotate the selected SNPs and plot genetic functions. SAIGE UKB (<https://pheweb.org/UKB-SAIGE/>) supplied phenome-wide association study (PheWAS) summary statistics of the selected SNPs, which included 1 403 ICD-based traits^[33]. Region plots^[34] were generated by LocusZoom. Genotype and expression data (transcript and protein levels) were merged to perform expression quantitative trait loci (eQTL) in colorectal tumor and normal tissues in a linear regression model^[35]. Details on genotype data collection, imputation, and processing have been published before^[19,36–37].

Causal inference analysis of SNP, pathway, and immune features

The associations between the candidate variant and immune features underlying the causal pathway, mediated by gene expression/pathway activity, were assessed based on four scenarios: (1) merging all associations between genetic variants and immunity (β_c); (2) merging all associations between genetic variants and gene expression/pathway activity (β_a); (3) merging all associations between gene expression/pathway activity and immunity (β_b); and (4) causal inference analysis. The indirect effect of genetic variants on immunity mediated by gene expression/pathway activity was estimated using the equation $\beta_{\text{indirect}} = \beta_a \times \beta_b$. The statistical P -value for

indirect effects was determined using bootstrapping method^[38] via the R package mediation^[39]. The correlation between flap structure-specific endonuclease 1 (*FEN1*) expression/pathway activity and immunity was analyzed by Spearman correlation underlying the R package psych^[40], and meta-analysis was performed via the R package metafor^[41].

Multiplex immunofluorescence analysis

Multiplex immunofluorescence staining was performed at Servicebio Technology Co., Ltd. (Wuhan, China), and a multiplexed tyramide signal amplification method was used on 4 µm sections of the tissue microarray (TMA) containing 30 pairs of colorectal tumors and adjacent normal tissues from the Nanjing cohort. The details of TMA were reported previously^[22]. After deparaffinization, antigen retrieval, and serum blocking, the TMA was sequentially stained with the primary antibodies anti-CD3 (GB13440, Servicebio), anti-CD45 (GB113885, Servicebio), anti-CD22 (66103-1-IG, Proteintech, Rosemont, IL, USA) and anti-FEN1 (14768-1-AP, Proteintech), followed by the appropriate secondary antibodies conjugated with horseradish peroxidase (HRP) and tyramine signal amplification (TSA) visualization. DAPI (4',6-diamidino-2-phenylindole) was finally added to stain nuclei until all four markers were labeled. Antibody details are shown in [Supplementary Table 5](#) (available online). The slide was scanned with the Panoramic MIDI (3DHISTECH, Hungary). Multispectral images were evaluated using the HALO Image Analysis Platform (Version 3.0.311.314, Indica, US). Basophils (CD22⁺), eosinophils (CD45⁺), T cells (CD3⁺) and FEN1-positive cells were quantified, and their average intensity in TMA was calculated.

Statistical analysis

All statistical analyses were performed using R software (version 4.0.5). The normality of the variables was assessed using the Shapiro-Wilk test. The Chi-squared test, Fisher's exact test, Student's *t*-test, and Wilcoxon test were used as appropriate for the respective variables. Correlations between variables were assessed using either Spearman's or Pearson's correlation, depending on the normal distribution of the data. The false discovery rate correction was used in the RNA-seq analysis to manage false positives in multiple tests, efficiently pinpointing genes with significantly differential expression levels. In the SNP analysis with lower-dimensional genotype data, Bonferroni correction was applied to adjust the significance level for the number of comparisons conducted, reducing the likelihood of

false discoveries. Bonferroni correction provides the most conservative results, emphasizing highly significant associations. A *P*-value less than 0.05, with a two-tailed test, was considered statistically significant. To perform the meta-analysis of the summary-level GWAS results for East Asian and European populations, we used METAL software, with the fixed-effects modeling and inverse-variance weighting^[42].

Results

Identification of differential biological pathways in colorectal tumors

The flowchart of the current study is shown in [Fig. 1](#). We first analyzed the differentially enriched pathways between colorectal tumors and adjacent normal tissues using the GSEA method. Although some pathways were specifically enriched in individual datasets, 38 pathways were consistently enriched in colorectal tumors ([Fig. 2A](#), [Supplementary Table 6](#) [available online]). We then used GSVA to estimate the pathway activity ([Supplementary Fig. 1A–1E](#), available online) and found four pathways differentially enriched between colorectal tumors and adjacent normal tissues across all datasets (*i.e.*, DNA replication, mismatch repair, proteasome, and RNA polymerase; [Fig. 2B](#)). Remarkably, these four pathways were the same as those identified by GSEA ([Fig. 2C](#)). Moreover, we performed GSEA ([Supplementary Table 7](#), available online) and GSVA ([Supplementary Fig. 1F](#), available online) on the Nanjing cohort proteome profiles, revealing that the four pathways enriched at the RNA level were also enriched at the protein level, thus being selected for further investigation ([Fig. 2D](#)).

Estimation of genetic effects in pathways on colorectal cancer susceptibility

We then assessed genomic information on colorectal cancer susceptibility to determine key genes in the four candidate pathways. Interestingly, *FEN1*, one of 108 genes in the four selected pathways (specifically the DNA replication pathway), was significantly associated with colorectal cancer susceptibility in East Asian populations ($P_{\text{Bonferroni}} = 0.025$; [Fig. 3A](#) and [Supplementary Table 2](#) [available online]), and was subsequently validated in European populations ($P = 4.41 \times 10^{-4}$; [Supplementary Table 2](#)). Variant-to-gene annotation ([Table 1](#), [Fig. 3B](#)) revealed that only rs4246215 at 11q12.2 in *FEN1* was associated with colorectal cancer susceptibility in both East Asian (OR = 0.92, 95% CI: 0.87–0.95, $P =$

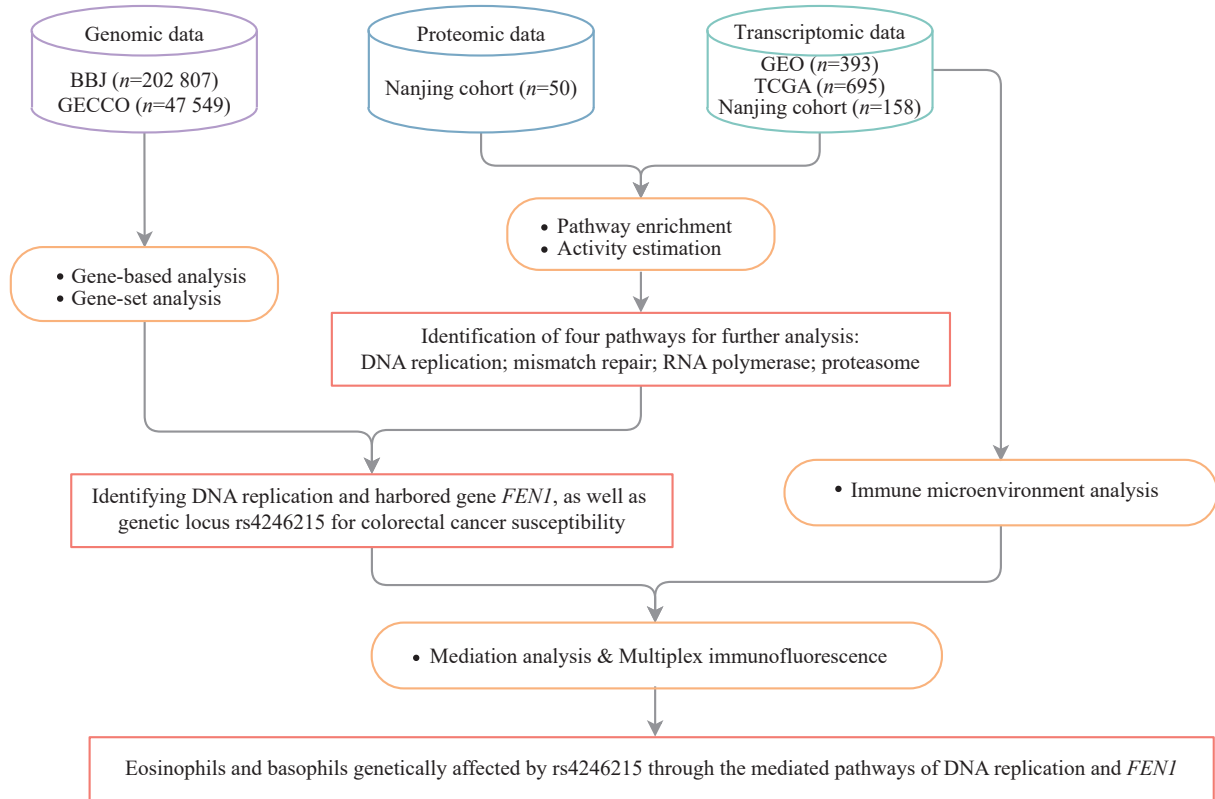


Fig. 1 Flow chart of the study design. Omics data and methods used in the current study are shown in cylinders and orange boxes, respectively. The key results of the study are displayed in red boxes. Four distinct pathways were derived by intersecting the results obtained from GSEA and GSVA using transcriptomic level and proteomic level datasets. Subsequently, MAGMA analysis was employed with genome-wide summary statistics to identify pathways that exhibited statistical significance. Additionally, a specific gene of interest, *FEN1*, and its corresponding SNP (rs4246215) were identified. Notably, the tumor immune microenvironment was also depicted, providing valuable insights into the interplay between genetic factors and the immune response within the context of tumorigenesis. To further explore causal associations, mediation analysis, and multiplex immunofluorescence were conducted, enabling a comprehensive understanding of the underlying mechanisms. Abbreviations: BBJ, Biobank Japan; GECCO, Genetics and Epidemiology of Colorectal Cancer Consortium; GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas; GSEA, gene set enrichment analysis; MAGMA, Multi-marker Analysis of GenoMic Annotation; SNPs, single nucleotide polymorphisms.

1.46×10^{-6}) and European (OR = 0.95, 95% CI: 0.92–0.98, $P = 4.41 \times 10^{-4}$) populations and reached genome-wide significance (OR = 0.94, 95% CI: 0.90–0.97, $P = 4.70 \times 10^{-9}$) without heterogeneity ($P_{\text{het}} = 0.27$, $I^2 = 0.20$). However, no pathway passed the significance threshold of $P < 0.05$ ([Supplementary Table 3](#), available online).

Functional annotation revealed that rs4246215 was located at the 3'-untranslated-region of *FEN1*, harboring strong functional signals of the enhancer activity, histone modification and transcription factor binding calculated by the FAVOR website ([Supplementary Table 8](#), available online). PheWAS for pleiotropy evaluation via SAIGE UKB revealed that rs4246215 was dramatically associated with multiple phenotypes, especially in solid tumors and hematologic system tumors ([Supplementary Fig. 2A](#), available online). Notably, both *FEN1* expression (RNA level: $\log_2\text{FC} = 0.32$, $P < 0.0001$, Nanjing cohort; protein-level: $\log_2\text{FC} = 0.45$, $P < 0.0001$,

Nanjing cohort) and DNA replication activity (RNA level: $\log_2\text{FC} = 0.39$, $P = 1.40 \times 10^{-19}$, Nanjing cohort; protein-level: $\log_2\text{FC} = 0.34$, $P = 6.43 \times 10^{-5}$, Nanjing cohort) were significantly increased in colorectal tumors and were positively associated with one another at both the RNA ($r = 0.85$, $P < 2.20 \times 10^{-16}$, Nanjing cohort) and protein ($r = 0.72$, $P = 2.50 \times 10^{-8}$, Nanjing cohort) levels ([Fig. 3C](#) and [Supplementary Fig. 3](#) [available online]). Subsequent multiplex immunofluorescence staining confirmed the high expression of *FEN1* in colorectal tumors ([Fig. 3D](#) and [3E](#)).

Evaluation of immune responses in colorectal tumorigenesis

We then assessed infiltrating immune cells in colorectal tumors and adjacent normal tissues to determine whether *FEN1* and DNA replication are involved in colorectal tumorigenesis via the immune pathway ([Supplementary Fig. 4](#), available online).

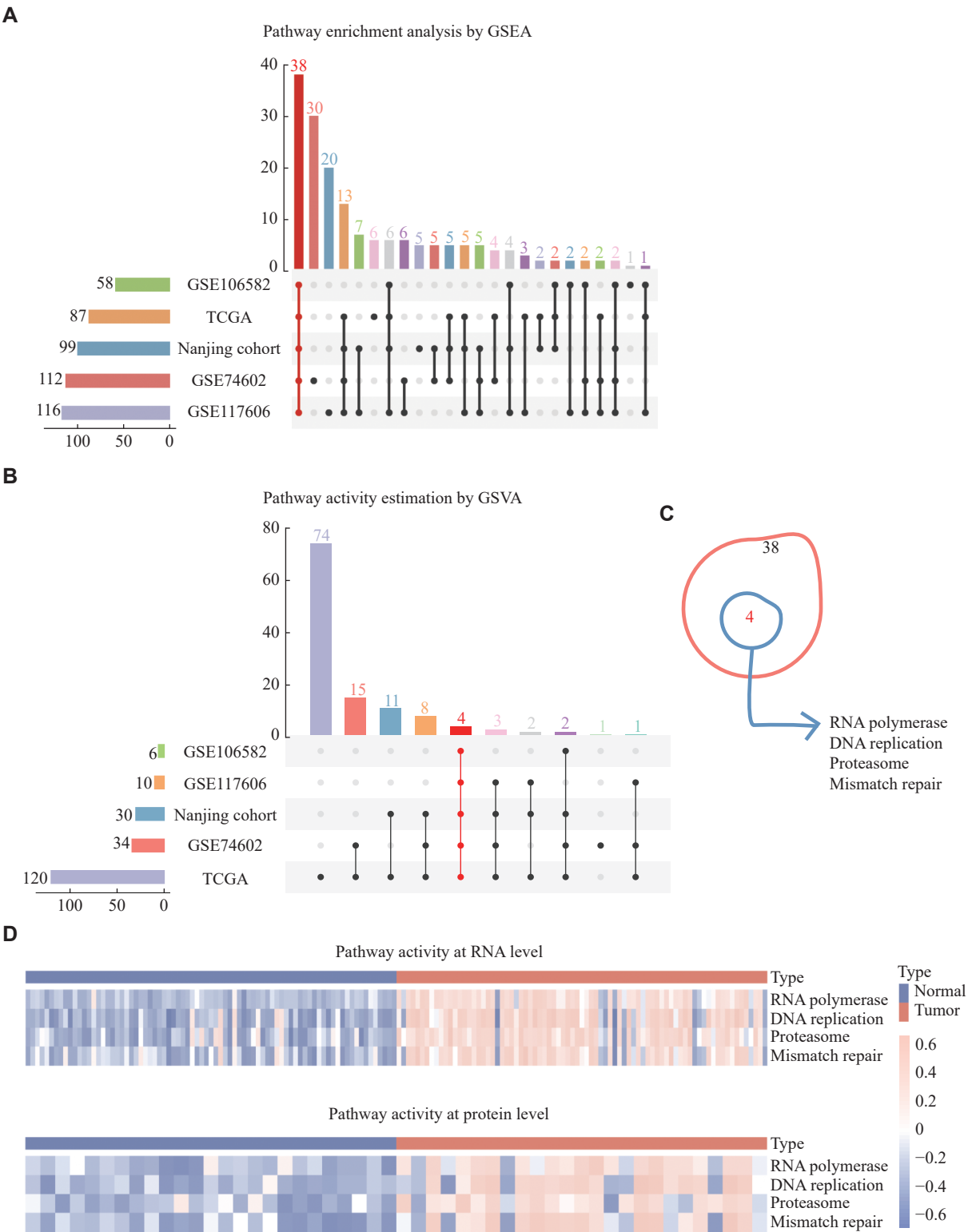


Fig. 2 Pathway enrichment analysis and activity estimation at both the RNA and protein levels. A: Significant pathways underlying the GSEA algorithm across five datasets at the RNA level. B: Significant pathways underlying the GSVA algorithm across five datasets at the RNA level. C: Four significant pathways were yielded from intersections between GSEA and GSVA analyses across five datasets. D: Differential activity of four representative pathways between colorectal tumors and normal tissues of the Nanjing cohort at the RNA level (top) and the protein level (bottom). Abbreviations: GSEA, gene set enrichment analysis; GSVA, gene set variation analysis; TCGA, The Cancer Genome Atlas.

Across the five datasets, a total of 26 distinct immune cells were consistently identified, albeit some cell

types being exclusively present in specific datasets (Fig. 4A). The proportions of most immune and

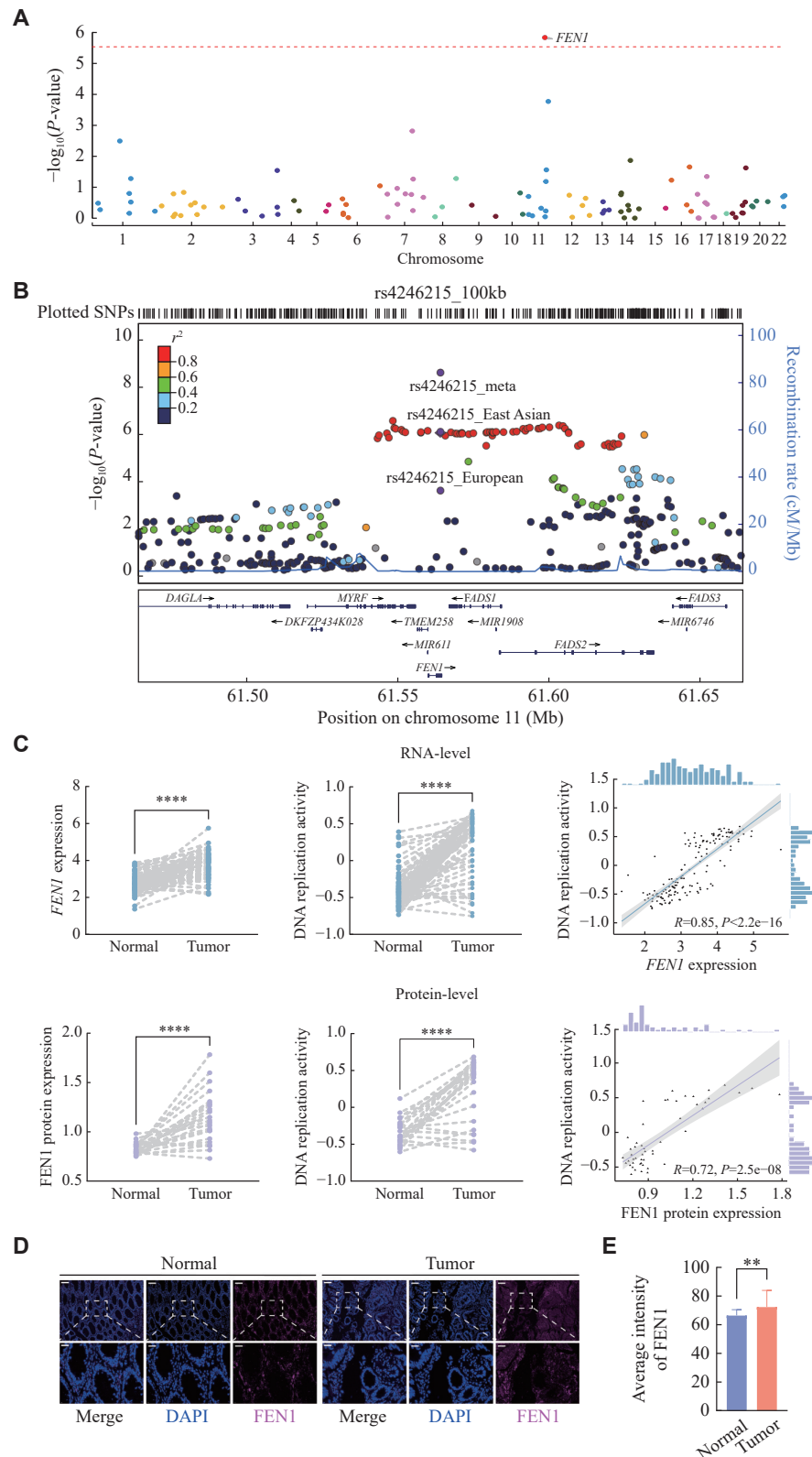


Fig. 3 Genetic effects on colorectal cancer susceptibility and expression pattern of *FEN1* and DNA replication activity. A: Manhattan plot of the genetic effect at the gene level via MAGMA. The red dotted line shows the significance threshold of Bonferroni-adjusted $P < 0.05$. B: Regional plot for the association of rs4246215 and colorectal cancer risk derived from BBJ, GECCO, and GWAS meta-analyses. Each unique locus was defined as ± 100 kilobases (kb) on either side of rs4246215. C: Expression pattern of *FEN1* and DNA replication activity of the Nanjing cohort at the RNA (top) and protein (bottom) levels. **** $P < 0.0001$. D: Representative multiple immunofluorescence images of *FEN1* in colorectal tumors and normal tissues. Scale bar: 100 μm (upper) and 25 μm (lower), respectively. E: The average intensity of *FEN1* in colorectal tumors and normal tissues. The P -value was calculated by Wilcoxon test. ** $P < 0.01$. Abbreviation: SNPs, single nucleotide polymorphisms.

Table 1 Association between rs4246215 and colorectal cancer susceptibility in Asian and European populations

Variant	CHR	BP	Reference/ effect allele	Gene	Populations	EAf	OR (95% CI)	P	P _{heterogeneity}	I ²
rs4246215	11	61564299	G/T	FEN1	East Asian	0.393	0.92 (0.87–0.95)	1.46×10 ⁻⁶		
					European	0.328	0.95 (0.92–0.98)	4.41×10 ⁻⁴		
					Combined		0.94 (0.90–0.97)	4.70×10 ⁻⁹	0.27	0.20

P-values of heterogeneity and meta-analysis were generated using METAL in fixed-effect inverse-variance. Abbreviations: CHR, chromosome; BP, base pair position in GRCh37/hg19; EAF, effect allele frequency; OR, odds ratio; CI, confidence interval.

stromal cells exhibited a decrease in colorectal tumors, but the proportions of Th1 cells, activated mast cells and M0 macrophages, showed an increase.

Furthermore, we conducted an additional analysis to determine the correlation of *FEN1* expression or DNA replication activity with the infiltration of immune cells. Interestingly, the infiltration levels of three immune cell types were significantly correlated with *FEN1*, and those of eight immune cell types were significantly correlated with DNA replication (Fig. 4B). Resting mast cells (*FEN1*: $r_{meta} = -0.42$, $P < 0.0001$; DNA replication: $r_{meta} = -0.43$, $P < 0.0001$) and megakaryocyte-erythroid progenitors (MEP; *FEN1*: $r_{meta} = 0.62$, $P < 0.0001$; DNA replication: $r_{meta} = 0.72$, $P < 0.0001$) exhibited consistent correlation patterns with *FEN1* expression and DNA replication activity.

Evaluation of the effect of rs4246215 on *FEN1* expression, DNA replication and the tumor immune microenvironment

The eQTL mapping was conducted to confirm the targets of rs4246215. Intriguingly, we observed that rs4246215 led to a downregulation of *FEN1* expression in colorectal tumors, while it conversely upregulated *FEN1* expression in normal tissues (Fig. 5A). In the Genotype-Tissue Expression Project (GTEx) database, a similar trend was observed in colon tissues ($\beta_{sigmoid} = -0.003$, $P = 0.90$; $\beta_{transverse} = -0.004$, $P = 0.80$), although the difference was not statistically significant. In addition, we found that rs4246215 decreased DNA replication activity in colorectal tumors but not in normal tissues (Fig. 5B) and dramatically influenced the proportions of five immune cells (i.e., basophils, CD4 Tem, eosinophils, mesangial cells, and Th2 cells; Fig. 5C and Supplementary Table 9 [available online]). However, the aforementioned trends were not significant in the Nanjing cohort (Supplementary Fig. 2B–2D, available online), potentially due to the limited sample size.

Subsequently, we performed multiplex immunofluorescence to visualize the colocalized biomarkers in the aforementioned immune cells (basophils, eosinophils and T cells) in the tumor

immune microenvironment. Representative immunofluorescence staining of basophils, eosinophils and T cells, represented by CD22⁺, CD45⁺ and CD3⁺ staining, is shown in Fig. 5D. In accordance with the immune infiltration analysis, the average levels of basophils (CD22⁺) and eosinophils (CD45⁺) infiltration in colorectal tumors were significantly lower than those in normal colorectal tissues ($P = 0.017$ and $P = 0.003$, respectively), but the difference in T cell (CD3⁺) infiltration levels was not statistically significant (Fig. 5E).

Causal cascade of rs4246215, DNA replication, and immunity in colorectal cancer susceptibility

We hypothesized that DNA replication and *FEN1* could act as mediators to influence the effect of rs4246215 on the immunity (Fig. 6). Upon the causal mediation analysis, we observed five significant indirect pathways, by which rs4246215 affects the immunity; both *FEN1* and DNA replication mediated the positive indirect effects of rs4246215 on CD4 Tem (*FEN1*: $\beta_{indirect} = 0.0002$, $P = 0.016$, 12.78% effects mediated; DNA replication: $\beta_{indirect} = 0.0002$, $P = 0.016$, 16.23% effects mediated), eosinophils (*FEN1*: $\beta_{indirect} = 0.0003$, $P = 0.004$, 21.42% effects mediated; DNA replication: $\beta_{indirect} = 0.0004$, $P = 0.016$, 24.83% effects mediated) and mesangial cells (*FEN1*: $\beta_{indirect} = 0.0003$, $P = 0.012$, 18.38% effects mediated; DNA replication: $\beta_{indirect} = 0.0006$, $P = 0.036$, 38.45% effects mediated), whereas they exerted negative indirect effects on basophils (*FEN1*: $\beta_{indirect} = -0.0008$, $P = 0.024$; DNA replication: $\beta_{indirect} = -0.0008$, $P = 0.036$) and Th2 cells (*FEN1*: $\beta_{indirect} = -0.0045$, $P = 0.004$, 56.40% effects mediated; DNA replication: $\beta_{indirect} = -0.0048$, $P = 0.016$, 60.95% effects mediated).

Discussion

In the current study, we investigated the potential causal protective locus rs4246215 G>T in *FEN1*, a locus associated with DNA replication, and its effect on the colorectal tumor immune microenvironment involved in tumorigenesis.

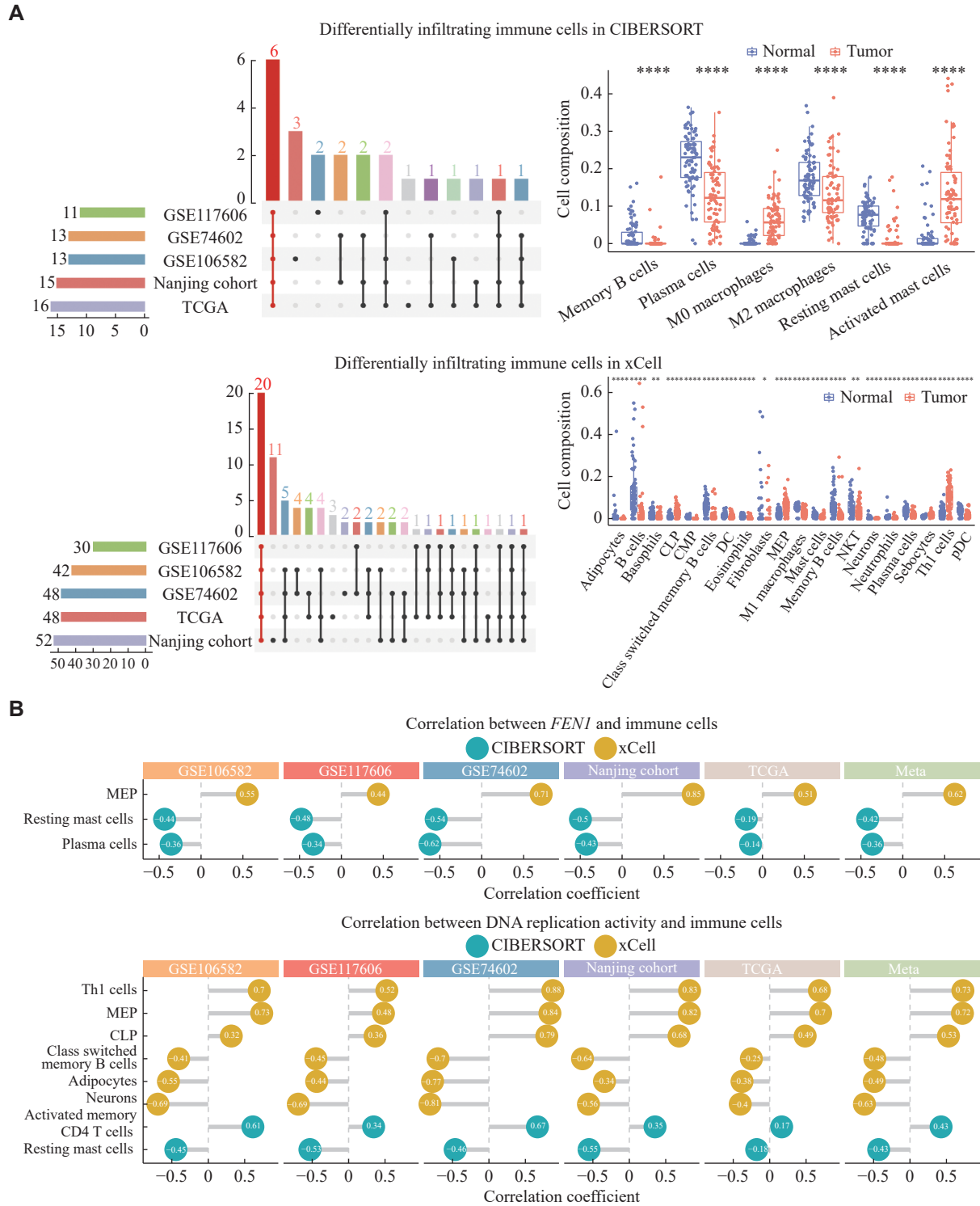


Fig. 4 Differentially infiltrating immune cells and their correlations with *FEN1* and DNA replication. A: Differentially infiltrating immune cells between colorectal tumors and normal tissues underlying the CIBERSORT (top) and xCell (bottom) algorithms across five datasets. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, and **** $P < 0.0001$. B: Correlations of *FEN1* expression (top) and DNA replication activity (bottom) with infiltrating immune cells. The meta value was derived by combining the results from the five datasets using a random effect model. The x-axis represents Spearman's rank correlation coefficient. Abbreviations: CLP, common lymphoid progenitors; CMP, common myeloid progenitors; DC, dendritic cells; MEP, megakaryocyte-erythroid progenitors; pDC, plasmacytoid dendritic cells.

Obstacles that impede DNA replication can lead to DNA replication stress, a phenomenon causing genetic mutations and instability^[43]. Chronic DNA replication stress is a hallmark of cancer cells, which

provides a potential therapeutic target for cancer treatment^[44]. Our findings highlight the significance of DNA replication in colorectal cancer and its potential as a therapeutic target. *FEN1*, a member of the DNA

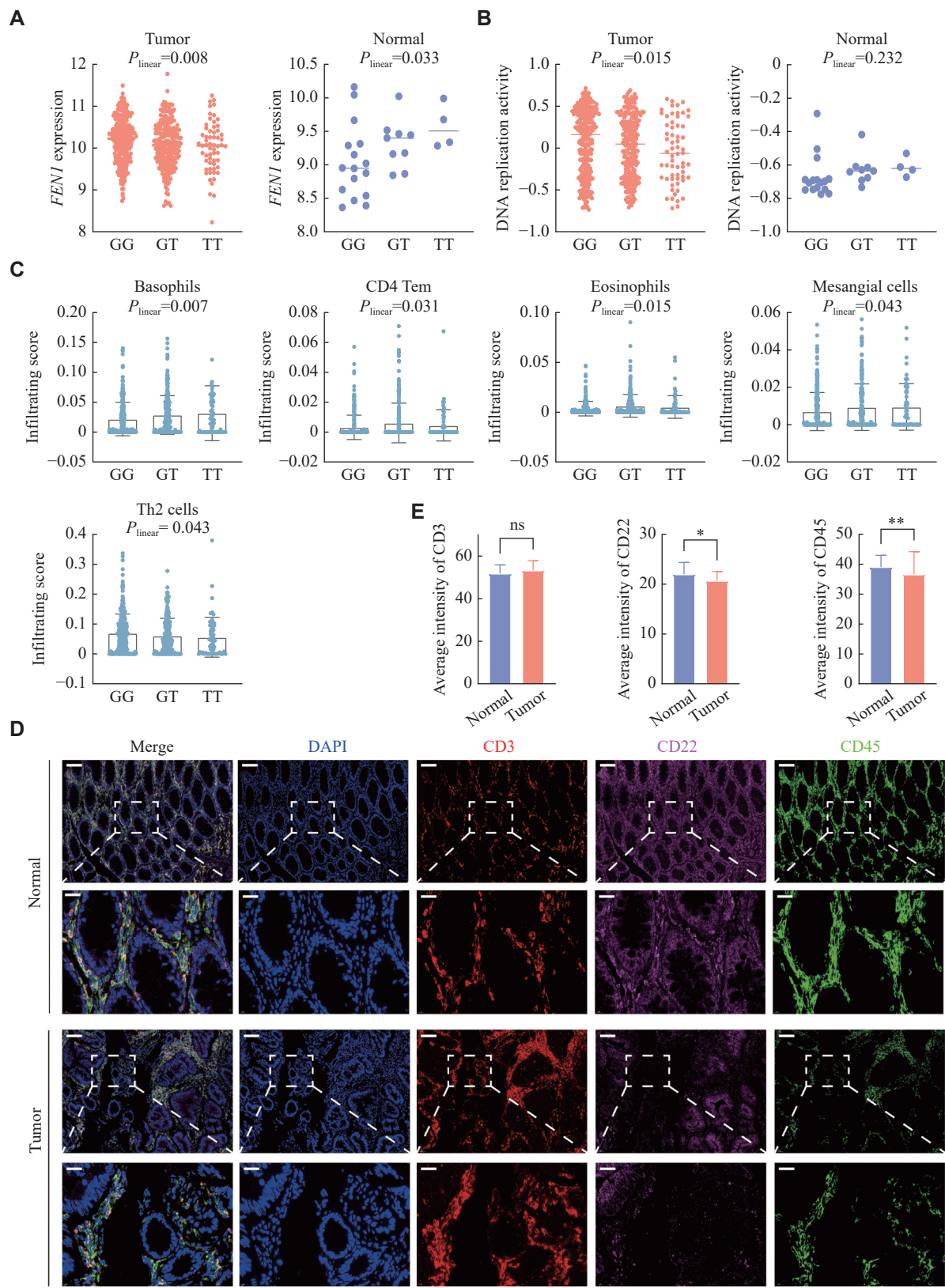
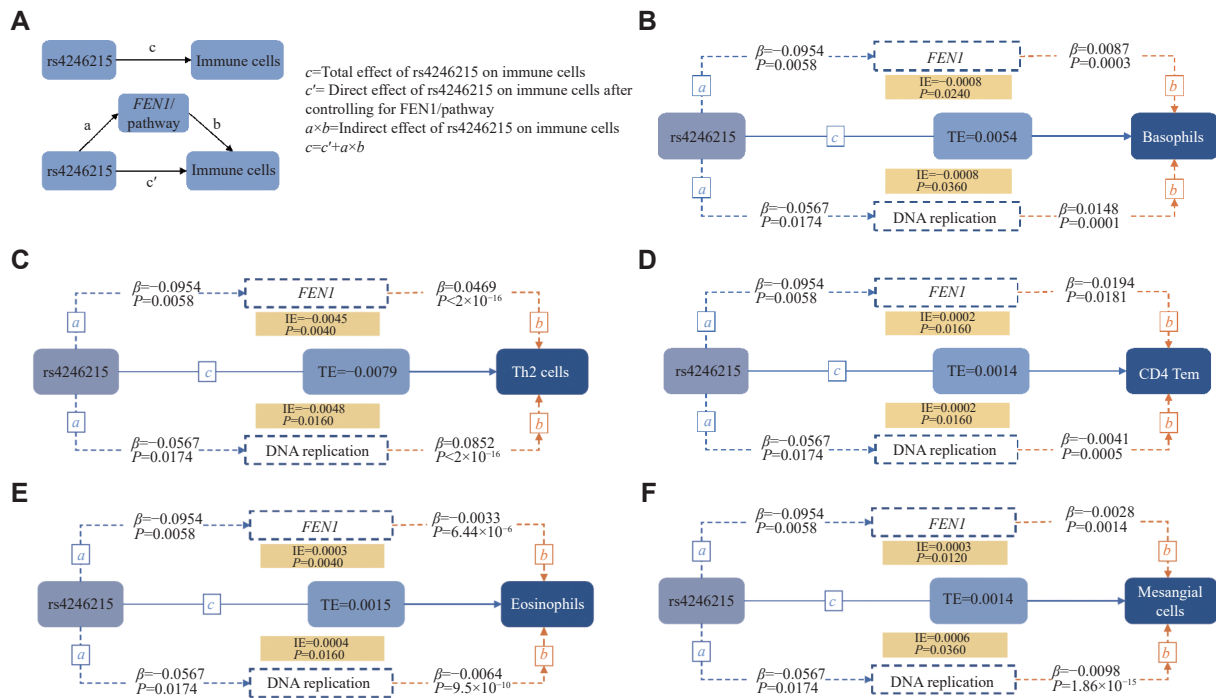


Fig. 5 Genetic effect of rs4246215 and multiplex immunofluorescence evaluation. A: The opposite genetic effect of rs4246215 on *FEN1* in colorectal tumors (left) and normal tissues (right), TCGA dataset. Gene expression data were log₂ transformed. B: Genetic effect of rs4246215 on DNA replication activity in colorectal tumors (left) and normal tissues (right), TCGA dataset. C: Genetic effect of rs4246215 on immunity (TCGA dataset). The *P*-value was calculated by linear regression. D: Representative multiple immunofluorescence images of CD3, CD22 and CD45 in colorectal tumors and normal tissues. Scale bar: 100 μ m for the first and third rows and 25 μ m for the second and forth rows, respectively. E: The average intensity of CD3, CD22 and CD45 in colorectal tumors and normal tissues. The *P*-value was calculated by the Wilcoxon test. **P* < 0.05 and ***P* < 0.01. Abbreviations: TCGA, The Cancer Genome Atlas; Tem, effector memory T cells.



replication pathway, is also involved in other DNA metabolic pathways, including telomere stability maintenance and apoptotic DNA fragmentation^[45–46]. *In vivo* and *in vitro* studies have demonstrated that *FEN1* is frequently overexpressed in various cancers, such as breast cancer, glioma, and hepatocellular cancer, and that its upregulation promotes tumorigenesis and cancer progression^[47–50]. Additionally, *FEN1* overexpression can lead to genome instability and impair DNA replication through its interaction with PCNA^[51]. We observed consistent FEN1 overexpression in colorectal tumors at both the RNA and protein levels, suggesting that FEN1 may serve as a promising biomarker for the diagnosis of colorectal cancer.

Tumor immune microenvironment consists of tumor cells and various non-tumor cells and plays a critical role in tumor initiation, progression, and response to therapy^[53]. In the current study, we

Recent studies have demonstrated that targeting DNA replication stress can modulate the immune microenvironment and improve immunotherapy efficacy^[63]. The current study contributes to this knowledge by revealing the intricate interplay among genetic variants, DNA replication, and the tumor immune microenvironment in colorectal cancer development. Investigating genetic variants that influence the infiltration of immune cells at the

genomic level may provide new insights into the regulation of immune microenvironments and facilitate the development of novel immunotherapeutic strategies.

Multiplex immunofluorescence is a powerful tool to study the spatial tumor immune microenvironment of limited tissue specimens, which improves the understanding of tumor-immune interactions^[64–65]. We validated our bioinformatic findings by using multiplex immunofluorescence to quantify immune cells and DNA replication activity in colorectal tumor and normal tissues at the protein level. Our results confirmed the spatial distribution of immune cells and tumor cells, and revealed differential infiltration of basophils, eosinophils, and FEN1 in colorectal tumors, compared with normal tissues. However, further studies are needed to better characterize other immune cell components in the colorectal tumor immune microenvironment.

It is important to acknowledge the limitations of the current study. We did not conduct *in vitro* and *in vivo* experiments to validate the differentially expressed genes found in the tumors and immune cell components. Future experiments should investigate immune-mediated cytotoxicity as a crucial mechanism in the tumor microenvironment. Additionally, the sample size of the cohort used to validate the data from public datasets was relatively small. Large-scale omics datasets are needed to confirm our results. Moreover, the current study focused on one SNP locus, and future studies should investigate other potential SNPs and their effects on immune microenvironment.

In summary, the current study uncovers the functional involvement of *FEN1* in colorectal cancer susceptibility through the DNA replication process and highlights the effect of the causal variant rs4246215 G>T on the colorectal tumor immune microenvironment. The current study contributes to the understanding of the pathogenesis of colorectal cancer and sheds light on the "SNP-gene/pathway-immunity" scheme that may have implications for clinical practice.

Fundings

This study was partly supported by the National Natural Science Foundation of China (Grant No. 82173601) and Yili & Jiangsu Joint Institute of Health (Grant No. y12021ms02).

Acknowledgments

We thank the participants who generously gave

their help in the study. We also thank the Genetics and Epidemiology of Colorectal Cancer Consortium and Biobank Japan for providing European and East Asian colorectal cancer GWAS data.

References

- [1] Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. *CA Cancer J Clin*, 2021, 71(3): 209–249.
- [2] Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2022[J]. *CA Cancer J Clin*, 2022, 72(1): 7–33.
- [3] Yurgelun MB, Kulke MH, Fuchs CS, et al. Cancer susceptibility gene mutations in individuals with colorectal cancer[J]. *J Clin Oncol*, 2017, 35(10): 1086–1095.
- [4] Jin K, Ren C, Liu Y, et al. An update on colorectal cancer microenvironment, epigenetic and immunotherapy[J]. *Int Immunopharmacol*, 2020, 89: 107041.
- [5] Zhang B, Bueno Venegas A, Hickson ID, et al. DNA replication stress and its impact on chromosome segregation and tumorigenesis[J]. *Semin Cancer Biol*, 2019, 55: 61–69.
- [6] Macheret M, Halazonetis TD. DNA replication stress as a hallmark of cancer[J]. *Annu Rev Pathol*, 2015, 10: 425–448.
- [7] Anderson NM, Simon MC. The tumor microenvironment[J]. *Curr Biol*, 2020, 30(16): R921–R925.
- [8] Hinshaw DC, Shevde LA. The tumor microenvironment innately modulates cancer progression[J]. *Cancer Res*, 2019, 79(18): 4557–4566.
- [9] Tan WCC, Nerurkar SN, Cai H, et al. Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy[J]. *Cancer Commun (Lond)*, 2020, 40(4): 135–153.
- [10] Taube JM, Akturk G, Angelo M, et al. The Society for Immunotherapy of Cancer statement on best practices for multiplex immunohistochemistry (IHC) and immunofluorescence (IF) staining and validation[J]. *J Immunother Cancer*, 2020, 8(1): e000155.
- [11] Xiao Y, Yu D. Tumor microenvironment as a therapeutic target in cancer[J]. *Pharmacol Ther*, 2021, 221: 107753.
- [12] Misra BB, Langefeld CD, Olivier M, et al. Integrated omics: tools, advances, and future approaches[J]. *J Mol Endocrinol*, 2019, 62(1): R21–R45.
- [13] Wörheide MA, Krumsiek J, Kastenmüller G, et al. Multi-omics integration in biomedical research - A metabolomics-centric review[J]. *Anal Chim Acta*, 2021, 1141: 144–162.
- [14] Karczewski KJ, Snyder MP. Integrative omics for health and disease[J]. *Nat Rev Genet*, 2018, 19(5): 299–310.
- [15] Du M, Garcia JGN, Christie JD, et al. Integrative omics provide biological and clinical insights into acute respiratory distress syndrome[J]. *Intensive Care Med*, 2021, 47(7): 761–771.
- [16] Gao Q, Zhu H, Dong L, et al. Integrated proteogenomic

- characterization of HBV-related hepatocellular carcinoma[J]. *Cell*, 2019, 179(2): 561–577.e22.
- [17] Chatsirisupachai K, Lesluyes T, Paraoan L, et al. An integrative analysis of the age-associated multi-omic landscape across cancers[J]. *Nat Commun*, 2021, 12(1): 2345.
- [18] Xin J, Du M, Gu D, et al. Combinations of single nucleotide polymorphisms identified in genome-wide association studies determine risk for colorectal cancer[J]. *Int J Cancer*, 2019, 145(10): 2661–2669.
- [19] Wang M, Gu D, Du M, et al. Common genetic variation in *ETV6* is associated with colorectal cancer susceptibility[J]. *Nat Commun*, 2016, 7: 11478.
- [20] Zheng R, Zhang K, Tan S, et al. Exosomal circLPA1 functions in colorectal cancer diagnosis and tumorigenesis through suppressing *BRD4* via METTL3-eIF3h interaction[J]. *Mol Cancer*, 2022, 21(1): 49.
- [21] Ishigaki K, Akiyama M, Kanai M, et al. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases[J]. *Nat Genet*, 2020, 52(7): 669–679.
- [22] Meng Y, Du M, Gu D, et al. Genome-wide association analyses identify CATSPERE as a mediator of colorectal cancer susceptibility and progression[J]. *Cancer Res*, 2022, 82(6): 986–997.
- [23] Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer[J]. *Nat Genet*, 2019, 51(1): 76–87.
- [24] Peters U, Jiao S, Schumacher FR, et al. Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis[J]. *Gastroenterology*, 2013, 144(4): 799–807.e24.
- [25] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles[J]. *Proc Natl Acad Sci U S A*, 2005, 102(43): 15545–15550.
- [26] Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data[J]. *BMC Bioinformatics*, 2013, 14: 7.
- [27] Yu GC, Wang LG, Han YY, et al. clusterProfiler: an R package for comparing biological themes among gene clusters[J]. *OMICS*, 2012, 16(5): 284–287.
- [28] Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments[J]. *Stat Appl Genet Mol Biol*, 2004, 3: 3.
- [29] MAQC Consortium. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements[J]. *Nat Biotechnol*, 2006, 24(9): 1151–1161.
- [30] de Leeuw CA, Mooij JM, Heskes T, et al. MAGMA: generalized gene-set analysis of GWAS data[J]. *PLoS Comput Biol*, 2015, 11(4): e1004219.
- [31] Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles[J]. *Nat Methods*, 2015, 12(5): 453–457.
- [32] Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape[J]. *Genome Biol*, 2017, 18(1): 220.
- [33] Hall MA, Verma A, Brown-Gentry KD, et al. Detection of pleiotropy through a Phenome-wide association study (PheWAS) of epidemiologic data as part of the Environmental Architecture for Genes Linked to Environment (EAGLE) study[J]. *PLoS Genet*, 2014, 10(12): e1004678.
- [34] Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results[J]. *Bioinformatics*, 2010, 26(18): 2336–2337.
- [35] Shabalín AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations[J]. *Bioinformatics*, 2012, 28(10): 1353–1358.
- [36] Gu D, Li S, Ben S, et al. Circadian clock pathway genes associated with colorectal cancer risk and prognosis[J]. *Arch Toxicol*, 2018, 92(8): 2681–2689.
- [37] Xin J, Du M, Jiang X, et al. Systematic evaluation of the effects of genetic variants on PIWI-interacting RNA expression across 33 cancer types[J]. *Nucleic Acids Res*, 2021, 49(1): 90–97.
- [38] Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models[J]. *Behav Res Methods*, 2008, 40(3): 879–891.
- [39] Tingley D, Yamamoto T, Hirose K, et al. Mediation: R package for causal mediation analysis[J]. *J Stat Software*, 2014, 59(5): 1–38.
- [40] Revelle W. Psych: procedures for psychological, psychometric, and personality research[EB/OL]. [2023-04-01] <https://CRAN.R-project.org/package=psych>.
- [41] Viechtbauer W. Conducting meta-analyses in R with the metafor package[J]. *J Stat Software*, 2010, 36(3): 1–48.
- [42] Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans[J]. *Bioinformatics*, 2010, 26(17): 2190–2191.
- [43] Saxena S, Zou L. Hallmarks of DNA replication stress[J]. *Mol Cell*, 2022, 82(12): 2298–2314.
- [44] Ubhi T, Brown GW. Exploiting DNA replication stress for cancer treatment[J]. *Cancer Res*, 2019, 79(8): 1730–1739.
- [45] Parrish JZ, Yang C, Shen B, et al. CRN-1, a *Caenorhabditis elegans* FEN-1 homologue, cooperates with CPS-6/EndoG to promote apoptotic DNA degradation[J]. *EMBO J*, 2003, 22(13): 3451–3460.
- [46] Saharia A, Guittat L, Crocker S, et al. Flap endonuclease 1 contributes to telomere stability[J]. *Curr Biol*, 2008, 18(7): 496–500.
- [47] Flach KD, Periyasamy M, Jadhav A, et al. Endonuclease FEN1 coregulates ER α activity and provides a novel drug interface in tamoxifen-resistant breast cancer[J]. *Cancer Res*, 2020, 80(10): 1914–1926.
- [48] He L, Luo L, Zhu H, et al. FEN1 promotes tumor progression and confers cisplatin resistance in non-small-cell lung cancer[J]. *Mol Oncol*, 2017, 11(9): 1302–1303.
- [49] Zhang J, Chen M, Pang Y, et al. Flap endonuclease 1 and

- DNA-PKcs synergistically participate in stabilizing replication fork to encounter replication stress in glioma cells[J]. *J Exp Clin Cancer Res*, 2022, 41(1): 140.
- [50] Bian S, Ni W, Zhu M, et al. Flap endonuclease 1 facilitated hepatocellular carcinoma progression by enhancing USP7/MDM2-mediated P53 inactivation[J]. *Int J Biol Sci*, 2022, 18(3): 1022–1038.
- [51] Becker JR, Gallo D, Leung W, et al. Flap endonuclease overexpression drives genome instability and DNA damage hypersensitivity in a PCNA-dependent manner[J]. *Nucleic Acids Res*, 2018, 46(11): 5634–5650.
- [52] Moazeni-Roodi A, Ghavami S, Ansari H, et al. Association between the flap endonuclease 1 gene polymorphisms and cancer susceptibility: an updated meta-analysis[J]. *J Cell Biochem*, 2019, 120(8): 13583–13597.
- [53] Galon J, Angell HK, Bedognetti D, et al. The continuum of cancer immunosurveillance: prognostic, predictive, and mechanistic signatures[J]. *Immunity*, 2013, 39(1): 11–26.
- [54] Xia Y, Rao L, Yao H, et al. Engineering macrophages for cancer immunotherapy and drug delivery[J]. *Adv Mater*, 2020, 32(40): 2002054.
- [55] Anderson NR, Minutolo NG, Gill S, et al. Macrophage-based approaches for cancer immunotherapy[J]. *Cancer Res*, 2021, 81(5): 1201–1208.
- [56] Mahmoud SMA, Lee AHS, Paish EC, et al. The prognostic significance of B lymphocytes in invasive carcinoma of the breast[J]. *Breast Cancer Res Treat*, 2012, 132(2): 545–553.
- [57] Edin S, Kaprio T, Hagström J, et al. The prognostic importance of CD20⁺ B lymphocytes in colorectal cancer and the relation to other immune cell subsets[J]. *Sci Rep*, 2019, 9(1): 19997.
- [58] Cabrita R, Lauss M, Sanna A, et al. Tertiary lymphoid structures improve immunotherapy and survival in melanoma[J]. *Nature*, 2020, 577(7791): 561–565.
- [59] Garnelo M, Tan A, Her Z, et al. Interaction between tumour-infiltrating B cells and T cells controls the progression of hepatocellular carcinoma[J]. *Gut*, 2017, 66(2): 342–351.
- [60] Wang S, Liu W, Ly D, et al. Tumor-infiltrating B cells: their role and application in anti-tumor immunity in lung cancer[J]. *Cell Mol Immunol*, 2019, 16(1): 6–18.
- [61] Sautès-Fridman C, Verneau J, Sun C, et al. Tertiary lymphoid structures and B cells: clinical impact and therapeutic modulation in cancer[J]. *Semin Immunol*, 2020, 48: 101406.
- [62] Mao Y, Feng Q, Zheng P, et al. Low tumor infiltrating mast cell density confers prognostic benefit and reflects immunoactivation in colorectal cancer[J]. *Int J Cancer*, 2018, 143(9): 2271–2280.
- [63] Zhang W, Liu W, Jia L, et al. Targeting KDM4A epigenetically activates tumor-cell-intrinsic immunity by inducing DNA replication stress[J]. *Mol Cell*, 2021, 81(10): 2148–2165.e9.
- [64] Taube JM, Roman K, Engle EL, et al. Multi-institutional TSA-amplified multiplexed immunofluorescence reproducibility evaluation (MITRE) study[J]. *J Immunother Cancer*, 2021, 9(7): e002197.
- [65] Yagi Y, Aly RG, Tabata K, et al. Three-dimensional histologic, immunohistochemical, and multiplex immunofluorescence analyses of dynamic vessel co-option of spread through air spaces in lung adenocarcinoma[J]. *J Thorac Oncol*, 2020, 15(4): 589–600.

Submit to the *Journal of Biomedical Research* at
<http://www.jbr-pub.org.cn/>

