Cancer Biol Med 2025. doi: 10.20892/j.issn.2095-3941.2025.0254

LETTER

Radiomic model for preoperative prediction of mismatch repair deficiency in gastric cancer: a multicenter study integrating tumor sub-region radiomics and transcriptomics

Siwei Pan[1,2,3]*, Enze Li[1]*, Guoliang Zheng[4]*, Yi You[5]*, Yanqiang Zhang[1,2,3], Mengxuan Cao[2,3,6], Ruolan Zhang[1], Qing Yang[1], Yizhou Wei[1], Weiwei Zhu[1], Ke Shen[1], Chencui Huang[5], Jingxu Xu[5], Lijing Wang[7], Zaisheng Ye[8], Zhiyuan Xu[1,2,3], Can Hu[1,2,3]

[1]Department of Gastric Surgery, Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou 310022, China; [2]Key Laboratory of Prevention, Diagnosis and Therapy of Upper Gastrointestinal Cancer of Zhejiang Province, Hangzhou 310022, China; [3]Zhejiang Provincial Research Center for Upper Gastrointestinal Tract Cancer, Zhejiang Cancer Hospital, Hangzhou 310022, China; [4]Department of Gastric Surgery, Cancer Hospital of China Medical University, Cancer Hospital of Dalian University of Technology, Liaoning Cancer Hospital & Institute, Shenyang 110042, China; [5]Department of Research Collaboration, R&D Center, Hangzhou Deepwise & League of PHD Technology Co., Ltd, Hangzhou 311101, China; [6]Department of Radiology, Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou 310022, China; [7]Department of Ultrasound, Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou 310022, China; [8]Department of Gastrointestinal Surgical Oncology, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fuzhou 350000, China

Gastric cancer (GC) remains a major global health challenge, because of its poor prognosis and limited treatment options in advanced stages[1,2]. Recent advancements in immunotherapy, highlighted by the findings of the CHECKMATE-649, ORIENT-16, and KEYNOTE-859 trials, have markedly transformed the treatment paradigm for advanced gastric cancer (AGC)[3-5]. Patients with mismatch repair deficient (dMMR)/microsatellite instability-high (MSI-H) AGC are highly responsive to immunotherapies such as pembrolizumab and nivolumab[6]. Clinical trials have revealed a 34–71% decrease in mortality risk for patients with dMMR/MSI-H AGC receiving immunotherapy, in both the cohort with a programmed cell death ligand 1 combined positive score ≥ 5 and the overall cohort[3,5,7]. However, patients with mismatch repair-proficient (pMMR)/microsatellite stable (MSS)/microsatellite instability-low (MSI-L) AGC showed lower mortality risk decreases of 18%–29% in the immunotherapy groups compared to dMMR/MSI-H AGC.

Currently, the identification of dMMR/MSI-H relies on immunohistochemistry (IHC) or polymerase chain reaction to evaluate the expression of MMR-related markers[8]. However, sufficient tumor tissue is not always available, and the cost and invasive nature of tissue-based molecular testing can hinder widespread clinical use. Therefore, an urgent need exists for a non-invasive, cost-effective approach to screen and diagnose dMMR/MSI-H in AGC. Recent advancements in radiomics and artificial intelligence have facilitated the development of imaging-based prediction models using features extracted from medical images[9]. On this basis, we sought to use radiomics data to predict the MMR status of patients with GC and to explore the biological relevance of radiomics features combined with transcriptomic data.

*These authors contributed equally to this work.
Correspondence to: Zaisheng Ye, Zhiyuan Xu and Can Hu
E-mail: flyingengel@sina.com, xuzy@zjcc.org.cn and hucan@zjcc.org.cn
ORCID ID: https://orcid.org/0000-0002-5648-9330, https://orcid.org/0000-0002-7257-6856 and https://orcid.org/0000-0002-8687-8310
Received May 17, 2025; accepted August 12, 2025; published online September 8, 2025.
Available at www.cancerbiomed.org
©2025 The Authors. Creative Commons Attribution-NonCommercial 4.0 International License

Overview of the study design and research cohorts

This study was aimed at developing a model for predicting MMR status in patients with GC by using enhanced computed tomography (CT) images obtained at initial diagnosis. Moreover, we sought to integrate transcriptomic and proteomic data to elucidate the biological relevance of the radiomic model (**Figure S1**). A multi-center clinical and radiomic dataset comprising 1,074 patients was used (detailed information in **Supplementary Methods**). The patients were divided into training (dMMR, 14.8%), internal validation (dMMR, 13.2%), and external validation (dMMR, 9.4%) cohorts (**Table S1**) for model construction and validation. To further investigate the biological relevance of the radiomic models and their features, we performed bulk RNA sequencing on tumor samples from 55 patients with GC in the institution I cohort.

Sub-region segmentation from the region of interest

All patients underwent enhanced CT examination before radical gastrectomy. Tumor segmentation and region of interest (ROI) delineation were performed by 2 experienced radiologists in ITK-SNAP software. Radiomics feature extraction in all 4 regions included shape features, first-order features, and texture features. Square, square root, logarithm, exponential, and gradient transformation were applied to the original features (**Table S2**). To achieve efficient and reproducible tumor segmentation, we implemented the Simple Linear Iterative Clustering (SLIC) algorithm, an adaptive k-means clustering method specifically optimized for medical image analysis. The SLIC approach involves 3 key steps: (1) initialization of cluster centroids (seeds) distributed across the tumor volume, (2) computation of 3D gradient distances between each voxel and its neighboring seeds in both spatial (x, y, z) and radiomic (CT intensity) domains, and (3) iterative centroid updating until convergence. Through this process, each tumor was automatically partitioned into 3 phenotypically distinct sub-regions: sub-region1 (low intensity), sub-region2 (intermediate intensity), and sub-region3 (high intensity). We further conducted a comparative analysis of radiomic features across the 3 sub-regions and visualized the top 20 most differentially expressed features (**Figure S2**).

Radiomic panel development, model construction, and diagnostic performance

Radiomic panels associated with MMR status were developed for different spatial sub-regions of the tumors. These features were subjected to correlation analysis and LASSO dimensionality reduction, thus resulting in the selection of 32 features for the entire region, 18 for sub-region1, 29 for sub-region2, and 25 for sub-region3. The performance of each model [logistic regression (LR), support vector machines (SVM), and XGBoost] was evaluated with the area under the receiver operating characteristic curve (AUC) in each sub-region (**Figure 1A**). LR demonstrated higher performance than SVM and XGBoost in predicting MMR status. Among the 4 regions, the LR model for sub-region3 consistently outperformed the others in both the internal (AUC = 0.828) and external (AUC = 0.800) validation sets, by exhibiting higher AUC, accuracy, and specificity than observed in the other sub-regions (**Figure 1B, Table S3**).

After the features of all 3 sub-regions were combined, the new model achieved the highest predictive performance, with an AUC of 0.863 and 0.849 in the internal and external validation cohorts, respectively (**Table S3**). With the optimal sub-region3 model and the combined regional model (sub-region1, sub-region2, and sub-region3), patients in the training set were stratified into low-, moderate-, and high-risk groups according to model scores. The same cutoff values were applied to stratify patients in the external validation set (**Figure 1C**). In the sub-region3 model, the dMMR risk rates for groups with low (< 0.3009), moderate (≥ 0.3009 and < 0.4817), and high (≥ 0.4817) risk were 6.1%, 26.1%, and 55.6%, respectively. Similarly, in the combined regional model, the corresponding risk rates were 3.7% (< 0.2067), 17.2% (≥ 0.2067 and < 0.4245), and 47.2% (≥ 0.4245). Furthermore, the calibration curves of the LR model for sub-region3 and the combined regional model, as well as the results of the decision curve analyses, all demonstrated excellent fitting performance and clinical application value (**Figure S3**).

A

Region	Training cohort			Internal validation cohort			External validation cohort		
	LR	SVM	XGBoost	LR	SVM	XGBoost	LR	SVM	XGBoost
All region	0.908	0.893	0.913	0.786	0.787	0.766	0.775	0.707	0.752
	(0.894–0.922)	(0.870–0.908)	(0.898–0.927)	(0.699–0.873)	(0.701–0.872)	(0.662–0.870)	(0.697–0.853)	(0.622–0.791)	(0.670–0.833)
Sub-region1	0.886	0.884	0.879	0.707	0.696	0.706	0.684	0.677	0.648
	(0.870–0.903)	(0.867–0.900)	(0.861–0.896)	(0.590–0.823)	(0.579–0.814)	(0.606–0.806)	(0.589–0.779)	(0.580–0.773)	(0.555–0.740)
Sub-region2	0.881	0.853	0.899	0.754	0.742	0.693	0.734	0.676	0.550
	(0.864–0.897)	(0.835–0.872)	(0.883–0.915)	(0.653–0.855)	(0.636–0.849)	(0.576–0.811)	(0.647–0.822)	(0.647–0.822)	(0.451–0.650)
Sub-region3	0.904	0.891	0.917	0.828	0.824	0.747	0.800	0.753	0.712
	(0.890–0.918)	(0.875–0.906)	(0.903–0.931)	(0.731–0.924)	(0.724–0.923)	(0.637–0.857)	(0.735–0.866)	(0.683–0.823)	(0.623–0.801)

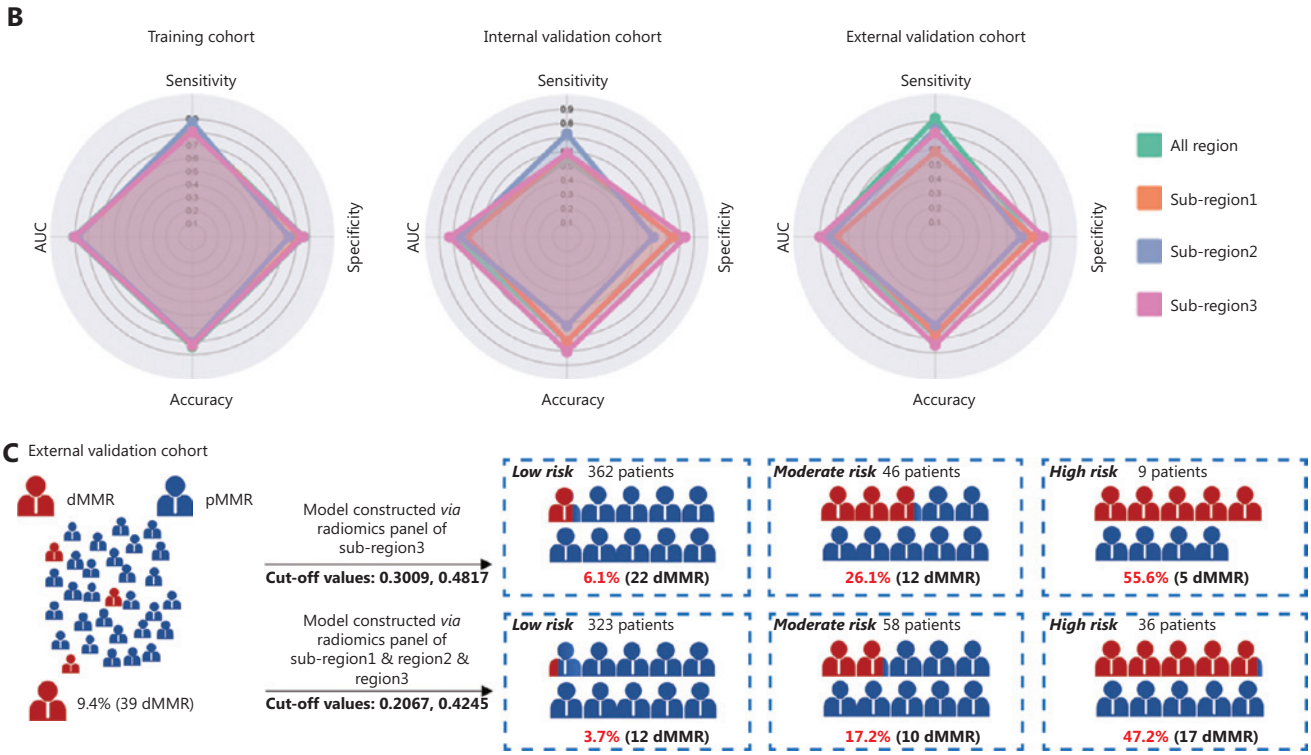


Figure 1 Continued

Transcriptomic alterations associated with radiomic panels of sub-regions

The radiomic features selected for model construction effectively predicted the MMR status of GC, thus suggesting potential biological implications underlying these features. To explore this hypothesis, we first performed gene set variation analysis on transcriptomic data from 55 patients with GC (**Figure 1D**) and observed significant differences between patients with dMMR and pMMR in the enriched functions, including the TGFβ signaling pathway,

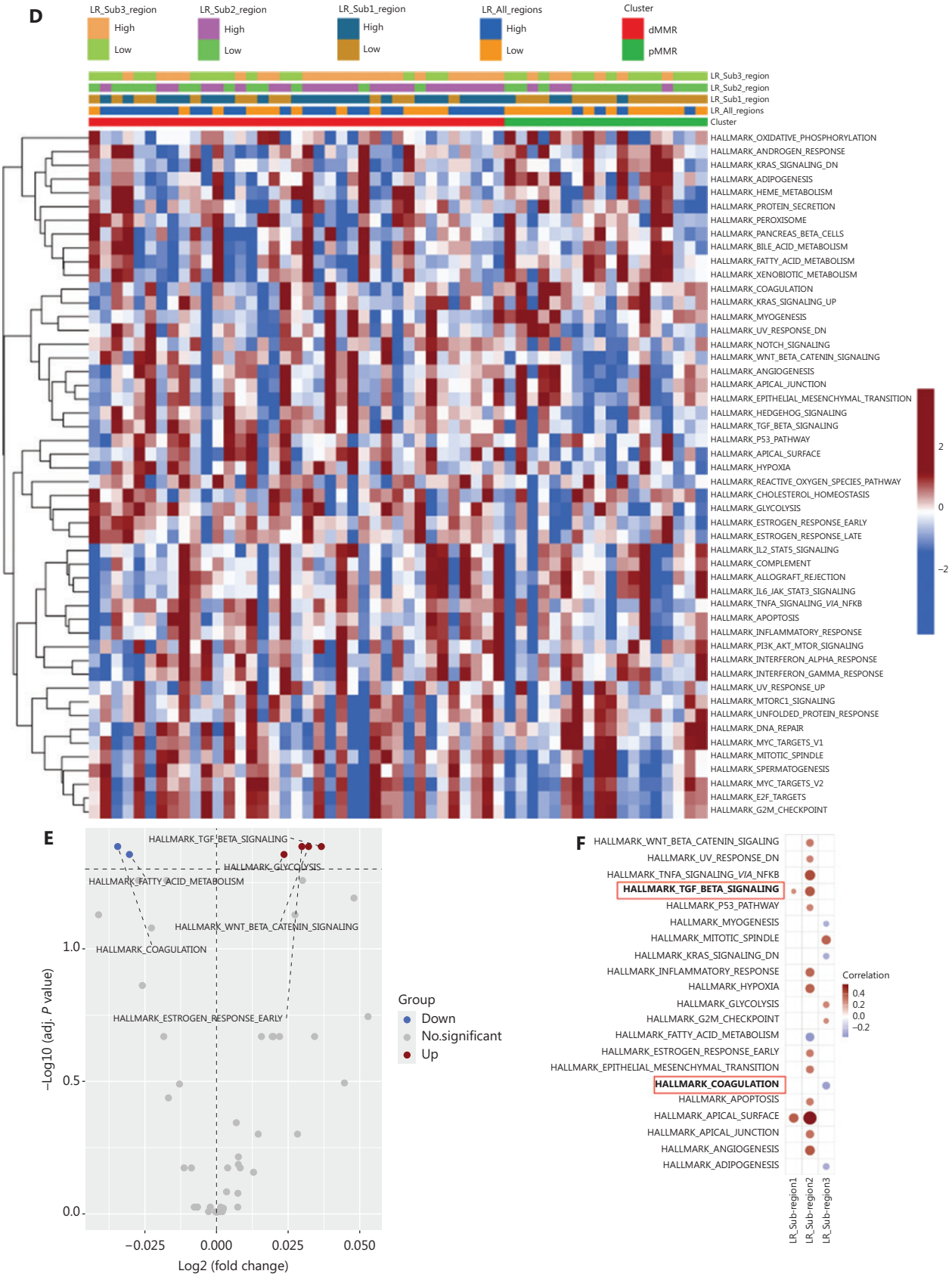


Figure 1 Continued

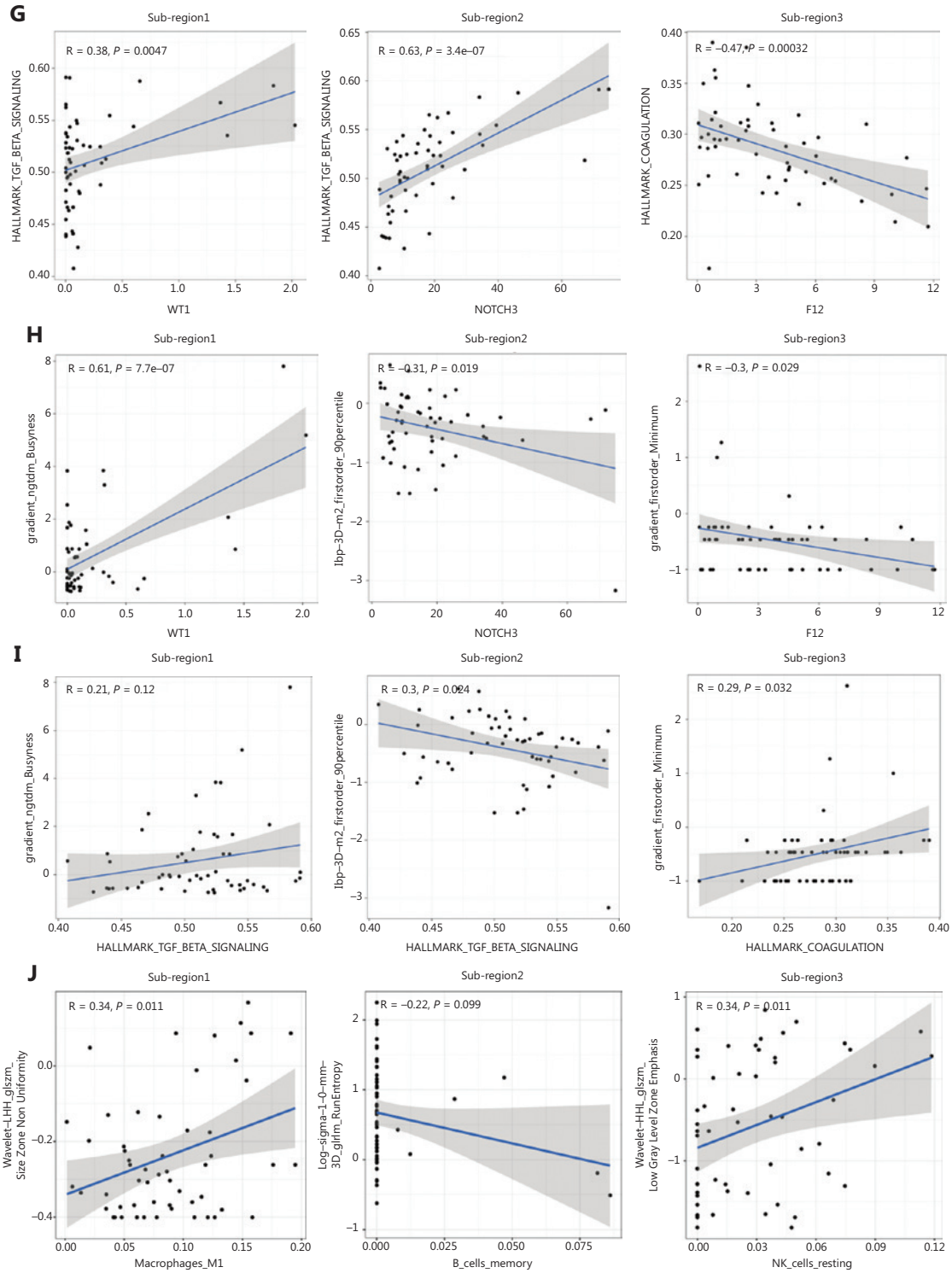


Figure 1 Continued

Wnt/β-catenin signaling pathway, estrogen response, fatty acid metabolism, and coagulation (**Figure 1E**). Correlation analyses between these enriched functions with significant differences between dMMR and pMMR and the radiomic models of the 3 sub-regions indicated that the sub-region1 model was significantly associated with TGFβ signaling,

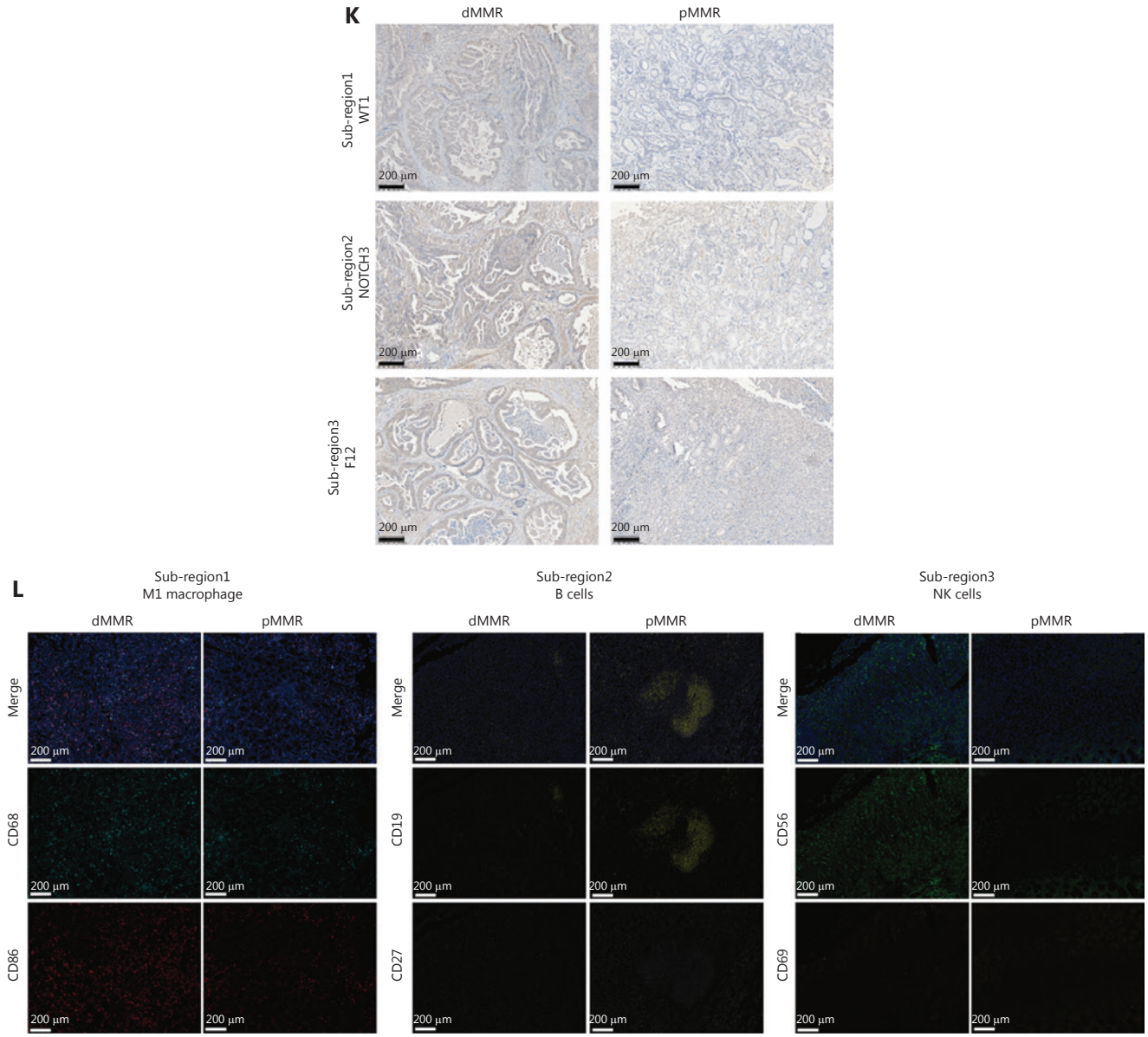


Figure 1 Radiomic models for preoperative prediction of dMMR in GC, linked with transcriptomic data. A. AUC values and 95% confidence intervals of each MMR-status prediction model constructed by 3 machine learning models in each sub-region and the entire region for each cohort. B. Radar maps of LR model performance in each sub-region and the entire region for each cohort. C. Clinical significance of models for the evaluation of dMMR risk. D. Heatmap of gene set variation analysis enrichment analyses. E. Differentially enriched functions between patients with dMMR vs. pMMR GC. F. Correlation analyses of radiomic models for MMR status prediction in each sub-region and enriched functions ($P < 0.05$ is represented by a dot). G. Correlation analyses between characteristic enriched functions and characteristic genes in each sub-region. H. Correlation analyses between selected radiomic features and characteristic genes in each sub-region. I. Correlation analyses between characteristic enriched functions and selected radiomic features in each sub-region. J. Correlation analyses between characteristic immune cells and selected radiomic features in each sub-region. K. IHC results of proteins coded by characteristic genes in each sub-region of patients with dMMR and pMMR. scale bar, 200 μm. L. Multiplex IHC results of the infiltration degree of characteristic immune cells in each sub-region in patients with dMMR or pMMR. scale bar, 200 μm. AUC, receiver operating characteristic curve; LR, logistic regression; SVM, support vector machines.

whereas the models for sub-region2 and sub-region3 were associated with TGFβ signaling and coagulation, respectively ($P < 0.05$, **Figures 1F and S4A**). Further investigation identified 3,770 differentially expressed genes (DEGs) between the tumor and adjacent normal tissues in dMMR GC and 3,192 DEGs between dMMR and pMMR GC tumor

tissues, with 975 overlapping DEGs (**Figure S4B**). Further correlation analyses between the expression of the aforementioned DEGs and the model's scores for the corresponding sub-regions identified 22 DEGs significantly associated with the radiomic model's score for sub-region1, 57 associated with sub-region2, and 106 associated with sub-region3 ($P < 0.01$, **Table S4**). Finally, the characteristic genes *WT1* for sub-region1, *NOTCH3* for sub-region2, and *F12* for sub-region3 were identified according to their strong correlations with radiomic features and functional pathways (all $P < 0.05$, **Figures 1G and S4C**). To further elucidate the microscopic biological relevance of radiomic features, we conducted correlation analyses, which indicated that feature gradient_ngtdm_Busyness in sub-region1, lbp.3D.m2_firstorder_90Percentile in sub-region2, and gradient_firstorder_Minimum in sub-region3 significantly correlated with the respective characteristic genes (*WT1*, *NOTCH3*, and *F12*) of their respective regions (**Figures 1H and S4D**). Moreover, the selected radiomic feature from sub-region2 ($P = 0.024$) and sub-region3 ($P = 0.032$) demonstrated significant correlations with their corresponding characteristic functions (**Figure 1I**).

Immune infiltration alterations associated with radiomic panels of sub-regions

Transcriptomic data were further analyzed to evaluate the extent of immune cell infiltration within the tumor microenvironment. Correlation analyses between the infiltration degree of immune cells and the radiomic models' scores for the 3 sub-regions identified the characteristic immune cells of M1 macrophages for sub-region1, B cells for sub-region2, and natural killer (NK) cells for sub-region3, on the basis of their high relative coefficients and significant correlations with the respective models ($P < 0.05$, **Figure S5A**). Further correlation analyses between radiomic panels and the infiltration levels of these characteristic cells revealed that wavelet.LHH_glszm_SizeZoneNonUniformity in sub-region1 and wavelet.HHL_glszm_LowGrayLevelZoneEmphasis in sub-region3 demonstrated the strongest correlations with M1 macrophages and NK cells, respectively, whereas log.sigma.1.0.mm.3D_glrlm_RunEntropy in sub-region2 showed the highest relative coefficient, with a potential but non-significant correlation with B cell infiltration ($P = 0.099$) (**Figures 1J and S5B**).

Validation of characteristic radiomic features, proteins, and immune cells

Further validation was performed on the proteins encoded by the characteristic genes and the associated immune cells identified in prior analyses. During the initial verification, we conducted targeted biopsies in each radiomics-defined sub-region from postoperative specimens. The IHC results mirrored the trends observed in the radiomic analysis. Specifically, WT1 expression in sub-region1, NOTCH3 in sub-region2, and F12 in sub-region3 were higher in the tumor tissues of patients with dMMR *vs*. pMMR (**Figure 1K**). Moreover, multicolor fluorescence *in situ* hybridization demonstrated that immune cells corresponding to each sub-region showed infiltration patterns consistent with the previously described trends, thus further supporting the correlations (**Figure 1L**).

Biological relevance and application prospects of the radiomics model and panels

The integration of radiomic models with transcriptomic data revealed distinct biological characteristics within each sub-region, which were associated with characteristic functions, genes, and immune cell infiltration associated with MMR status. For instance, WT1 in sub-region1, NOTCH3 in sub-region2, and F12 in sub-region3 emerged as potential biomarkers for identifying MMR status at specific tumor sites. Moreover, differences in immune cell infiltration, such as M1 macrophages in sub-region1, B cells in sub-region2, and NK cells in sub-region3, were observed between patients with dMMR *vs*. pMMR. These findings suggested that targeting characteristic immune cells might offer new opportunities to enhance immunotherapy efficacy in patients with pMMR, a subgroup typically showing poor response rates to immunotherapy. Integrating multi-omics approaches to identify novel therapeutic targets for pMMR GC is a promising research direction. However, this study had several limitations, such as the lack of spatial transcriptomic data based on spatially targeted biopsies for deeper exploration of the biological relevance across different radiomic sub-regions. This aspect is a direction for future research.

In conclusion, this study used initial CT images from patients with GC to construct and validate a radiomics model

that effectively predicts MMR status through a multi-regional combined approach. By integrating spatial heterogeneity within the radiomic model with transcriptomic data, we elucidated key biological differences between patients with dMMR *vs*. pMMR across sub-regions. Furthermore, the characteristic functions, genes, and immune cell types identified in each sub-region provide valuable insights into the molecular and cellular bases of GC heterogeneity.

Grant support

This study was supported by The National Key Research and Development Program of China (Grant no. 2021YFA0910100), Healthy Zhejiang One Million People Cohort (Grant no. K-20230085), Post-doctoral Innovative Talent Support Program (Grant no. BX2023375), Lingyan Project of Zhejiang Provincial Department of Science and Technology (Grant no. 2025C02059), the National Natural Science Foundation of China (Grant nos. 82304946, 82473489, and 82403546), Natural Science Foundation of Zhejiang Province (Grant nos. LR21H280001, LGF22H160056, ZCLQN25H1602, and LMS25H160006), Medicine and Health Science Fund of Zhejiang Province Health Commission (Grant nos. 2025KY047 and 2022KY658), and Traditional Chinese Medicine Science and Technology Project of Zhejiang Provincial Health Commission (Grant no. 2022ZA023).

Conflict of interest statement

No potential conflicts of interest are disclosed.

Author contributions

Guarantors of integrity for the entire study: CH, ZYX, ZSY.
Study concept/design: CH, ZYX, ZSY, SWP.
Data acquisition: all authors.
Contribution of data or analysis tools: YY, YQZ, MXC, CCH, JXX, KS.
Data analysis and interpretation: SWP, GLZ, EZL, JQZ, JY.
Manuscript drafting: SWP, EZL, GLZ, YY.
Manuscript revision for important intellectual content: CH, ZYX, ZSY, SWP.
All authors read and approved the final manuscript.

Data availability statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. Some of the data results are presented in the **Supplementary Material**.

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024; 74: 12-49.
2. Joshi SS, Badgwell BD. Current treatment and recent progress in gastric cancer. CA Cancer J Clin. 2021; 71: 264-79.
3. Janjigian YY, Ajani JA, Moehler M, Shen L, Garrido M, Gallardo C, et al. First-line nivolumab plus chemotherapy for advanced gastric, gastro-esophageal junction, and esophageal adenocarcinoma: 3-year follow-up of the phase III CheckMate 649 trial. J Clin Oncol. 2024; 42: 2012-20.
4. Xu J, Jiang H, Pan Y, Gu K, Cang S, Han L, et al. Sintilimab plus chemotherapy for unresectable gastric or gastroesophageal junction cancer: the ORIENT-16 randomized clinical trial. JAMA. 2023; 330: 2064-74.
5. Rha SY, Oh DY, Yañez P, Bai Y, Ryu MH, Lee J, et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for HER2-negative advanced gastric cancer (KEYNOTE-859): a multicentre, randomised, double-blind, phase 3 trial. Lancet Oncol. 2023; 24: 1181-95.
6. Gervaso L, Ciardiello D, Oliveira RA, Borghesani M, Guidi L, Benini L, et al. Immunotherapy in the neoadjuvant treatment of gastrointestinal tumors: is the time ripe? J Immunother Cancer. 2024; 12: e008027.
7. Qiu MZ, Oh DY, Kato K, Arkenau T, Tabernero J, Correa MC, et al. Tislelizumab plus chemotherapy versus placebo plus chemotherapy as first line treatment for advanced gastric or gastro-oesophageal junction adenocarcinoma: RATIONALE-305 randomised, double blind, phase 3 trial. Br Med J. 2024; 385: e078876.
8. Strickland MR, Lander EM, Gibson MK, Ilson DH, Ajani JA, Klempner SJ. Gastroesophageal adenocarcinomas with defective mismatch repair: current knowledge and clinical management. J Natl Compr Canc Netw. 2024; 22: e237103.
9. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. Nat Rev Clin Oncol. 2019; 16: 703-15.

Cite this article as: Pan S, Li E, Zheng G, You Y, Zhang Y, Cao M, et al. Radiomic model for preoperative prediction of mismatch repair deficiency in gastric cancer: a multicenter study integrating tumor sub-region radiomics and transcriptomics. Cancer Biol Med. 2025; 22: 1210-1217. doi: 10.20892/j.issn.2095-3941.2025.0254