



OPEN ACCESS

EDITED BY

Chao Wang,
Shanghai Jiao Tong University, China

REVIEWED BY

Namuunaa Juramt,
Harvard Medical School, United States
Dulguun Juramt,
Charité University Medicine Berlin, Germany

*CORRESPONDENCE

Peng Li

✉ lipenglimo@163.com

Yifei Pan

✉ 13506641535@139.com

Xiaodong Zhang

✉ zhangxiaodong777@wmu.edu.cn

†These authors have contributed equally to this work and share last authorship

‡These authors share first authorship

RECEIVED 03 May 2025

ACCEPTED 25 June 2025

PUBLISHED 21 July 2025

CORRECTED 22 July 2025

CITATION

Zheng C, Zhu B, Chen Y, Shahid N, Hu Y, Ali Husain HMA, Ou B, Zhang Q, Jin H, Zheng Y, Li P, Pan Y and Zhang X (2025) Integrating proteomics and machine learning reveals characteristics and risks of lymph node-independent distant metastasis in colorectal cancer. *Front. Immunol.* 16:1622528. doi: 10.3389/fimmu.2025.1622528

COPYRIGHT

© 2025 Zheng, Zhu, Chen, Shahid, Hu, Ali Husain, Ou, Zhang, Jin, Zheng, Li, Pan and Zhang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Integrating proteomics and machine learning reveals characteristics and risks of lymph node-independent distant metastasis in colorectal cancer

Chenxiao Zheng^{1†}, Baiwang Zhu^{1†}, Yanyu Chen², Numan Shahid³, Yiwang Hu¹, Hajar Mansoor Ahmed Ali Husain⁴, Binbin Ou⁵, Qiongying Zhang⁶, Haobo Jin⁷, Yating Zheng¹, Peng Li^{4*†}, Yifei Pan^{1*†} and Xiaodong Zhang^{1,8*†}

¹Department of Colorectal Anal Surgery, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China, ²Clinical Medical College, Hangzhou Medical College, Hangzhou, Zhejiang, China, ³Department of Hepato-Pancreato-Biliary (HPB) Surgery, King's College Hospital, Denmark Hill, London, United Kingdom, ⁴Department of Hepatology, King's College Hospital, Denmark Hill, London, United Kingdom, ⁵General Practice Department, Hangzhou Gongshu Hospital of Integrated Traditional and Western Medicine, Hangzhou, Zhejiang, China, ⁶Department of Pathology, the First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China, ⁷Laboratory Animal Centre, Wenzhou Medical University, Wenzhou, Zhejiang, China, ⁸National Key Clinical Specialty (General Surgery), The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China

Background: Metastatic colorectal cancer (mCRC) poses significant treatment challenges, especially liver metastasis (CRLM). A notable proportion of CRC has synchronous metastasis independent of lymph node metastasis (LNM). The biological traits of lymph node-independent metastasis in CRC are unclear, and early synchronous metastasis is hard to predict with current imaging or clinicopathological methods.

Method: We collected samples from 12 CRC patients with synchronous distant metastasis without LNM (T1-3N0M1). Data-Independent Acquisition Mass Spectrometry (DIA-MS), multi-omics data integration, and machine learning were used to develop a Lymph node-Independent Metastasis Genes (LIMGs) signature to predict synchronous distant metastasis risk in stage I-II CRC patients and validate it in multi-cohort. Immune microenvironment across risk subgroups was calculated by Estimating Relative Subsets of RNA Transcripts (CIBERSORT). Tumor Mutation Burden (TMB), Microsatellite Instability (MSI) score, immune functions and immune checkpoint gene expression were analyzed to evaluate immunotherapy response. Single cell RNA sequencing (scRNA-seq) analysis illustrated the expression profile of integrin $\alpha 11$ (ITGA11) in CRC. Immunohistochemistry (IHC) confirmed its expression pattern, while wound healing and transwell assays elucidated the role of ITGA11 in CRC metastasis.

Results: The LIMGs signature demonstrated strong predictive performance of lymph node-independent synchronous metastasis across cohorts. The high-risk subgroup exhibited enhanced extracellular matrix (ECM) remodeling, epithelial-mesenchymal transition (EMT) and correlated with immunosuppressive tumor

microenvironment (TME), lower TMB and MSI score, indicating worse immunotherapy response. Additionally, machine learning reveal ITGA11's pivotal role in lymph node-independent metastasis. IHC scores showing significant discriminatory ability of ITGA11 across different samples. Wound healing and transwell assays reveal that the knockdown of ITGA11 hinders the migration and invasion of CRC SW480 cells.

Conclusion: Our findings suggest that EMT-related signature LIMGs significantly affects lymph node-independent distant metastasis in CRC and effectively predicts non-LNM synchronous metastasis in stage I-II CRC patients. LIMG ITGA11 may promote early metastasis by enhancing migration and invasion. These offering insights into precise risk stratification and treatment for CRC patients.

KEYWORDS

colorectal cancer, proteomics, machine learning, synchronous metastasis, immune microenvironment, Itga11

Introduction

CRC currently ranks as the third most common cancer worldwide and the second leading cause of cancer-related deaths, with over 1,800,000 new cases and nearly 900,000 deaths annually worldwide (1). Metastatic colorectal cancer (mCRC) is one of the challenging aspects in the treatment of CRC, with the liver being the primary site for metastasis (CRLM). Synchronous metastases refer to metastasis detected before or at the time of CRC diagnosis (2). 15%–25% of CRC patients present with distant metastasis at diagnosis, and the vast majority (80%–90%) of CRLM are initially unresectable (3). Liver metastasis is also the leading cause of death in CRC patients, resulting in a significant social burden.

Traditionally, it has been believed that cancer progression involves sequential spread of the tumor to local lymph nodes followed by distant metastasis. However, a considerable number of mCRC patients do not exhibit early systemic spread. Among these, CRLM often occur without lymph node metastasis (LNM). Data indicate that approximately 23% of synchronous liver metastases originate from stage I-II (N0) CRC, and 44% of metachronous metastases arise from N0 CRC (4). A study on resection of CRLM showed that among over 12,000 patients, 37% had no LNM (5). Furthermore, there was no difference in the incidence of liver metastases between patients with and without LNM (6). At the molecular level, CRC metastasis are often proven to originate from a dominant clone within the primary tumor and sharing a high degree of consistency in mutated genes. In contrast, polyclonal origins are more commonly observed in LNM, with 65% of cases showing that LNM and distant metastases arise from independent subclones within the primary tumor (7). Moreover, LNM exhibits a high rate of inconsistency in mutations compared to the primary tumor (8), suggesting lymph nodes may not always

be involved in distant metastasis. Animal models have further confirmed that CRC dissemination to the liver can occur independently of LNM, with direct hematogenous spread being a route for CRLM (9). This may imply that stage III and IV CRC may be considered as parallel progression from stage II disease rather than sequential progression.

An incidence model based on tumor size, time, and mutations shows that early metastasis in the majority (80%) of mCRC patients may occur before the primary tumor is clinically detectable (10). As disseminated tumor cells (DTCs) frequently colonize distant organs by the time of primary tumor detection, and they are undetectable with clinical imaging and patients remain asymptomatic regarding subclinical disease. Circulating tumor DNA (ctDNA) and circulating tumor cells (CTCs) show promise as biomarkers for micrometastasis but require enhanced sensitivity and clinical feasibility (11). Effective biomarkers based on tissue-based protein/RNA detection are needed, combining single-cell analysis, detection of ctDNA epigenetic modification, CTC, exosome, immune cell, cytokine may enable real-time predictive biomarker development.

Recent proteomic studies in CRC have revealed novel protein traits, molecular subtypes, and metastasis markers, underscoring molecular heterogeneity across clinicopathological subgroups (12). However, proteomic research on lymph node-independent distant metastasis in CRC remains limited. Epithelial-mesenchymal transition (EMT), which drives early CRC progression by diminishing cell-cell adhesion and apical polarity while enhancing invasion, is of particular interest (13). Here, we hypothesized that lymph node-independent distant metastasis in CRC arises from EMT-related micrometastasis and hematogenous routes. Our study aims to develop a predictive signature for direct distant metastasis risk in early-stage (I-II) CRC by integrating multi-omics data and

machine learning, thus refining risk stratification and guiding therapy. To this end, we analyzed 12 synchronous distant metastasis patients (T1-T3N0M1) using DIA-MS. Our findings identify EMT-linked LIMGs as key drivers of lymph node-independent metastasis, with high-risk samples exhibiting a more immunosuppressive tumor microenvironment that may facilitate early distant metastasis.

Materials and methods

Patients

For the DIA-MS analysis, the patient cohort was sourced from the Colorectal and Anal Surgery Department of the First Affiliated Hospital of Wenzhou Medical University, with the study having secured ethical approval (KY2022-183) from the hospital's Ethics Committee. Our study screened 271 mCRC patients who underwent simultaneous radical resection of primary tumors and distant metastases between 2018 and 2024. From them, 12 patients with a pathological stage of T1 - 3N0M1 were selected for specimen collection, as shown in [Figure 1A](#). The inclusion criteria were age 18 - 80, clinical diagnosis of synchronous distant metastasis, having undergone radical surgery, histopathological confirmation of colorectal adenocarcinoma, and classification as T1 - 3N0M1 stage according to the 8th edition of the AJCC/UICC TNM staging system. Exclusion criteria included lymph node metastasis, an insufficient number of examined lymph nodes (< 12), a history of other primary malignancies, neoadjuvant therapy, and multiple distant metastases. A detailed overview of the clinicopathological characteristics of the study cohort is presented in [Figure 1B](#) and [Supplementary Table S1](#).

Sample preparation

Formalin-fixed paraffinembedded (FFPE) samples of adjacent normal tissues, primary tumors, and distant metastases were collected from 12 CRC patients. Pathological examination by a pathologist confirmed the tumor areas and using hematoxylin-eosin-stained pathologic slides as reference. All pathological reports were cross diagnosed by two senior pathologists and reviewed by a third. To minimize specimen loss, the same type of tissue sections (4µm) from different patients were prepared and mixed into four composite samples for testing.

Protein extraction and peptide enzymatic digestion

For protein extraction, each sample was supplemented with an appropriate volume of SDT lysis buffer (4% SDS, 100 mM Tris-HCl, pH 7.6), followed by protein quantification using the BCA method. Subsequently, 15 µg of protein from each sample was mixed with 5×

loading buffer, boiled for 5 minutes, and resolved via SDS-PAGE on a 4%–20% precast gradient gel under a constant voltage of 180 V for 45 minutes; the gel was stained with Coomassie Brilliant Blue R-250. To generate a quality control (QC) sample, equal amounts of protein from all samples were pooled into a “Pool sample.” All samples, including the QC Pool sample, underwent trypsin digestion using the Filter-Aided Proteome Preparation (FASP) method, after which the resulting peptide fragments were desalted via C18 Cartridge columns, lyophilized, and reconstituted in 40 µL of 0.1% formic acid. Peptide concentrations were determined by measuring absorbance at 280 nm (OD280), and an appropriate quantity of iRT standard peptides was added to each sample prior to analysis by data-independent acquisition (DIA) mass spectrometry using an Astral high-resolution mass spectrometer.

DIA mass spectrometry analysis

Data-Independent Acquisition Mass Spectrometry (DIA-MS) analysis involved a two-step workflow: (1) chromatographic separation of samples using the Vanquish Neo system (Thermo Fisher) with nanoliter flow rates via nano-HPLC, followed by (2) DIA-MS analysis on the Astral high-resolution mass spectrometer (Thermo Scientific) in positive ion mode (parent ion scan range: 380–980 m/z). First-order mass spectrometry parameters included 240,000 resolutions at 200 m/z, 500% Normalized AGC Target, and 5 ms Maximum Injection Time (IT). DIA data acquisition utilized 299 scan windows (2 m/z isolation window, 25 eV HCD collision energy, 500% Normalized AGC Target, 3 ms IT for MS2). The raw DIA data were processed using DIA-NN software with trypsin digestion (max 1 missed cleavage site), carbamidomethyl (C) as fixed modification, and oxidation (M) and acetyl (N-terminal protein) as dynamic modifications. Database search results were filtered to retain only proteins with a False Discovery Rate (FDR) below 1% ([14](#), [15](#)).

Data resources

The RNA-seq, proteome datasets and clinical data for CRC patients were obtained from Gene Expression Omnibus (GEO) database, The Cancer Genomic Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>), and Li et al.'s study cohort CCRC ([16](#)), totaling 1,479 samples across GSE39582 (n=585), CCRC (n=146), GSE38832 (n=122), and TCGA-COADREAD (n=626). Differential expression genes (DEGs) were identified using the limma package (Fold change < 0.67 or >1.5, p < 0.05). Overlaps of DEGs in primary tumors and distant metastases were visualized using the “Venn” tool. The CCRC cohort which containing N0M1 CRLM (n=23) and N0M0 (stage I-II, n=49) patients, was used as the training set, validated by GSE39582 and GSE38832, TCGA cohort were utilized for analyzing mutation frequency, TMB, MSI, CNVs and conducting survival analysis, while CRC_EMTAB8107 (n=7) was used for scRNA-seq data analysis.

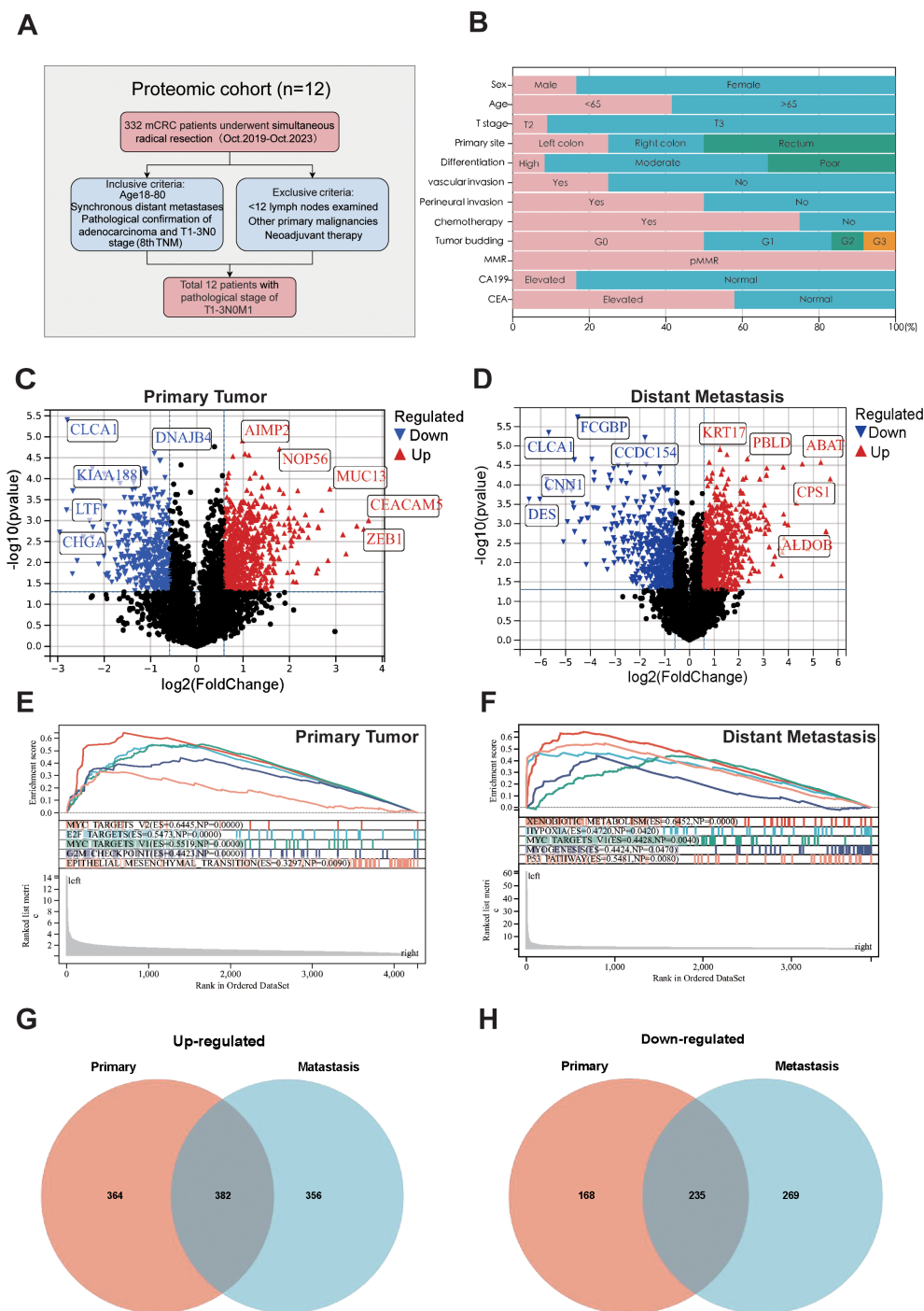


FIGURE 1 Sample selection and proteomics landscape of T1-T3N0M1 CRC. (A) Flow chart of the selection process. (B) Clinicopathological parameters are shown in histogram. Volcano plot of the differential expressed proteins in the primary tumors (C), distant metastases (D) compared to adjacent normal tissues. GSEA analysis for the differential expressed protein in primary tumors (E) and distant metastases (F). Venn plot of up-regulated (G) and down-regulated (H) proteins in primary tumors and distant metastases.

Protein-protein interaction network

The STRING database (<http://string-db.org/>) was employed to explore the interaction relationships among target proteins. Cytoscape software (version 3.10.0) and the GeneMANIA database (<http://genemania.org/>) were then used to construct a

Protein - Protein Interaction (PPI) network, which helped identify the co-expression patterns and interactions of key proteins. By leveraging the Molecular Complex Detection (MCODE, version 2.0.3) plugin (<https://apps.cytoscape.org/apps/mcode>), we extracted potentially densely interconnected gene modules from the PPI network.

Biological function and pathway enrichment analysis

To unravel the biological functions and pathways associated with differentially expressed genes (DEGs) and the core cluster within the PPI network, we utilized the “ClusterProfiler” package and Gene Set Enrichment Analysis (GSEA) software, which can be accessed at <https://www.gsea-msigdb.org/gsea/index.jsp>. With these tools, we performed Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), and GSEA analyses. To evaluate the correlation between gene expression levels and biological pathways or molecular mechanisms, we downloaded the h.all.v7.4.symbols.gmt subset from the Molecular Signatures Database (MSigDB), available at <https://www.gsea-msigdb.org>.

Machine learning identifies LIMGs prognostic biomarkers

Wilcoxon test identified differentially expressed genes between N0M0 (stage I-II) and N0M1 patients, Lasso regression eliminated redundant genes through ten-fold cross-validation using the glmnet package (17). Logistic analysis was used after Z-score transforming of the expression data to determine the odds ratio (OR) of potential hub genes and understand their contribution to the metastasis. Finally, 9 genes were identified as LIMGs. The diagnostic performance of LIMGs was validated using ten machine learning algorithms including Logistic, Support Vector Machine (SVM), Gradient Boosting Machine (GBM), Neural Network, Random Forest (RSF), XGboost, K-Nearest Neighbors (KNN), Adaptive Boosting (Adaboost), Light Gradient Boosting Machine (Light GBM), and Categorical Boosting (CatBoost). We applied them to the CCRC cohort for training, and further validated on external datasets (GSE39582, GSE38832), ROC curves generated by the pROC package were utilized to evaluate the accuracy of the model in diagnosing lymph node-independent distant metastasis of I-II stage CRC. For each patient, LIMG score was calculated for each sample and stratified them into subgroups based on the median score, Kaplan-Meier (KM) survival analysis and nomogram (rms package) assessed prognostic significance of LIMGs.

Mutation analysis and immune microenvironment

Based on the LIMGs Score, risk subgroups are classified in TCGA-COADREAD cohort. Utilizing the “mafTools” R package we analyzed the differences in somatic mutations, TMB between high-risk and low-risk groups, as well as mutation frequency in 9 LIMGs across all samples. The total count of non-synonymous somatic mutations per megabase across the entire genome was computed to assess the TMB. The CNV data and MSI score of CRC patients were downloaded using the TCGA bio links package. Using the CIBERSORT algorithm (18), we evaluated the abundance of 24

immune cell subsets in different risk subgroups. The immune-related functions and expression differences of immune checkpoint genes between subgroups calculated by ssGSEA package predicted immunotherapy response. Correlations between ITGA11 and immune cells were calculated using TIMER (19), QUANTISEQ (20), MCPcounter (21), EPIC (22), and CIBERSORT (23).

Drug sensitivity analysis

Based on Cancer Therapeutics Response Portal (CTRP, <https://portals.broadinstitute.org/ctrp.v2.1/>) and Genomics of Drug Sensitivity in Cancer (GDSC, <https://www.cancerrxgene.org/>), the “OncoPredict” R package was used to conduct a half-maximal inhibitory concentration (IC50) analysis of drugs for high-risk and low-risk groups of CRC patients.

Single cell RNA sequencing analysis

We acquired a CRC dataset (CRC_EMTAB8107) from the Tumor Immune Single Cell Hub 2.0 (TISCH 2.0) database (<http://tisch.compgenomics.org/>) (24), comprising 23,176 cells from 7 tumor samples. Subsequent analyses included scRNA-seq for the ITGA11 and Cell-Cell Interaction (CCI) analysis and visualize the expression and distribution of ITGA11, and the interactions between target gene-enriched cell subpopulations and others.

Antibodies, plasmids, cell lines and culture

In this study, two antibodies were utilized: ITGA11 (#DF8992, Affinity Biosciences, USA) and GAPDH (#2118, Cell Signaling Technology, USA). The SW480 cell line (#CBP60019), authenticated by short tandem repeat (STR) profiling, was procured from the Chinese Academy of Sciences (CAS). Cells were cultured in RPMI 1640 medium (#C11875500BT, Gibco, USA) supplemented with 10% fetal bovine serum (FBS) and 10,000 U/ml penicillin-streptomycin (#15140122, Gibco, USA) in a humidified incubator with 5% CO₂. A knockdown plasmid targeting ITGA11 was synthesized by Miaoling Bioscience (Wuhan, China).

Immunohistochemical assay

Tissue specimens were fixed in 4% paraformaldehyde, embedded in paraffin, and sectioned into 4 μm-thick slices for slide preparation. After gradient deparaffinization and rehydration, antigen retrieval was performed using a microwave method with citrate buffer (100°C, four cycles of 7 minutes each). The slides were washed extensively with PBS and then blocked for 30 minutes to minimize nonspecific binding. The primary antibody was incubated

overnight at 4°C, followed by incubation with the secondary antibody at room temperature. Color development was achieved using DAB chromogen, and the sections were counterstained with hematoxylin.

Wound healing assay

Cells were plated in 6-well plates and grown to confluence. A sterile pipette tip scratched the monolayer, which was then washed with PBS to remove any dislodged cells. Culture medium with 1% (fetal bovine serum) FBS was added. Images of cell migration were taken at 0, 24, and 48 hours post-wounding. The wound closure area was calculated as: $\text{Migration Area (\%)} = (X_0 - X_n)/X_0 \times 100$, where X_0 is the initial wound area and X_n is the area at a specific time.

Transwell assay

The invasive and metastatic potential of SW480 cells was assessed using a Matrigel-coated Transwell assay. Briefly, 3×10^4 cells were seeded in the upper chamber of a Transwell with serum-free medium, while the lower chamber contained 10% FBS-supplemented medium. After 24 hours of incubation at 37°C, cells in the upper chamber were fixed with methanol and stained with Giemsa for quantitative microscopic analysis of invasion and migration.

Western blot assay

Cell proteins were extracted using a lysis buffer (10 mM TRIS-HCl, pH 7.4, 1% SDS, 1 mM Na_3VO_4) and lysed via ultrasonic treatment. Protein concentration was quantified using a microspectrophotometer. Samples, mixed with loading buffer and a molecular weight marker, were loaded onto an 8% SDS-PAGE gel and subjected to electrophoresis at 80 V for 30 minutes, followed by 120 V for 90 minutes. Proteins were transferred to a PVDF membrane (25 V, 120 minutes) and blocked in buffer at 4°C for 3 hours. The membrane was then incubated overnight with primary antibodies at 4°C and for 3 hours with secondary antibodies. Protein bands were visualized using an ECF developer (RPN5785, GE Healthcare) and captured using a chemiluminescent imaging system (GE Healthcare).

Statistical analysis

All statistical analyses were conducted using R software (version 4.4.2). The Wilcoxon test compared variables between groups, the Chi-square test assessed categorical variable differences, Pearson correlation analyzed variable correlations, and KM survival analysis with log-rank test evaluated differences. Statistical significance was

set at $p < 0.05$: *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.0001$; NS: non-significant.

Ethics approval

This study was conducted in accordance with the ethical standards outlined in the Helsinki Declaration and certified by the Ethics Committee of The First Affiliated Hospital of Wenzhou Medical University (KY2022-183). Given the retrospective nature of the study, informed consent was waived.

Results

Proteomic characteristics of T1-T3N0M1 patients

To identify the protein signatures and pathways associated with T1-T3N0M1 CRC, we used adjacent normal tissue as a control and analyzed the differentially expressed proteins in primary tumor and distant metastasis and conducted a comprehensive comparison of biological pathways and functions. Differential analysis revealed 746 upregulated and 403 downregulated proteins in primary tumors (Figure 1C), and 751 upregulated and 321 downregulated in distant metastases (Figure 1D). GSEA was performed to analyze the features of the proteins detected in primary tumors and distant metastases in terms of biological pathways and molecular mechanisms. Results showed enrichment in MYC targets V2, E2F targets, MYC targets V1, G2M Checkpoints, EMT pathway in primary tumors (Figure 1E). Xenobiotic metabolism, Hypoxia, MYC targets V1, Myogenesis, P53 pathway enrichment in distant metastases (Figure 1F). Analysis of GO and KEGG pathway of these proteins is provided in Supplementary Figure S1. Venn diagrams visualized the intersections of differentially expressed proteins among primary tumors and metastases (Figures 1G, H).

Construction of PPI networks and module identification for biomarker discovery

To construct a PPI network for biomarker identification, we uploaded 617 differentially expressed proteins shared between primary tumors and distant metastases into the STRING database. The resulting network was visualized using Cytoscape software and the MCODE plugin, enabling the identification of the five most functionally significant modules (Supplementary Figure S2). Among these, Cluster 2 emerged as a key module, comprising 40 nodes and 214 edges (Figure 2A). GO analysis reveals that Cluster2 is mainly enriched in ECM organization, cell adhesion, collagen-containing ECM, and ECM structural constituent (Figure 2B). The KEGG analysis indicates the enrichment of Focal adhesion, ECM-receptor interaction, and PI3K-Akt signaling pathway (Figures 2C, D). Hallmark pathway enrichment

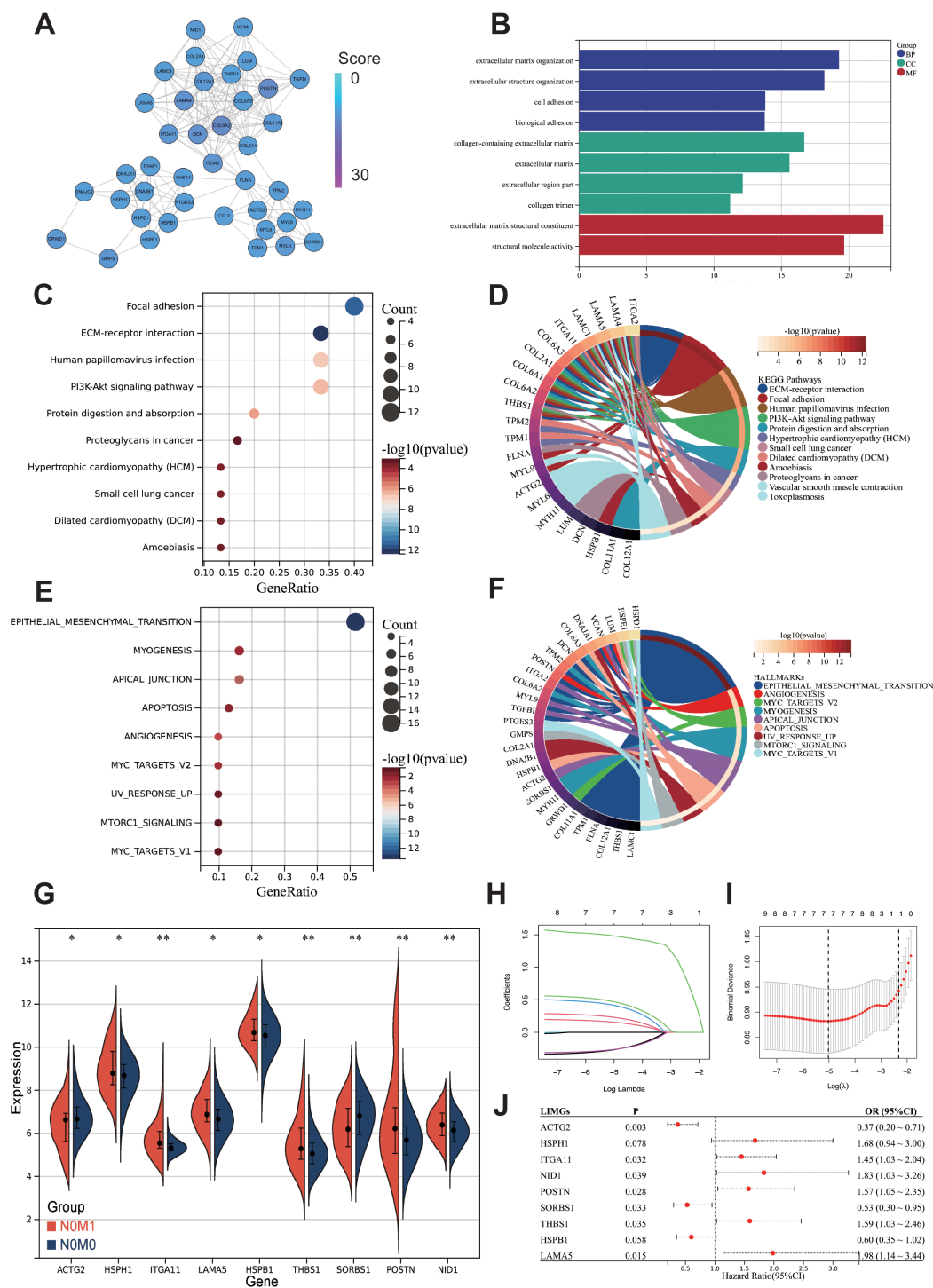


FIGURE 2

Preliminary screening of core biomarkers associated with lymph-node independent metastasis. (A) MCODE in Cytoscape identified a module consisting of 40 nodes from the PPI network. The GO (B), KEGG (C) and Hallmark gene sets (E) enrichment analysis of the 40 genes from cluster2. Chord diagrams of KEGG (D), and Hallmark gene sets (F) enrichments show associations of 40 genes across different biological aspects. Differentially expressed genes between NOM0 and NOM1 (G). The Lasso regression path plot (H) and cross-validation plot (I) illustrate the gene selection process. The univariate logistic regression results of LIMGs (J). Statistical significance: $p < 0.05$; $**$: $p < 0.01$; $***$: $p < 0.0001$; NS: non-significant.

analysis shows enrichment in EMT pathway, Myogenesis, Apical junction, Apoptosis, and Angiogenesis (Figures 2E, F).

Machine learning identifying LIMGs signature and constructing diagnostic model

To further identify core biomarkers and establish an accurate diagnostic model, we identified differentially expressed genes in Cluster2 between stage I-II and N0M1 CRC (Wilcoxon test, $p < 0.05$) (Figure 2G), after eliminating redundant genes using Lasso regression (Figures 2H, I), 9 genes were selected as LIMGs (ACTG2, HSPH1, ITGA11, LAMA5, HSPB1, THBS1, SORBS1, POSTN, NID1). Univariate logistic regression highlighted the importance of LIMGs via OR (Figure 2J). Ten machine learning algorithms including Logistic, SVM, GBM, Neural Network, Random Forest, XGboost, KNN, Adaboost, Light GBM, and CatBoost were then applied to assess the diagnostic efficacy of LIMGs in the CCRC training set, ROC curves (Figure 3A), DCA (Figure 3D), confirmed robust diagnostic performance, with external validation in two cohorts (GSE39582, Figures 3B, E), (GSE38832, Figures 3C, F). The Neural Network model demonstrated consistent performance across cohorts, with diagnostic efficacy displayed by the confusion matrix (Figures 3G–I). Feature importance analysis of the top eight models in the training cohort identified ITGA11 as the key factor influencing lymph node-independent metastasis (Figure 3J).

LIMGs correlate with poor prognosis and clinicopathological features

The GSVA method scored GSE39582 (Figure 4A), GSE38832 (Figure 4B) and TCGA-COADREAD (Figure 4C) samples based on LIMGs expression, classifying risk subgroups by the median score. KM curves revealed significant survival difference between risk subgroups. KM curves for subgroups based on ITGA11 median expression revealed its significant impact on overall survival (OS), Disease-free survival (DFS), and Progression-free interval (PFI) (Figure 4D). To further explore the association between LIMGs and metastasis of patients, we employed the R package rms to integrate data on metastasis-free survival (MFS), survival status, and eight relevant features of CCRC cohort. A nomogram was constructed using the Cox method, and the prognostic significance of these features was assessed in 143 samples of CCRC cohort (Figure 4E). Kaplan-Meier curves (Figure 4F) and ROC curves for 1- and 3-years MFS (Figure 4G) underscored the predictive accuracy of LIMGs, highlighting its value in predicting metastasis. Further evaluation for the association between the LIMGs and other pathological characteristics reveals that higher LIMG score is significant associated with advance AJCC stage (Figure 5A), N stages, (Figure 5B), MSI status (Figures 5E, F), KRAS-WT (Figure 5G), and the left-sided colorectal cancer (Figure 5H) (all $p < 0.05$). Furthermore, CRC is molecularly

classified into six subtypes by Marisa et al. (25) including C1 (downregulation of immune pathway), C2 (MSI subtype), C3 (KRAS mutant), C4 (chromosomal instability and stem-like), C5 (Wnt pathway upregulation) and C6 (derived from serrated tumors). We found that higher LIMG score correlated with C4–C6 molecular subtypes (Figure 5D). Additionally, no significant differences in LIMG score were observed across different T stages (Figure 5C), ages, sexes, vascular invasion statuses, histological types, and BRAF, TP53 mutation statuses (Supplementary Figure S3).

Analysis of LIMGs interaction and correlation with EMT

The hallmark enrichment revealed that LIMGs are mainly enriched in the EMT pathway (Figure 6A). EMT drives tumor invasion and metastasis through induction of stemness, modulation of the TME, angiogenesis promotion, and metabolic reprogramming. We investigated the correlation between LIMGs and the EMT pathway by 200 EMT-related genes from the MSigDB database v7.1. The correlation between LIMGs and EMT gene signatures was analyzed using Pearson correlation analysis on the GEPIA2 (<http://gepia2.cancer-pku.cn/#index>). The results indicated that ITGA11 has the strongest correlation with EMT (Figure 6D). Then we used GeneMANIA to analyze the interactions among LIMGs (Figure 6B) and the PPI network centered on ITGA11 (Figure 6C). The results of GO and KEGG enrichment analyses of LIMGs are shown in Supplementary Figure S4.

Mutation landscape and immune activity in different risk groups

To elucidate the distinct mutational patterns among different risk groups, we utilized the “mafTools” R package to analyze the distribution of top 20 somatic mutations between risk groups and mutation status in 9 LIMGs based on TCGA-COADREAD data. Our findings revealed that APC, TP53, TTN, and KRAS exhibited high mutation frequencies across different subgroups, with APC being identified as the most frequently mutated gene across subgroups (Figures 7A, C). LAMA5 showed the highest mutation frequencies among LIMGs (Figure 7B). Meanwhile, copy number variation (CNV) plays a crucial role in cancer occurrence and development. We found that the highest CNV in LIMGs was also found in LAMA5 (Figure 7F). Given the significance of TMB, MSI status, immune cell infiltration, immune functions, and immune checkpoint gene expression in immunotherapy response, we examined their relationship with LIMGs. Immune infiltration analysis by CIBERSORT revealed that the high-risk group had lower proportions of memory B cells, plasma cells, CD4+ T cells, NK cells, dendritic cells and eosinophils but higher proportions of M0 and M2 macrophages (Figure 7H). Moreover, the high-risk group exhibited greater immunological function, including higher levels of Type I and II IFN Response and APC co-stimulation.

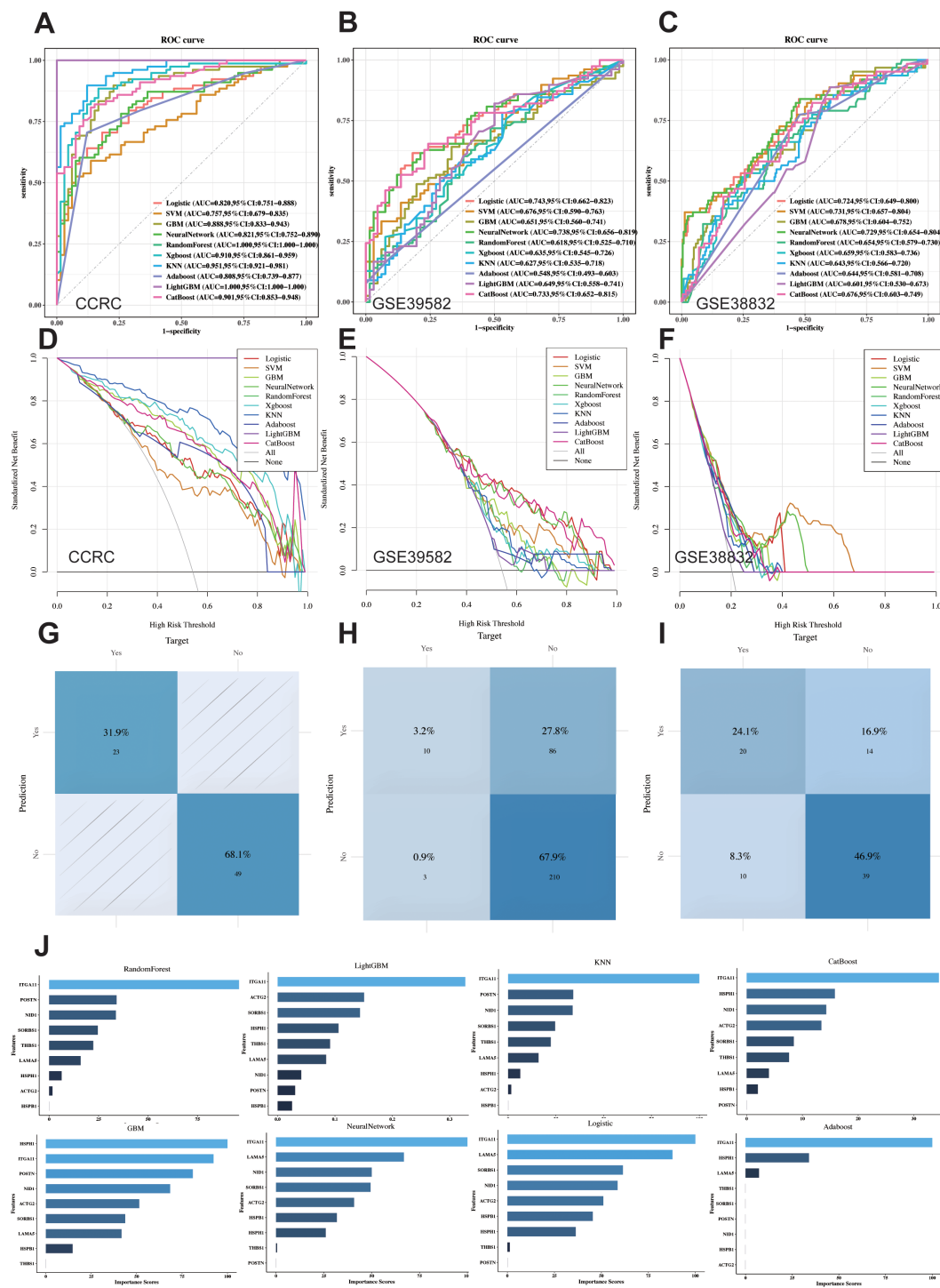


FIGURE 3 Ten Machine learning methods assess the diagnostic performance of LIMG signature. ROC curves of ten machine learning methods (Logistic, SVM, GBM, Neural Network, RF, XGboost, KNN, Adaboost, Light GBM, and CatBoost) applied in CCRC training cohort (A) and external validation cohort GSE39582 (B) and GSE38832 (C). Cost-benefit decision curves in training (D) and validation (E, F) cohorts. Classification confusion matrix of the Neural Network model in training (G) and validation cohort (H, I). (J) The feature importance bar chart illustrates variable contributions to the top 8 models in training cohort.

(Figure 7I). The high-risk group also exhibited a significantly lower MSI proportion (Figure 7D) and lower TMB (Figure 7E). Conversely, the TMB and MSI status was higher in low-risk group, suggesting better immunotherapy response. Analysis of

immune checkpoint expression showed higher expression of PDCD1 in low-risk group and higher expression of TIGIT, ICOS and CTLA4 in high-risk group (Figure 7G). Furthermore, the positive correlation between ITGA11 expression and various

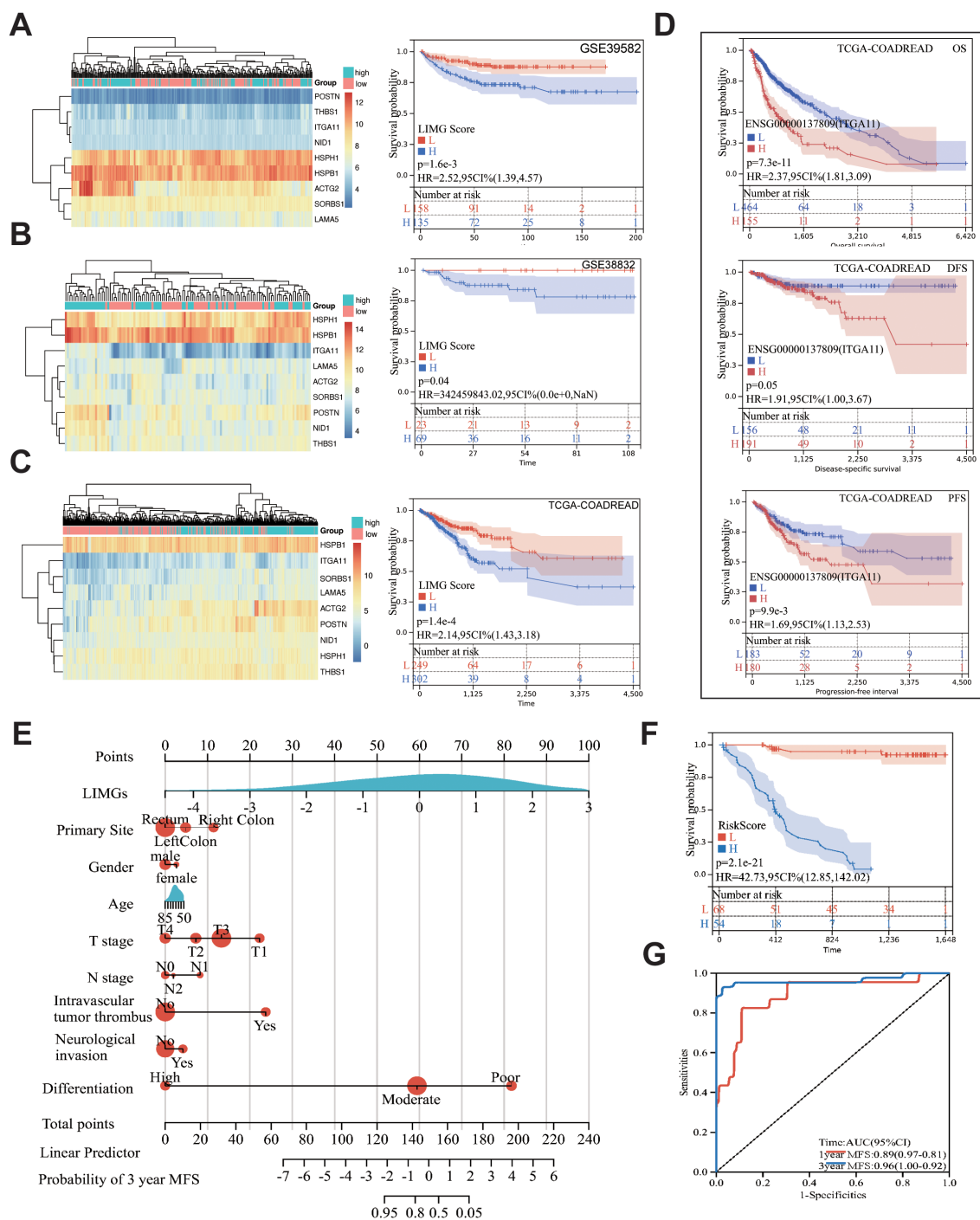


FIGURE 4
 The correlation between the LIMGs and survival. Heatmaps of LIMGs expression and the KM curves stratified by high-risk and low-risk groups based on median LIMG Score in the GSE39582 (A), GSE38832 (B), and TCGA cohorts (C). (D) KM curves for overall survival (OS), disease-free survival (DFS), and progression-free interval (PFI) in high-risk and low-risk groups stratified by the optimal cutoff value of ITGA11 expression in TCGA-COADREAD. (E) Nomogram for predicting Metastasis-free survival (MFS) was constructed using multivariate Cox regression. (F) KM curves compared high-risk vs. low-risk stratified by median risk score. (G) ROC curves evaluated 1-year/3-year MFS prediction accuracy of the Nomogram.

immune cells was calculated by five algorithms (TIMER, QUANTISEQ, MCPcounter, EPIC, CIBERSORT) (Figure 7), specifically with high levels of CAFs and TAMs. This correlation may indicate poor prognosis in CRC patients with higher level of

ITGA11+CAFs and ITGA11+TAMs. In summary, the high-risk group exhibited lower TMB, MSI status and immunosuppressive TME, suggesting less favorable immunotherapy outcomes compared to the low-risk group.

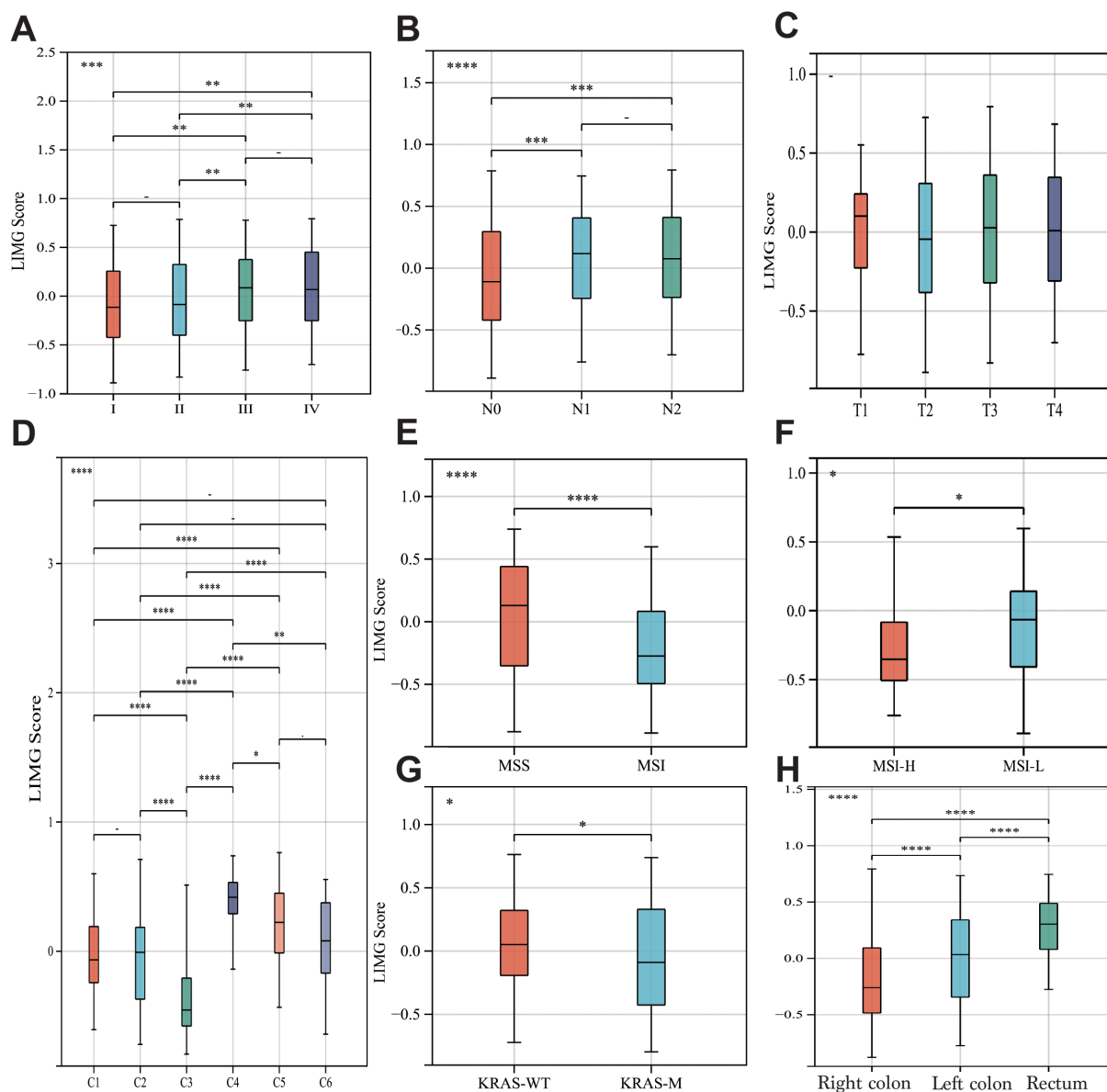


FIGURE 5
The correlation between LIMGs and clinicopathological features. The correlation between the LIMG Score and AJCC stage (A), N stage (B), T stage (C), Molecular subtypes (D), Microsatellite stability (E, F), KRAS mutation status (G), and tumor location (H). Statistical significance: *: p<0.05; **: p<0.01; ***: p<0.001; ****: p<0.0001; NS, non-significant.

LIMGs associate with lower chemotherapy sensitivity

To predict drug sensitivity and identify potential therapeutic drugs for high-risk CRC patients, we calculated IC50 values for three commonly used CRC chemotherapy drugs (5-Fluorouracil, Oxaliplatin, Irinotecan) in different risk subgroups and assess the correlation between LIMG score and drug sensitivity. The results showed that high-risk patients had poorer sensitivity to 5-Fluorouracil (Figure 8A), Oxaliplatin (Figure 8B), and Irinotecan (Figure 8C), with

IC50 values positively correlated with risk scores. Conversely, high risk patients exhibited higher sensitivity to Dasatinib (Figure 8D), Doramapimod (Figure 8E), and PRKDC inhibitor NU7441 (Figure 8F), with IC50 values negatively correlated with LIMG score.

ScRNA-seq analysis of ITGA11

To explore the expression and distribution of ITGA11 in the TME at the single-cell level, we conducted scRNA-seq analysis using the TISCH

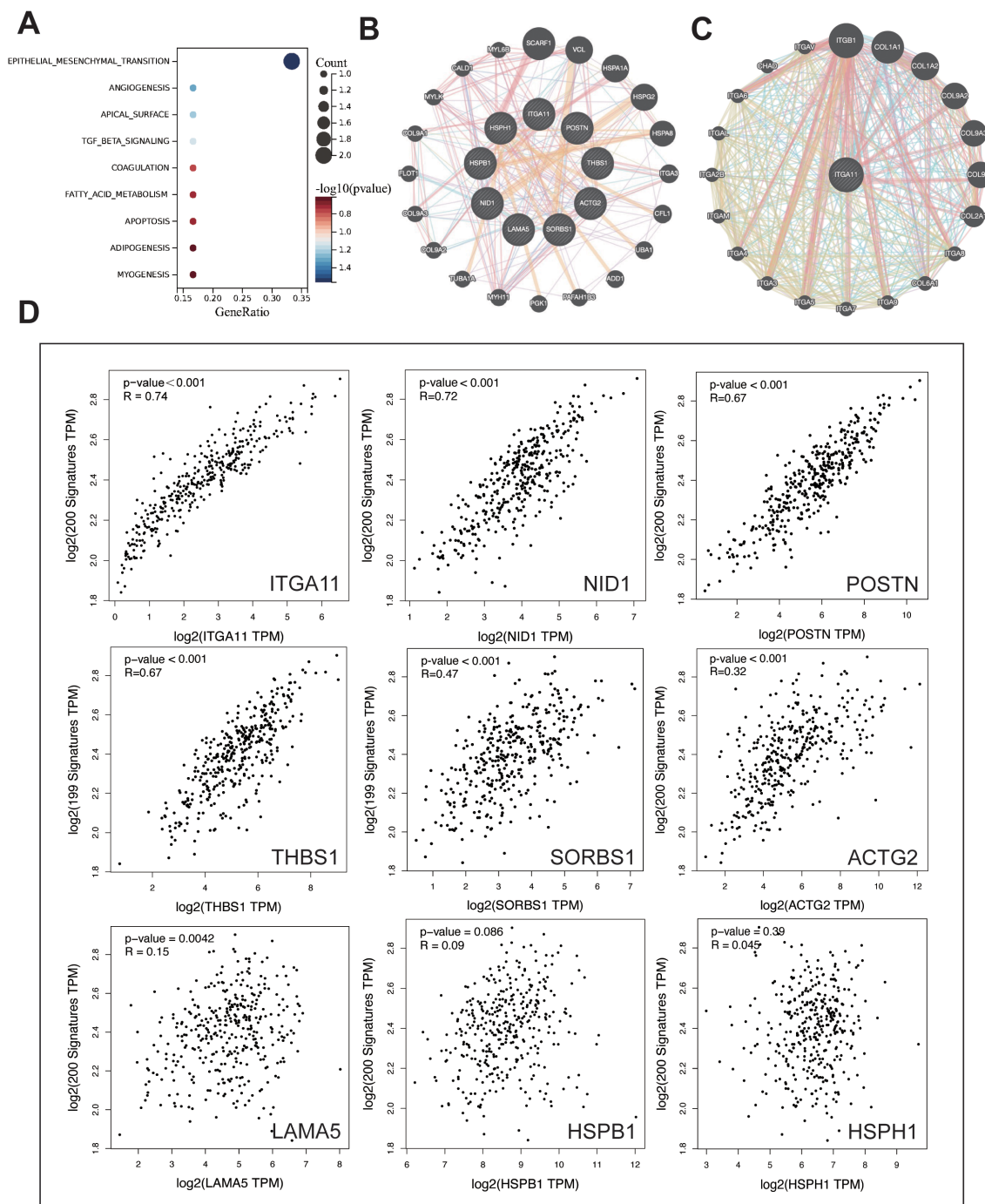


FIGURE 6
 The enrichment and interaction analysis of LIMGs and correlation analysis between LIMGs expression and the EMT pathway. **(A)** Hallmark enrichment analysis plot of LIMGs. **(B)** Interactions among LIMGs. **(C)** The PPI network centered on ITGA11. **(D)** Correlation between LIMGs expression and the EMT pathway gene set.

2.0 database. Analysis of scRNA-seq data from the CRC_EMTAB8107 dataset revealed the identification of 20 cell clusters and 12 cell types within CRC tissues (Figure 9A). We observed a significant enrichment of ITGA11 in CAFs (Figure 9B), especially within clusters C3 and C10 (Figure 9F). The analysis of Cell-Cell Interactions (CCI) revealed that both C3 and C10 CAFs mainly interacted with C9 malignant cells and C19 endothelial cells (Figures 9D, E).

ITGA11 promotes migration and invasion of colorectal cancer cells

In this study, we unveiled the crucial role of LIMGs in distant metastasis of CRC, primarily associated with cell adhesion and EMT. Notably, LIMG ITGA11 is the gene most strongly correlated with EMT. Although ITGA11 overexpression has been reported in

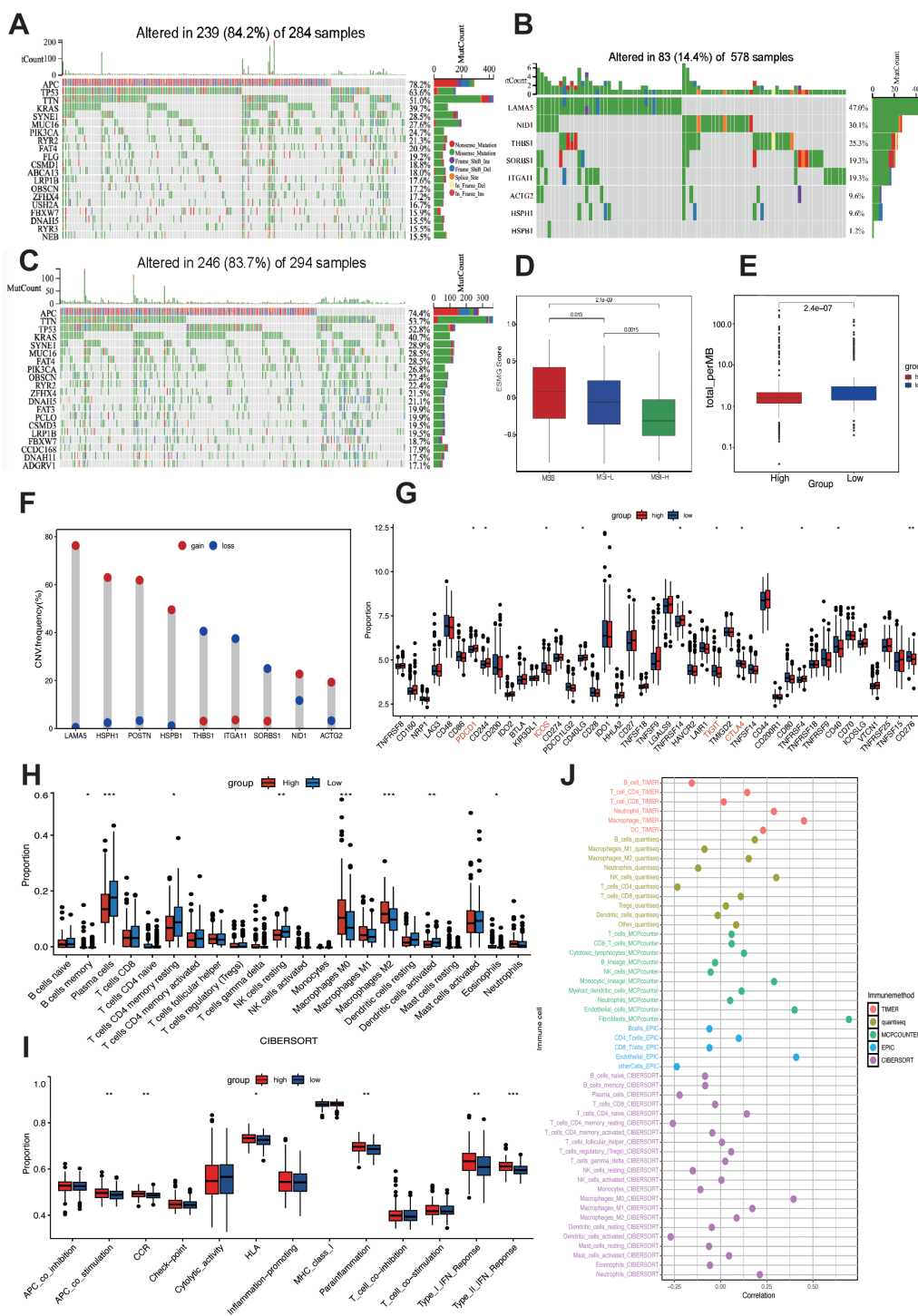


FIGURE 7 Mutation landscape and immune activity analysis. Top 20 mutated genes in high-risk (A) and low-risk (C) subgroups. (B) Mutation frequency of 9 LIMGs. (D) The LIMG Score of CRC patients with microsatellite instability-high (MSI-H) and microsatellite stability (MSS). (E) Comparison of TMB in high- and low- risk subgroups. (F) The CNV frequency of each LIMG signature genes. (G) Differentially expressed immunecheckpoint genes across risk subgroups. (H) Differences in immune cell infiltration across risk subgroups. (I) Immune-related functions in the high- and low- risk subgroups. (J) Correlation between ITGA11 expression and immune cells. Statistical significance: *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.0001$; NS: non-significant.

several tumors, its impact on CRC cell migration and invasion remains unexplored. The radar chart illustrates the expression levels of LIMGs (Log2(FC)) based on our proteomics data (Supplementary Figure S5). The results reveal the expression

across primary tumors, and distant metastasis for 9 LIMGs. Among them, ITGA11 showed higher expression in distant metastasis than in primary tumors. The IHC score further confirmed higher ITGA11 expression in both primary tumors

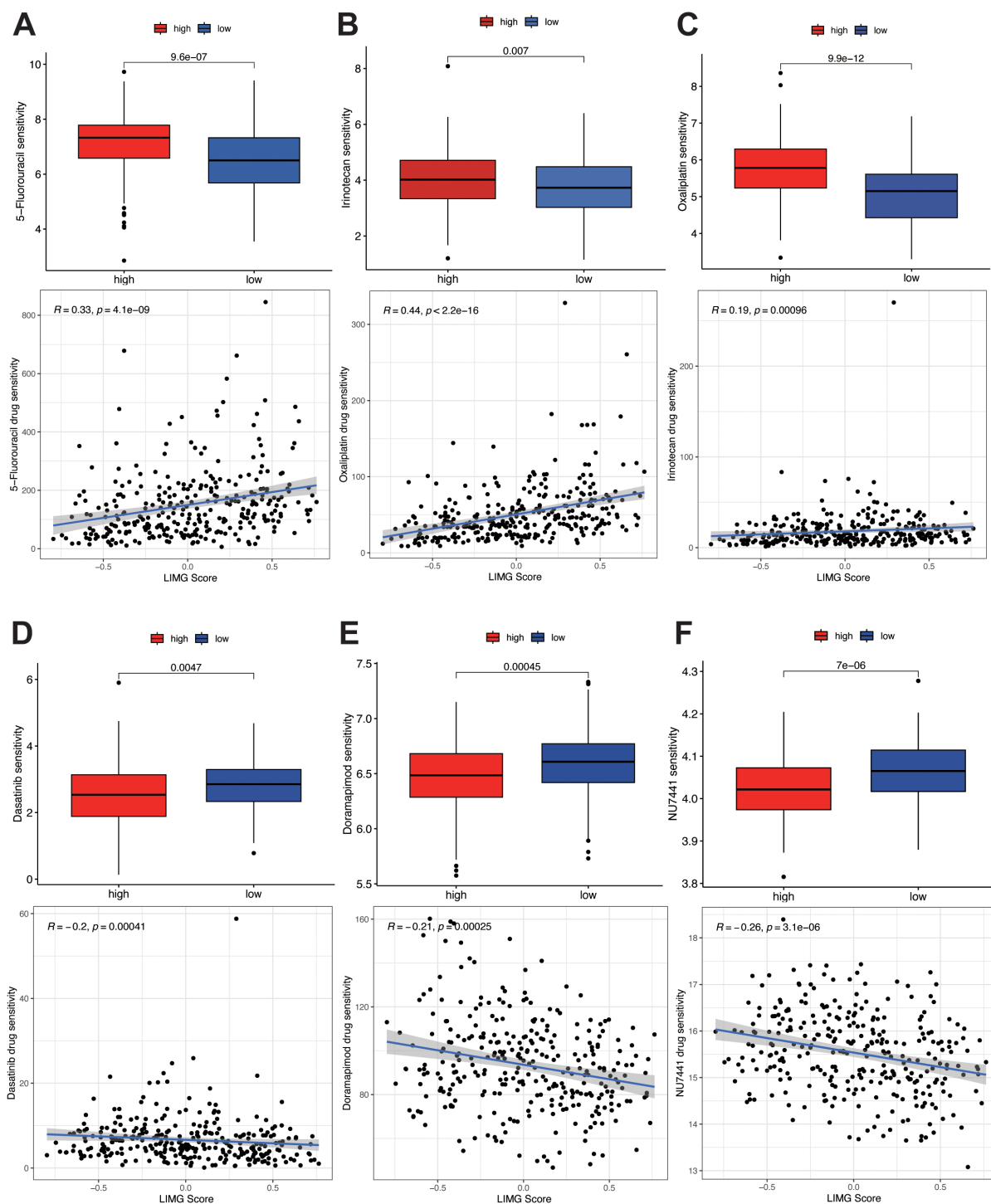
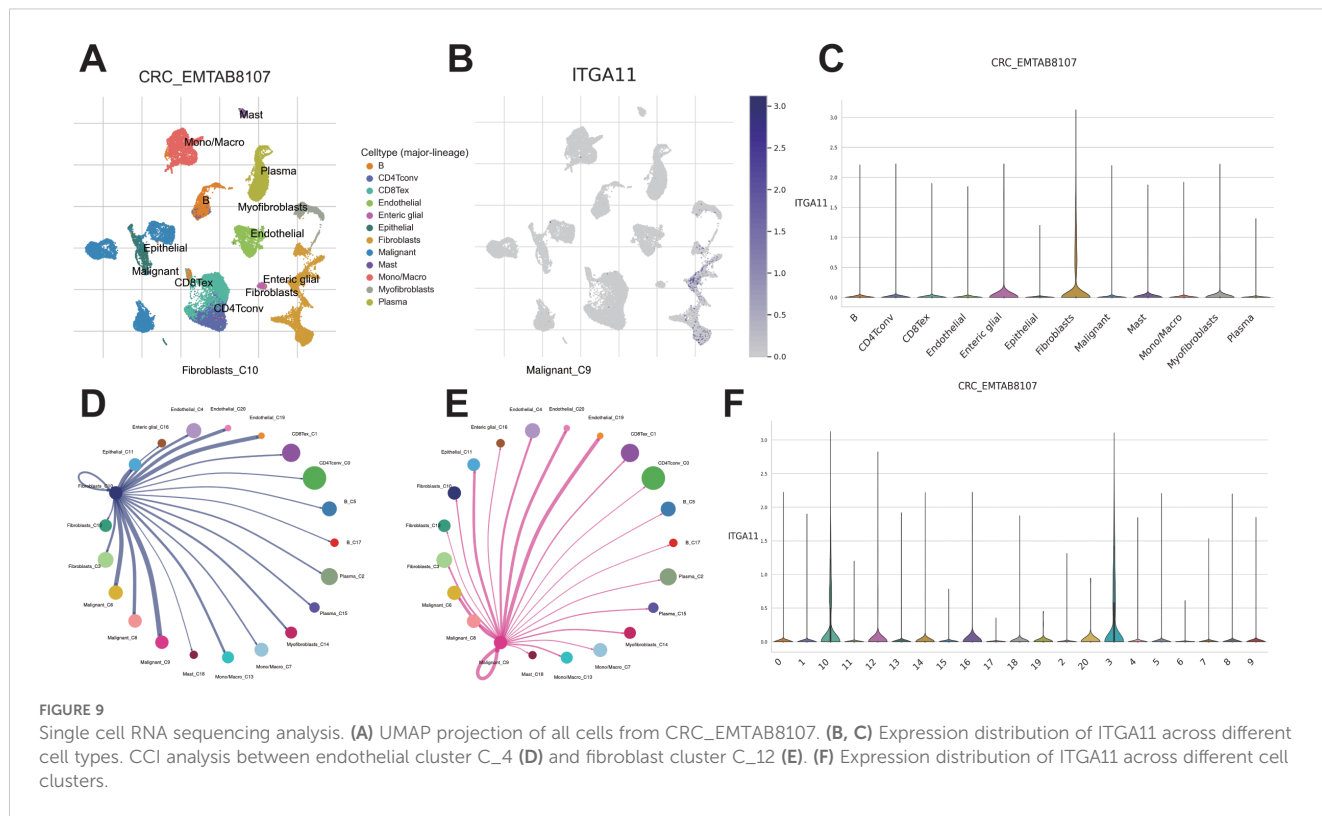


FIGURE 8 Drug sensitivity analysis. Sensitivity analysis of 5-fluorouracil (A), oxaliplatin (B), and irinotecan (C) in different risk groups. Sensitivity analysis of Dasatinib (D), Doramapimod (E), and PRKDC inhibitor NU7441 (F) in different risk groups.

and distant metastasis compared to normal tissues, with significantly higher levels in distant metastasis compared to primary tumors (Figure 10B, $p < 0.05$). The ROC curve shows that ITGA11 significantly differentiates primary tumors from distant metastasis (Figures 10C–E). Furthermore, we achieved stable knockdown of ITGA11 in the human colon cancer cell line

SW480 and subsequently evaluated the efficiency of this knockdown via Western blot analysis (Figure 10F). To assess the functional implications, we performed wound healing and transwell assays. The results of these assays demonstrated that ITGA11 knockdown significantly compromised the migratory ability of SW480 cells (Figure 10A; t - test, $p < 0.05$) and led to a



substantial decrease in the number of invading cells (Figure 10B; t-test, $p < 0.05$). Collectively, these findings underscore the pivotal role of ITGA11 in the migration and invasion processes of CRC.

Discussion

Approximately 20% of newly diagnosed CRC patients have synchronous distant metastasis and generally face a poor prognosis (26). While LNM signifies advanced disease, tumor cells may spread hematogenously before lymphatic metastasis. Previous reports indicate that about 18% of mCRC patients lack local lymph node involvement (27), and a novel CRC mouse model shows distant metastases can develop without prior lymph node involvement (9). Recurrence patterns in patients with CRLM undergoing liver transplantation without other metastasis suggest tumor cells may persist in circulation post-resection of primary and metastatic tumors. A more plausible explanation is that undetectable pre-operative metastasis account for most post-operative metastasis. Early hepatic metastases are often missed or undiagnosed by imaging, and by the time typical metastatic signs appear, radical surgery is usually no longer an option (28). Thus, identifying potential synchronous metastases or metastasis risks at early stage primary CRC is crucial (29).

The early occurrence of metastasis may stem from pre-existing, undetectable tumor dissemination prior to diagnosis or treatment. The primary tumor not only generates disseminated tumor cells but also establishes the pre-metastatic niche and modulates the immune response (30). Identifying the genetic traits of its stromal and extracellular matrix (ECM) components is vital for metastasis

prediction (31). In cancer, EMT enables cancer cells to lose cell polarity and acquire a mesenchymal phenotype with enhanced stemness and migratory ability through complex interactions between fibers and proteins (13). Continuous remodeling of the ECM and actin cytoskeleton is closely associated with EMT, with integrins acting as physical linkers between the ECM and actin cytoskeleton, mediating mechanotransduction through interactions with major ECM components like collagen and fibronectin (31, 32). ITGA11, identified among LIMGs, shows the strongest correlation with EMT and high importance in various models predicting synchronous metastasis in CRC. ITGA11 promotes CAF invasion and CAF-induced tumor cell invasion, and associates with high-grade tumors and poor prognosis (33, 34). Mechanistically, ITGA11's pro-invasive activity may stem from its ligand-dependent interaction with PDGFR β , promoting downregulated JNK activation and ECM changes, including increased deposition of a strongly co-expressed pro-invasive stromal protein (tenascin-C, TNC) (35). PDGFR α + ITGA11+ CAFs are associated with lymphovascular invasion (LVI) and early metastasis in early-stage bladder cancer, promoting lymphangiogenesis by recognizing the ITGA11 receptor SELE on lymphatic endothelial cells. Additionally, CHI3L1 from the CAF aligns the surrounding stroma to facilitate cancer cell intravasation and promote early tumor metastasis (36). Laminin LAMA5, a glycoprotein in the ECM, has been identified as a specific molecular target in mCRC (37). It is a key component of the vascular basement membrane, forming a scaffold for endothelial cell adhesion in conjunction with collagen IV, and is linked to the angiogenesis and tumor growth in CRLM (31, 38). Notably, LAMA5 exhibits the highest mutation frequency and CNVs in LIMGs, it is reported that genetic

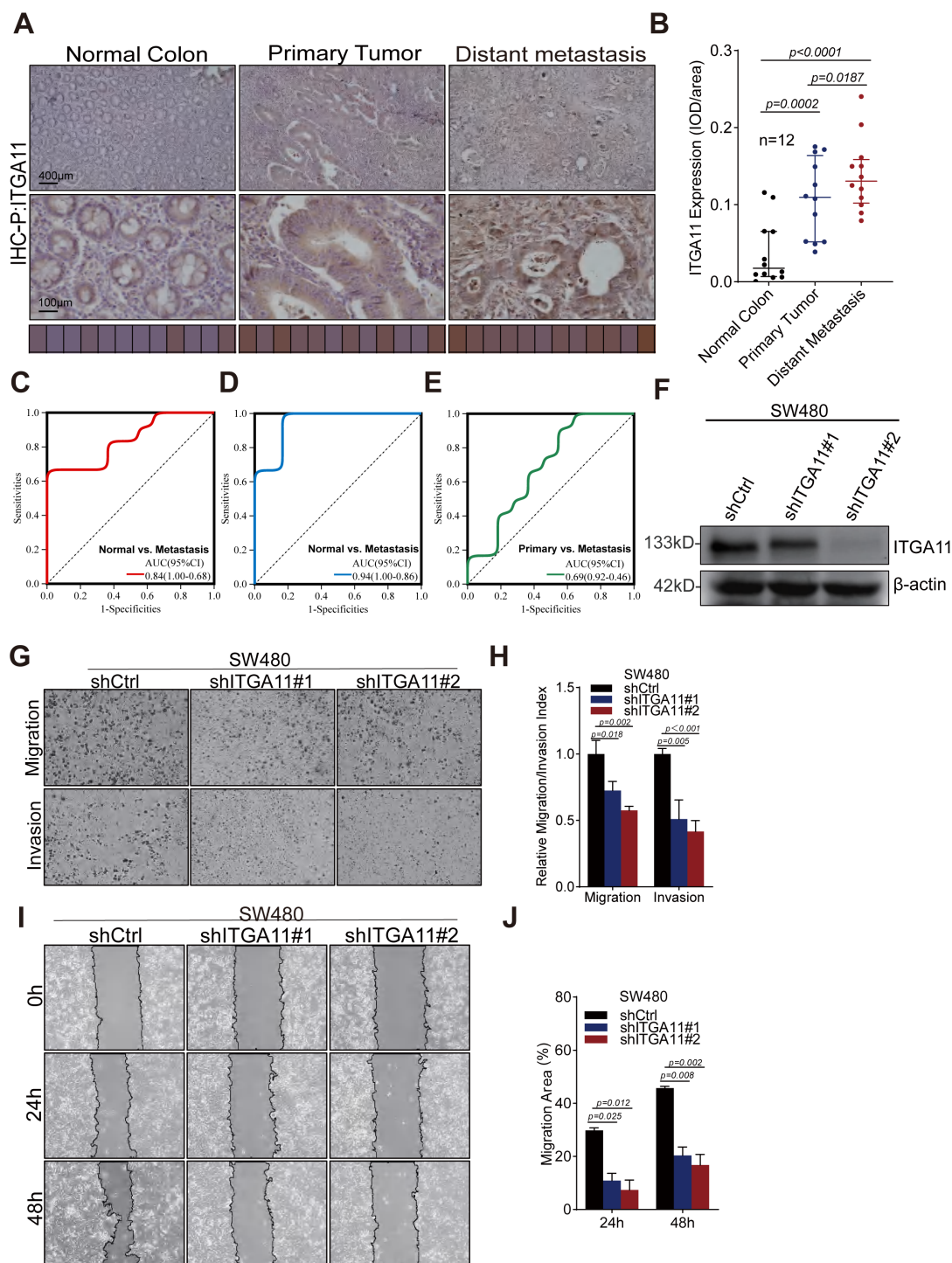


FIGURE 10
 ITGA11 expression variation across tissues and its role in the migration and invasion of CRC cell. **(A)** Immunohistochemical analysis of ITGA11 expression in normal tissues, primary tumors, and distant metastases. **(B)** The Wilcoxon test indicated a significant difference ($P < 0.05$) in ITGA11 expression across normal tissues, primary tumors, and distant metastases. **(C–E)** The ROC curves indicated that ITGA11 expression effectively differentiated normal tissues, primary tumors, and distant metastases. **(F)** The ITGA11 expression in human colon cancer cells SW480 was measured by western blotting. **(G, H)** Results of the transwell assay. **(I, J)** Results of the wound healing assay.

variant rs4925386 in chromosomal region 20q13.3 (LAMA5) significantly associated with CRC susceptibility (OR=0.93) (39). Periostin (POSTN), secreted by CAFs, accelerates angiogenesis, tumor invasion, and EMT via integrin interaction (40). Aberrant

POSTN expression in CRC correlates strongly with peritoneal and distant organ metastasis. Meanwhile, POSTN+ CAFs significantly promote CRC cell migration and proliferation through hypoxia induced POSTN expression and secretion (41). The cbl-associated

protein (CAP), encoded by the sorbin and SH3 domain-containing 1 (SORBS1) gene, plays a role in actin cytoskeleton regulation, receptor tyrosine kinase signaling, and cell adhesion. Overexpression of SORBS1 inhibits the PI3K/AKT pathway, blocks EMT, and promotes M1 macrophage polarization (42). Conversely, SORBS1 silencing accelerates EMT, boosts Filopodium-like Protrusion (FLP) formation via JNK/c-Jun activation in cancer cells, and elevates chemosensitivity by enhancing p53 protein accumulation (43). Nidogen1 (NID1), directly induced by SNAIL/SNAI-1 transcription factor, promotes EMT. It connects laminin, collagen, and proteoglycans to cell receptors, regulating cell polarization, migration, and invasion (44). Actin gamma 2 (ACTG2) is aberrantly expressed in cancers (45), with low levels in CRC, its overexpression inhibits CRC cell proliferation, migration, and invasion (46). Thrombospondin-1 (THBS1) inhibits angiogenesis and immune activity (47), but has complex, contradictory roles in carcinogenesis. THBS1 expression correlates with CRC mesenchymal phenotype, immunosuppression, and poor prognosis, promoting metastasis by exhausting cytotoxic T cells and impairing angiogenesis, especially at metastatic sites (48).

Pathological analysis of early-stage CRC aids in risk identification and treatment guidance. Factors like T4 stage, poor differentiation, intestinal perforation, lymphovascular/perineural invasion, inadequate lymph node examined, and positive surgical margins heighten disease progression risk (49). Our study revealed no significant LIMG score differences in T stage, vascular invasion, and histological type. However, a higher LIMG score correlated with tumor location, MSI status, lower KRAS mutation frequency, and lower TMB. Study shows that synchronous CRLM exhibit poorer prognosis and biological traits than metachronous ones (2, 50), with synchronous CRLM showing lower TMB (51). Moreover, patients with LNM- mCRC typically have fewer high-risk pathological features than LNM+ mCRC (52), indicating clinicopathological factors may inadequately assess the lymph node-independent metastasis in CRC, potentially leading to misdiagnosis. LIMG Score variations across primary tumor sites may stem from tumor site-genomic alteration correlations in mCRC. Left-sided CRC is more prone to synchronous liver metastases (7.1% vs. 11.6%), which may be anatomically influenced by venous shunting (53). Molecularly, right-sided primary tumor-derived MSS-type mCRC has a higher median TMB, with oncogenic alterations like KRAS, BRAF, and PIK3CA enriched, while APC and TP53 are more enriched in left-sided tumors (54).

Immunotherapy benefits CRC patients but is limited by the complex immunosuppressive TME and tumor heterogeneity (55). Our study found that high-risk patients have lower infiltration of anti-tumor immune cells (memory B cells, plasma cells, CD4+ T cells, NK cells, DCs), while exhibiting higher levels of M2-type TAMs that promote tumor growth and immunosuppression. Plasma cells, as terminal effector B cells, eliminate tumor cells via antibody-dependent cell-mediated cytotoxicity (ADCC) (56), forming an immunological chain with DCs and participating in tertiary lymphoid structures (TLS) formation (57). Reduced plasma cell and DC infiltration may indicate weakened antibody-mediated anti-tumor effects, TLS deficiency, and inadequate immune surveillance, potentially with increased Bregs or M2-type TAMs,

leading to insufficient CD8⁺ T cell activation, further exacerbating immune escape, and diminished immunotherapy response.

Furthermore, we have identified ITGA11 as a critical factor in lymph node-independent metastasis in CRC, though its precise mechanism remains unclear. Our study demonstrated that knocking down ITGA11 significantly inhibits CRC cell migration and invasion. The mechanism behind ITGA11's involvement in lymph node-independent synchronous metastasis may encompass multiple pathways, with EMT being a potential key player, which we aim to explore further.

Study limitations are several. A notable drawback is the small sample size, stemming from our single-center study and the scarcity of specimens meeting our criteria. This challenge weakened our study's statistical power and robustness. Also, the heterogeneity of extensive stage II disease (N0), encompassing tumors confined to the serosa (T3) and those extending beyond it (T4), representing diverse histopathological risks. Additionally, even when an adequate number of lymph nodes are examined, there is a possibility of lymph node micrometastasis, as conventional histopathological examination cannot detect the presence of isolated tumor cells (ITCs) or micrometastases (MMs) within regional lymph nodes, and we did not perform ultra-staging for all these cases. Despite the limitations, we are committed to promoting multi-center, large-sample studies and employing multi-omics analysis in future research to better uncover the mechanisms underlying lymph node-independent distant metastasis and offer more reliable insights.

In summary, we integrated proteomics, multi-omics analysis, and machine learning to identify molecular features and developed an LIMGs signature based on nine genes, effectively predicting synchronous distant metastasis risk in stage I-II CRC patients. We also analyzed associations between the LIMG Score and pathological features, immune microenvironment and activity, and drug responses, offering insights into precise stratification and personalized therapy for CRC. Our findings also position ITGA11 as a crucial prognostic indicator for CRC metastasis.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

Ethics statement

The studies involving humans were approved by Ethics Committee of The First Affiliated Hospital of Wenzhou Medical University (KY2022-183). The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal

guardians/next of kin because given the retrospective nature of the study, informed consent was waived.

Author contributions

BZ: Writing – original draft, Conceptualization, Investigation. CZ: Writing – original draft, Investigation, Formal Analysis, Methodology. YC: Conceptualization, Investigation, Writing – original draft, Methodology. NS: Writing – original draft, Validation. YH: Validation, Visualization, Writing – original draft. HH: Visualization, Writing – original draft. BO: Writing – original draft, Visualization. QZ: Writing – original draft, Investigation, Supervision, Conceptualization, Validation. HJ: Supervision, Writing – original draft, Project administration. YZ: Writing – original draft, Supervision, Validation. PL: Conceptualization, Writing – review & editing, Supervision, Writing – original draft, Investigation. YP: Conceptualization, Writing – original draft. XZ: Methodology, Conceptualization, Writing – review & editing, Formal Analysis, Writing – original draft, Project administration.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study is funded by Wenzhou Science and Technology Public Project (NO. Y20220184) and Wenzhou Science and Technology Public Project (NO. Y2020145).

Acknowledgments

The authors thank all participants in the study.

References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer J Clin.* (2021) 71:209–49. doi: 10.3322/caac.21660
- Adam R, de Gramont A, Figueras J, Kokudo N, Kunstlinger F, Loyer E, et al. Managing synchronous liver metastases from colorectal cancer: A multidisciplinary international consensus. *Cancer Treat Rev.* (2015) 41:729–41. doi: 10.1016/j.ctrv.2015.06.006
- Ihnát P, Vávra P, Vra PZ. Treatment strategies for colorectal carcinoma with synchronous liver metastases: Which way to go? *World J Gastroenterol.* (2015) 21:7014–21. doi: 10.3748/wjg.v21.i22.7014
- van Gestel YRBM, de Hingh IHJT, van Herk-Sukel MPP, van Erning FN, Beerepoot LV, Wijsman JH, et al. Patterns of metachronous metastases after curative treatment of colorectal cancer. *Cancer Epidemiol.* (2014) 38:448–54. doi: 10.1016/j.canep.2014.04.004
- Siddiqui M, Nagtegaal I, Santiago I, Knijn N, Berho M, Mirnezami A, et al. Session 2: What causes liver metastases – lymph nodes or is it something else? *Colorectal Dis.* (2018) 20:39–42. doi: 10.1111/codi.14077
- Knijn N, van Erning FN, Overbeek LIH, Punt CJ, Lemmens VE, Hugen N, et al. Limited effect of lymph node status on the metastatic pattern in colorectal cancer. *Oncotarget.* (2016) 7:31699–707. doi: 10.18632/oncotarget.9064
- Naxerova K, Reiter JG, Brachtel E, Lennerz JK, van de Wetering M, Rowan A, et al. Origins of lymphatic and distant metastases in human colorectal cancer. *Science.* (2017) 357:55–60. doi: 10.1126/science.aai8515
- Knijn N, Mekenkamp LJM, Klomp M, Vink-Börger ME, Tol J, Teerenstra S, et al. KRAS mutation analysis: a comparison between primary tumours and matched liver metastases in 305 colorectal cancer patients. *Br J Cancer.* (2011) 104:1020–6. doi: 10.1038/bjc.2011.26
- Enquist IB, Good Z, Jubb AM, Fuh G, Wang X, Junttila MR, et al. Lymph node-independent liver metastasis in a model of metastatic colorectal cancer. *Nat Commun.* (2014) 5:3530. doi: 10.1038/ncomms4530
- Hu Z, Ding J, Ma Z, Sun R, Seoane JA, Scott Shaffer J, et al. Quantitative evidence for early metastatic seeding in colorectal cancer. *Nat Genet.* (2019) 51:1113–22. doi: 10.1038/s41588-019-0423-x
- Tivey A, Church M, Rothwell D, Dive C, Cook N. Circulating tumour DNA — looking beyond the blood. *Nat Rev Clin Oncol.* (2022) 19:600–12. doi: 10.1038/s41571-022-00660-y
- Mani DR, Krug K, Zhang B, Satpathy S, Clauser KR, Ding L, et al. Cancer proteogenomics: current impact and future prospects. *Nat Rev Cancer.* (2022) 22:298–313. doi: 10.1038/s41568-022-00446-5
- Bates RC, Mercurio A. The epithelial-mesenchymal transition (EMT) and colorectal cancer progression. *Cancer Biol Ther.* (2005) 4:371–6. doi: 10.4161/cbt.4.4.1655

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Correction note

This article has been corrected with minor changes. These changes do not impact the scientific content of the article.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2025.1622528/full#supplementary-material>

14. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods*. (2007) 4:207–14. doi: 10.1038/nmeth1019
15. Deutsch EW, Lane L, Overall CM, Bandeira N, Baker MS, Pineau C, et al. Human proteome project mass spectrometry data interpretation guidelines 3.0. *J Proteome Res*. (2019) 18:4108–16. doi: 10.1021/acs.jproteome.9b00542
16. Li C, Sun YD, Yu GY, Cui JR, Lou Z, Zhang H, et al. Integrated omics of metastatic colorectal cancer. *Cancer Cell*. (2020) 38:734–747.e9. doi: 10.1016/j.ccell.2020.08.002
17. Engebretsen S, Bohlin J. Statistical predictions with glmnet. *Clin Epigenetics*. (2019) 11:123. doi: 10.1186/s13148-019-0730-1
18. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. In: von Stechow L, editor. *Cancer Systems Biology: Methods and Protocols*. Clifton, New Jersey: Springer (2018). p. 243–59. doi: 10.1007/978-1-4939-7493-1_12
19. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res*. (2017) 77:e108–10. doi: 10.1158/0008-5472.CAN-17-0307
20. Plattner C, Finotello F, Rieder D. Chapter Ten - Deconvoluting tumor-infiltrating immune cells from RNA-seq data using quanTIseq. In: Galluzzi L, Rudqvist NP, editors. *Methods in Enzymology*, vol. 636. Tumor Immunology and Immunotherapy - Integrated Methods Part B. Amsterdam, Netherlands; Boston, MA: Academic Press (2020). p. 261–85. doi: 10.1016/bs.mie.2019.05.056
21. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol*. (2016) 17:218. doi: 10.1186/s13059-016-1070-5
22. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *eLife*. (2017) 6:e26476. doi: 10.7554/eLife.26476
23. *Profiling Tumor Infiltrating Immune Cells with CIBERSORT*. SpringerLink.
24. Han Y, Wang Y, Dong X, Sun D, Liu Z, Yue J, et al. TISCH2: expanded datasets and new tools for single-cell transcriptome analyses of the tumor microenvironment. *Nucleic Acids Res*. (2023) 51:D1425–31. doi: 10.1093/nar/gkac959
25. Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med*. (2013) 10(5):e1001453. doi: 10.1371/journal.pmed.1001453
26. Nitzkorski JR, Farma JM, Watson JC, Siripurapu V, Zhu F, Matteotti RS, et al. Outcome and natural history of patients with stage IV colorectal cancer receiving chemotherapy without primary tumor resection. *Ann Surg Oncol*. (2012) 19:379–83. doi: 10.1245/s10434-011-2028-1
27. Ahmed S, Leis A, Chandra-Kanthan S, Fields A, Zaidi A, Abbas T, et al. Regional lymph nodes status and ratio of metastatic to examined lymph nodes correlate with survival in stage IV colorectal cancer. *Ann Surg Oncol*. (2016) 23:2287–94. doi: 10.1245/s10434-016-5200-9
28. Hagness M, Foss A, Egge TS, Dueland S. Patterns of recurrence after liver transplantation for nonresectable liver metastases from colorectal cancer. *Ann Surg Oncol*. (2014) 21:1323–9. doi: 10.1245/s10434-013-3449-9
29. Veereman G, Robays J, Verleye L, Leroy R, Rolfo C, Van Cutsem E, et al. Pooled analysis of the surgical treatment for colorectal cancer liver metastases. *Crit Rev Oncology/Hematology*. (2015) 94:122–35. doi: 10.1016/j.critrevonc.2014.12.004
30. Zhao H, Achreja A, Jessi E, Logozzi M, Mizzoni D, Di Raimo R, et al. The key role of extracellular vesicles in the metastatic process. *Biochim Biophys Acta (BBA) - Rev Cancer*. (2018) 1869:64–77. doi: 10.1016/j.bbcan.2017.11.005
31. Gordon-Weeks A, Lim SY, Yuzhalin A, Lucotti S, Vermeer JAF, Jones K, et al. Tumour-derived laminin $\alpha 5$ (LAMA5) promotes colorectal liver metastasis growth, branching angiogenesis and notch pathway inhibition. *Cancers*. (2019) 11:630. doi: 10.3390/cancers11050630
32. Zeltz C, Gullberg D. The integrin–collagen connection – a glue for tissue repair? *J Cell Sci*. (2016) 129:1284. doi: 10.1242/jcs.188672
33. Yang X, Wei M, Huang Y, Yang X, Yuan Z, Huang J, et al. ITGA11, a prognostic factor associated with immunity in gastric adenocarcinoma. *IJGM*. (2024) 17:471–83. doi: 10.2147/IJGM.S444786
34. Iwai M, Tulafu M, Togo S, Kawaji H, Kadoya K, Namba Y, et al. Cancer-associated fibroblast migration in non-small cell lung cancers is modulated by increased integrin $\alpha 11$ expression. *Mol Oncol*. (2021) 15:1507–27. doi: 10.1002/1878-0261.12937
35. Primac I, Maquoi E, Blacher S, Heljasvaara R, Van Deun J, Smeland HY, et al. Stromal integrin $\alpha 11$ regulates PDGFR β signaling and promotes breast cancer progression. *J Clin Invest*. (2019) 129:4609–28. doi: 10.1172/JCI125890
36. Zheng H, An M, Luo Y, Diao X, Zhong W, Pang M, et al. PDGFR α +ITGA11+ fibroblasts foster early-stage cancer lymphovascular invasion and lymphatic metastasis via ITGA11-SELE interplay. *Cancer Cell*. (2024) 42:682–700.e12. doi: 10.1016/j.ccell.2024.02.002
37. Bartolini A, Cardaci S, Lamba S, Oddo D, Marchiò C, Cassoni P, et al. BCAM and LAMA5 mediate the recognition between tumor cells and the endothelium in the metastatic spreading of KRAS-mutant colorectal cancer. *Clin Cancer Res*. (2016) 22:4923–33. doi: 10.1158/1078-0432.CCR-15-2664
38. Yousif LF, Di Russo J, Sorokin L. Laminin isoforms in endothelial and perivascular basement membranes. *Cell Adhesion Migration*. (2013) 7:101–10. doi: 10.4161/cam.22680
39. Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, Howarth K, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat Genet*. (2010) 42:973–7. doi: 10.1038/ng.670
40. Ueki A, Komura M, Koshino A, Wang C, Nagao K, Homochi M, et al. Stromal POSTN enhances motility of both cancer and stromal cells and predicts poor survival in colorectal cancer. *Cancers*. (2023) 15:606. doi: 10.3390/cancers15030606
41. Qin J, Hu S, Chen Y, Xu M, Xiao Q, Lou J, et al. Hypoxia promotes Malignant progression of colorectal cancer by inducing POSTN+ Cancer-associated fibroblast formation. *Mol Carcinogenesis*. (2025) 64:716–32. doi: 10.1002/mc.23882
42. Feng K, Di Y, Han M, Yan W, Wang Y. SORBS1 inhibits epithelial to mesenchymal transition (EMT) of breast cancer cells by regulating PI3K/AKT signaling and macrophage phenotypic polarization. *Aging*. (2024) 16:4789–810. doi: 10.18632/aging.205632
43. Song L, Chang R, Dai C, Wu Y, Guo J, Qi M, et al. SORBS1 suppresses tumor metastasis and improves the sensitivity of cancer to chemotherapy drug. *Oncotarget*. (2016) 8:9108–22. doi: 10.18632/oncotarget.12851
44. Miosge N, Holzhausen S, Zelent C, Sprysch P, Herken R. Nidogen-1 and nidogen-2 are found in basement membranes during human embryonic development. *Histochem J*. (2001) 33:523–30. doi: 10.1023/A:101499523521
45. Beck AH, Lee CH, Witten DM, Gleason BC, Edris B, Espinosa I, et al. Discovery of molecular subtypes in leiomyosarcoma through integrative molecular profiling. *Oncogene*. (2010) 29:845–54. doi: 10.1038/onc.2009.381
46. Tang G, Wu D, Guo M, Li H. LncRNA MIR497HG inhibits colorectal cancer progression by the miR-3918/ACTG2 axis. *J Genet*. (2022) 101:27. doi: 10.1007/s12041-022-01367-w
47. Sweetwyne MT, Murphy-Ullrich JE. Thrombospondin1 in tissue repair and fibrosis: TGF- β -dependent and independent mechanisms. *Matrix Biol*. (2012) 31:178–86. doi: 10.1016/j.matbio.2012.01.006
48. Ramchandani D, Mittal V. Thrombospondin in tumor microenvironment. In: Birbrair A, editor. *Tumor Microenvironment: Extracellular Matrix Components – Part B*. Cham, Switzerland: Springer International Publishing (2020). p. 133–47. doi: 10.1007/978-3-030-48457-6_8
49. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. *CA: A Cancer J Clin*. (2017) 67:93–9. doi: 10.3322/caac.21388. Amin.
50. Cossu ML, Ginesu GC, Feo CF, Fancellu A, Pinna A, Vargiu I, et al. Surgical outcomes in patients with hepatic synchronous and metachronous colorectal metastases. *Annali Italiani di Chirurgia*. (2017) 88:497–504.
51. Wang HW, Yan XL, Wang LJ, Zhang MH, Yang CH, Wei-Liu , et al. Characterization of genomic alterations in Chinese colorectal cancer patients with liver metastases. *J Trans Med*. (2021) 19:313. doi: 10.1186/s12967-021-02986-0
52. Kuo YT, Tsai WS, Hung HY, Hsieh PS, Chiang SF, Lai CC, et al. Prognostic value of regional lymph node involvement in patients with metastatic colorectal cancer: palliative versus curative resection. *World J Surg Oncol*. (2021) 19:150. doi: 10.1186/s12957-021-02260-z
53. Holch JW, Ricard I, Stintzing S, Modest DP, Heinemann V. The relevance of primary tumour location in patients with metastatic colorectal cancer: A meta-analysis of first-line clinical trials. *Eur J Cancer*. (2017) 70:87–98. doi: 10.1016/j.ejca.2016.10.007
54. Yaeger R, Chatila WK, Lipsyc MD, Hechtman JF, Cercek A, Sanchez-Vega F , et al. Clinical sequencing defines the genomic landscape of metastatic colorectal cancer. *Cancer Cell*. (2018) 33:125–136.e3. doi: 10.1016/j.ccell.2017.12.004
55. Orhan A, Justesen TF, Raskov H, Qvortrup C, Gögenur I. Introducing neoadjuvant immunotherapy for colorectal cancer: advancing the frontier. *Ann Surgery*. (2025) 281:95. doi: 10.1097/SLA.0000000000006443
56. Laumont CM, Banville AC, Gilardi M, Hollern DP, Nelson BH. Tumour-infiltrating B cells: immunological mechanisms, clinical impact and therapeutic opportunities. *Nat Rev Cancer*. (2022) 22:414–30. doi: 10.1038/s41568-022-00466-1
57. Ghorbaninezhad F, Nour MA, Farzam OR, Saeedi H, Vanan AG, Bakhshivad M, et al. The tumor microenvironment and dendritic cells: Developers of pioneering strategies in colorectal cancer immunotherapy? *Biochim Biophys Acta (BBA) - Rev Cancer*. (2025) 1880:189281. doi: 10.1016/j.bbcan.2025.189281