Frontiers in Bioscience Landmark

Front. Biosci. (Landmark Ed) **2026**; 31(11): 55483
https://doi.org/10.31083/FBL55483

Original Research

Identification of LAP3 as a Potential Diagnostic Biomarker Associated With Tumor Cell Proliferation in Colorectal Carcinoma

Shaopeng Wang[1,2], Jiaqi Zhai[1], Jiang Ni[3], Chen Wang[1,2], Yong Q. Chen[1,4,*], Rong Wang[1,4,*]

[1]Jiangnan University Medical Center, Wuxi School of Medicine, Jiangnan University, 214000 Wuxi, Jiangsu, China
[2]Department of Information, Wuxi No.2 People's Hospital, 214000 Wuxi, Jiangsu, China
[3]Department of Pharmacy, Affiliated Hospital of Jiangnan University, Jiangnan University, 214000 Wuxi, Jiangsu, China
[4]School of Food Science and Technology, Jiangnan University, 214000 Wuxi, Jiangsu, China
*Correspondence: yqchen@jiangnan.edu.cn (Yong Q. Chen); 8202305007@jiangnan.edu.cn (Rong Wang)
Academic Editor: Sung Eun Kim
Submitted: 1 July 2026 Revised: 11 October 2026 Accepted: 21 October 2026 Published: 16 November 2026

Abstract

Background: Colorectal carcinoma (CRC) remains a major cause of cancer-related mortality worldwide, and microsatellite instability (MSI) is an important molecular determinant of treatment response and CRC subtyping. The identification of potential MSI-related biomarkers could facilitate the development of clinical diagnostic tools to accurately predict the MSI status of patients with CRC. In addition, investigating the associations of candidate biomarkers with the immune microenvironment and immunotherapy-related features may help identify immune factors that influence treatment responses and outcomes in patients with CRC. Furthermore, analyzing the prognostic significance of these biomarkers, potential targeted drugs, and drug responses could enhance their translational relevance. Finally, investigating their biological roles in lipid metabolism, amino acid metabolism, immune regulation, oxidative stress, and epithelial–mesenchymal transition may further clarify their functional involvement in CRC metastasis and progression. **Methods:** An integrated analysis of 671 bulk transcriptomic samples and 18 single-cell samples comprising 27,414 cells was performed. Five machine-learning algorithms were applied to screen for candidate biomarkers associated with MSI status and CRC progression and to develop diagnostic models interpreted using the SHapley Additive exPlanations (SHAP) framework. The differential expression patterns of candidate genes were further validated in several tumor cell lines and patient tissues. *In silico* molecular docking and *in vitro* drug screening, combined with flow cytometry, were subsequently performed to identify and evaluate potential therapeutic agents. **Results:** The expression patterns of three candidate genes (*APOL3*, *PLA2G4F*, and *LAP3*) were validated across integrated multi-omics datasets, and the expressions of *PLA2G4F* and *LAP3* were significantly associated with patient prognosis. The best-performing models based on these three genes demonstrated diagnostic performance in internal and external validation datasets, with area under the curve (AUC) values ranging from 0.712 to 0.854, and provided interpretable predictions. Further experiments revealed consistent upregulation of *LAP3*, which encodes a metabolic metallopeptidase, at the mRNA and protein levels across several cell lines and patient tissues. Mitoxantrone, an FDA-approved topoisomerase II inhibitor identified through drug screening, was predicted to bind to LAP3 and inhibited CRC cell proliferation *in vitro*. **Conclusions:** Mitoxantrone was predicted to interact with LAP3 with favorable binding energy. However, the potential of the LAP3–mitoxantrone target–drug pair as a therapeutic strategy for CRC requires further computational and experimental validation. These findings provide insights into MSI-related biomarkers and their potential utility in improving molecular classification, prognostic assessment, and precision treatment of CRC.

Keywords: leucine aminopeptidase 3; mitoxantrone; colorectal neoplasms; microsatellite instability; machine learning

1. Introduction

Colorectal carcinoma (CRC) accounts for approximately 1.9 million new cases annually worldwide, and is the second leading cause of cancer-related mortality [1] with incidence is higher in men and increases up to middle age, then declines after 70 years of age [2,3]. Current treatments include chemotherapy with agents such as oxaliplatin and irinotecan, targeted therapies including cetuximab and immunotherapies including pembrolizumab. CRC development is associated with multiple environmental, acquired and genetic factors including microsatellite instability (MSI) [2]. Consequently, its early diagnosis and precision treatment remain challenging. MSI results from the dysfunction of the DNA mismatch repair system and is characterized by high frequencies of point mutations and insertion-deletion mutations within microsatellite regions [4]. Several human mismatch repair proteins have been identified based on their homology to proteins in *E. coli*, including components of the MutS and MutL complexes, exonuclease 1 (EXO1) and replication protein A (RPA) [5]. CRC can be classified as microsatellite instability (MSI)-stable (MSS), MSI-low (MSI-L), and MSI-high (MSI-H). MSI pathways occurs in approximately 15% of sporadic CRCs and in most CRCs associated with hered-

Copyright: © 2026 The Author(s). Published by IMR Press.
This is an open access article under the CC BY 4.0 license.

Publisher's Note: IMR Press stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

itary non-polyposis CRC. These observations underscore the importance of MSI status in the molecular classification and prognostic assessment of CRC [6,7]. Furthermore, CRC patients with different MSI statuses may exhibit distinct responses to pharmacological treatments [8,9], suggesting that the identification and functional characterization of MSI-related genes contribute to precision treatment.

In addition to MSI, chromosomal instability (CIN) and the CpG island methylator phenotype (CIMP) are major molecular pathways involved in CRC pathogenesis [10]. CIN, which occurs in approximately 84% of sporadic CRCs, is characterized by widespread changes in chromosome number, chromosomal rearrangements, and frequent somatic copy-number alterations [11,12]. Transcriptomic analyses by The Cancer Genome Atlas (TCGA) and the consensus molecular subtypes (CMS) classification have further characterized CRC heterogeneity. The major CMS groups include CMS2, the canonical subtype characterized by WNT and MYC signaling activation (37%), CMS3, characterized by metabolic dysregulation (13%), and CMS4, the mesenchymal subtype (23%) [7]. These subtypes may differ in their responses to immunotherapy and targeted treatments. CIMP frequently occurs in MSI CRC and promotes epigenetic instability through aberrant promoter hypermethylation and the silencing of tumor-suppressor genes [13]. CIMP-high sporadic CRCs show frequent methylation of *p16* and *THBS1* [14]. Molecular subtyping and biomarker identification based on CIN, CIMP and CMS may therefore improve the understanding of the molecular and immunological heterogeneity of CRC and facilitate prognostic assessment and the selection of immunotherapies, and targeted treatments [12,15].

Characterizing MSI-related genes and their association with the immune microenvironment, prognosis, and treatment response may improve the understanding of their potential applications in precision treatment and clinical practice. Furthermore, by revealing gene-expression heterogeneity at single cell resolution, single cell RNA sequencing (scRNA-seq) can help characterize tumor cell populations and elucidate mechanisms underlying tumorigenesis and progression.

Machine learning (ML) algorithms have attracted considerable attention in recent years and have been used to develop models that assess the discriminatory performance of biomarkers in multiple cancer types [16,17,18,19,20,21]. These algorithms can also identify complex or subtle relationships within heterogeneous datasets [22,23,24,25,26]. However, ML models frequently have limited interpretability and are therefore regarded as “black-boxes” [27]. The SHapley Additive exPlanation (SHAP) method [28] can improve model interpretability by providing both global explanations of overall feature importance and local explanations of individual predictions.

In this study, six datasets comprising 671 bulk transcriptomic samples and 18 single-cell samples were integrated to identify candidate MSI-related biomarkers in CRC. Three genes, apolipoprotein L3 (*APOL3*), phospholipase A2 group IVF (*PLA2G4F*), and leucine aminopeptidase 3 (*LAP3*), were selected as a candidate biomarker panel. Their gene and protein expression patterns were subsequently evaluated in CRC cell lines and patient tissues. In addition, *in silico* molecular docking and *in vitro* drug screening were performed to evaluate potential therapeutic agents. Among the candidate biomarkers, *LAP3*, which encodes a cytosolic metallopeptidase, was further investigated as a potential biomarker and therapeutic target. Molecular docking predicted a potential interaction between LAP3 and mitoxantrone, whereas *in vitro* experiments showed that mitoxantrone inhibited CRC cell proliferation. However, direct target engagement and the mechanistic relationship between LAP3 and the antiproliferative activity of mitoxantrone require further validation.

2. Materials and Methods

2.1 Data Sources

Transcriptomic data from 147 patients with CRC and known MSI status, previously reported by Kim et al. [29] were obtained from TCGA [30] through the University of California, Santa Cruz (UCSC) Xena platform [31] and designated the primary CRC dataset, which included 97 patients with MSS, 23 with MSI-L, and 27 with MSI-H tumors. An endometrial carcinoma dataset reported by Kim et al. [29], conprising 130 patients with known MSI status, was used for cross-cancer external validation. Additionally, four tumor-normal datasets obtained from TCGA and the Gene Expression Omnibus (GEO) [32] (GSE271719 [33], GSE32323 [34] and GSE8671 [35]) as well as one single-cell RNA-sequencing dataset (GSE144735 [36]) were used to validate the three candidate genes. Further information is provided in **Supplementary Method 1**.

2.2 Differentially Expressed Genes (DEGs) and Enrichment Analyses

The R package ‘limma’ (version 3.58.1) was used to identify two sets of DEGs in the primary CRC dataset. The first set comprises DEGs between the combined MSI group (MSIC; including patients with MSI-H and MSI-L tumors) and the MSS group. The second set comprises DEGs between the MSI-H and MSI-L groups. The detailed description of DEGs and enrichment analyses is provided in **Supplementary Method 1**.

2.3 Weighted Gene Co-Expression Network Analysis (WGCNA)

WGCNA (version 1.73) [37] was performed using the gene expression profiles from the primary CRC dataset to identify gene co-expression modules associated with the MSIC group relative to the MSS group. Genes belonging to modules positively correlated with the MSIC phenotype were selected for subsequent analyses. The detailed

description with setting parameters is provided in **Supplementary Method 1**.

2.4 Screening of Potential MSI-Related Biomarkers and Analysis of the Immune Microenvironment and Immunotherapy-Related Features

Least Absolute Shrinkage and Selection Operator (LASSO) regression ('glmnet' library, version 4.1.8) and random forest (RF) algorithms ('randomForest' library, version 4.7.1.2) were used to screen for potential MSI-related biomarkers. For LASSO regression, the lambda value that minimized 10 fold cross-validation error was selected and genes were ranked according to their coefficient values in lasso model. For the RF analysis, potential biomarkers were selected based on the out-of-bag errors with 'ntree' set to 110 and genes were ordered based on feature split criteria of gini impurity. Differences in immune-cell infiltration between MSS and MSIC groups were analyzed using the CIBERSORT (version 0.1.0) [38]. T-cell states and enrichment of immune-related gene sets were evaluated using TCellSI (version 1.2.0) [39], and ssGSEA (version 1.50.5) [40], respectively, with the parameters described in **Supplementary Method 1**. Associations between the identified potential MSI-related biomarkers and features of the immune microenvironment including macrophage polarization and dendritic cells, T-cell exhaustion, cytokine networks, and immune checkpoint molecules were analyzed. Immunotherapy-related features including tumor mutation burden (TMB), PD-1/PD-L1, CTLA-4, mismatch repair proteins, *etc*. were assessed at the gene and patient levels using TCGAplot (version 8.0.0) [41] and ssGSEA, as detailed in **Supplementary Method 1**.

2.5 Synthetic Minority Over-Sampling Technique (SMOTE)

The SVMSMOTE method in 'imblearn' library (version 0.12.4) was applied exclusively to the training set to address class imbalance and reduce prediction bias. Models trained on the balanced training set was subsequently evaluated using untouched internal and external validation sets [42,43]. A brief description of the SMOTE procedure is provided in **Supplementary Method 1**.

2.6 Predictive Performance, Model Validations, and Model Interpretation

The construction and evaluation of five machine-learning models implemented in 'scikit-learn' (version 1.2.2) and 'xgboost' (2.0.2) libraries, their predictive performance in datasets including samples classified by MSI status or tumor-normal status and the survival analyses ('survival library', version 3.5.5) are described in **Supplementary Method 1**. The major hyper-parameters of models were optimized using grid search combined with manual adjustment. Model performance was primarily evaluated based on the area under the curve (AUC) of receiver operating characteristic (ROC). Ten-fold cross validation was applied to the training set, and those achieving higher AUC values were selected for subsequent validation. To address the limited interpretability of complex machine-learning models, SHAP (version 0.44.1) was used to interpret the above selected models on the training set at both the global and local levels, thereby facilitating the understanding of their decision-making processes. Additional details of the SHAP analysis are provided in **Supplementary Method 1**.

2.7 scRNA-seq Analysis and In Silico Molecular Docking

The scRNA-seq dataset comprises 27,414 cells obtained from tumor core regions, tumor border regions and matched normal mucosa from 6 patients, yielding a total of 18 samples. The dataset was retrieved from GEO and used to investigate cellular heterogeneity and mechanisms potentially involved in CRC pathogenesis. The MSI-related candidate genes and top-50 DEGs between CRC and normal tissues were used to distinguish malignant from nonmalignant epithelial cells with 'AddModuleScore_UCell' function ('UCell' library, version 2.6.2) and threshold of zero.

Large-scale molecular docking of drug-like compounds from three compound databases against the proteins encoded by the three candidate genes was performed to identify compounds with favorable predicted binding energies. Given the absence of experimentally resolved protein structures, the structures of the three genes were retrieved from the AlphaFold and AlphaFill databases. The docking software, the preparation steps for protein structures and compounds, and the scoring functions are described in **Supplementary Method 2**. The top 100 compounds with the most favorable predicted binding energies were retained, and the predicted binding interactions of the three highest-ranking compounds were visually examined in detail. The drug efficacy of the three screened compounds against each protein were subsequently evaluated through *in vitro* experiments.

2.8 Cell Culture, Imaging, and Viability Analysis

Human CRC cell lines Caco-2 (ATCC HTB-37), HCT116 (CCL-247), and the normal human colon fibroblast cell line CCD-18Co (ATCC CRL-1459) were cultured in Minimum Essential Medium (MEM; Thermo Fisher Scientific, CA, USA, 11095080) supplemented with 10% fetal bovine serum (FBS; Thermo) and 1% penicillin-streptomycin at 37 °C in a humidified atmosphere containing 5% CO_2. The cell lines were authenticated by short tandem repeat analysis and tested for mycoplasma contamination using the PCR Mycoplasma Detection Set (Takara, OT, JP) and tested negative for mycoplasma. The effect of solabegron (MCE, NJ, USA, HY-19436), chiauranib (MCE, HY-124526), arbutamine (MCE, HY-16056), N^2-methylguanine (MCE, HY-101412), pixantrone dimaleate (MCE, HY-13727A), harmine (MCE, HY-N0737A), cyclen (MCE, HY-W007656), and mitoxantrone (MCE, HY-13502) on cell viability was evaluated after treatment at

different concentrations for 24 h. Cyclen was dissolved in water, whereas all other compounds were dissolved in dimethyl sulfoxide (DMSO).

Cell viability was assessed using the Cell Counting Kit (CCK-8; MCE, NJ, USA, HY-K0301) according to the manufacturer's instructions. Briefly, cells were trypsinized and stained with trypan blue, and viable cells were counted using a hemocytometer. Cells were seeded into a 96-well plate at a concentration of 4000 cells/200 µL/well and incubated overnight. After treatment, 10 µL of CCK-8 solution was added to each well and cells were incubated for 4 h at 37 °C. Absorbance was measured at 450 nm using a microplate spectrophotometer (Thermo, CA, USA). Three independent experiments were performed, each with three technical replicates.

2.9 Quantitative Real-Time PCR (qRT-PCR) and Western Blotting (WB)

qRT-PCR was used to quantify the mRNA expression of the three candidate genes, whereas WB was used to assess protein expression. Total RNA was extracted using the FastPure cell/tissue total RNA isolation kit (Vazyme Biotech Co., Ltd., Nanjing, China, RC101) according to the manufacturer's instructions. One µg of total RNA for each sample was reverse-transcribed into cDNA using a cDNA reverse transcription kit (Vazyme, R323). Real-time PCR was performed in triplicate using gene-specific primers and a Bio-Rad CFX96 PCR system. Gene expression was normalized to that of GAPDH. The primer sequences were as follows: APOL3-F: GGCACCATGGACTCAGTAGAA, APOL3-R: GCCGCAGTCACGAATCTCT; PLA2G4F-F: CCCACTCAACCACCAGGATT, PLA2G4F-R: ACCCCGTTGGTGATGACTTC; LAP3-F: AAGCCGGGGGATGTTGTTAG, LAP3-R: TCGAAGAGAGGCATCCTCCA; GAPDH-F: GGAGCGAGATCCCTCCAAAAT, GAPDH-R: GGCTGTTGTCATACTTCTCATGG.

Colorectal tissues of patients were homogenized in RIPA lysis buffer (Beyotime, Shanghai, China, P0013B) using a tissue grinder. Cells were lysed using cell lysis buffer for Western and IP (Beyotime, P0013) and the resulting lysate was heated at 95 °C for 10 min. Proteins were separated by SDS-PAGE gel and transferred to a PVDF membrane (Millipore, NH, USA, IPVH00010). The membrane was incubated with the indicated primary antibodies, followed by horseradish peroxidase-conjugated secondary antibodies, and immunoreactive bands were detected using an enhanced chemiluminescence reagent (ECL; Sigma, MO, USA, WBULS0500). Primary antibodies against APOL3 (28591-1-AP; 1:800), LAP3 (66417-1-Ig; 1:3000), and GAPDH (60004-1-Ig; 1:50,000) were purchased from Proteintech (IL, USA). PLA2G4F (PA5-31736; 1:1000) was purchased from Invitrogen (MA, USA). The horseradish peroxidase-conjugated secondary antibodies were also purchased from Proteintech (IL, USA; SA00001-2, SA00001-1; 1:10,000).

2.10 Hematoxylin and Eosin (H&E) Staining and Immunohistochemistry (IHC)

Tumor and adjacent non-tumor tissues were collected from 10 patients diagnosed with CRC. The study was performed in accordance with the protocol approved by the Ethics Committee of Affiliated Hospital of Jiangnan University approval No. LS2021028, and written informed consent was obtained from all participants. H&E staining was performed according to a routine protocol. Briefly, 5 µm-thick paraffin-embedded longitudinal tissue sections were deparaffinized, rehydrated and stained with hematoxylin solution for 5 min. Sections were dipped five times in 1% acid ethanol (1% HCl in 70% ethanol) and rinsed in distilled water. Subsequently, sections were stained with eosin solution for 3 min, dehydrated in a graded ethanol series, cleared in xylene and mounted. Slides were examined and digitized using a Pannoramic Scanner (3DHISTECH, Budapest, Hungary).

As regards IHC, paraffin-embedded CRC tissue blocks were sectioned and stained with antibodies against APOL3 (ab154869; 1:100; Abcam, Cambridge, UK) and LAP3 (66417-1-Ig; 1:200; Proteintech, IL, USA). Slices were digitized using a Pannoramic Scanner. Staining intensity was quantified using Aipathwell software and expressed as the mean integrated optical density (IOD) value.

2.11 Flow Cytometry

Flow cytometry was performed to determine the apoptotic rate. Cells were stained using Annexin V-YSFluor™ 488/PI Apoptosis Detection Kit (Yeasen, Shanghai, China, 40305ES50) according to the manufacturer's instruction and analyzed using BD Accuri C6 flow cytometer (BD Biosciences, NJ, USA). Flow cytometry data were analyzed using FlowJo 10.8.1 software and gating strategy was illustrated in **Supplementary Fig. 10B**.

2.12 Statistical Analysis

Statistical analysis was performed using R software (version 4.3.0). The comparison between groups was performed using a two-tailed Student's *t*-test or Mann-Whitney test, as appropriate. Differences between survival curves were assessed using the Log-rank test. The statistical significance in enrichment analysis was performed by hypergeometric test. A value of $p < 0.05$ was considered statistically significant.

3. Results

3.1 DEG Analysis Between Groups

The study workflow is shown in Fig. 1. A two-stage screening process was used to identify DEGs in the primary CRC dataset. In the first stage, the differential expression analysis was performed between the MSIC and MSS groups with criteria of 'adjusted p value < 0.05' and '$|\log_2$ fold change$|$ >0.5' and a total of 105 DEGs met the selection criteria were retained, including 55 upregulated and 50

downregulated genes. The corresponding volcano plot and heatmap are shown in Fig. 2A,B. The results of the GO, KEGG, and GSEA are shown in Fig. 2C–F. The GO analysis indicated that the DEGs were predominantly enriched in major histocompatibility complex (MHC)-related molecular functions (MF) and cellular component (CC) terms (Fig. 2C). The KEGG analysis identified the enrichment of pathways associated with intestinal immunity, inflammation, and metabolism (Fig. 2D). In the second stage, 235 DEGs were identified between the MSI-H and MSI-L groups, including 101 upregulated and 134 downregulated genes. The results of this analysis are presented in **Supplementary Fig. 1**.

3.2 Identification of a Gene Module Associated With MSIC Status

WGCNA was performed to identify gene modules associated with MSIC status (**Supplementary Fig. 2A–C** and Fig. 3A–C). The soft-thresholding power β was set to 8 in building similarity matrix and the minimum module size set to 30 to combine similar gene modules. A total of 127 genes in the purple module that was positively correlated with MSIC status were retained and denoted as the core MSIC associated genes for further analysis. The analysis details of WGCNA was provided in **Supplementary Method 1**.

3.3 Identification of MSI-Related Hub Genes and Candidate Biomarkers

Three gene sets were obtained, comprising 105 DEGs from the MSIC versus MSS comparison, 235 DEGs from the MSI-H versus MSI-L comparison, and 127 genes from the WGCNA purple module (Fig. 3D). The intersection yielded 11 genes, which were designated candidate MSI-related hub genes for subsequent analyses (Fig. 3E). The expression patterns of these 11 genes across all CRC samples are shown in **Supplementary Fig. 2D**. Differences in their expression between MSIC and MSS groups and between the MSI-H and MSI-L groups were assessed using Student's *t*-test and visualized using box-and-whisker plot (Fig. 3F,G). In both comparisons, nine genes showed significantly higher expression in the MSIC or MSI-H group, whereas *PLA2G4F* and *B3GNT8* showed significantly lower expression. These findings suggested that the 11 genes might represent candidate MSI-related gene signatures in CRC.

LASSO regression and RF analysis were subsequently used to prioritize the 11 hub genes. LASSO regression identified three genes with non-zero coefficients at the lambda value associated with the minimum cross-validation error (**Supplementary Fig. 2E** and Fig. 3H). The RF analysis selected the same three genes on the basis of out-of-bag error and ranked them highest according to Mean Decrease Accuracy and Mean Decrease Gini (**Supplementary Fig. 2F** and Fig. 3I). Accordingly, *APOL3*, *PLA2G4F*, and *LAP3* were selected as a three-gene panel of candidate MSI-related biomarkers in CRC. Principal component analysis (PCA) of the MSIC and MSS samples was also performed using the 11 hub genes and the three-gene panel (**Supplementary Fig. 2G**). Pairwise scatter plots showed the expression relationships among *APOL3*, *PLA2G4F*, and *LAP3* (Fig. 3J). The fitted models yielded R^2 values ranging from 0.16 to 0.40, indicating limited-to-moderate shared variation among the three genes.

3.4 Immune Microenvironment and Immunotherapy-Related Feature Analyses

CIBERSORT was used to investigate differences in the estimated immune-cell composition between the MSIC and MSS cohorts. **Supplementary Fig. 3A** shows the distribution of 22 immune-cell types across the CRC samples. The MSIC cohort showed higher estimated proportions of memory B cells, activated and resting dendritic cells, resting mast cells, NK cells activated, and regulatory T cells compared with the MSS cohort (**Supplementary Fig. 3B**). The correlation analysis was subsequently performed between the expression of the three candidate genes and the estimated proportions of the immune-cell population. *LAP3* expression was significantly associated with several T cell subsets, suggesting a potential relationship between *LAP3* expression and the immune microenvironment in CRC. *APOL3* and *PLA2G4F* expression was positively correlated with the estimated proportion of M1 macrophages and negatively correlated with that of M2 macrophages. By contrast, *LAP3* expression was positively correlated with estimated proportion of both M1 and M2 macrophages. *APOL3*, *PLA2G4F* and *LAP3* expression was also positively correlated with the estimated proportions of resting and activated dendritic cells (**Supplementary Fig. 3C**). These associations describe the relationship between gene expression and estimated cell proportions in immune microenvironment.

TCellSI was used to estimate eight T cell state scores in the MSI cohort. Helper and terminal-exhaustion scores differed significantly between the MSIC and MSS groups, with higher scores observed in the MSIC group (**Supplementary Fig. 4A**). *APOL3* and *LAP3* expression was positively correlated with terminal-exhaustion scores (r = 0.48 and r = 0.49, respectively; both $p < 0.001$; **Supplementary Fig. 4B**). Using ssGSEA, sample-level enrichment scores were calculated for cytokine network and immune-checkpoint gene signatures. *APOL3* expression showed a weak positive association with the cytokine-network score, whereas PLA2G4F expression showed a weak negative association with the immune-checkpoint score (**Supplementary Fig. 4C**). The other correlations shown in **Supplementary Fig. 4C** were not statistically significant.

At the gene level, pan-cancer analysis using TCGAplot examined the association between the three candidate genes and immunotherapy-related features, including

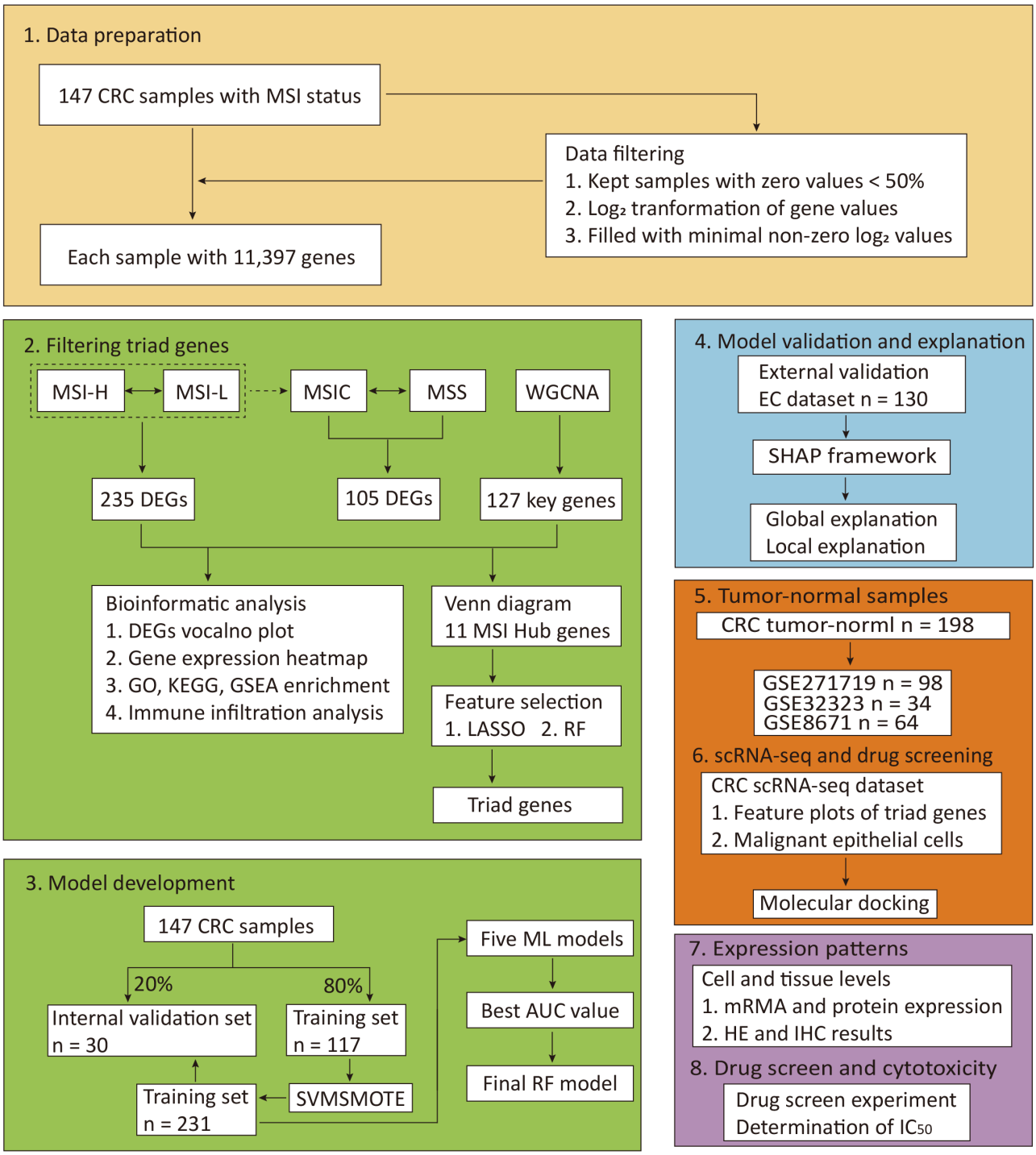


Fig. 1. Overview of the study workflow. The workflow comprised data preprocessing, screening of MSI-related candidate genes, machine-learning model development and validation, analysis of tumor–normal and scRNA-seq datasets, molecular docking, experimental validation of gene and protein expression, and drug screening. CRC, colorectal adenocarcinoma; MSI, microsatellite instability; MSI-H, MSI-high; MSI-L, MSI-low; MSIC, combined MSI group; MSS, MSI-stable; WGCNA, weighted gene co-expression network analysis; DEGs, differentially expressed genes; GO, gene ontology; KEGG, kyoto encyclopedia of genes and genomics; LASSO, Least Absolute Shrinkage and Selection Operator; RF, Random Forest; ML, machine learning; AUC, area under the curve; SHAP, SHapley Additive exPlanation; HE, hematoxylin eosin; IHC, immunohistochemistry; GSEA, gene set enrichment analysis.

tumor mutation burden (TMB), immune checkpoint genes, immune inhibitors, and immune scores. **Supplementary Fig. 5A** shows the correlation with TMB across cancer types, including colon adenocarcinoma and rectal adeno-

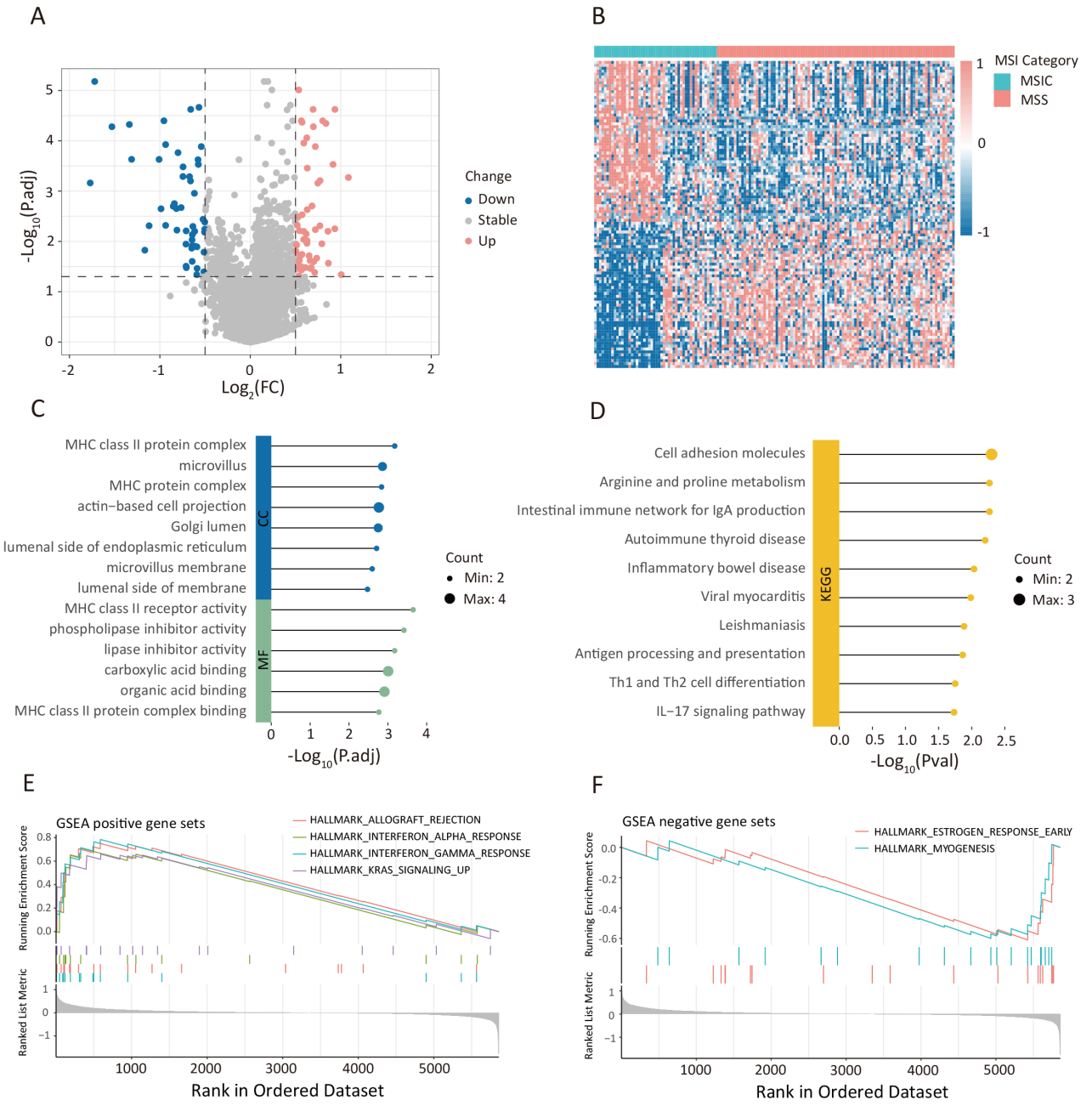


Fig. 2. DEGs between MSIC and MSS groups and enrichment analysis. (A) Volcano plot showing the DEGs between MSIC and MSS groups with blue and red points indicating significant down- and up-regulated genes met the given criterion. (B) Heatmap showing the expression patterns of the identified DEGs between MSIC and MSS groups. (C) Significantly enriched CC and MF terms in GO database. (D) Significantly enriched KEGG pathways. (E,F) Positively and negatively enriched gene sets, respectively, identified by GSEA based on the fold changes of ranked genes. CC, cell components; MF, molecular functions.

carcinoma. Predominantly positive correlation in these two cancer types was observed between the candidate genes and several immune-checkpoint genes and immune inhibitors (**Supplementary Fig. 5B,C**). *APOL3* and *LAP3* expression was also positively significant correlated with immune scores in CRC (**Supplementary Fig. 5D**). At the patient level, ssGSEA was applied to the 147 patient samples using a gene set comprising PD-1/PD-L1 and CTLA-4-related markers, mismatch repair genes, and the three candidate genes. The resulting enrichment scores were higher in the MSIC group than in the MSS group (**Supplementary Fig. 5E**). These scores represented enrichment of the selected gene set and expressions of immunotherapy-related genes in patients.

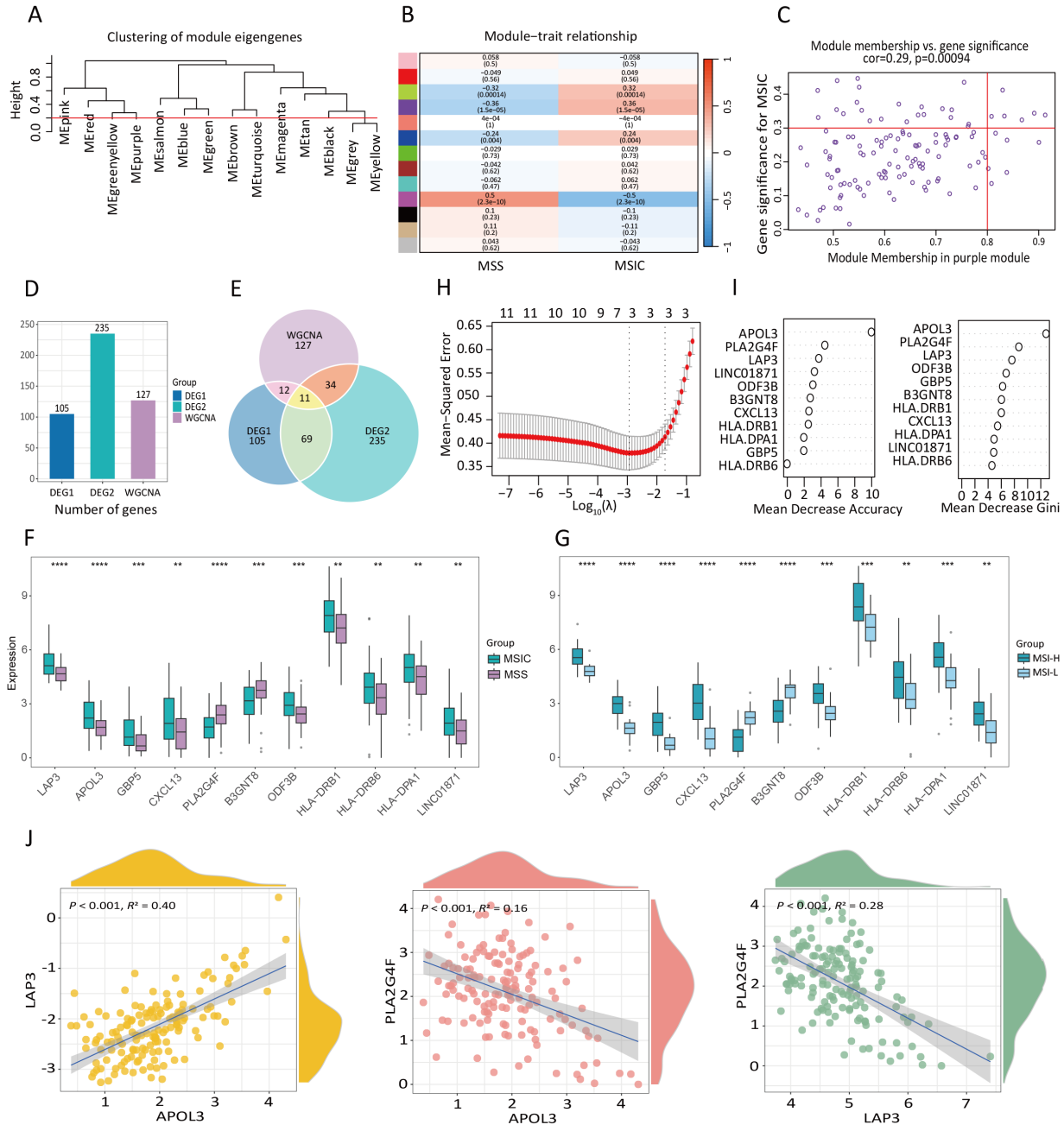


Fig. 3. WGCNA and machine-learning analysis for the identification of MSI-related hub genes. (A) Hierarchical clustering of module eigengenes with similar modules being combined. (B) Correlations between module eigengenes and the MSS and MSIC phenotypes. (C) Relationship between gene significance for MSIC status and module membership in the purple module. (D) Number of genes identified in the two differential expression analyses and the WGCNA purple module. (E) Venn diagram showing the intersections among the three gene sets and the 11 shared hub genes. (F,G) Gene expression patterns of the 11 hub genes in the MSIC versus MSS and MSI-H versus MSI-L comparison groups, respectively. (H) Cross-validation error as a function of the logarithm of lambda in the LASSO analysis with lower errors achieved with the triad genes. (I) RF variable-importance ranking based on Mean Decrease Accuracy and Mean Decrease Gini. (J) Pairwise expression relationships among *APOL3*, *PLA2G4F*, and *LAP3*, with fitted regression lines and R^2 values. $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$.

3.5 Model Development, Validation, and Interpretation

Five machine-learning algorithms were used to develop diagnostic models for classifying CRC samples according to MSI status. In the 10-fold cross-validation analysis, the model based on the three-gene panel showed better discriminatory performance than the models based on

the individual genes. For example, the RF model based on the three-gene panel in the 10-fold cross-validation analysis achieved a better AUC of 0.970 ± 0.021 (Fig. 4D and **Supplementary Fig. 6D**) than that of other models and was selected as the final model. In addition, compared with AUCs of RF models from *APOL3* (0.814 ± 0.031), *PLA2G4F* (0.849 ± 0.066), and *LAP3* (0.709 ± 0.058), respectively, the model based on three-gene panel also achieved better performance (Fig. 4A–C and **Supplementary Fig. 6A–C**). These results suggested that the three-gene panel provided greater discriminatory information for classifying CRC samples according to MSI status than the individual genes. The five algorithms were subsequently evaluated using the internal validation set. The RF model achieved AUCs of 0.854 and 0.814 in the internal and external validation sets, respectively (Fig. 4E,F). Complete model-performance metrics are provided in **Supplementary Dataset 1**.

SHAP was used to interpret the final RF model at the global and local levels. For prediction of the MSI-H class, the absolute mean SHAP-based importance values of the three genes are shown in descending order in Fig. 4G. The SHAP summary plot showed the distribution of the SHAP values for each gene across the samples, with feature values represented by a color scale from low to high (Fig. 4H). SHAP dependence plots showed the relationships between the expression value of each gene and its contribution to the model output for the MSI-H class (Fig. 4I). Positive SHAP values indicated contributions that increased the model output toward the MSI-H class, whereas negative values indicated contributions away from this class.

Local SHAP explanations were used to determine how the feature values contributed to predictions for individual samples. For the MSI-H sample shown in Fig. 4J, the predicted probabilities of the MSS, MSI-L, and MSI-H classes were 0.3%, 2.1%, and 97.6%, respectively. The waterfall plots showed the contributions of the three genes to each class-specific prediction. Corresponding local explanations for samples classified as MSI-L or MSS are shown in **Supplementary Figs. 7,8**, respectively.

3.6 Model Validation Using Tumor-Normal Datasets

Four tumor-normal datasets were analyzed to evaluate the ability of the hub genes to distinguish CRC tissues from normal tissues. The primary CRC tumor-normal dataset was divided into training and internal validation sets at a ratio of 8:2 followed by SVMSMOTE method to exclusively balanced training set. Based on the gene ranking shown in Fig. 3I, an incremental feature selection with 10-fold cross-validation was performed using the training set (Fig. 5A). Models incorporating all 11 hub genes achieved AUC values >0.900. The Gradient boosting decision tree (GBDT) models based on the three-gene panel and the 11 hub genes achieved cross-validated AUC values of 0.844 ± 0.044 and 0.976 ± 0.026, respectively (Fig. 5B,C).

Five machine-learning models based on either the three-gene panel or the 11 hub genes were subsequently evaluated using the internal validation set (Fig. 5D,E). Among the five algorithms, the GBDT models showed the highest AUC values for both gene sets. Fig. 5F shows the expression patterns of the 11 hub genes in the primary CRC tumor-normal dataset, nine of which were differentially expressed between tumor and normal tissues.

Three independent CRC tumor-normal datasets obtained from GEO were used for further validation of the three-gene panel and the hub gene signature (Fig. 5G–I). Since LINC01871 was not represented in these datasets, the hub-gene signature comprised the remaining ten genes. Complete performance metrics are provided in **Supplementary Dataset 2**. LAP3 was consistently upregulated in tumor and adenomas tissues across the three datasets.

3.7 Survival Analysis, Identification of Malignant Cells and Molecular Docking

CRC patients with available overall survival (OS) data from TCGA were stratified separately for each gene into high- and low-expression groups using the third quartile as the cutoff. The Kaplan-Meier survival analysis and log-rank test were then performed for each gene (Fig. 6A). OS differed significantly between the high- and low-expression groups for *PLA2G4F* ($p = 0.004$) and *LAP3* ($p = 0.014$) but not for *APOL3* ($p = 0.17$). High *PLA2G4F* expression was associated with better OS, whereas high *LAP3* expression was associated with poorer OS.

Low-quality cells and doublets were removed in the CRC scRNA-seq dataset (**Supplementary Fig. 9A**). PCA was performed using highly variable genes after log-normalization and scaling of the expression matrix (**Supplementary Fig. 9B**) and the first 15 PCs were selected for downstream analysis (**Supplementary Fig. 9C**). Thirteen cell clusters were identified (**Supplementary Fig. 9D**) and annotated as patients, tissue source (**Supplementary Fig. 9E**) and major cell types, which were visualized using t-distributed stochastic neighbor embedding (t-SNE) (Fig. 6C). The single-cell expression patterns of the three genes across various cell types are shown in Fig. 6B. The 50 most up- and down-regulated genes between CRC and normal samples, together with MSI hub genes, were used to classify epithelial cells as malignant (red points) or nonmalignant (blue points) (Fig. 6D).

Large-scale virtual molecular docking procedure was subsequently performed to identify compounds with favorable predicted binding energies against the proteins encoded by the three candidate genes. The preparation and screening procedure are described in **Supplementary Method 2**, and the 100 compounds with the most favorable predicted binding energies are presented in **Supplementary Dataset 3**. Detailed docking results for the three highest-ranked compounds for each protein are presented in Fig. 6E–G and **Supplementary Fig. 9F,G**. The fig-

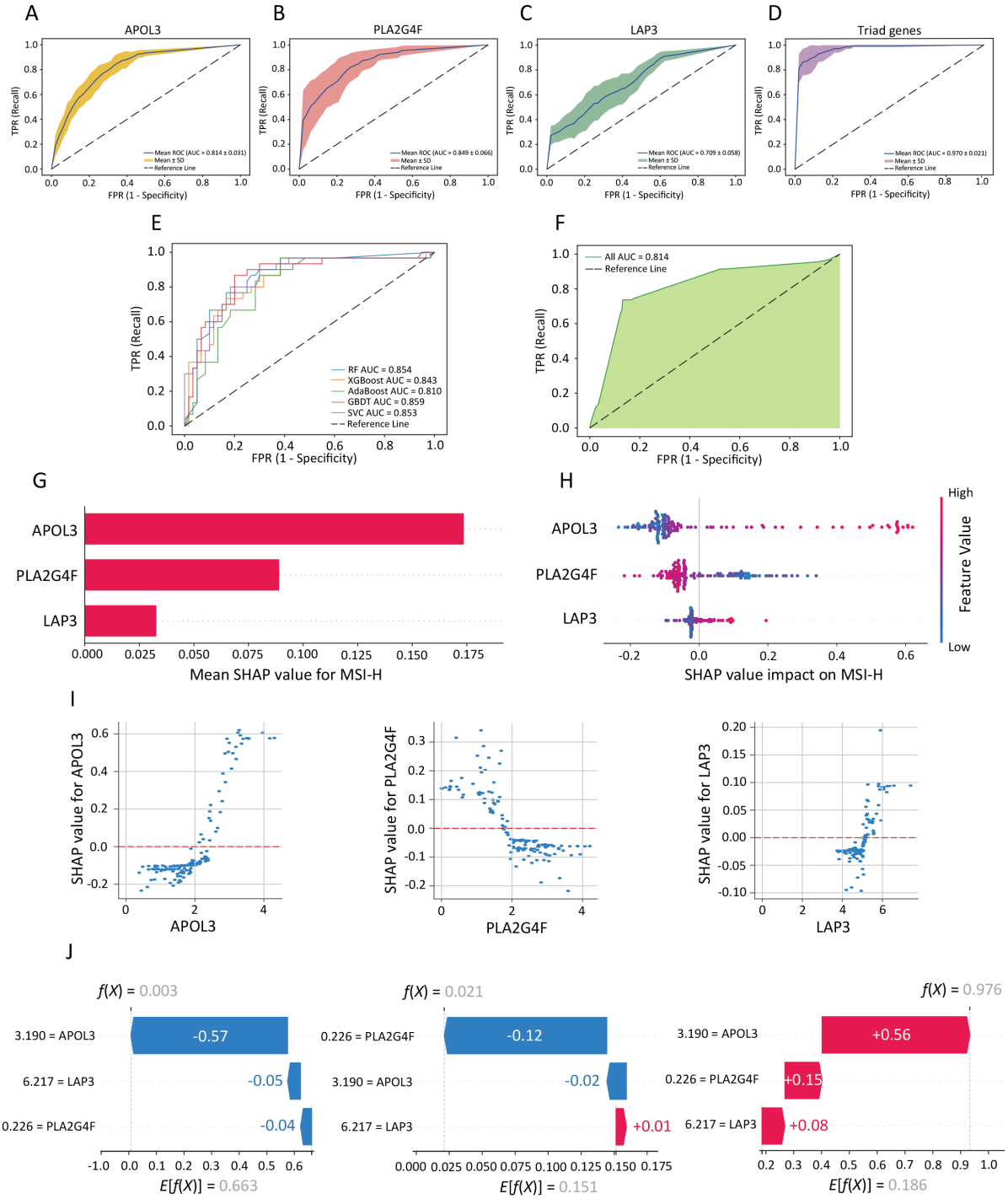


Fig. 4. Development, validation and SHAP-based interpretation of the RF model for MSI classification. (A–C) Mean ROC curves and variability across the 10-fold cross-validation analyses of models based individually on expressions of *APOL3* (A), *PLA2G4F* (B), and *LAP3* (C), respectively. (D) Mean ROC curve and variability across the 10-fold cross-validation analysis of the model based on the three-gene panel. (E) ROC curves and AUC values for five machine-learning models evaluated using the internal validation sets. (F) ROC curve and AUC value for the final RF model evaluated using the external validation set. (G) Mean SHAP-based feature-importance values for prediction of the MSI-H class. (H) SHAP summary plot showing the distribution of the SHAP values and feature values of the three genes across the samples; red and blue indicate high and low feature values, respectively. (I) SHAP dependence plots showing the relationships between individual gene-expression values and their contributions to the model output for the MSI-H class. (J) Class-specific SHAP waterfall plots for a sample with an observed MSI-H status, showing model-predicted probabilities of 0.3%, 2.1%, and 97.6% for the MSS, MSI-L, and MSI-H classes, respectively. ROC, receiver operating characteristic curve.

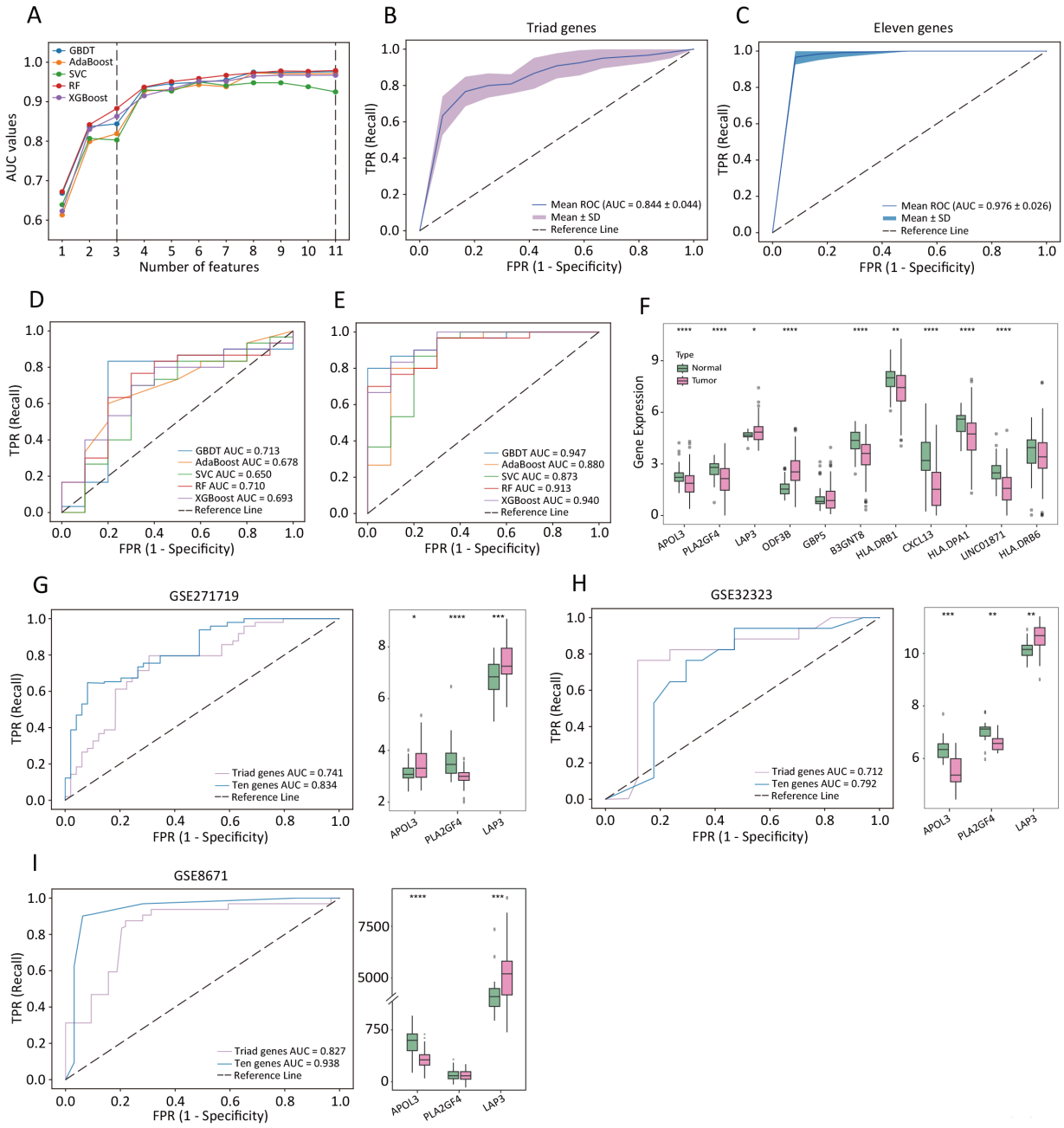


Fig. 5. Development and validation of tumor-normal classification models. (A) Incremental feature-selection and 10-fold cross-validation results obtained using the training set of the primary CRC tumor-normal dataset. (B,C) Mean ROC curves and variability across the cross-validation analyses of the GBDT models based on the three-gene panel (B) and the 11 hub genes (C), respectively. (D,E) ROC curves and AUC values for five machine-learning models based on the three-gene panel (D) and the 11 hub genes (E), respectively, evaluated using the internal validation set. (F) Expression patterns of the 11 hub genes in the primary CRC tumor-normal dataset. (G–I) ROC curves for models based on the three-gene panel and the ten available hub genes in GSE271719 (G), GSE32323 (H), and GSE8671 (I), respectively, together with the expression patterns of *APOL3*, *PLA2G4F*, and *LAP3* in tumor and normal tissues. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. GBDT, Gradient boosting decision tree.

ures show the predicted binding residues, intermolecular interactions and interaction distances. Among the screened compounds, pixantrone dimaleate and mitoxantrone, both topoisomerase II inhibitors and selected as potential drugs that targeted to LAP3, were selected for further experimental validation. However, the docking results represented computational predictions and did not by themselves demonstrate direct target engagement or therapeutic effi-

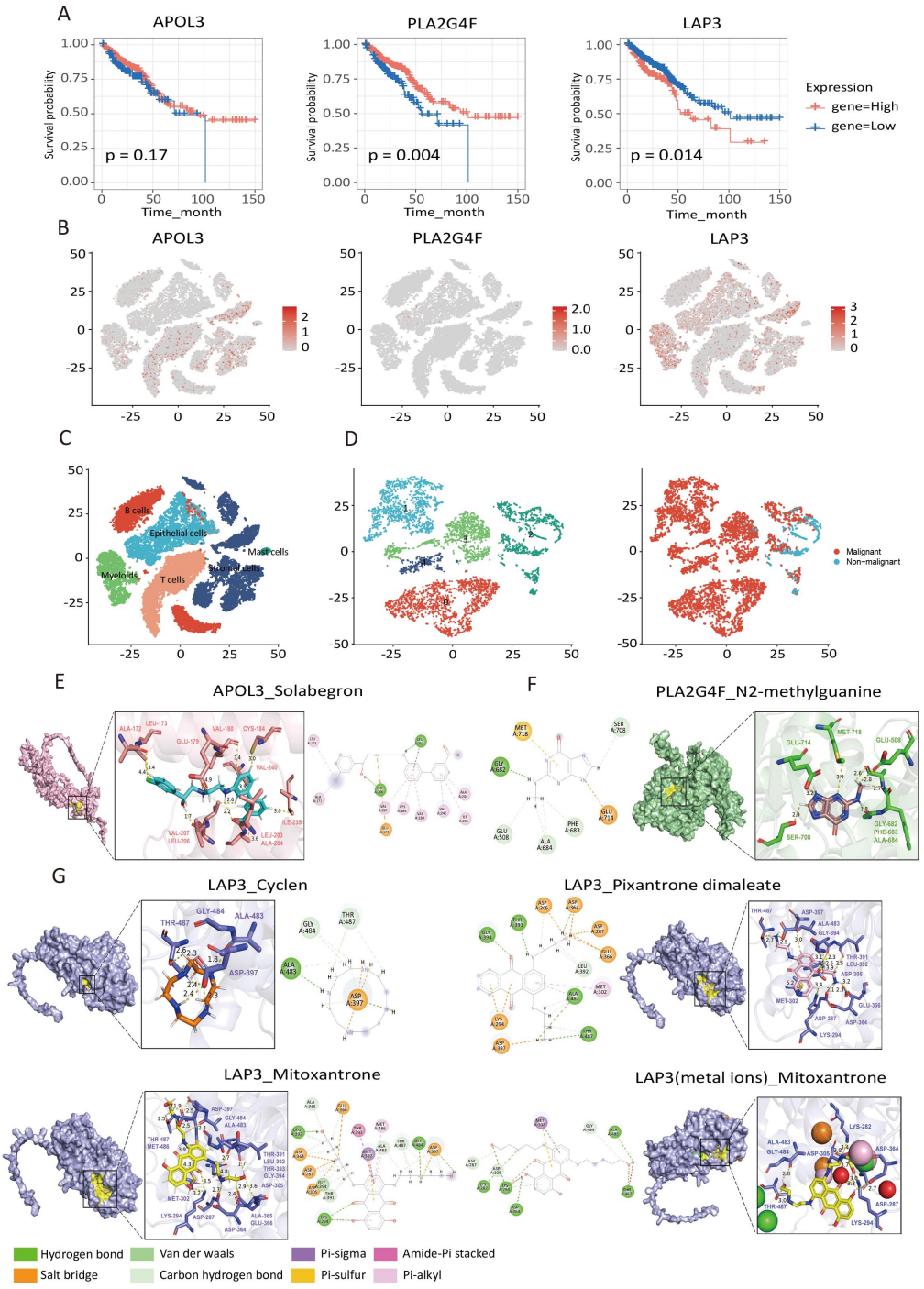


Fig. 6. Survival analysis, single cell expression patterns, malignant-cell classification and molecular docking. (A) Kaplan-Meier OS curves and log-rank tests for CRC patients stratified into high- and low-expression groups for *APOL3*, *PLA2G4F*, and *LAP3*. (B) t-SNE feature plots showing the single-cell expression patterns of *APOL3*, *PLA2G4F*, and *LAP3* across various cell types. (C) t-SNE plot of CRC single cells annotated according to their major cell types. (D) t-SNE plots showing epithelial-cell clusters (clustering into five major categories) and their predicted classification as malignant (red points) or nonmalignant (blue points) based on CRC-associated differentially expressed genes and MSI hub genes. (E) Predicted docking poses and molecular interactions of the one highest-ranked compounds, solabegron, with APOL3. (F) Predicted docking poses and molecular interactions of the one highest-ranked compounds, N2-methylguanine, with PLA2G4F. (G) Predicted docking poses and molecular interactions of the three highest-ranked compounds, cyclen, pixantrone dimaleate, and mitoxantrone, with LAP3. As a Zn^{2+}-dependent M17 metallopeptidase, the LAP3 structure with metal ions is virtually bound to mitoxantrone to mimic their interaction *in vivo*. OS, overall survival; t-SNE, t-distributed stochastic neighbor embedding.

cacy in CRC. Since LAP3 is a Zn^{2+}-dependent M17 metallopeptidase, the metal ions were retained in the LAP3 protein structure, and molecular docking was repeated for the 3 highest-ranked compounds. The revised docking pose and predicted interactions between *LAP3* and mitoxantrone are shown in Fig. 6G. Compared to the docking result of structure without metal ions, the retained metal ions changed protein conformation of LAP3, decreased number of interacted residues, and increased docking score to about 1 kcal/-mol, which showed a certain degree of differences between two docking results.

3.8 Experimental Validation of Candidate Genes and Screened Compounds

The mRNA and protein expressions of the three candidate genes were examined in human epithelial carcinoma Caco-2 and nonmalignant human colon fibroblast CCD-18Co cell lines. Caco-2 cells showed significantly lower PLA2G4F expressions and significantly higher LAP3 expressions compared with CCD-18Co cells, whereas no significant difference of high expression were observed in APOL3 (Fig. 7A,B). To select potential inhibitors targeting the screened proteins, PLA2G4F was excluded from further analysis because of its relatively lower expression in epithelial carcinoma Caco-2 cells. The expression levels of APOL3 and LAP3 were further compared between tumor and adjacent normal tissues. Tumor tissues and matched adjacent normal tissues were collected from 10 patients with CRC. H&E staining showed altered glandular architecture and nuclear pleomorphism, elongation, and pseudostratification in the CRC tissues (Fig. 7C). IHC analysis showed lower APOL3 staining and higher LAP3 staining in tumor tissues than in adjacent normal tissues (Fig. 7D,E). Consistent expression patterns were observed by qRT-PCR and WB (Fig. 7F,G). Therefore, LAP3 was identified as a potential protein target for further screening of binding inhibitors.

Through virtual screening, eight compounds selected from three compound databases, comprising a total of 10,966 compounds, were further evaluated for their drug efficacy and cytotoxic effects at the cellular level. Caco-2 and HCT116 cells were treated with each compound at 10 μM and cell viability was subsequently measured. Mitoxantrone (Mito) produced the greatest reduction in cell viability in both cell lines (Fig. 7H) and was therefore selected for further experimental evaluation. Dose-response analysis yielded IC_{50} values of approximately 1 μM in both cell lines (Fig. 7I). HCT116 cells exhibited marked morphological changes after treatment with 10 μM mitoxantrone for 24 h, such as cell shrinkage and loss of cellular integrity (Fig. 7J). Flow cytometry analysis of mitoxantrone-treated Caco-2 cells showed a total apoptotic-cell proportion of 32.52% with provided fluorescence channels and cell death percentages (Fig. 7K). In our result, *LAP3* mRNA expression and LAP3 protein expression in both cell lines were reduced after mitoxantrone treatment (Fig. 7L), which provided evidence for a potential interaction between LAP3 and mitoxantrone. Although mitoxantrone was ranked third in the molecular docking analysis with LAP3, it exhibited a stronger efficacy in inhibiting CRC cell proliferation and was therefore selected as a promising drug. The uncropped gels and blots in this study are provided in **Supplementary Fig. 10A**.

4. Discussion

In this study, an integrated bioinformatics, explainable machine learning, and experimental validation pipeline was used to identify a candidate three-gene panel associated with MSI status. The *APOL3–PLA2G4F–LAP3* panel may capture the involvement of lipid metabolism, immune regulation, and protein degradation in the MSI-associated inflammatory microenvironment compared with established mismatch repair proteins, including MLH1, MSH2, and PMS2 that are used to assess MSI status [8]. Xu et al. [44] conducted a bioinformatic analysis of the TCGA database to screen DEGs between MSI-H and MSS, as well as metastatic and non-metastatic CRC, using the 'DESeq2' library, which yielded five biologically plausible DEGs. In a separate study, Yu et al. [45] analyzed 46 patients from the GEO database (GSE25071) using the WGCNA method and identified five key DEGs and associated pathways distinguish MSI-H form MSS CRC. Lanza et al. [46] found 27 DEGs, including 8 miRNAs, based on samples from 23 MSS and 16 MSI-H patients, and validated them with qRT-PCR and Northern blotting. In the present analysis, DEGs were selected between the CRC patients with three MSI statuses. Furthermore, five machine learning models were developed and validated to evaluate the diagnostic efficacy of triad genes. Our results also displayed models from three-gene panel showed better discriminatory performance than models based on the individual genes. The use of explainable machine-learning algorithms provided information about the contributions of the individual genes to the model predictions. Further experimental validations of gene and protein expressions in CRC tumor cells and tissues as well as *in silico* and *in vitro* drug screening provided molecular and functional characterization of these genes, which might provide a research basis for investigating therapeutic strategies and target drugs in CRC. Furthermore, the analysis of the tumor immune microenvironment and immunotherapy-related features identified immune-cell populations and core immune response factors associated with the expression of the triad genes, which may contribute to the characterization of patients with different MSI statuses. Although gastric, lung, and ovarian cancers may also exhibit MSI, the present study focused on CRC. The generalizability utility and biological role of *LAP3* in other MSI-associated solid tumors will be investigated in future studies.

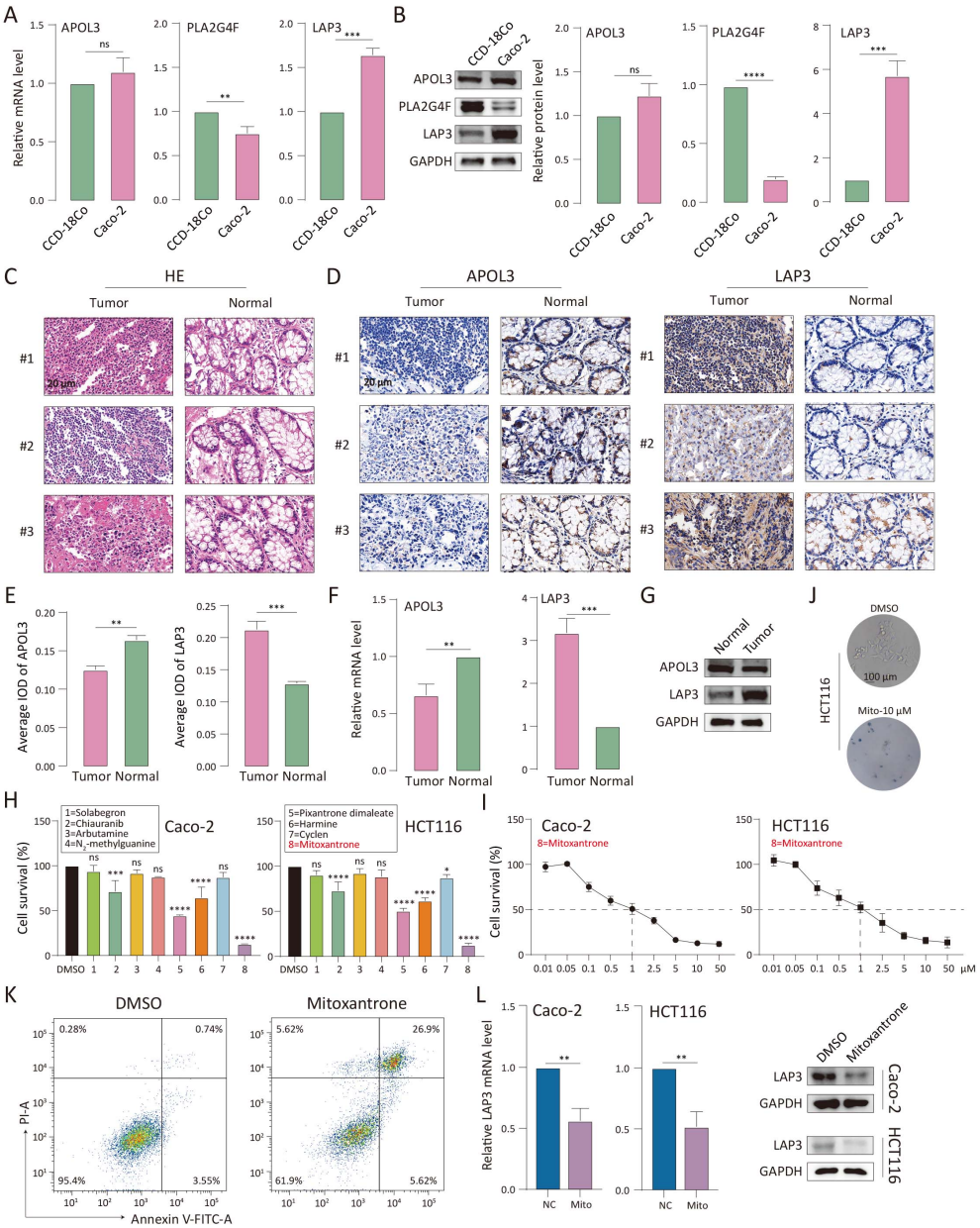


Fig. 7. Experimental validation of candidate-gene expression and cytotoxic effects of the screened compounds. (A) Relative mRNA expression of *APOL3*, *PLA2G4F*, and *LAP3* in human CRC Caco-2 cells and nonmalignant human colon fibroblast CCD-18Co cells. (B) APOL3, PLA2G4F, and LAP3 protein expression in Caco-2 and CCD-18Co cells. (C) Representative H&E staining images of paired CRC and adjacent normal tissues of 10 patients. Scale bar = 20 μm. (D) Representative IHC images showing APOL3 (left panel) and LAP3 (right panel) expression in paired CRC and adjacent normal tissues of 10 patients. Scale bar = 20 μm. (E) Quantification of APOL3 and LAP3 IHC staining based on average integrated optical density (IOD). (F) Relative *APOL3* and *LAP3* mRNA expression in 10 pairs of CRC tissues and matched adjacent normal tissues, as determined by qRT-PCR. (G) APOL3 and LAP3 protein expression in 10 pairs of CRC tissues and matched adjacent normal tissues, as determined by WB. (H) Viability of Caco-2 and HCT116 cells after treatment with each of the eight screened compounds at 10 μM. (I) Dose–response curves for Caco-2 and HCT116 cells treated with mitoxantrone. (J) Representative images of HCT116 cells treated with DMSO or 10 μM mitoxantrone for 24 hours. Scale bar = 100 μm. (K) Annexin V/PI flow-cytometric analysis of apoptosis in Caco-2 cells treated with mitoxantrone or DMSO. The total apoptotic-cell proportion after mitoxantrone treatment was detected by flow cytometry and was 32.52%, calculated as the sum of the early- and late-apoptotic populations. (L) Relative LAP3 mRNA and protein expression in Caco-2 and HCT116 cells after mitoxantrone treatment. GAPDH was used as the loading control. Results are expressed as mean ± SD of triplicates by Student's *t*-test. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, ns, not significant. H&E, hematoxylin and eosin; qRT-PCR, Quantitative real-time Polymerase chain reaction; WB, western blotting.

It has been reported that approximately 60% of patients with CRC are diagnosed at an advanced stage, that 22% present with distant metastases, and that the 5-year relative survival rate is only 14% among patients with distant metastases [47]. MSI status is associated with CRC pathogenesis, treatment response, and molecular classification. Thus, the identification of candidate MSI-associated biomarkers may support clinical diagnosis, molecular stratification and precision treatment in patients with CRC. The present study identified a three-gene panel associated with MSI status by integrating multi-omics analysis and explainable machine-learning models compared to our previous study [48]. Mitoxantrone induced apoptosis and reduced viability in CRC cells, and LAP3 expression was reduced after mitoxantrone treatment.

APOL3 has also been associated with membrane modeling (traffic, fission and fusion) and infection-induced inflammation through the control of PI(4)P synthesis [49]. *APOL3* is associated with the inflammatory response in endothelial cells [50] and antitumor immunity in CRC [51] and couples lysosomal damage to mitochondrial DNA efflux and activation of the cGAS/STING pathway, thereby inducing type I IFN production and linking innate immune responses with inflammatory signaling [52]. A multicohort study identified *APOL3*, together with other four genes, as candidate biomarkers associated with the response to immune-checkpoint inhibitors in non-small cell lung cancer [53]. In addition, *APOL3* was reported to be differentially methylated in a population with early onset CRC, providing further information about the epigenetic features of these patients [54], and *APOL3* overexpression promotes anti-tumor T cell immunity, resulting in the decrease of viability, clonogenicity, proliferation, and migration of melanoma cells in mice [55]. *PLA2G4F* has calcium-dependent phospholipase and lysophospholipase activities and hydrolyzes the ester bond of acyl or arachidonoyl groups at the sn-2 position of phospholipids. Fatty acid release mediated by *PLA2G4F* may influence cellular redox balance and tolerance to oxidative damage. Furthermore, lipid mediators including lysophosphatidic acid can activate G protein-coupled receptors and induce the expression of epithelial-mesenchymal transition biomarkers, thereby promoting the migration and invasion of tumor cells [56,57]. A study on laser-capture microdissected prostate tumor foci reported an inverse association between *PLA2G4F* expression and metastatic prostate cancer [58]. *LAP3* catalyzes the *N*-terminal removal of hydrophobic amino acids, with a preference for leucine residues, from peptides [59], being involved in protein degradation and amino acid cycle. Previous studies indicated that *LAP3* contributes to the IFN-γ-induced malignant transformation of bovine mammary epithelial cells through arginine depletion [60], malignant transformation of bovine mammary epithelial cells [60] and Epstein-Barr virus-associated carcinogenesis in gastric cancer [61]. Additionally, high *LAP3* expression increases vimentin expression and Erk1/2 phosphorylation, thereby promoting the migration and invasion of breast cancer cells [62]. Collectively, these studies associate the three candidate genes individually with lipid and amino acid metabolism, immune and inflammatory signaling, oxidative stress, epithelial-mesenchymal transition, and tumor progression of several cancer types, suggesting their potential utility in tumor treatment.

The final RF and GBDT models yielded AUC values ranging from 0.712 to 0.854 across the reported internal and external validation analyses. However, these performance estimates should be interpreted in light of the potential data leakage, cohort overlap, and overfitting discussed below. Some linear models shared are intrinsically interpretable because the fitted coefficients can provide information about the contribution of each feature. By contrast, models based on more complex machine-learning algorithms, such as SVM and tree-based ensembles, may be more difficult to interpret. The post hoc SHAP framework used Shapley values to quantify the contribution of each input feature to the model output and can provide both global and local interpretations of a fitted model [28]. For global interpretation, the absolute SHAP values can be averaged among samples to estimate the overall contribution of each feature. A larger mean absolute SHAP value indicates a greater contribution of that feature to the model output. SHAP summary and dependence plots can also show the relationships between feature values and the direction of their contributions. At the local level, SHAP values indicate how the feature values of an individual sample contribute to its predicted output [63]. These complementary interpretations may improve the transparency of complex machine-learning models and thereby enable precision treatment for MSI CRC patients based on the gene contributions revealed in SHAP results. Furthermore, model calibration, prospective validation, comparison with standard MSI testing, and clinical-utility analyses would be required before the clinical implementation of SHAP models. LIME uses a local surrogate model to approximate the behavior of a complex model around an individual prediction compared with other explainability methods [64], whereas integrated gradients attribute the output of a differentiable deep model to its input features [65]. Permutation importance is a model-agnostic method that assesses changes in model performance after the values of a feature are permuted. Compared to the above explainable method, SHAP provides additive feature attributions, feature dependences, and includes implementations applicable to several classes of models, including linear, kernel-based, tree-based, gradient-based and deep-based models [66]. The prediction direction of model ouputs can be explained in single feature level with SHAP values. Nejadshamsi et al. [67] reported that the use of TreeSHAP with XGBoost provides advantages over LIME in a COVID-19 severity-classification application. Despite its advantages, SHAP incurs high computational costs when

applied to large-scale datasets with extensive features and complex models. Furthermore, the interpretation and visualization of SHAP values remain challenging for large language models and highly correlated features.

At the scRNA-seq level, the three candidate genes and CRC-associated differentially expressed genes were used to classify epithelial cells as nonmalignant and malignant epithelial cells. The expression profiles of the three candidate genes also showed cell-type-specific distributions. These observations provided descriptive information about the expression of the three genes across the CRC cellular microenvironment but did not establish their functional involvement in malignant transformation. Furthermore, the *in silico* and *in vitro* screening procedures identified compounds for further experimental investigation. Molecular docking predicted a favorable binding energy between mitoxantrone and the modeled LAP3 structure. However, docking predictions alone do not establish direct binding, target engagement, or potential clinical efficacy. Irinotecan is a topoisomerase I inhibitor used to treat metastatic CRC by interfering with DNA replication in tumor cells [68,69]. By contrast, mitoxantrone is an established topoisomerase II inhibitor and therefore does not have the same molecular target as irinotecan, although both compounds interfere with topoisomerase-mediated DNA processing. Among the eight compounds evaluated, mitoxantrone exhibited the greatest reduction in viability in the two tumor cell lines. Although mitoxantrone is FDA-approved for other clinical indications, further mechanism, target-engagement, efficacy, as well as *in vitro* and *in vivo* safety would be required before its potential use in CRC could be considered. The present experiments indicated that mitoxantrone reduced CRC cell viability, induced apoptosis and decreased *LAP3* expression. However, the interaction mechanism underlying the pairing of *LAP3* and mitoxantrone as a gene-therapeutic combination, as well as direct evidence, warrant further investigation through targeted rescue experiments and application of the cellular thermal shift assay (CETSA) combined with drug affinity-responsive target stability (DARTS) assay.

5. Limitation

Although the present study produced several *in silico* and *in vitro* findings, several limitations remain. First, the roles of the three candidate genes in CRC require further investigation using additional omics data and functional experiments. Their relevance to other tumor types also remains to be established. Second, the inclusion of more diverse samples and datasets would improve assessment of the generalizability and potential clinical applicability of the diagnostic models. Third, although LAP3 upregulation and cytotoxicity of mitoxantrone were evaluated at the cellular and tissue levels, *in vivo* experiments are required for further validation. Moreover, whether mitoxantrone binds directly to LAP3 to inhibit CRC cell proliferation and interacts with the predicted binding residues remains to be determined. Finally, the identified candidate genes in bioinformatic-based process and feature space evaluated as internal validation were selected using the same 147-patient cohort that was subsequently used for internal validation, resulting in potential feature-selection leakage. Although stratified data splitting and application of SMOTE only to the training set prevented synthetic samples from entering the internal validation set, these procedures did not eliminate leakage arising from feature selection using the complete cohort. Consequently, the internally validated performance estimates may be optimistically biased. Thus, hub gene screening and all other feature-selection procedures should be restricted to the training data and repeated within each cross-validation or resampling fold.

6. Conclusion

This study used an integrated multi-omics and machine learning pipeline to identify a candidate three-gene panel associated with MSI status in CRC. *LAP3* was upregulated in CRC cells and tissues, and its expression was associated with immune-cell infiltration, immunotherapy-related features, and patient prognosis. Molecular docking predicted a possible interaction between the FDA-approved drug mitoxantrone and LAP3, whereas *in vitro* experiments showed that mitoxantrone reduced CRC cell proliferation, induced apoptosis and might contribute to decreased *LAP3* expression. However, these findings did not demonstrate that mitoxantrone bound directly to LAP3 or that its cytotoxic effects were directly mediated through LAP3. Further independent validation, mechanistic investigation, and direct target-engagement experiments are required to determine the potential diagnostic and therapeutic relevance of these findings.

Abbreviations

ACC, Accuracy; AdaBoost, Adaptive boosting; APOL3, Apolipoprotein L3; AUC, Area under the curve; BP, Biological process; CC, Cell components; CIMP, CpG island methylator phenotype; CIN, Chromosomal instability; CMS, Consensus molecular subtypes; CRC, Colorectal adenocarcinoma; DEGs, Differentially expressed genes; FPKM, Fragments per kilobase of exon model per million mapped fragments; GBDT, Gradient boosting decision tree; GEO, Gene expression Omnibus; GO, Gene ontology; HE, Hematoxylin eosin; IHC, Immunohistochemistry; KEGG, Kyoto encyclopedia of genes and genomics; LAP3, Leucine aminopeptidase 3; LASSO, Least absolute shrinkage and selection operator; MCC, Matthew's correlation coefficient; MF, Molecular functions; ML, Machine learning; MSI, Microsatellite instability; PCA, Principal component analysis; PLA2G4F, Phospholipase A2 group IVF; qRT-PCR, Quantitative real-time Polymerase chain reaction; RF, Random Forest; ROC, Receiver operating characteristic curve; scRNA-seq, Single cell RNA sequencing; SHAP,

Shapley additive explanation; SMOTE, Synthetic minority over-sampling technique; SVM, Support vector machine; TCGA, The cancer genome atlas; TMB, Tumor mutation burden; t-SNE, t-distributed stochastic neighbor embedding; UCSC, University of California, Santa Cruz; WB, Western blotting; WGCNA, Weighted gene co-expression network analysis; XGBoost, Extreme gradient boosting.

Availability of Data and Materials

The data that support the findings of this study are publicly available from TCGA (https://www.cancer.gov/ccg/research/genome-sequencing/tcga), UCSC Xena platform (https://xenabrowser.net/datapages/), and GEO (https://www.ncbi.nlm.nih.gov/geo/). Other data that support the findings of this study are available from the corresponding authors upon reasonable request.

Author Contributions

Conceptualization: SPW, YQC, and RW; Methodology: SPW, JN, and RW; Investigation: SPW, CW, and YQC; Visualization: SPW, JQZ, and RW; Data curation: SPW, JQZ, and JN; Software: SPW, JQZ, and RW; Supervision: SPW, JN, CW, and YQC; Project administration: CW, YQC, and RW; Funding acquisition: JN and RW; Writing - original draft: SPW, JQZ, and RW; writing - review & editing: YQC and RW. All authors have read and approved the final version of the manuscript. All authors contributed to the editorial changes in the manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

This study was approved by the Ethics Committee of Affiliated Hospital of Jiangnan University (LS2021028) in accordance with the Declaration of Helsinki. Informed consents were received from all participants.

Acknowledgment

Sincerely thanks to Prof. Ninghan Feng for his assistance in conceptual inspiration and expert guidance in this study. Thanks to associate Prof. Xiaoying Wang for her generous grant of colorectal tumor cells.

Funding

This study was supported by the Jiangsu Funding Program for Excellent Postdoctoral Talent (No. 2024ZB069).

Conflicts of Interest

The authors declare no conflicts of interest.

Declaration of AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, DeepSeek-V4 was used to assist with the Abstract refinement and structure editing. The AI-generated content was thoroughly reviewed and edited by authors and take full responsibility for the integrity of the publication.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.31083/FBL55483.

References

[1] Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. Translational Oncology. 2021; 14: 101174. https://doi.org/10.1016/j.tranon.2021.101174

[2] Kim SE, Paik HY, Yoon H, Lee JE, Kim N, Sung MK. Sex- and gender-specific disparities in colorectal cancer risk. World Journal of Gastroenterology. 2015; 21: 5167–5175. https://doi.org/10.3748/wjg.v21.i17.5167

[3] Li L, Sun K, Li X, Yao Y, Li H, Wang S, et al. Global landscape and temporal trends in lifetime risk of colorectal cancer in 185 countries: a population-based study. Cancer Biology & Medicine. 2026; 23: 1004–1015. https://doi.org/10.20892/j.issn.2095-3941.2025.0851

[4] Kloor M, von Knebel Doeberitz M. The Immune Biology of Microsatellite-Unstable Cancer. Trends in Cancer. 2016; 2: 121–133. https://doi.org/10.1016/j.trecan.2016.02.004

[5] Li GM. Mechanisms and functions of DNA mismatch repair. Cell Research. 2008; 18: 85–98. https://doi.org/10.1038/cr.2007.115

[6] Moufarrij S, Gazzo A, Rana S, Selenica P, Abu-Rustum NR, Ellenson LH, et al. Concurrent POLE hotspot mutations and mismatch repair deficiency/microsatellite instability in endometrial cancer: A challenge in molecular classification. Gynecologic Oncology. 2024; 191: 1–9. https://doi.org/10.1016/j.ygyno.2024.09.008

[7] Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nature Medicine. 2015; 21: 1350–1356. https://doi.org/10.1038/nm.3967

[8] Kawakami H, Zaanan A, Sinicrope FA. Microsatellite instability testing and its role in the management of colorectal cancer. Current Treatment Options in Oncology. 2015; 16: 30. https://doi.org/10.1007/s11864-015-0348-2

[9] Sinicrope FA. The role of microsatellite instability testing in management of colorectal cancer. Clinical Advances in Hematology & Oncology: H&O. 2016; 14: 476–479.

[10] Pino MS, Chung DC. The chromosomal instability pathway in colon cancer. Gastroenterology. 2010; 138: 2059–2072. https://doi.org/10.1053/j.gastro.2009.12.065

[11] Poulogiannis G, Ichimura K, Hamoudi RA, Luo F, Leung SY, Yuen ST, et al. Prognostic relevance of DNA copy number changes in colorectal cancer. The Journal of Pathology. 2010; 220: 338–347. https://doi.org/10.1002/path.2640

[12] Dunne PD, Arends MJ. Molecular pathological classification of colorectal cancer-an update. Virchows Archiv: an International Journal of Pathology. 2024; 484: 273–285. https://doi.org/10.1007/s00428-024-03746-3

[13] Gay LJ, Arends MJ, Mitrou PN, Bowman R, Ibrahim AE, Happerfield L, et al. MLH1 promoter methylation, diet, and lifestyle factors in mismatch repair deficient colorectal cancer patients from EPIC-Norfolk. Nutrition and Cancer. 2011; 63: 1000–1010. https://doi.org/10.1080/01635581.2011.596987

[14] Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer. Proceedings of the National Academy of Sciences of the

United States of America. 1999; 96: 8681–8686. https://doi.org/10.1073/pnas.96.15.8681

[15] Rhee YY, Kim KJ, Kang GH. CpG Island Methylator Phenotype-High Colorectal Cancers and Their Prognostic Implications and Relationships with the Serrated Neoplasia Pathway. Gut and Liver. 2017; 11: 38–46. https://doi.org/10.5009/gnl15535

[16] Wang J, He X, Mi Y, Chen YQ, Li J, Wang R. PSAT1 enhances the efficacy of the prognosis estimation nomogram model in stage-based clear cell renal cell carcinoma. BMC Cancer. 2024; 24: 463. https://doi.org/10.1186/s12885-024-12183-z

[17] Liu H, Zhang C, Peng S, Yin Y, Xu Y, Wu S, et al. Prognostic models of immune-related cell death and stress unveil mechanisms driving macrophage phenotypic evolution in colorectal cancer. Journal of Translational Medicine. 2025; 23: 127. https://doi.org/10.1186/s12967-025-06143-9

[18] Fujimoto D, Shibaki R, Kimura K, Haratani K, Tamiya M, Kijima T, et al. Identification of key gene signatures for predicting chemo-immunotherapy efficacy in extensive-stage small-cell lung cancer using machine learning. Lung Cancer (Amsterdam, Netherlands). 2025; 199: 108079. https://doi.org/10.1016/j.lungcan.2024.108079

[19] Cai ZR, Zheng YQ, Hu Y, Ma MY, Wu YJ, Liu J, et al. Construction of exosome non-coding RNA feature for non-invasive, early detection of gastric cancer patients by machine learning: a multi-cohort study. Gut. 2025; 74: 884–893. https://doi.org/10.1136/gutjnl-2024-333522

[20] Deng H, Wang Z, Zhang Q. Identification of Diagnostic Biomarkers Associated With M1 Macrophage in Lung Squamous Cell Carcinoma via Machine Learning. Frontiers in Bioscience (Landmark Edition). 2025; 30: 44661. https://doi.org/10.31083/FBL44661

[21] Zeng WJ, Zhang L, Cao H, Li D, Zhang H, Xia Z, et al. A novel inflammation-related lncRNAs prognostic signature identifies LINC00346 in promoting proliferation, migration, and immune infiltration of glioma. Frontiers in Immunology. 2022; 13: 810572. https://doi.org/10.3389/fimmu.2022.810572

[22] Jiang C, Qian C, Jiang Q, Zhou H, Jiang Z, Teng Y, et al. Virtual biopsy for non-invasive identification of follicular lymphoma histologic transformation using radiomics-based imaging biomarker from PET/CT. BMC Medicine. 2025; 23: 49. https://doi.org/10.1186/s12916-025-03893-7

[23] Captier N, Lerousseau M, Orlhac F, Hovhannisyan-Baghdasarian N, Luporsi M, Woff E, et al. Integration of clinical, pathological, radiological, and transcriptomic data improves prediction for first-line immunotherapy outcome in metastatic non-small cell lung cancer. Nature Communications. 2025; 16: 614. https://doi.org/10.1038/s41467-025-55847-5

[24] Wang R, Liu Q, You W, Wang H, Chen Y. A transformer-based deep learning survival prediction model and an explainable XGBoost anti-PD-1/PD-L1 outcome prediction model based on the cGAS-STING-centered pathways in hepatocellular carcinoma. Briefings in Bioinformatics. 2024; 26: bbae686. https://doi.org/10.1093/bib/bbae686

[25] Chen D, Liu P, Lin J, Zang L, Liu Y, Zhai S, et al. A Distinguished Roadmap of Fibroblast Senescence in Predicting Immunotherapy Response and Prognosis Across Human Cancers. Advanced Science (Weinheim, Baden-Wurttemberg, Germany). 2025; 12: e2406624. https://doi.org/10.1002/advs.202406624

[26] Liu X, Wei N, Chen H. Development of a Novel Prognostic Panel for Colorectal Cancer Based on Cancer Functional Status, and Validation of STC2 as a Promising Biomarker. Frontiers in Bioscience (Landmark Edition). 2024; 29: 245. https://doi.org/10.31083/j.fbl2907245

[27] Azodi CB, Tang J, Shiu SH. Opening the Black Box: Interpretable Machine Learning for Geneticists. Trends in Genetics: TIG. 2020; 36: 442–455. https://doi.org/10.1016/j.tig.2020.03.005

[28] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence. 2020; 2: 56–67. https://doi.org/10.1038/s42256-019-0138-9

[29] Kim TM, Laird PW, Park PJ. The landscape of microsatellite instability in colorectal and endometrial cancer genomes. Cell. 2013; 155: 858–868. https://doi.org/10.1016/j.cell.2013.10.015

[30] Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012; 487: 330–337. https://doi.org/10.1038/nature11252

[31] Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nature Biotechnology. 2020; 38: 675–678. https://doi.org/10.1038/s41587-020-0546-8

[32] Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Research. 2013; 41: D991–D995. https://doi.org/10.1093/nar/gks1193

[33] Belder N, Charyyeva S, Abaci Oruc EE, Kawalya H, Sahar NE, Omidvar N, et al. Transformative insights from transcriptome analysis of colorectal cancer patient tissues: identification of four key prognostic genes. PeerJ. 2025; 13: e19852. https://doi.org/10.7717/peerj.19852

[34] Khamas A, Ishikawa T, Shimokawa K, Mogushi K, Iida S, Ishiguro M, et al. Screening for epigenetically masked genes in colorectal cancer Using 5-Aza-2'-deoxycytidine, microarray and gene expression profile. Cancer Genomics & Proteomics. 2012; 9: 67–75.

[35] Sabates-Bellver J, Van der Flier LG, de Palo M, Cattaneo E, Maake C, Rehrauer H, et al. Transcriptome profile of human colorectal adenomas. Molecular Cancer Research: MCR. 2007; 5: 1263–1275. https://doi.org/10.1158/1541-7786.MCR-07-0267

[36] Lee HO, Hong Y, Etlioglu HE, Cho YB, Pomella V, Van den Bosch B, et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. Nature Genetics. 2020; 52: 594–603. https://doi.org/10.1038/s41588-020-0636-z

[37] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9: 559. https://doi.org/10.1186/1471-2105-9-559

[38] Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nature Methods. 2015; 12: 453–457. https://doi.org/10.1038/nmeth.3337

[39] Yang JM, Zhang N, Luo T, Yang M, Shen WK, Tan ZL, et al. TCellSI: A novel method for T cell state assessment and its applications in immune environment prediction. IMeta. 2024; 3: e231. https://doi.org/10.1002/imt2.231

[40] Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013; 14: 7. https://doi.org/10.1186/1471-2105-14-7

[41] Liao C, Wang X. TCGAplot: an R package for integrative pan-cancer analysis and visualization of TCGA multi-omics data. BMC Bioinformatics. 2023; 24: 483. https://doi.org/10.1186/s12859-023-05615-3

[42] Wang S, Wang D, Li J, Huang T, Cai YD. Identification and analysis of the cleavage site in a signal peptide using SMOTE, dagging, and feature selection methods. Molecular Omics. 2018; 14: 64–73. https://doi.org/10.1039/c7mo00030h

[43] Abudurexiti N, Liu B, Wang S, Dong Q, Batuer M, Liu Z, et al. Machine Learning-Based Prediction of Post-Operative Systemic Inflammatory Response Syndrome Following Pediatric Percutaneous Nephrolithotripsy. Journal of Inflammation Research. 2025; 18: 7067–7081. https://doi.org/10.2147/JIR.S518631

[44] Xu Y, Wang X, Chu Y, Li J, Wang W, Hu X, et al. Analysis of transcript-wide profile regulated by microsatellite instability of colorectal cancer. Annals of Translational Medicine. 2022; 10: 169. https://doi.org/10.21037/atm-21-6126
[45] Yu C, Hong H, Zhang S, Zong Y, Ma J, Lu A, et al. Identification of key genes and pathways involved in microsatellite instability in colorectal cancer. Molecular Medicine Reports. 2019; 19: 2065–2076. https://doi.org/10.3892/mmr.2019.9849
[46] Lanza G, Ferracin M, Gafà R, Veronese A, Spizzo R, Pichiorri F, et al. mRNA/microRNA gene expression profile in microsatellite unstable colorectal cancer. Molecular Cancer. 2007; 6: 54. https://doi.org/10.1186/1476-4598-6-54
[47] Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. CA: a Cancer Journal for Clinicians. 2023; 73: 233–254. https://doi.org/10.3322/caac.21772
[48] Chen L, Pan X, Hu X, Zhang YH, Wang S, Huang T, et al. Gene expression differences among different MSI statuses in colorectal cancer. International Journal of Cancer. 2018; 143: 1731–1740. https://doi.org/10.1002/ijc.31554
[49] Pays E. Apolipoprotein-L Functions in Membrane Remodeling. Cells. 2024; 13: 2115. https://doi.org/10.3390/cells13242115
[50] Horrevoets AJ, Fontijn RD, van Zonneveld AJ, de Vries CJ, ten Cate JW, Pannekoek H. Vascular endothelial genes that are responsive to tumor necrosis factor-alpha in vitro are expressed in atherosclerotic lesions, including inhibitor of apoptosis protein-1, stannin, and two novel genes. Blood. 1999; 93: 3418–3431.
[51] Lv Y, Tang W, Xu Y, Chang W, Zhang Z, Lin Q, et al. Apolipoprotein L3 enhances CD8+ T cell antitumor immunity of colorectal cancer by promoting LDHA-mediated ferroptosis. International Journal of Biological Sciences. 2023; 19: 1284–1298. https://doi.org/10.7150/ijbs.74985
[52] Ritacco DA, Shahnawaz H, Oduguwa A, Hawk J, Vizcaino B, Farber DL, et al. The human antibacterial factor APOL3 couples lysosomal damage to mitochondrial DNA efflux and type I IFN induction. Molecular Cell. 2026; 86: 1116–1133.e8. https://doi.org/10.1016/j.molcel.2026.01.029
[53] Chen S, Liu F, Fu Y, Deng CK, Borgia JA, Ayman AG, et al. A prospective multi-cohort study identifies and validates a 5-gene peripheral blood signature predictive of immunotherapy response in non-small cell lung cancer. Molecular Cancer. 2024; 23: 247. https://doi.org/10.1186/s12943-024-02160-2
[54] Li JS, Riggins K, Yang L, Chen C, Castro P, Alfarkh W, et al. DNA methylation profiling at base-pair resolution reveals unique epigenetic features of early-onset colorectal cancer in underrepresented populations. Clinical Epigenetics. 2025; 17: 11. https://doi.org/10.1186/s13148-025-01817-z
[55] Liao Y, Yao W, Liu Z, Peng Q, Zhou X, Xie M, et al. Antigen presentation-related protein APOL3 promotes anti-tumor T cell immunity and suppresses melanoma cell growth in mice. Autoimmunity. 2025; 58: 2531833. https://doi.org/10.1080/08916934.2025.2531833
[56] Nadhan R, Nath K, Basu S, Isidoro C, Song YS, Dhanasekaran DN. Decoding lysophosphatidic acid signaling in physiology and disease: mapping the multimodal and multinodal signaling networks. Signal Transduction and Targeted Therapy. 2025; 10: 337. https://doi.org/10.1038/s41392-025-02423-4
[57] Mäusbacher N, Schreiber TB, Daub H. Glycoprotein capture and quantitative phosphoproteomics indicate coordinated regulation of cell migration upon lysophosphatidic acid stimulation. Molecular & Cellular Proteomics: MCP. 2010; 9: 2337–2353. https://doi.org/10.1074/mcp.M110.000737
[58] Whitlock NC, White ME, Capaldo BJ, Ku AT, Agarwal S, Fang L, et al. Progression of prostate cancer reprograms MYC-mediated lipid metabolism via lysine methyltransferase 2A. Discover Oncology. 2022; 13: 97. https://doi.org/10.1007/s12672-022-00565-3
[59] Matsushima M, Takahashi T, Ichinose M, Miki K, Kurokawa K, Takahashi K. Structural and immunological evidence for the identity of prolyl aminopeptidase with leucyl aminopeptidase. Biochemical and Biophysical Research Communications. 1991; 178: 1459–1464. https://doi.org/10.1016/0006-291x(91)91057-j
[60] Li L, Li F, Hu X, Wu Z, Ren W, Wang T, et al. LAP3 contributes to IFN-γ-induced arginine depletion and malignant transformation of bovine mammary epithelial cells. BMC Cancer. 2022; 22: 864. https://doi.org/10.1186/s12885-022-09963-w
[61] Zhou H, Jing S, Liu Y, Wang X, Duan X, Xiong W, et al. Identifying the key genes of Epstein-Barr virus-regulated tumour immune microenvironment of gastric carcinomas. Cell Proliferation. 2023; 56: e13373. https://doi.org/10.1111/cpr.13373
[62] Wang X, Ji S, Ma Y, Xing X, Zhou Y, Xu X, et al. Vimentin plays an important role in the promotion of breast cancer cell migration and invasion by leucine aminopeptidase 3. Cytotechnology. 2020; 72: 639–647. https://doi.org/10.1007/s10616-020-00402-x
[63] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems. 2017; 30.
[64] Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). San Francisco California USA. ACM: New York, NY, USA. 2016. https://doi.org/10.1145/2939672.2939778
[65] Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. In International conference on machine learning (pp. 3319–3328). PMLR. 2017.
[66] Rodríguez-Pérez R, Bajorath J. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. Journal of Computer-aided Molecular Design. 2020; 34: 1013–1026. https://doi.org/10.1007/s10822-020-00314-0
[67] Nejadshamsi S, Chu C, McGilton KS, Li X, Ronquillo C, Abbasgholizadeh-Rahimi S. Evaluation and improvement of algorithmic fairness for COVID-19 severity classification using Explainable Artificial Intelligence-based bias mitigation. JAMIA Open. 2025; 9: ooaf171. https://doi.org/10.1093/jamiaopen/ooaf171
[68] Berretta M, Lleshi A, Zanet E, Bearz A, Simonelli C, Fisichella R, et al. Bevacizumab plus irinotecan-, fluorouracil-, and leucovorin-based chemotherapy with concomitant HAART in an HIV-positive patient with metastatic colorectal cancer. Onkologie. 2008; 31: 394–397. https://doi.org/10.1159/000132360
[69] Hurwitz H, Fehrenbacher L, Novotny W, Cartwright T, Hainsworth J, Heim W, et al. Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. The New England Journal of Medicine. 2004; 350: 2335–2342. https://doi.org/10.1056/NEJMoa032691