

Identifying clinically relevant transcriptional signatures and methylation profiles in the course, treatment and outcome of colorectal cancer

Edited by

Zora Lasabova and Pooneh Mokarram

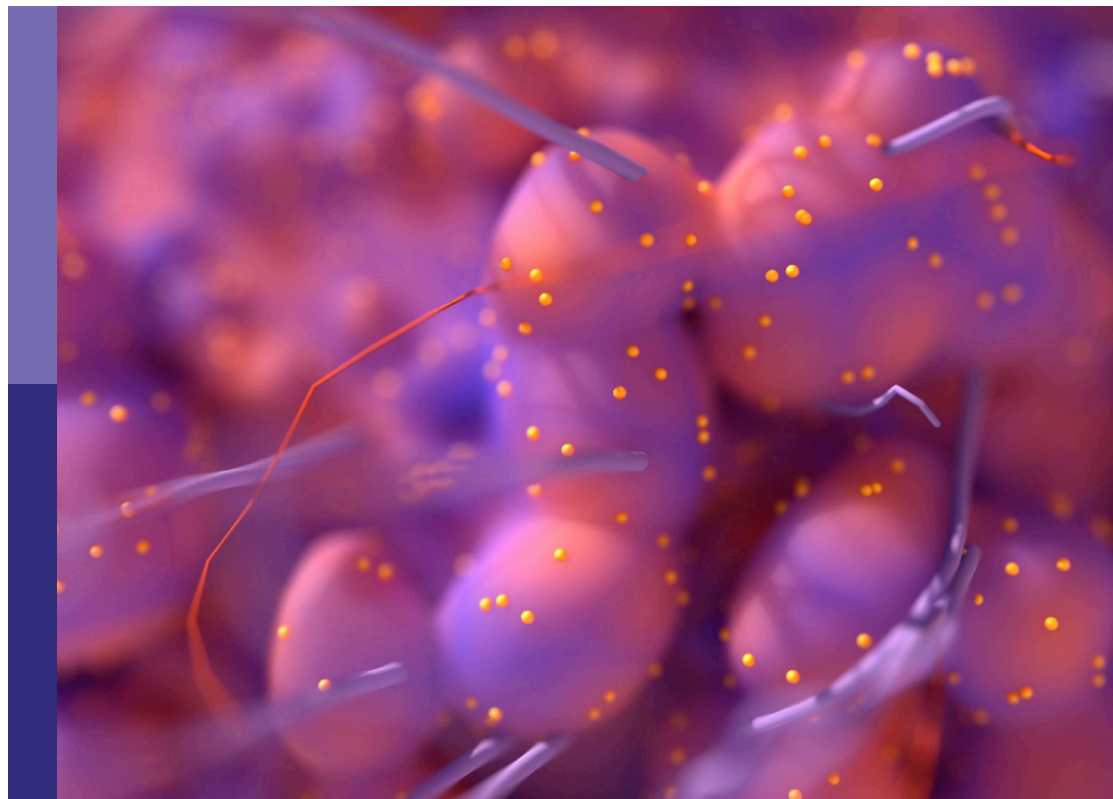
Coordinated by

Martina Poturnajova and Mozhdeh Zamani

Published in

Frontiers in Oncology

Frontiers in Immunology



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-4695-6
DOI 10.3389/978-2-8325-4695-6

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Identifying clinically relevant transcriptional signatures and methylation profiles in the course, treatment and outcome of colorectal cancer

Topic editors

Zora Lasabova — Comenius University, Slovakia

Pooneh Mokarram — Shiraz University of Medical Sciences, Iran

Topic Coordinators

Martina Poturnajova — Biomedical Research Center, Slovak Academy of Sciences, Slovakia

Mozhdeh Zamani — Shiraz University of Medical Sciences, Iran

Citation

Lasabova, Z., Mokarram, P., Poturnajova, M., Zamani, M., eds. (2024). *Identifying clinically relevant transcriptional signatures and methylation profiles in the course, treatment and outcome of colorectal cancer*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-4695-6

Table of contents

- 05 Editorial: Identifying clinically relevant transcriptional signatures and methylation profiles in the course, treatment and outcome of colorectal cancer
Pooneh Mokarram and Mozhddeh Zamani
- 08 Immune-infiltrating signature-based classification reveals CD103⁺CD39⁺ T cells associate with colorectal cancer prognosis and response to immunotherapy
Yang Luo, Yunfeng Zong, Hanju Hua, Meiting Gong, Qiao Peng, Chen Li, Dante Neculai and Xun Zeng
- 25 Clinicopathological characteristics of high microsatellite instability/mismatch repair-deficient colorectal cancer: A narrative review
Wei-Jian Mei, Mi Mi, Jing Qian, Nan Xiao, Ying Yuan and Pei-Rong Ding
- 38 Developing an m5C regulator-mediated RNA methylation modification signature to predict prognosis and immunotherapy efficacy in rectal cancer
Rixin Zhang, Wenqiang Gan, Jinbao Zong, Yufang Hou, Mingxuan Zhou, Zheng Yan, Tiegang Li, Silin Lv, Zifan Zeng, Weiqi Wang, Fang Zhang and Min Yang
- 58 Expression and prognostic value of PRDX family in colon adenocarcinoma by integrating comprehensive analysis and *in vitro* and *in vivo* validation
He Zhou, Lifa Li, Jia Chen, Songlin Hou, Tong Zhou and Yongfu Xiong
- 74 Differences in genome, transcriptome, miRNAome, and methylome in synchronous and metachronous liver metastasis of colorectal cancer
Josef Horak, Ondrej Kubecek, Anna Siskova, Katerina Honkova, Irena Chvojikova, Marketa Krupova, Monika Manethova, Sona Vodenkova, Sandra García-Mulero, Stanislav John, Filip Cecka, Ludmila Vodickova, Jiri Petera, Stanislav Filip and Veronika Vymetalkova
- 90 Expression of *OCT4* isoforms is reduced in primary colorectal cancer
Eva Turyova, Peter Mikolajcik, Marian Grendar, Eva Kudelova, Veronika Holubekova, Michal Kalman, Juraj Marcinek, Matej Hrniciar, Michal Kovac, Juraj Miklusica, Ludovit Laca and Zora Lasabova
- 101 Hypermethylated *GRIA4*, a potential biomarker for an early non-invasive detection of metastasis of clinically known colorectal cancer
Eva Lukacova, Tatiana Burjanivova, Petar Podlesniy, Marian Grendar, Eva Turyova, Ivana Kasubova, Ludovit Laca, Peter Mikolajcik, Eva Kudelova, Andrea Vanochova, Juraj Miklusica, Sandra Mersakova and Zora Lasabova

- 112 **Combining methylated SDC2 test in stool DNA, fecal immunochemical test, and tumor markers improves early detection of colorectal neoplasms**
Tao Zeng, Zhongchao Huang, Xufa Yu, Li Zheng, Tao Liu, Boyu Tian, Siyu Xiao and Jiahui Huang
- 123 **Molecular characteristics, clinical significance, and immune landscape of extracellular matrix remodeling-associated genes in colorectal cancer**
Wenlong Chen, Yiwen Wang, Haitao Gu, Yi Zhang, Cong Chen, Tingting Yu and Tao Chen
- 138 **Differential gene expression of immunity and inflammation genes in colorectal cancer using targeted RNA sequencing**
Veronika Holubekova, Dusan Loderer, Marian Grendar, Peter Mikolajcik, Zuzana Kolkova, Eva Turyova, Eva Kudelova, Michal Kalman, Juraj Marcinek, Juraj Miklusica, Ludovit Laca and Zora Lasabova



OPEN ACCESS

EDITED AND REVIEWED BY
Yun Dai,
Peking University, China

*CORRESPONDENCE
Pooneh Mokarram
✉ mokaram2@gmail.com

RECEIVED 16 December 2023

ACCEPTED 01 March 2024

PUBLISHED 19 March 2024

CITATION

Mokarram P and Zamani M (2024) Editorial:
Identifying clinically relevant transcriptional
signatures and methylation profiles in the
course, treatment and outcome of
colorectal cancer.
Front. Oncol. 14:1356765.
doi: 10.3389/fonc.2024.1356765

COPYRIGHT

© 2024 Mokarram and Zamani. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License](#)
(CC BY). 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.

Editorial: Identifying clinically relevant transcriptional signatures and methylation profiles in the course, treatment and outcome of colorectal cancer

Pooneh Mokarram* and Mozhdeh Zamani

Autophagy Research Center, Department of Biochemistry, Shiraz University of Medical Sciences, Shiraz, Iran

KEYWORDS

CRC, epigenetic, genetic, immunotherapy, biomarkers

Editorial on the Research Topic

Identifying clinically relevant transcriptional signatures and methylation profiles in the course, treatment and outcome of colorectal cancer

Colorectal Cancer (CRC) is currently one of the most aggressive neoplasms with increased incidence and mortality rates in the world. Conventionally, the key therapeutic policies in CRC include surgery, chemotherapy, radiotherapy, and targeted therapy. Among these therapeutic approaches, surgery is valid for patients with lesions in the early stages of diagnosis, and 20% of CRC patients miss surgery due to metastasis. Chemotherapy is also the main treatment option for CRC patients with metastatic states. Recently, targeted therapy has improved the process of colon cancer treatment, which has resulted in remarkable improvements in progression-free survival in metastatic CRC. Finding specific biomarkers for CRC screening and treatment is the prospect of scientific biomarker development in colorectal cancer-targeted therapy. In this regard, two related topics will be detected in this volume of Frontiers in oncology; the first is the role of triggering immune system by immunotherapy to prevent or treat cancer; and the second is novel genetic and epigenetic modifications as promising biomarkers for screening, prediction of prognosis, drug response, and metastasis in CRC. Aberrant DNA methylation and dysregulation of other epigenetic markers such as histone modification, MicroRNAs (miRNAs), Long non-coding RNAs (lncRNAs), etc. have been widely recognized as the most common characteristics of epigenetic alterations in CRC, having a role in CRC patients prognosis and treatment in the early stages (Figure 1). The most important and promising prognostic factor for CRC is the stage, the lower the stage of disease finding, the better the outcome. Emerging genetic and epigenetic biomarkers within the primary or metastatic tumor that are associated with prediction, recurrence, or favorable response have served as CRC biomarkers. However, in clinical decision making, heterogeneity is a vital factor affecting biomarker identification implementation

Epigenetic tumor biomarkers

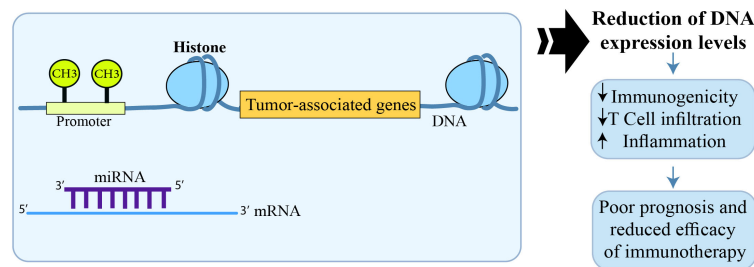


FIGURE 1

The effect of epigenetic modifications on CRC prognosis and response to immunotherapy.

and strategies. Molecular alterations cause CRC, which is a heterogeneous disease, such as gene methylation, and driver mutations, and is affected by the TME (Tumor Microenvironment), which causes infiltrates of immune cells through pro- and anti-tumor effects. Therefore, to develop effective immunotherapeutic strategies for CRC, it is important to study the interactions between the TME and the immune system (IS).

Consistent changes in the expression of cell surface receptors, the presence of pro-inflammatory cytokines, chronic inflammation, and alterations in the downstream signaling pathways of Map kinase (MAPK) and TGF- β have been reported in the analysis of metastatic tumor tissue compared to the primary tumor (Holubekova et al.). Given the influence of the immune system on signal transduction pathways, it may be related to tumor-associated macrophage activity. The subsequent macrophage response supports the inflammatory and immune responses. The constant production of proinflammatory cytokines and chemokines leads to further changes in cellular pathways. Selection of tumor clones with blocked apoptosis, and the emergence of mutations cause mutation accumulation, cell cycle errors, and uncontrolled proliferation of tumor cells. On the other hand, genetic alterations such as KRAS mutations specifically affect adaptive and innate cytokine production and the immune system (Mei et al.). In addition, different infiltrations of six immune cells (macrophages, CD8+ T cells, neutrophils, B cells, dendritic cells, and CD4+ T cells) along with the change in expression of Peroxiredoxins (PRDXs) were clinically significant for the prognostic status of CRC which potentiate to be therapeutic targets or new markers for CRC (Zhou et al., Luo et al., Chen et al.). Meanwhile, the Cancer Genome Atlas (TCGA) cohort revealed differences in Transient Receptor Potential (TRP) channel expression between normal tissues and CRC. Furthermore, a poor prognosis was reported in CRC patients with low expression of Transient Receptor Potential Cation Channel Subfamily A Member 1 (TRPA1) and overexpression of Transient receptor potential cation channel subfamily M member 5 (TRPM5) and Transient Receptor Potential Cation Channel Subfamily V Member 4 (TRPV4), with these genes being considered as hub genes. Nevertheless, more detailed research is essential to fully realizing the role of TRP channels in cancer in order to plan

novel, more precise, and valuable pharmacological tools. Moreover, activation of antitumor immunity occurred through TRP channels and upregulated M1 macrophages and CD4 + activated memory T cells. As one of the defining immune markers in CRC, CD103+CD39+ T cells may be a critical factor for antitumor immunity (Luo et al.). In these studies, the perspective of the immune infiltration landscape and personalized prognostic signatures have been identified.

In parallel to immunological alterations, novel genetic and epigenetic modifications have been introduced as promising biomarkers for screening and prediction of prognosis and drug response in CRC. Authors have indicated that successful cancer screening approaches contribute to early detection, improved prognosis, and a reduction in cancer-related deaths. In this regard, higher colorectal cancer (CRC) detection rates have been achieved through a combination of syndecan 2 (SDC2) gene methylation testing and fecal immunochemical testing (FIT) using non-invasive stool samples (Zeng et al.). The addition of the serum carcinoembryonic antigen (CEA) test to the above-combined tests further increased the detection sensitivity. Therefore, stool-methylated SDC2 along with FIT and serum CEA can be considered a sensitive screening approach for the early detection of CRC (Zeng et al.). Hypermethylated glutamate ionotropic receptor AMPA type subunit 4 (GRIA4) has also been introduced as a potential tissue-specific biomarker for the early detection of CRC (Lukacova et al.).

Metastasis is also a critical factor leading to a remarkably low survival rate. Therefore, finding reliable metastasis biomarkers is essential for cancer management. Plasma hypermethylated GRIA4 has been considered a diagnostic target for non-invasive detection of metastasis (Lukacova et al.). On the other hand, tissue hypomethylated and overexpressed importin 5 (IPO5), upregulated hsa-miR-200, hsa-miR-135b-3p, and -5p, and also downregulated hsa-miR-548 were potential epigenetic biomarkers for colorectal cancer liver metastases (CRCLMs) (Horak et al.). Deregulated genes related to PI3K/AKT and WNT signaling pathways are potential biomarkers associated with CRCLMs. To differentiate Synchronous metastatic CRC (SmCRC) from Metachronous metastatic CRC (MmCRC) epigenetic modifications such as deregulated has-miR-1269-3p and hsa-miR-

625-3p may be helpful (Horak et al.). The higher mutation burden in MmCRC and the significantly downregulated SMOC2 and PPP1R9A genes in SmCRC are other markers to distinguish SmCRC from MmCRC (Horak et al.).

Predicting tumor prognosis and drug response is particularly important for treatment planning and patient management. A novel risk model signature was constructed based on three genes of m5C regulator-mediated RNA methylation modification, including NOP2/Sun RNA methyltransferase 4 (NSUN4), NSUN7, and DNA methyltransferase 1 (DNMT1) (Zhang et al.). The prediction of prognosis and response to immunotherapy in rectal cancer patients is applicable using this novel signature. SPP1-CD44 and ALCAM-CD6 are two other immunological biomarkers of poor prognosis in rectal CRC (rCRC) patients (Jie et al.).

It could be concluded that epigenomic alterations in tumor cells are precise and promising predictors of the outcome of CRC patients. In addition, combination therapy with epigenetic drugs can achieve epigenetic changes to truly improve the efficacy response of cancer patients to targeted therapy which is applicable using novel biomarker signatures.

Author contributions

PM: Writing – original draft. MZ: Writing – review & editing.

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.

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.



OPEN ACCESS

EDITED BY

Pooneh Mokarram,
Shiraz University of Medical Sciences,
Iran

REVIEWED BY

Ying Liang,
Jiangxi Agricultural University, China
Tuoxian Tang,
University of Pennsylvania,
United States
Youtao Lu,
University of Pennsylvania,
United States
Zeguo Sun,
Icahn School of Medicine at Mount
Sinai, United States
Jinyang Cai,
Virginia Commonwealth University,
United States
Zishan Wang,
Icahn School of Medicine at Mount
Sinai, United States
Xin Wang,
National Institutes of Health (NIH),
United States

*CORRESPONDENCE

Chen Li
chenli2012@zju.edu.cn
Dante Neculai
dneculai@zju.edu.cn
Xun Zeng
xunzeng@zju.edu.cn

[†]These authors have contributed
equally to this work

SPECIALTY SECTION

This article was submitted to
Cancer Immunity
and Immunotherapy,
a section of the journal
Frontiers in Immunology

RECEIVED 04 August 2022

ACCEPTED 26 September 2022

PUBLISHED 12 October 2022

CITATION

Luo Y, Zong Y, Hua H, Gong M,
Peng Q, Li C, Neculai D and Zeng X
(2022) Immune-infiltrating signature-
based classification reveals
CD103⁺CD39⁺ T cells associate with
colorectal cancer prognosis and
response to immunotherapy.
Front. Immunol. 13:1011590.
doi: 10.3389/fimmu.2022.1011590

Immune-infiltrating signature- based classification reveals CD103⁺CD39⁺ T cells associate with colorectal cancer prognosis and response to immunotherapy

Yang Luo^{1,2†}, Yunfeng Zong^{1,3†}, Hanju Hua⁴, Meiting Gong⁵,
Qiao Peng^{1,3}, Chen Li^{6,7*}, Dante Neculai^{2*} and Xun Zeng^{1,3*}

¹State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, National Medical Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China, ²Department of Cell Biology, Department of Pathology Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, ³Research Units of Infectious disease and Microecology, Chinese Academy of Medical Sciences, Hangzhou, China, ⁴Colorectal Surgery Department, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China, ⁵Zhejiang University-University of Edinburgh Institute (ZJU-UoE Institute), Zhejiang University School of Medicine, Zhejiang University, Haining, China, ⁶Department of Human Genetics, Women's Hospital, Zhejiang University School of Medicine, Hangzhou, China, ⁷Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare, Hangzhou, China

Background: Current stratification systems for tumor prognostic prediction and immunotherapeutic efficacy evaluation are less satisfying in colorectal cancer (CRC). As infiltrating immune cells in tumor microenvironment (TME) played a key role in tumor progression and responses to immune checkpoint blockade (ICB) therapy, we want to construct an immune-related scoring system with detailed immune profiles to stratify CRC patients.

Methods: We developed a scoring system based on immune-related signatures and validated its ability to predict prognosis and immunotherapeutic outcomes in CRC. CD45⁺ cells from CRC patients were sorted to investigate detailed immune profiles of the stratification system using mass cytometry. A single-cell RNA sequencing dataset was used to analyze transcriptomic profiles.

Results: We constructed an immune-related signature score (IRScore) based on 54 recurrence-free survival (RFS)-related immune signatures to stratify CRC patients. We revealed that IRScore was positively correlated with RFS and favorable outcomes in ICB treatment. Moreover, we depicted a detailed immune profile in TME using mass cytometry and identified that CD103⁺CD39⁺ T cells, characterized by an exhaustive, cytotoxic and proliferative phenotype, were enriched in CRC patients with high IRScore. As

a beneficial immune signature, CD103⁺CD39⁺ T cells could predict prognosis and responses to ICB therapy in CRC.

Conclusions: All the analyses above revealed that IRScore could be a valuable tool for predicting prognosis and facilitating the development of new therapeutic strategies in CRC, and CD103⁺CD39⁺ T cells were one of defined immune signatures in IRScore, which might be a key factor for antitumor immunity.

KEYWORDS

colorectal cancer, immunotherapy, prognosis, high-dimensional single-cell analysis, immune cell diversity

Introduction

Colorectal cancer (CRC) ranks the third most commonly diagnosed malignant human cancer and the second leading cause of death worldwide (1). In China, 592,232 new colorectal cases and 309,114 CRC-related deaths were estimated to occur in 2022 (2). Currently, main therapeutic methods of CRC usually involve local treatments, including surgery removal, radiation therapy and systematic therapies like chemotherapy and targeted therapy. The most widely used CRC stratification system is the AJCC/TNM staging system (3). However, due to the significant heterogeneity in patients with CRC identified by the genetic or epigenetic investigation and transcriptomic profiles, patients with the same tumor stage may respond differently to the same treatment and thus lead to varying clinical outcomes. Therefore, there is a clinical requirement to establish a method for predicting CRC prognosis and evaluating the efficacy of immunotherapy.

Since its successful application in melanoma, tumor immunotherapy, especially immune checkpoint blockade (ICB), has been increasingly becoming a preferred consideration for various cancers (4–7). ICB treatment can reactivate exhausted functional cells and elicit durable antitumor responses. In this scenario, immune cells in tumor microenvironment (TME) play a key role in tumor progression and can influence the efficiency of ICB therapy. A positive response to ICB therapy relies on the context of TME and its interactions with tumor cells (8). TME is a complex and heterogeneous mixture of tumor cells, non-tumor cells such as infiltrating lymphocytes, macrophages, fibroblasts, stromal cells and other surrounding host cells, and non-cellular components like extracellular matrix (ECM) and secreted factors (9–12). These immune and non-immune cells with distinct functions can suppress or promote tumor progression. For example, higher

level of cytotoxic T lymphocytes (CTLs) infiltration has been shown to be positively correlated with better antitumor responses and prognosis (13, 14). Thus, depicting the detailed infiltrating immune profiles and understanding the role of each immune subset in TME are required to improve the efficacy of ICB therapy. On the other hand, due to the complexity of TME, most patients, including CRC, do not benefit from current ICB therapy strategies, resulting in wasted healthcare resources and poor prognosis. Therefore, an effective stratification system, which can distinguish responders and non-responders for ICB therapy, can facilitate precision and personalized medicine. Since the infiltrating immune cells in TME played a key role in tumor prognosis and immunotherapeutic outcomes, an infiltrating immune cell signature-based subtyping system will be useful for tumor prognosis prediction and immunotherapeutic efficacy evaluation in CRC.

This study focused on immune-related signatures in TME and investigated the relationship between immune signatures and prognosis, and the association of immune signatures with response to ICB therapy in CRC. We evaluated 151 immune-related signatures in TME collected from public studies in 7 CRC cohorts and developed immune-related signature score (IRScore) based on 54 identified prognosis-related signatures. We depicted a detailed immune profile of IRScore by performing cytometry by time of flight (CyTOF) analyses of tumor infiltrating lymphocytes (TILs) in CRC. We revealed that CD103⁺CD39⁺ T cells were one of defined immune signatures in IRScore, and the CD45⁺CD3⁺CD103⁺CD39⁺ signature could predict patients' prognosis and responses to ICB therapy (Figure S1A). Our study provides a useful tool for predicting CRC prognosis and response to ICB therapy and uncovers important clues to key immune subsets that affect tumor progression and outcomes of ICB therapy.

Materials and methods

Acquisition and pre-processing of CRC cohorts

Gene expression data and clinical information of CRC cohorts used in this study were acquired from UCSC Xena (15) and Gene Expression Omnibus (GEO) (16). Two CRC single-cell datasets were downloaded from GEO under the accession number GSE178341 and GSE108989. The metastatic urothelial cancer treated with anti-PD-L1 agent (atezolizumab) cohort IMvigor210 was obtained via R package *IMvigor210CoreBiologies* (17). The melanoma cohort undergoing anti-PD-1 checkpoint inhibition therapy was downloaded from GEO under accession number GSE78220. Only primary tumors and treatment-naïve patients were included.

Microarray data (GSE103479, GSE17538, GSE33113, GSE37892, GSE38832, GSE39084 and GSE39582 used to develop IRScore; GSE39395 and GSE39396 used to draw signatures of fibroblasts, endothelial cells and epithelial cells) were downloaded and processed using the R package *GEOquery* (18). For TCGA cohort, log-transformed (on a base 2 scale) gene expression data were downloaded from UCSC Xena (<https://xenabrowser.net/datapages/>). Processed expression data of GSE178341, GSE108989 and GSE78220 were directly downloaded from GEO. Counts data and patients' information from the IMvigor210 cohort were obtained by function *counts* and *pData*, and the gene counts were transformed into TPM for the following analysis. All datasets used in this study were listed in Table S1.

Collection of immune-related signatures

One hundred forty-eight immune-related signatures were collected from previously published studies through a literature search (Table S2) (19–22). The signature genes of fibroblasts, endothelial cells and epithelial cells were obtained by performing differential analysis in GEO cohorts GSE39395 (immune cells: CD45⁺Epcam⁻, epithelial cells: CD45⁻Epcam⁺, stromal cells: CD45⁻Epcam⁻) and GSE39396 (immune cells: CD45⁺EPCAM⁻CD31⁻FAP⁻, epithelial cells: CD45⁻EPCAM⁺CD31⁻FAP⁻, endothelial cells: CD45⁻EPCAM⁻CD31⁺FAP⁻, cancer-associated fibroblasts: CD45⁻EPCAM⁻CD31⁻FAP⁺).

IRScore calculation

First, we calculated a single sample gene set enrichment score for each patient using the *gsva* function implemented in the R package *GSVA* (23) and scaled the enrichment score to

draw a normalized enrichment score (NES). Patients were divided into high and low groups according to the median value. Then Univariate Cox regression was performed to examine the relationship between NES and RFS in each CRC cohort. A meta-analysis implemented in the R package *meta* (24) was used to evaluate the Hazard Ratio (HR) and *P*-value. Only signatures with a *P*-value less than 0.05 were included. Totally we identified 54 immune-related signatures and classified them into 30 prognostically good signatures (HR < 1) and 24 prognostically bad signatures (HR > 1). We thus defined and calculated a so-called IRScore for each sample as:

$$\text{IRScore} = \sum_{i=1}^M \text{NES}_i - \sum_{j=1}^N \text{NES}_j$$

where *NES_i* represents *NES* of *i*th prognostically good signature and *NES_j* is *NES* of *j*th prognostically bad signature; *M* and *N* denote the number of prognostically good and bad signatures, respectively.

Prediction of immunotherapeutic response

Prediction of ICB therapy response for TCGA and GSE39582 cohorts was conducted using the subclass mapping method (SubMap) (25). The SubMap module implemented in GenePattern (<https://www.genepattern.org/>) was used to conduct the prediction. Besides, a melanoma cohort treated with sequential CTLA-4 and PD-1 blockade was used to help predict patients' responses to anti-CTLA-4 and anti-PD-1 treatment (26).

Differential gene expression analysis and gene set enrichment analysis

Differential gene expression analysis for microarray data was performed by R package *limma* (27), and for TCGA counts data, *DESeq2* (28) was introduced for analysis. Gene set enrichment analysis and KEGG analysis was performed by R package *clusterProfiler* (29). HALLMARK gene sets and KEGG gene sets used in GSEA analysis were downloaded from the Molecular Signatures Database (MSigDB) (30). The curated signatures were obtained and summarized from previously published studies and provided in Table S3 (31–34).

Single-cell RNA-seq data analysis

Following the standard analysis procedures, scRNA-seq data were analyzed using the R package *Seurat* (35). For scRNA-seq data downloaded from GSE178341, 64 clusters were obtained

using function *FindClusters*. IRScore was calculated for each cluster and each patient based on average expression data derived by the *AverageExpression* function. Clusters and patients were classified into high and low IRScore groups respectively. As for scRNA-seq data of CRC T cells from GSE108989, we calculated IRScore for each cell and likewise classified them into high and low IRScore groups according to the median value. UMAP and tSNE algorithms implemented in *Seurat* were used to visualize high-dimensional data.

Bulk RNA sequencing and CyTOF

Thirteen fresh CRC tumor samples were collected from the First Affiliated Hospital, Zhejiang University School of Medicine. Clinical information of CRC patients was recorded in [Table S4](#). All participants, or their legally authorized representatives, provided written informed consent upon enrollment. Each CRC tumor tissue was divided into two parts, one for RNA-seq and one for CyTOF. Tumor tissues were kept in RNAlater and total RNA was extracted using TRIzol reagent (Thermo Fisher Scientific). RNA-seq was performed at Beijing Genomics Institute (BGI) (Shenzhen, Guangdong, China) using the DNBSEQ system. UCSC hg38 reference genome was used to map the paired-end transcriptome reads. FPKM and read counts were generated for subsequent analysis.

The rest of tumor tissues from the same patients were transferred to MACS Tissue Storage Solution (Miltenyi Biotec), digested and prepared into single-cell suspensions as previously reported ([36](#)). Briefly, Samples were washed in RPMI 1640 (Thermo Fisher Scientific), suspended in 5 ml Hank's Balanced Salt Solution (HBSS) (Thermo Fisher Scientific) with 1 μ M DTT, 5 mM EDTA and incubated at 37°C shaker at 145 rpm for 30 minutes. After washing twice with RPMI 1640, samples were then mechanically dissociated with a sterile scalpel and digesting in a buffer cocktail containing 2 mg/ml collagenase IV (Sigma), 20 μ g/ml DNase (Sigma) in RPMI 1640 for 2 hours in a 37°C shaker at 145 rpm in gentleMACS C tubes (Miltenyi Biotec), followed by dissociating on the gentleMACS Dissociator (Miltenyi Biotec) for 30 minutes. Tissue samples were filtered through a 100 μ M cell strainer, washed, and enriched using 36% Percoll (GE Healthcare) at 2000 rpm for 10 minutes. Single cell suspensions were washed twice with PBS and stained with 5 μ M 103Rh (Fluidigm) for 5 minutes at RT for viability. Cells were fixed in Fix I buffer (Fluidigm) at RT for 10 minutes and resuspended in freezing solution (90% FBS, 10% DMSO) after washing, and were stored at -80 °C for future use. A mass cytometry panel with 41 metal isotope-tagged antibodies ([Table S5](#)) was used to profile immune signatures in CRC samples. CyTOF was performed at Zhejiang Puluoting Health Technology Co., Ltd (Hangzhou, Zhejiang, China) by Helios (Fluidigm) with 300 events/s. Data were exported as FCS files. Fcs files were read into R by *read.FCS* function and signal

intensities were arcsinh transformed with a cofactor of 5. The R function *metaClustering_consensus* implemented in package *FlowSOM* ([37](#)) was used to cluster all cells into 36 clusters. The tSNE algorithm was performed on 13,000 randomly selected cells (1000 cells per sample) to demonstrate high-dimensional data. The 99th percentile of marker intensity was defined as the maximum to exclude extreme value, and then all markers' intensities were rescaled ranging between 0 to 1. The cluster-marker expression heatmap was generated by the R package *pheatmap* based on the median expression value.

Statistical analyses

Log-rank test implemented in R package *survival* was used to evaluate differences in recurrence-free survival between high and low groups. The Kaplan-Meier curves were drawn using the R package *survminer*. R function *coxph* implemented in *survival* was used to compute the Cox proportional hazards regression model. The circular heatmap was visualized by R package *circlize* ([38](#)). Spearman's correlation was calculated by function *rcorr* implemented in R package *Hmisc*. All data except mentioned above were displayed using R package *ggplot2* ([39](#)). Wilcoxon signed-rank test was used to compare the difference between two groups, and the Kruskal-Wallis test was used to compare differences among three or more groups. All analyses were conducted using R software (version 4.1.1). *P*-value < 0.05 was considered statistically significant unless explicitly noted.

Results

Construction of IRScore and the association with clinical and molecular phenotypes

To develop a predictive scoring system of recurrence-free survival (RFS) for CRC, single sample gene set enrichment analysis (ssGSEA) was performed to calculate enrichment score for each patient from 7 CRC cohorts based on 151 curated immune-related signatures. A univariate Cox regression model was applied to evaluate the predictive value of normalized enrichment score (NES) in each cohort. After leveraging 7 CRC cohorts, 54 immune-related signatures significantly associated with prognosis were identified (*P*-value < 0.05) ([Figure 1A](#)). We calculated IRScore for each patient in each cohort and stratified patients into high IRScore group and low IRScore group according to the median value, that is, CRC patients with IRScore higher than the median value were allocated as high IRScore group, and those lower than the median value as low IRScore group. Survival analysis revealed that patients in the high IRScore group had longer survival time without recurrence than those in the low IRScore group ([Figure](#)

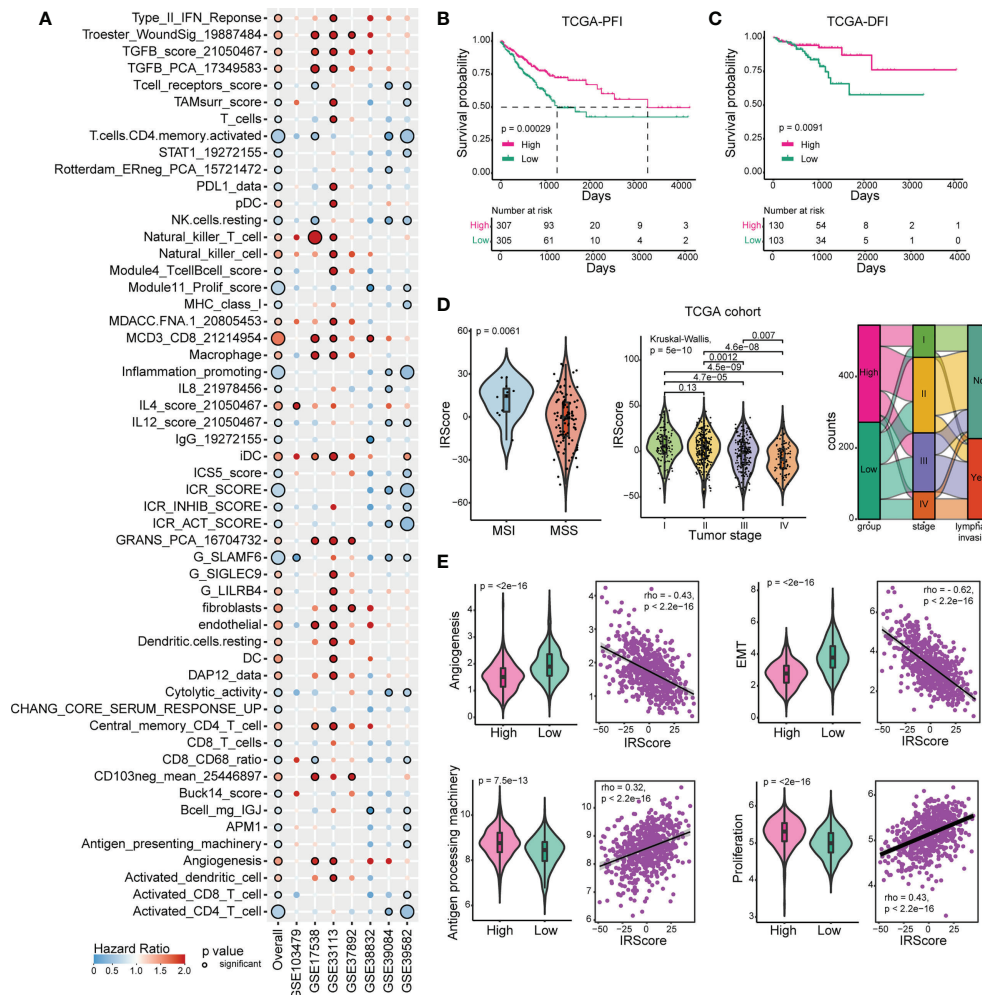


FIGURE 1

Construction and exploration of IRScore in CRC cohorts. (A) Immune-related signatures with survival significance in CRC. Circles with black border represent prognostic significance of the signature in corresponding cohort, and the size of circles represents the significance level. Red circles represent Hazard Ratio (HR) > 1, implying prognostically "bad" signatures, and blue circles represent HR < 1, implying prognostically "good" signatures. (B, C) Kaplan-Meier curves of progression-free interval (PFI), (B) and disease-free interval (DFI), (C) based on IRScore groups in TCGA cohort. (D) Violin plots showing the relationship between IRScore and MSI/MSS (left), tumor stages (middle), and Sankey diagram illustrating the relationship between IRScore and CRC subtypes in TCGA cohort. (E) The association and correlation between IRScore and gene signatures (angiogenesis, EMT, antigen presenting machinery and proliferation) in TCGA cohort illustrated by violin plots (left) and scatter plots with trend lines (right) in TCGA cohort.

S1B). Same phenomena were observed in the combined dataset and in TCGA cohort regardless of progression-free interval (PFI), disease-free interval (DFI), disease-specific survival (DSS) or overall survival (OS) (Figures 1B, C, S1B-E).

Then we investigated the relationships between IRScore and clinical features. Non-metastatic CRCs with microsatellite instability/deficient mismatch repair (MSI/dMMR) had been reported to have better prognosis and immunotherapeutic outcomes than microsatellite stability/proficient mismatch repair (MSS/pMMR) (40). We revealed that CRC patients harboring MSI/dMMR had significantly higher IRScore than those with MSS/pMMR. IRScore differed among four summarized AJCC stages,

with higher IRScore in early-stage patients and lower IRScore in advanced-stage patients, indicating that IRScore might be associated with tumor progression in CRC. The imbalanced components of high and low IRScore groups in tumor stages, lymphatic invasion and MMR were summarized by Sankey diagrams (Figures 1D, S1F). We also explored associations between IRScore and different molecular signatures. We observed that IRScore was negatively correlated with angiogenesis, epithelial-mesenchymal transition (EMT), and positively correlated with antigen processing machinery and proliferation (Figure 1E). Moreover, IRScore was positively associated with immune signatures such as CD8 T effector, cytotoxicity, immune checkpoint, proliferation-related

signatures, and negatively associated with pan-fibroblast TGF- β response signature (pan-F-TBRS), naiveness and plasminogen inhibitor (Figure S1G). These data suggested that IRScore was positively correlated with factors for good prognosis in CRC.

IRScore could be an indicator to predict responses to ICB therapy

The efficacy of ICB therapy targeting programmed cell death 1 (PD-1) or cytotoxic T-lymphocyte associated protein 4 (CTLA-4) was limited in CRC, so we examined the predictive and immunotherapeutic efficacy of IRScore. We used SubMap and a melanoma cohort treated with sequential CTLA-4 and PD-1 blockade to predict patients' responses. Patients in the high

IRScore group might respond to PD-1 inhibitors in both cohorts, while there was a possibility that patients in the low IRScore group in GSE39582 were likely to respond to CTLA-4 inhibitors (Figures S2A, B).

We further used the published IMvigor210 cohort to investigate the predictive efficacy of IRScore. Patients with high IRScore significantly had longer survival time than those with low IRScore (Figure 2A). The inflamed phenotype presented the highest IRScore than desert or excluded phenotypes, and tumors in the high IRScore group had higher neoantigen burdens (Figures 2B, C). The results showed that the CR/PR group had the highest IRScore and patients in the high IRScore group displayed better responses to ICB (Figures 2D-F). The predictive efficacy of IRScore was also testified in GSE78220 cohort. High IRScore patients presented favorable responses and

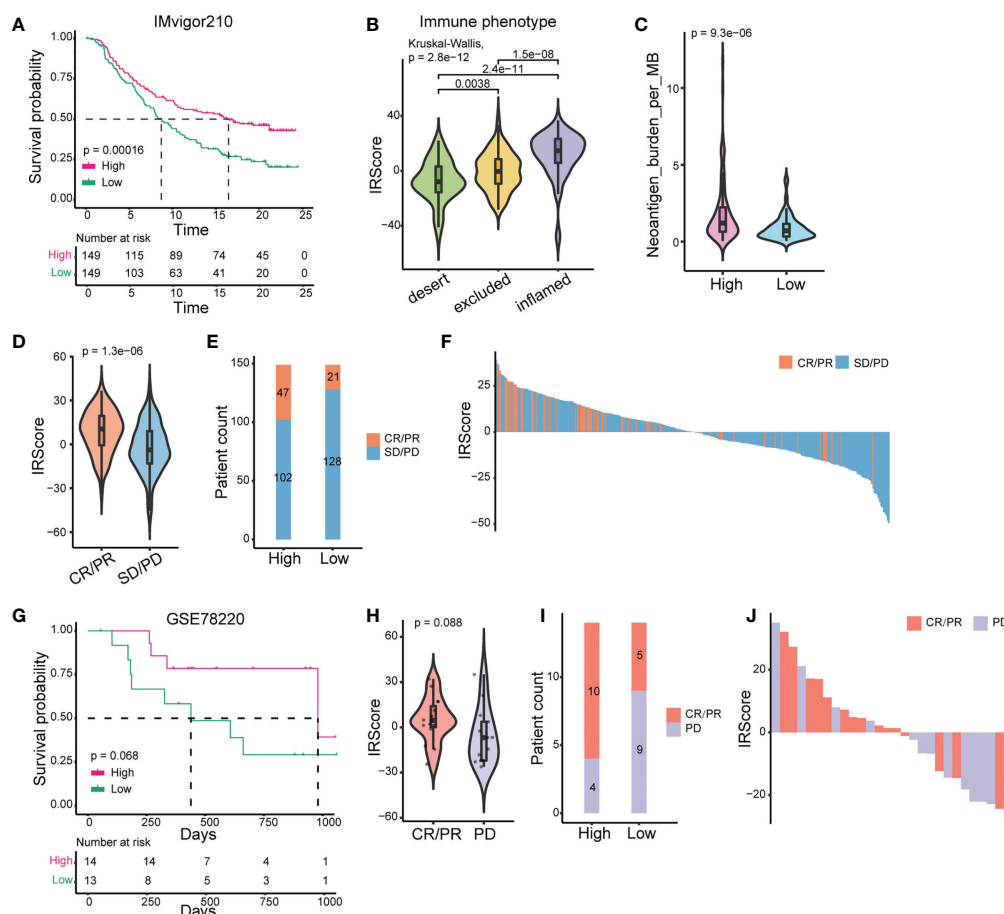


FIGURE 2

Prediction of patients' responses to immunotherapy by IRScore. (A) Kaplan-Meier curve of survival probability based on IRScore groups in IMvigor210 cohort. (B) Comparison of IRScore among different immune phenotypes. (C) Comparison of neoantigen burden between high and low IRScore groups. (D) Violin plots of IRScore between CR/PR and SD/PD. (E) Compositions of patients' responses to PD-L1 inhibitor treatment between high and low IRScore groups (P-value < 0.05, Fisher's exact test). (F) Waterfall plots illustrating IRScore according to immunotherapeutic responses in IMvigor210 cohort. (G) Kaplan-Meier curve of survival probability based on IRScore groups in GSE78220 cohort. (H) Violin plot of IRScore between CR/PR and PD. (I) Compositions of patients' responses to PD-1 inhibitor treatment in the two IRScore groups. (J) Waterfall plot illustrating IRScore according to immunotherapeutic responses in GSE78220 cohort.

prolonged survival (Figures 2G-J). Together, these results implied that higher IRScore was associated with better responses and longer survival time in ICB treatment patients.

Transcriptomic, genomic and immune signatures of high and low IRScore groups in CRC

Underlying changes accompanied the phenotypic differences between high and low IRScore groups. We assessed transcriptomic, genomic and immune features between high

and low IRScore groups to investigate such changes. We first examined Cancer Hallmark gene sets in the two groups by GSEA. Gene sets related to proliferation and inflammation significantly contributed to the positive side, indicating that they were significantly enriched in up-regulated genes when comparing the high IRScore group to the low IRScore group. Epithelial-mesenchymal transition (EMT) and signaling pathways known to induce EMT, including TGF- β , Notch, Wnt- β -catenin and Hedgehog, were enriched in low IRScore tumors (Figures 3A, S3A). We further analyzed differentially expressed genes (DEGs, absolute log2FoldChange > 1, adjusted P -value < 0.05) between the two groups coupled with Kyoto

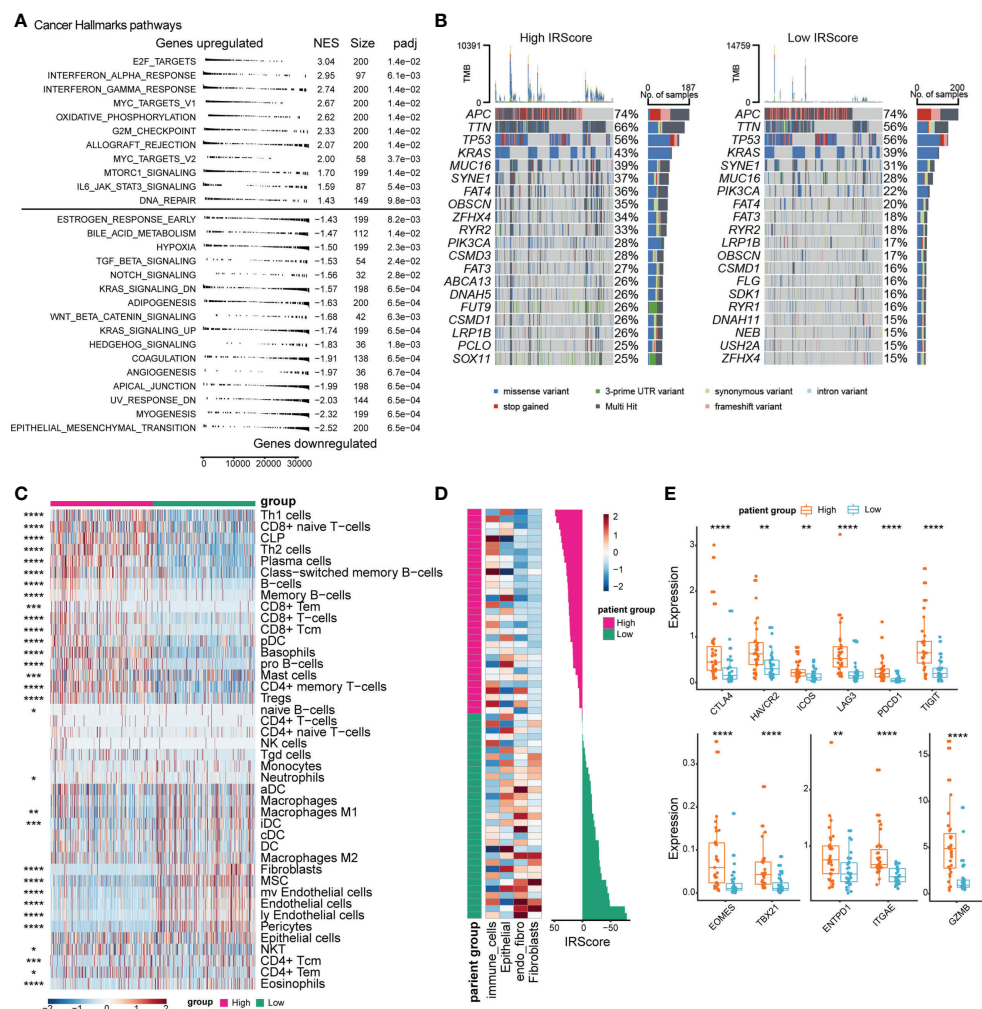


FIGURE 3

Comparison of transcriptomic profile, genomic alteration and immune infiltration between high and low IRScore groups. (A) Gene set enrichment analysis of Cancer Hallmark gene sets identified significantly enriched pathways in high and low IRScore groups. (B) Oncoplot showing the top 20 most frequently mutated genes in high and low IRScore groups in TCGA cohort. (C) Abundance of different cell types in high and low IRScore groups was estimated by xCell in TCGA cohort. Z-scored results were depicted in heatmap. (D) Abundance of different cell types in high and low IRScore groups in the single-cell dataset GSE178341. Z-scored results were depicted in heatmap (left). Waterfall plot illustrating IRScore of each patient according to high and low groups (right). (E) Box plots showing the expression level of indicated genes in the two IRScore groups. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

Encyclopedia of Genes and Genomes (KEGG) pathways enrichment analysis. Consistently, the most enriched KEGG pathway represented by low IRScore tumors was ECM-receptor interaction. In high IRScore tumors, up-expressed genes were enriched for immune-related pathways implicated in processes crucial for host innate/adaptive immune responses (Figures S3A, B). Moreover, we analyzed gene mutation conditions of the two groups in TCGA cohort. The top 6 mutated genes were identical in the two groups but with distinct mutation frequencies except for APC (74%) and TP53 (56%). Besides, most genes tended to have higher mutation frequencies in the high IRScore group, in tune with the result that tumors with high IRScore had higher mutation burdens (Figures 3B, S3C).

Next, we investigated the TME characteristics of the two groups. Using xCell (41), we estimated immune and stromal cell compositions for patients with different IRScore in TCGA cohort. The high IRScore group was enriched with several CD8⁺ T cell types, Type 1 T helper (Th1), Type 2 T helper (Th2), B cell types, plasma cells, plasmacytoid dendritic cells (pDCs), and mast cells, but less fibroblasts, mesenchymal stem cells (MSCs), endothelial cells, pericytes (Figure 3C). To evaluate whether tumor-infiltrating cells displayed similar patterns as in TCGA cohort in a single-cell perspective, we utilized a single cell dataset (GSE178341) and classified tumor cells into 64 clusters. We identified 22 immune cell clusters, 33 epithelial clusters, 3 fibroblast clusters, 4 clusters expressing both endothelial and fibroblast markers (referred to as endo_fibro clusters) and 2 clusters expressing both epithelial and immune cell markers (referred to as other) (Figures S3D, E). We calculated cluster-level IRScore and classified 64 clusters into high and low IRScore groups. The high IRScore group consisted of 16 immune cell clusters, 15 epithelial clusters and 1 other cluster, and the low IRScore group comprised 6 immune cell clusters, 18 epithelial clusters, 3 fibroblast clusters, and 4 endo_fibro clusters and 1 other cluster (Figure S3G). Consistently, the immune cell clusters had the highest IRScore, followed by epithelial clusters. Fibroblast and endo_fibro clusters presented the lowest IRScore (Figures S3F, H). Moreover, we calculated IRScore for each patient based on average gene expression and classified 62 patients into high and low IRScore groups. Patients with high IRScore had higher frequency of immune cells, while fibroblasts were enriched in low IRScore group (Figure 3D). We also evaluated gene expression between the two groups, and it showed that patients in the high IRScore group displayed significantly higher expression of immune checkpoint genes *CTLA4*, *HAVCR2*, *ICOS*, *LAG3*, *PDCD1* and *TIGIT*. Besides, expression of T cell development-associated genes *TBX21* and *EOMES*, tumor reactivity-associated genes *ITGAE* and *ENTPD1*, and cytotoxic gene *GZMB* were also significantly higher in patients with high IRScore (Figure 3E).

Collectively, we observed significant differences in transcriptomic, genomic and immune characteristics between

the two groups. Mesenchymal and tumor-promoting phenotypes represented the low IRScore tumors, and in contrast, the high IRScore tumors displayed immune-active characteristics.

CRC TILs' clustering and subtype analysis

Although high IRScore closely correlated with several immune signatures (Figure 3C), the detailed immune phenotype was still missing. We collected 13 treatment-naïve CRC samples and performed bulk RNA-seq and CyTOF for TILs. We calculated IRScore for each patient according to RNA-seq data and divided them into high and low IRScore groups. Using the t-distributed stochastic neighbor embedding dimensionality reduction algorithm, we visualized the diversity of CD45⁺ tumor-infiltrating cells. We identified 4 major clusters: T cells, B cells, natural killer (NK) cells and myeloid cells (Figures 4A, B). PCA analysis of cell frequencies showed that patients in high and low IRScore groups were clearly separated (Figure 4C). T cells were the most abundant immune cell population among TILs, with a mean of 70% across samples, followed by B cells with a mean of 24% (Figure 4D). Besides, frequency of T cells was higher in the high IRScore group than that in the low IRScore group (P -value = 0.073). No significant difference was observed for frequencies of B cells, NK cells, or myeloid cells between the two groups (Figure 4E).

To further investigate functional subtypes of the overall TILs, an in-depth clustering analysis was conducted, and those TILs were finally classified into 36 clusters (Figure 4F). Specifically, we identified 4 B cell clusters (B01-B04), 2 myeloid cell clusters (M01-M02), 2 NK cell clusters (NK01-NK02), 16 CD4⁺ T cell clusters (T01-T16), 9 CD8⁺ T cell clusters (T17-T25), 2 $\gamma\delta$ T cell clusters (T27-T28) and 1 CD4⁺CD8⁺ T cell cluster (T26). Most T cell clusters presented effector memory phenotype (Tem: CCR7⁻, CD45RA⁻) or central memory phenotype (Tcm: CCR7⁺, CD45RA⁻) (42) and expressed classical activation marker CD69 (43) (Figure 4F). We compared frequency of each cluster between high and low IRScore groups. Except for clusters M01, T07, T12, T14, T17 and T25, most clusters showed no significant difference between the two groups (Figures 4G, S4A). Intriguingly, clusters with higher frequencies in the high IRScore group (T12, T14, T17 and T25) were positive for CD103 and CD39 (Figure 4H). Frequencies of total CD103⁺CD39⁺ TILs were also significantly higher in the high IRScore group (Figure 4G, P -value = 0.073). Previous studies revealed that co-expression of CD103 and CD39 identified a unique population of tumor-reactive CD8⁺ TILs in solid human tumors with an exhausted tissue-resident memory phenotype (44, 45). T25 exhibited high expression of exhaustion markers PD-1, CTLA4 and low expression of CCR7, CD45RA, CD127 and CD28, representative of an effector-memory phenotype (Figure 4I).

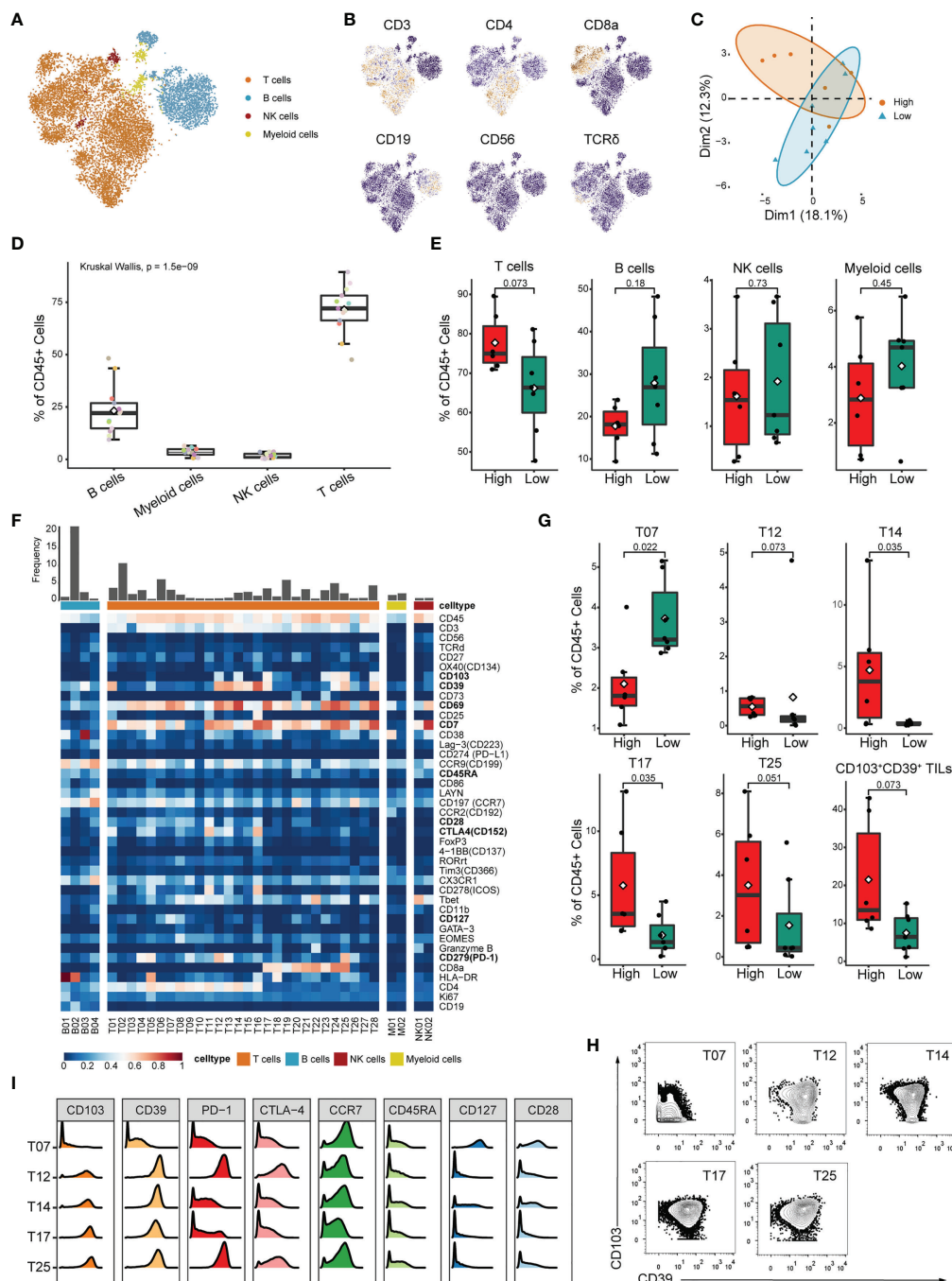


FIGURE 4

Identification of detailed immune profiles of CRC TILs by mass-cytometry. **(A)** tSNE plot displaying T cells, B cells, NK cells and myeloid cells based on manual annotation. **(B)** tSNE plots of normalized expression of markers used to annotate main immune clusters. **(C)** PCA analysis of cluster frequencies in the two IRScore groups. Each dot represented one patient. **(D)** Boxplot showing the frequency of B cell, myeloid cell, NK cell and T cell for each sample. **(E)** Comparison of frequencies of T cell, B cell, NK cell and myeloid cell between high and low IRScore groups. **(F)** Heatmap of normalized marker expression for 36 immune clusters. **(G)** Boxplots showing significant differences of frequency in T07, T12, T14, T17, T25 and total CD103⁺CD39⁺ TILs (T12+T14+T17+T24+T25) between high and low IRScore groups. **(H)** Contour plots showing expression of CD103 and CD39 in T07, T12, T14, T17 and T25. **(I)** Histograms showing expression of indicated markers in T07, T12, T14, T17 and T25.

T12 was the CD4⁺ counterpart of T25, with an almost identical marker expression pattern (Figures 4F, I). CD39⁺CD4⁺ TILs shared similar activated, tissue-resident and effector cell-associated signatures with CD39⁺CD103⁺CD8⁺ TILs (46). The CD4⁺ T14 and CD8⁺ T17 were also inter-counterparts (Figures 4F, I). Together, clusters co-expressing CD103 and CD39 had higher frequencies in the high IRScore group. This phenotype might be relevant to better prognosis in CRC patients with higher IRScore.

The characteristics of CD103⁺CD39⁺CD4⁺/CD8⁺ T cells and relationship with IRScore

To further evaluate whether CD103⁺CD39⁺ T cells represented the immune signature of the high IRScore group, we calculated IRScore of tumor-infiltrating CD4⁺/CD8⁺ T cells from a single-cell dataset (GSE108989) and investigated gene expression patterns of CD4⁺/CD8⁺ TILs co-expressing CD103 (*ITGAE*) and CD39 (*ENTPD1*). We classified those TILs into high and low IRScore groups. Besides, we defined cells expressing CD103 and CD39 as double-positive cells (DP), cells expressing neither CD103 nor CD39 as double-negative cells (DN), and cells expressing either CD103 or CD39 as single-positive cells (SP).

Sixteen clusters were identified, including 8 CD4⁺ clusters and 8 CD8⁺ clusters (Figures 5A, S5A). CD8⁺ clusters were identified as activated effector cytotoxic cells based on the expression of canonical cytotoxic markers *GZMA/B/H*, *PRF1*, *IFNG* and *NKG7*. CD8_2, CD8_6, and CD8_8 also exhibited expression of exhaustion markers *PDCD1* and *HAVCR2*, indicating dual characteristics. Among CD4⁺ clusters, CD4_2 and CD4_4 specifically expressed naive marker genes such as *CCR7*, *IL7R* and *SELL*, thus representing naive T cells; CD4_1, CD4_3, CD4_5 and CD4_8 were characterized by high expression of *FOXP3* and *IL2RA*, suggestive of the identity of regulatory T cells (Tregs); CD4_6 and CD4_7 were comprised of CD4⁺ T cells with high expression of exhausted marker genes *HAVCR2*, *PDCD1* and cytotoxic molecules *GZMA*, *GZMB*, *PRF1*, indicative of the status of exhausted and cytotoxic CD4⁺ T cells (Figure S5B).

DP cells comprised 46.2% of total detected cells and were significantly enriched in the high IRScore group with absolute predominance over DN cells. In contrast, DN cells only constituted 13.3% of whole cells. There were relatively equal proportion of DP and DN cells in the low IRScore group (Figures 5B, E). Besides, DP cells were enriched in 6 CD4⁺ and 7 CD8⁺ clusters (Figure 5C). Surprisingly, a roughly coincident ratio could be observed when comparing the distribution of high and low IRScore TILs in each cluster (Figure 5D). Moreover, DP cells were composed of the most significant proportion of high IRScore cells, and DN cells were the least, consistent with the

observation that DP cells displayed the highest IRScore, with DN cells being the lowest, and SP cells having intermediate IRScore (Figures 5F, H). We also compared IRScore between CD4⁺ and CD8⁺ cells and found that CD8⁺ cells had significantly higher IRScore than CD4⁺ cells, on account of a higher percentage of DP cells among CD8⁺ cells (Figures 5G, I).

We examined the expression of well-known naive, cytotoxic and exhausted markers among DP, DN and SP cells (45). Intriguingly, DP cells displayed dual phenotypes with higher expression of cytotoxic and exhausted marker genes, whereas DN cells had higher expression of naive genes; SP cells exhibited an intermediate status (Figure 5J). To better understand the function of DP cells, we compared them with DN cells at the transcriptome level. Among CD8⁺ T cells, DP cells highly expressed a set of 898 genes (adjusted *P*-value < 0.05, log2FC > 0), including exhausted markers (*HAVCR2*, *LAYN*, *TIGIT*, *PDCD1*, *CTLA4*), cytotoxic markers (*GZMA*, *GZMB*, *PRF1*, *NKG7*) and proliferation-related genes MCMs (Figure 5K). The concurrence of an exhausted phenotype with cytotoxic and proliferative characteristics in DP cells was further confirmed by GSEA analysis of HALLMARK gene sets and the curated signatures. Genes up-expressed among DP cells were significantly enriched for processes associated with inflammation, proliferation, cytotoxicity and presented an exhausted phenotype (Figures 5L, S5E). CD4⁺ DP cells also exhibited an exhausted and proliferative phenotype (Figures S5C, D). GSEA analysis of HALLMARK gene sets and the curated gene signatures revealed that immune-related pathways were enriched considerably in up-expressed genes in CD4⁺ DP cells, and an unfavorable status of hypoxia and unfolded protein response was likewise found in CD4⁺ DN cells (Figure S5F). Together, these data indicated that DP cells were accumulated in the high IRScore group, characteristic of a cytotoxic, exhausted and proliferative phenotype.

CD45⁺CD3⁺CD103⁺CD39⁺ signature could predict the CRC prognosis and responses to ICB therapy

As identified in CyTOF and scRNA-seq analyses, high IRScore could be partially represented by a CD45⁺CD3⁺CD103⁺CD39⁺ signature, we further examined its predictive value and efficacy in predicting response to ICB therapy. Similarly, we calculated NES of this CD45⁺CD3⁺CD103⁺CD39⁺ signature and divided patients into two groups (M1 and M2) according to the median value. In TCGA cohort, patients with higher NES presented prolonged survival than those with lower NES (Figures 6A, B). In ICB therapy cohort IMvigor210, patients with higher NES had prolonged overall survival, and NES of patients with better responses (CR/PR/SD) to ICB therapy were significantly higher than those with progressive disease (PD). Besides, the higher NES group (M2) consisted of more patients that benefited from ICB therapy (Figures 6C, S6A-C). We

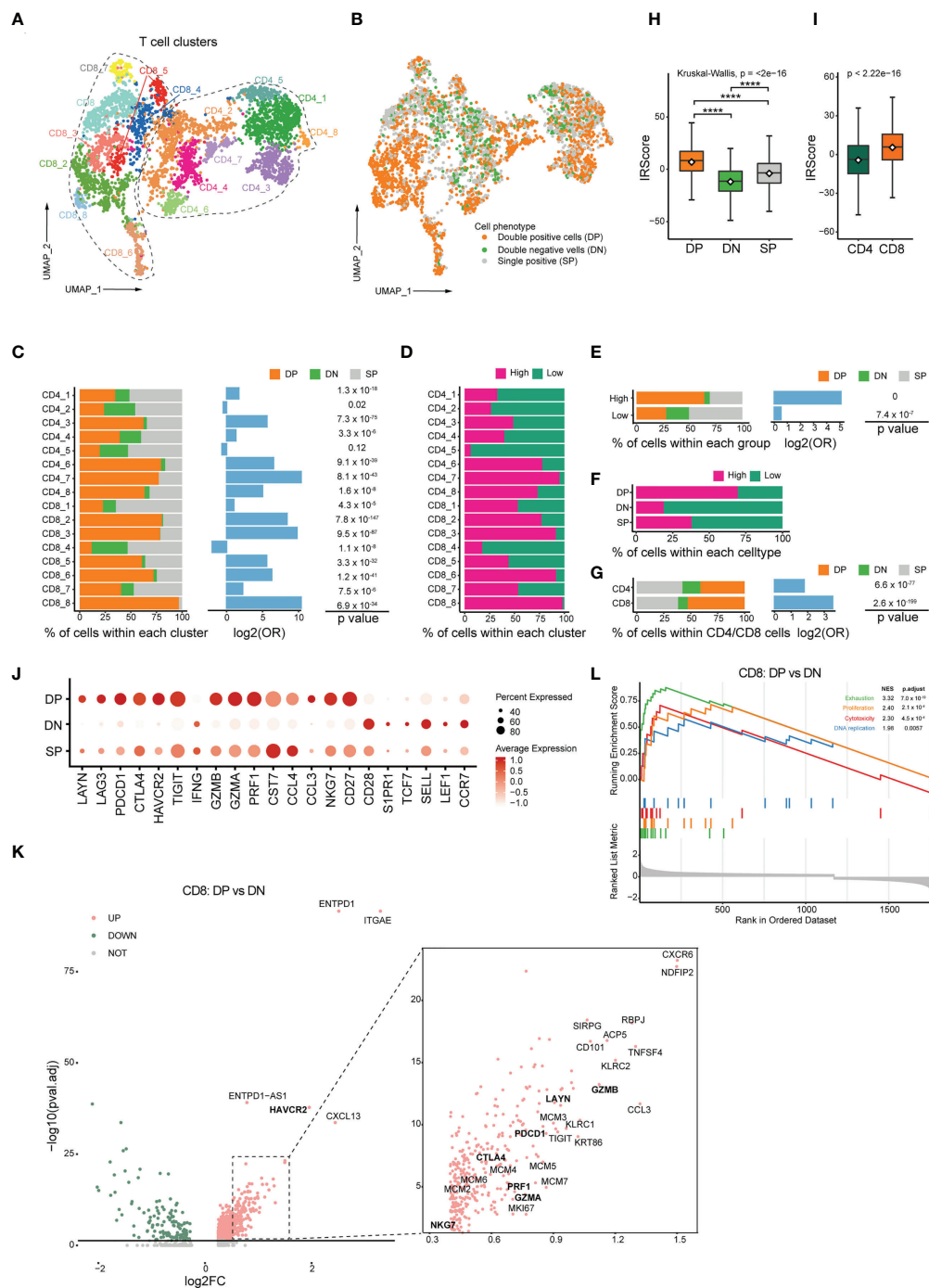


FIGURE 5

Transcriptomic profiles of CD103⁺CD39⁺ T cells and association with IRScore. **(A)** UMAP projection of CD4⁺ and CD8⁺ TILs, showing 8 CD8⁺ clusters and 8 CD4⁺ clusters in different colors. **(B)** UMAP projection of CD103⁺ and CD39⁺ double-positive cells (DP), double-negative cells (DN) and single-positive cells (SP). **(C)** Stacked barplot showing the percentile of DP, DN and SP cells in each clusters (left), log₂ odds ratio (DP versus DN, middle) and p value (right). **(D)** Percentile of high and low IRScore TILs in each cluster. **(E)** Stacked barplot showing the percentile of DP, DN and SP cells in high and low IRScore groups (left), log₂ odds ratio (DP versus DN, middle) and p value (right). **(F)** Percentile of high and low IRScore TILs among three cell types. **(G)** Stacked barplot showing the percentile of DP, DN and SP cells in CD4⁺ and CD8⁺ TILs (left), log₂ odds ratio (DP versus DN, middle) and p value (right). **(H)** Comparison of IRScore among DP, DN and SP cells. **(I)** Comparison of IRScore between CD4⁺ and CD8⁺ TILs. **(J)** Dotplot showing expression of exhausted, cytotoxic and naive markers in DP, DN and SP cells. **(K)** Volcano plot of differentially expressed genes between CD8⁺ DP and DN cells. **(L)** Enrichment plot for the curated gene signatures in CD8⁺ TILs by GSEA. ****P < 0.0001.

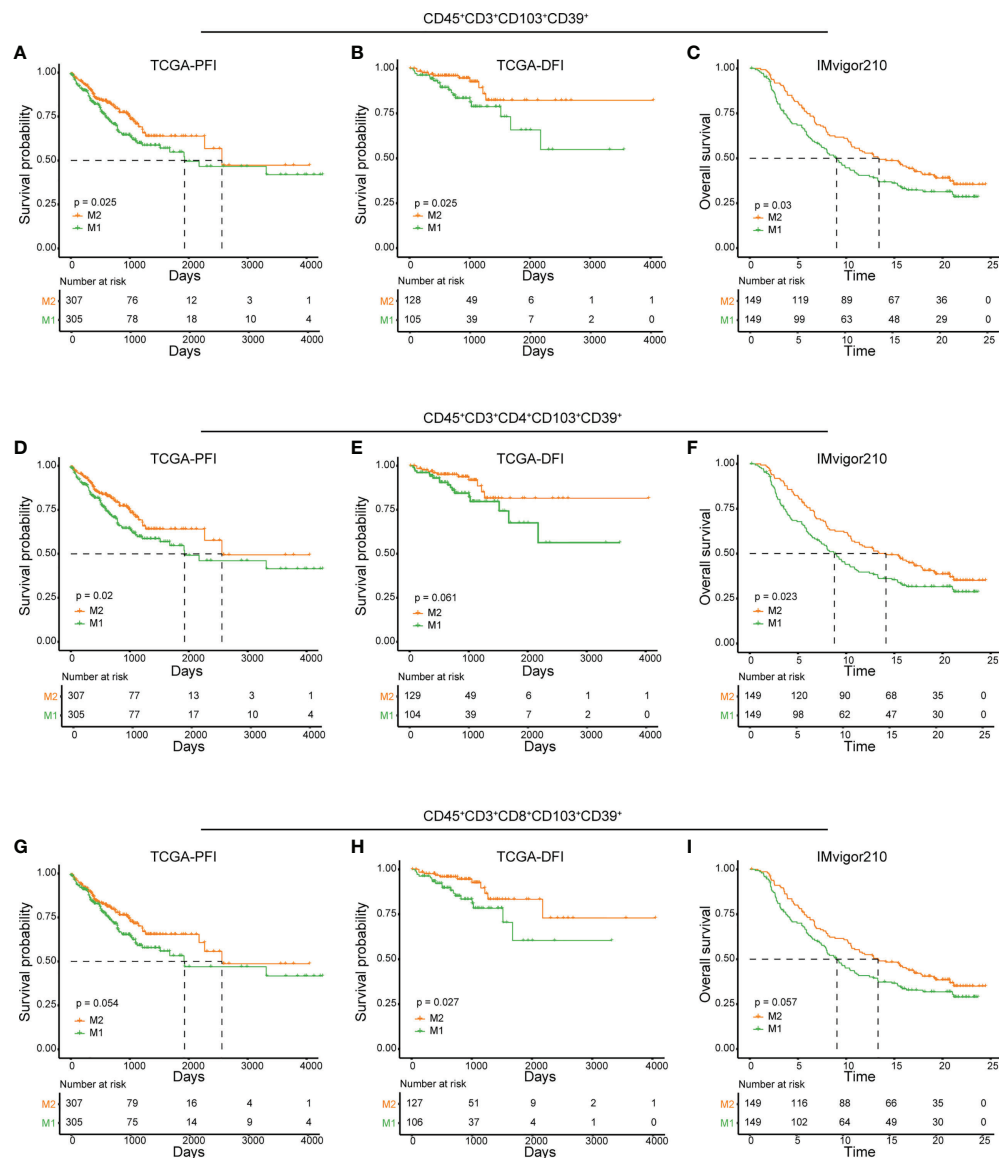


FIGURE 6

Predictive value of CD45⁺CD3⁺CD103⁺CD39⁺ phenotype in prognosis and immunotherapy. (A, D, G) Kaplan-Meier plots of relationship between the immune signatures from RNAseq dataset (CD45⁺CD3⁺CD103⁺CD39⁺ (A), CD45⁺CD3⁺CD4⁺CD103⁺CD39⁺ (D) and CD45⁺CD3⁺CD8⁺CD103⁺CD39⁺ (G) and PFI in TCGA cohort. Patients were stratified according to the median of signature NES, where M1 was the lower NES and M2 was the higher NES. B, E, (H) Kaplan-Meier plots of relationship between the immune signatures from RNAseq dataset (CD45⁺CD3⁺CD103⁺CD39⁺ (B), CD45⁺CD3⁺CD4⁺CD103⁺CD39⁺ (E) and CD45⁺CD3⁺CD8⁺CD103⁺CD39⁺ (H) and DFI in TCGA cohort. Patients were stratified according to the median of signature NES, where M1 was the lower NES and M2 was the higher NES. C, F, (I) Kaplan-Meier plots of relationship between immune signatures from RNAseq dataset (CD45⁺CD3⁺CD103⁺CD39⁺ (C), CD45⁺CD3⁺CD4⁺CD103⁺CD39⁺ (F) and CD45⁺CD3⁺CD8⁺CD103⁺CD39⁺ (I) and OS in IMvigor210 cohort. Patients were stratified according to the median of signature NES, where M1 was the lower NES and M2 was the higher NES.

also showed that CD103⁺CD39⁺ CD4⁺/CD8⁺ T cells had certain ability in predicting prognosis and responses to ICB therapy respectively (Figures 6D-I, S6D-I). Thus, the CD45⁺CD3⁺CD103⁺CD39⁺ signature was a beneficial immune signature that, to a certain extent, could predict the prognosis and efficacy of ICB therapy.

Discussion

The TME is a heterogeneous mixture of tumor cells, infiltrating and resident host cells, extracellular matrix and secreted cytokines (9–12). Cross-talk between TME components significantly affects tumor development and

progression (47). This study utilized immune-related signatures to develop IRScore for CRC patients' stratification. Investigations regarding the relationship between this classification strategy and prognosis, transcriptomic profiles, immune activities etc., will allow for a better understanding of the TME and improved instructions for ICB therapy (Figure S1A).

In this study, we developed the IRScore system based on 54 prognosis-related immune signatures. Our research revealed that IRScore was positively correlated with immune and proliferative signatures and negatively associated with EMT and naiveness (Figure 1E). The proliferation signature of high IRScore might be a result of CD8⁺ TILs, however, alternative interpretation is the proliferating tumor cells. Actively proliferating cells are more likely to be accumulated with mutations, leading to a heavier mutation burden. T cells' recognition of mutation-derived neoantigens in tumors is critical for antitumor activity. Moreover, tumor neoantigens are associated with therapeutic benefits in PD-1 or CTLA-4 blockade. Besides, IRScore was positively correlated with prolonged survival and beneficial responses to ICB therapy, which might be explained by favorable transcriptomic features and active antitumor activities in the high IRScore group. Better responses to ICB treatment in the high IRScore group allowed us to focus on the differences of infiltrating immune signatures in TME between the two groups.

As indispensable components to the TME, TILs show a critical role in tumor progression. The antitumor function involves CD8⁺ cytotoxic cells, NK cells, and CD4⁺ Th1 cells, while MDSCs, TAMs and Tregs inhibit antitumor responses (48–51). We showed that the high IRScore group was more abundant for several CD8⁺ T cell populations and Th1 cells, whereas the low IRScore group had more fibroblasts, endothelial cells and pericytes (Figure 3C). We found more Tregs in the high IRScore group, which might be due to Tregs displaying positive effects and associating with good prognosis in specific cancer types (52–54). Analyses of a CRC single-cell dataset confirmed the fact that endothelial and fibroblast clusters had the lowest IRScore while immune cell clusters were the highest (Figures 3C, E).

Recently, molecular subtyping systems focusing on immune infiltrating signatures have emerged to stratify patients and predict prognosis and immunotherapeutic outcomes in various tumors (55–57). However, these subtyping systems failed to depict detailed immune profiles in TME, which are important for understanding the mechanisms by which immune cells modulate tumor progression and response to ICB therapy. To solve this problem, we used mass cytometry to perform in-depth immune profiling of tumor samples from CRC patients. Among the 36 identified clusters, T12, T14, T17 and T25 were significantly abundant in the high IRScore group, and all of them displayed a CD103⁺CD39⁺ phenotype (Figures 4F–H). Since high IRScore was associated with favorable prognosis and antitumor response, these clusters might represent tumor-reactive populations. T25 exhibited increased expression of PD-1, CTLA4 and low

expression of CCR7, CD45RA, CD127 and CD28 (Figure 4I), consistent with previous studies that a subset of tumor-reactive CD8⁺ TILs were positive for CD103 and CD39 and exhibited an exhausted tissue-resident memory phenotype (44, 45). We further validated CD4⁺ and CD8⁺ TILs co-expressing CD103 and CD39 at the single-cell level. Duhon and colleagues found that CD103⁺CD39⁺CD8⁺ TILs were enriched in CRCs with MSI/dMMR and displayed more elevated exhaustion markers *CTLA4*, *PDCD1*, *HAVCR2* and lower expression of naive markers *SIPRI*, *SELL*, *TCF7* (44). We showed that CD8⁺ DP TILs up-regulated genes related to proliferation, exhaustion, cytotoxicity markers and down-regulated genes related to naiveness (Figures 5J–L). Moreover, CD103⁺CD39⁺CD8⁺ TILs displayed more clonal expansion and better tumor antigen recognition than CD103⁺CD39⁺CD8⁺ TILs, indicating that CD103⁺CD39⁺CD8⁺ TILs were enriched with tumor specific T cells (44, 58). *CXCL13*, a B-cell recruiting chemokine, was the most significantly up-regulated gene except *HAVCR2*, *ITGAE* and *ENTPD1* in CD8⁺ DP cells (Figure 5K). He et al. showed that *CXCL13* was the unique marker for tumor antigen specific CD4⁺/CD8⁺ T cells in multiple tumors (59). Besides, *CXCL13*⁺CD103⁺CD8⁺ TILs were potentially associated with B cell recruitment, neoantigen load and tertiary lymphoid structures (TLSs) formation in human tumors (60), and were identified as tumor antigen specific T cells in lung cancer (59, 61). Thus, CD103 and CD39 positive cells might shape a tumor microenvironment suitable for B-cell antitumor activities.

Other researchers had shown the effects of CD103⁺CD39⁺ TILs in various tumor types. This population of T cells were enriched in genes associated with exhaustion, and may represent a prognostic marker of cancer progression (44, 58). Higher frequencies of CD39⁺CD103⁺CD8⁺ TILs in patients with head and neck cancer were associated with better overall survival, and *in vitro* studies showed that co-expression of CD39 and CD103 were strongly enriched in tumor-recognizing and -killing CD8 T cells (44). Circulating CD103⁺CD39⁺CD8⁺ T cells was significantly enriched in nasopharyngeal carcinoma patients without distant metastasis, and those patients had better PFS (62). CD103⁺CD39⁺ TILs could also be considered as a potential biomarker for predicting patients' response to novel ICB approaches in various tumors. CD103⁺CD39⁺ TILs could serve as a potential biomarker of anti-OX40 clinical activity in patients with head and neck cancer, and might represent a biomarker of RFS following anti-PD-1 therapy in melanoma (63, 64). In our study, we also revealed that the CD45⁺CD3⁺CD103⁺CD39⁺ signature could predict prognosis and response to ICB (Figure 6). These findings suggest that co-expression of CD39 and CD103 could serve as useful markers of tumor specific CD8⁺ T cells, and could be exploited for the development of targeted immunotherapies.

However, due to the limitation of CyTOF samples and markers, we only observed limited immune features that were significantly differentiated between high and low groups. Other

potential phenotypes might have a significant difference if more samples and markers were included. Besides, more in-depth investigations and basic experimental research are required to explain further the underlying mechanisms of the event in the future.

In conclusion, we developed IRScore to stratify CRC patients and explored various profiles contributing to the differences between high and low IRScore groups. We further characterized the detailed immune signature, CD103⁺CD39⁺ T cells, in IRScore system, which may offer important clues to mechanisms of antitumor immune responses in CRC.

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](#). The RNA-seq data of 13 CRC patients presented in the study are deposited in NCBI Sequence Read Archive, accession number PRJNA884190.

Ethics statement

The studies involving human participants were reviewed and approved by Institutional Review Board of the First Affiliated Hospital of Zhejiang University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

XZ, DN, and CL led and supervised the study as joint authors. YL performed public datasets collection, bioinformatic analyses and generated the figures and tables; YZ performed CyTOF experiments; HH collected tumor samples; MG and QP assisted the experiments. YL wrote the manuscript and XZ, DN, and CL revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by National Natural Science Foundation of China (31870899, 32070899 to XZ; 82272126 and 31301100 to CL; 82103304 to QP), CAMS Innovation Fund for Medical Sciences (2019-I2M-5-045) to XZ and the Independent Task of State Key Laboratory for Diagnosis and Treatment of Infectious Diseases (2022zz07 to QP).

Acknowledgments

We would like to thank all staff members and website maintainers of the TCGA program, GEO data repository and UCSC Xena data portal, and all authors who made their research work public. We thank Dr. Kangxin He for collecting clinical samples. We thank the technical support for CyTOF by Zhejiang Puluoting Health Technology Co., Ltd (Hangzhou, Zhejiang, China). We thank the technical support by the Core Facilities, Zhejiang University School of Medicine. We thank Dr. Hangjun Wu in the Center of Cryo-Electron Microscopy (CCEM), Zhejiang University for his technical assistance on computer clustering. CL is supported by Alibaba Cloud.

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.

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.2022.1011590/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Association between IRScore and clinical factors and molecular signatures. (A). Workflow of this study: IRScore construction and characterization, and IRScore-based detailed immune signatures exploration. (B). Kaplan-Meier curves of recurrence-free survival based on IRScore groups in 7 GEO cohorts. (C). Kaplan-Meier curves of recurrence-free survival based on IRScore groups in a combined dataset. (D, E). Kaplan-Meier curves of disease-specific survival (DSS, D) and overall survival (OS, E) based on IRScore groups in TCGA cohort. (F). Violin plots showing the relationship between IRScore and dMMR/pMMR (left), tumor stages (middle) and Sankey diagram (right) illustrating the relationship between IRScore and CRC subtypes in GSE39582 cohort. (G). The association and correlation between IRScore and different gene signatures (Pan-F-TBRS, naiveness, plasminogen inhibitor, CD8 T effector, cell cycle, cytotoxicity, DNA damage repair, DNA replication, homologous recombination, immune checkpoint, mismatch repair, nucleotide excision repair and exhaustion) in TCGA cohort.

SUPPLEMENTARY FIGURE 2

Prediction of response to immunotherapy using SubMap. (A, B). SubMap prediction of response to anti-PD1 and anti-CTLA4 treatment in TCGA cohort (A) and GSE39582 cohort (B).

SUPPLEMENTARY FIGURE 3

Transcriptomic, genomic and immune characterizations of IRScore. (A). Circos plots of GSEA-based Cancer Hallmarks pathways (top) and KEGG pathways (bottom) differences between high and low IRScore groups (from the innermost to the outmost: TCGA, GSE39582, GSE39084, GSE38832, GSE37892, GSE33113, GSE17538, GSE103479). The grids were colored according to NES (high versus low IRScore group, $p_{\text{adj}} < 0.1$). A blank grid meaning the pathway difference is not significant ($p_{\text{adj}} > 0.1$). (B). Volcano plot showing up- or down-regulated genes of high versus low IRScore group in TCGA cohort ($p_{\text{adj}} < 0.05$, absolute $\log_2\text{FoldChange} > 1$). Corresponding bar plots showing the enriched KEGG pathways in up- or down-regulated genes. (C). Boxplot showing the difference of mutation burden between high and low IRScore groups in TCGA cohort. (D). Dotplot showing the expression of genes used to identify endothelial cells, epithelial cells, fibroblasts and immune cells. (E). The tSNE projection of endo-fibro cells, epithelial cells, fibroblasts and immune cells. (F). The tSNE projection of cluster-level IRScore. (G). Stacked bar chart showing the cluster compositions in high and low IRScore groups. (H). Comparison of IRScore among main cell types. Each point represented one cell subset.

SUPPLEMENTARY FIGURE 4

Difference of cluster frequencies between high and low IEScore groups. (A). Boxplots showing the difference of frequencies in 31 immune clusters between high and low IRScore groups.

SUPPLEMENTARY FIGURE 5

Transcriptomic profiles of CD103⁺CD39⁺ T cells. (A). Violin plot showing the expression of CD69, CD4, CD8A and CD8B. (B). Heatmap of CD4⁺ and CD8⁺ T cells, with 16 main clusters identified, each containing a unique set of signature genes. (C). Volcano plot of differentially expressed genes between CD4⁺ DP and DN cells. (D). Enrichment plot for the curated gene signature in CD4⁺ TILs by GSEA. (E, F). Gene set enrichment analysis of Cancer Hallmark gene sets identified significantly enriched pathways in CD8⁺ (E) and CD4⁺ (F) DP and DN cells.

SUPPLEMENTARY FIGURE 6

Prediction of CD45⁺CD3⁺CD103⁺CD39⁺ phenotype to immunotherapy. (A, D, G). Violin plots showing the difference of NES (CD45⁺CD3⁺CD103⁺CD39⁺ (A), CD45⁺CD3⁺CD4⁺CD103⁺CD39⁺ (D) and CD45⁺CD3⁺CD8⁺CD103⁺CD39⁺ (G)) between CR/PR/SD and PD in IMvigor210 cohort. (B, E, H). Composition of patients' responses to PD-L1 inhibitor treatment between M1 and M2 in IMvigor210 cohort ((B): P -value = 0.010; (E): P -value = 0.019; (H): P -value = 0.035, Fisher's exact test). (C, F, I). Waterfall plots illustrating NES (CD45⁺CD3⁺CD103⁺CD39⁺ (C), CD45⁺CD3⁺CD4⁺CD103⁺CD39⁺ (F) and CD45⁺CD3⁺CD8⁺CD103⁺CD39⁺ (I)) according to immunotherapeutic responses in IMvigor210 cohort.

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 Cancer J Clin* (2021) 71:209–49. doi: 10.3322/caac.21660
- Xia C, Dong X, Li H, Cao M, Sun D, He S, et al. Cancer statistics in China and united states, 2022: profiles, trends, and determinants. *Chin Med J (Engl)* (2022) 135:584–90. doi: 10.1097/CM9.0000000000002108
- 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 Cancer J Clin* (2017) 67:93–9. doi: 10.3322/caac.21388
- Wei SC, Duffy CR, Allison JP. Fundamental mechanisms of immune checkpoint blockade therapy. *Cancer Discovery* (2018) 8:1069–86. doi: 10.1158/2159-8290.CD-18-0367
- Robert C, Long GV, Brady B, Dutriaux C, Maio M, Mortier L, et al. Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med* (2015) 372:320–30. doi: 10.1056/NEJMoa1412082
- Brahmer J, Reckamp KL, Baas P, Crino L, Eberhardt WE, Poddubskaya E, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N Engl J Med* (2015) 373:123–35. doi: 10.1056/NEJMoa1504627
- Garon EB, Rizvi NA, Hui R, Leighl N, Balmanoukian AS, Eder JP, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med* (2015) 372:2018–28. doi: 10.1056/NEJMoa1501824
- Tang H, Qiao J, Fu YX. Immunotherapy and tumor microenvironment. *Cancer Lett* (2016) 370:85–90. doi: 10.1016/j.canlet.2015.10.009
- Henke E, Nandigama R, Ergun S. Extracellular matrix in the tumor microenvironment and its impact on cancer therapy. *Front Mol Biosci* (2019) 6:160. doi: 10.3389/fmolb.2019.00160
- Anderson NM, Simon MC. The tumor microenvironment. *Curr Biol* (2020) 30:R921–5. doi: 10.1016/j.cub.2020.06.081
- Arneth B. Tumor microenvironment. *Med (Kaunas)* (2019) 56:1–21. doi: 10.3390/medicina56010015
- Bussard KM, Mutkus L, Stumpf K, Gomez-Manzano C, Marini FC. Tumor-associated stromal cells as key contributors to the tumor microenvironment. *Breast Cancer Res* (2016) 18:84. doi: 10.1186/s13058-016-0740-2
- Dahlin AM, Henriksson ML, Van Guelpen B, Stenling R, Öberg Å, Rutegård J, et al. Colorectal cancer prognosis depends on T-cell infiltration and molecular characteristics of the tumor. *Modern Pathol* (2011) 24:671–82. doi: 10.1038/modpathol.2010.234
- Nosho K, Baba Y, Tanaka N, Shima K, Hayashi M, Meyerhardt JA, et al. Tumor-infiltrating T-cell subsets, molecular changes in colorectal cancer, and prognosis: cohort study and literature review. *J Pathol* (2010) 222:350–66. doi: 10.1002/path.2774
- Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the xena platform. *Nat Biotechnol* (2020) 38:675–8. doi: 10.1038/s41587-020-0546-8
- Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* (2002) 30:207–10. doi: 10.1093/nar/30.1.207
- Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* (2018) 554:544–8. doi: 10.1038/nature25501
- Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. *Bioinformatics* (2007) 23:1846–7. doi: 10.1093/bioinformatics/btm254
- Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep* (2017) 18:248–62. doi: 10.1016/j.celrep.2016.12.019
- Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol (Clifton N.J.)* (2018) 1711:243–59. doi: 10.1007/978-1-4939-7493-1_12
- Şenbabaoğlu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol* (2016) 17:231. doi: 10.1186/s13059-016-1092-z
- He Y, Jiang Z, Chen C, Wang X. Classification of triple-negative breast cancers based on immunogenomic profiling. *J Exp Clin Cancer Res: CR* (2018) 37:327. doi: 10.1186/s13046-018-1002-1
- Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf* (2013) 14:7–7. doi: 10.1186/1471-2105-14-7
- Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health* (2019) 22:153–60. doi: 10.1136/ebmental-2019-300117

25. Hoshida Y, Brunet JP, Tamayo P, Golub TR, Mesirov JP. Subclass mapping: identifying common subtypes in independent disease data sets. *PLoS One* (2007) 2: e1195. doi: 10.1371/journal.pone.0001195
26. Roh W, Chen PL, Reuben A, Spencer CN, Prieto PA, Miller JP, et al. Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci Transl Med* (2017) 9:1–11. doi: 10.1126/scitranslmed.aah3560
27. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43:e47. doi: 10.1093/nar/gkv007
28. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* (2014) 15:550. doi: 10.1186/s13059-014-0550-8
29. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* (2021) 2:100141. doi: 10.1016/j.xinn.2021.100141
30. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102
31. Zhang L, Yu X, Zheng L, Zhang Y, Li Y, Fang Q, et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* (2018) 564:268–72. doi: 10.1038/s41586-018-0694-x
32. Maynard A, McCoach CE, Rotow JK, Harris L, Haderk F, Kerr DL, et al. Therapy-induced evolution of human lung cancer revealed by single-cell RNA sequencing. *Cell* (2020) 182:1232–1251.e22. doi: 10.1016/j.cell.2020.07.017
33. Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med* (2018) 24:978–85. doi: 10.1038/s41591-018-0045-3
34. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGF β attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* (2018) 554:544–8. doi: 10.1038/nature25501
35. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell* (2021) 184:3573–3587.e29. doi: 10.1016/j.cell.2021.04.048
36. de Vries NL, van Unen V, IJsselstein ME, Abdelaal T, van der Breggen R, Farina Sarasqueta A, et al. High-dimensional cytometric analysis of colorectal cancer reveals novel mediators of antitumour immunity. *Gut* (2020) 69:691–703. doi: 10.1136/gutjnl-2019-318672
37. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P, Dhaene T, et al. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytom A* (2015) 87:636–45. doi: 10.1002/cyto.a.22625
38. Gu Z, Gu L, Eils R, Schlesner M, Brors B. Cirdize implements and enhances circular visualization in R. *Bioinformatics* (2014) 30:2811–2. doi: 10.1093/bioinformatics/btu393
39. Hadley W. *Ggplot2: Elegant graphics for data analysis*. 1st ed. New York: Springer (2009). Corr. 3rd printing 2010 edition.
40. Jin Z, Sinicrope FA. Prognostic and predictive values of mismatch repair deficiency in non-metastatic colorectal cancer. *Cancers (Basel)* (2021) 13:1–16. doi: 10.3390/cancers13020300
41. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* (2017) 18:220. doi: 10.1186/s13059-017-1349-1
42. Mullen KM, Gocke AR, Allie R, Ntranos A, Grishkan IV, Pardo C, et al. Expression of CCR7 and CD45RA in CD4+ and CD8+ subsets in cerebrospinal fluid of 134 patients with inflammatory and non-inflammatory neurological diseases. *J Neuroimmunol* (2012) 249:86–92. doi: 10.1016/j.jneuroim.2012.04.017
43. Cibrián D, Sánchez-Madrid F. CD69: from activation marker to metabolic gatekeeper. *Eur J Immunol* (2017) 47:946–53. doi: 10.1002/eji.201646837
44. Duhon T, Duhon R, Montler R, Moses J, Moudgil T, de Miranda NF, et al. Co-expression of CD39 and CD103 identifies tumor-reactive CD8 T cells in human solid tumors. *Nat Commun* (2018) 9:2724. doi: 10.1038/s41467-018-05072-0
45. Yang R, Cheng S, Luo N, Gao R, Yu K, Kang B, et al. Distinct epigenetic features of tumor-reactive CD8+ T cells in colorectal cancer patients revealed by genome-wide DNA methylation analysis. *Genome Biol* (2019) 21:2. doi: 10.1186/s13059-019-1921-y
46. Kortekaas KE, Santegeerts SJ, Sturm G, Ehsan I, van Egmond SL, Finotello F, et al. CD39 identifies the CD4(+) tumor-specific T-cell population in human cancer. *Cancer Immunol Res* (2020) 8:1311–21. doi: 10.1158/2326-6066.CIR-20-0270
47. Hinshaw DC, Shevde LA. The tumor microenvironment innately modulates cancer progression. *Cancer Res* (2019) 79:4557–66. doi: 10.1158/0008-5472.CAN-18-3962
48. Li F, Zhao Y, Wei L, Li S, Liu J. Tumor-infiltrating treg, MDSC, and IDO expression associated with outcomes of neoadjuvant chemotherapy of breast cancer. *Cancer Biol Ther* (2018) 19:695–705. doi: 10.1080/15384047.2018.1450116
49. Raskov H, Orhan A, Christensen JP, Gogenur I. Cytotoxic CD8(+) T cells in cancer and cancer immunotherapy. *Br J Cancer* (2021) 124:359–67. doi: 10.1038/s41416-020-01048-4
50. Tosolini M, Kirilovsky A, Mlecnik B, Fredriksen T, Mauger S, Bindea G, et al. Clinical impact of different classes of infiltrating T cytotoxic and helper cells (Th1, Th2, Th17) in patients with colorectal cancer. *Cancer Res* (2011) 71:1263–71. doi: 10.1158/0008-5472.CAN-10-2907
51. Sun JC, Beilke JN, Lanier LL. Adaptive immune features of natural killer cells. *Nature* (2009) 457:557–61. doi: 10.1038/nature07665
52. Winerdal ME, Marits P, Winerdal M, Hasan M, Rosenblatt R, Tolf A, et al. FOXP3 and survival in urinary bladder cancer. *BJU Int* (2011) 108:1672–8. doi: 10.1111/j.1464-410X.2010.10020.x
53. Schreck S, Friebe D, Buettner M, Distel L, Grabenbauer G, Young LS, et al. Prognostic impact of tumour-infiltrating Th2 and regulatory T cells in classical Hodgkin lymphoma. *Hematol Oncol* (2009) 27:31–9. doi: 10.1002/hon.878
54. Badoual C, Hans S, Rodriguez J, Peyrard S, Klein C, Agueznay Nel H, et al. Prognostic value of tumor-infiltrating CD4+ T-cell subpopulations in head and neck cancers. *Clin Cancer Res* (2006) 12:465–72. doi: 10.1158/1078-0432.CCR-05-1886
55. Lin Y, Pan X, Zhao L, Yang C, Zhang Z, Wang B, et al. Immune cell infiltration signatures identified molecular subtypes and underlying mechanisms in gastric cancer. *NPJ Genom Med* (2021) 6:83. doi: 10.1038/s41525-021-00249-x
56. Wang S, Zhang Q, Yu C, Cao Y, Zuo Y, Yang L. Immune cell infiltration-based signature for prognosis and immunogenomic analysis in breast cancer. *Brief Bioinform* (2021) 22:2020–31. doi: 10.1093/bib/bbaa026
57. Zeng D, Li M, Zhou R, Zhang J, Sun H, Shi M, et al. Tumor microenvironment characterization in gastric cancer identifies prognostic and immunotherapeutically relevant gene signatures. *Cancer Immunol Res* (2019) 7:737–50. doi: 10.1158/2326-6066.CIR-18-0436
58. Simoni Y, Becht E, Fehlings M, Loh CY, Koo SL, Teng KWW, et al. Bystander CD8(+) T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature* (2018) 557:575–9. doi: 10.1038/s41586-018-0130-2
59. He J, Xiong X, Yang H, Li D, Liu X, Li S, et al. Defined tumor antigen-specific T cells potentiate personalized TCR-T cell therapy and prediction of immunotherapy response. *Cell Res* (2022) 32:530–42. doi: 10.1038/s41422-022-00627-9
60. Workel HH, Lubbers JM, Arnold R, Prins TM, van der Vlies P, de Lange K, et al. A transcriptionally distinct CXCL13(+)CD103(+)CD8(+) T-cell population is associated with b-cell recruitment and neoantigen load in human cancer. *Cancer Immunol Res* (2019) 7:784–96. doi: 10.1158/2326-6066.CIR-18-0517
61. Hanada KI, Zhao C, Gil-Hoyos R, Gartner JJ, Chow-Parmer C, Lowery FJ, et al. A phenotypic signature that identifies neoantigen-reactive T cells in fresh human lung cancers. *Cancer Cell* (2022) 40:479–493.e6. doi: 10.1016/j.ccell.2022.03.012
62. Dong DN, Fan PW, Feng YN, Liu GH, Peng YC, Dong T, et al. Association between circulating CD39+CD8+ T cells pre-chemoradiotherapy and prognosis in patients with nasopharyngeal carcinoma. *Chin Med J (Engl)* (2021) 134:2066–72. doi: 10.1097/CM9.0000000000001745
63. Duhon R, Ballesteros-Merino C, Frye AK, Tran E, Rajamanickam V, Chang SC, et al. Neoadjuvant anti-OX40 (MEDI6469) therapy in patients with head and neck squamous cell carcinoma activates and expands antigen-specific tumor-infiltrating T cells. *Nat Commun* (2021) 12:1047. doi: 10.1038/s41467-021-21383-1
64. Attrill GH, Owen CN, Ahmed T, Vergara IA, Colebatch AJ, Conway JW, et al. Higher proportions of CD39+ tumor-resident cytotoxic T cells predict recurrence-free survival in patients with stage III melanoma treated with adjuvant immunotherapy. *J Immunother Cancer* (2022) 10:1–15. doi: 10.1136/jitc-2022-004771

COPYRIGHT

© 2022 Luo, Zong, Hua, Gong, Peng, Li, Neculai and Zeng. 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.

Glossary

CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CR	Complete response
CRC	Colorectal cancer
CTLA4	Cytotoxic T-lymphocyte associated protein 4
CTLs	Cytotoxic T lymphocytes
CyTOF	Cytometry by time of flight
DEGs	Differentially expressed genes
dMMR	Deficient mismatch repair
DN	Double-negative cells
DP	Double-positive cells
ECM	Extracellular matrix
EMT	Epithelial-mesenchymal transition
GEO	Gene Expression Omnibus
HBSS	Hank's Balanced Salt Solution
HGSC	High-grade serous ovarian cancer
HR	Hazard ratio
ICB	Immune checkpoint blockade
IRScore	Immune-related signature score
MCM	Minichromosome maintenance
MSCs	Mesenchymal stem cells
MSI	Microsatellite instability
MSS	Microsatellite stability
MSigDB	Molecular Signatures Database
NES	Normalized enrichment score
NK	Natural killer cell
pan-F-TBRS	Pan-fibroblast TGF- β response signature
PD	Progressive disease
PD-1	Programmed cell death 1
pDCs	Plasmacytoid dendritic cells
PR	Partial response
pMMR	Proficient mismatch repair
RFS	Recurrence-free survival
scRNA-seq	Single-cell RNA sequencing
SD	Stable disease
SP	Single-positive cells
ssGSEA	Single sample gene set enrichment analysis
SubMap	Subclass mapping
Tcm	Central memory T cell
Tem	Effector memory T cell
Th1	Type 1 T helper cell
Th2	Type 2 T helper cell
TILs	Tumor infiltrating lymphocytes
TME	Tumor microenvironment
Tregs	Regulatory T cells
UPR	Unfolded protein response



OPEN ACCESS

EDITED BY

Michele Ghidini,
IRCCS Ca 'Granda Foundation
Maggiore Policlinico Hospital, Italy

REVIEWED BY

Mohd Farhan,
King Faisal University, Saudi Arabia
Youtao Lu,
University of Pennsylvania,
United States
Daiwei Wan,
The First Affiliated Hospital of
Soochow University, China
Zheng Liu,
National Cancer Center of China,
China
Gabriel B. K. Sasa
Tianjin University, China

*CORRESPONDENCE

Ying Yuan
yuanying1999@zju.edu.cn
Pei-Rong Ding
dingpr@sysucc.org.cn

[†]These authors have contributed
equally to this work and share
first authorship

SPECIALTY SECTION

This article was submitted to
Cancer Immunity
and Immunotherapy,
a section of the journal
Frontiers in Immunology

RECEIVED 15 August 2022

ACCEPTED 28 November 2022

PUBLISHED 23 December 2022

CITATION

Mei W-J, Mi M, Qian J, Xiao N, Yuan Y
and Ding P-R (2022)
Clinicopathological characteristics of
high microsatellite instability/mismatch
repair-deficient colorectal cancer: A
narrative review.
Front. Immunol. 13:1019582.
doi: 10.3389/fimmu.2022.1019582

Clinicopathological characteristics of high microsatellite instability/mismatch repair-deficient colorectal cancer: A narrative review

Wei-Jian Mei^{1†}, Mi Mi^{2†}, Jing Qian³, Nan Xiao³, Ying Yuan^{2,4,5*}
and Pei-Rong Ding^{1*}

¹Department of Colorectal Surgery, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Sun Yat-sen University, Guangzhou, China, ²Department of Medical Oncology (Key Laboratory of Cancer Prevention and Intervention, China National Ministry of Education, Key Laboratory of Molecular Biology in Medical Sciences), The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China, ³Global Medical Affairs, MSD China, Shanghai, China, ⁴Zhejiang Provincial Clinical Research Center for CANCER, Hangzhou, China, ⁵Cancer Center of Zhejiang University, Hangzhou, China

Colorectal cancers (CRCs) with high microsatellite instability (MSI-H) and deficient mismatch repair (dMMR) show molecular and clinicopathological characteristics that differ from those of proficient mismatch repair/microsatellite stable CRCs. Despite the importance of MSI-H/dMMR status in clinical decision making, the testing rates for MSI and MMR in clinical practice remain low, even in high-risk populations. Additionally, the real-world prevalence of MSI-H/dMMR CRC may be lower than that reported in the literature. Insufficient MSI and MMR testing fails to identify patients with MSI-H/dMMR CRC, who could benefit from immunotherapy. In this article, we describe the current knowledge of the clinicopathological features, molecular landscape, and radiomic characteristics of MSI-H/dMMR CRCs. A better understanding of the importance of MMR/MSI status in the clinical characteristics and prognosis of CRC may help increase the rates of MMR/MSI testing and guide the development of more effective therapies based on the unique features of these tumors.

KEYWORDS

colorectal cancer, high microsatellite instability, deficient mismatch repair, microsatellite instability, characteristics, prognosis, therapy development, review

1 Introduction

Colorectal cancers (CRCs) with high microsatellite instability (MSI-H) and deficient mismatch repair (dMMR) are a unique subgroup of cancers of the colon and rectum. The molecular and clinicopathological characteristics of MSI-H/dMMR CRCs are distinct from those of proficient mismatch repair (pMMR)/microsatellite stable (MSS) CRCs (1).

Because of their unique etiology and clinicopathological characteristics, MSI-H/dMMR and pMMR/MSS CRCs respond differently to treatment (2). This is particularly true for immune checkpoint inhibition, as MSI-H/dMMR CRCs are more immunogenic and show a better response to immunotherapy than pMMR/MSS CRCs (3, 4). Recent clinical studies showed that, in patients with advanced or metastatic MSI-H/dMMR CRC, pembrolizumab treatment led to an objective response rate (ORR) of 40.0% (95% confidence interval [CI], 12.0–74.0) (5) and the ORR in patients treated with the combination of ipilimumab and nivolumab was 54.6% (95% CI, 45.2–63.8) (6). In addition, pembrolizumab led to a significantly longer progression-free survival (PFS) than chemotherapy when administered as first-line therapy for metastatic MSI-H/dMMR CRC (hazard ratio [HR] for disease progression or death, 0.10; $P < 0.001$), with fewer treatment-related adverse events (5). KEYNOTE-177, a phase 3 study of 307 previously untreated patients with metastatic MSI-H/dMMR CRC showed that first-line pembrolizumab was superior to chemotherapy in improving PFS (HR for progression, 0.60; 95% CI, 0.45–0.80; $P = 0.0002$) and ORR (43.8% [95% CI, 35.8–52.0] vs. 33.1% [95% CI, 25.8–41.1]) (7). The identification of MSI-H/dMMR as a potential biomarker for response to immunotherapy in patients with CRC has led to the initiation of various clinical trials evaluating the use of neoadjuvant immunotherapy in patients with early-stage disease. Preliminary findings from the exploratory NICHE study (NCT03026140) suggest that neoadjuvant immunotherapy with nivolumab plus ipilimumab may be a suitable regimen for patients with dMMR early-stage colon cancer (8). The ability of neoadjuvant treatment with nivolumab plus ipilimumab to improve outcomes was confirmed in patients with locally advanced dMMR colon cancer (9). Furthermore, neoadjuvant treatment with immunotherapy (nivolumab plus ipilimumab) in combination with the COX-2 inhibitor celecoxib in patients with non-metastatic dMMR CRC led to a major pathologic response in 97% of patients (95% CI, 91–100; 31 of 32) (10). The efficacy of immunotherapy in combination with other treatments (e.g., chemotherapy and radiotherapy) is also being investigated in multiple ongoing trials, including VOLTAGE-A (NCT02948348), AVANA (NCT03854799), NRG-GI002 (NCT02921256), and PANDORA (NCT04083365) (11–14). Neoadjuvant immunotherapy alone or in combination with other therapies may provide new treatment options for patients with early-stage CRC, especially in MSI-H/dMMR CRC.

Despite the importance of MSI-H/dMMR and pMMR/MSS status in clinical decision making, the rates of microsatellite instability (MSI) and mismatch repair (MMR) testing in clinical practice remain low, even in high-risk populations (15, 16). Consequently, the real-world prevalence of MSI-H/dMMR CRC may be higher than that reported in the literature. Insufficient MSI and MMR testing leads to failure to identify patients with MSI-H/dMMR CRC who could benefit from immunotherapy (5, 6). Additionally, because of differences in the epidemiological, molecular, anatomical, and histological characteristics of MSI-H/dMMR and pMMR/MSS CRCs, failure to distinguish between these subgroups may lead to discrepancies in CRC diagnostic and prognostic features (1, 17).

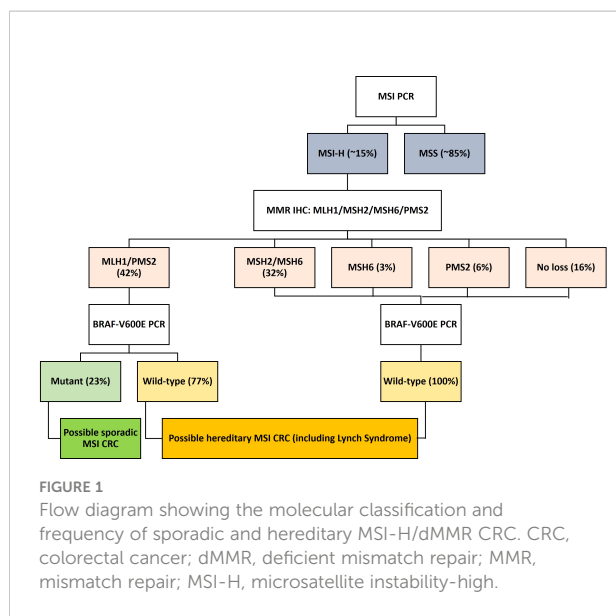
In this article, we comprehensively review the current knowledge of the clinicopathological characteristics, molecular landscape, and radiological findings of MSI-H/dMMR tumors among patients with CRC. This overview of the role of MMR and MSI status in CRCs could increase the understanding of MSI-H/dMMR CRCs, help clinicians identify this subgroup of patients using available approaches besides MSI/MMR testing, and guide the development of more effective therapies based on the unique molecular characteristics of these tumors.

2 Molecular mechanisms of MSI-H/dMMR in CRC

Inactivation of an MMR gene by mutation or transcriptional silencing results in deficient function of the MMR system, leading to the accumulation of errors during DNA replication (18). Multiple proteins that mediate DNA repair are involved in the MMR pathway, including the MutS family (MSH2, MSH3, and MSH6) and the MutL family (MLH1, MLH3, PMS1, and PMS2). Among these proteins, MLH1, MSH2, MSH6, and PMS2 are the most important regulators of MMR (18).

Studies have identified two distinct molecular pathways comprising germline or somatic mutations that contribute to the inactivation of MMR genes. Germline mutations in an MMR gene followed by a second hit to the wild-type copy due to point mutations, loss of heterozygosity (LOH), or methylation (18) can inactivate the gene. Inherited colorectal syndromes contribute to the development of approximately 5% of all CRCs, of which Lynch syndrome is the most common (Figure 1) (18). Mutations in *MLH1* and *MSH2* are found in approximately 70% of patients with Lynch syndrome, whereas mutations in *MSH6* and *PMS2* are less common and are found in only 15% of patients (19).

DNA methylation, also referred to as CpG island methylator phenotype (CIMP), occurs in 20% of CRCs and results in a non-familial form of MSI (19). Such DNA hypermethylation results in gene silencing in most cases (19) or upregulation under certain circumstances (20, 21). Sporadic CRCs are mainly due



to loss of *MLH1* expression caused by hypermethylation of the *MLH1* promoter in a CIMP background (19).

3 Clinical characteristics of MSI-H/dMMR CRCs

3.1 Demographic characteristics and MSI/MMR status in CRCs

Findings from multiple studies suggest that dMMR status is associated with early onset disease among patients with CRC, as dMMR CRCs are more frequent in younger patients than in older patients. A retrospective analysis of 133 patients with CRC showed that mutations in *MLH1*, *MSH2*, *MSH6*, and *PMS2* were significantly associated with age (22). A subsequent retrospective study of 61 patients with stage I–III CRC confirmed a significant association between dMMR status and patient age (23). A recent real-world study revealed that, among patients with dMMR CRC, dMMR tumors were observed in both older (≥ 60 years) and younger (< 50 years) patients. The frequency of *MSH6*/*MSH2*, *MSH6*, and *PMS2* loss was higher in younger patients than in older patients. However, the statistical significance of this finding could not be determined because the expected expression values were low in $> 20\%$ of the cells (24). Among patients with Lynch syndrome, the median age at CRC diagnosis was ten years higher for carriers of *MSH6* mutations than for those carrying *MLH1* and *MSH2* mutations (25).

Similar associations have been reported for dMMR status and sex; in most studies, the percentage of women in the dMMR CRC group was higher than the percentage of men. For example, a large-scale study of 535 patients with CRC showed that tumors from women had a higher frequency of *MLH1*/*PMS2* loss than

tumors from men (26). Consistently, Viñal et al. (27) reported that the percentage of women was significantly higher among patients with dMMR CRC than among those with pMMR CRC (55% [$n = 55/100$] vs. 38% [$n = 351/914$]; $P = 0.001$).

3.2 Tumor characteristics and MSI/dMMR status in CRC

MSI-H/dMMR status has been associated with various CRC tumor characteristics, including the location of the primary tumor, tumor diameter, T stage, and distant metastasis. Several retrospective studies have shown a significant association between dMMR/MSI-H status and early onset disease, maximum tumor diameter, large tumor volume, primary tumor site, and advanced T stage in patients with stage (including tumor, node, metastasis [TNM] stage) I–III or I–IV CRC (23, 27–29).

A retrospective study of 245 patients with CRC showed that the incidence of MSI-H was higher in patients with right colon cancer and TNM stage I–II disease (30). Another retrospective analysis of 268 patients with CRC showed a high incidence of dMMR in patients with locally advanced (T4b) tumors without distant metastasis (31). Additionally, a recent analysis of 1,014 patients with CRC (100 [9.8%] with dMMR and 914 [90.2%] with pMMR tumors) indicated that advanced-stage tumors were significantly more common among patients with pMMR CRC than among those with dMMR CRC (stage IV: 21% vs. 3%; $P < 0.001$) (27). Similarly, Kang et al. (29) found a significant association between MSI-H and earlier-stage tumors in patients with CRC. These findings suggest that dMMR may play a protective role in CRC.

In a retrospective case series, Li et al. found that mutations in *MLH1*, *MSH2*, and *MSH6* were significantly associated with primary tumor location among patients with dMMR CRC; h*MLH1* or *PMS2* loss was more common on the right side, whereas h*MSH2* or h*MSH6* loss was more common on the left side (22). Similarly, a retrospective analysis of 795 patients found that proximal lesions were a predictor for MSI, with a multivariate odds ratio (OR [95% CI]) of 0.419 (0.223–0.784; $P = 0.007$) (32).

However, Yan et al. found that larger tumor size was associated with MSI (OR [95% CI], 1.300 [1.076–1.572]; $P = 0.007$), as did Liang et al. (median diameters, 6.0 cm in the dMMR group compared with 4.5 cm in the pMMR group; $P < 0.01$) (23, 32).

3.3 Histopathological and pathomorphological characteristics of MSI-H/dMMR CRCs

MSI-H/dMMR CRCs and pMMR/MSS/MSI-L CRCs differ in their histopathological and pathomorphological

characteristics. For instance, in a study of 312 patients with colorectal adenocarcinomas, mucinous adenocarcinomas were more common among patients with dMMR CRC than among those with pMMR CRC (33). Most dMMR CRCs show aggressive histological features, including an expansile growth pattern, a high degree of tumor cell infiltration, poor tumor differentiation, and a medullary pattern (34), as summarized in Figure 2. Consistently, Liang et al. (23) reported a significantly higher frequency of poorly differentiated tumors in patients with dMMR CRC than in those with pMMR CRC (41.0% [n = 25/61] vs. 10.9% [n = 20/183]; $P < 0.05$), although no significant differences in the rates of lymphovascular invasion and extranodal extension were observed. In contrast, localized disease at diagnosis (97% vs. 79%; $P < 0.001$) and histological grade 3 (20% vs. 8%; $P < 0.001$) were more frequent in patients with dMMR CRC than in those with pMMR CRC (27).

While both serrated and non-serrated sporadic colorectal adenocarcinomas can present MSI-H (35), studies have shown that MSI is more common in conventional carcinomas than in colorectal serrated adenocarcinomas (36). Other histological and morphological features of dMMR/MSI-H CRCs include high numbers of tumor-infiltrating lymphocytes, Crohn's-like inflammatory reaction, mucinous/focal signet ring cell differentiation, and lack of dirty necrosis within the tumor lumen (37–39).

MSI-H/dMMR status in patients with CRC has also been correlated with high infiltration levels of immune cells, including T helper 1 (Th1) cells and cytotoxic T cells, which may explain the favorable response to immunotherapy among patients with MSI-H/dMMR tumors. The high degree of immune cell infiltration in dMMR CRCs may be attributed to the high mutational burden and neoantigen load of these tumors (3, 40), making dMMR/MSI-H CRCs amenable to immunotherapy. Despite the durable responses observed in some patients with MSI-H/dMMR CRC treated with immune checkpoint inhibitors, many CRCs are resistant to immunotherapy (41, 42). High intratumoral heterogeneity due to the high rate of mutations in MSI-H/dMMR CRCs may contribute to the generation of immune escape clones, leading to the development of immunotherapy resistance (41, 43). Although tumor mutational burden (TMB) and the expression levels of PD-1/PD-L1 have been proposed as determinants of differential responses to anti-PD-1/PD-L1 treatment among CRC patients with different MSI statuses (44), further studies are required to determine the roles of PD-1/PD-L1, *BRAF/RAS* mutations, TMB, and T-cell phenotype as biomarkers of response to immune checkpoint inhibitors in patients with MSI-H/dMMR CRC.

Despite the significant association between dMMR status and certain clinicopathological and tumor histologic characteristics as mentioned above, the role of MSI and MMR

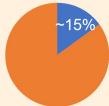

	MSI-H/dMMR	MSS/pMMR
		
Pathways and programs	JAK/STAT→Immune evasion CIMP+	Epithelial or EMT CIMP ^{low} WNT/β-catenin → Myc TGFβ→Angiogenesis Integrin β3→ECM remodelling
Histological features	More aggressive behavior with expansile growth pattern and poor tumor differentiation	
Selected molecular features	MSI+ BRAF ^{V600E} High TMB	CIN+ KRAS mutations Low TMB
Immunological features	High TIL levels Immune infiltrate (e.g. cytotoxic T cells)	Low TIL levels High Treg infiltration

FIGURE 2

Proposed relationship between tumor features, molecular profiles, clinicopathological characteristics, and immunological features of colorectal cancer according to MSI subtype. CIMP, CpG island methylator phenotype; CIN, chromosomal instability; dMMR, mismatch repair deficient; ECM, extracellular matrix; EMT, epithelial–mesenchymal transition; JAK/STAT, Janus kinase/signal transducer and activator of transcription; MSI, microsatellite instability; MSS, microsatellite stable; pMMR, MMR proficient; TGF, transforming growth factor; TIL, tumor-infiltrating lymphocyte; TMB, tumor mutational burden; Tregs, T-regulatory cells; VEGF, vascular endothelial growth factor.

in clinicopathological features is complex and may be confounded by multiple factors. The higher prevalence of dMMR in earlier-stage tumors indicates a protective role of dMMR, while dMMR CRCs show aggressive histological features, including an expansile growth pattern, a high degree of tumor cell infiltration, poor tumor differentiation, and a medullary pattern.

Several factors may contribute to these paradoxical observations regarding the relationship between MSI and tumor characteristics in CRC. Most studies evaluating the relationship between MSI and tumor or clinicopathological features in patients with CRC included small cohorts because dMMR CRC is relatively rare. Additionally, there are no widely established criteria for the diagnosis of MSI-H/dMMR tumors. Variations in the evaluation of MMR status may contribute to contradictory findings regarding the predictive and prognostic roles of MSI-H/dMMR status in CRC. There is also evidence to suggest heterogeneous characteristics among the different subgroups of MSI-H/dMMR CRCs. For example, hereditary and sporadic MSI-H CRCs differ in their histological and morphological characteristics (45). Substantial racial differences in the tumor microenvironment of CRCs have also been reported (46). Moreover, because of defects in DNA repair pathways, dMMR CRCs have substantial genetic instability, which could lead to intertumoral molecular heterogeneity (17, 47). Accumulation of genetic mutations during the progression of MSI-H/dMMR CRCs can also lead to the acquisition of more aggressive features. These and other confounding factors must be accounted for in studies evaluating the role of MSI in the characteristics of CRC.

3.4 Pathomics and artificial-intelligence–assisted prediction of MSI-H/dMMR status in CRC

Advances in artificial intelligence (AI) have augmented the development of pathomics and AI-assisted methods for the characterization of tumors from patients with CRC. Most research efforts have focused on the development of models to automate the analysis and increase the accuracy of gland segmentation, tumor classification, tumor microenvironment characterization, and prognostication (48, 49).

Significant progress has also been made in the development of AI-assisted models to predict MSI status in CRC based on the distinct histomorphological features of MSI-H/dMMR tumors (Table 1) (61). For example, an open-source AI system that was trained using routine pathology slides from eight multicenter cohorts facilitated accurate and fully automated prediction of MSI status, yielding an area under the curve (AUC) value of 0.96 (95% CI, 0.94–0.98) (50). The system was successfully applied as a rule-out test to predict MSS/pMMR and identify patients with CRC for whom molecular MSI testing is not required. Similarly,

Cao et al. (51) developed a pathomics-based deep-learning model trained using histological data from The Cancer Genome Atlas (TCGA) and an Asian CRC cohort. The model accurately predicted MSI status from histopathological images, with an AUC of 0.8848 (95% CI, 0.8185–0.9512) in the TCGA cohort and 0.8504 (95% CI, 0.7591–0.9323) in the Asian CRC cohort. The model accurately captured various characteristics of MSI-H tumors, including poor differentiation and high TMB (51). In an effort to improve the performance of AI algorithms in predicting MSI, Saillard et al. (52) developed a self-supervised deep-learning model that was trained using histology images from the TCGA dataset. The model predicted MSI status in CRC with high sensitivity and specificity, achieving an AUC of 0.92 (95% CI, 0.84–0.99) and outperforming previous supervised deep-learning models.

AI-assisted algorithms have also been developed to predict TMB-H status, which is strongly correlated with MSI-H/dMMR status. Shimada et al. (53) developed a convolutional neural network-based algorithm to predict TMB-H status (defined as MSI-H, high TMB, or both) from hematoxylin and eosin (H&E)-stained slides of CRC tissues. The model integrated various histomorphological features of MSI-H/dMMR CRCs, including increased lymphocytic infiltration, abundance of peritumoral lymphocytes, mucinous features, Crohn's-like inflammatory reaction, and medullary features. The model accurately predicted TMB-H status, providing an AUC of 0.934 (range, 0.835–0.981) (53).

Current testing strategies for MSI-H/dMMR status include polymerase chain reaction (PCR), next-generation sequencing (NGS), and immunohistochemistry (IHC). High testing costs and limited resources are critical factors that hinder the wider application of MSI testing in patients with CRC. Kacew et al. (62) used a representative population-based sample of individuals receiving first-line treatment for metastatic CRC (N = 32,549) in the US to estimate the clinical and financial consequences of predicting MSI status in CRC using AI-assisted methods instead of conventional methods. Their model showed that, compared with current testing strategies, MSI testing using AI followed by confirmatory PCR or IHC testing for patients testing dMMR/MSI-H-positive by AI resulted in the lowest population-level diagnostic costs (including testing and first-line drug costs) in this cohort (\$400 million [12.9%] lower than that of NGS alone). The method also maintained 91% diagnostic accuracy and facilitated timely diagnosis (62). However, a key limitation of this study was that population-based costs were estimated using a model, and no real-world validation of costs was conducted. Hence, it is possible that AI-related costs were underestimated, as additional costs of applying AI in real-world clinical settings (e.g., internal validation, maintenance of hardware and software, and scanning slides) were not taken into account. Further validation of AI-assisted MSI prediction algorithms in large datasets and real-world cohorts is required to support the clinical adoption of these models in routine practice.

TABLE 1 Emerging non-invasive methods for predicting MSI-H/dMMR status in CRC.

Method	Application	AUC (95% CI)	Reference
AI-based MSI/dMMR detector trained using routine pathology slides from eight multicenter cohorts	Rule-out test for predicting MSS/pMMR and identifying patients with CRC for whom molecular MSI testing is not required	0.96 (0.94–0.98)	(50)
Pathomics-based deep learning model trained using histological data from TCGA and an Asian CRC cohort	Prediction of MSI status from histopathology images	TCGA cohort: 0.8848 (0.8185–0.9512) Asian CRC cohort: 0.8504 (0.7591–0.9323)	(51)
Self-supervised deep learning model that was trained using histology images from TCGA	Prediction of MSI status from histology slides	0.92 (0.84–0.99)	(52)
Convolutional neural network-based algorithm	Prediction of TMB-H status (defined as MSI-H, high TMB, or both) from H&E-stained slides of CRC tissues	0.934 (0.835–0.981)	(53)
Serum CEA levels	Prediction of dMMR status in patients with CRC	0.546	(54)
Serum CA 72-4 levels	Prediction of dMMR status in patients with CRC	0.583	(54)
Combination of serum levels of CA 72-4 and CEA with patient age, histology type, tumor size, tumor location, degree of differentiation, LN metastasis, and peripheral nerve invasion	Prediction of dMMR status in patients with CRC	0.849	(54)
Age, tumor diameters, histology, tumor location, perineural invasion, the number of sampled LNs and positive LNs	Prediction of dMMR status in patients with CRC	Primary cohort: 0.756 (0.722–0.789) Validation cohort: 0.754 (0.715–0.793)	(55)
Combination of serum levels of CA 72-4 and CEA with age, tumor diameters, histology, tumor location, perineural invasion, the number of sampled LNs, and positive LNs	Prediction of dMMR status in patients with CRC	Primary cohort: 0.805 (0.774–0.835) Validation cohort: 0.796 (0.758–0.835)	(55)
CT-based nomogram consisting of six radiomic features and 11 clinical characteristics	Prediction of MSI status in patients with stage II CRC	0.752	(56)
Preoperative triphasic enhanced CT radiomics signatures consisting of 32 features	Prediction of MSI status in patients with CRC	Primary cohort: 0.898 (0.860–0.937) Validation cohort: 0.964 (0.919–1.000)	(57)
Nomogram integrating clinical, pathological, and radiomics data	Prediction of MSI status in patients with rectal cancer	0.757 (0.726–0.787)	(58)
Machine learning model trained using both tumoral and peritumoral radiomic signatures	Prediction of MSI status in patients with rectal cancer	Primary cohort: 0.817 (0.772–0.856) Validation cohort: 0.726 (0.648–0.796)	(59)
Model integrating six MRI-derived radiomic features and clinical characteristics	Preoperative prediction of MSI status in patients with rectal cancer	0.895 (0.838–0.938)	(60)

AI, artificial intelligence; AUC, area under the curve; CEA, carcinoembryonic antigen; CRC, colorectal cancer; CT, computed tomography; H&E, hematoxylin and eosin; LN, lymph node; MRI, magnetic resonance imaging; MSI-H/dMMR CRC, microsatellite instability-high/deficient mismatch repair colorectal cancer; MSI, microsatellite instability; MSS, microsatellite stable; pMMR, mismatch repair proficient; TCGA, The Cancer Genome Atlas.

3.5 Molecular characteristics and MSI/MMR status in CRCs

MSI-H tumors have high genetic instability, and dMMR/MSI-H CRCs exhibit extensive intratumoral and intertumoral molecular heterogeneity (47). A key feature of MSI-H CRCs is the lack of MMR proteins or deletions in MMR-related genes. Hence, MSI-H/dMMR CRCs can be distinguished from MSS/pMMR CRCs using IHC, PCR, or NGS. Epigenetic mechanisms (e.g., DNA methylation) also contribute to the loss of MMR proteins in dMMR/MSI-H CRC. Thus, numerous methods have

been developed to detect epigenetic alterations in clinical samples from patients with CRC.

The frequencies of *TP53* loss, *MLH1* promoter methylation, and *KRAS* and *BRAF* mutations vary between MSI-H and MSS CRCs (Table 2) (1, 63). *BRAF*-V600E mutations are more common in MSI-H CRCs than in MSS CRCs. In contrast, *KRAS* mutations and *TP53* loss are more frequent in MSS than in MSI-H CRCs (1, 63). Approximately 80% of dMMR CRCs exhibit *MLH1* promoter methylation (67). Because most sporadic dMMR CRCs exhibit *MLH1* promoter methylation and many also have *BRAF* mutations, *MLH1* promoter methylation

TABLE 2 Frequency of common mutations in MSI-H and MSS CRCs.

Genetic alteration	Mutation Frequency, %		Reference
	MSI-H CRC	MSS CRC	
<i>TP53</i> loss	31.6	46.4	(63)
<i>KRAS</i> mutations	36.8	2.3	(63)
<i>BRAF</i> -V600E mutation	36.8	1.0	(63)
Cytokeratin 7 loss	94 ^a	61 ^a	(64)
Cytokeratin 20 loss	30 ^a	7 ^a	(64)
<i>CTNNB1</i> mutations	10	0.7	(65)
<i>HNF1A</i> mutations	32	0.2	(65)
<i>BRCA1</i> mutations	19	5	(65)
<i>BRCA2</i> mutations	50	14	(65)
Thymidylate synthase upregulation	85	31	(65)
<i>PTEN</i> upregulation	71	48	(65)
<i>HER2</i> mutations	5.6	3.7	(66)

^aTumors harbored *BRAF* mutations.

analysis can be used to discriminate sporadic tumors from Lynch syndrome in patients without *BRAF* mutations. Absence of *BRAF* mutations and *MLH1* promoter methylation in tumors is associated with hereditary forms of CRC (67, 68). In contrast, the *BRAF*-V600E mutation in patients with dMMR CRC is strongly associated with sporadic tumors (69). Several clinicopathological characteristics, including age at diagnosis, tumor location, and patient survival, differ between patients with *MLH1*-deficient/*BRAF*-V600E-mutated dMMR CRC and those with *MLH1*-deficient/*BRAF* wild-type dMMR CRC (70, 71). Interestingly, the characteristics of *BRAF*-mutated MSS CRCs appear to be distinct from those of *BRAF*-mutated MSI-H CRCs and *BRAF* wild-type MSS CRCs. Landau et al. analyzed 205 CRCs and found that stage IV tumors at diagnosis were significantly more common among patients with *BRAF*-mutated MSS CRCs than among those with *BRAF*-mutated MSI-H CRCs and *BRAF* wild-type MSS CRCs ($P < 0.001$) (64). They also found that cytokeratin 7 (CK7) loss was significantly more common in *BRAF*-mutated MSI-H and *BRAF* wild-type MSS CRCs than in *BRAF*-mutated MSS CRCs ($P = 0.0001$). Furthermore, cytokeratin 20 (CK20) loss was more common in *BRAF*-mutated MSI-H CRCs than in *BRAF*-mutated MSS and *BRAF* wild-type MSS CRCs ($P = 0.001$) (64). *BRAF* mutations in patients with metastatic dMMR/MSI-H CRC have been associated with poor outcomes, including shorter overall survival (OS) (72).

Because MSI-H tumors are more genetically unstable than MSS tumors, MSI-H CRCs tend to accumulate mutations in various oncogenes and tumor suppressor genes, including *BRAF*, *CTNNB1*, *HNF1A*, *PTEN*, *BRCA1*, and *BRCA2* (65, 73). Consequently, MSI-H CRCs have a higher TMB and neoantigen load than MSS CRCs (66, 74). Advances in NGS have contributed to the identification of several mutations associated with dMMR/MSI-H status in CRC. NGS analysis of

tissues from 430 patients with CRC showed that mutations in MAPK pathway genes (e.g., *KRAS*, *NRAS*, *BRAF*) and *HER2* were significantly more frequent in MSI-H CRCs than in MSS tumors (83.6% vs. 58.4%, $P = 0.0003$) (66).

In line with the high TMB of dMMR/MSI-H CRCs, NGS analysis of tumor samples from 64 patients with CRC showed that MSI-H tumors harbored a total of 1756 alterations (mean, 125; range 63–302) across 447 genes, whereas MSS tumors had only 493 alterations (mean, 10; range 1–26) across 186 genes (75). Among the total of 633 mutated genes, only 165 were altered in both groups. Both MSI-H and MSS tumors harbored mutations in *APC*, *TP53*, and *KRAS*, which are among the most frequently mutated genes in CRC. The most commonly altered genes that were mutated only in MSI-H tumors were *ANKRD11* (78.6%), *ARID1A* (71.4%), *KMT2B* (71.4%), *BCORL1* (64.3%), *IGF1R* (50.0%), *KDM5* (50.0%), *POLD1* (50.0%), and *TSC1* (50.0%). Additionally, mutations in microsatellite loci (mononucleotide repeats) were more frequent in MSI-H CRCs than in MSS tumors.

Interestingly, the serum levels of molecular tumor markers, including CEA, CA 19-9, and CA 72-4, have also been associated with MSI/MMR status in CRC (55, 63). A retrospective analysis of samples from 2279 patients with CRC indicated that dMMR status was associated with normal CEA serum levels and elevated CA 72-4 levels (54). The use of serum CEA levels to predict dMMR status yielded AUC scores of 0.546 in the entire cohort and 0.554 in the TNM II/III subgroup. Similarly, serum CA 72-4 levels showed a modest ability to predict dMMR status, with an AUC score of 0.583 (54). Although the ability of individual serum tumor markers to predict MSI/MMR status is limited, the combination of serum markers and clinicopathological features may help identify patients with MSI-H/dMMR CRC (55). The combination of serum levels of CA 72-4 and CEA with patient age, histology type, tumor size,

tumor location, degree of differentiation, lymph node metastasis, and peripheral nerve invasion to predict dMMR status in patients with CRC provided an AUC score of 0.849, which was considerably higher than the AUC scores of individual markers (54). Another retrospective analysis of 3,274 patients with CRC confirmed that the addition of CEA and CA 72-4 to dMMR prediction models significantly improved the discriminative ability of a pathology-based model in the primary (AUC: 0.805 [95% CI, 0.774–0.835] vs. 0.756 [95% CI, 0.722–0.789]; $P < 0.001$) and validation cohorts (AUC: 0.796 [95% CI, 0.758–0.835] vs. 0.754 [95% CI, 0.715–0.793]; $P < 0.001$) (55).

3.6 Radiomic characteristics of MSI-H/dMMR CRCs

Using non-invasive methods to predict MSI status prior to treatment or surgery remains an unmet clinical need. The usefulness of traditional radiological evaluation of CRCs using computed tomography (CT) to predict MSI status is limited. Therefore, novel CT technologies and radiomic features to predict MSI/MMR status in CRC have been evaluated in several studies (Table 1).

In a recent radiomics analysis of iodine-based material decomposition images captured by dual-energy CT imaging, a nomogram based on a combination of clinical factors and radiomics scores predicted MSI status in pretreatment patients with CRC (76). Preliminary findings from a retrospective study indicated that a CT-based nomogram consisting of six radiomic features and 11 clinical characteristics could predict MSI status in patients with stage II CRC, yielding an AUC of 0.752 (sensitivity, 0.663; specificity, 0.842) (56). Consistently, a multicenter study demonstrated that preoperative triphasic enhanced CT radiomics signatures consisting of 32 features could predict MSI status in 502 patients with CRC (57). Delayed-phase models were superior to arterial- or venous-phase models in predicting MSI status. Although these studies demonstrated the feasibility of using radiomic features to predict MSI status in CRC, the nomograms were developed based on data from relatively small patient cohorts.

To develop an MSI-predictive nomogram based on radiomics data from a large cohort, Pei et al. (77) used texture analytical software to extract pelvic CT radiomic features from 762 patients with CRC. Patients with MSI-H tumors showed significantly higher radiomics nomogram scores, suggesting that pretreatment radiomic features can be used as a non-invasive method to predict MSI status in patients with CRC (77). In contrast, another study involving the development and validation of a model to predict MSI status by integrating clinical, pathological, and radiomics data from a large cohort of patients with rectal cancer ($n = 788$) showed that the nomogram provided a moderate ability to predict MSI status. However, the model provided a higher AUC than clinical,

pathological, and radiomic features alone (AUCs: 0.757, 0.584, 0.585, and 0.737, respectively) (58). In contrast to previous efforts to develop MSI-prediction models based on tumoral CT-based radiomics, Ma et al. (59) developed a machine-learning model to predict MSI status using both tumoral and peritumoral radiomic signatures. The model predicted MSI status in rectal cancer, achieving AUCs of 0.817 (95% CI, 0.772–0.856) and 0.726 (95% CI, 0.648–0.796) in the training and validation sets, respectively.

Positron emission tomography (PET)/CT has also been used as a non-invasive method to predict MSI status in patients with CRC and analysis of quantitative imaging markers *via* statistical modelling may further reflect pathophysiology and allow objective evaluation of tumor heterogeneity, giving more information than a single IHC/PCR assay. Li et al. (78) conducted a radiomics analysis using preoperative ^{18}F FDG PET/CT images to predict MSI/MMR status in 173 patients with CRC. They identified one PET radiomic feature (wavelet-LHH_firstorder_Skewness_PET) and one CT radiomic feature (wavelet-HHL_firstorder_RootMeanSquared_CT) that were associated with MSI status (both $P < 0.05$), providing a quantitative and non-invasive approach to identify patients with MSI-H/dMMR CRC (78). Metabolic parameters derived from preoperative ^{18}F FDG PET/CT images have also been found to predict MSI status in 44 patients with CRC (79). Metabolic tumor volume (MTV)_{30%}, MTV_{40%}, MTV_{50%}, MTV_{60%}, total lesion glycolysis (TLG)_{50%}, and TLG_{60%} differed significantly between the MSI and MSS groups (all $P < 0.05$). Among these parameters, MTV_{50%} was the strongest predictor of MSI (79). Although PET is expensive and may not be readily available in all clinics, predicting MSI status using PET/CT is non-invasive and does not require tissue biopsy. Therefore, it could be adopted for patients with insufficient biopsy tissues for IHC/PCR testing or without biopsy tissue. Even though IHC is inexpensive and widely available, variations in IHC fixation and staining protocols and objectivity in scoring may influence its accuracy in identifying dMMR/MSI-H tumors.

MRI-based radiomic features have also been used to develop non-invasive models to predict MSI status in CRC. For example, Zhang et al. (60) developed a model integrating six MRI-derived radiomic features and clinical characteristics to predict MSI preoperatively in 491 patients with rectal cancer. The combined model yielded an AUC of 0.895 (95% CI, 0.838–0.938), which was significantly higher than that obtained using clinical characteristics alone (AUC: 0.685 [95% CI, 0.608–0.755]; $P = 0.015$).

3.7 Racial disparities in MSI-H/dMMR CRC

Racial differences in the prevalence of MSI-H/dMMR CRC have been reported (Table 3) (102–104). For example, the

TABLE 3 Incidence of MSI-H/dMMR colorectal cancer in different populations.

Population	Incidence, %	Reference
Egyptians	37.0	(80)
Europeans		
Greece	5.0	
Romania	21.1	(81)
Germany	23.7	(82)
Scandinavia	7.0	(83)
The Netherlands	3.5	(84)
UK	22.9	(85)
African Americans	12.0–45.0	(86, 87)
Caucasian Americans	19.0	(88)
US Latino/Hispanic individuals	13.0	(89)
Mexicans	21.3–27.1	(90, 91)
Japanese	4.0–20.0	(92, 93)
Chinese	5.0–15.0	(94–96)
Koreans	16.0	(97)
Indians	1.0–29.5	(98–101)

incidence of MSI-H CRC is relatively high among Egyptians (37%), African Americans (12%–45%), Europeans (5%–24%), and Caucasian Americans of European descent (8%–20%) (73, 80, 86–88, 105). In contrast, the reported incidence of MSI-H CRC in Asian countries is relatively low, ranging from 3.8% to 20.0% in Japan (92, 93) and from 4.5% to 15.0% in China (94, 95). Furthermore, the reported frequency of dMMR in synchronous CRCs is lower in Japanese patients than in Western patients (106).

In line with racial differences in the prevalence of MSI-H/dMMR CRC, accumulating evidence suggests that racial/ethnic disparities also exist in the genetic profiles of CRCs (102, 107, 108). Zhang et al. sequenced tumors from 1,110 Chinese patients with CRC to identify oncogenic mutations. They found that 45.4%, 3.9%, 3.1%, and 3.5% of tumors harbored mutations in *KRAS*, *NRAS*, *BRAF*, and *PIK3CA*, respectively (109). Interestingly, the frequency of the *BRAF* V600E mutation was 3.1%, which is lower than that reported in studies conducted in Western countries. To identify racial differences in the tumor microenvironment of colon cancers, Paredes et al. analyzed gene expression in tumor tissues in a US cohort. They found that tumors from African American patients had higher expression levels of *FOXP3*, *IL1B*, and *IL8* than tumors from Caucasian Americans (all $P < 0.05$) (46). In contrast, tumors from Caucasian Americans had higher expression levels of markers associated with antitumor immune responses, including *GZMB*, *IFNG*, *CD274* (encoding PD-L1), and *CTLA4* (all $P < 0.05$).

The combination of non-modifiable genetic factors (e.g., family history, genetic polymorphisms) and environmental factors (e.g., diet, body weight, sedentary lifestyle, and exercise) may contribute to racial and ethnic disparities in the incidence and mortality of CRC (109–111).

3.8 Disparities of MSI-H/dMMR CRCs in hereditary vs. sporadic CRCs

Family history, age at disease onset, and prognosis are among the clinical characteristics that differ between patients with Lynch syndrome and those with sporadic MSI-H CRC. In contrast to patients with Lynch syndrome, those with sporadic MSI-H CRC often have no family history of CRC (45). Consistent with the role of germline mutations in hereditary colorectal syndromes, Lynch syndrome typically presents earlier in life than sporadic MSI-H CRCs (45, 112). Additionally, patients with stage I–III MSI CRC have a lower OS rate than those with Lynch syndrome; however, no significant differences in recurrence-free survival rates have been reported (113).

Histological and morphological differences also exist between sporadic MSI-H CRCs and Lynch syndrome. Sporadic MSI-H CRCs typically have cytoplasmic eosinophilia, and large, round, vesicular nuclei with a prominent nucleolus. In contrast, the cytological features of Lynch syndrome are similar to those of conventional adenomas (45). Furthermore, lymphocytic infiltration, tumor-cell de-differentiation, and presence of adenomas are more common in Lynch syndrome than in sporadic MSI-H CRCs. In contrast, mucin secretion, poor tumor differentiation, high intratumoral heterogeneity, Crohn's-like reaction, glandular serration, and the presence of serrated polyps are more frequent in sporadic MSI-H CRCs than in Lynch syndrome (45, 113, 114).

Genetic factors predisposing individuals to DNA methylation may contribute to differences in clinicopathological characteristics between hereditary and sporadic MSI-H CRCs (45). Methylation of the *MLH1* promoter and *BRAF* V600E mutations are frequently detected in sporadic MSI-H CRCs but

not in patients with Lynch syndrome, suggesting that testing for *BRAF* mutations and *MLH1* promoter methylation may help differentiate Lynch syndrome from sporadic MSI-H/dMMR CRC (115–117).

4 Conclusions

MSI CRCs possess distinct clinicopathological and molecular characteristics compared to MSS CRCs. In recent years, the development and exploration of novel testing technologies and methods, including NGS, AI-based histological algorithms, and image-based radiomic analysis, have shown promise for further defining and identifying this unique subgroup of CRCs. Combining multiple parameters with machine learning is a promising strategy to improve the performance of predictive models. The integration of histopathological and clinicopathological characteristics may improve the identification of patients with MSI-H/dMMR CRC. Progress in testing methods and predictive models has led to a deeper understanding of the disease and has important implications for patient management. The future development and utilization of these methods may hold promise for improving patient outcomes and for the development of novel therapeutics for patients with MSI-H/dMMR CRC.

Author contributions

All authors substantially contributed to planning, gathering, and interpreting the information or ideas used in the paper.

References

1. Kawakami H, Zaanen A, Sinicrope FA. Microsatellite instability testing and its role in the management of colorectal cancer. *Curr Treat Options Oncol* (2015) 16:30. doi: 10.1007/s11864-015-0348-2
2. Jin Z, Sinicrope FA. Prognostic and predictive values of mismatch repair deficiency in non-metastatic colorectal cancer. *Cancers (Basel)* (2021) 13:300. doi: 10.3390/cancers13020300
3. Ooki A, Shinozaki E, Yamaguchi K. Immunotherapy in colorectal cancer: Current and future strategies. *J Anus Rectum Colon* (2021) 5:11–24. doi: 10.23922/jarc.2020-064
4. Manz SM, Losa M, Fritsch R, Scharl M. Efficacy and side effects of immune checkpoint inhibitors in the treatment of colorectal cancer. *Therap Adv Gastroenterol* (2021) 14:17562848211002018. doi: 10.1177/17562848211002018
5. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med* (2015) 372:2509–20. doi: 10.1056/NEJMoa1500596
6. Overman MJ, Lonardi S, Wong KYM, Lenz HJ, Gelsomino F, Aglietta M, et al. Durable clinical benefit with nivolumab plus ipilimumab in DNA mismatch repair-Deficient/Microsatellite instability-high metastatic colorectal cancer. *J Clin Oncol* (2018) 36:773–9. doi: 10.1200/JCO.2017.76.9901
7. Andre T, Shiu KK, Kim TW, Jensen BV, Jensen LH, Punt C, et al. Pembrolizumab in microsatellite-Instability-High advanced colorectal cancer. *N Engl J Med* (2020) 383:2207–18. doi: 10.1056/NEJMoa2017699
8. Chalabi M, Fanchi LF, Dijkstra KK, Van den Berg JG, Aalbers AG, Sikorska K, et al. Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. *Nat Med* (2020) 26:566–76. doi: 10.1038/s41591-020-0805-8
9. Chalabi M, Verschoor YL, van den Berg J. Neoadjuvant immune checkpoint inhibition in locally advanced MMR-deficient colon cancer: The NICHE-2 study. *Ann Oncol* (2022) 33:S1389. doi: 10.1016/j.annonc.2022.08.016
10. Verschoor YL, Berg Jvd, Beets G, Sikorska K, Aalbers A, Lent Av, et al. Neoadjuvant nivolumab, ipilimumab, and celecoxib in MMR-proficient and MMR-deficient colon cancers: Final clinical analysis of the NICHE study. *J Clin Oncol* (2022) 40:3511–1. doi: 10.1200/JCO.2022.40.16_suppl.3511
11. Inamori K, Togashi Y, Bando H, Tsukada Y, Fukuoka S, Suzuki A, et al. Translational research of VOLTAGE-a: Efficacy predictors of preoperative chemoradiotherapy and consolidation nivolumab in patients with both microsatellite stable and microsatellite instability-high locally advanced rectal cancer. *J Clin Oncol* (2021) 39:100–0. doi: 10.1200/JCO.2021.39.3_suppl.100

During the process, they substantially contributed to providing suggestions for revision or critically reviewing subsequent iterations of the manuscript and ensured that questions related to the accuracy or integrity of any part of the work were appropriately investigated and resolved. Finally, they all reviewed and approved the final version of the paper.

Acknowledgments

Editorial assistance for this review article was provided by Christos Evangelou PhD and Jake Burrell PhD (Rude Health Consulting). This assistance was funded by MSD China.

Conflict of interest

Authors JQ and NX are employed by MSD China.

The remaining 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.

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.

12. Salvatore L, Bensi M, Corallo S, Bergamo F, Pellegrini I, Rasola C, et al. Phase II study of preoperative (PREOP) chemoradiotherapy (CTRT) plus avelumab (AVE) in patients (PTS) with locally advanced rectal cancer (LARC): The AVANA study. *J Clin Oncol* (2021) 39:3511–1. doi: 10.1200/JCO.2021.39.15_suppl.3511
13. George TJ, Yothers G, Hong TS, Russell MM, You YN, Parker W, et al. NRG-GI002: A phase II clinical trial platform using total neoadjuvant therapy (TNT) in locally advanced rectal cancer (LARC)—first experimental arm (EA) initial results. *J Clin Oncol* (2019) 37:3505–5. doi: 10.1200/JCO.2019.37.15_suppl.3505
14. Tambari S, Grassi E, Corbelli J, Papiani G, Barbera M. A., Zingaretti C, et al. A phase II study of capecitabine plus concomitant radiation therapy followed by durvalumab (MEDI4736) as preoperative treatment in rectal cancer: PANDORA study first-stage. *J Clin Oncol* (2021) 39:3607–7. doi: 10.1200/JCO.2021.39.15_suppl.3607
15. Shaikh T, Handorf EA, Meyer JE, Hall MJ, Esnaola NF. Mismatch repair deficiency testing in patients with colorectal cancer and nonadherence to testing guidelines in young adults. *JAMA Oncol* (2018) 4:e173580. doi: 10.1001/jamaoncol.2017.3580
16. Thiebault Q, Defossez G, Karayan-Tapon L, Ingrand P, Silvain C, Tougeron D. Analysis of factors influencing molecular testing at diagnostic of colorectal cancer. *BMC Cancer* (2017) 17:765. doi: 10.1186/s12885-017-3759-6
17. Evrard C, Tachon G, Randrian V, Karayan-Tapon L, Tougeron D. Microsatellite instability: Diagnosis, heterogeneity, discordance, and clinical impact in colorectal cancer. *Cancers (Basel)* (2019) 11:1567. doi: 10.3390/cancers11101567
18. Kanth P, Grimmett J, Champine M, Burt R, Samadder NJ. Hereditary colorectal polyposis and cancer syndromes: A primer on diagnosis and management. *Am J Gastroenterol* (2017) 112:1509–25. doi: 10.1038/ajg.2017.212
19. McCabe M, Perner Y, Magobo R, Magangane P, Mirza S, Penny C. Microsatellite instability assessment in black south African colorectal cancer patients reveal an increased incidence of suspected lynch syndrome. *Sci Rep* (2019) 9:15019. doi: 10.1038/s41598-019-51316-4
20. Rauluseviute I, Drablos F, Rye MB. DNA Hypermethylation associated with upregulated gene expression in prostate cancer demonstrates the diversity of epigenetic regulation. *BMC Med Genomics* (2020) 13:6. doi: 10.1186/s12920-020-0657-6
21. Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* (2014) 26:577–90. doi: 10.1016/j.ccr.2014.07.028
22. Li J, Xu Q, Luo C, Chen L, Ying J. Clinicopathologic characteristics of resectable colorectal cancer with mismatch repair protein defects in Chinese population: Retrospective case series and literature review. *Med (Baltimore)* (2020) 99:e20554. doi: 10.1097/MD.00000000000020554
23. Liang Y, Cai X, Zheng X, Yin H. Analysis of the clinicopathological characteristics of stage I–III colorectal cancer patients deficient in mismatch repair proteins. *Onco Targets Ther* (2021) 14:2203–12. doi: 10.2147/OTT.S278029
24. Yamada A, Matsuoka Y, Minamiguchi S, Yamamoto Y, Kondo T, Sunami T, et al. Real-world outcome of universal screening for lynch syndrome in Japanese patients with colorectal cancer highlights the importance of targeting patients with young-onset disease. *Mol Clin Oncol* (2021) 15:247. doi: 10.3892/mco.2021.2409
25. Watson P, Vasen HFA, Mecklin JP, Bernstein I, Aarnio M, Jarvinen HJ, et al. The risk of extra-colonic, extra-endometrial cancer in the lynch syndrome. *Int J Cancer* (2008) 123:444–9. doi: 10.1002/ijc.23508
26. Ye JX, Liu Y, Qin Y, Zhong HH, Yi WN, Shi XY. KRAS and BRAF gene mutations and DNA mismatch repair status in Chinese colorectal carcinoma patients. *World J Gastroenterol* (2015) 21:1595–605. doi: 10.3748/wjg.v21.i5.1595
27. Viñal D, Martínez-Pérez D, Martínez-Recio S, Ruiz I, Jiménez Bou D, Peña J, et al. Clinicopathological characteristics and outcomes of patients with deficient mismatch repair colorectal cancer. *J Clin Oncol* (2022) 40:181–1. doi: 10.1200/JCO.2022.40.4_suppl.181
28. Guo TA, Wu YC, Tan C, Jin YT, Sheng WQ, Cai SJ, et al. Clinicopathologic features and prognostic value of KRAS, NRAS and BRAF mutations and DNA mismatch repair status: A single-center retrospective study of 1,834 Chinese patients with stage I–IV colorectal cancer. *Int J Cancer* (2019) 145:1625–34. doi: 10.1002/ijc.32489
29. Kang S, Na Y, Joung SY, Lee SI, Oh SC, Min BW. The significance of microsatellite instability in colorectal cancer after controlling for clinicopathological factors. *Med (Baltimore)* (2018) 97:e0019. doi: 10.1097/MD.00000000000010019
30. Zheng J, Huang B, Nie X, Zhu Y, Han N, Li Y. The clinicopathological features and prognosis of tumor MSI in East Asian colorectal cancer patients using NCI panel. *Future Oncol* (2018) 14:1355–64. doi: 10.2217/fon-2017-0662
31. Han K, Tang JH, Liao LE, Jiang W, Sui QQ, Xiao BY, et al. Neoadjuvant immune checkpoint inhibition improves organ preservation in T4bM0 colorectal cancer with mismatch repair deficiency: A retrospective observational study. *Dis Colon Rectum* (2022). doi: 10.1097/DCR.0000000000002466
32. Yan WY, Hu J, Xie L, Cheng L, Yang M, Li L, et al. Prediction of biological behavior and prognosis of colorectal cancer patients by tumor MSI/MMR in the Chinese population. *Onco Targets Ther* (2016) 9:7415–24. doi: 10.2147/OTT.S117089
33. Chen J, Zhou L, Gao J, Lu T, Wang J, Wu H, et al. Clinicopathological characteristics and mutation spectrum of colorectal adenocarcinoma with mucinous component in a Chinese cohort: Comparison with classical adenocarcinoma. *Front Oncol* (2020) 10:917. doi: 10.3389/fonc.2020.00917
34. Sacdalan DL, Garcia RL, Diwa MH, Sacdalan DB. Clinicopathologic factors associated with mismatch repair status among Filipino patients with young-onset colorectal cancer. *Cancer Manag Res* (2021) 13:2105–15. doi: 10.2147/CMAR.S286618
35. Tuppurainen K, Makinen JM, Junttila O, Liakka A, Kyllonen AP, Tuominen H, et al. Morphology and microsatellite instability in sporadic serrated and non-serrated colorectal cancer. *J Pathol* (2005) 207:285–94. doi: 10.1002/path.1850
36. Garcia-Solano J, Conesa-Zamora P, Carbonell P, Trujillo-Santos J, Torres-Moreno DD, Pagan-Gomez I, et al. Colorectal serrated adenocarcinoma shows a different profile of oncogene mutations, MSI status and DNA repair protein expression compared to conventional and sporadic MSI-h carcinomas. *Int J Cancer* (2012) 131:1790–9. doi: 10.1002/ijc.27454
37. Echle A, Laleh NG, Schrammen PL, West NP, Trautwein C, Brinker TJ, et al. Deep learning for the detection of microsatellite instability from histology images in colorectal cancer: A systematic literature review. *Immunoinformatics* (2021) 3–4:100008. doi: 10.1016/j.immuno.2021.100008
38. Shia J, Schultz N, Kuk D, Vakiani E, Middha S, Segal NH, et al. Morphological characterization of colorectal cancers in the cancer genome atlas reveals distinct morphology-molecular associations: clinical and biological implications. *Mod Pathol* (2017) 30:599–609. doi: 10.1038/modpathol.2016.198
39. Greenon JK, Huang SC, Herron C, Moreno V, Bonner JD, Tomsho LP, et al. Pathologic predictors of microsatellite instability in colorectal cancer. *Am J Surg Pathol* (2009) 33:126–33. doi: 10.1097/PAS.0b013e31817ec2b1
40. Lizardo DY, Kuang C, Hao S, Yu J, Huang Y, Zhang L. Immunotherapy efficacy on mismatch repair-deficient colorectal cancer: From bench to bedside. *Biochim Biophys Acta Rev Cancer* (2020) 1874:188447. doi: 10.1016/j.bbcan.2020.188447
41. Sahin IH, Akce M, Alese O, Shaib W, Lesinski GB, El-Rayes B, et al. Immune checkpoint inhibitors for the treatment of MSI-H/MMR-D colorectal cancer and a perspective on resistance mechanisms. *Br J Cancer* (2019) 121:809–18. doi: 10.1038/s41416-019-0599-y
42. Gallois C, Taieb J, Sabouret A, Broudin C, Karoui M, Garinet S, et al. Upfront progression under pembrolizumab followed by a complete response after encorafenib and cetuximab treatment in BRAF V600E-mutated and microsatellite unstable metastatic colorectal cancer patient: A case report. *Genes Chromosomes Cancer* (2022) 61:114–8. doi: 10.1002/gcc.23012
43. Wu W, Liu Y, Zeng S, Han Y, Shen H. Intratumor heterogeneity: The hidden barrier to immunotherapy against MSI tumors from the perspective of IFN-gamma signaling and tumor-infiltrating lymphocytes. *J Hematol Oncol* (2021) 14:160. doi: 10.1186/s13045-021-01166-3
44. Cui G. The mechanisms leading to distinct responses to PD-1/PD-L1 blockades in colorectal cancers with different MSI statuses. *Front Oncol* (2021) 11:573547. doi: 10.3389/fonc.2021.573547
45. Jass JR. HNPCC and sporadic MSI-h colorectal cancer: A review of the morphological similarities and differences. *Fam Cancer* (2004) 3:93–100. doi: 10.1023/B:FAME.0000039849.86008.b7
46. Paredes J, Zabaleta J, Garai J, Ji P, Imtiaz S, Spagnardi M, et al. Immune-related gene expression and cytokine secretion is reduced among African American colon cancer patients. *Front Oncol* (2020) 10:1498. doi: 10.3389/fonc.2020.01498
47. Kim JH, Kang GH. Molecular and prognostic heterogeneity of microsatellite-unstable colorectal cancer. *World J Gastroenterol* (2014) 20:4230–43. doi: 10.3748/wjg.v20.i15.4230
48. Thakur N, Yoon H, Chong Y. Current trends of artificial intelligence for colorectal cancer pathology image analysis: A systematic review. *Cancers (Basel)* (2020) 12:1884. doi: 10.3390/cancers12071884
49. Jiang W, Mei WJ, Xu SY, Ling YH, Li WR, Kuang JB, et al. Clinical actionability of triaging DNA mismatch repair deficient colorectal cancer from biopsy samples using deep learning. *EBioMedicine* (2022) 81:104120. doi: 10.1016/j.ebiom.2022.104120
50. Echle A, Ghaffari Laleh N, Quirke P, Grabsch HI, Muti HS, Saldanha OL, et al. Artificial intelligence for detection of microsatellite instability in colorectal cancer—a multicentric analysis of a pre-screening tool for clinical application. *ESMO Open* (2022) 7:100400. doi: 10.1016/j.esmoop.2022.100400

51. Cao R, Yang F, Ma SC, Liu L, Zhao Y, Li Y, et al. Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in colorectal cancer. *Theranostics* (2020) 10:11080–91. doi: 10.7150/thno.49864
52. Saillard C, Dehaene O, Marchand T, Moindrot O, Kamoun A, Schmauch B, et al. Self supervised learning improves dMMR/MSI detection from histology slides across multiple cancers. *MICCAI 2021 COMPAY* (2021).
53. Shimada Y, Okuda S, Watanabe Y, Tajima Y, Nagahashi M, Ichikawa H, et al. Histopathological characteristics and artificial intelligence for predicting tumor mutation burden-high colorectal cancer. *J Gastroenterol* (2021) 56:547–59. doi: 10.1007/s00535-021-01789-w
54. Zhao N, Cao Y, Yang J, Li H, Wu K, Wang J, et al. Serum tumor markers combined with clinicopathological characteristics for predicting MMR and KRAS status in 2279 Chinese colorectal cancer patients: A retrospective analysis. *Front Oncol* (2021) 11:582244. doi: 10.3389/fonc.2021.582244
55. Cao Y, Peng T, Li H, Yang M, Wu L, Zhou Z, et al. Development and validation of MMR prediction model based on simplified clinicopathological features and serum tumour markers. *EBioMedicine* (2020) 61:103060. doi: 10.1016/j.ebiom.2020.103060
56. Fan S, Li X, Cui X, Zheng L, Ren X, Ma W, et al. Computed tomography-based radiomic features could potentially predict microsatellite instability status in stage II colorectal cancer: A preliminary study. *Acad Radiol* (2019) 26:1633–40. doi: 10.1016/j.acra.2019.02.009
57. Cao Y, Zhang G, Zhang J, Yang Y, Ren J, Yan X, et al. Predicting microsatellite instability status in colorectal cancer based on triphasic enhanced computed tomography radiomics signatures: A multicenter study. *Front Oncol* (2021) 11:687771. doi: 10.3389/fonc.2021.687771
58. Yuan H, Yu P, Li J, Song N, Wan Z. A., Chen B, et al. An integrating of clinical, pathological, and radiomic approach to predict the microsatellite instability in rectal carcinoma. (2021). doi: 10.21203/rs.3.rs-957494/v1
59. Ma Y, Yuan H, Yu P, Xu X, Tu S, Wei Y. A tumoral and peritumoral CT-based radiomics and machine learning approach to predict the microsatellite instability of rectal carcinoma. *Res Square* (2022) 14:2409–18. doi: 10.21203/rs.3.rs-1305223/v1
60. Zhang W, Huang Z, Zhao J, He D, Li M, Yin H, et al. Development and validation of magnetic resonance imaging-based radiomics models for preoperative prediction of microsatellite instability in rectal cancer. *Ann Transl Med* (2021) 9:134. doi: 10.21037/atm-20-7673
61. Park JH, Kim EY, Luchini C, Eccher A, Tizaoui K, Shin JI, et al. Artificial intelligence for predicting microsatellite instability based on tumor histomorphology: A systematic review. *Int J Mol Sci* (2022) 23:2462. doi: 10.3390/ijms23052462
62. Kacew AJ, Strohbehn GW, Saulsberry L, Laiterapong N, Cipriani NA, Kather JN, et al. Artificial intelligence can cut costs while maintaining accuracy in colorectal cancer genotyping. *Front Oncol* (2021) 11:630953. doi: 10.3389/fonc.2021.630953
63. Chang SC, Lin JK, Yang SH, Wang HS, Li AF, Chi CW. Relationship between genetic alterations and prognosis in sporadic colorectal cancer. *Int J Cancer* (2006) 118:1721–7. doi: 10.1002/ijc.21563
64. Landau MS, Kuan SF, Chiosea S, Pai RK. BRAF-mutated microsatellite stable colorectal carcinoma: An aggressive adenocarcinoma with reduced CDX2 and increased cytokeratin 7 immunohistochemical expression. *Hum Pathol* (2014) 45:1704–12. doi: 10.1016/j.humpath.2014.04.008
65. Gatalica Z, Vijayvergia N, Vranic S, Xiu J, Reddy SK, Snyder CL, et al. Therapeutic biomarker differences between MSI-h and MSS colorectal cancers. *J Clin Oncol* (2015) 33:3597–7. doi: 10.1200/jco.2015.33.15_suppl.3597
66. Xiao J, Li W, Huang Y, Huang M, Li S, Zhai X, et al. A next-generation sequencing-based strategy combining microsatellite instability and tumor mutation burden for comprehensive molecular diagnosis of advanced colorectal cancer. *BMC Cancer* (2021) 21:282. doi: 10.1186/s12885-021-07942-1
67. Maloberti T, De Leo A, Sanza V, Merlo L, Visani M, Acquaviva G, et al. BRAF and MLH1 analysis algorithm for the evaluation of lynch syndrome risk in colorectal carcinoma patients: Evidence-based data from the analysis of 100 consecutive cases. *J Mol Pathol* (2022) 3:115–24. doi: 10.3390/jmp3030011
68. Capper D, Voigt A, Bozokova G, Ahadova A, Kickingeder P, von Deimling A, et al. BRAF V600E-specific immunohistochemistry for the exclusion of lynch syndrome in MSI-h colorectal cancer. *Int J Cancer* (2013) 133:1624–30. doi: 10.1002/ijc.28183
69. Toon CW, Walsh MD, Chou A, Capper D, Clarkson A, Sioson L, et al. BRAFV600E immunohistochemistry facilitates universal screening of colorectal cancers for lynch syndrome. *Am J Surg Pathol* (2013) 37:1592–602. doi: 10.1097/PAS.0b013e31828f233d
70. Mahesh SA, Hanna E, Khan MS, Ravichandran P, Slezak F. Incidence and characteristics of BRAF V600E mutation in colorectal cancer (CRC) with mismatch repair (MMR) protein defect due to loss of MLH1: A prospective evaluation of 104 consecutive patients. *J Clin Oncol* (2011) 29:3539–9. doi: 10.1200/jco.2011.29.15_suppl.3539
71. Sahin IH, Goyal S, Pumpalova Y, Sonbol MB, Das S, Haraldsdottir S, et al. Mismatch repair (MMR) gene alteration and BRAF V600E mutation are potential predictive biomarkers of immune checkpoint inhibitors in MMR-deficient colorectal cancer. *Oncologist* (2021) 26:668–75. doi: 10.1002/onco.13741
72. Tan E, Whiting J, Xie H, Imanirad I, Carballido E, Felder S, et al. BRAF mutations are associated with poor survival outcomes in advanced-stage mismatch repair-deficient/Microsatellite high colorectal cancer. *Oncologist* (2022) 27:191–7. doi: 10.1093/oncolo/oyab055
73. Vilar E, Gruber SB. Microsatellite instability in colorectal cancer—the stable evidence. *Nat Rev Clin Oncol* (2010) 7:153–62. doi: 10.1038/nrclinonc.2009.237
74. Zhou Z, Li K, Wei Q, Chen L, Shuai Y, Wang Y, et al. Tumor mutation burden determined by a 645-cancer gene panel and compared with microsatellite instability and mismatch repair genes in colorectal cancer. *J Gastrointest Oncol* (2021) 12:2775–87. doi: 10.21037/jgo-21-572
75. Zheng K, Wan H, Zhang J, Shan G, Chai N, Li D, et al. A novel NGS-based microsatellite instability (MSI) status classifier with 9 loci for colorectal cancer patients. *J Transl Med* (2020) 18:215. doi: 10.1186/s12967-020-02373-1
76. Wu J, Zhang Q, Zhao Y, Liu Y, Chen A, Li X, et al. Radiomics analysis of iodine-based material decomposition images with dual-energy computed tomography imaging for preoperatively predicting microsatellite instability status in colorectal cancer. *Front Oncol* (2019) 9:1250. doi: 10.3389/fonc.2019.01250
77. Pei Q, Yi X, Chen C, Pang P, Fu Y, Lei G, et al. Pre-treatment CT-based radiomics nomogram for predicting microsatellite instability status in colorectal cancer. *Eur Radiol* (2022) 32:714–24. doi: 10.1007/s00330-021-08167-3
78. Li J, Yang Z, Xin B, Hao Y, Wang L, Song S, et al. Quantitative prediction of microsatellite instability in colorectal cancer with preoperative PET/CT-based radiomics. *Front Oncol* (2021) 11:702055. doi: 10.3389/fonc.2021.702055
79. Liu H, Ye Z, Yang T, Xie H, Duan T, Li M, et al. Predictive value of metabolic parameters derived from (18)F-FDG PET/CT for microsatellite instability in patients with colorectal carcinoma. *Front Immunol* (2021) 12:724464. doi: 10.3389/fimmu.2021.724464
80. Soliman AS, Bondy ML, El-Badawy SA, Mokhtar N, Eissa S, Bayoumy S, et al. Contrasting molecular pathology of colorectal carcinoma in Egyptian and Western patients. *Br J Cancer* (2001) 85:1037–46. doi: 10.1054/bjoc.2001.1838
81. Lungulescu C, Croitoru VM, Volovat SR, Cazacu IM, Turcu-Stolica A, Gheonea DI, et al. An insight into deficient mismatch repair colorectal cancer screening in a Romanian population—a bi-institutional pilot study. *Medicina (Kaunas)* (2021) 57:847. doi: 10.3390/medicina57080847
82. Noepel-Duenebecke S, Juethe H, Feder IS, Kluxen L, Basara N, Hiller W, et al. High microsatellite instability (MSI-h) is associated with distinct clinical and molecular characteristics and an improved survival in early colon cancer (CC); real world data from the AIO molecular registry colopredict plus. *Z Gastroenterol* (2020) 58:533–41. doi: 10.1055/a-1156-4433
83. Aasebo KO, Dragomir A, Sundstrom M, Mezheyeuski A, Edqvist PH, Eide GE, et al. Consequences of a high incidence of microsatellite instability in Egyptian and BRAF-mutated tumors: A population-based cohort of metastatic colorectal cancer patients. *Cancer Med* (2019) 8:3623–35. doi: 10.1002/cam4.2205
84. Koopman M, Kortman GA, Mekenkamp L, Ligtnerberg MJ, Hoogerbrugge N, Antonini NF, et al. Deficient mismatch repair system in patients with sporadic advanced colorectal cancer. *Br J Cancer* (2009) 100:266–73. doi: 10.1038/sj.bjc.6604867
85. Loughrey MB, McGrath J, Coleman HG, Bankhead P, Maxwell P, McGready C, et al. Identifying mismatch repair-deficient colon cancer: near-perfect concordance between immunohistochemistry and microsatellite instability testing in a large, population-based series. *Histopathology* (2021) 78:401–13. doi: 10.1111/his.14233
86. Kumar K, Brim H, Giardiello F, Smoot DT, Nouraei M, Lee EL, et al. (V600E) and KRAS mutations in high microsatellite instability sporadic colorectal cancer in African americans. *Clin Cancer Res* (2009) 15:1155–61. doi: 10.1158/1078-0432.CCR-08-1029
87. Ashktorab H, Ahuja S, Kannan L, Llor X, Ellis NA, Xicola RM, et al. A meta-analysis of MSI frequency and race in colorectal cancer. *Oncotarget* (2016) 7:34546–57. doi: 10.18632/oncotarget.8945
88. Klaassen ZWA, Waller J, Gu L, De Hoedt AM, Wang T, Amonkar M, et al. DNA Mismatch repair and microsatellite instability in colorectal tumors: An observational study in the veterans affairs health care system. *J Clin Oncol* (2021) 39:e15562–2. doi: 10.1200/JCO.2021.39.15_suppl.e15562
89. Ricker CN, Hanna DL, Peng C, Nguyen NT, Stern MC, Schmit SL, et al. DNA Mismatch repair deficiency and hereditary syndromes in Latino patients with colorectal cancer. *Cancer* (2017) 123:3732–43. doi: 10.1002/cnrc.30790
90. Rios-Valencia J, Cruz-Reyes C, Galindo-Garcia TA, Rosas-Camargo V, Gamboa-Dominguez A. Mismatch repair system in colorectal cancer: frequency, cancer phenotype, and follow-up. *Rev Gastroenterol Mex (Engl Ed)* (2022) 87:432–8. doi: 10.1016/j.rgmex.2022.05.017

91. Lopez-Correa PE, Lino-Silva LS, Gamboa-Dominguez A, Zepeda-Najar C, Salcedo-Hernandez RA. Frequency of defective mismatch repair system in a series of consecutive cases of colorectal cancer in a national cancer center. *J Gastrointest Cancer* (2018) 49:379–84. doi: 10.1007/s12029-018-0132-1
92. Akagi K, Oki E, Taniguchi H, Nakatani K, Aoki D, Kuwata T, et al. Real-world data on microsatellite instability status in various unresectable or metastatic solid tumors. *Cancer Sci* (2021) 112:1105–13. doi: 10.1111/cas.14798
93. Lorenzi M, Amonkar M, Zhang J, Mehta S, Liaw K-L. Epidemiology of microsatellite instability high (MSI-h) and deficient mismatch repair (dMMR) in solid tumors: A structured literature review. *J Oncol* (2020) 2020:1807929. doi: 10.1155/2020/1807929
94. Zhang C, Ding H, Sun S, Luan Z, Liu G, Li Z. Incidence and detection of high microsatellite instability in colorectal cancer in a Chinese population: a meta-analysis. *J Gastrointest Oncol* (2020) 11:1155–63. doi: 10.21037/jgo-20-487
95. Jiang W, Sui QQ, Li WL, Ke CF, Ling YH, Liao LE, et al. Low prevalence of mismatch repair deficiency in Chinese colorectal cancers: a multicenter study. *Gastroenterol Rep (Oxf)* (2020) 8:399–403. doi: 10.1093/gastro/goaa006
96. Jiang W, Cai MY, Li SY, Bei JX, Wang F, Hampel H, et al. Universal screening for lynch syndrome in a large consecutive cohort of Chinese colorectal cancer patients: High prevalence and unique molecular features. *Int J Cancer* (2019) 144:2161–8. doi: 10.1002/ijc.32044
97. Choi YY. Genomic analysis for discovering genetic alterations in young Korean patients with double primary cancers of the stomach and colon, graduate school, yonsei university. (2019).
98. Vg DR, Nayanar SK, Gopinath V, Philip KJ, Aryadan N, Nair V, et al. Testing for microsatellite instability in colorectal cancer - a comparative evaluation of immunohistochemical and molecular methods. *Gulf J Oncolog* (2022) 1:70–8.
99. Ostwal V, Pande NS, Engineer R, Saklani A, deSouza A, Ramadwar M, et al. Low prevalence of deficient mismatch repair (dMMR) protein in locally advanced rectal cancers (LARC) and treatment outcomes. *J Gastrointest Oncol* (2019) 10:19–29. doi: 10.21037/jgo.2018.10.01
100. Paulose RR, Ail DA, Biradar S, Vasudevan A, Sundaram KR. Prognostic and predictive significance of microsatellite instability in stage II colorectal carcinoma: An 8-year study from a tertiary center in south India. *Indian J Cancer* (2019) 56:302–8. doi: 10.4103/ijc.IJC_365_18
101. Rai PR, Shetty N, Rai PR, Shet D, Shetty A. A study on the frequency and clinicopathological correlates of mismatch repair-deficient colorectal cancer. *J Cancer Res Ther* (2020) 16:S183–8. doi: 10.4103/jcrt.JCRT_526_18
102. 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 Cancer J Clin* (2021) 71:209–49. doi: 10.3322/caac.21660
103. Wong MCS, Huang J, Lok V, Wang J, Fung F, Ding H, et al. Differences in incidence and mortality trends of colorectal cancer worldwide based on sex, age, and anatomic location. *Clin Gastroenterol Hepatol* (2021) 19:955–66.e61. doi: 10.1016/j.cgh.2020.02.026
104. Wong MC, Ding H, Wang J, Chan PS, Huang J. Prevalence and risk factors of colorectal cancer in Asia. *Intest Res* (2019) 17:317–29. doi: 10.5217/ir.2019.00021
105. Sougklakos I, Athanasiadis E, Boukovinas I, Karamouzis M, Koutras A, Papakotoulas P, et al. Treatment pathways and associated costs of metastatic colorectal cancer in Greece. *Cost Eff Resour Alloc* (2022) 20:7. doi: 10.1186/s12962-022-00339-2
106. Nakano K, Yamamoto H, Fujiwara M, Koga Y, Tsuruta S, Ihara E, et al. Clinicopathologic and molecular characteristics of synchronous colorectal carcinoma with mismatch repair deficiency. *Am J Surg Pathol* (2018) 42:172–82. doi: 10.1097/PAS.0000000000000947
107. Jackson CS, Oman M, Patel AM, Vega KJ. Health disparities in colorectal cancer among racial and ethnic minorities in the united states. *J Gastrointest Oncol* (2016) 7:S32–43. doi: 10.3978/j.issn.2078-6891.2015.039
108. Petrick JL, Barber LE, Warren Andersen S, Florio AA, Palmer JR, Rosenberg L. Racial disparities and sex differences in early- and late-onset colorectal cancer incidence, 2001–2018. *Front Oncol* (2021) 11:734998. doi: 10.3389/fonc.2021.734998
109. Zhang J, Zheng J, Yang Y, Lu J, Gao J, Lu T, et al. Molecular spectrum of KRAS, NRAS, BRAF and PIK3CA mutations in Chinese colorectal cancer patients: analysis of 1,110 cases. *Sci Rep* (2015) 5:18678. doi: 10.1038/srep18678
110. Ollberding NJ, Nomura AM, Wilkens LR, Henderson BE, Kolonel LN. Racial/ethnic differences in colorectal cancer risk: The multiethnic cohort study. *Int J Cancer* (2011) 129:1899–906. doi: 10.1002/ijc.25822
111. Carethers JM, Doubeni CA. Causes of socioeconomic disparities in colorectal cancer and intervention framework and strategies. *Gastroenterology* (2020) 158:354–67. doi: 10.1053/j.gastro.2019.10.029
112. Son IT, Kim DW, Kim MH, Shin YK, Ku JL, Oh HK, et al. Comparison of oncologic outcomes between patients with lynch syndrome and sporadic microsatellite instability-high colorectal cancer. *Ann Surg Treat Res* (2021) 101:13–9. doi: 10.4174/ast.2021.101.1.13
113. Nakayama Y, Iijima T, Inokuchi T, Kojika E, Takao M, Takao A, et al. Clinicopathological features of sporadic MSI colorectal cancer and lynch syndrome: a single-center retrospective cohort study. *Int J Clin Oncol* (2021) 26:1881–9. doi: 10.1007/s10147-021-01968-y
114. Peng J, Xiao-ming M, Jian-qiu S, Zi-tao W, Lei F, He-juan A, et al. Clinicopathological features of non-familial colorectal cancer with high-frequency microsatellite instability. *Chin Med Sci J* (2010) 25:228–32. doi: 10.1016/S1001-9294(11)60007-2
115. Leclerc J, Vermaut C, Buisine MP. Diagnosis of lynch syndrome and strategies to distinguish lynch-related tumors from sporadic MSI/dMMR tumors. *Cancers (Basel)* (2021) 13:467. doi: 10.3390/cancers13030467
116. Sato K, Kawazu M, Yamamoto Y, Ueno T, Kojima S, Nagae G, et al. Fusion kinases identified by genomic analyses of sporadic microsatellite instability-high colorectal cancers. *Clin Cancer Res* (2019) 25:378–89. doi: 10.1158/1078-0432.CCR-18-1574
117. Cohen R, Buhard O, Cervera P, Hain E, Dumont S, Bardier A, et al. Clinical and molecular characterisation of hereditary and sporadic metastatic colorectal cancers harbouring microsatellite instability/DNA mismatch repair deficiency. *Eur J Cancer* (2017) 86:266–74. doi: 10.1016/j.ejca.2017.09.022

COPYRIGHT

© 2022 Mei, Mi, Qian, Xiao, Yuan and Ding. 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.



OPEN ACCESS

EDITED BY

Apostolos Zaravinos,
European University Cyprus, Cyprus

REVIEWED BY

Jun Li,
The University of Sydney, Australia
Xianhuo Wang,
Tianjin Medical University Cancer Institute
and Hospital, China

*CORRESPONDENCE

Min Yang
✉ minyang@imm.ac.cn

SPECIALTY SECTION

This article was submitted to
Cancer Immunity
and Immunotherapy,
a section of the journal
Frontiers in Immunology

RECEIVED 27 September 2022

ACCEPTED 09 February 2023

PUBLISHED 22 February 2023

CITATION

Zhang R, Gan W, Zong J, Hou Y, Zhou M,
Yan Z, Li T, Lv S, Zeng Z, Wang W, Zhang F
and Yang M (2023) Developing an m5C
regulator-mediated RNA methylation
modification signature to predict
prognosis and immunotherapy
efficacy in rectal cancer.
Front. Immunol. 14:1054700.
doi: 10.3389/fimmu.2023.1054700

COPYRIGHT

© 2023 Zhang, Gan, Zong, Hou, Zhou, Yan,
Li, Lv, Zeng, Wang, Zhang and Yang. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). 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.

Developing an m5C regulator–mediated RNA methylation modification signature to predict prognosis and immunotherapy efficacy in rectal cancer

Rixin Zhang¹, Wenqiang Gan¹, Jinbao Zong^{2,3}, Yufang Hou¹,
Mingxuan Zhou¹, Zheng Yan¹, Tiegang Li¹, Silin Lv¹, Zifan Zeng¹,
Weiqi Wang¹, Fang Zhang¹ and Min Yang^{1*}

¹State Key Laboratory of Bioactive Substances and Function of Natural Medicine, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China,

²Clinical Laboratory, The Affiliated Hospital of Qingdao University, Qingdao, China, ³Qingdao Hospital of Traditional Chinese Medicine, The Affiliated Qingdao Hiser Hospital of Qingdao University, Qingdao, China

Background: Currently, a very small number of patients with colorectal cancer (CRC) respond to immune checkpoint inhibitor (ICI) treatment. Therefore, there is an urgent need to investigate effective biomarkers to determine the responsiveness to ICI treatment. Recently, aberrant 5-methylcytosine (m⁵C) RNA modification has emerged as a key player in the pathogenesis of cancer. Thus, we aimed to explore the predictive signature based on m⁵C regulator-related genes for characterizing the immune landscapes and predicting the prognosis and response to therapies.

Methods: The Cancer Genome Atlas (TCGA) cohort was used as the training set, while GEO data sets, real-time quantitative PCR (RT-qPCR) analysis from paired frozen tissues, and immunohistochemistry (IHC) data from tissue microarray (TMA) were used for validation. We constructed a novel signature based on three m⁵C regulator-related genes in patients with rectal adenocarcinoma (READ) using a least absolute shrinkage and selection operator (LASSO)-Cox regression and unsupervised consensus clustering analyses. Additionally, we correlated the three-gene signature risk model with the tumor immune microenvironment, immunotherapy efficiency, and potential applicable drugs.

Results: The m⁵C methylation-based signature was an independent prognostic factor, where low-risk patients showed a stronger immunoreactivity phenotype and a superior response to ICI therapy. Conversely, the high-risk patients had enriched pathways of cancer hallmarks and presented immune-suppressive state, which demonstrated that they are more insensitive to immunotherapy. Additionally, the signature markedly correlated with drug susceptibility.

Conclusions: We developed a reliable m⁵C regulator-based risk model to predict the prognosis, clarify the molecular and tumor microenvironment

status, and identify patients who would benefit from immunotherapy or chemotherapy. Our study could provide vital guidance to improve prognostic stratification and optimize personalized therapeutic strategies for patients with rectal cancer.

KEYWORDS

rectal cancer, prognosis, tumor immune microenvironment, immunotherapy, m⁵C RNA methylation regulator

Introduction

By blocking programmed cell death 1/programmed cell death ligand 1 (PD1/PDL1) axis, immune checkpoint inhibitors (ICIs) have introduced a new era of antitumor therapy that could elicit durable responses and significantly improve survival in several tumors (1, 2). However, the contexture and organization of the immune environment can be highly heterogeneous among tumors, even within the same cancer type, leading to a complex crosstalk within the tumor immune microenvironment (TIME) (3). The overall status of tumor-infiltrating lymphocytes (TILs) in TIME closely correlates with the efficacy of immunotherapy. According to the immune cell status in TIME, tumor immune infiltration pattern could be broadly classified into “hot tumor” (indicating presence of CD8+ and CD4+ T cells accompanied by high expression of immune checkpoint molecules) and “cold tumor” (representing the deficiency of immune cells within the tumor parenchyma) (4, 5). The former has a potential antitumor efficacy, while the latter barely benefits from the ICI therapy (6). At present, patients with deficient mismatch repair (dMMR)/microsatellite instability-high (MSI-H) have more immune cell infiltration accompanied by high tumor mutational burden (TMB), while microsatellite stable (MSS)/microsatellite instability-low (MSI-L) patients have low abundance of TILs and low TMB (7, 8). Moreover, according to the KEYNOTE-016 study, 62% of colorectal cancer (CRC) patients with MSI-H phenotype achieve an objective response, while patients with MSS/MSI-L tumors cannot achieve objective response, indicating a better efficacy of immunotherapy in patients with dMMR/MSI-H tumors (9). Nonetheless, dMMR/MSI-H tumors account for only 15% of all patients with CRC (7, 10). Therefore, establishing effective predictive biomarkers is essential for the improvement of immunotherapeutic strategy.

RNA modification plays an important role in the regulation of gene expression. More than 150 RNA modifications containing N⁶-methyladenosine (m⁶A), 5-methylcytosine (m⁵C), and N¹-methiadenosine (m¹A) have been investigated (11, 12). Among these modifications, m⁵C is one of the most intensively researched epigenetic modifications, and overall, 95391 m⁵C sites in the human genome have been identified (13). The m⁵C methylation landscape is regulated by a dynamic process that integrates methyltransferases (“writer”), binding proteins (“readers”), and demethylases (“erasers”) (14, 15). Although m⁵C is widely recognized for its

essential function as an epigenetic marker for DNA, research into its functional roles in RNA is beginning to emerge. It has been shown that a vast majority of azacitidine (5-AZA), widely used to treat hematologic malignancies, is incorporated into RNA instead of DNA of treated tumor cells (16). Therefore, the potential use of m⁵C RNA modification as a novel therapeutic target for various types of cancers is a current topic of research.

RNA methylation impacts the efficacy of tumor immunotherapy by modulating immune activity in a range of tumors (17). Recently, several studies have uncovered the close relationship between TIME-infiltrating immune cells and m⁵C RNA methylation. Pan et al. found that NOP2/Sun RNA methyltransferase 4 (NSUN4) and NOP2/Sun RNA methyltransferase 3 (NSUN3) were closely related to the infiltration by six major immune cells that could regulate TIME in lung squamous cell carcinoma (18). Gao et al. showed that m⁵C RNA modification patterns could predict and affect TIME in oral squamous cell carcinoma (19). Despite these facts, the relationship between RNA methylation and tumor immunotherapy is still in its infancy. In the current study, we integrated multiple data sets and developed a novel signature based on the expression of m⁵C RNA methylation regulators, which could be used to evaluate risk status and predict prognosis of patients with rectal adenocarcinoma. Furthermore, we comprehensively explored the correlations between the m⁵C RNA methylation regulator-based signature having immune characteristics, mutational burden, and immunotherapeutic and chemotherapeutic sensitivity in READ (rectal adenocarcinoma) patients. Our results suggested that the established signature based on m⁵C RNA methylation regulators could be used as a robust biomarker to predict the clinical prognosis and therapeutic effect among patients with rectal cancer.

Materials and methods

Acquisition and processing of data sets

The RNA-sequencing transcriptome data (TPM value) and corresponding clinical annotation were retrieved from The Cancer Genome Atlas (TCGA) database (<http://gdcc.portal.nci.nih.gov/>). After patients without survival information were excluded, a total of 434 colon adenocarcinoma (COAD) and 157 READ samples were integrated for further analysis. The

validation data set was retrieved from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) under the accession number GSE87211 (n=190) (20) and GSE133057 (n=17) (21). The copy number variations (CNV) of READ used in our research were retrieved from the UCSC Xena browser (<http://xena.ucsc.edu/>), where genes with CNV values smaller than -0.3 were categorized as a “loss,” while CNV values larger than 0.3 were categorized as a “gain.” The messages of simple nucleotide variations (SNV) were retrieved from the TCGA database, R package *maftools* was used to analyze the level 4 mutation data, and the *mafCompare* function of *maftools* was used to identify the differentially mutated genes (DMGs) (22). The neoantigens and mutation loads for READ were accessed from The Cancer Immunome Atlas (<https://tcia.at/>) database (23). Information on CMS subtyping calls and sample annotations were retrieved from the Colorectal Cancer Subtyping Consortium Synapse (24). The STRING database can predict the functional links between proteins based on a variety of algorithms. The genes with the highest confidence scores were identified as the functional partners of specific genes (25). The Gene_DE module of Tumor Immune Estimation Resource (TIMER, cistrome.shinyapps.io/timer) can be utilized to examine the mRNA expression profiles between the tumor tissues and the normal tissues (26). We used the Human Protein Atlas (HPA) database to analyze the protein expression levels of candidate genes in tumor tissues and corresponding normal tissues (27).

Construction of gene signature and survival analysis

The least absolute shrinkage and selection operator (LASSO) model is a linear regression method applying L1-regularization, which could accurately contract some regression coefficients to zero to achieve sparseness and feature selection (28). The LASSO model was generated through R package *glmnet*. At the penalty coefficient ($\lambda_{\min} = 0.036$), the optimal risk model was established based on three m⁵C regulatory genes. Next, the R package *survival* was used to calculate the risk scores for rectal cancer samples. The following formula was used:

$$\text{Risk score} = e^{\left(\text{constant} + \sum_i \text{coefficient}(mRNA_i) \times \text{expression}(mRNA_i) \right)}$$

Patients from the TCGA training cohort were separated into a high-risk and a low-risk group according to the median value of the calculated risk score. Patients from the GEO validation data set were grouped based on the optimal cutoff decided by *cutp* function of the R package *survMisc*. The Kaplan–Meier method was employed to compare the survival probability between the two risk subgroups.

Functional enrichment analysis

Differentially expressed genes (DEGs) between the subgroups were identified by R package *limma*. Metascape ([\[metascape.org\]\(http://metascape.org\)\), a web tool comprising Gene Ontology \(GO\) and Kyoto Encyclopedia of Genes and Genomes \(KEGG\) analysis \(29\), was used to identify the terms across different ontology sources enriched based on the screened DEGs. A GOCircle plot was depicted to show the enriched terms by R package Goplot \(30\). To further investigate pathways enriched in specific subgroups, we performed Gene Set Variation Analysis \(GSVA\) by R package GSVA. GSVA is a gene set enrichment method that estimates variation of pathway activity over a sample population in an unsupervised manner \(31\). The gene set of “c5.go.v7.4.symbols” was downloaded from MSigDB database, and gene markers of epithelial–mesenchymal transition \(EMT\) including EMT1, EMT2, and EMT3; angiogenesis; pan-fibroblast TGFβ; and type I IFN response were obtained from previous studies for GSVA analysis \(32, 33\).](http://</p>
</div>
<div data-bbox=)

Immune cell infiltration analysis

A total of 28 immune cell types were collected for GSVA analysis (23). A web server TIMER, integrating multiple algorithms (TIMER, Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts [CIBERSORT], European Prospective Investigation into Cancer and Nutrition [EPIC]) was used to estimate the abundances of immune cell types based on the gene expression profiles (26, 34, 35). The ratios between immune-stimulatory signatures and immune-inhibitory signatures (CD8 +/CD4+ regulatory T cells, pro-/anti-inflammatory cytokines, and M1/M2 macrophages) were also compared between the subgroups based on the average expression levels of the marker genes (36). The immune system–related genes were obtained from previous studies (23, 37–39). The Pearson correlation was calculated and then depicted by R package *corrplot*.

Prediction of the efficacy of immunotherapy and chemotherapy

A web platform named Tumor Immune Dysfunction and Exclusion (TIDE, <http://tide.dfci.harvard.edu>) was used to evaluate the anti-PD1 and anti-CTLA4 immunotherapeutic response based on the gene expression profiles of the TCGA-READ cohort (40). To validate the correlation between immunotherapeutic efficacy and three genes–based risk model, another data set was retrieved, which included 348 patients with metastatic urothelial cancer who were treated with an anti-PD-L1 agent (32). The R package *oncoPredict* can be used to discover drug sensitivity *in vitro* and *in vivo* contexts (41). The half-maximal inhibitory concentration (IC50) was calculated to predict the chemotherapeutic response in READ patients. The Cancer Therapeutics Response Portal (CTRP, <https://portals.broadinstitute.org/ctrp/>) (42) and Profiling Relative Inhibition Simultaneously in Mixtures (PRISM, <https://depmap.org/portal/prism/>) (43) were both developed to access the associations between drug sensitivity and gene expression. The calcPhenotype function of R package *oncoPredict* was used to calculate the AUC (Area Under Curve) value of each drug based on the CTRP and PRISM databases. Lower AUC value indicates higher sensitivity to therapeutic drugs.

Consensus clustering analysis

To further validate the reliability of discriminating patients with rectal cancer into two subgroups based on the three m⁵C regulatory genes, DEGs were identified through R package *limma* between the low- and high-risk groups. Furthermore, univariate Cox regression analysis was carried out by R package *survival* to filter prognostic genes on the basis of DEGs. Ultimately, unsupervised clustering analysis was conducted by using R package *ConsensusClusterPlus*, which was repeated 1,000 times to identify different risk gene clusters (44).

Tissue microarray–based immunohistochemistry validation

From Superbiotech in Shanghai, China (#REC1601), we acquired a TMA of 80 paired rectal cancers and corresponding normal tissues. Surgical samples from the patients were taken between May 2008 and December 2012 through operations. The patients' median survival duration was 81.5 months, ranging from 14 to 130 months. For every case, clinicopathological information including overall survival time, survival status, age, gender, tumor size, pathological T, N, and M stage, and grade was accessible. Based on this commercial TMA, we conducted a retrospective analysis.

For immunohistochemistry (IHC) process, the TMA slides were deparaffinized, rehydrated, and incubated by 3% hydrogen peroxide to block the endogenous peroxidase activity for 10 min at room temperature. Antigens were restored by boiling in a pressure cooker containing sodium citrate buffer for 90 s. The slides were incubated in bovine serum albumin (BSA) for 30 min to reduce nonspecific background. Then, they were incubated with rabbit monoclonal NSUN4 antibody (HPA028489, Sigma), NSUN7 antibody (HPA020653, Sigma), and DNMT1 antibody (HPA002694, Sigma) at 4°C overnight. Next, secondary antibody was incubated with the slides for 1 h at 37°C. Finally, the slides were developed in 3, 3'-diaminobenzidine (DAB) and stained with hematoxylin.

The slides were assessed digitally with the APERIO ScanScope (Leica Biosystems, Germany) and the APERIO ImageScope (Leica Biosystems, Germany) using the positive pixel counting algorithm. The IHC staining results were interpreted by both the intensity of staining and the staining positive area. Each sample was assigned a score according to the intensity of the staining (0 = no staining; 1 = weak staining; 2 = moderate staining; and 3 = strong staining) and the proportion of stained cells (0 = 0%; 1 = 1%–25%; 2 = 25%–50%; 3 = 50%–75%; 4 = 75%–100%). The final score was calculated as the staining intensity multiplying positive area score, ranging from 0 to 12. The IHC results of TMA-rectal cancer were independently reviewed by two experienced pathologists who were blinded to the clinical parameters.

Real-time quantitative PCR validation

For the RT-qPCR experiment, tissue samples from 26 rectal cancer patients and matched nearby normal tissue samples (proximity to the cancer larger than 5 cm) were collected at the

Affiliated Hospital of Qingdao University. The inclusion requirements were as follows (1): a pathological analysis and imaging-based diagnosis of rectal cancer; (2) radical resection; (3) available information on clinicopathological indexes, such as tumor size, pathological stage, and pathological TNM; (4) pathological TNM in accordance with the 8th edition of the American Joint Committee on Cancer; and (5) lack of a prior history of other malignancies. Patients with recurrent rectal cancer and nonprimary malignancies as well as those who had had neoadjuvant chemotherapy and/or radiation prior to surgery were disqualified. All of the included patients gave their informed permission. The Affiliated Hospital of Qingdao University's Research Ethics Committee approved the study, and it was completed in conformity with the 1964 Helsinki Declaration and its later amendments.

Total RNA was extracted using RNeasy kit (Beyotime, Shanghai, China, R0027) in accordance with the manufacturer's instructions. Then, total RNA (1 µg) was quantified, followed by reverse-transcription by the SuperScript II reverse transcriptase (Takara, Japan, RR047). Quantitative PCR analysis was operated using SYBR Green Mix (Takara, Japan, RR820) with ABI 7900 HT Real-Time PCR system. The primer sequences are listed below: NSUN4, 5'-CCAAACCCTGGCAAAGGTG-3', 5'-GCGTGCCG GTCATAGAAGAA-3'; NSUN7, 5'-CCAGATCATTTGAGCAGT CTTATT-3', 5'-GGTCTCTACTTCTTGAACCTTCTGA-3'; DNMT1, 5'-ATCCGAGGAGGGCTACCTG-3', 5'-ACTTCTT GCTTGGTTCCCGT-3'; GAPDH, 5'-CTGACTTCAACAG CGACACC-3', 5'-TGAGCTTGACAAAGTGGTCGT-3'. mRNA levels were determined relatively according to the expression of GAPDH.

Statistical analysis

The t-test or Wilcoxon test was adopted for comparisons of two groups, and one-way ANOVA or Kruskal–Wallis test was adopted for comparisons of three or more groups. The choice of t-test vs. Wilcoxon test, or one-way ANOVA vs. Kruskal–Wallis test, was based on the normality of the variables. Chi-squared tests were used to analyze the distribution of variables among different subgroups. Multivariate Cox regression analysis was carried out by R package *survival*. Receiver operating characteristic (ROC) analysis was used to evaluate the predictive power of the established model. We constructed nomograms to predict survival probability using R package *rms*. *P* value less than 0.05 was recognized as significant in this research.

Results

Construction of m⁵C RNA methylation regulator–based signature for READ patients

The schematic diagram summarizes the study design of the current research (Figure 1). m⁵C RNA modification regulators (NOP2 nucleolar protein [NOP2], NOP2/Sun RNA

To promote the clinical application, LASSO-penalized Cox analysis was performed to enhance the forecast accuracy and explainability of the statistical model. In the current model, the optimal penalty coefficient ($\lambda = 0.036$, $\log \lambda = -3.33$) was identified with the minimum criterion (Figure 3A). In Figure 3B, each curve indicates the track of a single gene, and the red dot represents the target lambda. We can see that three genes (DNMT1, NSUN4, and NSUN7) were retained after the shrinking process. Then, the produced three prognostic indicators were employed to predict clinical results.

To confirm the effectiveness of the established model, we carried out Kaplan–Meier survival analyses. We found a superior survival status in the low-risk group compared with the corresponding high-risk group in both the TCGA dataset and two GEO cohorts, illustrating that the built model could significantly predict the



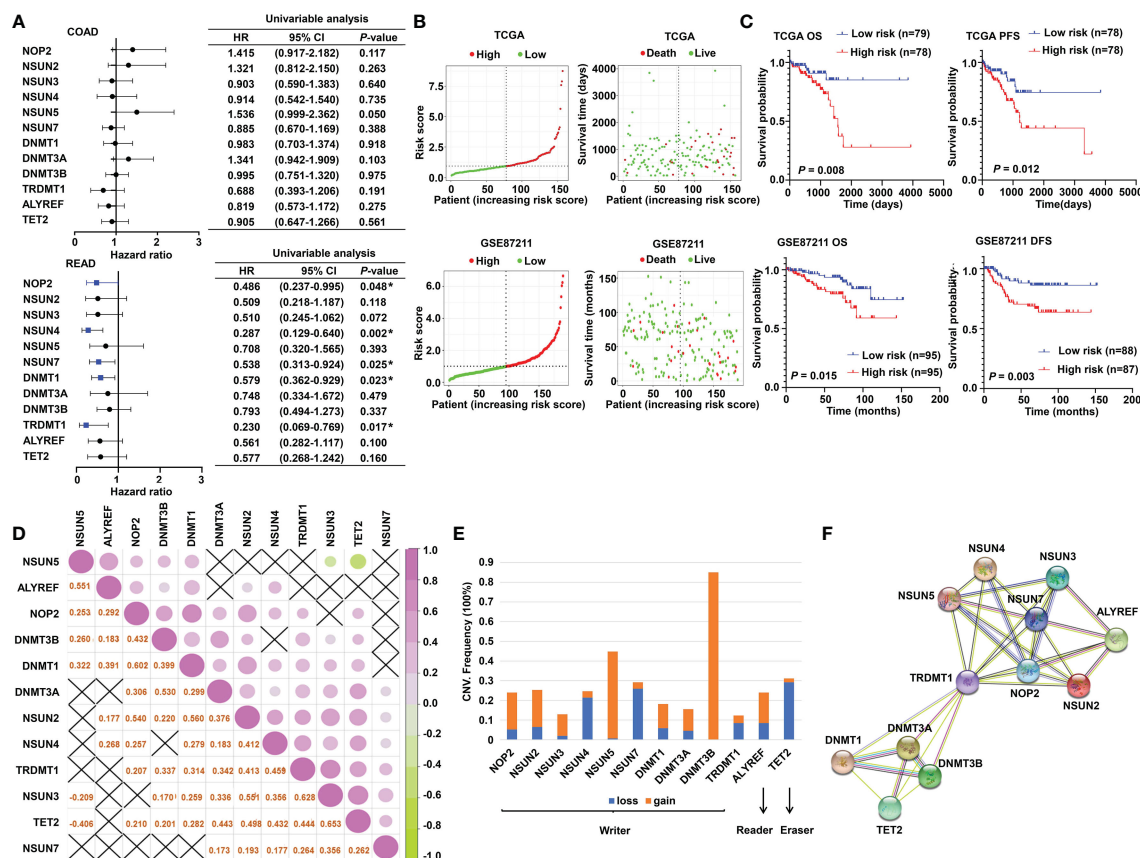


FIGURE 2

The prognostic value of m⁵C RNA methylation regulators. (A) Forest plot of the prognostic ability of the m⁵C regulator genes in COAD and READ separately. (B) The risk score distribution and patient survival status are shown in ranked dot and scattered plots based on the expression of m⁵C regulator genes. (C) The Kaplan–Meier curves for OS, PFS, and DFS between the high-risk and the low-risk groups in READ from the TCGA and GSE87211 cohorts. (D) Pearson correlation among the m⁵C regulators in READ patients. (E) The CNV variation frequency of the m⁵C regulators in the TCGA cohort. The orange rectangle = the amplification frequency; the blue rectangle = the deletion frequency. (F) The PPI network depicted for m⁵C regulators. CI, confidence interval; DFS, disease-free survival; HR, hazard ratio; OS, overall survival; PFS, progression-free survival. **P* < 0.05.

prognosis of READ patients (Figure 3C). Similar processes were applied to the samples of COAD patients, and no factor was retained after LASSO analysis (Figure S1A). The three factors identified in the READ patients were repurposed for COAD samples; as expected, the survival curves of the two groups were highly crossed (Figure S1B). To further explore the relationship of the prognostic risk model of the three m⁵C regulators and clinical features in READ, univariate and multivariate Cox regression analyses were conducted. To facilitate the understanding of the patients' clinical and genetic background, a table including basic information about the low- and high-risk groups is displayed in Table S1. The results of Cox regression analysis revealed that the risk score was an independent prognostic factor for READ, unrelating to clinicopathological parameters, such as pathologic N and age (Figure 3D). We further investigated whether the risk score could further subdivide the pathologic N and age parameters. The results showed that the established risk score further distinguished the risk pattern in subgroups differentiated by age, successfully stratified the patients in the N0 pathological stage, and exhibited a tendency to differentiate patients in the N1 pathological stage due to small sample size (Figure 3E). To visualize the expression pattern of

m⁵C regulators, a heat map was depicted. To our expectation, the majority of the methylation regulators displayed a significant high expression module in the low-risk group (Figure S2), which is reasonable due to their protective ability in READ. Thus, this powerful and accurate model symbolized a potential clinical parameter for patients with READ.

Construction and validation of a nomogram combined with clinical parameters

To make the m⁵C regulator-based risk signature more clinically adapted and available, a prognostic nomogram was depicted integrating the risk factors and independent identified parameters of READ. The aim was to establish a quantitative analytic algorithm that could be put into practice for survival prediction. In the current case, the pathologic N, age, and risk score were integrated to calculate the corresponding score, which could be used as an index for matching the one-, three-, and five-year death probabilities (Figure 4A). To reinforce the superior capability of the established nomogram, the ROC analyses

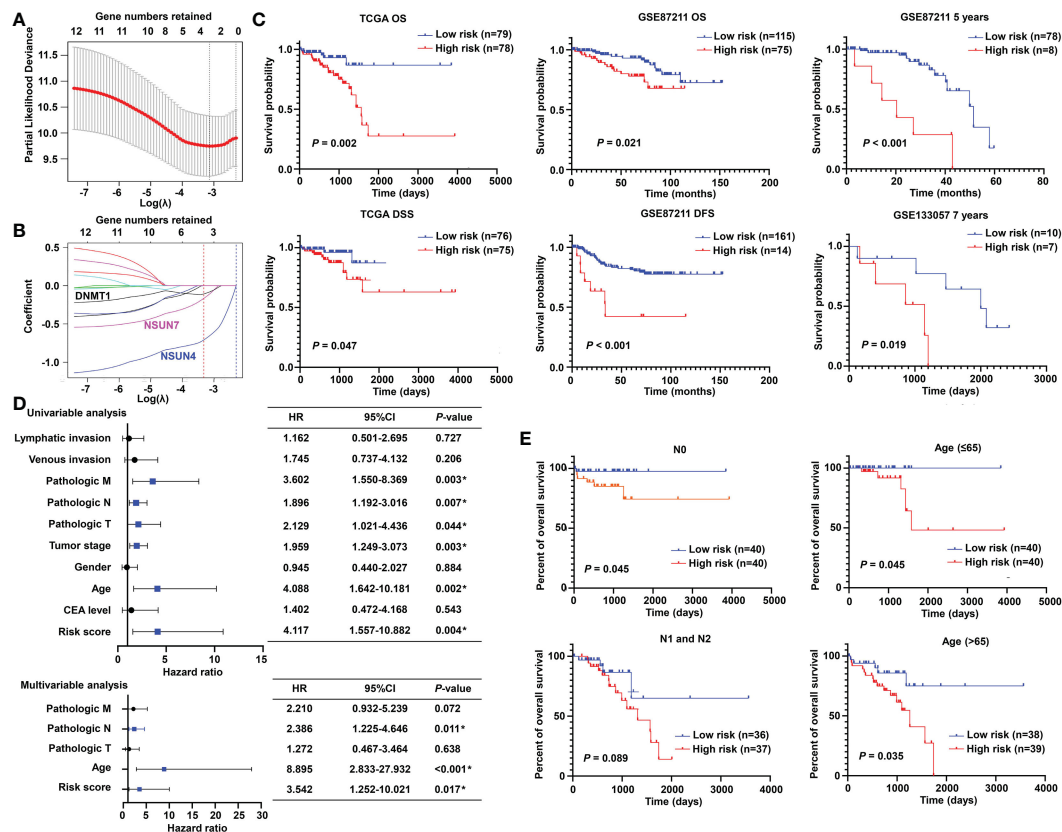


FIGURE 3

Prognostic significance of the m^5C methylation-based signature in READ patients. (A) The process of LASSO regression based on the TCGA cohort and the identification of “lambda” for best selection of gene signature. (B) The curves indicate the tracks of single genes; the red dot line represents the target lambda. The blue track refers to NSUN4, pink track refers to NSUN7, and black line is DNMT1. (C) The Kaplan–Meier curves for OS and DSS between two categories in READ from the TCGA data set; the Kaplan–Meier curves for OS, 5-year survival based on the GSE87211 data set; the Kaplan–Meier curves for 7-year survival based on the GSE133057 data set. (D) The univariate and multivariate Cox analysis integrating risk score and clinicopathological indexes based on the TCGA cohort. (E) The prognostic ability of the risk score in distinguishing the overall survival status for READ patients with or without lymph node metastasis. The prognostic ability of the risk score in differentiating the overall survival status in patients with age less than 65 years or those with 65 years or more. CI, confidence interval; DFS, disease-free survival; DSS, disease-specific survival; HR, hazard ratio; OS, overall survival. * $P < 0.05$.

were used to compare the prognostic accuracy and specificity. The results indicated that the nomogram was superior to other independent clinical factors for predicting the overall survival (OS) of READ patients in the TCGA cohort (AUC of one-year OS = 0.803; AUC of three-year OS = 0.855; AUC of five-year OS = 0.838; AUC of overall survival = 0.834; Figure 4B). The calibration curve was drawn to confirm the consistency between the nomogram-predicted and the actual probability. The calibration curves were close to the optimal performance in the one-, three-, and five-year nomogram (Figure 4C), indicating the accuracy of the constructed nomogram. These results implied that the three-gene signature was capable and reliable to make prediction for READ patients.

Functional enrichment analysis of m^5C methylation-based signature between low- and high-risk READ patients

To explore the underlying molecular mechanisms of the m^5C -based signature, GO, GSEA, and GSVA analyses were performed.

DEGs were identified using the limma algorithm, and the result is displayed as a volcano plot (Figure S3). Next, the screened DEGs were put into the GO analysis. The GO pathway enrichment analysis revealed that the most significantly changed pathways in the high-risk subgroup were mainly related to cancer and immune-targeted processes, such as epithelial-mesenchymal transition, angiogenesis, hypoxia, regulation of leukocyte migration, and regulation of macrophage activation; however, cell cycle-related pathways, including G2M checkpoints, sister chromatid segregation, and signal transduction in response to DNA damage, were mainly converged in the low-risk group (Figure 5A). The GSEA analysis confirmed these findings and showed some extent of overlap with the GO analysis results (Figure 5B). In order to clarify the specific roles of these pathways according to the risk categories, a series of related gene sets were collected to further carry out the GSVA analysis. Importantly, the GSVA results revealed that the process of angiogenesis, EMT, and pan-fibroblast TGF β were consistently upregulated in the high-risk category (Figure 5C). Meanwhile, the GSVA analysis indicated that many biological functions in the high-risk group primarily correlated with

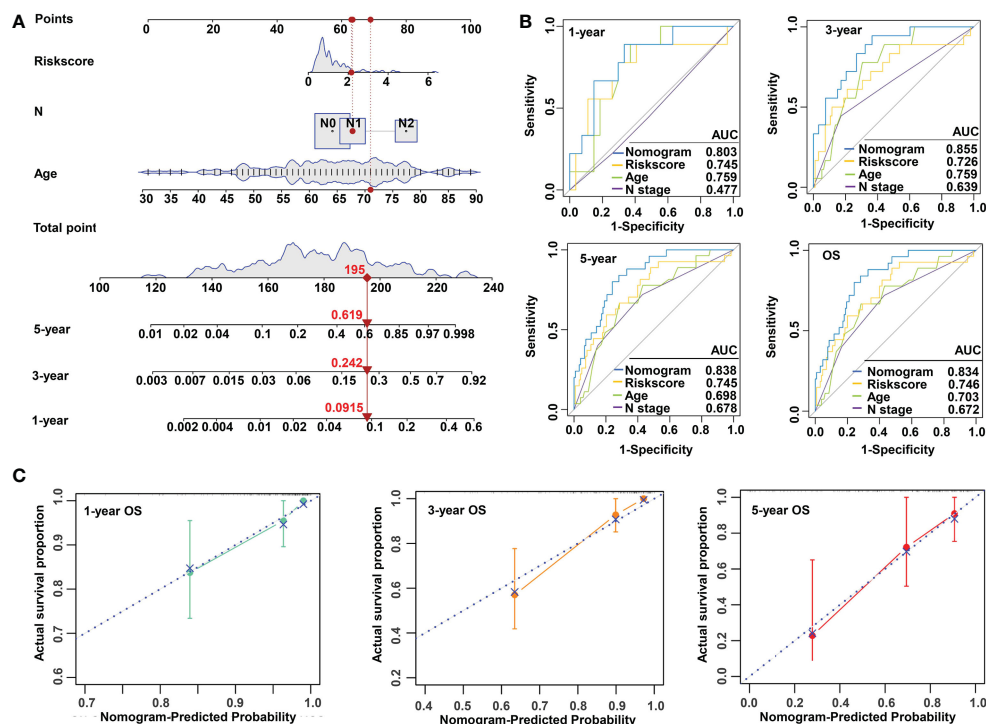


FIGURE 4

Construction and validation of a risk model based on m^5C methylation regulators. (A) The predictive nomogram integrating the risk score and clinicopathological parameters for 1-, 3-, and 5-year OS in READ patients from the TCGA cohorts. (B) The ROC for nomogram and independent clinical parameters for 1-, 3-, and 5-year OS based on the TCGA cohort in READ patients. (C) The calibration curve depicted for 1-, 3-, and 5-year nomogram in TCGA. OS, overall survival.

inflammatory responses and carcinogenic reactions, while in the low-risk group, RNA methylation process and drug response were significantly enriched (Figure 5D). These features gave the hint that cancer–immunity interaction is the potential mechanism of the m^5C -based risk signature, and the efficacy of the established model was further validated by the above results.

The immune characteristics of the m^5C regulator–based signature in READ

Due to the close relationship between the built model and immune process, the detailed connection between the risk signature and immune cell abundance was studied. The GSVA and deconvolution algorithms including CYBERSORT, TIMER, and EPIC were used to evaluate the extent of infiltrating immune cells. CD4+ T cells, B cells, CD8+ T cells, dendritic cells, and T helper cells exhibited higher expression in the low-risk category; meanwhile, the abundance of myeloid-derived suppressor cells (MDSC) and regulatory T cells (Tregs) was elevated in the high-risk category compared with the low-risk group (Figures 6A–C, E). Furthermore, additional investigations were conducted to substantiate the above findings. The ratio between the immune stimulatory signatures (including CD8+ T cells, proinflammatory cytokines, and M1 macrophages) and the immune inhibitory signatures (integrating CD4+ regulatory T cells, anti-inflammatory cytokines, and M2 macrophages) was significantly

increased in the low-risk category (Figure 6D), which was consistent with the above results, indicating an immune-inhibiting environment in the high-risk group and a proinflammatory status in the low-risk group. We collected the signatures of cancer–immunity cycle and immune stimulators. The heat maps showed that the majority of genes exhibited higher expression in the low-risk group (Figure 6F) and the established risk score correlated negatively with the expression of most of the immune stimulators (Figure 6G). According to the obtained evidence, the low-risk group belongs to activated immune microenvironment, while the high-risk group shows a suppressed immune phenotype.

The mutational landscape for the m^5C regulator–based signature in READ

Considering the evidence that hot tumor is more sensitive to immune therapy, we hypothesized that the low-risk group of our established model might be more readily responsive to immune therapies than the high-risk group. Previous studies have revealed that high somatic mutation and neoantigens represent a higher possibility to response. Thus, we investigated the differences in mutation status between the two groups. First, we identified the top 10 mutated genes in rectal cancer using the maftools R package (Figure S4A), and these genes were subsequently compared between the two subgroups. A significantly higher mutational rate of RYR2 was observed in the low-risk group, while the other genes showed

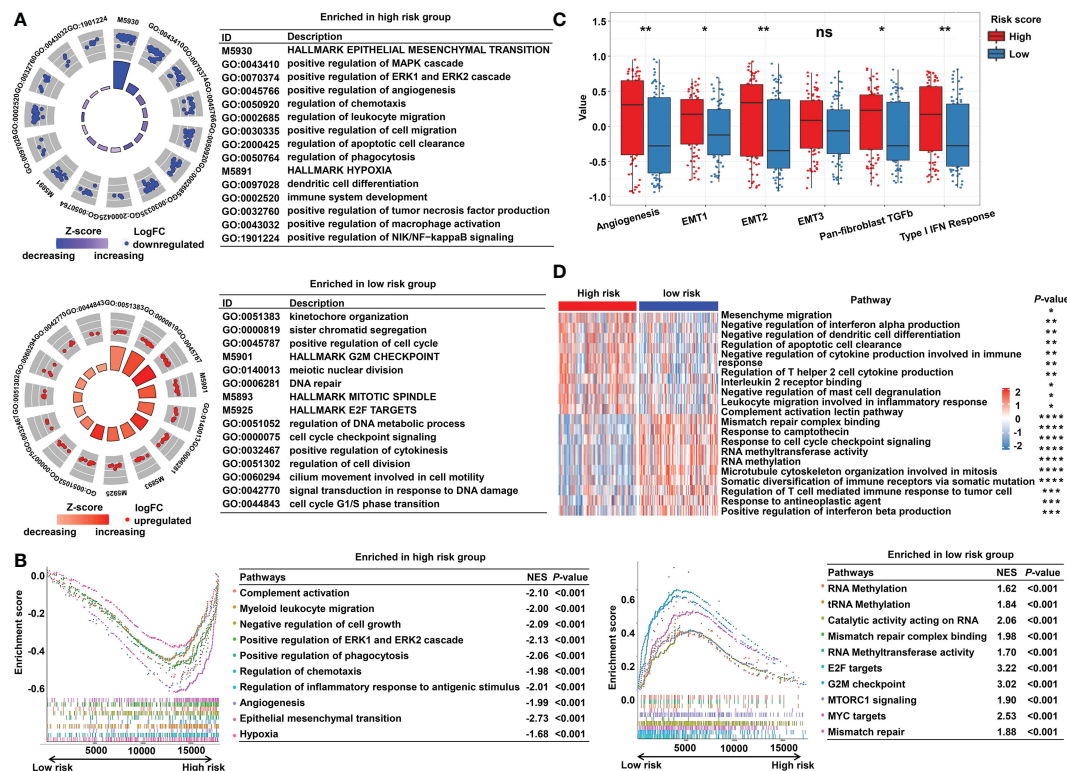


FIGURE 5

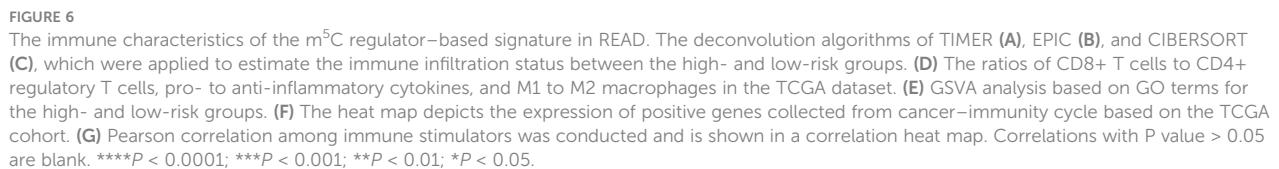
Functional enrichment analysis of the m⁵C methylation-based signature between low- and high-risk READ patients. (A) The enriched pathways including GO and HALLMARK terms are displayed by GOcircle plots. The red and blue dots represent the genes upregulated in the low-risk and high-risk categories separately. (B) GSEA enrichment plots for the two subgroups in the TCGA cohort. (C) The GSVA analysis for hallmarks of cancer in the TCGA cohort. (D) The heat map drawn for GSVA analysis based on GO terms. DEGs, differentially expressed genes. **** $P < 0.0001$; *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; ns, not significant.

no statistical differences (Figure S4B). Then, we used the *mafCompare* function to identify the DMGs. Interestingly, we found overall higher mutational rates in the low-risk group (Figure 7A), indicating that the built model did not affect the frequently mutated genes but exerted a cumulative effect of low-frequency mutations. We also found that the low-risk group was accompanied by more neoantigens. However, TMB only exhibited an elevated tendency (Figure 7B). Moreover, we combined the m⁵C-based model with neoantigens and TMB and found that neoantigens and TMB cannot effectively distinguish the survival status in patients with rectal cancer (riskscore-L + NEO-L vs. riskscore-L + NEO-H, $P = 0.655$; riskscore-L + TMB-L vs. riskscore-L + TMB-H, $P = 0.748$), although possessing high neoantigen levels showed a tendency of better overall survival compared with possessing low neoantigen levels (riskscore-H + NEO-L vs. riskscore-H + NEO-H, $P = 0.083$). The constructed risk score showed significant efficacy in stratifying patients with a same status of neoantigens and TMB (riskscore-L + NEO-L vs. riskscore-H + NEO-L, $P = 0.012$; riskscore-L + NEO-H vs. riskscore-H + NEO-H, $P = 0.050$; riskscore-L + TMB-L vs. riskscore-H + TMB-L, $p = 0.005$), confirming the superiority of this model over current biomarkers. In addition, riskscore-L + TMB-H vs. riskscore-H + TMB-H ($P = 0.064$) showed a strong tendency without significant difference. We also found that combining risk score with neoantigens (riskscore-L + NEO-H vs. riskscore-H + NEO-L, $P =$

0.002) could achieve a higher efficiency for predicting the prognosis of patients with rectal cancer (Figure 7C).

Prediction of immunotherapeutic response for distinct subgroups in READ

The obtained findings promoted us to further examine the relationship between the m⁵C-based signature and immunotherapy. First, we compared the expression of the immune checkpoints in the two subgroups. No significant differences were found, as shown in Figure S5. Next, we investigated the relationship between model factors and immune infiltration cells. Interestingly, we found that DNMT1 and NSUN4 were moderately positively correlated with CD4⁺ T cells, natural killer cells, dendritic cells, and T helper cells; meanwhile, NSUN7 was weakly negatively correlated with MDSC and Tregs (Figure 8A), substantiating the close connection between the risk model based on the above three m⁵C regulatory genes and the tumor immune microenvironment. Next, we investigated the relationship between model factors and immune checkpoints. Higher expression of NSUN4 was accompanied by a higher expression of immune checkpoints, and patients with high DNMT1 expression showed a trend of elevated expression of immune checkpoints. However, low expression of NSUN7 was associated with only weakly elevated immune checkpoint



investigation by adopting an additional data set with the therapeutic information. We compared the survival rates of two subgroups by conducting Kaplan-Meier analysis, and found that the low-risk group had prolonged survival compared with the high-risk group despite an insignificant P value ($P = 0.121$, **Supplementary Figure S7**). The expression of immune checkpoints was higher in the low-risk group, which represents higher sensitivity toward ICI treatment (**Figure 8E**). Accordingly, the proportion of complete response/partial response (CR/PR) was remarkably higher in the low-risk group (**Figure 8F**), and the risk score was lower in the CR/PR subgroup (**Figure 8G**). Interestingly, compared with the immune-excluded high-risk group, the low-risk group revealed an immune inflammation phenotype (**Figure 8H**). These results solidly certified that the established signature had the ability to efficiently predict the immunotherapeutic efficacy for READ patients.

To further investigate the heterogeneity of different m⁵C RNA methylation regulator patterns, we identified 950 DEGs between the high-risk and low-risk groups. Subsequently,

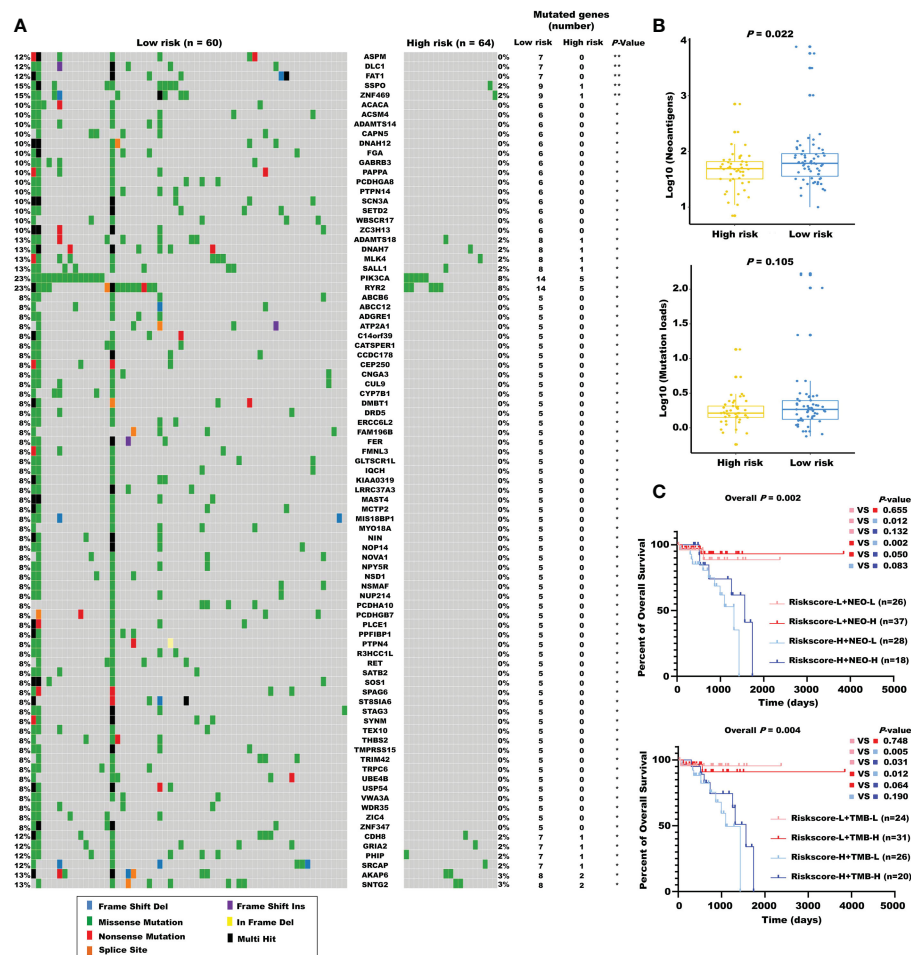


FIGURE 7

The mutational landscape for the m^5C regulator-based signature in READ. (A) The waterfall plot of differentiated somatic mutation features between the high- and low-risk groups using the TCGA-READ data set. (B) The neoantigens and mutation loads between the two subgroups are displayed. (C) Survival analyses for READ patients stratified by both the risk score and neoantigen burden or mutation loads using Kaplan-Meier curves. NEO, neoantigen burden; H, high; L, low. $**P < 0.01$; $*P < 0.05$.

univariate Cox regression analysis was conducted to certify the genes with prognostic value, and finally, a total of 173 m^5C RNA methylation regulator risk model-related genes were identified (Figure 9A). Unsupervised clustering analysis based on the expression of these 173 genes separated READ patients into two clusters, which we referred to as m^5C RNA methylation gene clusters (Figure 9B). Survival analysis indicated that cluster 2 had a better prognosis (Figure 9C). Moreover, we found that cluster 1 had a higher risk score than that in cluster 2 (Figure 9D), and chi-squared tests also revealed a significant difference between the two clusters (Figure 9E). CMS stratification is considered a robust classification system and is currently used for CRC with distinguished features; among the four CMS subtypes, CMS4 mesenchymal tumors display worse overall survival and relapse-free survival (24). To evaluate the CMS status in different m^5C regulator-based subgroups, we further compared the proportion of the CMS phenotypes by chi-squared tests. The high-risk group and cluster 1 category displayed a higher proportion of CMS4 compared with other categories (Figures 9F, G).

In line with the previous findings, cluster 1 was enriched mainly in cancer and immune system-related pathways, while process related to the functions of RNA methylation played an important role in cluster 2 (Figure 9I). Patients of cluster 2 had higher abundance of CD4+ T cells and helper T cells, while cluster 1 exhibited higher amount of immune-inhibiting cells, such as Tregs and macrophages (Figure 9J). The relationship of the survival status, m^5C regulator-based risk model, m^5C regulator gene clusters, and CMS phenotypes is summarized in a Sankey diagram (Figure 9H). The TIDE algorithm was carried out to predict the immunotherapeutic response relating with the clustering system. Accordingly, there were more responders in cluster 2 (Figure 9K), and the index integrating the lower TIDE score, higher MSI score, lower extent of T-cell dysfunction and exclusion, and lower abundance of cancer-associated fibroblast (CAF) consistently indicated a better responsive rate for patients of cluster 2 (Figures 9L–P). To make the outline clear, a Sankey diagram connecting with both the risk classification and clustering system was depicted (Figure 9Q). Above all, these results reinforced the notion that there were indeed two different m^5C regulator-

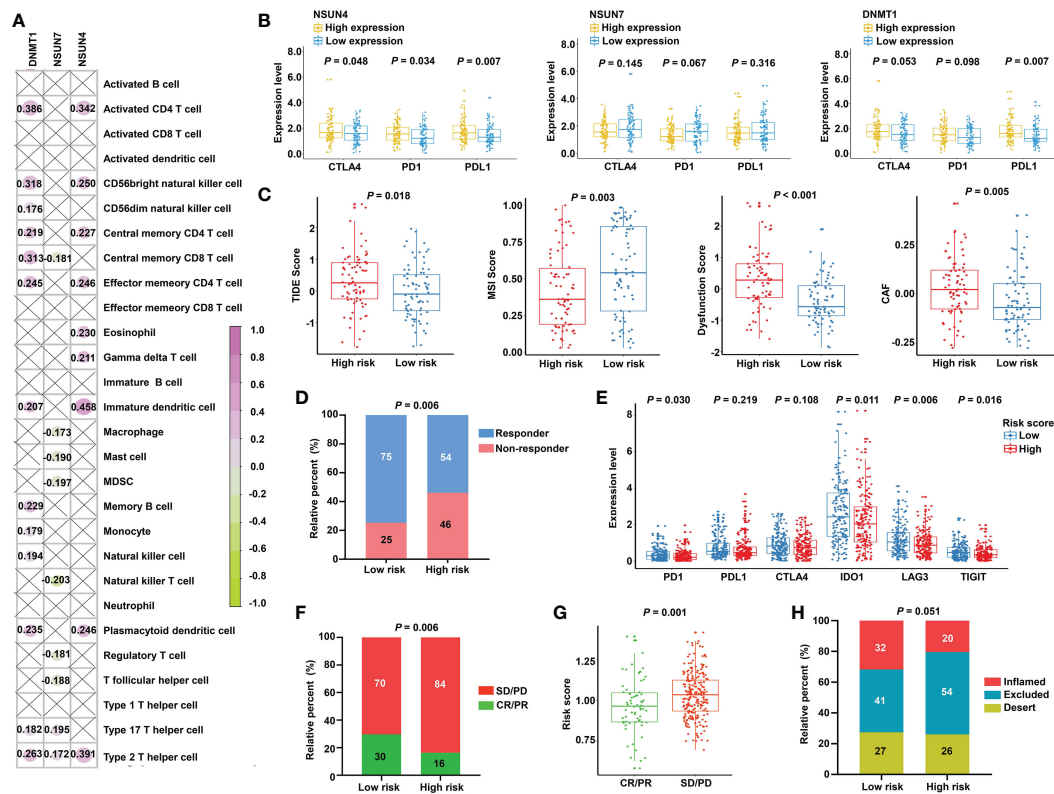


FIGURE 8

Prediction of immunotherapeutic response for distinct subgroups in READ. (A) Pearson correlation between three signature factors and 28 types of immune cells is illustrated by a correlation heat map. Correlations with P value > 0.05 are marked by a cross. (B) The differences in the three signature factors between distinct subgroups classified by the expression level of three immune checkpoints, including CTLA4, PD1, and PDL1. (C) The distribution of TIDE score, MSI score, T-cell dysfunction score, and abundance of CAF between the low- and high-risk categories. (D) The proportion of READ patients with response to ICI therapy in the high- and low-risk groups based on TIDE prediction. (E) The differential analysis for immune checkpoints between the two categories in IMvigor210 cohort. (F) The proportion of patients with response to PD-L1 treatment in the high- and low-risk groups based on IMvigor210 cohort. (G) Distribution of the risk score between CR/PR and SD/PD groups. (H) The proportion of immune phenotype in the high- and low-risk groups. CR, complete response; PD, progressive disease; PR, partial response; SD, stable disease.

based groups in READ, which represented different clinical and immune features.

Validation of the m⁵C methylation-based signature by TMA in patients with rectal cancer

To demonstrate the robustness and repeatability of the prognostic value of the established model, different laboratory assays were adopted. RT-qPCR was conducted to detect the mRNA expression of the signature's factors in 26 pairs of rectal cancer tissues and corresponding normal tissues. The results showed that NSUN4 was highly expressed in normal tissue (Figure S8A). Examination of the correlation between the risk score and clinical parameters revealed a higher proportion of patients with no lymph node metastasis in the low-risk group compared with the high-risk group (Figure S8B).

Next, we detected the protein expression levels of NSUN4, NSUN7, and DNMT1 *via* IHC staining in a tissue microarray containing 80 paired normal and tumor tissues. The clinical features of tissue microarray as the validation cohort are displayed in Table

S2. The protein expression levels of the three m⁵C regulatory genes were analyzed using IHC staining, substantiating the findings obtained using the TCGA-READ dataset. The following analyses were based on the protein expression levels detected *via* IHC. The results revealed significant elevation of NSUN7 and DNMT1 in normal tissues compared with tumor tissues, while NSUN4 showed no obvious difference between the two groups (Figure 10A). Next, we investigated the relationship between the three genes using the Pearson correlation analysis. High correlation coefficients (> 0.7) shown in the correlation plot indicate that the protein expression levels of the three genes were closely associated (Figure 10B).

Importantly, the KM survival curves demonstrated that the survival probability was significantly increased in the high expression group compared to the low expression group, according to the protein expression of an individual gene in the risk model (Figure 10D). We constructed a signature based on the protein expression of the three genes, in which the low-risk group showed prolonged survival compared with the high-risk group (Figure 11A). Remarkably, based on the IHC protein expression data, the risk score was correlated with clinical characteristics including pathologic TNM, gender, grade, and clinical stage (Figure 10C); this was further confirmed by a Wilcoxon test

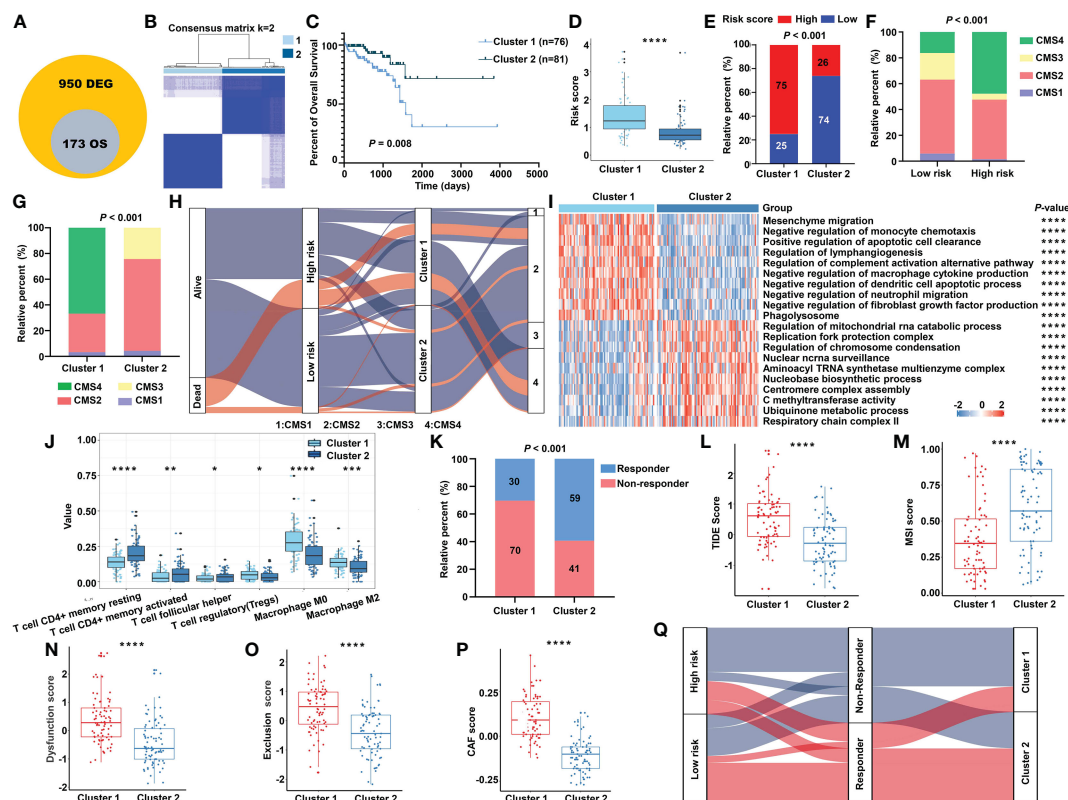


FIGURE 9

The transcriptomic characteristics of the m^5C methylation-based gene clusters. (A) The intersection of DEGs and prognostic genes. (B) The unsupervised consensus cluster of the identified 173 genes. (C) The Kaplan–Meier survival curve for two clusters in the TCGA cohort. (D) The m^5C signature-based risk score distribution between two clusters. (E) The proportion of READ patients with different risk status in cluster 1 and cluster 2 from the TCGA cohort. The CMS distribution among the risk groups (F) and clusters (G) separately. (H) Sankey diagram depicting the relationship of survival status, risk groups, clusters, and CMS classification. (I) The functional enrichment analysis on GO terms of the two clusters. (J) The immune cells infiltration between different clusters. (K) The proportion of responsive patients in the two clusters based on TIDE prediction. The distribution of TIDE score (L), MSI score (M), T-cell dysfunction score (N), T-cell exclusion score (O), and abundance of CAF (P) in cluster 1 and cluster 2. (Q) Sankey diagram connecting the two classification systems with the immune response. **** $P < 0.0001$; *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$.

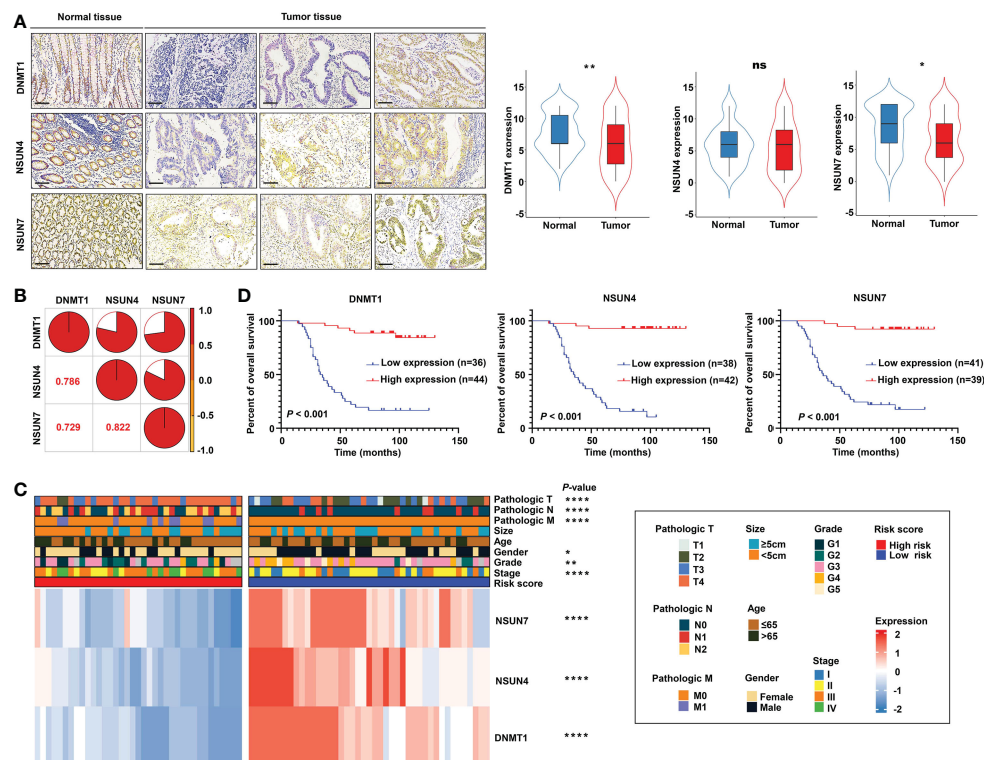
between the two subgroups (Figure 11D). To examine the significance of the established risk score, univariate and multivariate Cox regression analyses were conducted. Risk score, grade, and pathologic M remained independent factors after the above tests (Figure 11C). ROC analysis was exploited to inspect the superiority of the built risk score over other indexes (AUC of risk score = 0.954; AUC of grade = 0.744; AUC of pathologic N = 0.764; AUC of pathologic M = 0.639; AUC of pathologic T = 0.749; Figure 11B). To validate the efficiency of the nomogram generated based on the TCGA-READ dataset, we integrated the model factors, including risk score, age, and pathological N, to construct a nomogram based on the IHC independent cohort (Figure 11E). The C-index of the nomogram was 0.840, indicating a stable and robust predictive power. The subsequent calibration plots also revealed high concordance between the predicted probability of three-, five-, and seven-year OS and actual OS (Figure 11F). These results reinforced that our classification based on the m^5C methylation regulators was potent and reliable in terms of prognostic significance for patients with rectal cancer.

In addition, the regulated genes associated with NSUN4, NSUN7 and DNMT1 using the STRING database were analyzed. Mitochondrial transcription termination factor 4 [MTERFD2],

NOP14 nucleolar protein [NOP14] and RB transcriptional corepressor 1 [RB1] were identified to be closely related to NSUN4, NSUN7 and DNMT1 respectively with the highest predicted scores. As shown in Supplementary Figure S9A, NOP14 was significantly upregulated in rectal cancer tissues compared with normal tissues, while both MTERFD2 and RB1 showed no differences. Consistent with our results in TCGA, the immunohistochemistry results of the HPA database presented that the protein expression level of NOP14 was elevated in the tumor cells compared with the corresponding glandular cells, and mainly localized to the cytoplasmic and membranous nuclear (Supplementary Figure S9B). However, the protein expression of MTERFD2 or RB1 exhibited no difference between cancer tissues and normal tissues (Supplementary Figure S9B).

Estimation of drug sensitivity for the m^5C methylation-based signature

Based on the potential role played by the established m^5C regulator signature in modulating the immunotherapies, we further investigated its clinical usefulness by measuring the IC50



value of different oncology drugs. According to the predictive model, we found that the effects of 10 commonly used drugs for READ were different between the two subgroups. Chemotherapeutic drugs, including camptothecin, 5-fluorouracil, cisplatin, oxaliplatin, and irinotecan, had a lower IC₅₀ in the low-risk group.

Similarly, cediranib, sorafenib, and axitinib, which belong to VEGFR-targeted angiogenesis drugs, exhibited a lower IC₅₀ in the low-risk group. EGFR/HER2 inhibitor lapatinib and BRAF inhibitor dabrafenib also followed that pattern (Figure 12A). To benefit high-risk patients, we further excavated both the CTRP and PRISM databases; two drugs specific to high-risk patients were found effective by intersecting the two sources (Figures 12B–D) and include chlorambucil and SKI.II. These results implied that our model could predict certain drug sensitivity that would be beneficial to different groups of READ patients.

Discussion

Accumulating studies have revealed that colon and rectal cancer have distinct metastatic patterns, spread ratio, and drug response in patients (46). In multiple trials, individuals with rectal or colon

cancer who received bevacizumab-containing regimens have shown different survival rates (47–49). In order to systematically distinguish colon and rectal cancer, Liang et al. even profiled specific biomarker and identified a key factor to tailor the medical treatment of patients with colon and rectal cancer (50). Available evidence indicates that colon and rectal cancer should be regarded as two specific cancers when considering clinical treatment. Therefore, we evaluated the prognostic significance of m⁵C regulators in COAD and READ separately. The results indicate that m⁵C might exert more impact on the prognosis of READ patients than COAD patients, which could be explained by the fact that colon and rectal cancer exhibit remarkably different genetic and epigenetic characteristics. A study enrolling 1,443 stage I–IV CRC patients revealed that the prevalence of MSI-high, BRAF mutations, and CIMP-high tumors rapidly decreased from the proximal colon to the rectum (51). Moreover, proximal tumors were more frequently MSI, hypermutated, BRAF mutant, and densely infiltrated by TIL, whereas distal tumors were CIN, HER1, and HER2 amplified, with active EGFR signaling and mostly non-BRAF-like characteristics according to an analysis of molecular features along anatomical sites in colon carcinomas of patients enrolled in the Pan European Trial Adjuvant Colon Cancer-3 (PETACC3) chemotherapy trial (52), indicating a great heterogeneity within CRC. Overall, the observation of significant

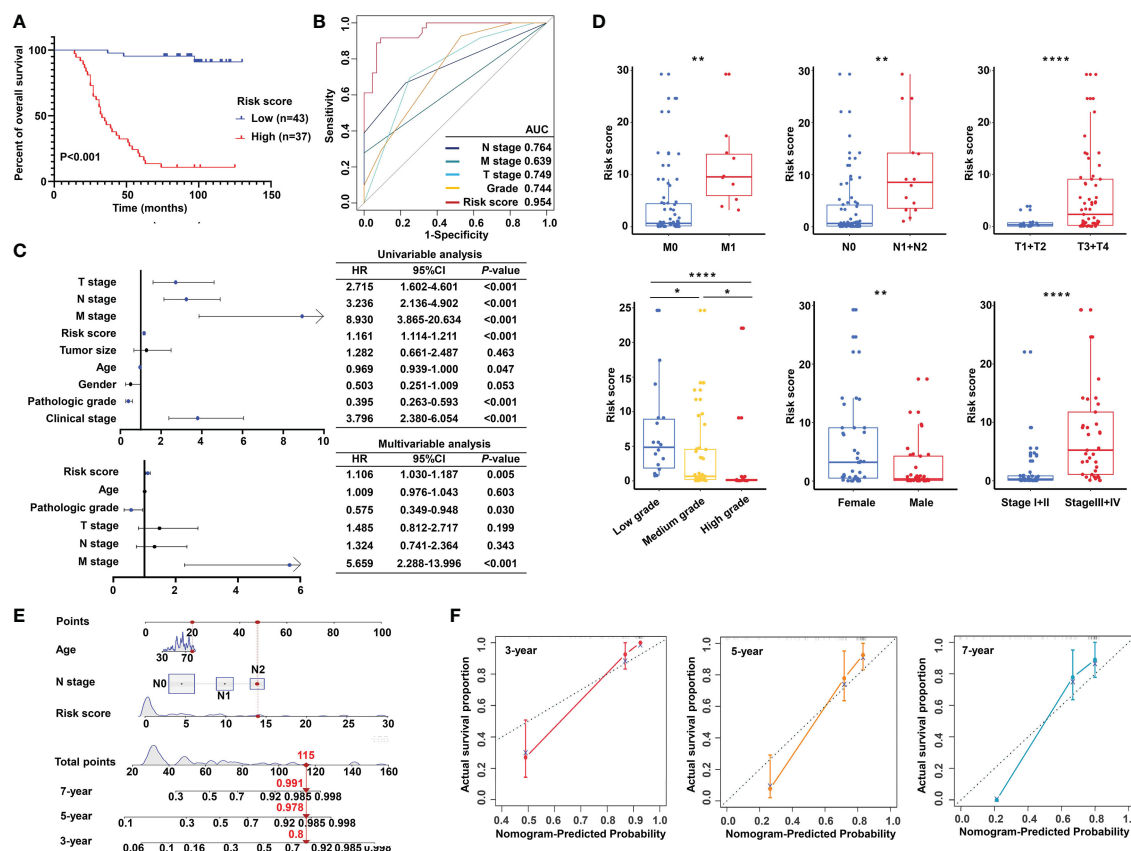


FIGURE 11

Validation of the m^5C methylation-based signature by tissue microarray (TMA). (A) Kaplan–Meier survival curves show overall survival for low- and high-risk patients based on the rectal cancer TMA cohort. (B) The ROC curves depicted for the risk score and common clinical diagnostic indexes. (C) The univariate and multivariate Cox analysis of the risk score and clinicopathological indexes in the rectal cancer TMA patients. (D) The distribution of the risk score among various parameters including pathological TNM, stage, grade, and gender. (E) A nomogram integrated age, pathological N, and risk score was constructed for 3, 5, and 7 years based on the rectal cancer TMA cohort. (F) The calibration curves show the discrepancy between actual and nomogram-predicted survival probability in 3-, 5-, and 7-year nomograms. **** $P < 0.0001$; ** $P < 0.01$; * $P < 0.05$.

difference between two types of cancer led us to focus our research on patients with rectal cancer.

RNA epigenetic modification is a crucial biological process. There is increasing evidence that the malfunction of RNA epigenetic modification leads to the deterioration of cancers. For example, NUSN4 has been found to affect the expression of mitochondrial DNA, which leads to a cascade of changes relating with the regulation of mammalian oxidative phosphorylation, finally resulting in the progression of cancers (53–55). The dysfunction of NSUN7 has been reported to result in male infertility (56), and NSUN7 is downregulated in prostate cancer compared with normal prostate tissue, acting as a protective factor in patients with prostate cancer (57). Additionally, DNMT1 is an important methyltransferase for the stable process of RNA methylation. It is associated with a series of cancers, including breast cancer, thyroid cancer, pancreatic cancer, and hepatocellular carcinoma (58–61). Here, based on the gene expression of m^5C regulators (NSUN4, NSUN7, DNMT1), we established a signature that could effectively distinguish the prognosis of READ patients. A weak positive correlation was found between the three genes based on the TCGA-READ, indicating the independence of the three genes in the current model, and their cumulative effect can endow

the model biological significance at the mRNA expression level. The constructed signature, age, and pathologic N act as independent prognostic factors in rectal cancer. Moreover, the signature could predict risk for patients of different age groups and N stages. Notably, the signature failed to distinguish the survival status of patients in the N1/N2 stage. At the advanced stage of the disease, colorectal cancer-associated immune infiltrates can be highly heterogeneous and can vary their phenotypes in a spatiotemporal manner (62, 63). Moreover, various factors such as intestinal obstruction, gastro-intestinal bleeding, malnutrition, liver metastasis, and other maladies can cause death in advanced colorectal cancer. All the above uncertainties could account for the reason that risk score only exhibited a trend ($P = 0.089$) when stratifying the overall survival of patients in the pathological N1/N2 stage due to small sample size. Differentiating the survival status of N0 patients is significant for early intervention. Colorectal cancer develops asymptotically, leading to the difficulty in diagnosis and thus progressing into the advanced stage, which requires considerable efforts to treat (64). The established risk score could efficiently evaluate the hazards of patients in the pathological N0 stage and predict patients who are at high risk of developing advanced stage cancer, and the emphasis placed on these patients

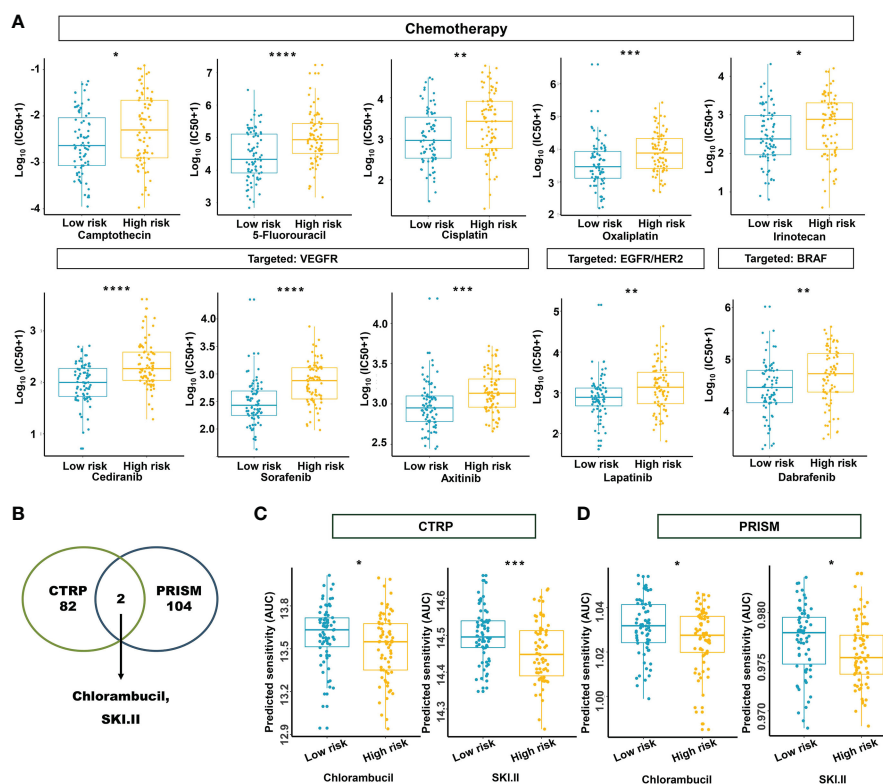


FIGURE 12

Estimation of drug sensitivity for the m⁵C methylation-based signature. **(A)** The evaluation of drug sensitivity including chemotherapeutics and small molecular drugs targeting VEGFR, EGFR/HER2, and BRAF. **(B)** Intersection of the identified drugs targeting high-risk patients between CTRP and PRISM databases. **(C, D)** The differential drug response analysis of CTRP- and PRISM-derived compounds targeting the high-risk group. IC₅₀: half-maximal inhibitory concentration. *****P* < 0.0001; ****P* < 0.001; ***P* < 0.01; **P* < 0.05.

will benefit them clinically. Since IHC enables a pathologist to examine gene expression at the protein level within the context of histologically interpretable tissue sections, it is a reliable method for confirming expression signatures discovered by RNA sequencing. Therefore, to further substantiate the results of the bioinformatics analysis, TMAs from patients with rectal cancer were immunohistochemically stained for NSUN4, NSUN7, and DNMT1. The stained slides were evaluated for calculating risk score. In concordance with the TCGA data mining, the risk score was able to differentiate the prognosis of patients with rectal cancer well and determine their survival as an independent prognostic factor, and the nomogram integrating risk score, age, and pathological N could serve as a reliable indicator in predicting the survival probability of patients with rectal cancer. Since IHC is carried out on commonly processed clinical tissue samples, validated IHC assays could be easily applied in clinical diagnostics. To facilitate the clinical use, we developed a nomogram with high accuracy and robustness. Our findings together suggest that the built signature based on m⁵C RNA regulators is highly involved in the progression of rectal cancer and could serve for effective risk stratification in patients with rectal cancer.

There is increasing evidence relating the m⁵C modification with innate immunity as well as antitumor effect through a complex crosstalk among various m⁵C regulators. We found that the

established signature could effectively determine the TIME infiltration patterns. The interplay between tumor and immunity begins when tumor antigens are presented by dendritic cells and activate CD8+ T cells and CD4+ T cells to exert cytotoxic effects (65). Moreover, cancer cells can suppress immune system, leading to an inhibitory TIME to escape immune surveillance with the increase of Tregs and MDSC. As revealed in our analysis integrating CYBERSORT, TIMER, EPIC, and ssGSEA algorithms, the low-risk group was characterized by the activation of adaptive immunity, with the increasing abundance of CD4+ T cells, CD8+ T cells, B cells, and myeloid dendritic cells. The high-risk group was characterized by the suppression of immunity, accompanied by upregulation of Tregs and MDSC. The ratio analysis further explained that compared with the high-risk group, the scales of CD8+/CD4+ regulatory T cells and pro-/anti-inflammatory cytokines were higher in the low-risk group. According to different functions, macrophages could be classified into two categories: classically activated macrophages (M1), mainly acting as a tumor-killer role, and alternatively activated macrophages (M2), which function to promote tumor cells (66). As indicated in our results, the ratio of M1 to M2 macrophages was elevated in the low-risk group. m⁵C RNA methylation regulators have already demonstrated the efficacy for predicting prognosis and regulating TIME in various cancers (18, 67, 68), suggesting the potential value in pan-cancer analysis. Consistent with the current knowledge, our

model showed a predictive accuracy in prognosis and in TIME cell infiltration characterization among READ patients.

The signatures derived from m6A/m5C/m1A RNA methylation regulators were widely explored in recent studies. Commonly, the signatures could characterize the immune landscape of cancer patients and further predict the efficacy of immunotherapy (69, 70). m6A modification is one of the most researched RNA methylation patterns. The “writer”, “reader”, and “eraser” of m6A modification correlated closely with immune infiltrating cells (71), giving rise to the application of m6A RNA methylation regulators in predicting immune efficacy. Two m6A RNA demethylases, FTO and ALKBH5 were targeted to develop inhibitors (72–74), providing insights into understanding the roles of m6A RNA methylation involved in multiple diseases. m5C RNA modification is regarded as a novel methylated process in eukaryotes. Small-molecular inhibitors targeting m5C RNA methylation regulators were conceived by proof-of-concept studies, while, specific m5C inhibitors have yet to be developed (75). m5C RNA methylation regulators can impact the process of tumorigenesis by regulating TIME in cancers, so that inspecting the roles involved in the immune system will give hints to personalized immunotherapy strategies making. m1A methylation modification is a new form of modification of RNA, thus, studies on m1A modification in tumorigenesis are rarely reported. Although several signatures based on m1A modification were built to guild effective immunotherapy strategies (70, 76), controversies remained when detecting the m1A methylation sites (77, 78). More efficient and accurate technologies need to be developed to uncover the m1A modification sites to fully exploit the value of m1A modification in anti-tumor immunotherapies. More effort is deserved to understand the complex network regulated by different kinds of RNA methylations in modulating tumor-immune interactions. However, in the current study, we focused on the prospects of m5C methylation regulator as the predictive biomarker for ICIs treatment.

The quantity of cancer mutations is reflected by TMB. Major histocompatibility complex proteins turn mutations into neoantigens and further present them to T cells. More neoantigens are produced by higher TMB, which in turn boosts the likelihood that T cells recognition will happen, clinically corresponding with improved ICI outcomes (79). Several studies have shown that high TMB and neoantigens correlated with better prognosis in non-small-cell lung cancer (NSCLC) and melanoma (80–82). In this study, the low-risk group possessed more mutations and higher level of neoantigens than the high-risk group, suggesting a better response to immunotherapy within the low-risk group. We also identified the stratifying efficiency of the model in patients with same status of neoantigens and TMB. The prognostic power of the established model was superior to neoantigens or TMB. These results indicated that our model had the potential to combine with or modify existing biomarkers, achieving improved accuracy in prognostic prediction. In addition to using neoantigens and TMB, immune checkpoints can be inhibited to enable T cell functions. By allowing T-cell reactivation, ICIs have revolutionized cancer treatment (83). The Food and Drug Administration (FDA)

has approved six inhibitors of the programmed cell death protein pathway (PD1/PD-L1) and an inhibitor of the CTLA-4 for use in treating various cancers (84–86). In our study, we observed a weak correlation between model factors and immune checkpoints except for NSUN4. In fact, immune checkpoints alone are not sufficient to predict the efficiency of the immunotherapy due to a highly complex immune tumor microenvironment, which could be generalized by a cancer immunity cycle (87). Several studies have suggested integrating multiple biomarkers to predict the immune response, including tumor-infiltrating lymphocytes, mutational burden, immune gene signatures, and multiplex immunohistochemistry (88, 89). TIDE is a reliable surrogate biomarker that could accurately predict immune checkpoint blockade (ICB) response by measuring the tumor immune escape, and it even performed better than PD-L1 expression in melanoma; that is, a higher TIDE score is associated with worse ICB response and worse patient survival under anti-PD1 and anti-CTLA4 therapies (90). According to our previous studies and others, the immune landscape is crucial in assessing the efficacy of immunotherapy and chemotherapy targeting CRC patients (91–94). However, the role of m⁵C RNA methylation regulators in patients with rectal cancer is still unclear. In the current research, we found that responders were proportionally more frequent in the low-risk group compared with the high-risk group. The lower TIDE prediction score, T-cell dysfunction score, CAF, and higher MSI score in the low-risk group indicate a good function of T cells with high infiltration by cytotoxic T lymphocyte (CTL), further explaining why the low-risk group was more sensitive to immunotherapy. In addition, in IMvigor210 cohort with the determined immune response, these results were well confirmed. Besides, drug sensitivity was examined between the low- and high-risk groups by performing the R package “oncoPredict”. Apparently, the majority of the chemotherapeutic agents achieved their efficacy among the low-risk group; nonetheless, drugs targeting specifically the high-risk group were also investigated by screening drugs of CTRP and PRISM databases. The AUC values between two risk groups were compared and drugs intended to the high-risk group were selected. Finally, chlorambucil and SKI.II were found in both the CTRP and PRISM databases. These results indicated the built risk model was a trustworthy and robust approach for a thorough evaluation of each patient’s therapeutic response, which could benefit the precision treatment combining immunotherapy and chemotherapy for patients with rectal cancer.

Furthermore, the mRNA transcriptome differences between the high- and low-risk groups have been investigated. They were highly involved in the cancer and immune system-related biological pathways. The DEGs with prognostic efficacy were considered m⁵C-related signature genes. Two genomic subgroups were discovered based on the m⁵C signature genes, which could significantly predict the survival and immune response of READ patients, and were substantially connected with immunological activity. These results were similar to the stratification of the risk model. This once again showed the power of the m⁵C regulator-based signature in shaping the landscapes of the READ patients. Thus, a thorough analysis of m⁵C alteration patterns will definitely

improve the precision classification and therapeutic strategy for patients with READ.

Despite the encouraging findings, the current study included several limitations. First, the gathered data were analyzed retrospectively, and multicenter research and large-scale prospective investigation are required to confirm and rectify our model. Second, the specific crosstalk between these m⁵C methylation regulators and corresponding immune characteristics remains unrevealed. The regulatory network of the three genes in rectal cancer needs to be further investigated. As for now, the genes regulated by NSUN4 and NSUN7 still need to be identified. Research related to the regulatory role of the three genes could provide novel insights into the mechanisms of the built signature. Third, the ability of this signature to predict immunotherapeutic or chemotherapeutic response was assessed indirectly due to the lack of data from patients with rectal cancer receiving related treatments. Research focusing on the therapeutic effect of the current signatures should be done *in vitro* and *in vivo* in the future. Fourth, the sizes of clinical tissue specimens for TMA and RT-qPCR assay used in our independent validation cohorts were limited, and more samples are expected to verify the m⁵C methylation regulator –based signature in the future.

In conclusion, the established risk model could be used to comprehensively evaluate the prognosis and the clinical response to adjuvant chemotherapy and immunotherapy among patients with rectal cancer. Moreover, the complex characteristic of the TIME cell infiltration could be effectively illustrated by the built signatures based on m⁵C regulators, producing a number of novel insights for cancer immunotherapy. Our research offers fresh approaches for predicting survival status, enhancing immunotherapy outcomes, disclosing various tumor immune phenotypes, and conclusively, advancing tailored cancer treatment in the future.

Data availability statement

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by the Research Ethics Committee of The Affiliated Hospital of Qingdao University. The patients/participants provided their written informed consent to participate in this study.

References

1. Gong J, Chehraz-Raffle A, Reddi S, Salgia R. Development of pd-1 and pd-L1 inhibitors as a form of cancer immunotherapy: A comprehensive review of registration trials and future considerations. *J immunother cancer* (2018) 6(1):8. doi: 10.1186/s40425-018-0316-z
2. Akinleye A, Rasool Z. Immune checkpoint inhibitors of pd-L1 as cancer therapeutics. *J Hematol Oncol* (2019) 12(1):92. doi: 10.1186/s13045-019-0779-5
3. Angell H, Galon J. From the immune contexture to the immunoscore: The role of prognostic and predictive immune markers in cancer. *Curr Opin Immunol* (2013) 25(2):261–7. doi: 10.1016/j.coi.2013.03.004
4. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature* (2017) 541(7637):321–30. doi: 10.1038/nature21349

Author contributions

MY and RZ conceived and designed the experiments. MY supervised the work. JZ, RZ, MZ, and ZY provisioned the study materials or patients. RZ, JZ, TL, and SL collected and assembled the data. RZ, ZZ, WW, and FZ analyzed and interpreted the data. MY and RZ wrote the article. The final manuscript was read and approved by all authors. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the grants from CAMS Innovation Fund for Medical Sciences (CIFMS) (No. 2021-I2M-1-028) and the Natural Science Foundation of China (NSFC) (No. 81773750).

Acknowledgments

The authors are thankful for the publicly available databases they used in this work.

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.

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.2023.1054700/full#supplementary-material>

5. Joyce JA, Fearon DT. T Cell exclusion, immune privilege, and the tumor microenvironment. *Sci (New York NY)* (2015) 348(6230):74–80. doi: 10.1126/science.aaa6204
6. Herbst RS, Soria JC, Kowanetz M, Fine GD, Hamid O, Gordon MS, et al. Predictive correlates of response to the anti-Pd-L1 antibody Mpd3280a in cancer patients. *Nature* (2014) 515(7528):563–7. doi: 10.1038/nature14011
7. Ganesh K, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, et al. Immunotherapy in colorectal cancer: Rationale, challenges and potential. *Nat Rev Gastroenterol hepatology*. (2019) 16(6):361–75. doi: 10.1038/s41575-019-0126-x
8. Llosa NJ, Cruise M, Tam A, Wicks EC, Hechenbleikner EM, Taube JM, et al. The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints. *Cancer discovery* (2015) 5(1):43–51. doi: 10.1158/2159-8290.cd-14-0863
9. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. Pd-1 blockade in tumors with mismatch-repair deficiency. *New Engl J Med* (2015) 372(26):2509–20. doi: 10.1056/NEJMoa1500596
10. Boland CR, Goel A. Microsatellite instability in colorectal cancer. *Gastroenterology* (2010) 138(6):2073–87.e3. doi: 10.1053/j.gastro.2009.12.064
11. Roundtree IA, Evans ME, Pan T, He C. Dynamic rna modifications in gene expression regulation. *Cell* (2017) 169(7):1187–200. doi: 10.1016/j.cell.2017.05.045
12. Boccaletto P, Machnicka MA, Purta E, Piatkowski P, Baginski B, Wirecki TK, et al. Modomics: A database of rna modification pathways. 2017 update. *Nucleic Acids Res* (2018) 46(D1):D303–d7. doi: 10.1093/nar/gkx1030
13. Tang Y, Chen K, Song B, Ma J, Wu X, Xu Q, et al. M6a-atlas: A comprehensive knowledgebase for unraveling the N6-methyladenosine (M6a) epitranscriptome. *Nucleic Acids Res* (2021) 49(D1):D134–d43. doi: 10.1093/nar/gkaa692
14. Lyko F. The DNA methyltransferase family: A versatile toolkit for epigenetic regulation. *Nat Rev Genet* (2018) 19(2):81–92. doi: 10.1038/nrg.2017.80
15. Kohli RM, Zhang Y. Tet enzymes, tdk and the dynamics of DNA demethylation. *Nature* (2013) 502(7472):472–9. doi: 10.1038/nature12750
16. Li LH, Olin EJ, Buskirk HH, Reineke LM. Cytotoxicity and mode of action of 5-azacytidine on L1210 leukemia. *Cancer Res* (1970) 30(11):2760–9.
17. Zhang M, Song J, Yuan W, Zhang W, Sun Z. Roles of rna methylation on tumor immunity and clinical implications. *Front Immunol* (2021) 12:641507. doi: 10.3389/fimmu.2021.641507
18. Pan J, Huang Z, Xu Y. M5c rna methylation regulators predict prognosis and regulate the immune microenvironment in lung squamous cell carcinoma. *Front Oncol* (2021) 11:657466. doi: 10.3389/fonc.2021.657466
19. Gao L, Chen R, Sugimoto M, Mizuta M, Zhou L, Kishimoto Y, et al. The rna methylation modification 5-methylcytosine impacts immunity characteristics, prognosis and progression of oral squamous cell carcinoma by bioinformatics analysis. *Front bioengineering Biotechnol* (2021) 9:760724. doi: 10.3389/fbioe.2021.760724
20. Hu Y, Gaedcke J, Emons G, Beissbarth T, Grade M, Jo P, et al. Colorectal cancer susceptibility loci as predictive markers of rectal cancer prognosis after surgery. *Genes Chromosomes cancer* (2018) 57(3):140–9. doi: 10.1002/gcc.22512
21. Ferrandon S, DeVecchio J, Duraes L, Chouhan H, Karagkounis G, Davenport J, et al. Coa synthase (Coasy) mediates radiation resistance Via P13k signaling in rectal cancer. *Cancer Res* (2020) 80(2):334–46. doi: 10.1158/0008-5472.can-19-1161
22. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* (2018) 28(11):1747–56. doi: 10.1101/gr.239244.118
23. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep* (2017) 18(1):248–62. doi: 10.1016/j.celrep.2016.12.019
24. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med* (2015) 21(11):1350–6. doi: 10.1038/nm.3967
25. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. String: A database of predicted functional associations between proteins. *Nucleic Acids Res* (2003) 31(1):258–61. doi: 10.1093/nar/gkg034
26. 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(21):e108–e10. doi: 10.1158/0008-5472.Can-17-0307
27. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. tissue-based map of the human proteome. *Sci (New York NY)* (2015) 347(6220):1260419. doi: 10.1126/science.1260419
28. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models Via coordinate descent. *J Stat Software* (2010) 33(1):1–22. doi: 10.18637/jss.v033.i01
29. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* (2019) 10(1):1523. doi: 10.1038/s41467-019-09234-6
30. Walter W, Sánchez-Cabo F, Ricote M. Gplot: An r package for visually combining expression data with functional analysis. *Bioinformatics* (2015) 31(17):2912–4. doi: 10.1093/bioinformatics/btv300
31. Hänzelmann S, Castelo R, Guinney J. Gsva: Gene set variation analysis for microarray and rna-seq data. *BMC Bioinf* (2013) 14:7. doi: 10.1186/1471-2105-14-7
32. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. Tgfb attenuates tumour response to pd-L1 blockade by contributing to exclusion of T cells. *Nature* (2018) 554(7693):544–8. doi: 10.1038/nature25501
33. Meng Q, Lu YX, Ruan DY, Yu K, Chen YX, Xiao M, et al. DNA Methylation regulator-mediated modification patterns and tumor microenvironment characterization in gastric cancer. *Mol Ther Nucleic Acids* (2021) 24:695–710. doi: 10.1016/j.omtn.2021.03.023
34. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* (2015) 12(5):453–7. doi: 10.1038/nmeth.3337
35. 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
36. Li L, Li M, Jiang Z, Wang X. Arid1a mutations are associated with increased immune activity in gastrointestinal cancer. *Cells* (2019) 8(7):678. doi: 10.3390/cells8070678
37. Xu L, Deng C, Pang B, Zhang X, Liu W, Liao G, et al. Tip: A web server for resolving tumor immunophenotype profiling. *Cancer Res* (2018) 78(23):6575. doi: 10.1158/0008-5472.CAN-18-0689
38. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T-H, et al. The immune landscape of cancer. *Immunity* (2018) 48(4):812–30.e14. doi: 10.1016/j.immuni.2018.03.023
39. Rooney Michael S, Shukla Sachet A, Wu Catherine J, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* (2015) 160(1):48–61. doi: 10.1016/j.cell.2014.12.033
40. Fu J, Li K, Zhang W, Wan C, Zhang J, Jiang P, et al. Large-Scale public data reuse to model immunotherapy response and resistance. *Genome Med* (2020) 12(1):21–. doi: 10.1186/s13073-020-0721-z
41. Maeser D, Gruener RF, Huang RS. Oncopredict: An r package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Briefings Bioinf* (2021) 22(6):bbab260. doi: 10.1093/bib/bbab260
42. Rees MG, Seashore-Ludlow B, Cheah JH, Adams DJ, Price EV, Gill S, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol* (2016) 12(2):109–16. doi: 10.1038/nchembio.1986
43. Beijersbergen RL. Old drugs with new tricks. *Nat Cancer* (2020) 1(2):153–5. doi: 10.1038/s43018-020-0024-8
44. Wilkerson MD, Hayes DN. Consensusclusterplus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* (2010) 26(12):1572–3. doi: 10.1093/bioinformatics/btq170
45. Yang X, Yang Y, Sun BF, Chen YS, Xu JW, Lai WY, et al. 5-methylcytosine promotes mrna export - Nsun2 as the methyltransferase and alyref as an M(5)C reader. *Cell Res* (2017) 27(5):606–25. doi: 10.1038/cr.2017.55
46. Riihimäki M, Hemminki A, Sundquist J, Hemminki K. Patterns of metastasis in colon and rectal cancer. *Sci Rep* (2016) 6:29765. doi: 10.1038/srep29765
47. Hurwitz H, Fehrenbacher L, Novotny W, Cartwright T, Hainsworth J, Heim W, et al. Bevacizumab plus capecitabine versus capecitabine alone in elderly patients with previously untreated metastatic colorectal cancer (Avex): An open-label, randomised phase 3 trial. *Lancet Oncol* (2013) 14(11):1077–85. doi: 10.1016/s1470-2045(13)70154-2
49. Grothey A, Sugrue MM, Purdie DM, Dong W, Sargent D, Hedrick E, et al. Bevacizumab beyond first progression is associated with prolonged overall survival in metastatic colorectal cancer: Results from a Large observational cohort study (Brite). *J Clin Oncol Off J Am Soc Clin Oncol* (2008) 26(33):5326–34. doi: 10.1200/jco.2008.16.3212
50. Liang JZ, Liang XL, Zhong LY, Wu CT, Zhang J, Wang Y. Comparative proteome identifies complement component 3-mediated immune response as key difference of colon adenocarcinoma and rectal adenocarcinoma. *Front Oncol* (2020) 10:617890. doi: 10.3389/fonc.2020.617890
51. Yamauchi M, Morikawa T, Kuchiba A, Imamura Y, Qian ZR, Nishihara R, et al. Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum. *Gut* (2012) 61(6):847–54. doi: 10.1136/gutjnl-2011-300865
52. Missiaglia E, Jacobs B, D'Ario G, Di Narzo AF, Soneson C, Budinska E, et al. Distal and proximal colon cancers differ in terms of molecular, pathological, and clinical features. *Ann Oncol Off J Eur Soc Med Oncol* (2014) 25(10):1995–2001. doi: 10.1093/annonc/mdl275
53. Cámara Y, Asin-Cayuela J, Park CB, Metodieva MD, Shi Y, Ruzzenente B, et al. Mterf4 regulates translation by targeting the methyltransferase Nsun4 to the mammalian mitochondrial ribosome. *Cell Metab* (2011) 13(5):527–39. doi: 10.1016/j.cmet.2011.04.002
54. Tan JL, Li F, Yeo JZ, Yong KJ, Bassal MA, Ng GH, et al. New high-throughput screening identifies compounds that reduce viability specifically in liver cancer cells

that express high levels of Sall4 by inhibiting oxidative phosphorylation. *Gastroenterology* (2019) 157(6):1615–29.e17. doi: 10.1053/j.gastro.2019.08.022

55. Rodríguez-Enríquez S, Pacheco-Velázquez SC, Marín-Hernández Á, Gallardo-Pérez JC, Robledo-Cadena DX, Hernández-Reséndiz I, et al. Resveratrol inhibits cancer cell proliferation by impairing oxidative phosphorylation and inducing oxidative stress. *Toxicol Appl Pharmacol* (2019) 370:65–77. doi: 10.1016/j.taap.2019.03.008

56. Chi L, Delgado-Olguin P. Expression of Noli1/Nop2/Sun domain (Nsun) rna methyltransferase family genes in early mouse embryogenesis. *Gene Expression Patterns* (2013) 13(8):319–27. doi: 10.1016/j.gexp.2013.06.003

57. Xing Q, Liu S, Luan J, Wang Y, Ma L. A novel 13 rna binding proteins (Rbps) signature could predict prostate cancer biochemical recurrence. *Pathol - Res Practice* (2021) 225:153587. doi: 10.1016/j.prp.2021.153587

58. Zhang Y, Sun B, Huang Z, Zhao DW, Zeng Q. Shikonin inhibits migration and invasion of thyroid cancer cells by downregulating Dnmt1. *Med Sci monitor* (2018) 24:661–70. doi: 10.12659/msm.908381

59. Wang P, Chu W, Zhang X, Li B, Wu J, Qi L, et al. Kindlin-2 interacts with and stabilizes Dnmt1 to promote breast cancer development. *Int J Biochem Cell Biol* (2018) 105:41–51. doi: 10.1016/j.biocel.2018.09.022

60. Peng DF, Kanai Y, Sawada M, Ushijima S, Hiraoka N, Kosuge T, et al. Increased DNA methyltransferase 1 (Dnmt1) protein expression in precancerous conditions and ductal carcinomas of the pancreas. *Cancer science* (2005) 96(7):403–8. doi: 10.1111/j.1349-7006.2005.00071.x

61. Gu X, Zhou H, Chu Q, Zheng Q, Wang J, Zhu H. Uncovering the association between M(5)C regulator-mediated methylation modification patterns and tumour microenvironment infiltration characteristics in hepatocellular carcinoma. *Front Cell Dev Biol* (2021) 9:727935. doi: 10.3389/fcell.2021.727935

62. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenaus AC, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* (2013) 39(4):782–95. doi: 10.1016/j.immuni.2013.10.003

63. Van den Eynde M, Mlecnik B, Bindea G, Fredriksen T, Church SE, Lafontaine L, et al. The link between the multiverse of immune microenvironments in metastases and the survival of colorectal cancer patients. *Cancer Cell* (2018) 34(6):1012–26.e3. doi: 10.1016/j.ccell.2018.11.003

64. Tepus M, Yau TO. Non-invasive colorectal cancer screening: An overview. *Gastrointestinal tumors* (2020) 7(3):62–73. doi: 10.1159/000507701

65. den Haan JM, Lehar SM, Bevan MJ. Cd8+ but not Cd8– dendritic cells cross-prime cytotoxic T cells in vivo. *J Exp Med* (2000) 192(12):1685–96. doi: 10.1084/jem.192.12.1685

66. Brown JM, Recht L, Strober S. The promise of targeting macrophages in cancer therapy. *Clin Cancer Res* (2017) 23(13):3241–50. doi: 10.1158/1078-0432.ccr-16-3122

67. Wu XR, Chen Z, Liu Y, Chen ZZ, Tang F, Chen ZZ, et al. Prognostic signature and immune efficacy of M(1) a-, M(5) c- and M(6) a-related regulators in cutaneous melanoma. *J Cell Mol Med* (2021) 25(17):8405–18. doi: 10.1111/jcmm.16800

68. Liu Y, Zheng S, Wang T, Fang Z, Kong J, Liu J. Identification of the expression patterns and potential prognostic role of 5-methylcytosine regulators in hepatocellular carcinoma. *Front Cell Dev Biol* (2022) 10:842220. doi: 10.3389/fcell.2022.842220

69. Zhang T, Liu H, Gao F, Gong W, Cui Y, He J, et al. M6a-regulator expression signatures identify a subset of follicular lymphoma harboring an exhausted tumor microenvironment. *Front Immunol* (2022) 13:922471. doi: 10.3389/fimmu.2022.922471

70. Gao Y, Wang H, Li H, Ye X, Xia Y, Yuan S, et al. Integrated analyses of M(1)a regulator-mediated modification patterns in tumor microenvironment-infiltrating immune cells in colon cancer. *Oncoimmunology* (2021) 10(1):1936758. doi: 10.1080/2162402x.2021.1936758

71. Zhang B, Wu Q, Li B, Wang D, Wang L, Zhou YL. M(6)a regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. *Mol cancer* (2020) 19(1):53. doi: 10.1186/s12943-020-01170-0

72. Xu C, Liu K, Tempel W, Demetriades M, Aik W, Schofield CJ, et al. Structures of human Alkbh5 demethylase reveal a unique binding mode for specific single-stranded N6-methyladenosine rna demethylation. *J Biol Chem* (2014) 289(25):17299–311. doi: 10.1074/jbc.M114.550350

73. Qiao Y, Zhou B, Zhang M, Liu W, Han Z, Song C, et al. A novel inhibitor of the obesity-related protein fto. *Biochemistry* (2016) 55(10):1516–22. doi: 10.1021/acs.biochem.6b00023

74. Chen B, Ye F, Yu L, Jia G, Huang X, Zhang X, et al. Development of cell-active N6-methyladenosine rna demethylase fto inhibitor. *J Am Chem Society* (2012) 134(43):17963–71. doi: 10.1021/ja3064149

75. Song H, Zhang J, Liu B, Xu J, Cai B, Yang H, et al. Biological roles of rna M(5)C modification and its implications in cancer immunotherapy. *biomark Res* (2022) 10(1):15. doi: 10.1186/s40364-022-00362-8

76. Liu J, Chen C, Wang Y, Qian C, Wei J, Xing Y, et al. Comprehensive of N1-methyladenosine modifications patterns and immunological characteristics in ovarian cancer. *Front Immunol* (2021) 12:746647. doi: 10.3389/fimmu.2021.746647

77. Safra M, Sas-Chen A, Nir R, Winkler R, Nachshon A, Bar-Yaacov D, et al. The M1a landscape on cytosolic and mitochondrial mrna at single-base resolution. *Nature* (2017) 551(7679):251–5. doi: 10.1038/nature24456

78. Schwartz S. M(1)a within cytoplasmic mrnas at single nucleotide resolution: A reconciled transcriptome-wide map. *RNA (New York NY)* (2018) 24(11):1427–36. doi: 10.1261/rna.067348.118

79. Jardim DL, Goodman A, de Melo Gagliato D, Kurzrock R. The challenges of tumor mutational burden as an immunotherapy biomarker. *Cancer Cell* (2021) 39(2):154–73. doi: 10.1016/j.ccell.2020.10.001

80. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Sci (New York NY)* (2016) 351(6280):1463–9. doi: 10.1126/science.aaf1490

81. Singal G, Miller PG, Agarwala V, Li G, Kaushik G, Backenroth D, et al. Association of patient characteristics and tumor genomics with clinical outcomes among patients with non-small cell lung cancer using a clinicogenomic database. *JAMA* (2019) 321(14):1391–9. doi: 10.1001/jama.2019.3241

82. Chan TA, Wolchok JD, Snyder A. Genetic basis for clinical response to ctla-4 blockade in melanoma. *New Engl J Med* (2015) 373(20):1984. doi: 10.1056/NEJMcl1508163

83. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. *Sci (New York NY)* (2018) 359(6382):1350–5. doi: 10.1126/science.aar4060

84. Bellmunt J, de Wit R, Vaughn DJ, Fradet Y, Lee JL, Fong L, et al. Pembrolizumab as second-line therapy for advanced urothelial carcinoma. *New Engl J Med* (2017) 376(11):1015–26. doi: 10.1056/NEJMoa1613683

85. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WE, Poddubskaya E, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *New Engl J Med* (2015) 373(2):123–35. doi: 10.1056/NEJMoa1504627

86. Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved survival with ipilimumab in patients with metastatic melanoma. *New Engl J Med* (2010) 363(8):711–23. doi: 10.1056/NEJMoa1003466

87. Chen DS, Mellman I. Oncology meets immunology: The cancer-immunity cycle. *Immunity* (2013) 39(1):1–10. doi: 10.1016/j.immuni.2013.07.012

88. Gibney GT, Weiner LM, Atkins MB. Predictive biomarkers for checkpoint inhibitor-based immunotherapy. *Lancet Oncol* (2016) 17(12):e542–e51. doi: 10.1016/s1470-2045(16)30406-5

89. Chae YK, Pan A, Davis AA, Raparia K, Mohindra NA, Matsangou M, et al. Biomarkers for pd-1/Pd-L1 blockade therapy in non-small-cell lung cancer: Is pd-L1 expression a good marker for patient selection? *Clin Lung Cancer* (2016) 17(5):350–61. doi: 10.1016/j.clcc.2016.03.011

90. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* (2018) 24(10):1550–8. doi: 10.1038/s41591-018-0136-1

91. Hou Y, Zhang R, Zong J, Wang W, Zhou M, Yan Z, et al. Comprehensive analysis of a cancer-immunity cycle-based signature for predicting prognosis and immunotherapy response in patients with colorectal cancer. *Front Immunol* (2022) 13:892512. doi: 10.3389/fimmu.2022.892512

92. Zhang R, Li T, Wang W, Gan W, Lv S, Zeng Z, et al. Indoleamine 2, 3-dioxygenase 1 and Cd8 expression profiling revealed an immunological subtype of colon cancer with a poor prognosis. *Front Oncol* (2020) 10:594098. doi: 10.3389/fonc.2020.594098

93. Li T, Yan Z, Wang W, Zhang R, Gan W, Lv S, et al. Sema6b overexpression predicts poor prognosis and correlates with the tumor immunosuppressive microenvironment in colorectal cancer. *Front Mol biosciences* (2021) 8:687319. doi: 10.3389/fmolb.2021.687319

94. Wang W, Zhong Y, Zhuang Z, Xie J, Lu Y, Huang C, et al. Multiregion single-cell sequencing reveals the transcriptional landscape of the immune microenvironment of colorectal cancer. *Clin Trans Med* (2021) 11(1):e253. doi: 10.1002/ctm2.253



OPEN ACCESS

EDITED BY
Pooneh Mokarram,
Shiraz University of Medical Sciences, Iran

REVIEWED BY
Peter Wang,
Bengbu Medical College, China
T. D. Wang,
University of Michigan, United States

*CORRESPONDENCE
Yongfu Xiong
✉ Xiong-Yongfu@ansmc.edu.cn

[†]These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to
Gastrointestinal Cancers:
Colorectal Cancer,
a section of the journal
Frontiers in Oncology

RECEIVED 03 January 2023
ACCEPTED 27 February 2023
PUBLISHED 09 March 2023

CITATION
Zhou H, Li L, Chen J, Hou S, Zhou T and
Xiong Y (2023) Expression and prognostic
value of PRDX family in colon
adenocarcinoma by integrating
comprehensive analysis and
in vitro and *in vivo* validation.
Front. Oncol. 13:1136738.
doi: 10.3389/fonc.2023.1136738

COPYRIGHT
© 2023 Zhou, Li, Chen, Hou, Zhou and
Xiong. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). 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.

Expression and prognostic value of PRDX family in colon adenocarcinoma by integrating comprehensive analysis and *in vitro* and *in vivo* validation

He Zhou^{1,2,3†}, Lifa Li^{1,2†}, Jia Chen^{1,2}, Songlin Hou^{1,2},
Tong Zhou^{1,2} and Yongfu Xiong^{2,4*}

¹The Second Department of Gastrointestinal Surgery, Affiliated Hospital of North Sichuan Medical College, Nanchong, China, ²Institute of Hepatobiliary, Pancreatic and Intestinal Disease, North Sichuan Medical College, Nanchong, China, ³Laboratory of Cancer Biology Department of Oncology, University of Oxford, Oxford, United Kingdom, ⁴The Department of Hepatobiliary Surgery, Affiliated Hospital of North Sichuan Medical College, Nanchong, China

Background: The peroxiredoxin family, a crucial regulator of redox reactions, is strongly associated with various tumorigenesis. However, the role of peroxiredoxin4 (PRDX4) in colon adenocarcinoma (COAD) remains poorly understood.

Methods: Multicenter databases, including GEPIA, HPA, UALCAN, cBioPortal, cancerSEA, STRING, CCLE, and LinkedOmics, comprehensively analyzed transcriptional expression, prognostic value, genetic alterations, signaling pathways, and associated genes of the PRDXs in COAD patients. Colony formation, transwell, flow cytometry, sphere formation, and xenograft assays were performed to validate further *in vitro* and *in vivo*.

Results: Members of the PRDX family were differentially expressed in COAD, with each member showing varying degrees of genetic alterations. Intriguingly, only PRDX4 significantly correlated with COAD prognosis and stage. The single-cell sequencing suggested that PRDX4 is positively correlated with proliferation, apoptosis, and invasion, whereas negatively correlated with stemness. Moreover, PRDX4 involved in a series of critical biological processes, such as cell growth. Furthermore, *in vivo* and *in vitro* analyses indicated that knocking down PRDX4 inhibits the proliferation and invasion of HCT116 cells while promoting apoptosis and stemness.

Conclusions: We identified PRDX4 expression as a novel potential prognostic marker in COAD.

KEYWORDS

peroxiredoxin family, bioinformatics analysis, biomarker, prognosis, colorectal cancer

Introduction

Colorectal cancer (CRC) remains the third most common cancer and the second leading cause of cancer-related death worldwide, with approximately 1.9 million new cases and 900,000 deaths reported in 2020 (1). The incidence of CRC decreased in developed countries as early diagnosis, radiotherapy, and chemotherapy became more prevalent (2). In contrast, CRC incidence and mortality have rapidly increased in many low- and middle-income countries, such as Asia and South America (3). CRC still ranks among the top five tumor-related causes of death, and the incidence and mortality of CRC continue to increase annually in China, making it a significant public health concern (4, 5). Despite the various surgical-based treatments being applied to CRC, the 5-year survival rate of CRC patients in China is still not ideal (6). More than half of the patients were diagnosed in the middle and advanced stages and lost the chance of surgery.

Evidence suggests that CRC develops through a multi-step process, where abnormal cell proliferation is the initiator, and the invasion of advanced cancer cells is a leading cause of poor prognosis (7, 8). Currently, the diagnosis of CRC is mainly based on microscopic examination of tumor specimens, tumor node metastasis (TNM) staging is a crucial criterion for predicting the prognosis of CRC patients, and immunohistochemistry can be used to determine the colorectal origin of metastasis or visualize the spread of tumor cells in surrounding tissues. However, none of these means can diagnose CRC at an early stage. There is an urgent need to identify several highly sensitive and specific tumor markers and to intervene in the intracellular stage of tumorigenesis, which is critical for early prevention and improvement of the prognosis of CRC.

The reactive oxygen species (ROS) level in biological systems is critical for maintaining intracellular homeostasis (9). Emerging evidence has shown that excessive ROS can cause additional accumulation of DNA mutations leading to the initiation and progression of cancer (10). Several studies have demonstrated that antioxidants can maintain the balance between cell proliferation and apoptosis by reducing intracellular ROS levels to prevent tumorigenesis (11). Peroxiredoxins (PRDXs), a family of such antioxidants, catalyze the reduction of peroxides to maintain the balance of intracellular ROS levels (12). Currently, six PRDX isoforms have been identified in mammalian cells, namely PRDX1-PRDX6. They can be further divided into three subgroups: 2-cysteine (PRDX1-4), atypical 2-cysteine (PRDX5), and 1-cysteine (PRDX6), according to the mechanism of response to cysteine dependence.

There are few studies on the expression and prognostic value of the PRDX family in CRC. Thus, we performed a comprehensive bioinformatics analysis integrating multiple databases to systematically investigate the transcriptional expression profile, gene co-expression, cell line expression, survival prognosis, gene mutation analysis, and protein interaction network of the PRDX family in CRC. Intriguingly, the combined analysis revealed that only PRDX4 was significantly associated with the prognosis and various clinical characteristics of CRC patients. Consequently, we

further validated the tumor-promoting properties of PRDX4 in CRC by *in vitro* and *in vivo* assays.

Materials and methods

Gene expression profiling interactive analysis

GEPIA (<http://gepia.cancer-pku.cn/>) is a newly developed interactive website that integrates TCGA and GTEX gene expression information, including extensive RNA raw sequencing expression data (13). Users can quickly find the expression of the target gene in the corresponding tumor by entering the gene symbol or Ensembl ID. In addition, GEPIA can also perform prognostic survival analysis, including two types of overall survival (OS) and disease-free survival (DFS). GEPIA can also analyze target genes and potential transcription factors by analyzing the correlation between two associated genes.

The human protein atlas

The HPA database (<https://www.proteinatlas.org/>), dedicated to providing tissue and cellular distribution information for all 24,000 human proteins, is freely available for public inquiry (14). Users can query gene expression (RNA and protein expression) in various tissues by entering the target gene. According to the HPA database, we determined the expression of PRDXs in various solid tumors, immunohistochemical images of normal and cancer tissues from CRC patients, and the localization images of PRDXs in cells.

The University of Alabama at Birmingham cancer data analysis portal

UALCAN (<http://ualcan.path.uab.edu/>) is a comprehensive and easily accessible web-based database built on TCGA (15). The database provides gene expression in different cancer tissues and corresponding normal tissues and allows comprehensive analysis of the association between target genes and multiple clinical characteristics. Our study utilized UALCAN to analyze the association between specific PRDXs and CRC clinic characteristics.

Cancer single-cell state atlas

The CancerSEA database covers 41,900 single-cell functional annotation information of 25 types of tumors involving 14 cell functional states such as proliferation, invasion, and apoptosis (16). It enables users to investigate the relationship between the functional phenotypes of genes of interest and various tumors. For example, the CancerSEA was utilized to analyze the association between the PRDX family and CRC cell function.

cBio cancer genomics portal

The cBioPortal database (<https://www.cbioportal.org>) combines 126 tumor genome studies, including the cancer cell line encyclopedia and TCGA (17). cBioPortal integrates genomic data such as somatic mutations, DNA copy-number alterations (CNAs), gene expression (mRNA, microRNA, protein), DNA methylation, and phosphorylated protein enrichment.

Search tool for the retrieval of interaction gene

The String (<https://www.string-db.org>) is a database for searching the interactions between known proteins and predicted proteins. The database includes over 5,000 species, 24 million protein species, and over 20 million protein interaction links (18). In addition, we collected and integrated proteins and potential interactions related to PRDX family members by constructing a protein-protein interaction (PPI) network.

LinkedOmics

The LinkedOmics database (<http://www.linkedomics.org>) includes multi-omics data and clinical data for 32 cancer types and data from 11,158 patients in the TCGA project (19). We can perform correlation analysis between target genes and other genes by inputting PRDX family genes, selecting cancer types and data sets, GO, KEGG, GSEA, and other analyses. Through this database, we can also obtain a series of visualization results, such as volcano plots of related genes and heat maps of positive and negative related genes.

Cancer cell line encyclopedia

The CCLE is a database of cancer cell lines maintained by the Broad Institute of MIT, which currently has more than 1,000 cell lines (20). The PRDX4 expression data in CRC cells were obtained directly from the CCLE website (<https://www.betastasis.com/>). We evaluated the expression of PRDX4 in 57 CRC cell lines by CCLE.

Cell culture and lentivirus transfection

HCT116 cells were purchased from the American type culture collection (ATCC) and cultured in RPMI-1640 medium (Hyclone, UT, USA) containing 10% fetal bovine serum (PAN-Biotech, Adenbach, Bavaria). The short hairpin RNA (shRNA) was synthesized by Sigma-Aldrich (NM_006406). The target sequence was as follows: PRDX4 5'-CCACACTCTTAGAGGTCTCTT-3' (TRCN000064818, Sigma). In addition, the shRNA vector targeting GFP (SHC005, Sigma) was used as a knockdown control. We observed and photographed the cells using immunofluorescence

microscopy, and the number of transfected cells more significant than 95% was considered successful. Moreover, RT-qPCR was applied to detect knockdown efficiency.

RT-qPCR

We used RNA (Promega, Madison, USA) kit to extract the total RNA. The extracted RNA was then reverse-transcribed into cDNA using the PrimeScript™ cDNA synthesis kit (Takara, Japan). Finally, the results were calculated using the $2^{-\Delta\Delta C_t}$ method.

Western blot

An extraction buffer containing protein lysis (Solarbio, Beijing, China) was used to obtain the total protein. Then the extracted protein was quantitatively analyzed using a BCA kit (Solarbio, Beijing, China). First, an equal amount of 30 µg proteins was added to each lane for electrophoresis and membrane transfer. Next, the protein bands were incubated with the primary antibody overnight at 4°C and the secondary antibody for one hour at room temperature. Finally, the western blot results were detected using the ECL chemical reagent method.

Sphere formation assay

A key method for identifying cancer stem cells (CSCs) *in vitro* is their capacity to form spheres. Therefore, we used the sphere formation assay to assess the stemness of HCT116 cells in this work. First, cells were trypsinized to make a single-cell suspension, then seeded into six-well ultra-low cluster plates (4×10^3 cells/well) and cultured in a serum-free medium (DMEM/F-12, Sigma, USA) containing 1% B27 (Gibco, USA), 0.02 µg/ml EGF (Peprotech, USA) and 0.02 µg/ml bFGF (Peprotech, USA) for 15 days. The sphere volume and number were observed under a microscope (Olympus, Japan) every four days, and a new medium was added.

Colony formation, apoptosis, and transwell assay

The specific methods are consistent with our previous work (21). Briefly, for colony formation, PRDX4 knockdown and control cells were seeded into 6-well plates at a density of 300 cells/well and incubated in a cell incubator for two weeks. Then, the cells were fixed with 4% formaldehyde for 30 min and stained with crystal violet for 20 min. Colonies were counted under a microscope and images were collected. For apoptosis assay, the PRDX4 knockdown and control cells were disaggregated with trypsin and washed once with PBS. Then, 1×10^6 cells were resuspended in PBS and stained using allophycocyanin and propidium iodide according to the manufacturer's instructions. For transwell assay, the upper chamber of the 24-well plate was coated with diluted Matrigel and serum-free medium was added, and the lower chamber was

filled with medium containing 20% FBS. The transfected cells (5×10^5 cells/well) were seeded into the serum-free medium and incubated in a cell incubator for 48 h. Then, the cells were fixed and stained with 4% formaldehyde and 0.2% crystal violet, respectively. The invaded cells were photographed and counted under an inverted microscope.

Xenograft assay and immunohistochemistry

The nude mice (BALB/c, four-week-old) were acquired from the Animal Center of North Sichuan Medical College. All experimental procedures were under the ARRIVE guidelines and approved by the Animal Ethics Committee of North Sichuan Medical College (No. SCXK2022-0018). All mice were randomly divided into two groups (sh-Control, $n=5$; sh-PRDX4, $n=5$). A total of 5×10^6 cells were subcutaneously inoculated into each nude mouse. Tumor volume and weight were recorded every four days for four weeks after injection. After treatment, all mice were euthanized to isolate tumors, and tumor weights were measured. The immunohistochemistry (IHC) method is consistent with that described in our previous work (21).

Statistical analysis

The web database generated all statistical analysis results, which were then calculated using the GraphPad Prism 9.0 software. Student's t-test analyzed differences between the control and transfection groups. All experiments were repeated more than three times. $P < 0.05$ represents a statistical difference.

Results

Differential expression of PRDX family members in various tumors

A total of 31 solid tumors were analyzed using the GEPIA database for the RNA expression of each member of the PRDX family (Figure 1). Specifically, PRDX1 was highly expressed in BRCA, CESC, DLBC, GBM, PAAD, TGCT, THYM, and UCEC, and lowly expressed in KICH and LAML (Figure 1A). PRDX2 was highly expressed in DLBC, LUSC, PAAD, TGCT, THYM and UCEC, while it was lowly expressed in LAML (Figure 1B). PRDX3 was highly expressed in DLBC, PAAD, SKCM, STAD, TGCT and THYM (Figure 1C). PRDX4 was highly expressed in BLCA, COAD, DLBC, GBM, KIRC, LGG, LUAD, LUSC, PRAD, READ, SKCM, THYM and UCEC, and was lowly expressed in LAML (Figure 1D). PRDX5 was highly expressed in COAD, DLBC, LIHC, PAAD, READ, THYM, UCEC and UCS, and low in ESCA and LAML (Figure 1E). PRDX6 was highly expressed in DLBC,

GBM, OV and THYM and low in LAML (Figure 1F). Furthermore, we focused on the expression of PRDXs in COAD. Although PRDX1, PRDX2, and PRDX3 were highly expressed in COAD, their expression was not statistically significant (Figures 2A–C). PRDX6 was lowly expressed in COAD, but also not statistically significant (Figure 2F). Among the PRDX family, only PRDX4 and PRDX5 were highly expressed in COAD with statistical significance (Figures 2D, E).

IHC results from the HPA database were used to analyze protein expression to confirm the PRDX family's gene expression trends in CRC. Based on the degree of staining, the protein expression of PRDX2 and PRDX4 was significantly higher in colon cancer tissues than in normal tissues (Figures 3B, D). There is no significant difference in the protein expression of PRDX1 and PRDX3 in normal tissues and cancer tissues (Figures 3A, C). Conversely, the protein expression of PRDX5 and PRDX6 was lower in colon cancer tissues (Figures 3E, F).

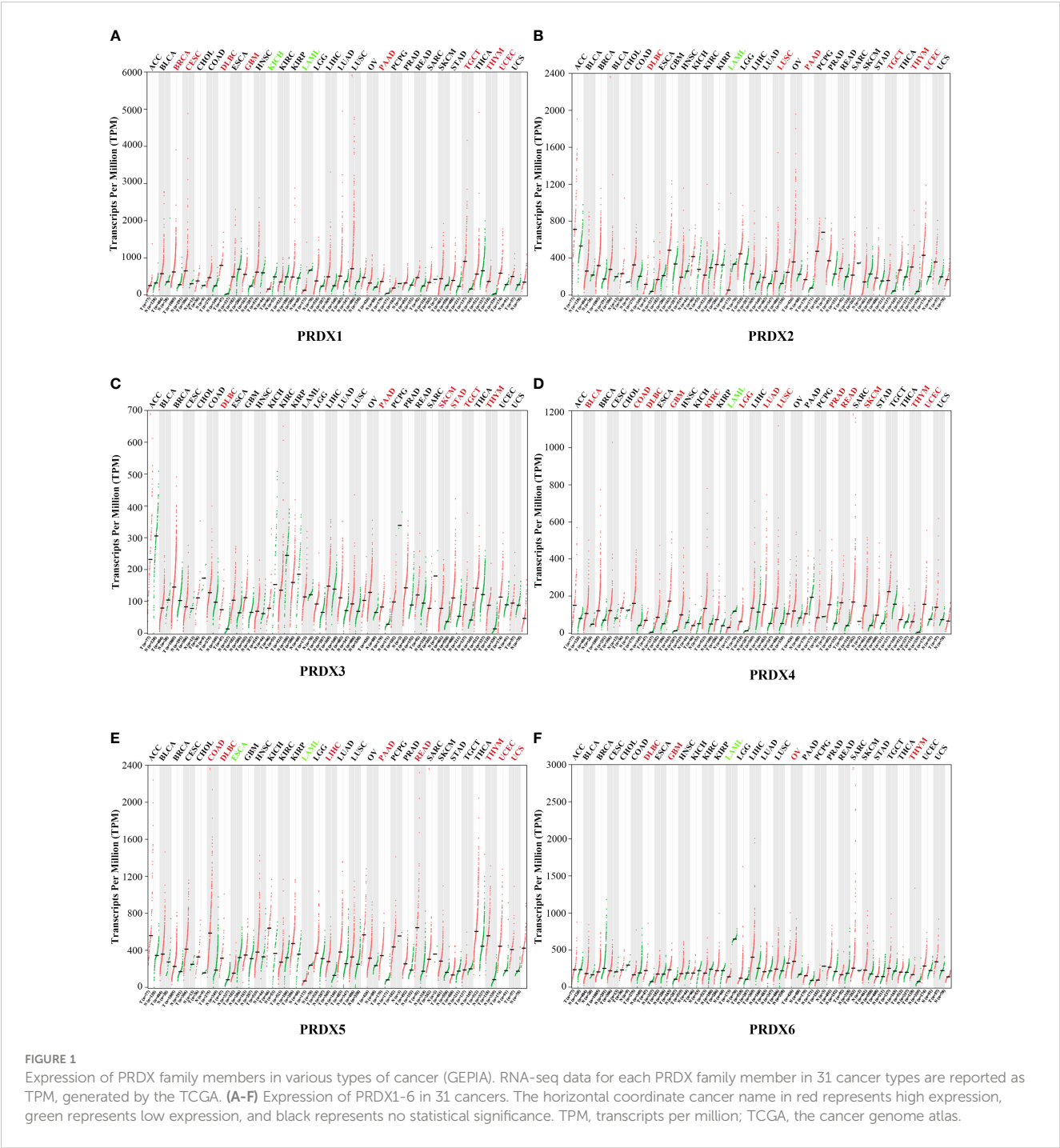
It is well known that tumor node metastasis (TNM) staging remains the crucial criterion for determining the prognosis of CRC patients. Therefore, we further evaluated the correlation between PRDX family members and CRC pathological stage. Interestingly, only the expression of PRDX2, PRDX3, and PRDX4 indicated significant variability ($P < 0.05$) (Figures 4B–D). In contrast, there was no correlation between PRDX1, PRDX5, and PRDX6 and the pathological stage of CRC ($P > 0.05$) (Figures 4A, E, F).

Association between PRDX family expression and prognosis in CRC patients

We investigated the relationship between mRNA expression and OS and DFS in CRC patients using the GEPIA database. As shown in Figure 5, PRDX family members showed no significant difference with DFS. Surprisingly, only PRDX4 expression was statistically significant in terms of OS, and low PRDX4 expression was associated with a poor prognosis (Figure 6D). Therefore, we further analyzed the relationship between PRDX4 and the clinical characteristics of CRC patients in the UALCAN database. The results revealed that the expression of PRDX4 was significantly correlated with individual cancer stages, race, gender, weight, age, histological subtypes, metastasis status, and TP53 mutation status ($P < 0.001$) (Figures 7A–H). Given these findings, we speculated that PRDX4 might be a potential prognostic marker for CRC.

Genetic alterations, protein expression and interaction analysis of PRDXs in CRC patients

PRDXs genetic alterations in CRC were analyzed using the cBioPortal database. The results revealed a minimum of four alteration types (PRDX3/5/6) and a maximum of six alteration types (PRDX1/2) (Figure 8A). Overall, PRDXs were altered in 61 of



the 107 CRC patient samples (Figure 8B). Furthermore, PRDX1, PRDX2, PRDX3, PRDX4, PRDX5, and PRDX6 were altered in 15%, 11%, 10%, 7%, 6%, and 8% of CRC samples, respectively (Figure 8B). Moreover, the mRNA expression heatmap of PRDXs revealed that PRDX1 and PRDX3 had higher mutation rates (Figure 8C).

Furthermore, we performed PPI network analysis for each PRDX using the STRING to explore potential interactions between proteins. The results showed that there were 61 protein

nodes related to the SFRP family in the constructed PPI network (Figure 8D).

Functional states of the PRDX family in CRC cells

RNA-seq technology's emergence allows for exploring tumor cells' functional heterogeneity. The association of each PRDX family

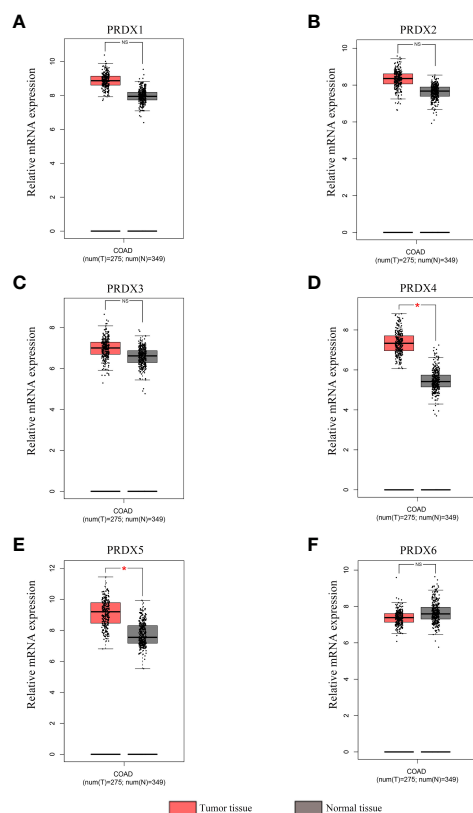


FIGURE 2

Expression analysis of PRDX family members in COAD and normal tissues (GEPiA). (A–F) The mRNA expression levels of each PRDX family member in COAD and corresponding normal tissues were compared in the box plot. * $P < 0.05$. NS, no significance; COAD, colon adenocarcinoma.

member with 14 functional states in CRC cells is shown in Figure 9. As a potential clinical indicator of CRC progression, PRDX4's role in cancer biology attracted our attention. According to cell function analysis, PRDX4 positively correlated with proliferation, apoptosis, and invasion of CRC cells but negatively correlated with stemness, which was closely associated with prognosis (Figure 9). Based on these results, it deserves to verify whether PRDX4 can have a similar effect on CRC cells by further *in vitro* analysis.

KEGG pathway analysis and cellular localization of PRDX4-related differentially expressed genes in CRC

We further analyzed RNA-seq sequencing data of PRDX4 from 379 CRC patients using the LinkedOmics database. The results of the volcano plot displayed that 11628 genes (red dots) were positively correlated with PRDX4, while 8200 genes were negatively correlated (green dots) (Figure 10A). Moreover, Figures 10C, D show that 50 significant genes were positively and negatively correlated with PRDX4. Among them, the strongest positive correlation was found between PRDX4 and UXT (Pearson correlation = 0.76, $P = 1.34e-71$), whereas the strongest negative correlation was found with SPTAN1 (Pearson correlation = 0.57, $P = 8.02e-34$). Furthermore, KEGG pathway analysis showed that PRDX4 positively correlated with several major cellular functional processes, including DNA replication (Figure 10B). Moreover, through cellular localization analysis, we found that PRDX4 is mainly localized to the endoplasmic reticulum (Figure 10E). In addition, it is partially localized to the cytosol (Figure 10F).

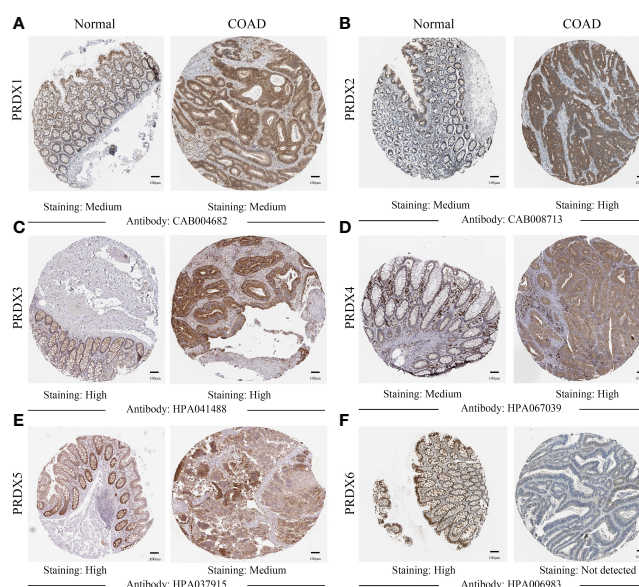


FIGURE 3

(A–F) IHC analysis of protein expression of each PRDX family member in COAD tissues and normal tissues (HPA). Brown areas in the image indicate positive expression, and blue areas indicate negative expression. IHC, immunohistochemical; COAD, colon adenocarcinoma. scale bar, 100 μ m.

Differential expression of PRDX4 in CRC cell lines

The CCLE database was applied to analyze the expression of PRDX4 in various CRC cell lines. The results suggested that among 57 CRC cell lines, the expression of PRDX4 in COLO320, HCT116, HCT15, KM12, LS411N, NCIH716, and RKO was significantly higher than in other cell lines (Figure 11A). Therefore, we selected one of the most common HCT116 cells for further *in vitro* and *in vivo* analysis. Furthermore, single-cell sequencing data (GSE81861) from the cancerSEA database revealed 14 functional cell states, such as apoptosis, proliferation, invasion, and stemness, in HCT116 cells without prior treatment (Figure 11B).

Knockdown of PRDX4 inhibits CRC cell proliferation and invasion, but promotes apoptosis and stemness

Further analyses of PRDX4 cellular functions were performed both *in vitro* and *in vivo* to validate the cancer SEA database findings. Firstly, HCT116 cells with stable knockdown expression PRDX4 were established by lentiviral transfection (Figure 12A). RT-qPCR and western blot assays were then performed to validate the knockdown efficiency of the lentivirus. Both mRNA and protein levels were significantly reduced by targeting shRNA to PRDX4 (Figures 12A, B).

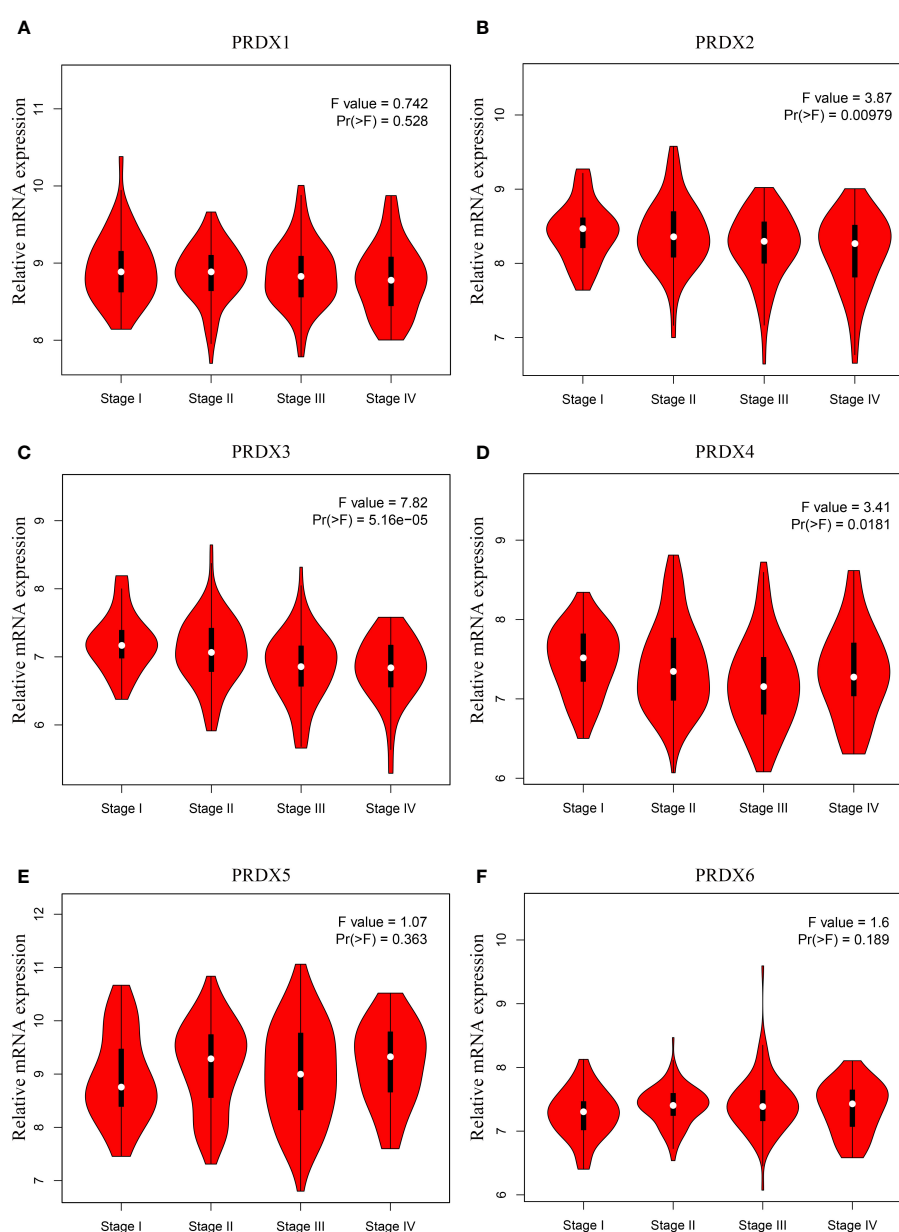


FIGURE 4

(A-F) Association between mRNA expression of PRDXs and tumor stage in COAD patients (GEPIC). The expression of PRDX2, PRDX3 and PRDX4 was significantly correlated with the pathological stage of COAD patients ($P < 0.05$). COAD, colon adenocarcinoma.

Numerous studies have found a strong link between the stemness of cancer cells and the prognosis of cancer patients (22). Consequently, a sphere formation assay was utilized to determine the effect of PRDX4 expression level on the stemness of cancer cells. The results showed that knocking down PRDX4 dramatically promoted the sphere-forming capability of cancer cells (Figure 12C). As shown in Figures 12D–F, knocking out PRDX4 inhibited HCT116 cell proliferation and invasion while promoting apoptosis. Notably, the experimental results of PRDX4 apoptosis were not consistent with the bioinformatics analysis in Figure 9.

Furthermore, the effect of PRDX4 on cancer cell proliferation *in vivo* was assessed by injecting sh-Ctrl or sh-PRDX4 cells into nude mice. The results suggested that the tumor volumes were significantly reduced in the sh-PRDX4 group compared to the control group (Figure 12G). Proliferating cell nuclear antigen (PCNA) was used to indicate tumor proliferation. IHC staining revealed that the expression of PCNA in the sh-PRDX4 group was significantly lower than in the control group (Figure 12H). Taken together, these results demonstrate the involvement of PRDX4 in regulating CRC cell progression.

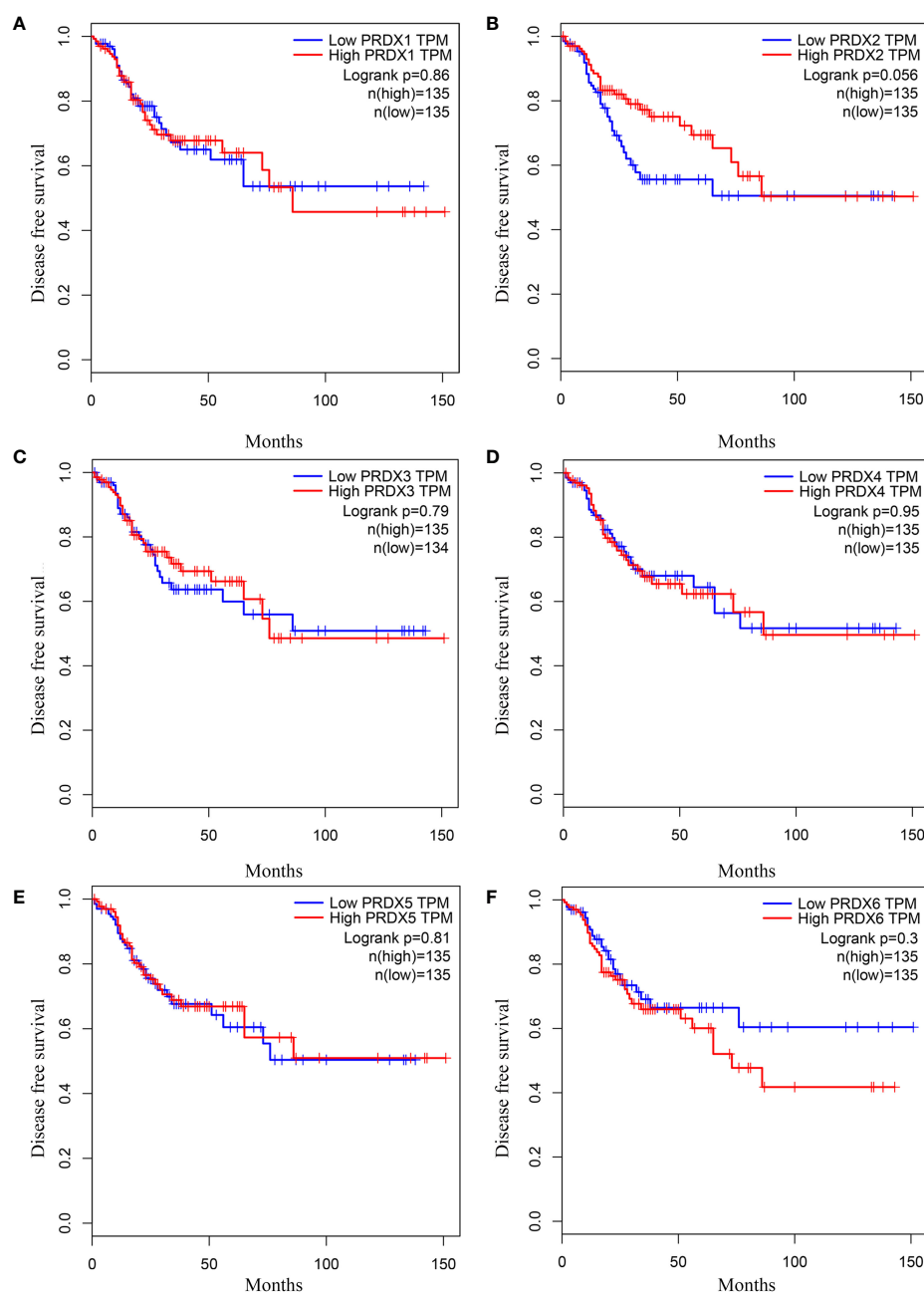


FIGURE 5

The DFS curves of PRDXs in COAD (GEPIA). (A–F) None of the PRDX family members were associated with DFS in COAD ($P > 0.05$). DFS, disease free survival; COAD, colon adenocarcinoma.

Discussion

Recent studies have noted that ROS, a crucial regulator of the intracellular environment, is strongly associated with tumorigenesis (23). In addition, ROS is involved in the growth process of tumor cells and can promote metastasis and thus affect patient prognosis (24). This has led to the suggestion that ROS modulators may be helpful in the primary prevention of cancer.

Increasing research has demonstrated that aberrantly expressed PRDXs are involved in a series of tumor cell processes, including cell proliferation, apoptosis, and invasion (25, 26). For instance, several studies have indicated that PRDX1 can function as a pro-

oncogene in lung cancer (27) and soft tissue sarcoma (28). *In vivo*, PRDX1 knockout mice developed several malignancies, including lymphomas and sarcomas (29). In addition, PRDX1, PRDX2, and PRDX6 have also been reported to be overexpressed in oral squamous cell carcinoma (30–32). Moreover, the tumor-promoting role of PRDX4 has been demonstrated in lung cancer, leukemia, and glioblastoma (33–35). As for PRDX5, it shows a dual function in tumors. Elamin et al. found that PRDX5 was the only member of the PRDX family markedly downregulated in breast cancer (36). In contrast, Gérard et al. reported the tumor-promoting role of highly expressed PRDX5 in thyroid cancer (37). Surprisingly little is known about the PRDX family in CRC,

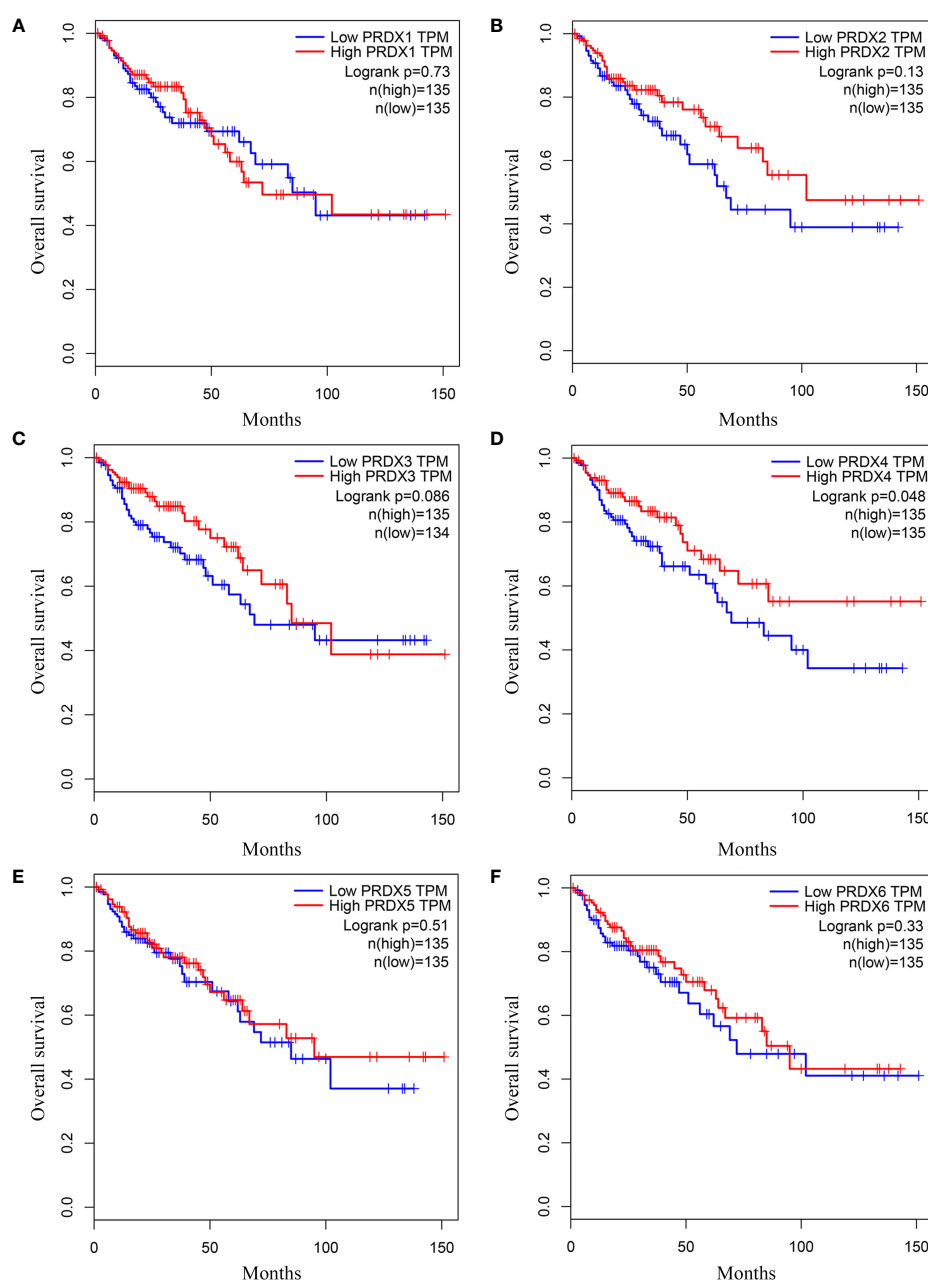


FIGURE 6

(A–F) The OS curves of PRDXs in COAD (GEPIC). (D) Only COAD patients with low expression of PRDX4 were significantly associated with shorter OS ($P < 0.05$). OS, overall survival; COAD, colon adenocarcinoma.

particularly PRDX4. Therefore, we conducted this study to address this issue.

PRDX was first introduced in 1994, and subsequent studies found that PRDX4 is located on human chromosome 10p22.13 (38, 39). Numerous studies have demonstrated the potential of PRDX4 as a biomarker for many diseases, such as diabetes and stroke (40). However, the role and prognostic value of PRDX4 in CRC remain

unknown, which drew our attention. To our knowledge, this is the first study to combine multicenter databases to analyze the expression and prognostic value of the entire PRDX family in CRC. The data in this study included cell lines and clinical samples and comprehensively revealed the expression and prognostic value of the PRDX family in CRC. First, we found that PRDXs are differentially expressed in colon adenocarcinoma (COAD) at the

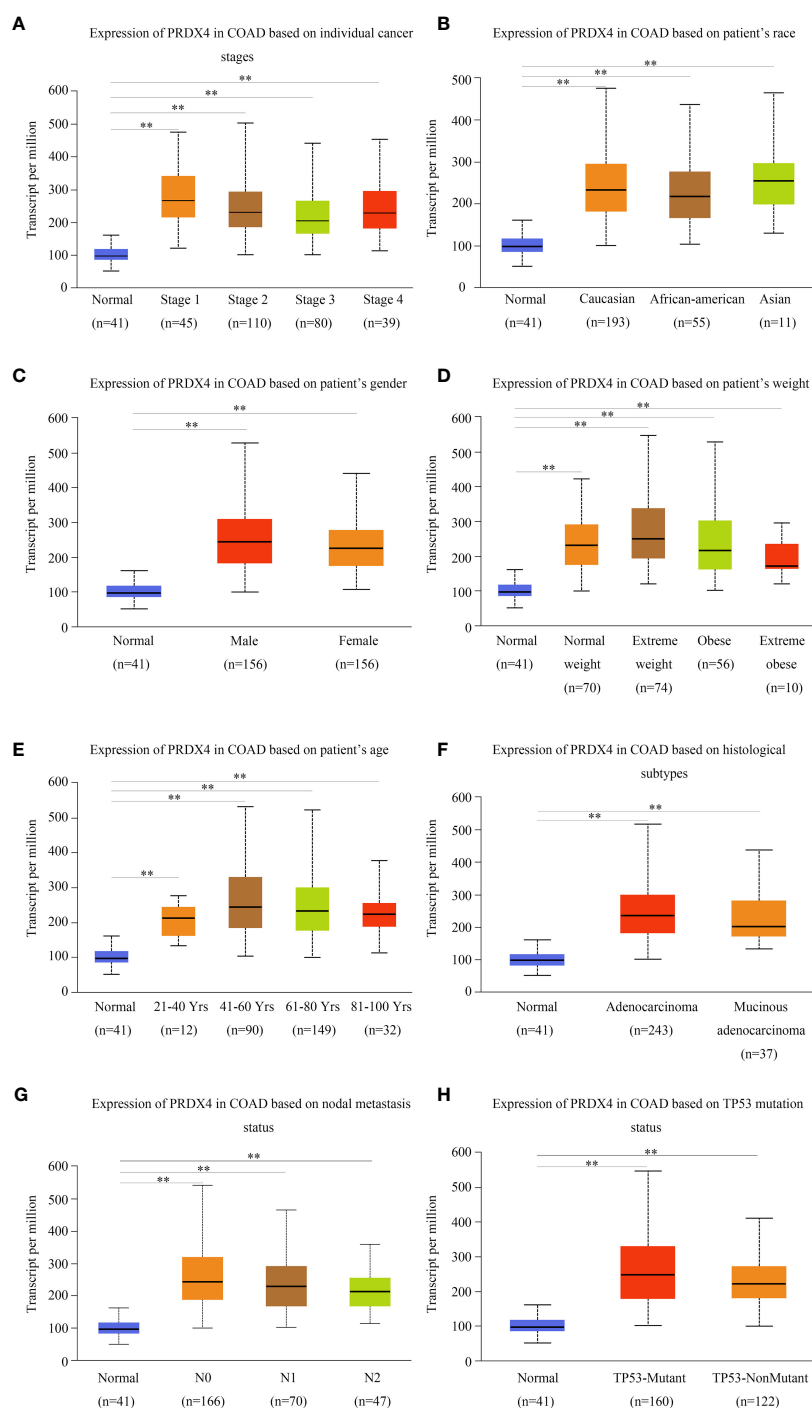


FIGURE 7

PRDX4 transcript levels in subgroups of COAD patients, stratified by individual cancer stages, race, gender, and other criteria (UALCAN). Compared with normal tissues, PRDX4 was highly expressed in all clinical subtypes shown in (A-H) and had a significant correlation with these clinical characteristics. $**P < 0.01$. COAD, colon adenocarcinoma.

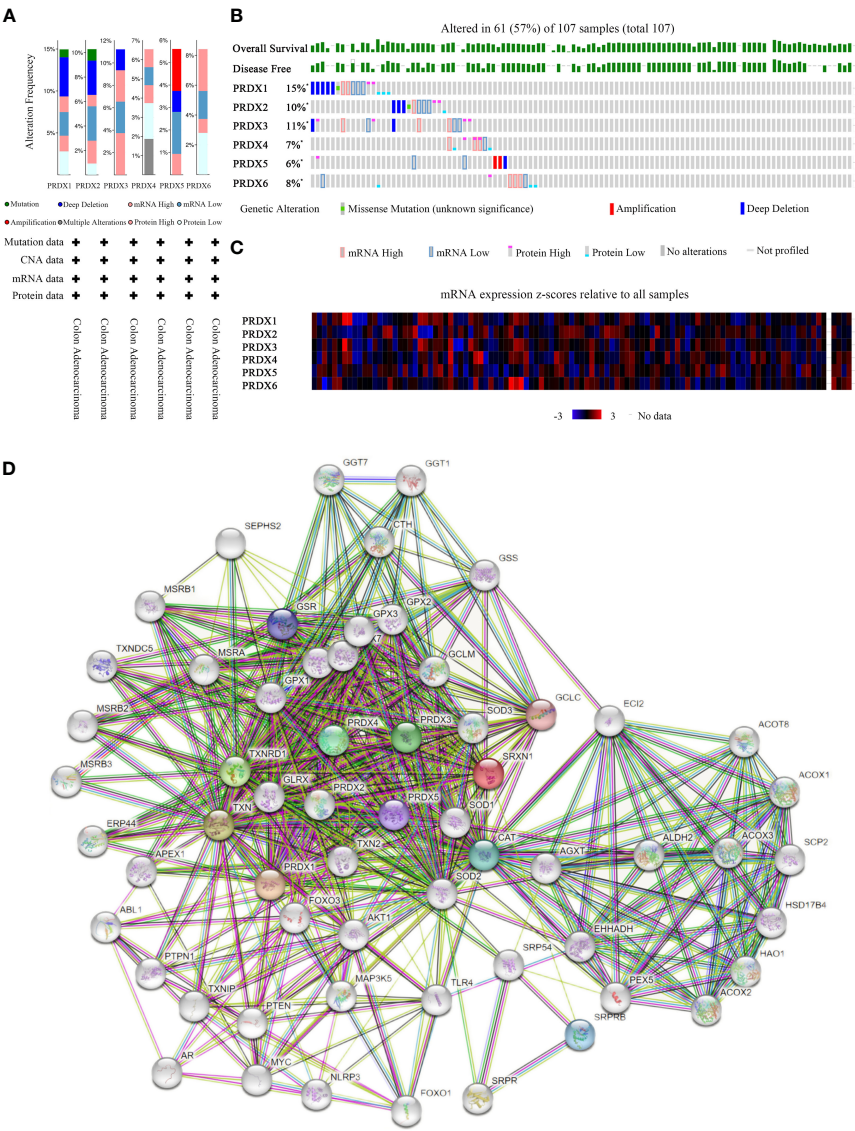


FIGURE 8 Mutation and expression analysis of PRDX gene in COAD (cBioPortal and STRING). **(A, B)** Summary of alterations in different expressed PRDXs in COAD. PRDXs were altered in 61 of the 107 COAD patient samples, accounting for 57%. **(C)** Heatmap showing mRNA expression z-scores relative to all samples (RNA Seq V2 RSEM UQ Log2). **(D)** Protein-protein interaction network of the PRDX family. COAD, colon adenocarcinoma.

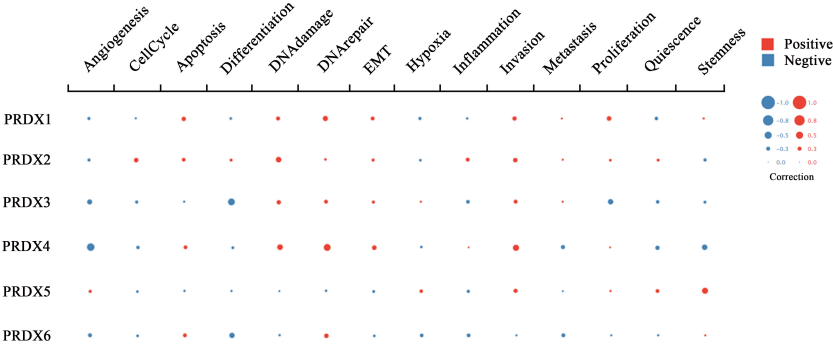


FIGURE 9 Functional state analysis of PRDX family members in CRC (cancerSEA). Functional relevance of each PRDX member in CRC cells. The size of the bubble indicates the strength of the correlation; red represents a positive correlation and blue represents a negative correlation. CRC, colorectal cancer.



Differentially expressed genes associated with PRDX4 in CRC, KEGG pathway analysis and cellular localization (LinkedOmics). **(A)** Pearson test was used to analyze the correlation between PRDX4 and differentially expressed genes in CRC. **(B)** KEGG enrichment analysis showing functional pathways associated with PRDX4. **(C, D)** Heatmap analysis showing genes positively and negatively correlated with PRDX4 in CRC (Top 50). Red represents positively correlated genes, and green represents negatively correlated genes. **(E, F)** Subcellular localization of PRDX4. Green represents the target protein, blue represents the nucleus, and red represents the cellular microtubules. KEGG, kyoto encyclopedia of genes and genomes.

associated with multiple clinical characteristics, including age, sex, pathological subtype, lymph node metastasis, etc., suggesting that PRDX4 may be a potential prognostic marker for COAD.

Recently, several studies have identified that mutations cause approximately 40% of CRC cases in tumor suppressors or oncogenes (41). Further genetic analyses revealed that differentially expressed PRDXs undergo frequent genetic alterations in COAD. These data

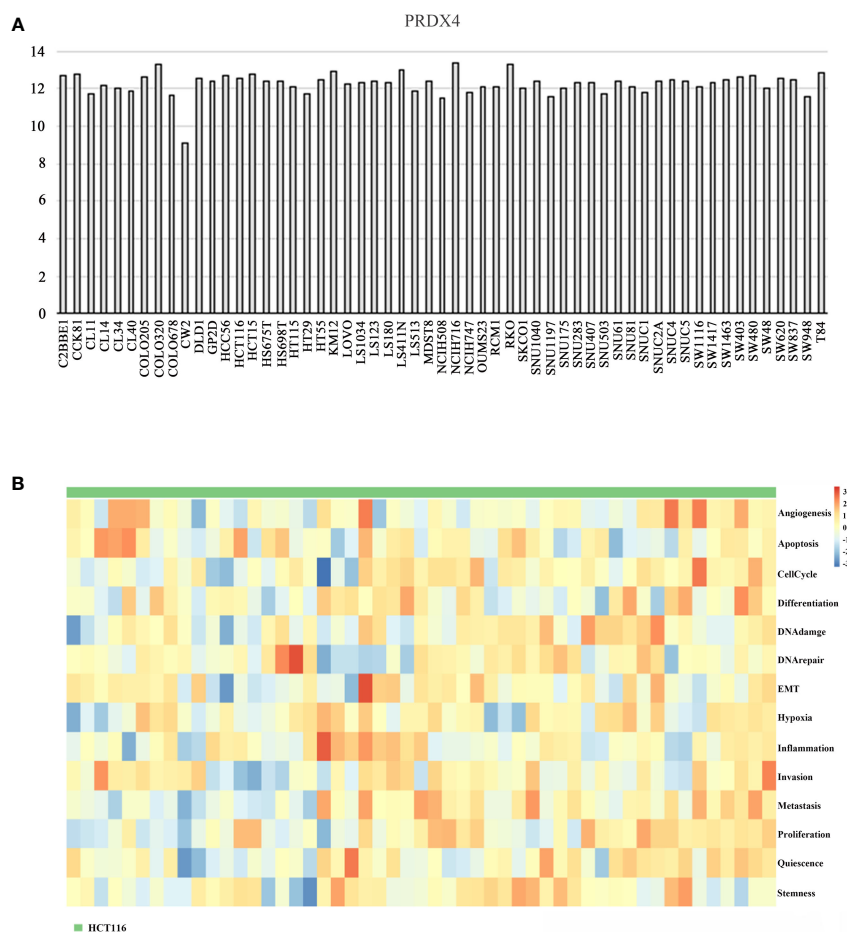


FIGURE 11

Differential expression of PRDX4 in different CRC cell lines (CCLE and cancerSEA). **(A)** The mRNA expression level of PRDX4 in various CRC cell lines, as determined by CCLE analysis. **(B)** The heatmap in this panel displays the activity of function states of cells in the selected dataset. Rows represent function states, columns represent hierarchically clustered cells. Column labels indicate the cell groups which cells belong to. CCLE, cancer cell line encyclopedia.

could partially explain the causes of colorectal carcinogenesis. By building a protein interaction network, we found that PRDXs are mostly associated with peroxides and cell proliferation proteins. We further focused on the functional states of PRDXs in CRC. The results suggested that PRDX4 was positively associated with apoptosis, proliferation, and invasion and negatively associated with stemness. We discovered that PRDX4 is primarily involved in the biological process of cell proliferation using KEGG enrichment analysis, which is consistent with previous research. Moreover, our *in vitro* and *in vivo* experiments further validated PRDX4's role in COAD. In HCT116 cells, knocking down the expression of PRDX4 inhibited proliferation and invasion and promoted apoptosis. These findings are consistent with the studies reported by Yi et al. (42). Extensive studies have demonstrated the involvement of CSCs in developing tumors, and their formation is one of the leading causes of cancer recurrence (43). Intriguingly, the downregulation of PRDX4 expression promotes stemness in HCT116 cells. This finding may explain the poor prognosis caused by the low expression of PRDX4.

Overall, our study systematically reveals the primary role of PRDXs, especially PRDX4, in the carcinogenesis of COAD and provides a promising strategy for treating COAD.

Our study has some undeniable limitations. All analytical data in our study were obtained from online databases, and additional clinical samples must be collected for subsequent validation. Furthermore, our *in vitro* and *in vivo* assays used only one type of COAD cell, and our findings need to be validated in more CRC cells.

Conclusion

Our current research mainly has the following innovations: (1) We used multiple databases for the first time to systematically examine the transcriptome, mutation, mRNA, and protein of the PRDX family in CRC. (2) The differential expression of PRDX4 was finally determined by screening to be closely related to the

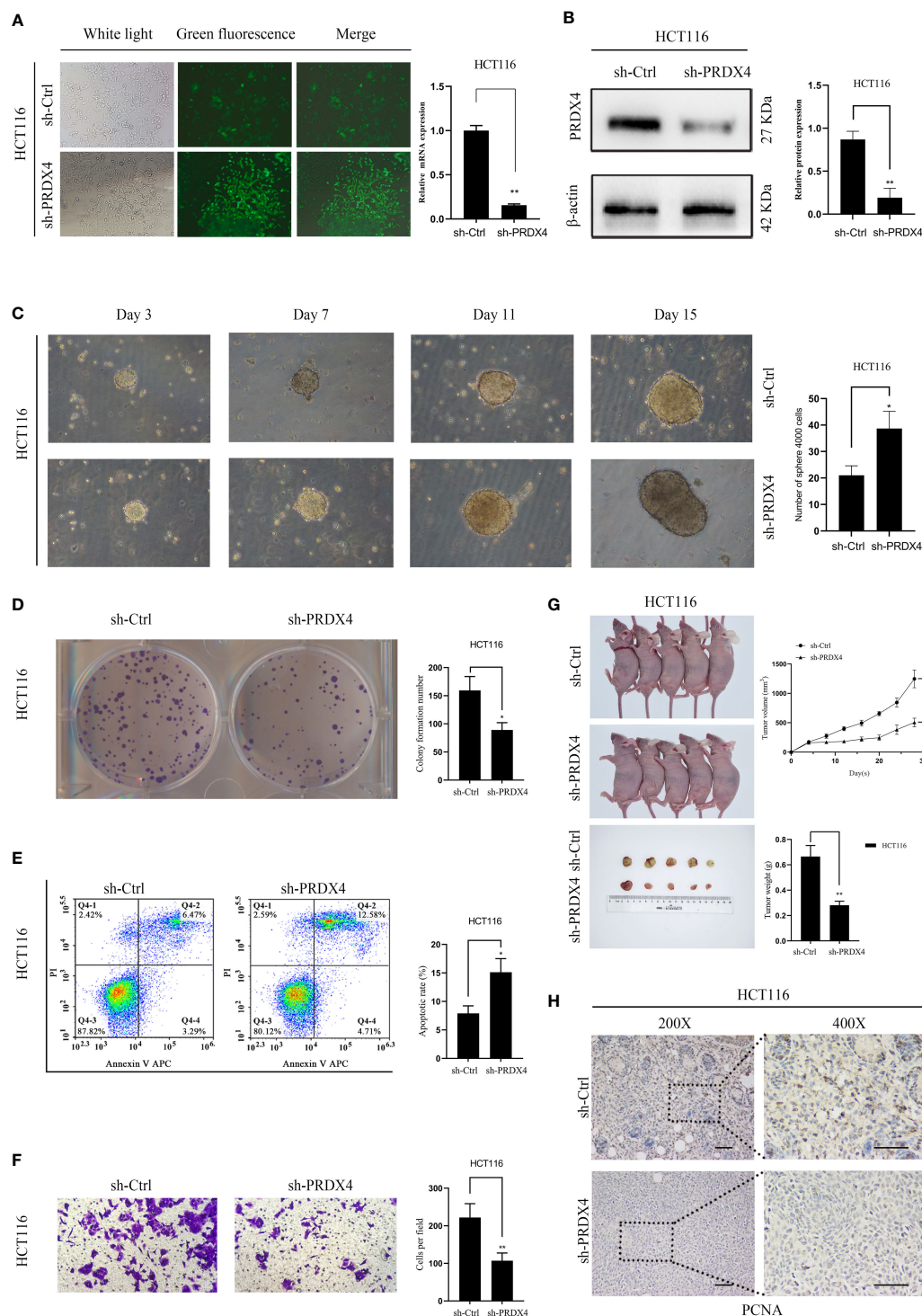


FIGURE 12

Knockdown of PRDX4 inhibits CRC cell proliferation and invasion and promotes CRC cell stemness and apoptosis. (A) The transfection efficiency was observed by fluorescence microscopy and detected by RT-qPCR. (B) Western blot was used to validate the expression of PRDX4 protein in HCT116 cells transfected with shRNA. (C) Sphere formation assay of HCT116 cells showing that knockdown of PRDX4 promoted cell stemness. (D) Colony formation assay of HCT116 cells showing that knockdown of PRDX4 inhibited cell proliferation. (E) Flow cytometry results showing that knockdown of PRDX4 promoted apoptosis in HCT116 cells. (F) Transwell assay of HCT116 cells showing that knockdown of PRDX4 inhibited cell invasion. (G) Knockdown of PRDX4 inhibited tumor growth of HCT116 cells *in vivo*. (H) Images of IHC staining of PCNA in xenograft tumors. scale bar, 50 μ m. * P < 0.05, ** P < 0.01. IHC, immunohistochemical; PCNA, proliferating cell nuclear antigen.

prognosis of COAD patients. (3) Our *in vitro* and *in vivo* analyses revealed that aberrant expression of PRDX4 has significant effects on the process of COAD, including proliferation, apoptosis, invasion, and stemness. Based on our comprehensive analysis and validation, PRDX4 may represent a new prognostic factor for COAD.

Data availability statement

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

Ethics statement

The animal study was approved by the Animal Ethics Committee of North Sichuan Medical College (license number: SCXK2022-0018).

Author contributions

HZ and YX designed the research, LL performed the experiments, SH, JC, and TZ analyzed the data, HZ wrote the article, and all authors revised and approved the manuscript. All authors contributed to the article and approved the submitted version.

References

1. 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 Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
2. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut* (2017) 66(4):683–91. doi: 10.1136/gutjnl-2015-310912
3. Center MM, Jemal A, Smith RA, Ward E. Worldwide variations in colorectal cancer. *CA Cancer J Clin* (2009) 59(6):366–78. doi: 10.3322/caac.20038
4. Cao W, Chen HD, Yu YW, Li N, Chen WQ. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. *Chin Med J* (2021) 134(7):783–91. doi: 10.1097/cm9.0000000000001474
5. Zhu J, Tan Z, Hollis-Hansen K, Zhang Y, Yu C, Li Y. Epidemiological trends in colorectal cancer in China: An ecological study. *Dig Dis Sci* (2017) 62(1):235–43. doi: 10.1007/s10620-016-4362-4
6. Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, et al. Cancer treatment and survivorship statistics, 2016. *CA Cancer J Clin* (2016) 66(4):271–89. doi: 10.3322/caac.21349
7. Wang H, Tian T, Zhang J. Tumor-associated macrophages (TAMs) in colorectal cancer (CRC): From mechanism to therapy and prognosis. *Int J Mol Sci* (2021) 22(16):8470. doi: 10.3390/ijms22168470
8. Chen K, Collins G, Wang H, Toh JWT. Pathological features and prognostication in colorectal cancer. *Curr Oncol* (2021) 28(6):5356–83. doi: 10.3390/curroncol28060447
9. Garza-Lombó C, Pappa A, Panayiotidis MI, Franco R. Redox homeostasis, oxidative stress and mitophagy. *Mitochondrion* (2020) 51:105–17. doi: 10.1016/j.mito.2020.01.002
10. Kumari S, Badana AK, Murali Mohan G, Shailender G, Malla R. Reactive oxygen species: A key constituent in cancer survival. *biomark Insights* (2018) 13:1177271918755391. doi: 10.1177/1177271918755391
11. Skoko JJ, Attaran S, Neumann CA. Signals getting crossed in the entanglement of redox and phosphorylation pathways: Phosphorylation of peroxiredoxin proteins sparks cell signaling. *Antioxidants* (2019) 8(2):29. doi: 10.3390/antiox8020029
12. Hopkins BL, Nadler M, Skoko JJ, Bertomeu T, Pelosi A, Shafaei PM, et al. A peroxidase peroxiredoxin 1-specific redox regulation of the novel FOXO3 microRNA target let-7. *Antioxid Redox Signal* (2018) 28(1):62–77. doi: 10.1089/ars.2016.6871
13. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* (2017) 45(W1):98–102. doi: 10.1093/nar/gkx247
14. Pontén F, Jirstrom K, Uhlen M. The human protein atlas—a tool for pathology. *J Pathol* (2008) 216(4):387–93. doi: 10.1002/path.2440
15. Chandrashekar DS, Karthikeyan SK, Korla PK, Patel H, Shovon AR, Athar M, et al. UALCAN: An update to the integrated cancer data analysis platform. *Neoplasia* (2022) 25:18–27. doi: 10.1016/j.neo.2022.01.001
16. Yuan H, Yan M, Zhang G, Liu W, Deng C, Liao G, et al. CancerSEA: a cancer single-cell state atlas. *Nucleic Acids Res* (2019) 47(D1):900–8. doi: 10.1093/nar/gky939
17. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* (2013) 6(269):pl1. doi: 10.1126/scisignal.2004088
18. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* (2019) 47(D1):607–13. doi: 10.1093/nar/gky1131

Funding

The study was supported by Sichuan Provincial Medical Research Youth Innovation Project (Q22040), Free Exploring Basic Research Project of Science and Technology of Sichuan Province (2021YJ0456), Scientific Research Project of Affiliated Hospital of North Sichuan Medical College (2021ZD001), Doctoral Research Startup Fund Project of North Sichuan Medical College (CBY20-QD03) and China Scholarship Council (202108515042). Youth National Natural Science foundation of China (82003147), Nanchong City School Cooperation Project (20YFZ0091).

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.

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.

19. Vasaikar SV, Straub P, Wang J, Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res* (2018) 46(D1):956–63. doi: 10.1093/nar/gkx1090
20. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* (2012) 483(7391):603–7. doi: 10.1038/nature11003
21. Zhou H, Xiong Y, Peng L, Wang R, Zhang H, Fu Z. LncRNA-cCSC1 modulates cancer stem cell properties in colorectal cancer via activation of the hedgehog signaling pathway. *J Cell Biochem* (2020) 121(3):2510–24. doi: 10.1002/jcb.29473
22. Walcher L, Kistenmacher AK, Suo H, Kitte R, Dluczek S, Strauß A, et al. Cancer stem cells-origins and biomarkers: Perspectives for targeted personalized therapies. *Front Immunol* (2020) 11:1280. doi: 10.3389/fimmu.2020.01280
23. de Sá Junior PL, Câmara DAD, Porcacchia AS, Fonseca PMM, Jorge SD, Araldi RP, et al. The roles of ROS in cancer heterogeneity and therapy. *Oxid Med Cell Longev* (2017) 2017:2467940. doi: 10.1155/2017/2467940
24. Reczek CR, Chandel NS. ROS promotes cancer cell survival through calcium signaling. *Cancer Cell* (2018) 33(6):949–51. doi: 10.1016/j.ccell.2018.05.010
25. Immenschuh S, Baumgart-Vogt E. Peroxiredoxins, oxidative stress, and cell proliferation. *Antioxid Redox Signal* (2005) 7(5-6):768–77. doi: 10.1089/ars.2005.7.768
26. Hofmann B, Hecht HJ, Flohé L. Peroxiredoxins. *Biol Chem* (2002) 383(3-4):347–64. doi: 10.1515/bc.2002.040
27. Chen MF, Keng PC, Shau H, Wu CT, Hu YC, Liao SK, et al. Inhibition of lung tumor growth and augmentation of radiosensitivity by decreasing peroxiredoxin I expression. *Int J Radiat Oncol Biol Phys* (2006) 64(2):581–91. doi: 10.1016/j.ijrobp.2005.10.012
28. Takahashi A, Nakayama R, Ishibashi N, Doi A, Ichinohe R, Ikuyo Y, et al. Analysis of gene expression profiles of soft tissue sarcoma using a combination of knowledge-based filtering with integration of multiple statistics. *PloS One* (2014) 9(9):e106801. doi: 10.1371/journal.pone.0106801
29. Neumann CA, Krause DS, Carman CV, Das S, Dubey DP, Abraham JL, et al. Essential role for the peroxiredoxin Prdx1 in erythrocyte antioxidant defence and tumour suppression. *Nature* (2003) 424(6948):561–5. doi: 10.1038/nature01819
30. Lee EY, Kang JY, Kim KW. Expression of cyclooxygenase-2, peroxiredoxin I, peroxiredoxin 6 and nuclear factor- κ B in oral squamous cell carcinoma. *Oncol Lett* (2015) 10(5):3129–36. doi: 10.3892/ol.2015.3705
31. Chuerduangphui J, Ekalaksananan T, Heawchaiyaphum C, Vatanasapt P, Pientong C. Peroxiredoxin 2 is highly expressed in human oral squamous cell carcinoma cells and is upregulated by human papillomavirus oncoproteins and arecoline, promoting proliferation. *PloS One* (2020) 15(12):e0242465. doi: 10.1371/journal.pone.0242465
32. Huang CF, Sun ZJ, Zhao YF, Chen XM, Jia J, Zhang WF. Increased expression of peroxiredoxin 6 and cyclophilin a in squamous cell carcinoma of the tongue. *Oral Dis* (2011) 17(3):328–34. doi: 10.1111/j.1601-0825.2010.01730.x
33. Wei Q, Jiang H, Xiao Z, Baker A, Young MR, Veenstra TD, et al. Sulfiredoxin-peroxiredoxin IV axis promotes human lung cancer progression through modulation of specific phosphokinase signaling. *Proc Natl Acad Sci* (2011) 108(17):7004–9. doi: 10.1073/pnas.1013012108
34. Palande KK, Beekman R, van der Meeren LE, Beverloo HB, Valk PJ, Touw IP. The antioxidant protein peroxiredoxin 4 is epigenetically down regulated in acute promyelocytic leukemia. *PloS One* (2011) 6(1):e16340. doi: 10.1371/journal.pone.0016340
35. Kim TH, Song J, Alcantara Llaguno SR, Murnan E, Liyanarachchi S, Palanichamy K, et al. Suppression of peroxiredoxin 4 in glioblastoma cells increases apoptosis and reduces tumor growth. *PloS One* (2012) 7(8):e42818. doi: 10.1371/journal.pone.0042818
36. Elamin A, Zhu H, Hassan AM, Xu N, Ibrahim ME. Peroxiredoxin V: A candidate breast tumor marker of population specificity. *Mol Clin Oncol* (2013) 1(3):541–9. doi: 10.3892/mco.2013.91
37. Gérard AC, Many MC, Daumerie C, Knoops B, Colin IM. Peroxiredoxin 5 expression in the human thyroid gland. *Thyroid* (2005) 15(3):205–9. doi: 10.1089/thy.2005.15.205
38. Chae HZ, Chung SJ, Rhee SG. Thioredoxin-dependent peroxide reductase from yeast. *J Biol Chem* (1994) 269(44):27670–8. doi: 10.1016/S0021-9258(18)47038-X
39. Wood ZA, Schröder E, Robin Harris J, Poole LB. Structure, mechanism and regulation of peroxiredoxins. *Trends Biochem Sci* (2003) 28(1):32–40. doi: 10.1016/s0968-0004(02)00003-8
40. Tavender TJ, Sheppard AM, Bulleid NJ. Peroxiredoxin IV is an endoplasmic reticulum-localized enzyme forming oligomeric complexes in human cells. *Biochem J* (2008) 411(1):191–9. doi: 10.1042/bj20071428
41. Dobre M, Salvi A, Pelisenco IA, Vasilescu F, De Petro G, Herlea V, et al. Crosstalk between DNA methylation and gene mutations in colorectal cancer. *Front Oncol* (2021) 11:697409. doi: 10.3389/fonc.2021.697409
42. Yi N, Xiao MB, Ni WK, Jiang F, Lu CH, Ni RZ. High expression of peroxiredoxin 4 affects the survival time of colorectal cancer patients, but is not an independent unfavorable prognostic factor. *Mol Clin Oncol* (2014) 2(5):767–72. doi: 10.3892/mco.2014.317
43. Ruhland MK, Coussens LM, Stewart SA. Senescence and cancer: An evolving inflammatory paradox. *Biochim Biophys Acta* (2016) 1865(1):14–22. doi: 10.1016/j.bbcan.2015.10.001



OPEN ACCESS

EDITED BY

Zora Lasabova,
Comenius University, Slovakia

REVIEWED BY

Peng-Chan Lin,
National Cheng Kung University, Taiwan
Raffaele Capasso,
University of Naples Federico II, Italy

*CORRESPONDENCE

Veronika Vymetalkova

✉ veronika.vymetalkova@iem.cas.cz

Stanislav Filip

✉ stanislav.filip@fnhk.cz

RECEIVED 29 December 2022

ACCEPTED 20 March 2023

PUBLISHED 27 April 2023

CITATION

Horak J, Kubecek O, Siskova A, Honkova K, Chvojekova I, Krupova M, Manethova M, Vodenkova S, García-Mulero S, John S, Cecka F, Vodickova L, Petera J, Filip S and Vymetalkova V (2023) Differences in genome, transcriptome, miRNAome, and methylome in synchronous and metachronous liver metastasis of colorectal cancer.

Front. Oncol. 13:1133598.

doi: 10.3389/fonc.2023.1133598

COPYRIGHT

© 2023 Horak, Kubecek, Siskova, Honkova, Chvojekova, Krupova, Manethova, Vodenkova, García-Mulero, John, Cecka, Vodickova, Petera, Filip and Vymetalkova. 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.

Differences in genome, transcriptome, miRNAome, and methylome in synchronous and metachronous liver metastasis of colorectal cancer

Josef Horak^{1,2}, Ondrej Kubecek³, Anna Siskova^{1,4}, Katerina Honkova⁵, Irena Chvojekova⁵, Marketa Krupova⁶, Monika Manethova⁶, Sona Vodenkova^{1,7}, Sandra García-Mulero^{8,9}, Stanislav John³, Filip Cecka¹⁰, Ludmila Vodickova^{1,4,7}, Jiri Petera³, Stanislav Filip^{3*} and Veronika Vymetalkova^{1,4,7*}

¹Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czechia, ²Department of Medical Genetics, Third Faculty of Medicine, Charles University, Prague, Czechia, ³Department of Oncology and Radiotherapy, Faculty of Medicine and University Hospital in Hradec Kralove, Charles University, Hradec Kralove, Czechia, ⁴Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, Prague, Czechia, ⁵Department of Genetic Toxicology and Epigenetics, Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czechia, ⁶The Fingerland Department of Pathology, University Hospital in Hradec Kralove, Hradec Kralove, Czechia, ⁷Biomedical Centre, Faculty of Medicine in Pilsen, Charles University, Pilsen, Czechia, ⁸Unit of Biomarkers and Susceptibility, Oncology Data Analytics Program (ODAP), Catalan Institute of Oncology (ICO)-Oncobell Programme, Bellvitge Biomedical Research Institute Oncobell Programme, Bellvitge Biomedical Research Institute (IDIBELL), Hospitalet de Llobregat, Barcelona, Spain, ⁹Program in Molecular Mechanisms and Experimental Therapy in Oncology (Oncobell), Oncobell Programme, Bellvitge Biomedical Research Institute (IDIBELL), Hospitalet de Llobregat, Barcelona, Spain, ¹⁰Department of Surgery, Faculty of Medicine and University Hospital in Hradec Kralove, Charles University, Hradec Kralove, Czechia

Despite distant metastases being the critical factor affecting patients' survival, they remain poorly understood. Our study thus aimed to molecularly characterize colorectal cancer liver metastases (CRCLMs) and explore whether molecular profiles differ between Synchronous (SmCRC) and Metachronous (MmCRC) colorectal cancer. This characterization was performed by whole exome sequencing, whole transcriptome, whole methylome, and miRNAome. The most frequent somatic mutations were in *APC*, *SYNE1*, *TP53*, and *TTN* genes. Among the differently methylated and expressed genes were those involved in cell adhesion, extracellular matrix organization and degradation, neuroactive ligand-receptor interaction. The top up-regulated microRNAs were hsa-miR-135b-3p and -5p, and the hsa-miR-200-family while the hsa-miR-548-family belonged to the top down-regulated. MmCRC patients evinced higher tumor mutational burden, a wider median of duplications and deletions, and a heterogeneous mutational signature than SmCRC. Regarding chronicity, a significant down-regulation of *SMOC2* and *PPP1R9A* genes in SmCRC compared to MmCRC was observed. Two miRNAs were deregulated between SmCRC and MmCRC, hsa-miR-625-3p and has-miR-1269-3p. The combined data identified the *IPO5* gene. Regardless of miRNA expression levels, the

combined analysis resulted in 107 deregulated genes related to relaxin, estrogen, PI3K-Akt, WNT signaling pathways, and intracellular second messenger signaling. The intersection between our and validation sets confirmed the validity of our results. We have identified genes and pathways that may be considered as actionable targets in CRCLMs. Our data also provide a valuable resource for understanding molecular distinctions between SmCRC and MmCRC. They have the potential to enhance the diagnosis, prognostication, and management of CRCLMs by a molecularly targeted approach.

KEYWORDS

colorectal cancer, MiRNAome, methylome, transcriptome, liver metastasis

Introduction

Metastasis, the process of spreading cancer cells from the primary site to distant organs, is a major cause of cancer mortality (1). The liver is the most common site of distant metastasis in colorectal cancer (CRC) patients. Colorectal cancer liver metastases (CRCLMs) are rather frequent. Even during the disease follow up, about 50% of CRC patients experience tumor occurrence at a distant site, resulting in a poor prognosis with a 14% five-year survival rate (2).

When the CRCLMs are not treated, the disease exhibits an unfavorable prognosis with the median overall survival (mOS) of less than 12 months. Even with the use of aggressive treatment (e.g. oxaliplatin-based treatment), the overall survival (OS) does not exceed 13–18 months (3–5). The only curative approach at present is surgical resection of isolated liver metastases. However, only 20% of resected patients attain long-term remission [1% of patients would develop liver metastasis during one-year follow-up, whereas 15% during five-year follow-up (5–11)], while 60–70% of patients experience local or distant recurrence (12). However, it remains clinically impossible to predict which patients are more likely to develop distant recurrence after resection of primary cancer. As the incidence and mortality of CRCLMs remain high, it is important and urgent to identify their etiology, molecular mechanisms, and biomarkers for early prediction and personalized treatment.

Patients with CRCLMs form a heterogeneous group. We can distinguish between Synchronous (SmCRC) and Metachronous (MmCRC) colorectal cancers. There is, however, still no consensus on the definition of SmCRC and MmCRC as used in the context of CRCLMs. Although by definition, all metastases are synchronous (occult or detectable at diagnosis), most definitions include detection at or before diagnosis or surgery of the primary tumor (13), while some also include metastases detected up to 3 (14, 15), 4 (16) or 6 months (17, 18) following diagnosis. It was suggested that both diseases might represent a distinct phenotype and impact therapy outcome, such as patients with SmCRC compared with patients with MmCRC have more adverse prognostic features and significantly shorter time to treatment failure (19). However, several studies did not confirm that (20, 21).

Over the past few decades, numerous preclinical and clinical studies have been conducted to uncover the underlying mechanisms of CRCLMs formation. However, only a limited number of results from these studies reached clinical practice, such as assessing PD-L1 expression in tumor cells using immunohistochemistry to provide prognostic information and predict response to treatment with PD-1/PD-L1 inhibitors (22). The lack of biomarkers in clinical practice is mainly because most studies focused on a single gene or a primary tumor.

In the present study, the high-throughput approaches with comprehensive bioinformatic analysis provided us with a platform for the analysis of metastatic tissue, both SmCRC and MmCRC, to find relevant hub genes and pathways and understand the molecular mechanism and characteristics of CRCLMs. In addition, we also investigated clinicopathologic information of SmCRC and MmCRC patients to improve patients' management and follow-up.

Materials and methods

Sample collection, transport, and storage

In this prospective single-center study, histologically confirmed CRCLM patients were recruited at the Department of Oncology and Radiotherapy (Hradec Kralove, Czech Republic) between June 2019 and December 2021. All CRCLM cases were monitored regularly until May 31st, 2022. Patients with any personal history of previous other malignancy or with CRC-associated well-defined inherited syndromes (including Lynch syndrome, familial adenomatous, and MUTYH-associated polyposis) were excluded from the study. In total, 10 CRCLM patients were included in the study.

Biological samples were immediately put in an RNA Later stabilization reagent (Invitrogen, USA) and stored at -80°C . Clinical data were collected from all subjects recruited in the study. The clinicopathological data for the patients recruited are reported in Table 1.

The local ethics committees at the Faculty Hospital in Hradec Kralove, Czech Republic (number of approval 201207- S01P) and

the Institute of Experimental Medicine, Prague, Czech Republic (number of approval 2018/05) approved the study. All patients provided written informed consent.

The colon mucosa from healthy individuals the fresh frozen samples were collected consecutively during the planned colonoscopy and were described in Jungwirth et al. (23).

Seed and soil mechanism and correction of bioinformatical data

The first cells invading the liver came from the primary CRC tumor. They thus also carry the genome/transcriptome of the primary tumor, respectively colorectum. Thus, comparing the genetic/molecular background of CRC cells with non-tumor liver cells could lead to results focused on organ specificity rather than on the distinctiveness of the metastasis relative to the healthy tissue. For this reason, we included a set of control colon tissues into the analyses apart from the patients' paired samples (metastases and adjacent liver tissue). Overall, control colorectal tissues (n=10) were used to eliminate false positive results from contamination of liver tissue. These tissues were used from healthy individuals with no previous or current cancer. This correction was used for DNA methylation assessment and RNA seq analysis. A more detailed description of this correction is explained in the bioinformatics section of the methods.

DNA isolation, quality, and quantity analyses

DNA isolation, quality and quantity analyses were all performed according to the manufacturer's instructions. For detailed description see the appendix.

Whole exome sequencing (WES)

We enriched protein-coding DNA using Human Whole Exome kit v7 (Agilent, USA) according to the manufacturer's instructions. For detailed description see the appendix.

The data curation was performed using the standard tools – FastQC, MultiQC (24) for quality control, Trimmomatic for data trimming (25), BWA for alignment (26), PicardTools for deduplication, GATK pipeline for variant calling (27), PureCN for tumor purity and CNV calling (28), SnpEff for variant annotations (29). For detailed description see the appendix. TMB was calculated as aTMB and fTMB as described in Zou et al. (30).

DNA methylation analysis

DNA methylation analyses were done by the Infinium Methylation EPIC Kit (Illumina, USA) according to the manufacturer's protocol and as described in Honkova et al. (31)

TABLE 1 Patients' characteristics.

Characteristics		N=10	Synchronous (N=7)	Metachronous (N=3)
Age	Years \pm SD	64.6 \pm 8.1	69.9 \pm 4.2	62.4 \pm 8.5
Gender	Men	7	5	2
	Women	3	2	1
Localization	Colon	5	4	1
	Rectum	5	3	2
Laterality ¹	Right-sided	2	2	1
	Left-sided	8	6	2
Histology	Adenocarcinoma, NOS	9	6	3
	Mucinous adenocarcinoma	1	1	0
Neoadjuvant therapy ²	Yes	2	2	0
	No	8	5	3
Adjuvant therapy ³	Yes	5	4	1
	No	5	3	2
Relapse after liver surgery	Yes	9	6	3
	No	1	1	0
Reoperation feasible	Yes	3	2	1
	No	6	4	2
	NA	1	1	0
Palliative therapy after liver surgery	Yes	5	3	2
	No	4	3	1
	NA	1	1	0
Living status	Alive	6	4	2
	Dead	4	3	1

¹Right- and left-sided primary tumors were defined as having their origin proximally or distally from the distant third of the transverse colon, respectively. ²Chemotherapy and/or targeted therapy before CRCLM resection. ³Chemotherapy after CRCLM resection. SD, standard deviation; NOS, not otherwise specified; NA, not applicable.

Raw data were processed with the minfi package (32). Data were normalized using the quantile method. A series of filtering was performed (probes with SNPs and crossreactive probes (33), probes to eliminate false positive results from tissue bias), resulting in the final number of 559,364 probes. SVA package was used for batch corrections (34). Differentially methylated loci were identified using limma package (35) and for the multiple testing of the false discovery rate (FDR), the p-values for the contrast of interest were adjusted to be below <0.01, which is regarded to be the most appropriate for microarray analysis (36). Annotatr package was used for the probe annotations (37) and clusterProfiler (38) and ReactomePA (39) for the functional analyses. For detailed description see the appendix.

RNA isolation, quality, and quantity analyses

RNA isolation, quality and quantity analyses were all performed according to the manufacturer's instructions. For detailed description see the appendix.

RNA-seq analysis

Ribosomal RNAs (rRNAs) were removed using NEBNext rRNA Depletion Kit (Human/Mouse/Rat, New England Biolabs, USA). RNA sequencing cDNA libraries were constructed according to the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina, as provided by the manufacturer (New England Biolabs, USA). For detailed description of sequencing see the appendix.

The data curation was performed using the standard tools – FastQC, MultiQC for quality control (24), Trimmomatic for data trimming (25), BBmap for ribosomal RNA filtering, STAR for alignment (40), RSEM for quantification (41). The tumor purity, immune and stromal proportion, and contamination of liver tissue were assessed by the ratio of somatic variants in non-differently expressed genes and by the ESTIMATE R package (42) ahead of the final expression calculation. The normalized counts of metastatic samples were then cleaned from the contamination of liver tissue by subtracting the calculated proportion of the liver tissue normalized counts. EdgeR was used for identifying differentially expressed genes (43). After Benjamin-Hochberg adjustment, gene was considered deregulated with FDR lower than 0.01 and absolute logFC greater than 1. clusterProfiler (38) and ReactomePA (39) were used for functional analyses. For detailed description see the appendix.

miRNA isolation, quality, and quantity analyses

RNA isolation, quality and quantity analyses were all performed according to the manufacturer's instructions. For detailed description see the appendix.

Small RNA-sequencing

The next-generation sequencing library preparation was carried out as described in Sabo et al. (44) and Cervená et al. (45). MiRNA libraries were constructed using the NEB Next Multiplex Small RNA Library Prep Set for Illumina (New England BioLabs, USA) according to the manufacturer's protocols.

FastQC and MultiQC (24) were used for quality control, Cutadapt tool (46) for trimming and BBMap for ribosomal RNA filtering. The miRge3.0 pipeline (47) was used for alignment and quantification. Significantly different miRNAs were identified by EdgeR R package (43). After BH adjustment, the miRNA was considered DE with a false discovery rate lower than 0.05.

External validation

For external validation of our data, the GSE62321 set was used (48, 49). This set comprises pairs of primary tumors and hepatic metastases before chemotherapy from 13 patients.

Bioinformatic in silico analysis

DepMap data analysis

To analyze the *IPO5* gene interactions, the correlation data were downloaded from the CRISPR (Avana) Public Depmap v20Q3 portal (<https://depmap.org/portal/download/>) for all cell lines in the database (1,078 cell lines). Gene correlation between *IPO5* knockout effect and gene expression was considered significant for a p-value lower than 0.05 and an absolute value of correlation coefficient higher than 0.1. Functional enrichment analysis of significantly correlated genes with the *IPO5* effect was performed using the clusterProfiler, and ReactomePA R packages. Only terms associated with a BH-adjusted p-value lower than 0.05 were considered.

Results

Patients' characteristics

The clinical and pathological characteristics of 10 patients with liver metastasis included in the study are described in Figure 1 and Table 1. Patients were predominantly males (70%) with a mean age of 64.6 ± 8.1 years. At the time of diagnosis, three patients had CRC stage III, while seven patients had stage IV disease with metastases confined to the liver. Regarding the therapy, two subjects received chemotherapy and/or radiation therapy prior to the surgical resection of the primary tumor and one after the primary tumor resection. All patients had CRCLM eligible for potential resection at the time of enrollment. In terms of chronicity, the group SmCRC patients comprised seven patients, while the MmCRC group three. Two patients received induction therapy with Bevacizumab/mFOLFOX6 prior to the CRCLM resection and six patients

received a fluoropyrimidine-based adjuvant chemotherapy following CRCLM resection. Nine patients relapsed with a median relapse-free survival (RFS) of 12.0 months (range 2.3–26.5 months). Three of these patients were eligible for subsequent surgery and five received palliative fluoropyrimidine-based chemotherapy and/or targeted therapy. Four patients had died during the follow-up period with a median OS of 19.4 months (range 18.6–35.6 months).

DNA-based analysis

Analysis of WES data

We have successfully analyzed the DNA isolated from 10 paired CRCLM and adjacent liver tissue samples.

After alignment, a range from 81 to 169 million properly aligned read pairs per sample were obtained. In summary, the percentage of adequately mapped read pairs reached 99% for all samples. The average read depth for each patient can be found in [Supplementary Table S1](#). Due to low data quality control, two out of ten patient samples had to be discarded from this analysis.

With the advantage of robust WES data, several features, such as mutational analysis, tumor mutation burden and copy number variations were analyzed with obtained WES data. The first analysis of all samples was performed, and then stratification for either SmCRC or MmCRC was presented.

Mutation analysis

WES in CRCLM and adjacent liver tissue DNA was performed to identify somatic tumor-specific single nucleotide variants (SNV) and short insertion/deletion (Indel) mutations characterizing liver metastasis.

In total, 1,900 heterozygous and homozygous SNVs and Indels (1,779 substitutions, 4 double substitutions, 58 short insertions, and 59 short deletions) have been discovered as different between these two groups: 104 with high, 361 with moderate, 645 with low and 816 with modifier putative effect. Regarding the SNV effect, the largest group of identified variants belonged to the missense and

intron variants, followed by synonymous variants. For the Polyphen2 calculation, we filtered out only non-synonymous exonic variants, and the results were as follows: benign prediction was among all samples in the range 32–48%, possibly damaging in the range 17–33%, and damaging prediction in the range 28–48%. After stratification for chromosome localizations, the observed variants' distribution was similar across all chromosomes ([Supplementary Figure 1](#)).

The most frequent somatic mutations at the time of enrolment were in *APC* (75%), *SYNE1* (50%), *TP53* (50%), and *TTN* (50%) genes. Only one *APC* mutation (rs121913333, c.2680C>T, p.Arg894*) was detected more than once (patients ID 5 and 22, [Supplementary Table S2](#)). No other mutations were observed more than once. The top genes with a mutation frequency of > 38% are presented in [Figure 2A](#).

In the terms of chronicity, we further compared the mutational landscape between SmCRC (n=5) and MmCRC (n=3) patients. No significant differences in distribution among these two groups were noticed after stratification for putative effect, SNV effect, chromosome distribution, or Polyphen2 characterization ([Supplementary Figure 2](#)).

Using the OncoPrint for visualization of differences in the distribution of genomic variations, the most observed difference between these two groups was noticed for *MUC16*, *RYR2* (SmCRC in 60% vs MmCRC 0%) and *USH2A*, *ACACA*, *ALS2CL*, *ATP10A*, *BRINP2*, *CACNA1L*, *CCNJ*, *CFAP46*, *DYNC2H1*, *ENPP7P13*, *KIA1549L*, *MED13L*, *PCDHA10*, *PRPF6*, *PTPRT* (0% vs 66%) genes. No other substantial differences were observed between these two groups ([Figure 2B](#)).

Tumor mutation burden analysis

The overall TMB representing the number of single nucleotide variations (SNVs), multiple nucleotide variations (MNVs), and InDels per Mb was calculated. TMB was defined as aTMB, which includes synonymous mutations, and fTMB, which excludes synonymous mutations. The average value of aTMB was 5.99 in all patients, and the median value of aTMB was 6.62 in all patients (SmCRC aTMB mean 5.84 and median 6.60, MmCRC aTMB mean

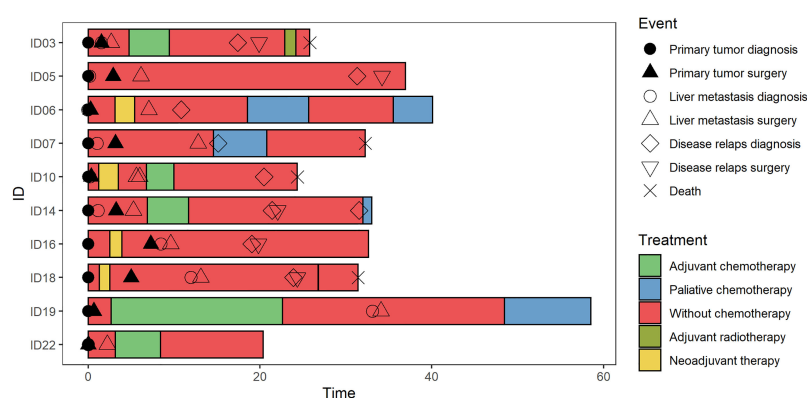


FIGURE 1

Swimmer plot. The clinical and pathological characteristics of 10 patients included in the study. Each bar represents one subject in the study, and symbols along each bar represent various relevant clinical events.

6.22 and median 6.64, respectively). The average value of fTMB was 5.21 in all patients, and the median value of aTMB was 5.75 in all patients (SmCRC fTMB mean 4.98 and median 5.59, MmCRC fTMB mean 5.59 and median 6, respectively).

Mutational signature

Mutational signature analysis revealed that transition mutations were more common than transversion mutations, and C > T substitutions were predominant in our samples (Supplementary Figure 3).

To better understand the pathogenesis of CRCLM, we performed mutational signature analysis on 2,696 (1,779 heterozygous/homozygous and 917 subclonal) SNVs by analyzing the six mutation classes (C > T, C > A, T > C, T > G, C > G, and T > A). According to the COSMIC SBS V3 database, we identified eight mutational signatures with a higher than 5% proportion in our CRCLM samples (in descending order SBS1, SBS6, SBS17b, SBS36, SBS30, SBS25, SBS3, and SBS9, Figure 2C).

In the terms of chronicity, SmCRC patients had a higher proportion of SBS1 (cause of mutations is age related) and SBS6 (cause is related to defective mismatch repair) mutational signatures, while MmCRC patients displayed a more

heterogeneous distribution of mutation signatures. Given the small number of patients in each group taking this result with caution is necessary.

CNV analysis

Using PureCN, we identified recurrent CNVs affecting 5,432 genes (4,564 gained genes and 868 lost genes, Supplementary Table S3) within at least 50% of our group, containing several already known and putative driver genes (Figure 2D). For all samples, the median duplication length was 2,675,594 bases, while the median deletion length was 783,537. The density of genomic sizes of structural variants in all CRCLM patients is depicted in Figure 3C. The deletions ranged from ~239 b to 92 Mb, with a distinct peak at ~496 kb. A broad range of differently sized duplications (~264b-147Mb) was observed, with a peak at 1.2 Mb.

In the terms of chronicity, for SmCRC, the median duplication length was 2,166,142 (range 264b-147Mb, peak at 1.2Mb) bases while the median deletion length was 116,0517 (range 241b-92Mb, peak at 929kb). For MmCRC, median duplication length was 3,267,214 and median deletion length 607,852 (range 269b-125Mb, peak at 1.6Mb, respectively 239b-66Mb, peak at 382kb, Supplementary Tables S4 and Supplementary Figure 4).

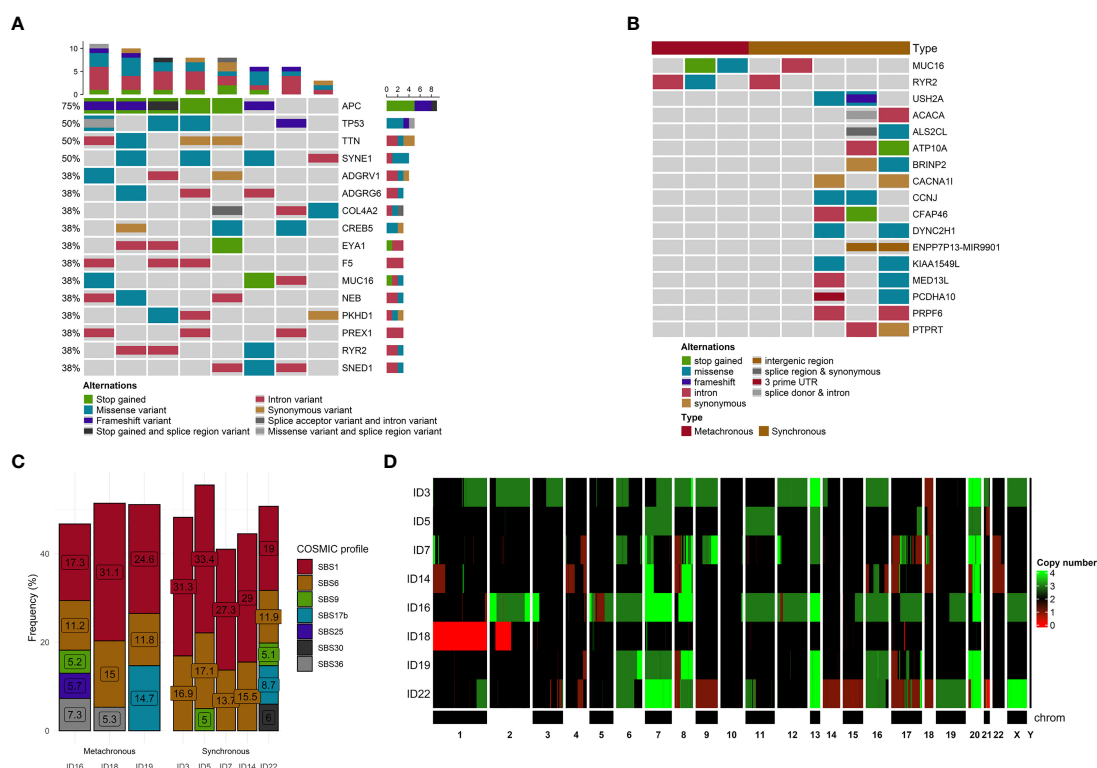


FIGURE 2

Mutational and copy number variants analyses. (A) The distribution of genetic alterations detected in all samples. Only genes affected, at least in 3 samples, are pictured. Each column represents a patient, and each row represents a gene. Different colors indicate different mutation types. The bar chart on the top shows the total number of the given gene mutations observed in the sample. (B) The distribution of genetic alterations detected between SmCRC and MmCRC groups. Each column represents either SMCRC or MmCRC group of patients, and each row represents a gene. Only genes with differences in relative group distribution difference of 50% or higher are pictured. Different colors indicate different mutation types. (C) The distribution of mutational signatures in percents according to the COSMIC SBS V3 database across all samples. Only signatures with distribution in single sample of more than 5% are pictured. Each column represents a patient. (D) Heatmap of CNV distribution across all samples. Each row is represented by a patient. A chromosome represents each column. The green color represents a gain, while red represents a loss within a chromosome locus.

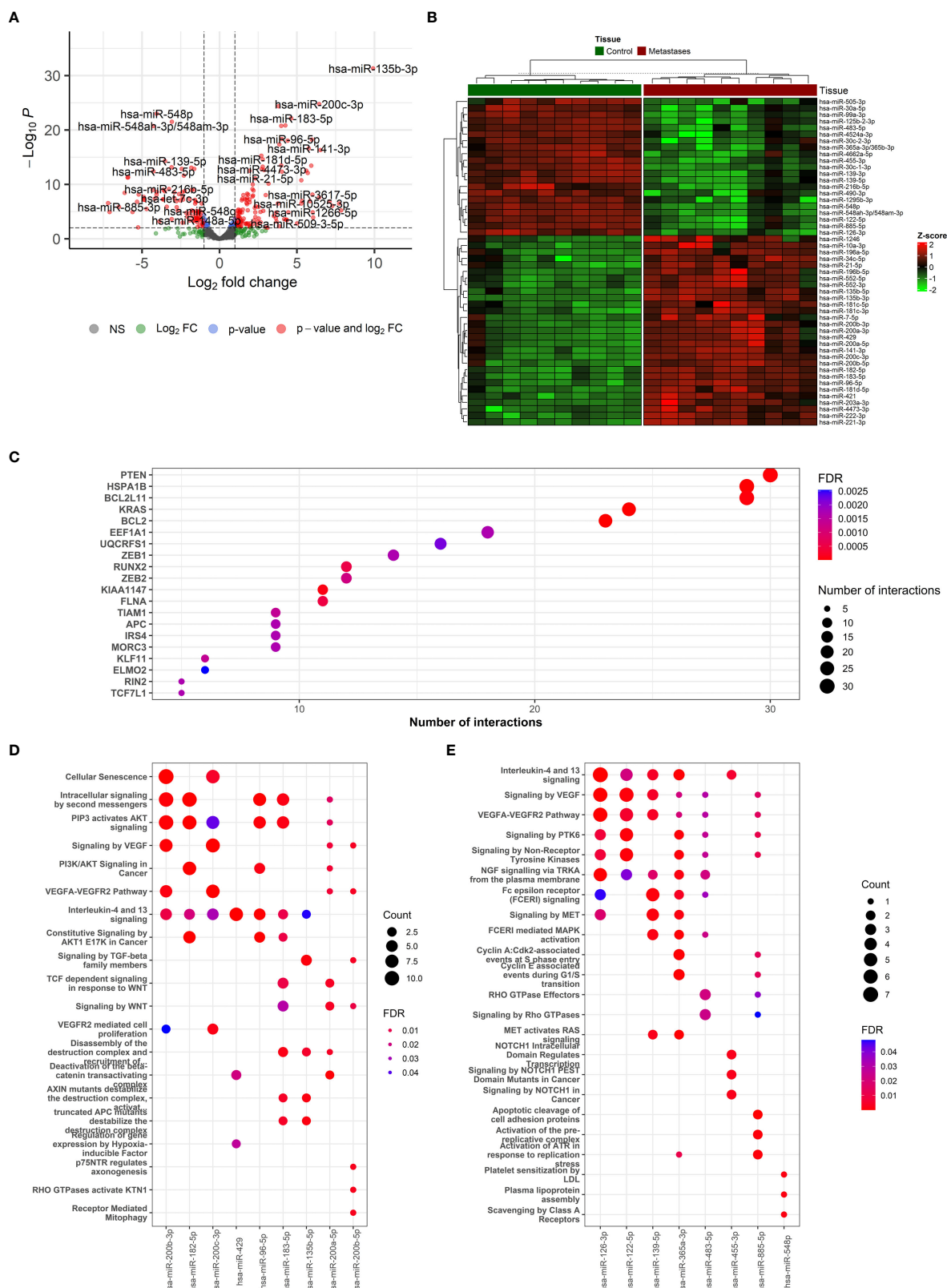


FIGURE 3
The description of miRNA-based results and functional enrichment of its targets. **(A)** Volcano plot reporting the \log_2 FC and the adjusted p-value in the DE analysis between metastasis and adjacent liver tissue. Thresholds: adjusted p-value < 0.0+ and $|\log_2$ FC| > 1. **(B)** Heatmap of top 50 differentially expressed miRNAs between metastasis and adjacent liver tissue. **(C)** The top miRNA targets from target enrichment analysis. **(D)** The Reactome enrichment analysis of the top 10 differentially up-regulated miRNAs. **(E)** The Reactome enrichment analysis of the top 10 differentially down-regulated miRNAs.

DNA methylation analysis

We have successfully analyzed the DNA isolated from 8 paired CRCLM and adjacent liver tissue samples.

To obtain an overview of DNA methylation profiles, first, the principal component analysis based on the differential DNA methylation of the CpG loci showed a clear separation among CRCLM and adjacent liver tissue (Supplementary Figure 5).

We next analyzed the major differences in DNA methylation distribution and found 25,865 CpG sites differentially methylated in CRCLM tissue compared to adjacent liver tissue. Among the differentially methylated CpG sites, 5,185 were hypermethylated and 20,680 hypomethylated and mapped to 3,412 different genes (1,043 hypermethylated and 2,369 hypomethylated). Hypermethylated CpG sites were predominantly observed in CpG islands (64.4%). On the other hand, hypomethylated CpG sites were mainly located in intergenic regions, the so-called open sea (79.8%) (Supplementary Figure 5).

Additionally, we identified those genes with the highest quantitative differences in methylation between CRCLM and adjacent tissue. The comparison was based on the analyzed $\Delta\beta$ -values and CpG sites with an adjusted p-value < 0.01. The distribution of the top 25 genes with the highest differences in methylation and hierarchical clustering is depicted in Figure 4A.

For the visualization, interpretation and analysis of affected pathways, the Reactome database was used for pathway enrichment in our list of differentially methylated genes (Supplementary Table S5 and Figure 4B). The hypomethylated genes were involved in

processes such as Cell adhesion and extracellular matrix (ECM) organization, and the hypermethylated genes in Neuroactive ligand-receptor interaction or calcium signaling pathways and various cancer-related pathways.

RNA-based analyses

RNA-seq analysis

We have successfully analyzed the RNA isolated from 9 paired CRCLM and adjacent liver tissue samples and 9 non-paired normal colon tissue samples.

The expression profiling of transcripts and differential expression of genes between the CRCLM and adjacent liver tissue samples were analyzed using RNA-seq data. As for WES, the first analysis of all samples was performed, and then stratification for either SmCRC or MmCRC was presented.

After the normalization of liver contamination in metastatic tissue by using paired liver samples, 14,121 genes remained in the analysis. A total of 2,711 differentially expressed genes (DEGs, adjusted p < 0.01 and absolute value of logFC > 1) were identified by an analysis of CRCLM compared with colon tissue. Of these, 1,719 were up-regulated, and 992 were down-regulated in CRCLM samples (Figure 5A and Supplementary Table S6). The stratification for the top 50 significant DEGs is depicted in Figure 5B.

To further characterize the changes in mRNA expression levels in CRCLM and adjacent liver tissue samples, functional pathway

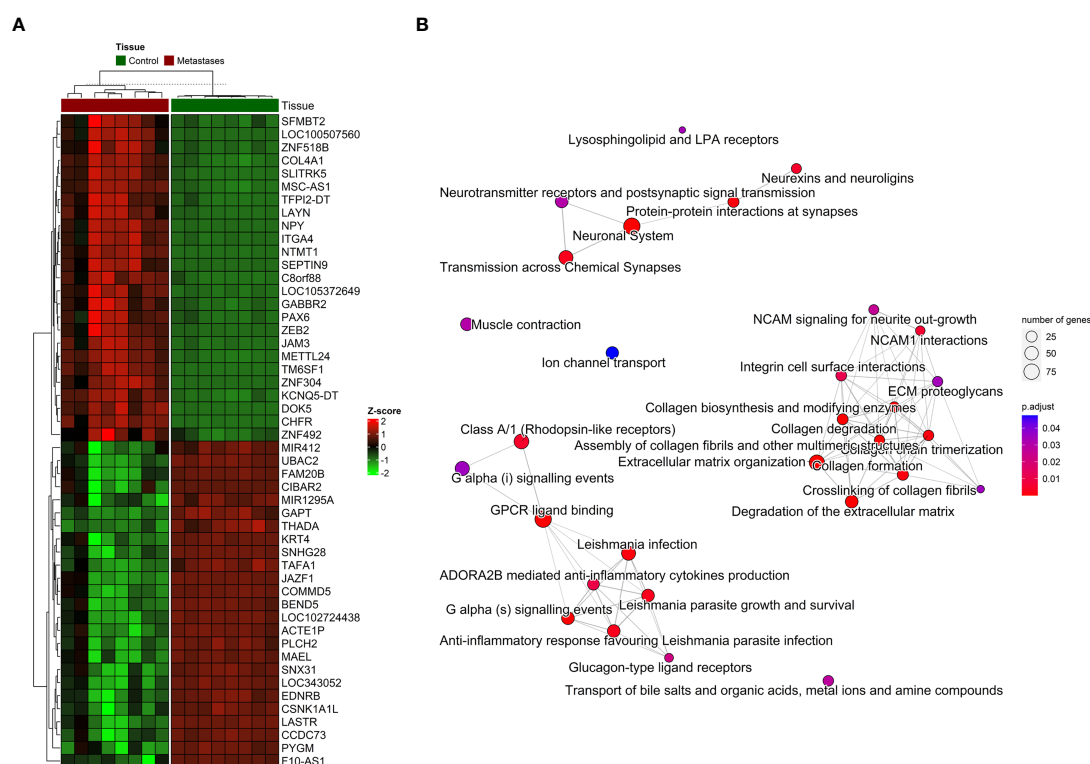


FIGURE 4

The distribution of DNA methylation patterns. (A) Heatmap of top 25 significantly hypermethylated and top 25 significantly hypomethylated genes between metastasis and adjacent liver tissue. (B) Emplot of the overrepresentation analysis (ORA) pathway enrichment according to the Reactome database.

enrichment analysis was performed to identify the main related biological functions. Gene set enrichment analysis (GSEA) uncovered 124 significantly deregulated pathways in the Reactome terms (adj. p-value < 0.05, top 80 in Figure 5C). Overrepresentation analysis (ORA) of up-regulated (logFC > 1) and down-regulated (logFC < -1) genes are reported in Supplementary Table S7. The most significant results were obtained for down-regulated genes, which were enriched in Reactome terms related to Ion channel transport, the citric acid cycle, and respiratory electron transport. Conversely, the up-regulated genes were enriched prevalently in ECM organization and degradation of ECM. Similarly, GSEA on KEGG terms uncovered 27 significant terms, ORA on upregulated genes described 15 significant KEGG terms, and ORA on downregulated genes described 17 significant KEGG terms (Supplementary Figure 6).

In the terms of chronicity, only two genes were significantly down-regulated in SmCRC compared to MmCRC (*PPP1R9A* logFC=2.8, FDR=0.05, and *SMOC2* logFC=2.6, FDR=0.05).

miRNA-seq analysis

We have successfully analyzed the miRNA isolated from 10 paired CRCLM and adjacent liver tissue samples.

To identify the miRNA signature for CRCLM detection, the miRNA expression profiling between the CRCLM and adjacent liver tissue sample was analyzed using miRNA-seq data. As for WES and RNA-seq, the first analysis of all samples was performed, and then stratification for either SmCRC or MmCRC was presented. Additionally, we also looked at valid miRNA targets that may be associated with differential miRNA expression levels in CRCLM.

After sequencing, an average of 57.3% of the reads were aligned to miRNA sequences, while 10.2% were aligned to other small non-coding RNAs (sncRNAs) (Supplementary Figure 7; Supplementary Table S8). In total, 1,603 miRNAs were detected, with an average of 845 miRNAs detected in each sample and 670 miRNAs with sufficient coverage across all samples for differential expression calculation. Other classes of sncRNAs were seen, including, tRNAs, and snoRNAs. The distribution of the detected non-miRNA sncRNAs in the sample group is shown in Supplementary Table S9. In total, 2,415 non-miRNA sncRNAs were identified in all subject groups.

In total, 123 miRNAs were up- and 92 down-regulated in CRCLM compared to adjacent liver tissue (Supplementary Table S10). Among the up-regulated top miRNAs were hsa-miR-135b-3p, hsa-miR-200c-3p, hsa-miR-200b-5p, hsa-miR-183-5p, hsa-miR-182-5p, hsa-miR-200b-3p, hsa-miR-96-5p, hsa-miR-135b-5p, hsa-miR-429 and hsa-miR-200a-5p. Among the top down-regulated were hsa-miR-548p, hsa-miR-99a-3p, hsa-miR-548ah-3p/548am-3p, hsa-miR-139-5p, hsa-miR-125b-2-3p, hsa-miR-30c-1-3p, hsa-miR-455-3p, hsa-miR-126-3p, hsa-miR-483-5p, hsa-miR-365a-3p/365b-3p. The levels of the most significant deregulated miRNAs are reported in Figures 3A, B.

In the terms of chronicity, only two miRNAs were deregulated between SmCRC and MmCRC (hsa-miR-625-3p and hsa-miR-1269-3p, Supplementary Table S11).

The miRNA binding to its target transcript does not necessarily lead to a downregulation of gene expression. In fact, most of the observed miRNA binding events have minor functional consequences. Thus, focusing on miRNA binding alone has limited value in predicting valid miRNA targets, i.e., downregulated targets. To alleviate this problem, we directly determined the target downregulation of miRNA-seq with RNA-seq. Among the top miRNA targets were observed *PTEN*, *HSPA1B*, *BCL2L1*, *KRAS*, and *BCL2* genes (Figure 3C). Gene set enrichment analysis of targeted mRNAs by miRNAs can support the role of differentially expressed miRNAs. The Reactome enrichment in Figures 3D, E and the KEGG enrichment analysis of differentially expressed miRNA-targeted pathways is presented in Supplementary Figure 7. Among these pathways Signaling pathways regulating pluripotency of stem cells, TGF-Beta, WNT, Ras, mTOR, and NOD-like receptor signaling pathway were the most prominent pathway targeted by deregulated miRNAs.

Integration of results from the genome, methylome, transcriptome, and miRNAome analyses

To go behind the source of gene deregulation in our set, we also performed integrated analysis for the regulatory relationships of DNA methylation, genetic selection, and miRNA expression data, together with mRNA expression data. The overlap of alterations discovered between the CRCLM across all platforms was performed (Figure 6A). We have identified 2,711 differentially DEGs, out of which 612 could be explained as deregulation due to the CNV effect, 22 due to the miRNA expression levels, and 212 due to methylation profile. CNV, methylation, miRNA, and DEGs overlapped at one gene, *IPO5* (high expression profile, high miRNA expression, hypomethylation profile, and gain in the CNV analysis, Figure 6B).

To analyze the *IPO5* gene interactions, the correlation data were downloaded from the CRISPR (Avana) Public Depmap v20Q3 portal (<https://depmap.org/portal/download/>) for all cell lines in the database (1,078 cell lines). With the threshold for adjusted p-value 0.05 and absolute value of correlation coefficient >0.1, 1,739 genes that interact with the *IPO5* gene were identified.

The most significant results obtained for *IPO5* gene enriched in Reactome were related to the cell cycle. The Reactome analysis for *IPO5* gene was visualized in Supplementary Figure 8 and Supplementary Table S12.

It is well known that one miRNA can regulate several genes and vice versa, multiple miRNAs can regulate one gene. For this reason, we were also interested in overlapped genes without considering the miRNA expression levels. This combined analysis resulted in 107 genes (Figure 6B and Supplementary Table S13). GSEA on Reactome and KEGG terms uncovered significant terms (Figure 6C and Supplementary Table S14) related to relaxin, Estrogen, PI3K-Akt, WNT signaling pathways and intracellular signaling by the second messenger.

Due to the low numbers in each group, no stratification for SmCRC and MmCRC was not performed.

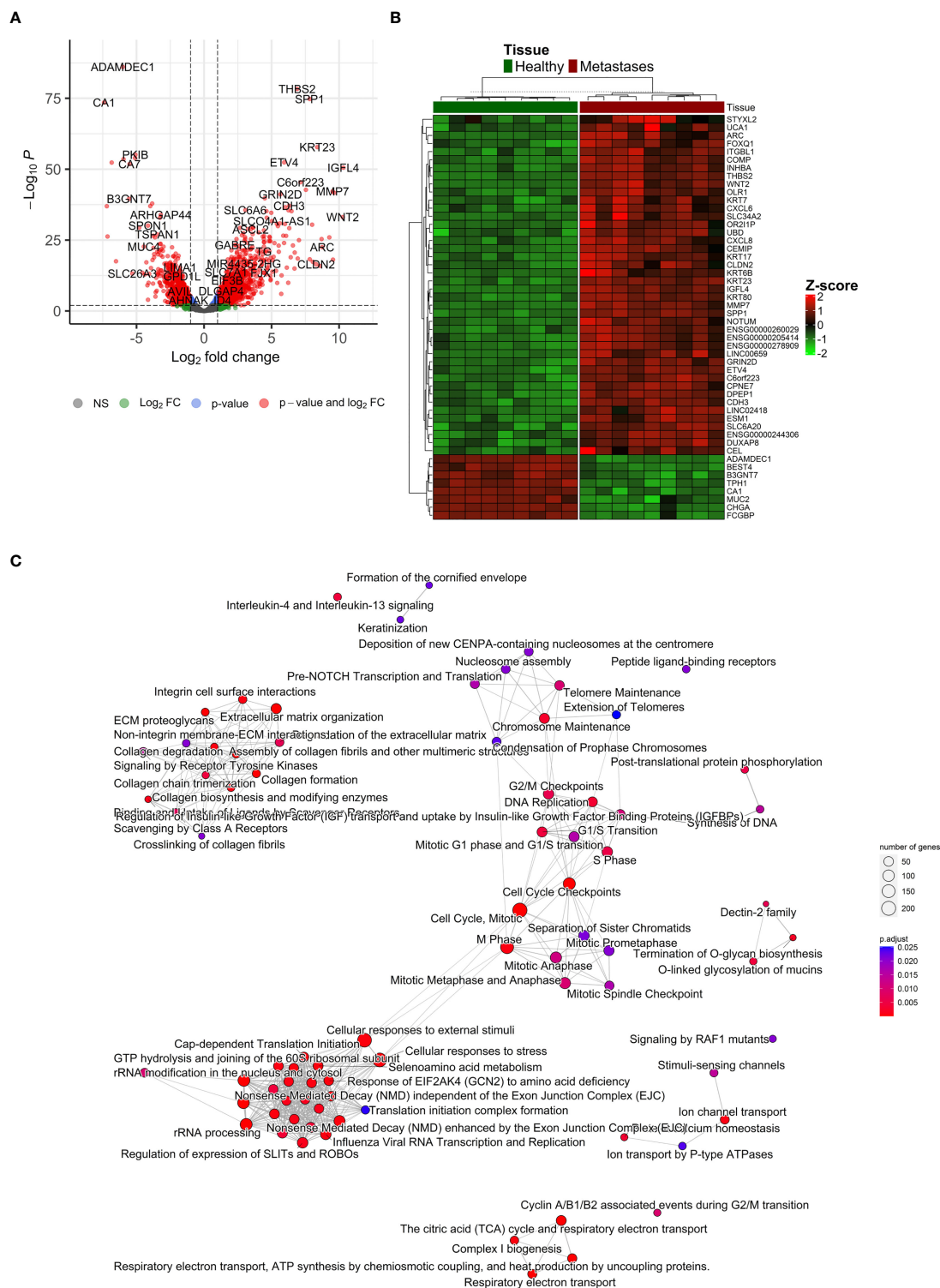


FIGURE 5

Differentially expressed genes 'characterization. (A) Volcano plot reporting the \log_2FC and the adjusted p-value in the DE analysis between metastasis and adjacent liver tissue. Thresholds: adjusted p-value < 0.01 and $|\log_2FC| > 1$. (B) Heatmap of top 50 differentially expressed genes between metastasis and control colon tissue. (C) Emaplot of the gene set enrichment analysis (GSEA) of the significantly deregulated pathways in the Reactome terms (adj. p-value < 0.05, top 80 terms).

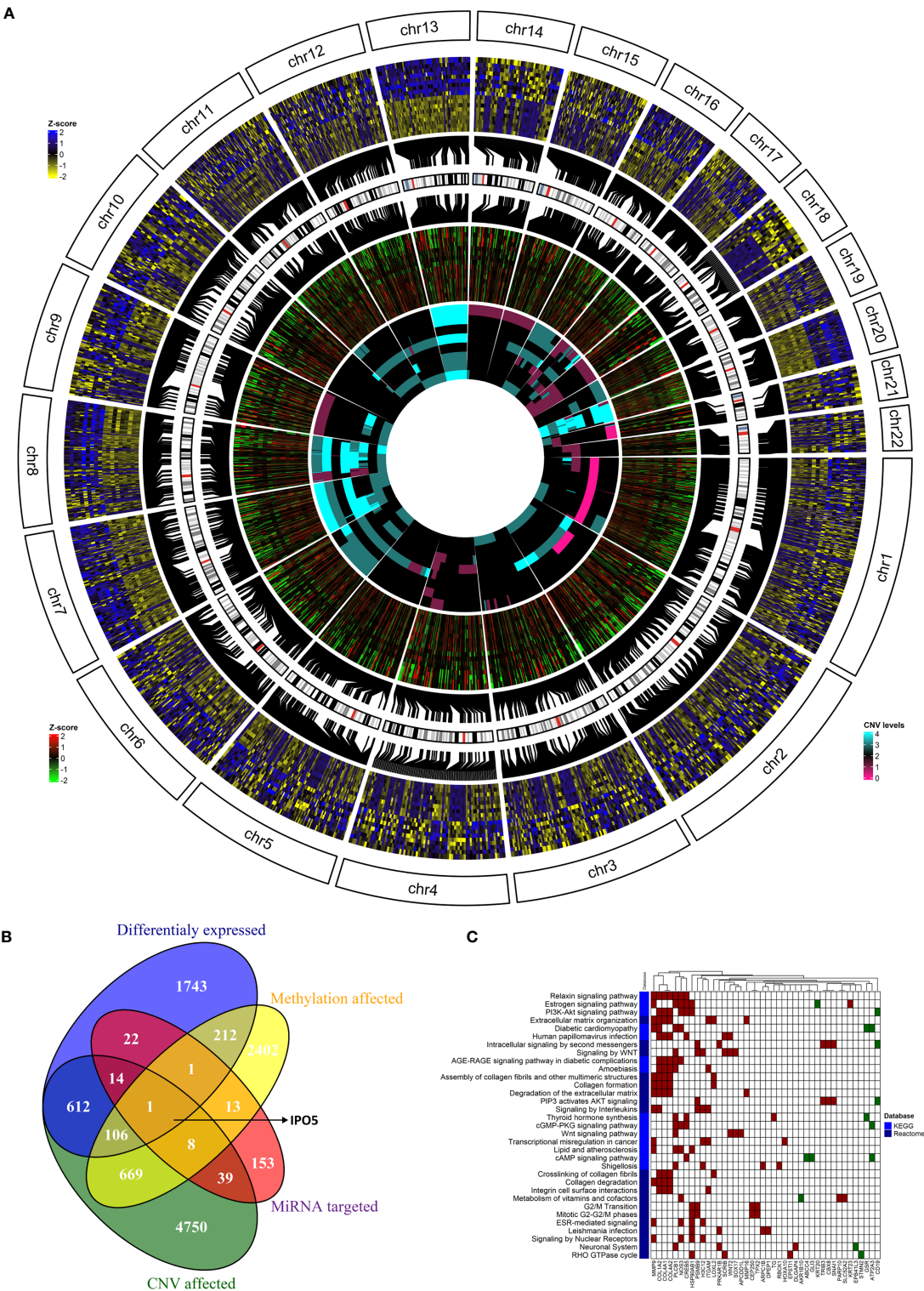


FIGURE 6
The integration of results from the genome, methylome, transcriptome, and miRNAome analyses. **(A)** The overlap of alterations discovered between the CRCLM patients across all platforms. The outside heatmap circle represents the significant DEGs, the blue color represents up-regulation, yellow color represents down-regulation. The middle heatmap circle represents the methylation level for the genes depicted in the outside circle, the red color represents hypermethylation, the green color hypomethylation, and the black color non-significant changes. The inner heatmap represents the CNV level for the genes depicted in the outside circle. The cyan color represents gains, the purple represents losses. **(B)** Venn diagram showing the overlap between DEGs obtained in the analysis stratified by methylation profile, CNV, and miRNA expression levels. **(C)** The heatmap of KEGG (light blue) and Reactome (dark blue) enriched pathways and genes from integrated analysis of genome, methylome and transcriptome analyses. Up-regulated genes have red square while down-regulated genes have green square. White square means that the gene is not present in the specific pathway.

External validation

To validate our results, the GSE62321 set was used (48, 49). After data processing 9,115 genes were left out of which 2,873 genes were significantly deregulated between colon and metastases. The intersection between our and this validation set resulted in 5,918 genes where 592 genes were significantly deregulated (332 up- and 260 down-regulated) in the validation set and 1,060 (624 up- and 436 down-regulated) in our set. The final intersection between these two sets resulted in 366 DEGs where 204 were up- and 162 down-regulated (Supplementary Table S15).

The same validation set also comprised primary tumors from patients with metastasis (105 significantly DEGs between primary tumor and metastasis). Comparing this set with our set, the intersection resulted in 9 significant DEGs (2 upregulated and 7 downregulated, Supplementary Table S16).

Discussion

In newly diagnosed CRC patients without metastases, elucidating potential molecular risk factors for the development of liver metastasis is paramount, as they could have important clinical implications. For this reason, we focused on CRCLM patients as, after all, metastasis constitutes the primary cause of death for >90% of patients with CRC (1). Understanding the dynamics of this process will help identify targets for molecular therapies that may halt or possibly reverse cancer growth and metastasis. By identifying these targets, therapies can be designed to target cancer cells more precisely through either selective disruption of pathways necessary for cancer cell survival/growth or artificial modulation of the patient's immune system to generate a response against cancer cells. Besides, given its potential impact on patient care, a better understanding of the development of distant metastases is critical.

This study aimed to molecularly characterize metastatic tissues in CRCLM patients and explore whether molecular profiles differ between SmCRC and MmCRC. This characterization was performed by high-throughput approaches, such as WES, whole transcriptome, whole methylome, and miRNAome and analyzed by comprehensive bioinformatic tools. We hypothesize that combining all these approaches into comprehensive molecular profiling would lead to more relevant findings than single analyses published previously. We have demonstrated the utility of this integrative and broad approach where each genomic platform, independently and in combination, contributed to detecting pathogenic variants in all samples. We first analyzed each approach separately and then combined all the analyzes together and summarized the results below.

By analyzing the results from the WES only, we revealed an APC mutation (rs121913333, c.2680C>T, p.Arg894*) that was detected more than once. No other mutation was observed in any other gene more than once. Similarly, Naxerova and co-authors reported that in 65% of cases, regional lymphatic and distant metastases arose from independent subclones from the primary tumor (50).

CRC cells acquire a capacity to evade the primary CRC through morphological changes such as EMT, migration through the ECM, invasion into the neighboring tissues, intravasation, survival in the circulation, extravasation and finally, colonization to distant liver forming more aggressive CRCLMs (51). The ability of cells to undergo all these steps in the metastatic cascade requires them to acquire specific characteristics connected to the 'hallmarks of cancer'. In addition, Vermeulen et al. (52) proposed liver metastasis as a heterogeneous tumor and classified CRC liver metastasis into 3 growth patterns, i.e., pushing, desmoplastic, and replacement, based on histological differences. This agrees with our findings, the up-regulated genes in our study were prevalently enriched in the ECM degradation and organization, and hypomethylated genes were involved in processes such as Cell adhesion and ECM organization.

Deregulation of gene expression in the metastatic process is a complex process involving several modalities, including genomic changes or epigenetic modifications. Although the mutations in a few driver genes, such as APC and TP53, are shown to be one of the drivers of CRC progression as well as a biomarker of CRC stage and resistance to various CRC therapies (53), the true heterogeneity between patients is represented by the CNV distribution. The differences between patients and even the tumor cells were observed as highly heterogeneous (54). Our results supported these findings, as only a relatively low number of genes was mutated in our cohort while CNV distribution affected more genes, and the range was higher. Although only 10% of affected genes by CNV had an impact on expression deregulation, it was the major source of gene deregulation in our set. The well-known gains and losses occurring in the CRC progression, specifically on chromosomes 8, 13, 20, and chromosomes 4 or 18, respectively (55) were also observed in our cohort.

Another gene expression deregulation source are small non-coding RNAs, such as miRNAs. In this study, among the up-regulated top miRNAs were hsa-miR-135b-3p and -5p, and the hsa-miR-200-family. Among the top down-regulated was hsa-miR-548-family. Serum hsa-miR-135b (56) and has-miR-200c (57) levels were already considered as diagnostic (56) or prognostic and metastasis-predictive (57) biomarkers in patients with CRC. High serum miR-200c demonstrated a significant positive correlation with lymph node metastasis, distant metastasis, and prognosis (57). Regarding the has-miR-548-family, has-miR-548b was already found to be down-regulated in CRC patients as well as in CRC cells and even lower in advanced stages. The overexpression of miR-548b suppressed cells' proliferation and induced apoptosis. Our observation means that the downregulation of miR-548b may lead to uncontrolled cell division. Besides, WNT2 was predicted to be the downstream gene binding miR-548b (58). Almost all CRC cases demonstrated hyperactivation of the WNT pathway, which is considered to be the initiating and driving event of CRC (59).

When the intersection of all our data is considered, the combined data from CNV analysis, DNA methylation profiles, and mRNA sequencing led to identifying the IPO5 gene. Several sources upregulated the IPO5 gene: in 75% of our samples, the gain in this gene was observed and concurrently was significantly hypomethylated. On the other hand, miRNAs targeting IPO5

were upregulated as well. We hypothesize that the upregulation of targeting miRNAs might be driven by counter-regulation to reduce the impact of the *IPO5* upregulation to establish a balanced state. The *IPO5* gene belongs to the Karyopherins family, which are crucial regulatory molecules of nuclear plasma transport and represent the most classic cellular transporter proteins, including both importins and exportins (60, 61). Transport proteins with a molecular weight greater than 40 kDa help to transport molecules between the cytoplasm and the nucleus through the nuclear pore complex, such as transcription factors, splicing factors, and other proteins (62, 63). Therefore, karyopherin dysfunction can lead to altered transport activity and cause abnormal localization of oncogenic factors, consequently leading to tumorigenesis (64). Recently, Zhang et al. (65) noticed that the expression of the *IPO5* gene was gradually growing with the increasing CRC stage. In addition, the high *IPO5* expression, especially in CRC cells, was also confirmed in the TCGA database and Oncomine database in their study. Besides, the functional assays revealed that *IPO5* promoted CRC growth *in vitro* and *in vivo* through the RASAL2 nuclear translocation, followed by the activation of the RAS signaling pathway. Inhibition of the nuclear transport system, therefore, has future potential for therapeutic intervention and could contribute to the elucidation of disease mechanisms. *IPO5* gene also seems to play an important role in the metastatic process *via* matrix metalloproteinase 7 (*MMP7*) regulation (66) and EMT induction (67). Furthermore, *IPO5* expression may affect the tumor immune microenvironment and mediate tumor immune response (68). The commensal microbiota has already been shown to influence immunity as well as tissue development. In the context of cancer, commensal bacteria have been shown to play a key role in modulating tumor microenvironment, which drives cancer therapy responses (69). Therefore, direct inhibition of *IPO5* function may uncover a promising targeted therapeutic strategy for CRCLM.

On the other hand, as one miRNA can regulate several genes, and conversely, multiple miRNAs can regulate one gene, we were also interested in intersection regardless of miRNA expression levels. This combined analysis resulted in 107 deregulated genes. GSEA on Reactome and KEGG terms revealed significant terms related to relaxin, estrogen, PI3K-Akt, WNT signaling pathways, and intracellular second messenger signaling. All these pathways were assigned to the CRC evolution and thus made our results valid even if they were obtained from a smaller number of patients.

The occurrence of CRCLM, such as SmCRC vs. MmCRC detection, was mainly investigated and reported in the surgical case series (21, 70, 71). The gene expression and molecular patterns of SmCRC and MmCRC are considered different. Synchronous liver metastases are similar to local invasion and are more inclined to become a disseminated disease (72). One of our main aims was the comparison of the molecular characteristics of these two groups. MmCRC patients evinced higher TMB, a wider median of duplications and deletions, as well as heterogeneous mutational signatures in our set of patients. We thus hypothesize that these differences could be explained as a consequence of the adaptation of resistant clones to escape previously applied chemotherapy in MmCRC patients. Similarly, higher TMB in MmCRC patients could be related to better immunosurveillance and thus delayed

relapse (compared to SmCRC), as higher TMB is associated with higher tumor immunogenicity (73). However, given the small number of patients in each group, it is necessary to take this result with caution.

Further, in terms of chronicity, we observed a significant down-regulation of *SMOC2* and *PPP1R9A* genes in SmCRC patients compared to MmCRC. *SMOC2* is an extracellular glycoprotein involved in a broad spectrum of cellular processes, including cell cycle, cell attachment and migration, angiogenesis, and others (74). It has a suppressive role in tumor growth, migration, colon and sphere formation role in CRC cells and could be considered a tumor suppressor in CRC progression (74). *PPP1R9A* gene codes a protein Neurabin-1 (75), which binds to protein phosphatase 1 (PP1) and inhibits its activity (76). Neurabins are highly concentrated in dendritic spines, and post-synaptic densities and their function is to regulate synaptic transmission in mammalian neurons (77). Despite there are no reports on the role of Neurabin-1 in carcinogenesis, the loss of Neurabin-2 (also known as Spinophilin or *PPP1R9B*), which has 80% homology in its sequence and similar biological functions to Neurabin-1, was associated with more aggressive histological phenotype, faster relapse, poor survival, and a low response to fluoropyrimidine-based chemotherapy in CRC (78, 79). Interestingly, the loss of Neurabin-2 is associated with an increase in the stemness properties (80) which may facilitate the spread to the liver. Given these circumstances, it could be hypothesized that the down-regulation of both *SMOC2* and *PPP1R9A* genes might be associated with rapid CRCLM formation in patients with SmCRC. Further research is warranted to evaluate the potential value of these genes as predictors of prognosis and risk of CRCLM formation in CRC patients.

Only two miRNAs were deregulated between SmCRC and MmCRC, hsa-miR-625-3p and has-miR-1269-3p. Interestingly, hsa-miR-625-3p induced oxaliplatin resistance by abrogating MAP2K6-p38-regulated apoptosis and cell cycle control networks (81). Bu et al. (82) observed that in stage II CRC patients, miR-1269a expression in their surgically removed primary tumors was strongly associated with the risk of CRC relapse and metastasis. The authors hypothesized that miR-1269a was a potential marker to contribute to adjuvant chemotherapy decisions for CRC patients and a potential therapeutic target to deter metastasis. The fact that we found these two miRNAs deregulated right between the SmCRC and MmCRC patients where the MmCRC patients had already undergone both surgical and adjuvant treatment strengthens our results considerably.

In summary, we have identified genes and pathways that may be considered actionable targets in CRCLMs (e.g. *IPO5* gene, hsa-miR-135b-3p and -5p, the hsa-miR-200-family, and hsa-miR-548-family, WNT and PI3K-Akt signaling). The data presented here also provide a valuable resource for understanding molecular distinctions between SmCRC and MmCRC, and have the potential to enhance the diagnosis, prognostication and management of CRCLMs by a molecularly targeted approach. The strength of this study lies in the sizeable high-throughput analysis, including the non-trial patients reflecting real-world patients. On the other hand, we are aware of certain limitations of our study, such as the lack of primary tumors due to the

prospective design. However, the intersection between our and validation sets confirmed the validity of our results.

Detailed molecular analysis of mechanisms that mediate metastatic expansion to the liver will contribute to early detection and prevention. Future research should focus on elucidating the origin of CRCLM based on the molecular mechanisms and clinical characteristics – this elucidation could thus guide clinical precision treatment. Targeted treatments to specific regulatory molecules, such as *IPO5*, make personalized cancer therapy possible. Many oncogenic cellular processes can intervene effectively, which is the promise of accuracy in eradicating cancer and better patient care. Defining high-risk factors for developing CRCLM, stratification of high-risk individuals and minimization of the controllable risk factors are essential to prevent CRCLM and further reduce CRC mortality.

Data availability statement

All the processed data are in the manuscript's supplementary data. Unfortunately, the raw data supporting the conclusions of this article can not be published in the public repository due to the lack of patient consent. Still, they will be made available by the authors on request. The validation data set was publicly available here GSE62321.

Ethics statement

The studies involving human participants were reviewed and approved by ethics committees at the Faculty Hospital in Hradec Kralove, Czech Republic (number of approval 201207- S01P) and the Institute of Experimental Medicine, Prague, Czech Republic (number of approval 2018/05). The patients/participants provided their written informed consent to participate in this study.

Author contributions

VV, JP, and LV secured the financial funding, VV and SF were responsible for the experimental design, OK, SJ, MM, and JP collected the patients' samples, MK was responsible for the pathological assessment of the samples, AS performed the isolation of the samples and preparation for sequencing methods, KH and IC performed the methylation array analyses, JH and SG-M were responsible for bioinformatical analyses of data from all methods, VV, JH, and OK wrote the paper with the input from all authors. All co-authors have participated in the planning and execution of the study as well as the preparation of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This project was supported by the Czech health research council of the Ministry of Health of the Czech Republic (NV19–09–00237),

by the Cooperatio Program, research area, Oncology and Haematology", Cooperatio No. 43 -Surgical Disciplines, and by the National Operation Programme: National Institute for Cancer Research LX22NPO05102. The methylation part was funded by the Ministry of Education, Youth and Sports of the Czech Republic (Research Infrastructure NanoEnviCZ, LM2018124) and The European Union - European Structural and Investments Funds in the frame of Operational Programme Research Development and Education - project Pro-NanoEnviCz (Project No. CZ.02.1.01/0.0/0.0/16_013/0001821).

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.

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/fonc.2023.1133598/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Descriptive mutational analysis (A) The distribution of genetic variation in percent according to the putative effect across all samples. Each column represents a patient. (B) The distribution of genetic variation in percent according to the single nucleotide variation effect across all samples. Each column represents a patient. (C) The distribution of genetic variation in percent according to PolyPhen2 across all samples. Each column represents a patient. (D) The distribution of genetic variation in percent across all samples after stratification for chromosome localizations. Each column represents a patient.

SUPPLEMENTARY FIGURE 2

Descriptive mutational analysis between SmCRC and MmCRC patients (A) The distribution of genetic variation in percent according to the putative effect between SmCRC and MmCRC patients. Each column represents a group either of SMCRC or MmCRC group of patients. (B) The distribution of genetic variation in percent according to the single nucleotide variation effect between SmCRC and MmCRC patients. Each column represents either of SMCRC or MmCRC group of patients. (C) The distribution of genetic variation in percent according to PolyPhen2 between SmCRC and MmCRC patients. Each column represents either of SMCRC or MmCRC group of patients. (D) The distribution of genetic variation in percent between SmCRC and MmCRC patients after stratification for chromosome localizations. Each column represents either of SMCRC or MmCRC group of patients.

SUPPLEMENTARY FIGURE 3

Mutational signatures (A) Mutational signature analysis in all samples. Each row is represented by a patient. (B) The distribution of mutational signatures in percent across all samples. Each column represents a patient.

SUPPLEMENTARY FIGURE 4

Densities of structural variants (A) The density of genomic sizes of structural variants in all patients. (B) The density of genomic sizes of structural variants in SmCRC patients. (C) The density of genomic sizes of structural variants in MmCRC patients.

SUPPLEMENTARY FIGURE 5

Genomic distribution of differentially methylated regions (A) The principal component analysis among metastasis (M) and adjacent liver tissue (H). (B) The Circos-Manhattan plot shows the CpG sites differentially methylated in CRCLM tissue compared to adjacent liver tissue, the x-axis shows the location in genome, and y-axis is $-\log(p\text{-value})$. (C) The distribution of differentially methylated CpG sites in percent in relation to the position in gene. Profile of all sites, significantly hypermethylated and significantly hypomethylated, is shown. (D) The distribution of differentially methylated CpG sites in percent in relation to the CpG island position. Profile of all sites, significantly hypermethylated and significantly hypomethylated, is shown.

SUPPLEMENTARY FIGURE 6

Functional pathway enrichment analysis in KEGG (A) Emapplot of the gene set enrichment analysis (GSEA) of the significantly deregulated pathways in the

Reactome terms (adj. p-value < 0.05, all 27 terms), (B) The overrepresentation analysis (ORA) of significantly up-regulated genes (adj. p-value < 0.05; all 15 terms), (C) The overrepresentation analysis (ORA) of significantly down-regulated genes (adj. p-value < 0.05; all 17 terms).

SUPPLEMENTARY FIGURE 7

Descriptives of miRNAs based results and KEGG based enrichments (A) The alignment of miRNA and other small non-coding RNAs (sncRNAs) sequences for each patient. Each column represents metastasis (M) and adjacent liver tissue (H), and the y-axis represents the number of reads to the sncRNAs groups. (B) The KEGG enrichment analysis of the top 10 differentially up-regulated miRNAs. (C) The KEGG enrichment analysis of the top 10 differentially down-regulated miRNAs.

SUPPLEMENTARY FIGURE 8

The Reactome *IPO5* gene interaction Emapplot of the overrepresentation analysis (ORA) of the interacted pathways in the Reactome terms for *IPO5* gene extracted from DepMap (adj. p-value < 0.05, top 50 terms).

References

- Seyfried TN, Huysentruyt LC. On the origin of cancer metastasis. *Crit Rev Oncog* (2013) 18:43–73. doi: 10.1615/CritRevOncog.v18.i1-2.40
- Filip S, Vymetalkova V, Petera J, Vodickova L, Kubecek O, John S, et al. Distant metastasis in colorectal cancer patients-do we have new predicting clinicopathological and molecular biomarkers? A comprehensive review. *Int J Mol Sci* (2020) 21:E5255. doi: 10.3390/ijms21155255
- Misiakos EP, Karidis NP, Kouraklis G. Current treatment for colorectal liver metastases. *World J Gastroenterol* (2011) 17:4067–75. doi: 10.3748/wjg.v17.i36.4067
- Nassabein R, Mansour L, Richard C, Vandenbroucke-Menu F, Aubin F, Ayoub J-P, et al. Outcomes of older patients with resectable colorectal liver metastases cancer (CRLM): Single center experience. *Curr Oncol* (2021) 28:1899–908. doi: 10.3390/curroncol28030176
- Stewart CL, Warner S, Ito K, Raoof M, Wu GX, Kessler J, et al. Cytorreduction for colorectal metastases: liver, lung, peritoneum, lymph nodes, bone, brain. when does it palliate, prolong survival, and potentially cure? *Curr Probl Surg* (2018) 55:330–79. doi: 10.1067/j.cpsurg.2018.08.004
- Manfredi S, Lepage C, Hatem C, Coatmeur O, Faivre J, Bouvier A-M. Epidemiology and management of liver metastases from colorectal cancer. *Ann Surg* (2006) 244:254–9. doi: 10.1097/01.sla.0000217629.94941.cf
- Ding Q, Chang C-J, Xie X, Xia W, Yang J-Y, Wang S-C, et al. APOBEC3G promotes liver metastasis in an orthotopic mouse model of colorectal cancer and predicts human hepatic metastasis. *J Clin Invest* (2011) 121:4526–36. doi: 10.1172/JCI45008
- Nakayama I, Hirota T, Shinozaki E. BRAF mutation in colorectal cancers: From prognostic marker to targetable mutation. *Cancers (Basel)* (2020) 12:E3236. doi: 10.3390/cancers12113236
- Tsilimigras DI, Ntanasis-Stathopoulos I, Bagante F, Moris D, Cloyd J, Spartalis E, et al. Clinical significance and prognostic relevance of KRAS, BRAF, PI3K and TP53 genetic mutation analysis for resectable and unresectable colorectal liver metastases: A systematic review of the current evidence. *Surg Oncol* (2018) 27:280–8. doi: 10.1016/j.suronc.2018.05.012
- Kostner AH, Kersten C, Löwenmark T, Ydsten KA, Peltonen R, Isoniemi H, et al. The prognostic role of systemic inflammation in patients undergoing resection of colorectal liver metastases: C-reactive protein (CRP) is a strong negative prognostic biomarker. *J Surg Oncol* (2016) 114:895–9. doi: 10.1002/jso.24415
- Hao M, Wang K, Ding Y, Li H, Liu Y, Ding L. Which patients are prone to suffer liver metastasis? A review of risk factors of metachronous liver metastasis of colorectal cancer. *Eur J Med Res* (2022) 27:130. doi: 10.1186/s40001-022-00759-z
- Jones RP, Jackson R, Dunne DFJ, Malik HZ, Fenwick SW, Poston GJ, et al. Systematic review and meta-analysis of follow-up after hepatectomy for colorectal liver metastases. *Br J Surg* (2012) 99:477–86. doi: 10.1002/bjs.8667
- Yin Z, Liu C, Chen Y, Bai Y, Shang C, Yin R, et al. Timing of hepatectomy in resectable synchronous colorectal liver metastases (SCLM): Simultaneous or delayed? *Hepatology* (2013) 57:2346–57. doi: 10.1002/hep.26283
- Moertel CG. Multiple primary malignant neoplasms: historical perspectives. *Cancer* (1977) 40:1786–92. doi: 10.1002/1097-0142(197710)40:4+<1786::AID-CNCR2820400803>3.0.CO;2-2
- Ruers T, Punt C, Van Coevorden F, Pierie JPN, Borel-Rinkes I, Ledermann JA, et al. Radiofrequency ablation combined with systemic treatment versus systemic treatment alone in patients with non-resectable colorectal liver metastases: a randomized EORTC intergroup phase II study (EORTC 40004). *Ann Oncol* (2012) 23:2619–26. doi: 10.1093/annonc/mds053
- Edge SB, Compton CC. The American joint committee on cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol* (2010) 17:1471–4. doi: 10.1245/s10434-010-0985-4
- Mekenkamp LJM, Koopman M, Teerenstra S, van Krieken JHJM, Mol L, Nagtegaal ID, et al. Clinicopathological features and outcome in advanced colorectal cancer patients with synchronous vs metachronous metastases. *Br J Cancer* (2010) 103:159–64. doi: 10.1038/sj.bjc.6605737
- Siriwardena AK, Mason JM, Mullanitha S, Hancock HC, Jegatheeswaran S. Management of colorectal cancer presenting with synchronous liver metastases. *Nat Rev Clin Oncol* (2014) 11:446–59. doi: 10.1038/nrclinonc.2014.90
- Laubert T, Habermann JK, Hemmelmann C, Kleemann M, Oevermann E, Bouchard R, et al. Metachronous metastasis- and survival-analysis show prognostic importance of lymphadenectomy for colon carcinomas. *BMC Gastroenterol* (2012) 12:24. doi: 10.1186/1471-230X-12-24
- Engstrand J, Nilsson H, Strömberg C, Jonas E, Freedman J. Colorectal cancer liver metastases – a population-based study on incidence, management and survival. *BMC Cancer* (2018) 18:78. doi: 10.1186/s12885-017-3925-x
- Engstrand J, Strömberg C, Nilsson H, Freedman J, Jonas E. Synchronous and metachronous liver metastases in patients with colorectal cancer-towards a clinically relevant definition. *World J Surg Onc* (2019) 17:228. doi: 10.1186/s12957-019-1771-9
- Yi M, Jiao D, Xu H, Liu Q, Zhao W, Han X, et al. Biomarkers for predicting efficacy of PD-1/PD-L1 inhibitors. *Mol Cancer* (2018) 17:129. doi: 10.1186/s12943-018-0864-3
- Jungwirth J, Urbanova M, Boot A, Hosek P, Bendova P, Siskova A, et al. Mutational analysis of driver genes defines the colorectal adenoma: *in situ* carcinoma transition. *Sci Rep* (2022) 12:2570. doi: 10.1038/s41598-022-06498-9
- Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* (2016) 32:3047–8. doi: 10.1093/bioinformatics/btw354
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* (2014) 30:2114–20. doi: 10.1093/bioinformatics/btu170
- Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics* (2010) 26:589–95. doi: 10.1093/bioinformatics/btp698
- van der Auwera G, O'Connor BD. *Genomics in the cloud: using docker, GATK, and WDL in Terra*. 1st ed. Sebastopol, CA: O'Reilly Media (2020). p. 467.
- Riester M, Singh AP, Brannon AR, Yu K, Campbell CD, Chiang DY, et al. PureCN: copy number calling and SNV classification using targeted short read sequencing. *Source Code Biol Med* (2016) 11:13. doi: 10.1186/s13029-016-0060-z
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* (2012) 6:80–92. doi: 10.4161/fly.19695
- Zou B, Guo D, Kong P, Wang Y, Cheng X, Cui Y. Integrative genomic analyses of 1,145 patient samples reveal new biomarkers in esophageal squamous cell carcinoma. *Front Mol Biosci* (2022) 8:792779. doi: 10.3389/fmolb.2021.792779
- Honkova K, Rossnerova A, Chvojikova I, Milcova A, Margaryan H, Pastorkova A, et al. Genome-wide DNA methylation in policemen working in cities differing by major sources of air pollution. *IJMS* (2022) 23:1666. doi: 10.3390/ijms23031666

32. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive bioconductor package for the analysis of infinium DNA methylation microarrays. *Bioinformatics* (2014) 30:1363–9. doi: 10.1093/bioinformatics/btu049
33. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the illumina infinium MethylationEPIC BeadChip. *Genom Data* (2016) 9:22–4. doi: 10.1016/j.gdata.2016.05.012
34. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* (2012) 28:882–3. doi: 10.1093/bioinformatics/bts034
35. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43:e47. doi: 10.1093/nar/gkv007
36. Li D, Xie Z, Le Pape M, Dye T. An evaluation of statistical methods for DNA methylation microarray data analysis. *BMC Bioinf* (2015) 16:217. doi: 10.1186/s12859-015-0641-x
37. Cavalcante RG, Sartor MA. Annotatr: genomic regions in context. *Bioinformatics* (2017) 33:2381–3. doi: 10.1093/bioinformatics/btx183
38. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* (2021) 2:100141. doi: 10.1016/j.xinn.2021.100141
39. Yu G, He Q-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst* (2016) 12:477–9. doi: 10.1039/C5MB00663E
40. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* (2013) 29:15–21. doi: 10.1093/bioinformatics/bts635
41. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinf* (2011) 12:323. doi: 10.1186/1471-2105-12-323
42. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612
43. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* (2010) 26:139–40. doi: 10.1093/bioinformatics/btp616
44. Sabo AA, Birolo G, Naccarati A, Dragomir MP, Aneli S, Allione A, et al. Small non-coding RNA profiling in plasma extracellular vesicles of bladder cancer patients by next-generation sequencing: Expression levels of miR-126-3p and piR-5936 increase with higher histologic grades. *Cancers* (2020) 12:1507. doi: 10.3390/cancers12061507
45. Cervena K, Novosadova V, Pardini B, Naccarati A, Opattova A, Horak J, et al. Analysis of MicroRNA expression changes during the course of therapy in rectal cancer patients. *Front Oncol* (2021) 11:702258. doi: 10.3389/fonc.2021.702258
46. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* (2011) 17:10. doi: 10.14806/ej.17.1.200
47. Patil AH, Halushka MK. miRge3.0: a comprehensive microRNA and tRF sequencing analysis pipeline. *NAR Genomics Bioinf* (2021) 3:lqab068. doi: 10.1093/nargab/lqab068
48. Del Rio M, Molina F, Bascoul-Molleli C, Copois V, Bibeau F, Chablos P, et al. Gene expression signature in advanced colorectal cancer patients select drugs and response for the use of leucovorin, fluorouracil, and irinotecan. *J Clin Oncol* (2007) 25:773–80. doi: 10.1200/JCO.2006.07.4187
49. Del Rio M, Mollevi C, Vezzio-Vie N, Bibeau F, Ychou M, Martineau P. Specific extracellular matrix remodeling signature of colon hepatic metastases. *PLoS One* (2013) 8:e74599. doi: 10.1371/journal.pone.0074599
50. 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
51. Zhou H, Liu Z, Wang Y, Wen X, Amador EH, Yuan L, et al. Colorectal liver metastasis: molecular mechanism and interventional therapy. *Sig Trans Target Ther* (2022) 7:70. doi: 10.1038/s41392-022-00922-2
52. Vermeulen PB, Colpaert C, Salgado R, Royers R, Hellemans H, Van Den Heuvel E, et al. Liver metastases from colorectal adenocarcinomas grow in three patterns with different angiogenesis and desmoplasia. *J Pathol* (2001) 195:336–42. doi: 10.1002/path.966
53. Lee CS, Song IH, Lee A, Kang J, Lee YS, Lee IK, et al. Enhancing the landscape of colorectal cancer using targeted deep sequencing. *Sci Rep* (2021) 11:8154. doi: 10.1038/s41598-021-87486-3
54. Mamlouk S, Childs LH, Aust D, Heim D, Melching F, Oliveira C, et al. DNA Copy number changes define spatial patterns of heterogeneity in colorectal cancer. *Nat Commun* (2017) 8:14093. doi: 10.1038/ncomms14093
55. Ried T, Meijer GA, Harrison DJ, Grech G, Franch-Expósito S, Briffa R, et al. The landscape of genomic copy number alterations in colorectal cancer and their consequences on gene expression levels and disease outcome. *Mol Aspects Med* (2019) 69:48–61. doi: 10.1016/j.mam.2019.07.007
56. Saidijam M, Saidijam M, Taherikalani M, Shabab N, Ghanbari R, Afshar S, et al. Serum and stool miR-135b levels as a potential diagnostic biomarker for colorectal cancer. *CEI* (2020) 1–6. doi: 10.31487/j.CEI.2020.01.04
57. Toiyama Y, Hur K, Tanaka K, Inoue Y, Kusunoki M, Boland CR, et al. Serum miR-200c is a novel prognostic and metastasis-predictive biomarker in patients with colorectal cancer. *Ann Surg* (2014) 259:735–43. doi: 10.1097/SLA.0b013e3182a6909d
58. Xu Y, Zhong Y-D, Zhao X-X. MiR-548b suppresses proliferative capacity of colorectal cancer by binding WNT2. *Eur Rev Med Pharmacol Sci* (2020) 24:10535–41. doi: 10.26355/eurrev_202010_23406
59. Schatoff EM, Leach BI, Dow LE. WNT signaling and colorectal cancer. *Curr Colorectal Cancer Rep* (2017) 13:101–10. doi: 10.1007/s11888-017-0354-9
60. Chook Y. Karyopherins and nuclear import. *Curr Opin Struct Biol* (2001) 11:703–15. doi: 10.1016/S0959-440X(01)00264-0
61. Çağatay T, Chook YM. Karyopherins in cancer. *Curr Opin Cell Biol* (2018) 52:30–42. doi: 10.1016/j.jceb.2018.01.006
62. Kau TR, Way JC, Silver PA. Nuclear transport and cancer: from mechanism to intervention. *Nat Rev Cancer* (2004) 4:106–17. doi: 10.1038/nrc1274
63. Tran EJ, King MC, Corbett AH. Macromolecular transport between the nucleus and the cytoplasm: Advances in mechanism and emerging links to disease. *Biochim Biophys Acta (BBA) - Mol Cell Res* (2014) 1843:2784–95. doi: 10.1016/j.bbamer.2014.08.003
64. Kosyna F, Depping R. Controlling the gatekeeper: Therapeutic targeting of nuclear transport. *Cells* (2018) 7:221. doi: 10.3390/cells7110221
65. Zhang W, Lu Y, Li X, Zhang J, Lin W, Zhang W, et al. IPO5 promotes the proliferation and tumorigenicity of colorectal cancer cells by mediating RASAL2 nuclear transportation. *J Exp Clin Cancer Res* (2019) 38:296. doi: 10.1186/s13046-019-1290-0
66. Li X-F, Aierken A-L-D, Shen L. IPO5 promotes malignant progression of esophageal cancer through activating MMP7. *Eur Rev Med Pharmacol Sci* (2020) 24:4246–54. doi: 10.26355/eurrev_202004_21004
67. Li M, Li X, Chen S, Zhang T, Song L, Pei J, et al. IPO5 mediates EMT and promotes esophageal cancer development through the RAS-ERK pathway. *Oxid Med Cell Longevity* (2022) 2022:1–17. doi: 10.1155/2022/8999899
68. Chao H-W, Lai Y-T, Lu Y-L, Lin C, Mai W, Huang Y-S. NMDAR signaling facilitates the IPO5-mediated nuclear import of CPEB3. *Nucleic Acids Res* (2012) 40:8484–98. doi: 10.1093/nar/gks598
69. Ağaçgündüz D, Coccoza E, Cemali Ö, Bayazit AD, Nani MF, Cerqua I, et al. Understanding the role of the gut microbiome in gastrointestinal cancer: A review. *Front Pharmacol* (2023) 14:1130562. doi: 10.3389/fphar.2023.1130562
70. Settmacher U, Dittmar Y, Knösel T, Schöne U, Heise M, Jandt K, et al. Predictors of long-term survival in patients with colorectal liver metastases: a single center study and review of the literature. *Int J Colorectal Dis* (2011) 26:967–81. doi: 10.1007/s00384-011-1195-7
71. Spelt L, Andersson B, Nilsson J, Andersson R. Prognostic models for outcome following liver resection for colorectal cancer metastases: A systematic review. *Eur J Surg Oncol* (2012) 38:16–24. doi: 10.1016/j.ejso.2011.10.013
72. Colloca GA, Venturino A, Guarneri D. Different variables predict the outcome of patients with synchronous versus metachronous metastases of colorectal cancer. *Clin Transl Oncol* (2020) 22:1399–406. doi: 10.1007/s12094-019-02277-7
73. Thomas A, Routh ED, Pullikuth A, Jin G, Su J, Chou JW, et al. Tumor mutational burden is a determinant of immune-mediated survival in breast cancer. *Oncimmunology* (2018) 7:e1490854. doi: 10.1080/2162402X.2018.1490854
74. Jang BG, Kim HS, Bae JM, Kim WH, Kim HU, Kang GH. SMOC2, an intestinal stem cell marker, is an independent prognostic marker associated with better survival in colorectal cancers. *Sci Rep* (2020) 10:14591. doi: 10.1038/s41598-020-71643-1
75. Nakabayashi K. Genomic imprinting of PPP1R9A encoding neurabin I in skeletal muscle and extra-embryonic tissues. *J Med Genet* (2004) 41:601–8. doi: 10.1136/jmg.2003.014142
76. McAvoy T, Allen PB, Obaishi H, Nakanishi H, Takai Y, Greengard P, et al. Regulation of neurabin I interaction with protein phosphatase 1 by phosphorylation. *Biochemistry* (1999) 38:12943–9. doi: 10.1021/bi991227d
77. Terry-Lorenzo RT, Carmody LC, Voltz JW, Connor JH, Li S, Smith FD, et al. The neuronal actin-binding proteins, Neurabin I and Neurabin II, Recruit specific isoforms of protein phosphatase-1 catalytic subunits. *J Biol Chem* (2002) 277:27716–24. doi: 10.1074/jbc.M203365200
78. Verdugo-Sivianes EM, Carnero A. SPINOPHILIN: A multiplayer tumor suppressor. *Genes Dis* (2022) 10(1):187–98. doi: 10.1016/j.gendis.2021.12.021
79. Estevez-Garcia P, Lopez-Calderero I, Molina-Pinelo S, Muñoz-Galvan S, Salinas A, Gomez-Izquierdo L, et al. Spinophilin loss correlates with poor patient prognosis in advanced stages of colon carcinoma. *Clin Cancer Res* (2013) 19:3925–35. doi: 10.1158/1078-0432.CCR-13-0057
80. Ferrer I, Verdugo-Sivianes EM, Castilla MA, Melendez R, Marin JJ, Muñoz-Galvan S, et al. Loss of the tumor suppressor spinophilin (PPP1R9B) increases the cancer stem cell population in breast tumors. *Oncogene* (2016) 35:2777–88. doi: 10.1038/onc.2015.341
81. Rasmussen MH, Lyskjær I, Jersie-Christensen RR, Tarpgaard LS, Primdal-Bengtson B, Nielsen MM, et al. miR-625-3p regulates oxaliplatin resistance by targeting MAP2K6-p38 signalling in human colorectal adenocarcinoma cells. *Nat Commun* (2016) 7:12436. doi: 10.1038/ncomms12436
82. Bu P, Wang L, Chen K-Y, Rakhilin N, Sun J, Closa A, et al. miR-1269 promotes metastasis and forms a positive feedback loop with TGF-β. *Nat Commun* (2015) 6:6879. doi: 10.1038/ncomms7879



OPEN ACCESS

EDITED BY

Matjaz Rokavec,
Ludwig Maximilian University of Munich,
Germany

REVIEWED BY

Roxana Pincheira,
University of Concepcion, Chile
Youtao Lu,
University of Pennsylvania, United States

*CORRESPONDENCE

Eva Turyova
✉ turyova10@uniba.sk
Zora Lasabova
✉ zora.lasabova@uniba.sk

RECEIVED 15 February 2023

ACCEPTED 02 May 2023

PUBLISHED 20 June 2023

CITATION

Turyova E, Mikolajcik P, Grendar M,
Kudelova E, Holubekova V, Kalman M,
Marcinek J, Hrneciar M, Kovac M,
Miklusica J, Laca L and Lasabova Z (2023)
Expression of *OCT4* isoforms is reduced in
primary colorectal cancer.
Front. Oncol. 13:1166835.
doi: 10.3389/fonc.2023.1166835

COPYRIGHT

© 2023 Turyova, Mikolajcik, Grendar,
Kudelova, Holubekova, Kalman, Marcinek,
Hrneciar, Kovac, Miklusica, Laca and Lasabova.
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.

Expression of *OCT4* isoforms is reduced in primary colorectal cancer

Eva Turyova^{1*}, Peter Mikolajcik², Marian Grendar³,
Eva Kudelova², Veronika Holubekova³, Michal Kalman⁴,
Juraj Marcinek⁴, Matej Hrneciar⁵, Michal Kovac⁵,
Juraj Miklusica², Ludovit Laca² and Zora Lasabova^{1*}

¹Department of Molecular Biology and Genomics, Jessenius Faculty of Medicine in Martin, Comenius University Bratislava, Martin, Slovakia, ²Clinic of Surgery and Transplant Center, Jessenius Faculty of Medicine in Martin and University Hospital Martin, Comenius University Bratislava, Martin, Slovakia, ³Biomedical Centre Martin, Jessenius Faculty of Medicine in Martin, Comenius University Bratislava, Martin, Slovakia, ⁴Department of Pathological Anatomy, Jessenius Faculty of Medicine in Martin and University Hospital Martin, Comenius University Bratislava, Martin, Slovakia, ⁵Department of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies, Slovak University of Technology, Bratislava, Slovakia

Introduction: Colorectal cancer (CRC) is one of the most common types of cancer worldwide. The carcinogenesis of CRC is indeed complex, and there are many different mechanisms and pathways that contribute to the development of malignancy and the progression from primary to metastatic tumors. The *OCT4A*, encoded by the *POU5F1* gene, is a transcription factor responsible for the phenotype of stem cells, maintaining pluripotency and regulation of differentiation. The *POU5F1* gene is made up of five exons that can create numerous isoforms through alternative promoter or alternative splicing. In addition to *OCT4A*, other isoforms called *OCT4B* are also translated into protein; however, their role in cells has been unclear. The aim of our work was to investigate the expression patterns of *OCT4* isoforms in primary and metastatic CRC, providing us with useful information about their role in the development and progression of CRC.

Methods: Surgical specimens from a total of 78 patients were collected and isolated from primary tumors ($n = 47$) and metastases ($n = 31$). The relative gene expression of *OCT4* isoforms was investigated using the RT-qPCR method together with the TaqMan probes for particular *OCT4* isoforms.

Results: Our results suggest significantly downregulated expression of the *OCT4A* and *OCT4Bs* isoforms in both primary ($p = 0.0002$ and $p < 0.0001$, respectively) and metastatic tumors ($p = 0.0006$ and $p = 0.00051$, respectively) when compared with the control samples. We also observed a correlation between reduced expression of all *OCT4* isoforms and both primary and left-sided tumors ($p = 0.001$ and $p = 0.030$, respectively). On the other hand, the expression of all *OCT4* isoforms was significantly upregulated in metastases compared with primary tumors ($p < 0.0001$).

Discussion: Unlike previous reports, we found out that the expression of *OCT4A*, *OCT4Bs*, and all *OCT4* isoforms was significantly reduced in primary tumors and metastases compared with control samples. On the other hand, we supposed that the expression rate of all *OCT4* isoforms may be related to the cancer type and side, as well as to liver metastases. However, further studies are required to investigate the detailed expression patterns and significance of individual *OCT4* isoforms in carcinogenesis.

KEYWORDS

colorectal cancer, gene expression, oct4, isoforms, carcinogenesis

1 Introduction

Colorectal cancer (CRC) is one of the most common types of cancer worldwide. The global statistics from 2020 show that colorectal cancer creates 10% of all newly diagnosed cases, which means that after breast and lung cancer, it is the third most common type (1). Regardless of the enormous interest in CRC research, the annual incidence and number of CRC-related deaths have been increasing worldwide. The incidence of CRC is higher in developed countries due to unhealthy lifestyle including high consumption of red meat and alcohol, smoking, sedentary lifestyle, and inflammatory bowel diseases (2–4).

The octamer binding transcription factor 4 (*OCT4*) isoforms are encoded by the *POU5F1* gene which is located at the short arm of chromosome 6 (5). Proteins from the POU protein family contain the so-called POU domains which allow them to bind to DNA and influence the gene expression, as well as interact with other transcription factors and cofactors. The POU domain recognizes and binds to the octamer consensus DNA sequence ATGCAAT, and in this way, the *OCT4* can regulate gene expression, making it the main regulator of maintaining pluripotency and self-renewal of the stem cells (6–8). The human *OCT4* gene can, through alternative transcription initiation or alternative splicing, create numerous different isoforms. These isoforms differ not only in nucleotide sequence but also in subcellular localization and properties. At the RNA level, *OCT4* creates four groups of variants, namely, *OCT4A*, *OCT4B*, *OCT4C*, and *OCT4D*, and each of them uses a unique transcription start site. Furthermore, the mechanism of alternative splicing is responsible for the emergence of other *OCT4B* and *OCT4C* isoforms (9, 10). The discovery of individual transcripts and isoforms was gradual, and to date, overall, 10 *OCT4* transcripts have been identified (9, 11–16). Not all transcripts have been identified also at the protein level, and there are only assumptions that the length of the protein product would be 164 amino acids (16). On the other hand, there is unequivocal evidence that the *OCT4A* and *OCT4B* variants generate distinct protein products that differ in their properties. The *OCT4A* isoform encodes the longest protein composed of 360 amino acids, which fulfills the role of the transcription factor. In contrast, the *OCT4B* transcript may be through the mechanism of alternative translation translated into three different proteins with

lengths of 265, 190, and 164 amino acids (17). *OCT4B* proteins differ from *OCT4A* in DNA-binding properties; as a consequence of the inability of *OCT4B* proteins to regulate gene expression, their role in the cell remains unknown (18).

The *OCT4A* is the most known isoform, and its correct and precise expression is necessary for the regulation of the expression of further genes, as well as for the repression of genes involved in differentiation (18). The *OCT4A* protein plays a pivotal role in pluripotency and its expression was confirmed in embryonic stem (ES) and embryonic cancer (EC) cells, tumor cells, and cell lines (19, 20), but its presence has not yet been detected in non-pluripotent cell types (11, 18). The definite role of *OCT4A* in pluripotent cells was clarified by Takahashi and Yamanaka who proved that *OCT4* is one of the four transcription factors needed for the reprogramming of somatic cells to pluripotent cells (12, 13, 21). In contrast to *OCT4A*, which is located exclusively in the nucleus, the *OCT4B* isoform is diffusely localized in both the cytoplasm and nucleus (17, 22). Atlasi et al. demonstrated that *OCT4B* is expressed in almost all cell types tested, including ES, EC, and somatic cell types (12). Interestingly, another identified *OCT4B* isoform, called *OCT4B1*, has properties and expression patterns more similar to the *OCT4A* than to the *OCT4B* isoform.

The role of *POU5F1*, especially the *OCT4A* isoform in stem cells, emphasizes its potential role in carcinogenesis. The so-called theory of stem cells in carcinogenesis claims that cancer stem cells are capable of inducing tumorigenesis and tumor growth due to their self-renewal ability, as well as giving rise to additional progenitor cells (23). Due to the great diversity of *OCT4* isoforms, little is known about their expression patterns and role in primary and metastatic colorectal cancer. In our study, we focused on determining the expression patterns of *OCT4A*, *OCT4Bs*, and all *OCT4* isoforms in primary and metastatic colorectal cancer, as well as the correlation between their expression and clinical parameters.

2 Methods

2.1 Clinical tissue samples

In total, 78 CRC patients were registered and underwent CRC resection. Primary tumor samples as well as metastatic

samples were collected in collaboration with the University Hospital Martin, the Clinic of General, Visceral and Transplant Surgery, and the Department of Pathological Anatomy (Martin, Slovakia). Control, adjacent non-tumorous samples were obtained from patients with primary tumor ($n = 30$). The inclusion criteria were confirmation of diagnosis by histopathological examination, TNM classification, and clinical stages I, II, and III for primary tumor samples ($n = 47$) and confirmation of diagnosis by histopathological examination, TNM classification, and clinical stage IV for samples of liver metastases ($n = 31$). The pathologist also decided to take a non-cancer control sample from the adjacent tissue as far as possible from the tumor. The exclusion criteria were age less than 40 years or the simultaneous presence of other cancer types or liver metastasis of unknown origin. The histopathological assessment such as staging, grading, and typing of tumors was done by experienced pathologists (MK and JM). The clinicopathological characteristics of the patients included in the study are listed in Table 1. Tumor surgical specimens were collected by the pathologists and on the same day embedded into a solution of Dulbecco's modified Eagle's medium (DMEM), fetal bovine serum (10%), and penicillin/streptomycin and stored at 4°

C. Immediately, the samples were frozen in RNAlater and then transferred to the Department of Molecular Biology and Genomics for RNA isolation.

2.2 RNA purification and cDNA preparation

Total RNA was purified using the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions, treated with DNase, and stored at -80°C . Total RNA was quantified using the Qubit 3.0 Fluorometer (Thermo Fisher, MA, USA), and RNA quality assessment was performed on Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, USA). Reverse transcription was performed for each sample using a High-Capacity cDNA Reverse Transcription Kit with RNase inhibitor (Thermo Fisher, MA, USA) according to the manufacturer's instructions in a total volume of 20 μl consisting of 2 μl of 10 \times RT buffer, 0.8 μl of 2 \times dNTP mix (100 mM), 2 μl of 10 \times RT random primers, 1 μl of MultiScribeTM Reverse Transcriptase, 1 μl of RNase inhibitor, 3.2 μl of nuclease-free water, and 10 μl of RNA diluted in nuclease-free water to an overall concentration of 500 ng/ μl . Reverse transcription was performed using the Bio-Rad MJ Mini Personal Thermal Cycler (Bio-Rad, Hercules, CA, USA), and the thermal conditions were as follows: 25°C for 10 min, followed by 37°C for 120 min and 85°C for 5 min. The prepared cDNA was stored at -20°C .

2.3 Relative quantification of the gene expression

For the RT-PCR analysis, 10 μl of Gene Expression Master Mix was mixed with 1 μl of *POU5F1* TaqMan assay, 1 μl of *GAPDH* TaqMan assay, and 1 μl of cDNA. The total volume of 20 μl per reaction was supplemented with nuclease-free water. Expression analysis was performed by real-time quantitative PCR (qPCR) using LightCycler ABI 3500 (Applied Biosystems, Foster City, CA, USA), TaqMan Gene Expression Master Mix, and several TaqMan assays (Thermo Fisher, MA, USA) for specific *OCT4* isoforms and endogenous control. We determined the expression of the *OCT4A* isoform, *OCT4B* isoforms, and all *OCT4* isoforms together using three different TaqMan assays (FAM-MGB) (assay ID Hs01654807_s1, Hs04195369_s1, and Hs04260367_gH, respectively). Individual assays were chosen based on genomic map analysis (Supplementary Figure 1) according to the manufacturer, as well as based on the relevant literature. The chosen assays have already been validated and used in several articles, for instance, Hs04260367_gH (24, 25), Hs01654807_s1 (26, 27), and Hs04195369_s1 (28, 29). The assay used for the detection of the *OCT4A* isoform did not recognize other *OCT4* isoforms due to its exon 1 specificity. The same principle was used in the experiment with assay specific for almost all *OCT4B* isoforms. Some of the *OCT4B* and *OCT4A* isoforms were not recognized by this assay due to the absence of a certain exon. The last assay was used to determine all *OCT4* isoforms that share the same last exon (Supplementary Figure 1). Glyceraldehyde-3-

TABLE 1 Clinicopathological characteristics of the patients included in the study.

	Primary tumor samples	Liver metastases samples
Patients	47	31
Females	20	8
Males	27	23
Average age	66.7 (SD \pm 11.2)	67.6 (SD \pm 9.1)
BMI	27.6 (SD \pm 5.3)	28.1 (SD \pm 4.9)
Grade		
G1	13	1
G2	21	16
G3	9	3
G n.a.	4	11
T stage		
T1	2	1
T2	10	5
T3	25	20
T4	10	2
T n.a.	0	3
N stage		
N0	28	7
N1	13	14
N2	6	6
N n.a.	0	4
Clinical stage		
I	9	0
II	20	0
III	14	0
IV	4	30

n.a., not available.

phosphate dehydrogenase (GAPDH) was used as an endogenous control, and its expression was determined by TaqMan assay (VIC-MGB, assay ID Hs99999905_m1), too. As a reference sample, we used the Total Human RNA Control (Thermo Fisher, MA, USA; Cat. number 4307281) which was also reverse-transcribed into cDNA. The thermal conditions of the reactions were as follows: 50°C for 2 min, 95°C for 10 min, followed by 40 cycles at 95°C for 15 s and 60°C for 60 s. All reactions were performed in duplicate. Contamination was controlled by using a no-template control without adding cDNA as a template molecule. The relative quantification of *OCT4* isoform expression in the primary and metastatic tumor samples but also in adjacent non-tumorous tissue was performed by the $\Delta\Delta C_t$ method (30). We calculated the relative expression of *OCT4A*, *OCT4B*, and all *OCT4* isoforms against the GAPDH expression separately and obtained a fold-change value ($\log_2(FC)$). For statistical analysis, $\log_2(FC)$ values were used.

2.4 Statistical analysis

Data were explored and analyzed in collaboration with the Biomedical Centre Martin in R ver. 4.0.5, with the aid of different libraries (31–41). Data were summarized as the mean, SD, min, quartiles, and max. Boxplot overlaid with a swarm plot and quantile–quantile plot with the 95% confidence band constructed by bootstrap were used to assess the normality of data. Welch's *t*-test was used to test the null hypothesis that the population mean of the $\log_2(FC)$ is 0. The regression model was used to model the association between $\log_2(FC)$ and clinical data. Using the Wilkinson–Rogers notation, the full model that we used can be written as $\log_2FC \sim \text{type} + \text{side} + T + N + M + \text{stage}$, where type is a factor with levels *p* (for primary tumor) and *t* (for metastasis); side is a 0/1 factor where the left side is coded as 1; *T*, *N*, and *M* are the factors, describing the amount and spread of cancer; and stage is the pathological stage. The Akaike information criterion (AIC) was used for model selection. The model selected by the AIC was subjected to standard diagnostic analyses. Adjusted R^2 was used to measure the effect size. Marginal means and *post hoc* pairwise comparisons were performed where the AIC-selected model contained other predictors aside from the intercept. The *post hoc* *p*-values were adjusted by the Tukey method.

2.5 TCGA data

The STAR-processed RNA sequencing level 1 data were downloaded from the GDC Portal of the NIH National Cancer Institute. Files were generated within the projects TCGA-COAD, TCGA-READ, CPTAC-2, and HCCMI-CMDC. The set contained tsv files of 737 samples from 616 individuals, of which 615 were tumor samples, 45 were labeled as normal tissue, and 11 were labeled as metastases. Sixty-six samples with no information of their origin were omitted from further analysis. Given that all files were open access level 1 data, no preprocessing was carried out.

3 Results

In our study, we focused on determining the expression of different *OCT4* isoforms in primary and metastatic colorectal cancer samples. We investigated the expression of the *OCT4A* isoform, *OCT4B* isoforms, and all *OCT4* isoforms together, as well as the relationship between isoform expression in certain sample types and clinical data. A total of 78 cancer samples (47 primary cancer samples and 31 liver metastases samples) were analyzed in duplicates for each isoform (*A*, *B*, and all isoforms). Control samples (*n* = 30) represented adjacent non-tumor tissues of the intestinal epithelium.

3.1 Relative quantification of the gene expression

Fold-change values were calculated based on qPCR results for individual isoforms in specific sample types. At first, we compared the expression of the *OCT4* isoform in all obtained tumor samples (primary + metastatic) vs. control (adjacent non-tumor tissue). The expression of all tested isoforms (namely, *OCT4A*, *OCT4B*, and all *OCT4* isoforms) was significantly downregulated in tumor tissues compared with controls. The $\log_2(FC)$ values for the *OCT4A* and *OCT4B* isoforms were -1.25 and -1.53 , respectively, which refers to significantly reduced expression in tumor samples compared with controls ($p < 0.0001$ and $p < 0.0001$, respectively). In the case of all *OCT4* isoforms, the $\log_2(FC)$ was -1.03 and the expression was lower in tumors compared with control samples ($p < 0.0001$) (Figure 1). All data are listed in Supplementary Table 1.

After getting these results, we tried to determine whether the primary or metastatic samples were responsible for such reduced expression of individual isoforms in tumor samples or if it was a result of a mutual effect. Thus, subsequently, we determined the changes in the expression in primary and metastatic tumors compared with the control separately. In primary tumors, we also observed significantly reduced expression of all tested isoforms. For the *OCT4A* isoform, the $\log_2(FC)$ value was -1.08 , referring to reduced expression in primary tumors compared with control samples ($p = 0.0002$). The expression of *OCT4B* isoforms was the lowest in primary tumors, and compared with the control samples, the $\log_2(FC)$ value was -1.69 ($p < 0.0001$). Furthermore, we obtained statistically significant results also for the remaining isoforms. For all *OCT4* isoforms, the $\log_2(FC)$ value was -1.55 , which means decreased expression in primary tumors compared with controls ($p < 0.0001$) (Figure 2).

Similar results were also obtained in metastatic samples; in other words, the expression of *OCT4A* and *OCT4B* isoforms was significantly downregulated in tumor samples compared with controls. Both the *OCT4A* and *OCT4B* isoforms had statistically significant results. The expression of *OCT4A* was reduced ($\log_2(FC) = -1.53$) in liver metastasis compared with control tissues ($p = 0.0006$). For *OCT4B*, the $\log_2(FC)$ value was -1.35 , indicating a lower expression in metastasis ($p = 0.00051$). On the other hand, although the $\log_2(FC)$ for all *OCT4* isoforms was -0.2 and indicated

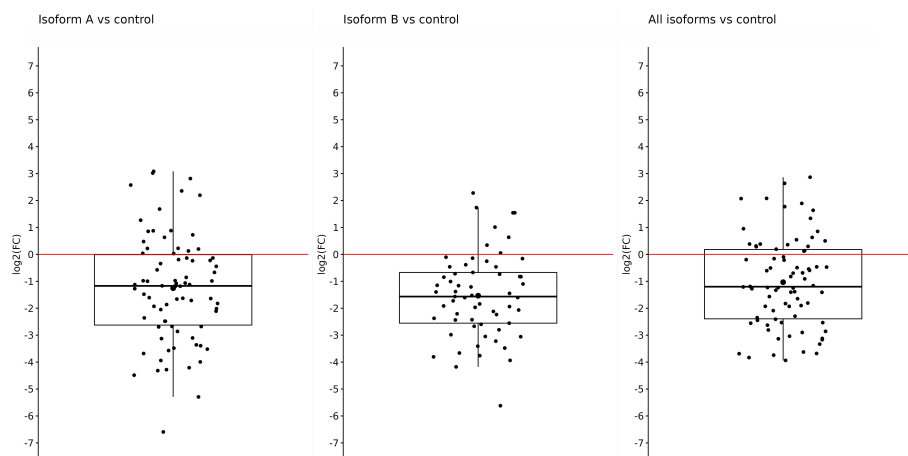


FIGURE 1

Expression of *OCT4A*, *OCT4B*, and all *OCT4* isoforms in tumor samples (primary + metastatic) compared with control samples. On the y-axis, we see $\log_2(\text{FC})$ values for *OCT4A* (−1.25), *OCT4B* isoforms (−1.53), and all *OCT4* isoforms (−1.03), which means a significantly reduced expression of all mentioned isoforms in primary tumors compared with control.

reduced expression in metastases compared with controls, this difference was not enough to be statistically significant ($p = 0.5$) (Figure 3).

After that, we did an expression comparison in primary and metastatic tumors. Surprisingly, the expression of *OCT4A* and *OCT4B* was not significantly different in metastases and primary tumors, and $\log_2(\text{FC})$ −0.44 and 0.34 did not show statistically significant results ($p = 0.27$ and $p = 0.33$, respectively). Nonetheless, the expression of *OCT4A* was lower in metastases compared with primary tumors. On the contrary, *OCT4B* isoforms were overexpressed in liver metastases. We obtained the only statistically significant result in the case of all *OCT4* isoforms which were significantly overexpressed in metastases compared with primary tumors ($\log_2(\text{FC}) = 1.357$; $p < 0.0001$) (Figure 4).

So, in the case of *OCT4A*, the most downregulated expression was observed in metastatic tumors. Regarding the *OCT4B* isoforms, the most downregulated expression was also detected in metastases, too. In the case of all *OCT4* isoforms, when comparing metastases and primary tumors, the expression was significantly upregulated in metastatic samples. Data are available on [data-mendeley.com/datasets \(42\)](https://data.mendeley.com/datasets/42).

3.2 Regression analysis

We further evaluated the relationship between the isoform expression and several clinical parameters, for instance, cancer type, tumor side, clinical, and TNM stage. Comparison of the

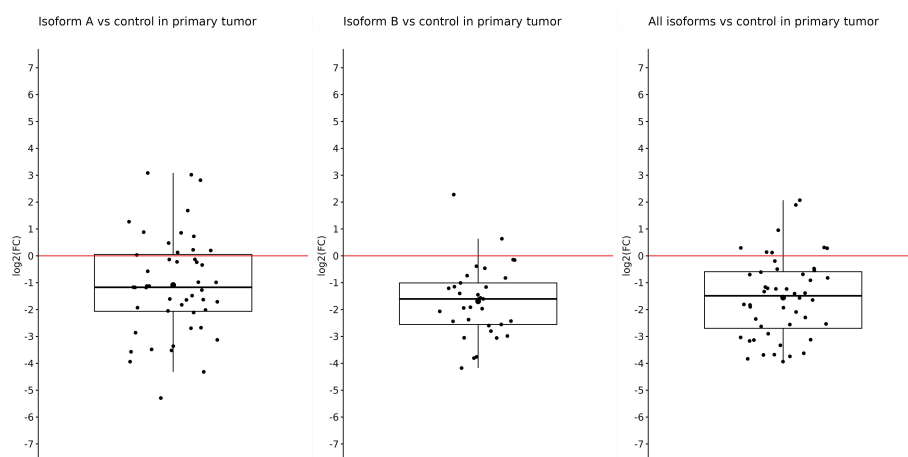


FIGURE 2

Expression of *OCT4A*, *OCT4B*, and all *OCT4* isoforms in primary tumors compared with control samples. In the artwork, we can see the significantly downregulated expression of all tested isoforms in primary tumors compared with control samples. The $\log_2(\text{FC})$ values are −1.08 for *OCT4A*, −1.69 for *OCT4B*, and −1.55 for all *OCT* isoforms.

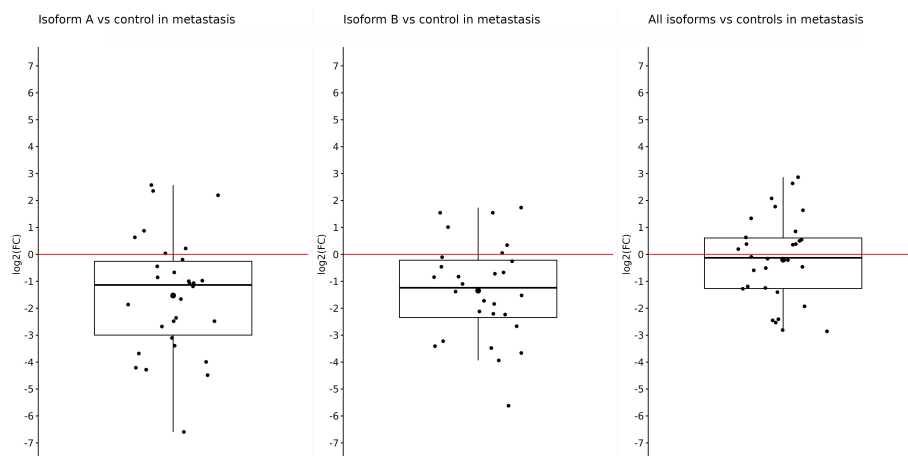


FIGURE 3

Expression of *OCT4A*, *OCT4B*, and all *OCT4* isoforms in metastases compared with control samples. When comparing metastases with controls, a significantly reduced expression was observed for only *OCT4A* ($\log_2(\text{FC}) = -1.53$) and *OCT4B* isoforms ($\log_2(\text{FC}) = -1.35$). For all *OCT4* isoforms, the $\log_2(\text{FC})$ value was the lowest (-0.2) and without statistical significance.

tumor samples (primary + metastatic) with the control showed a correlation between significantly downregulated expression of all *OCT4* isoforms and primary tumors ($p = 0.001$). Regression analysis also revealed the relationship between significantly reduced expression of all *OCT4* isoforms and left-sided tumors ($p = 0.030$). Other correlations regarding the expression of certain isoforms and the type or side of tumors were not observed.

3.3 TCGA results

We have downloaded the STAR-processed RNA sequencing level 1 data of 671 samples (colorectal tumors: $n = 615$, metastases: $n = 11$, unpaired samples from normal tissue: $n = 45$) from the Cancer

Genome Atlas Consortium Portal and used them to extract transcripts per kilobase million (TPM) counts for *POU5F1* (Ensemble ID ENSG00000204531.20). *POU* class 5 homeobox 1 has seven splice variants, which unfortunately are not discriminated by level 1 data. The level of gene expression in normal, tumor, and metastatic tissues is summarized in [Supplementary Table 3](#) and [Supplementary Figure 2](#). The TCGA results ([Supplementary Table 3](#)) indicate a rise in the expression of the *POU5F1* gene from non-tumor samples to primary tumors and subsequently to metastases, although without statistical significance. Given the low number of observations for two out of three groups and the skewness toward higher TPM values, it is inconclusive at any significance level to say that there is an increased expression from normal to primary tumors and to metastases, although there might

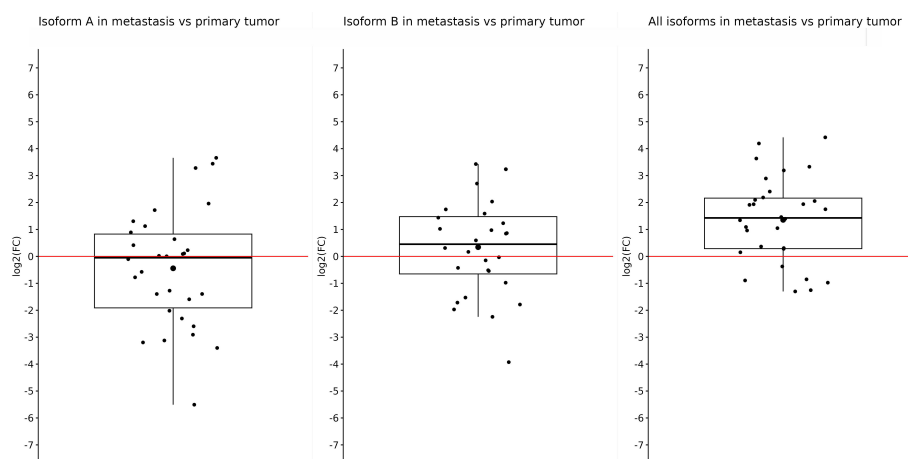


FIGURE 4

Expression of *OCT4A*, *OCT4B*, and all *OCT4* isoforms in metastases compared with primary tumor samples. Comparison of the expression in metastases and primary tumors proved only one statistically significant result, which showed us that the expression of all *OCT4* isoforms is significantly upregulated in metastasis ($\log_2(\text{FC}) = 1.357$). For the *OCT4A* and *OCT4B* isoforms, there were slight downregulation and upregulation ($\log_2(\text{FC}) = -0.44$ and 0.34 , respectively).

be some trend in this direction. Welch's *t*-test was used for pairwise testing of the null hypothesis that the population means of TPM in the normal, tumor, and metastatic groups are the same. Contrary to our previous expectations, in no tested combination did we reach the statistical significance at the level of <0.05 (normal vs. tumor: $p = 0.17$; normal vs. metastatic: $p = 0.05$, tumor vs. metastatic tissue: $p = 0.06$), probably owing to the high number of outliers.

4 Discussion

It is well known that the *OCT4A* isoform is expressed in numerous cell types including ES, EC, and cancer cell lines. Based on very little evidence about its expression in adult tissues (12, 43) and the study of Feldman et al., who found out that *OCT4A* transcription is turned off during gastrulation due to its promoter and enhancer methylation (44), it was generally thought that *OCT4A* is not expressed in adult somatic tissues. The *OCT4A* was shown to be an irreplaceable factor for stem-like cell phenotype, maintaining pluripotency and self-renewal of the stem cells and also for precise embryonic development, making it one of the most popular transcription factors ever (45–48). At present, a large number of articles demonstrate the presence of *OCT4A* in cancer stem cells (CSCs), and it is thought that *OCT4A* is responsible for stem-like cell properties of the cancer cells such as self-renewal, resistance, and the possibility to give rise to progenitor cells as well as epithelial–mesenchymal transition (19, 49). These assumptions make it a suitable target for treatment (50). In the past, it was proven that downregulation of *OCT4* expression resulted in inhibited tumorigenesis, reduced drug resistance, and induced G2/M phase arrest (51). Furthermore, *OCT4* appears to play a role in the angiogenesis and conversion of human fibroblasts to functional endothelial cells (52–54), and its expression was confirmed in all 13 CRC cell lines established from patients with both primary and metastatic tumors (55). Altogether, it appears that *OCT4A* contributes to tumor initiation, cancer growth, metastasis, and therapy resistance (56).

To date, 10 different *OCT4* isoforms have been identified at the RNA level, but not all are translated into protein (15). At the protein level, we can distinguish *OCT4A* and *OCT4B* isoforms which differ in their exon composition, nucleotide sequence, cell localization, and properties (11) (Supplementary Table 2). Due to the inability to bind to DNA, the *OCT4B* isoform is unable to regulate gene expression, and therefore, the role of *OCT4B* isoforms in cells remains unknown and unclear (18). Even though there are numerous *OCT4B* transcripts, the existence of only three protein isoforms has been confirmed (57). *OCT4B* proteins have been shown to play a role in stress response in two different ways. *OCT4B*-190 protects cells against apoptosis after heat shock (17), and in contrast, *OCT4B*-265 promotes apoptosis in reaction to genotoxic stress through the p53 signaling pathway (58).

In our study, we demonstrated an expression pattern of different *OCT4* isoforms in different sample types, as well as the relationship between isoform expression and several clinical parameters.

Comparing the expression of all determined isoforms (A, B, and all isoforms) in the control and tumor samples, regardless of the

type, we observed significant overexpression in the control samples (Figures 1–3). Distinct results were presented by Liu et al. (59) who found out that *OCT4* was overexpressed in tumor tissue compared with their matched normal counterparts of CRC. However, the authors did not distinguish primary and metastatic samples nor the expression of individual isoforms because they used primers specific for *OCT4A* as well as *OCT4B* isoforms (59). On the other hand, similar results for the expression of all *OCT4* isoforms, such as reduced gene expression in tumor tissue compared with the control and regardless of the isoforms, were also obtained in breast cancer (60). Aside from the other isoforms, *OCT4A* also had a higher expression in the control samples. On the other hand, the most reduced expression was detected in metastases compared with control samples, and the lowest difference was observed in the comparison of primary tumors with control tissues. In earlier investigations, devoted to *OCT4A* expression, the authors pointed out the possible distortion of the results due to the existence of numerous *OCT4* pseudogenes, so we should also consider the fact that our expression data may be influenced by pseudogenes expression (61–63). To date, eight *OCT4* pseudogenes have been identified and the transcription of these pseudogenes can have a confusing effect on research and knowledge of the *OCT4* gene expression (64). On the contrary, Saha et al. published results in which the expression of several *OCT4* pseudogenes has the same trend compared with the *OCT4* expression which indicates that pseudogene expression should not have an impact on the overall direction of expression (60). It could be precisely pseudogenes that are responsible for such high *OCT4A* expression, but even though our results indicate that *OCT4A* expression is the highest in the control samples, our raw expression data demonstrated that *OCT4A* expression is not as high as it can be seen in the controls (average Ct approximately 31). We also observed that the expression of *OCT4A* was mildly decreased in metastases compared with primary tumors. We suppose that it is related to the change in *OCT4B* expression. Li et al. (65) found out that *OCT4B* functions as a non-coding RNA, modulating *OCT4A* expression by competitive binding with microRNAs. This may highlight the role of *OCT4B* in miRNA regulation of *OCT4A* expression (65). Hypothetically, this could be the reason why the expression of *OCT4A* was reduced in primary tumors, and with the mild increase in *OCT4B* expression, we observed a little but statistically non-significant decrease in *OCT4A* expression in metastases.

We also observed significantly upregulated expression of all *OCT4* isoforms in metastasis compared with primary tumors. Such increased expression may be caused by using a probe that specifically recognizes a mutual exon shared by all of the already identified *OCT4* isoforms. Thus, we were probably able to detect the expression of not only the *OCT4A* and *OCT4B* isoforms but also fewer known isoforms such as *OCT4C* and *OCT4D*, whose existence at the protein level has not yet been confirmed (15, 16). Several research groups have published results that emphasize the role of *OCT4* in the aggressive behavior of CRC and its contribution to forming liver metastasis in CRC, especially *OCT4* in the high-expression group, which is consistent with our results (66–68). Furthermore, *OCT4* expression was denoted as an independent prognostic biomarker for predicting worse disease-specific survival

and overall survival in CRC (69). On the other hand, the expression of all *OCT4* isoforms in both primary and metastatic samples than in control samples was significantly downregulated. Similar results, such as reduced gene expression of *OCT4* in tumor tissue compared with control and regardless of the isoforms, were also obtained from breast cancer (60).

For the *OCT4B* isoforms, there is a typical expression in various non-pluripotent cell types and differentiated tissues at different levels based on a specific isoform. There is also unequivocal evidence that *OCT4A* is also expressed in adult human stem cells and differentiated somatic cells, in addition to pluripotent cells, but at a much lower level (14). Surprisingly, *OCT4B1* has a similar expression pattern to *OCT4A* (18). As a result of determining the expression of all *OCT4B* isoforms together, we were not able to designate which isoform had a higher expression and vice versa. In primary tumors, the expression of the *OCT4B* isoforms was significantly reduced compared with controls. When compared with metastases, the expression was slightly reduced but without a statistically significant result.

Although we used data from the TCGA and despite our efforts, we were not able to confirm in the TCGA dataset our findings of reduced *POU5F1* gene expression in tumor samples compared with non-tumor controls at a statistically significant level. The TCGA results (Supplementary Table 3) indicate a non-significant trend of the increase of the *POU5F1* expression from normal tissue to primary tumor and to metastases. Our analyses demonstrate only a significant increase in gene expression solely between the primary and metastatic samples. In contrast, we have observed a decrease in gene expression in tumor samples relative to non-tumor samples. Although it would have been ideal to have paired TCGA data, the number of non-cancerous samples available for comparison was rather limited. To avoid any contradictions in our research, we believe that it is important to conduct further validation studies with a larger sample size and better access to the TCGA data. It will also be necessary to differentiate between individual *OCT4* isoforms, which were not distinguished in many previous studies, to obtain more conclusive results on the gene expression of *POU5F1* and its isoforms. These efforts will provide a better understanding of the role of *POU5F1* in colorectal cancer.

The role of *OCT4* expression as a prognostic marker, as well as its role in metastatic CRC, has already been explored. Patients with high *OCT4* expression had a poorer prognosis, making it a potential marker for the diagnosis and assessment of the prognosis of CRC. In addition, the results also indicate that *OCT4* expression was correlated with clinical stage, tumor grade, metastasis, and TNM stage (69–71). In our study, we did not observe any correlation between *OCT4* isoform expression and clinical or TNM tumor stage. The authors of the studies mentioned above did not recognize between individual *OCT4* isoforms which may cause inconsistency in the results. Our results also suggest a correlation between reduced expression of all *OCT4* isoforms and both primary and left-sided tumors. On the other hand, Talebi et al. did not observe *OCT4* expression in any of the tissues tested (normal, polyp, and cancer tissue) and concluded that the diagnostic power of the *OCT4* gene is

not enough to identify cancer (72). Even though the exact role of *OCT4B* isoforms in the cell is still under investigation, several studies indicate that in some way they may contribute to the properties of cancer cells, such as invasion, having antiapoptotic properties, and resistance to chemotherapy (73, 74). In addition, Gazouli et al. confirmed the expression of the *OCT4B* and *OCT4B1* isoforms in CRC samples and observed that the level of *OCT4B1* mRNA was correlated with poorly and moderately differentiated CRC and with the progression of cancer stage (62). In our study, no correlation was detected between *OCT4B* expression and clinical data as well as the type or side of the tumor or TNM stage. Another research group showed that *OCT4B1* has a potential role in regulating the self-renewal of colorectal CSC through its involvement in epithelial–mesenchymal transition (75). Furthermore, Simó-Riudalbas et al. demonstrated a pro-oncogenic effect of *OCT4B1* through its association with protein kinases and subsequent activation of intracellular signaling events as well as cytoskeletal rearrangements (76). All these findings can signify a potential role of *OCT4B*, especially *OCT4B1* as a marker of tumor-initiating cells or CSCs, and not only *OCT4A* but all *OCT4* isoforms might play a significant role in carcinogenesis.

Based on previous results as well as our results, the expression level of *OCT4* isoforms could be a useful tool not only for diagnosis, especially metastatic disease, but also for prognosis prediction. As the results of the TCGA analysis and our own analyses are not fully consistent, further studies will be needed to clarify these differences. Thus, these results emphasize the importance of precise characterization of individual *OCT4* isoforms whether transcriptional or protein, as well as their expression patterns in colorectal cancer.

5 Conclusion

Unlike previous reports, we found out that the expression of *OCT4A*, *OCT4Bs*, and all *OCT4* isoforms is significantly reduced in primary tumors and metastases compared with control samples. On the other hand, the expression of all *OCT4* isoforms was significantly upregulated in metastases compared with primary tumors, which was not caused by the upregulation of the *OCT4A* isoform and only partially by the *OCT4B* isoform, emphasizing the role of less-known isoforms in metastatic CRC. Furthermore, the reduced expression of all *OCT4* isoforms was correlated with primary and left-sided tumors. Based on these results, we supposed that the expression rate of all *OCT4* isoforms can be related to the cancer type and side, as well as to liver metastasis. However, further studies are required to investigate the detailed expression patterns and significance of individual *OCT4* isoforms in carcinogenesis.

Data availability statement

The datasets presented in this study are available on <https://data.mendeley.com/datasets/2c5296ytbr> (42).

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of the Jessenius Faculty of Medicine in Martin, Comenius University in Bratislava by decision number 1863/2016. The patients/participants provided their written informed consent to participate in this study.

Author contributions

ET was responsible for the experimental design, performed the experiments, and wrote the manuscript. ZL secured the financial funding, was responsible for the experimental design, supervised the data analysis and interpretation, and critically reviewed and revised the manuscript drafts. VH and ZL assisted in conducting the experiments. MG performed the bioinformatic analyses and generated the figures. MH and MKo performed the gene analysis from the TCGA raw reads and wrote the TCGA part of the manuscript. MKa and JMa performed the histopathological evaluation. EK and PM collected the patients' samples. JMi and LL critically reviewed and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from The Ministry of Education, Science, Research and Sport of the Slovak Republic (grant VEGA no. 1/0269/22) and by The Slovak Research and Development Agency (grant APVV no. 16-0066 and no. 21-0448). Michal Kovac received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie COFUND SASPro2 grant agreement no. 2207/02/01.

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. *Cancer J Clin* (2021) 7:209–49. doi: 10.3322/caac.21660
- Chang VC, Cotterchio M, De P, Tinmouth J. Risk factors for early-onset colorectal cancer: a population-based case–control study in Ontario, Canada. *Cancer Cause Control* (2021) 32:1063–83. doi: 10.1007/s10552-021-01456-8
- Botteri E, Borroni E, Sloan E, Bagnardi V, Bosetti C, Peveri G, et al. Smoking and colorectal cancer risk, overall and by molecular subtypes: a meta-analysis. *Am J Gastroenterol* (2020) 115:1940–9. doi: 10.14309/ajg.0000000000000803
- Gausman V, Dornblaser D, Anand S, Hayes RB, O'Connell K, Du M, et al. Risk factors associated with early-onset colorectal cancer. *Clin Gastroenterol Hepatol* (2020) 18:2752–9. doi: 10.1016/j.cgh.2019.10.009
- Guillaudeux T, Mattei MG, Depetris D, Le Bouteiller P, Pontarotti P. *In situ* hybridization localizes the human OTF3 to chromosome 6p21.3 ?p22 and OTF3L to 12p13. *Cytogenet Cell Genet* (1993) 63:212–4. doi: 10.1159/000133537
- Herr W, Sturm RA, Clerc RG, Corcoran LM, Baltimore D, Sharp PA, et al. The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and caenorhabditis elegans unc-86 gene products. *Genes Dev* (1988) 2:1513–6. doi: 10.1101/gad.2.12a.1513
- Verijzer CP, Alkema MJ, van Weperen WW, Van Leeuwen HC, Strating JJ, van der Vliet PC. The DNA binding specificity of the bipartite POU domain and its subdomains. *EMBO J* (1992) 11:4993–5003. doi: 10.1002/j.1460-2075.1992.tb05606.x
- Shi G, Jin Y. Role of Oct4 in maintaining and regaining stem cell pluripotency. *Stem Cell Res Ther* (2010) 1:39. doi: 10.1186/scrt39
- Mehra V, Ghaemimaneh F, Poursani EM. An overview on the complexity of OCT4: at the level of DNA, RNA and protein. *Stem Cell Rev Rep* (2021) 17:1121–36. doi: 10.1007/s12015-020-10098-3
- Poursani EM, Mehra V, Shahryari A, Mowla SJ, Mohammad Soltani B. Alternative splicing generates different 5' UTRs in OCT4B variants. *Avicenna J Med Biotechnol* (2017) 9:201–4.
- Takeda J, Seino S, Bell GI. Human Oct3 gene family: cDNA sequences, alternative splicing, gene organization, chromosomal location, and expression at low levels in adult tissues. *Nucleic Acids Res* (1992) 20:4613–20. doi: 10.1093/nar/20.17.4613
- Atlasi Y, Mowla SJ, Ziaee SAM, Gokhale PJ, Andrews PW. OCT4 spliced variants are differentially expressed in human pluripotent and nonpluripotent cells. *Stem Cells* (2008) 26:3068–74. doi: 10.1634/stemcells.2008-0530
- Poursani EM, Mehra V, Soltani BM, Mowla SJ. OCT4B2, a novel alternative spliced variant of OCT4, is significantly upregulated under heat-stress condition and downregulated in differentiated cells. *Tumour Biol* (2017) 39. doi: 10.1177/1010428317724280
- Zhao FQ, Misra Y, Li DB, Wadsworth MP, Krag D, Weaver D, et al. Differential expression of Oct3/4 in human breast cancer and normal tissues. *Int J Oncol* (2018) 52:2069–78. doi: 10.3892/ijo.2018.4341

Acknowledgments

We would like to thank Mrs. Andrea Vaňochová for her professional technical assistance.

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.

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/fonc.2023.1166835/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Areas within the *POU5F1* gene which are recognized by specific TaqMan probe.

SUPPLEMENTARY FIGURE 2

Gene expression of *POU5F1* based on Transcript Per Kilobase Million (TPM) counts from RNA sequencing experiments. Conformity to the Gaussian distribution was assessed and was subsequently rejected in each group, owing to the substantial enrichment of samples with high expression (Shapiro-Wilk test: normal tissue: $P = 2.901e-10$, tumor tissue: $P < 2.2e-16$, metastases: $P = 0.0009742$).

15. Mehravar M, Poursani EM. Novel variant of OCT4, named OCT4B5, is highly expressed in human pluripotent cells. *Stem Cell Rev Rep* (2021) 17:1068–73. doi: 10.1007/s12015-020-10093-8
16. Malakootian M, Azad FM, Naeli P, Pakzad M, Fouani Y, Bajgan ET. Novel spliced variants of OCT4, OCT4C and OCT4C1, with distinct expression patterns and functions in pluripotent and tumor cell lines. *Eur J Cell Biol* (2017) 96:347–55. doi: 10.1016/j.ejcb.2017.03.009
17. Wang X, Zhao Y, Xiao Z, Chen B, Wei Z, Wang B, et al. Alternative translation of OCT4 by an internal ribosome entry site and its novel function in stress response. *Stem Cells* (2009) 27:1265–75. doi: 10.1002/stem.58
18. Lee J, Kim HK, Rho JY, Han YM, Kim J. The human OCT-4 isoforms differ in their ability to confer self-renewal. *J Biol Chem* (2006) 281:33554–65. doi: 10.1074/jbc.M603937200
19. Tai MH, Chang CC, Kiupel M, Webster JD, Olson LK, Trosko JE. Oct4 expression in adult human stem cells: evidence in support of the stem cell theory of carcinogenesis. *Carcinogenesis* (2005) 26:495–502. doi: 10.1093/carcin/bgh321
20. Palmieri SL, Peter W, Hess H, Schöler HR. Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. *Dev Biol* (1994) 166:259–67. doi: 10.1006/dbio.1994.1312
21. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science* (2007) 318:1917–20. doi: 10.1126/science.1151526
22. Cauffman G, Liebaers I, Van Steirteghem A, Van de Velde H. POU5F1 isoforms show different expression patterns in human embryonic stem cells and preimplantation embryos. *Stem Cells* (2006) 24:2685–91. doi: 10.1634/stemcells.2005-0611
23. O'Flaherty JD, Barr N, Fennell D, Richard D, Reynolds J, O'Leary J, et al. The cancer stem-cell hypothesis: its emerging role in lung cancer biology and its relevance for future therapy. *J Thorac Oncol* (2012) 7:1880–90. doi: 10.1097/JTO.0b013e31826bfb6c
24. Ilieva M, Dufva M. SOX2 and OCT4 mRNA-expressing cells, detected by molecular beacons, localize to the center of neurospheres during differentiation. *PLoS One* (2013) 8:e73669. doi: 10.1371/journal.pone.0073669
25. Losada-García A, Salido-Guandarrama I, Cortes-Ramirez SA, Cruz-Burgos M, Morales-Pacheco M, Vazquez-Santillan. SFRP1 induced a stem cell phenotype in prostate cancer cells. *Front Cell Dev Biol* (2023) . 11:1096923. doi: 10.3389/fcell.2023.1096923
26. Yamashita A, Yoshitomi H, Kihara S, Toguchida J, Tsumaki N. Culture substrate-associated YAP inactivation underlies chondrogenic differentiation of human induced pluripotent stem cells. *Stem Cells Transl Med* (2021) 1:115–27. doi: 10.1002/sctm.20-0058
27. Okuno M, Aoki S, Kawai S, Imataki R, Abe Y, Harada K, et al. Effect of non-thermal atmospheric pressure plasma on differentiation potential of human deciduous dental pulp fibroblast like cells. *Appl Sci* (2021) . 21:10119. doi: 10.3390/app112110119
28. Gil-Kulik P, Chomik P, Krzyzanowski A, Radzikowska-Buchner E, Maciejewski R, Kwasniewska A, et al. Influence of the type of delivery, use of oxytocin, and maternal age on POU5F1 gene expression in stem cells derived from wharton's jelly within the umbilical cord. *Oxid Med Cell Longev* (2019) 2019:1027106. doi: 10.1155/2019/1027106
29. Timmerman DM, Eleveld TF, Gillis AJM, Friedrichs CC, Hillenius S, Remmers TL, et al. The role of TP53 in cisplatin resistance in mediastinal and testicular germ cell tumors. *Int J Mol Sci* (2021) . 22:11774. doi: 10.3390/ijms222111774
30. Rao X, Huang X, Zhou Z, Lin X. An improvement of the 2^{−(delta delta CT)} method for quantitative real-time polymerase chain reaction data analysis. *Bioinforma Bioma* (2013) 3:71–85.
31. Gagatelewski M. Stringi: fast and portable character string processing in r. j. stat. *Software* (2022) 103:2. doi: 10.18637/jss.v103.i02
32. Goode K, Rey K. ggResidpanel: panels and interactive versions of diagnostic plots using 'ggplot2'. R package (2022). Available at: <https://goodekat.github.io/ggResidpanel/>.
33. Kassambara A. Ggpubr: 'ggplot2' based publication ready plots. R package (2020). Available at: <https://rpkgs.datanovia.com/ggpubr>.
34. Kassambara A. Rstatix: pipe-friendly framework for BasicStatistical tests. R package (2021). Available at: <https://CRAN.R-project.org/package=rstatix>.
35. Lenth R. Emmeans: estimated marginal means, aka least-SquaresMeans. R package (2022). Available at: <https://CRAN.R-project.org/package=emmeans>.
36. Lüdtke D. sjPlot: data visualization for statistics in social science. R package (2023). Available at: <https://CRAN.R-project.org/package=sjPlot>.
37. Ogle DH, Doll JC, Wheeler AP, Dinno A. FSA: Simple FisheriesStock assessment methods. R package (2023). Available at: <https://CRAN.R-project.org/package=FSA>.
38. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing (2021). Available at: <https://www.R-project.org/>.
39. Wickham H. Reshaping data with the reshape package. *J Stat Software* (2007) 21:12. doi: 10.18637/jss.v021.i12
40. Wickham H. GGplot2: elegant graphics for data analysis (2016). Available at: <https://ggplot2.tidyverse.org>.
41. Zhu H. kableExtra: construct complex table with 'kable' and Pipe syntax. R package (2021). Available at: <https://cran.r-project.org/web/packages/kableExtra/index.html>.
42. Turyova E, Mikolajcik P, Grendar M, Kudelova E, Holubekova V, Kalman M, et al. Supplement to expression of OCT4 isoforms is reduced in primary colorectal cancer. *Mendeley Data* (2023) V1. doi: 10.17632/2c5296ytrb.1
43. Okamoto K, Okazawa H, Okuda A, Sakai M, Muramatsu M, Hamada H. A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* (1990) 60:461–72. doi: 10.1016/0092-8674(90)90597-8
44. Feldman N, Gerson A, Fang J, Li E, Zhang Y, Shinkai Y, et al. G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat Cell Biol* (2006) 8:188–94. doi: 10.1038/ncb1353
45. Nichols J, Zevnik B, Anastasiadis K, Niwa H, Klewe-Nebenius D, Chambers I, et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* (1998) 95:379–91. doi: 10.1016/S0092-8674(00)81769-9
46. Fogarty NME, McCarthy A, Snijders KE, Powell BE, Kubikova N, Blakeley P, et al. Genome editing reveals a role for OCT4 in human embryogenesis. *Nature* (2017) 550:67–73. doi: 10.1038/nature24033
47. Mulas C, Chia G, Jones KA, Hodgson AC, Stirparo GG, Nichols J. Oct4 regulates the embryonic axis and coordinates exit from pluripotency and germ layer specification in the mouse embryo. *Development* (2018) 145. doi: 10.1242/dev.159103
48. Lo JHH, Edwards M, Langerman J, Sridharan R, Plath K, Smale ST. Oct4:Sox2 binding is essential for establishing but not maintaining active and silent states of dynamically regulated genes in pluripotent cells. *Genes Dev* (2022) 36:1079–95. doi: 10.1101/gad.350113.122
49. Zhang Q, Han Z, Zhu Y, Chen J, Li W. The role and specific mechanism of OCT4 in cancer stem cells: a review. *Int J Stem Cells* (2020) 13:312–25. doi: 10.15283/ijsc.20097
50. Frank NY, Schatton T, Frank MH. The therapeutic promise of the cancer stem cell concept. *J Clin Invest* (2010) 120:41–50. doi: 10.1172/JCI41004
51. Ruan Z, Yang X, Cheng W. OCT4 accelerates tumorigenesis through activating JAK/STAT signaling in ovarian cancer side population cells. *Cancer Manag Res* (2018) 11:389–99. doi: 10.2147/CMAR.S180418
52. Li J, Huang NF, Zou J, Laurent TJ, Lee JC, Okogbaa J, et al. Conversion of human fibroblasts to functional endothelial cells by defined factors. *Arterioscler Thromb Vasc Biol* (2013) 33:1366–75. doi: 10.1161/ATVBAHA.112.301167
53. Hess DL, Kelly-Goss MR, Cherepanova OA, Nguyen AT, Baylis RA, Tkachenko S, et al. Perivascular cell-specific knockout of the stem cell pluripotency gene Oct4 inhibits angiogenesis. *Nat Commun* (2019) 10:967. doi: 10.1038/s41467-019-08811-z
54. Liu HL, ting TH, lin YH, Deng TT, Xu YP, Xu SQ, et al. Oct4 regulates the transition of cancer stem-like cells to tumor endothelial-like cells in human liver cancer. *Front Cell Dev Biol* (2020) 8:563316. doi: 10.3389/fcell.2020.563316
55. Ku JL, Shin YK, Kim DW, Kim KH, Choi JS, Hong SH, et al. Establishment and characterization of 13 human colorectal carcinoma cell lines: mutations of genes and expressions of drug-sensitivity genes and cancer stem cell markers. *Carcinogenesis* (2010) 31:1003–9. doi: 10.1093/carcin/bgq043
56. Wang YJ, Herlyn M. The emerging roles of Oct4 in tumor-initiating cells. *Am J Physiol Cell Physiol* (2015) 309:C709–718. doi: 10.1152/ajpcell.00212.2015
57. Gao Y, Wang X, Han J, Xiao Z, Chen B, Su G, et al. The novel OCT4 spliced variant OCT4B1 can generate three protein isoforms by alternative splicing into OCT4B. *J Genet Genomics* (2010) 7:461–5. doi: 10.1016/S1673-8527(09)60065-5
58. Gao Y, Wei J, Han J, Wang X, Su G, Zhao Y, et al. The novel function of OCT4B isoform-265 in genotoxic stress. *Stem Cells* (2012) 4:665–72. doi: 10.1002/stem.1034
59. Liu YH, Li Y, Liu XH, Sui HM, Liu YX, Xiao ZQ, et al. A signature for induced pluripotent stem cell-associated genes in colorectal cancer. *Med Oncol* (2013) 1:426. doi: 10.1007/s12032-012-0426-2
60. Saha SK, Jeong Y, Cho S, Cho SG. Systematic expression alteration analysis of master reprogramming factor OCT4 and its three pseudogenes in human cancer and their prognostic outcomes. *Sci Rep* (2018) 8:14806. doi: 10.1038/s41598-018-33094-7
61. Suo G, Han J, Wang X, Zhang J, Zhao Y, Zhao Y, et al. Oct4 pseudogenes are transcribed in cancers. *Biochem Biophys Res Commun* (2005) 337:1047–51. doi: 10.1016/j.bbrc.2005.09.157
62. Gazouli M, Roubelakis MG, Theodoropoulos GE, Papailiou J, Vaiopoulou A, Pappa KI, et al. OCT4 spliced variant OCT4B1 is expressed in human colorectal cancer. *Mol Carcinog* (2012) 51:165–73. doi: 10.1002/mc.20773
63. Poursani EM, Mohammad Soltani B, Mowla SJ. Differential expression of OCT4 pseudogenes in pluripotent and tumor cell lines. *Cell J* (2016) 18:28–36. doi: 10.22074/cellj.2016.3984
64. Pain D, Chirn GW, Strassel C, Kemp DM. Multiple retrospseudogenes from pluripotent cell-specific gene expression indicates a potential signature for novel gene identification. *J Biol Chem* (2005) 280:6265–8. doi: 10.1074/jbc.C400587200
65. Li D, Yang ZK, Bu JY, Xu CY, Sun H, Tang JB, et al. OCT4B modulates OCT4A expression as ceRNA in tumor cells. *Oncol Rep* (2015) 33:2622–30. doi: 10.3892/or.2015.3862

66. Dai X, Ge J, Wang X, Qian X, Zhang C, Li X. OCT4 regulates epithelial-mesenchymal transition and its knockdown inhibits colorectal cancer cell migration and invasion. *Oncol Rep* (2013) 29:155–60. doi: 10.3892/or.2012.2086
67. Fujino S, Miyoshi N. Oct4 gene expression in primary colorectal cancer promotes liver metastasis. *Stem Cells Int* (2019) 2019:7896524. doi: 10.1155/2019/7896524
68. Roudi R, Barodabi M, Madjd Z, Roviello G, Corona SP, Panahi M. Expression patterns and clinical significance of the potential cancer stem cell markers OCT4 and NANOG in colorectal cancer patients. *Mol Cell Oncol* (2020) 7:1788366. doi: 10.1080/23723556.2020.1788366
69. Fang W, Ni M, Zhang M, Chen H. Prognostic value of OCT4 in colorectal cancer: analysis using immunohistochemistry and bioinformatics validation. *biomark Med* (2020) 14:1473–84. doi: 10.2217/bmm-2020-0069
70. Zhou H, Hu YU, Wang W, Mao Y, Zhu J, Zhou B, et al. Expression of Oct-4 is significantly associated with the development and prognosis of colorectal cancer. *Oncol Lett* (2015) 2:691–6. doi: 10.3892/ol.2015.3269
71. Miyoshi N, Fujino S, Ohue M, Yasui M, Takahashi Y, Sugimura K, et al. The POU5F1 gene expression in colorectal cancer: a novel prognostic marker. *Surg Today* (2018) 7:709–15. doi: 10.1007/s00595-018-1644-9
72. Talebi A, Kianersi K, Beiraghdar M. Comparison of gene expression of SOX2 and OCT4 in normal tissue, polyps, and colon adenocarcinoma using immunohistochemical staining. *Adv BioMed Res* (2015) 4:234. doi: 10.4103/2277-9175.167958
73. Wen KM, Zhang GH, Li J, Chen ZQ, Cheng YL, Su X, et al. OCT4B1 promotes cell growth, migration and invasion suppressing sensitivity to oxaliplatin in colon cancer. *Oncol Rep* (2015) 34:2943–52. doi: 10.3892/or.2015.4286
74. Wen K, Fu Z, Wu X, Feng J, Chen W, Qian J. Oct-4 is required for an antiapoptotic behavior of chemoresistant colorectal cancer cells enriched for cancer stem cells: effects associated with STAT3/Survivin. *Cancer Lett* (2013) 333(1):56–65. doi: 10.1016/j.canlet.2013.01.009
75. min ZJ, qing H, Jiang H, lin CY, hong FJ, quan CZ, et al. OCT4B1 promoted EMT and regulated the self-renewal of CSCs in CRC: effects associated with the balance of miR-8064/PLK1. *Mol Ther – Oncolytics* (2019) 15:7–20. doi: 10.1016/j.omto.2019.08.004
76. Simó-Riudalbas L, Offner S, Planet E, Duc J, Abrami L, Dind S, et al. Transposon-activated POU5F1B promotes colorectal cancer growth and metastasis. *Nat Commun* (2022) 1:4913. doi: 10.1038/s41467-022-32649-7



OPEN ACCESS

EDITED BY

David Gibbons,
St. Vincent's University Hospital, Ireland

REVIEWED BY

Bhavna Murali,
Cerner, United States
Yanqiang Li,
Boston Children's Hospital and Harvard
Medical School, United States

*CORRESPONDENCE

Tatiana Burjanivova
✉ BTatiana@seznam.cz
Zora Lasabova
✉ zora.lasabova@uniba.sk

[†]These authors have contributed
equally to this work

RECEIVED 14 April 2023

ACCEPTED 12 June 2023

PUBLISHED 05 July 2023

CITATION

Lukacova E, Burjanivova T, Podlesniy P,
Grendar M, Turyova E, Kasubova I, Laca L,
Mikolajcik P, Kudelova E, Vanochova A,
Miklusica J, Mersakova S and Lasabova Z
(2023) Hypermethylated *GRIA4*, a potential
biomarker for an early non-invasive
detection of metastasis of clinically
known colorectal cancer.
Front. Oncol. 13:1205791.
doi: 10.3389/fonc.2023.1205791

COPYRIGHT

© 2023 Lukacova, Burjanivova, Podlesniy,
Grendar, Turyova, Kasubova, Laca, Mikolajcik,
Kudelova, Vanochova, Miklusica, Mersakova
and Lasabova. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). 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.

Hypermethylated *GRIA4*, a potential biomarker for an early non-invasive detection of metastasis of clinically known colorectal cancer

Eva Lukacova^{1†}, Tatiana Burjanivova^{1*†}, Petar Podlesniy²,
Marian Grendar³, Eva Turyova¹, Ivana Kasubova⁴, Ludovit Laca⁵,
Peter Mikolajcik⁵, Eva Kudelova⁵, Andrea Vanochova¹,
Juraj Miklusica⁵, Sandra Mersakova⁴ and Zora Lasabova^{1*}

¹Department of Molecular Biology and Genomics, Comenius University in Bratislava, Jessenius Faculty of Medicine in Martin (JFM CU), Martin, Slovakia, ²Centro Investigacion Biomedica en Red Enfermedades Neurodegenerativas (CiberNed), Madrid, Spain, ³Laboratory of Bioinformatics and Biostatistics, Biomedical Center Martin JFM CU, Comenius University in Bratislava, Jessenius Faculty of Medicine in Martin (JFM CU), Martin, Slovakia, ⁴Biomedical Center Martin, Jessenius Faculty of Medicine in Martin, Comenius University in Bratislava, Martin, Slovakia, ⁵Clinic of Surgery and Transplant Center, Jessenius Faculty of Medicine in Martin, Comenius University in Bratislava, Martin, Slovakia

Introduction: Colorectal cancer (CRC) can develop through several dysregulated molecular pathways, including the serrated pathway, characterized by CpG island methylator (CIMP) phenotype. Although the tumor tissue is a commonly tested material, sample types such as stool or plasma, bring a new, non-invasive approach. Several cancer-related methylated genes have been identified in CRC patients, including gene *GRIA4*, showing promising diagnostic potential. The aim of our study was to develop a sensitive droplet digital PCR (ddPCR) assay to examine *GRIA4* hypermethylation status in CRC patients and evaluate its diagnostic potential in tissue and liquid biopsy samples.

Methods: In total, 23 patients participated in this study, 7 patients with primary CRC and 16 patients with liver metastasis of clinically known CRC. We obtained tumor and non-tumor tissues (N=17), blood samples pre- and post-surgery (N=22), and blood of five volunteers without a personal cancer history. We have developed and optimized a ddPCR assay for *GRIA4* hypermethylation detection, from tissue and plasma samples.

Results: We detected significantly increased *GRIA4* methylation in tumor tissues compared to their adjacent non-tumor tissue, $p < 0.0001$. Receiver operating characteristic (ROC) analysis defined cutoff values to separate primary tumors and metastases from non-tumor colon/rectum, specifically 36.85% for primary tumors and 34.81% for metastases. All primary tumors were above this threshold. When comparing the methylation levels of metastatic vs. non-tumor tissue, a smaller increase was observed in liver metastasis versus colon tissue (3.6x gain; $p = 0.001$), then in liver metastasis versus adjacent liver tissue (17.4x gain; $p < 0.0001$). On average, *GRIA4* hypermethylation in primary tumor plasma was

2.8-fold higher ($p=0.39$), and in metastatic plasma, 16.4-fold higher ($p=0.0011$) compared to healthy individuals. Hypermethylation in metastatic plasma was on average 5.9 times higher ($p=0.051$) than in primary tumor plasma. After tumor removal surgery, average hypermethylation decrease in plasma was 1.6x for primary ($p=0.037$) and 4.5x for metastatic patients ($p=0.023$).

Discussion: Based on our data, it can be inferred that *GRIA4* serves as a tissue specific biomarker for the colon/rectum tissue, thus is suitable for cancer classification. This biomarker showed the potential to be an attractive target for early non-invasive detection of metastases of clinically known CRC, although additional analysis has to be performed.

KEYWORDS

colorectal cancer, methylation, ctDNA, metastasis, liquid biopsy

1 Introduction

Colorectal cancer (CRC) is the third most prevalent cancer in men, after lung and prostate cancer, and the second most common cancer in women, after breast cancer (1). According to Globocan 2020 (2), 4,821 new cases of colorectal cancer were diagnosed in 2020 in Slovakia, which makes it the most common form of cancer in the Slovak Republic. The most common site of distant metastasis for colorectal cancer is the liver, due to the direct connection through the portal vein (3). Approximately 50% of patients with colorectal cancer develop liver metastases during the course of their disease (4).

The development of CRC can proceed through the accumulation of genetic and epigenetic changes (5). DNA methylation is the most common epigenetic modification. It ensures cell-specific gene expression for normal development, cell functioning, and tissue stability. On the other hand, in somatic cells, hypermethylation/hypomethylation within the specific promoter region can contribute to neoplastic cell transformation (6). In 1999, Toyota et al. (7) proposed the term CIMP (CpG island methylator phenotype) to describe a subset of CRCs that show extensive hypermethylation of CpG dinucleotides.

Multiple sample types can be used for the identification of epigenetic alterations in CRC patients. Although tumor tissue is the predominant choice, non-invasive approaches, such as stool or plasma sampling, are progressively being incorporated into clinical practice. Circulating tumor DNA (ctDNA) is tumor-derived DNA, present in body fluids, such as blood, stool, urine, and saliva (8, 9). It is released into the bloodstream by tumor cells undergoing apoptosis/necrosis or *via* active secretion (10). Due to its origin, ctDNA provides comprehensive genetic and epigenetic information about tumor, and its concentration can greatly vary, depending on tumor size and type, proliferative stage, response to the treatment, level of vascularization, etc. (11). Circulating tumor DNA can be used to monitor disease dynamics non-invasively. It has the potential to assess therapy response and efficacy (12–14) and predict and improve early relapse detection (15, 16).

A large number of cancer-related methylated genes have been identified in CRC patients, for instance, *MLH1*, *CDKN2A*, *MGMT* (17, 18), *SFRP2* (19), *Vimentin* (20), *BMP3* (21), *Sept9* (22), *NDRG4* (23), and many others, in the last few decades. Currently, a limited number of assays for non-invasive hypermethylation detection are commercially available, for example, assays for the detection of methylated *Sept9* (Epi proColon®) from plasma (24), *Vimentin* (ColoSure™) (25), and *BMP3* together with *NDRG4* (Cologuard®) from fecal DNA (26). When searching for other epigenetic biomarkers, gene *GRIA4* (glutamate ionotropic receptor AMPA subunit 4) has great diagnostic potential in patients with colorectal cancer, although studies on its biological properties are quite limited. It was published recently that 100% tissue and 71.3% plasma samples of metastatic CRC patients had a higher methylation profile for *GRIA4* gene (27). In another study, gene *GRIA4* showed hypermethylation in 99.1% of experimental tissue samples (28). The presence of methylated *GRIA4* promoter was also observed in stool specimens, indicating its potential utility as a biomarker for the early detection of colorectal cancer from stool samples (29). The investigations conducted by and Sun et al. (30) and Hauptman et al. (28), both in 2019, employed the TCGA dataset to examine CRC methylation biomarkers, including hypermethylated *GRIA4*. Sun et al. aimed to validate a previously identified markers, while Hauptman et al. sought to identify potential CRC biomarkers from TCGA data. Among 198 genes, *GRIA4* exhibited the most significant methylation difference among six selected genes. All four promoter probes of the *GRIA4* gene displayed high methylation difference, with two of them present in 98.4% of the samples. Furthermore, *GRIA4* was found to be downregulated in 98.1% of the samples within the TCGA dataset.

Circulating DNA can be extracted from plasma and identified using a variety of molecular techniques. Analyzing tumor material acquired by liquid biopsies necessitates very sensitive assays. Multiple ctDNA analysis platforms are currently available, but PCR-based techniques are still the backbone of all detection strategies. Droplet digital PCR (ddPCR) is a sensitive, low-cost detection method that has been commercially available since 2011 (31). This technique is suitable

for targeting specific mutations/methylations/alterations on DNA fragments present at very low concentrations. However, the complexity of laboratory protocols, the constraint in the number of targets being tested, and the variability in analytical sensitivity could potentially impose limitations on the application of this technique (32). The aim of our study was to develop sensitive droplet digital assay to examine *GRIA4* methylation status in CRC patients, either with a primary tumor or with metastases. The next step was to evaluate its diagnostic potential using non-invasive liquid biopsy samples.

2 Patients and methods

2.1 Patients

In total, 23 patients participated in this study (Supplementary Table S1), 7 patients with primary tumor (Supplementary Table S2) and 16 patients with liver metastasis of clinically known colorectal cancer. Tumor tissues of 17 CRC patients (5 primary and 12 metastatic) and blood samples of 22 CRC patients (6 primary and 16 metastatic) and of 5 volunteers without a personal cancer history were obtained in collaboration with Clinic of General, Visceral and Transplantation Surgery and Department of Pathology at the Jessenius Faculty of Medicine (Comenius University), University Hospital in Martin. Blood samples of patients were taken before and after surgical removal of the tumor. The histopathological diagnosis was conducted by the experienced pathologists. Histological typing, grading, localization, and staging of tumors were determined using the recommendation according to WHO and Union for International Cancer Control (UICC) (33). This study was approved by the Ethics Review Board of the Jessenius Faculty of Medicine.

2.2 DNA isolation and quantification, bisulfite conversion

Tumor and non-tumor samples were obtained after resection surgery with subsequent evaluation of the tissue by an experienced pathologist. Genomic DNA from tumor tissues was isolated with the commercial kit DNeasy Blood and Tissue (Qiagen, Hilden, Germany) according to manufacturer's instructions. Isolated DNA was eluted into 60 µl and stored at −20°C. EpiTect Bisulfite Kit (Qiagen, Hilden, Germany) was used for bisulfite conversion of gDNA. Converted DNA

was eluted into 30 µl and stored at −20°C. Blood samples were collected to ethylenediaminetetraacetic acid (EDTA) tubes and centrifuged at 2,200g for 8 min at 4°C. Plasma was pipetted into new 1.5-ml tubes and centrifuged one more time at 20,000g for 8 min at 4°C. Plasma samples were then stored at −80°C until cfDNA extraction was performed. Circulating DNA extraction from plasma samples and bisulfite conversion were performed according to the manufacturer's instructions of Epi proColon 2.0 CE Plasma Quick Kit (Epigenomics AG, Berlin, Germany) from 3.5 ml of plasma. Bisulfite-converted DNA was then eluted into 60 µl, stored at 4°C, and used for analysis within 24 h. The rest of the eluate was stored at −20°C. For DNA quantification, Qubit dsDNA HS assay kit (Life Technologies, CA, USA) and Qubit 2.0 fluorometer were used.

2.3 Droplet digital PCR

Droplet digital PCR was performed in 20 µl ddPCR reactions, containing 10 µl Supermix for Probes (No dUTP) (Bio-Rad Laboratories, Hercules, CA, USA), 1.4 µl primers (final concentration, 225 nM) and probes (final concentration 125 nM) (Table 1), and 8.6 µl of circulating DNA or 0.6 µl of genomic DNA adjusted with water up to volume 8.6 µl. The probe complementary to the methylated sequence of the *GRIA4* promoter is referred as M-Probe, and the one complementary to the unmethylated sequence is referred as U-Probe (Table 1). As controls, commercially available, fully methylated and fully unmethylated, EpiTect DNA controls (Qiagen, Hilden, Germany) and ultrapure water were used to check for template contamination. The reaction (20 µl) from the previous step was transferred to the middle rows of a DG8 (Bio-Rad Laboratories, Hercules, CA, USA) cartridge. After that, 70 µl of Droplet Generation Oil for Probes was loaded into the bottom wells of DG8. Cartridge was then placed into the QX200 droplet generator, which produces approximately 20,000 droplets per sample. Created droplet emulsion (40 µl) was then pipetted from the top wells of the cartridge into 96-well plate. The PCR plate was covered with pierceable foil and heat sealed using Bio-Rad's PX1. It was placed in a T100 thermal cycler (Bio-Rad Laboratories, Hercules, CA, USA), and the protocol was initialized with denaturation (95°C, 10 min), following 40 cycles of denaturation (94°C, 30 s), annealing/extension (50°C–62°C during optimization, then 56°C, 1 min) and droplet stabilization (98°C, 10 min.) with ramps of 2°C/s. After PCR, the product was held at 4°C with cooling ramp set up ~1°C/s, until the next step of analysis.

TABLE 1 Primer and probe sequences for *GRIA4* gene.

Primer	Sequence		Tm °C
Forward	5'-CACCACAACCACCACACACA-3'		55.2
Reverse	5'-CCTTACTTTCTCACATACACACAA-3'		54.6
Probe	Sequence	Tm°C	54.6
U-Probe	5'-CACCACAACCACCACACACA-3'	61.0	54.6
M-Probe	5'-CGCCGCGACCGCCACAC-3'	67.2	54.6

U-Probe is complementary to the unmethylated sequence of the *GRIA4* promoter, M-Probe is complementary to the methylated sequence of the *GRIA4* promoter.

2.4 Droplet analysis using QX200™ droplet reader and data interpretation in QuantaSoft™ software

After amplification, a 96-well plate was loaded to the QX200 Droplet Reader (Bio-Rad Laboratories, Hercules, CA, USA), where droplet analysis of each well was carried out. Each droplet was analyzed using QuantaSoft software (Bio-Rad Laboratories, Hercules, CA, USA) and divided into four clusters according to fluorescence emission analysis in HEX or FAM wavelengths. Droplets containing methylated DNA with high FAM amplitude, droplets containing unmethylated DNA with high HEX amplitude, and droplets with both types of DNA with high HEX and FAM amplitudes and empty droplets without target DNA. Data obtained from QX200 Droplet Reader were analyzed and interpreted by QuantaSoft v.1.7 Software (Bio-Rad Laboratories, Hercules, CA, USA). The correlation coefficient (R^2) was calculated from serial dilutions of 100% methylated EpiTect control DNA and 100% unmethylated control DNA into water with 8,000, 4,000, 2,000, 1,000, 500, 250, 125, and 62.5 copies/per reaction as a correlation between two variables, expected number of copies and detected number of copies. Selectivity of each probe was calculated from the analysis of diluted controls in a pair with the opposite probe, more precisely a methylated control with a U-probe and an unmethylated control with an M-probe. Detected copies were divided by number of copies detected for each control in pair with the complementary probe, more precisely a methylated control with an M-probe and an unmethylated control with a U-probe. Specificity was calculated for each dilution, and the final number was an average of all values determined. Threshold values were defined during assay development and optimization processes, 1,500 for FAM and 2,500 for HEX (genomic DNA) and 1,500 for FAM and 2,700 for HEX (circulating DNA). The quantity of methylation was expressed in percentage as the ratio of methylated sequences to the sum of methylated and unmethylated sequences.

2.5 Statistical analysis

Data were explored and analyzed in R, ver. 4.0.5, with the aid of different libraries (34–48). Data and R script to reproduce the presented results are available at Mendeley data repository (49). For exploratory data analysis, data were summarized by the mean, SD, min, quartiles, and max. Spaghetti plot was used to visualize the methylation values in pairs. The boxplot overlaid with swarmplot and quantile–quantile plot with the 95% confidence band constructed by bootstrap was used to assess normality of the data. For the regression model, linear mixed model (LMM) was used to implement repeated measures ANOVA with non-constant variance, i.e., to model the association between methylation and interaction of group and specimen. Non-homogeneity was taken into account using the weights in the `lme()` function of `nlme` library. For the specimen data, the methylation was log-transformed to obtain a fit passing diagnostic analysis. In addition, one subject with extremely high methylation in plasma was excluded from the data

prior to the model fitting. Effect size was quantified by the marginal and conditional R^2 . Marginal means were estimated given the fitted model and the grid of the factors. Contrasts, based on the *a priori* research questions, were specified, and the resulting p-values were adjusted using the Benjamini–Hochberg correction. The interaction plot was used to visualize the marginal means and their 95% confidence intervals. For the LMM model of methylation in plasma, the marginal means were back-transformed. Methylation in plasma before surgery for primary tumor vs. healthy controls was compared by the two-sample t-test, after assessing normality of the data by the quantile–quantile plot with the 95% confidence band constructed by bootstrap and by boxplot overlaid with swarmplot. For the case of metastasis, the methylation in plasma before surgery appeared skewed to the right; hence, the data were log-transformed prior to performing the two-sample t-test. EDA suggested that the log-transformation was appropriate. The cutoff on methylation to best separate between tumor and normal (separately in primary; in metastases) was obtained as the cutoff corresponding to the Youden index on the empirical receiver operating characteristic (ROC) curve.

3 Results

3.1 Droplet digital PCR as specific and sensitive method for detection of methylated DNA in tissue and plasma sample

Completely unmethylated and completely methylated DNA controls were tested with both probes (M-probe and U-probe) separately to identify optimal annealing temperature and temperatures when non-specific binding occurs. Within an annealing temperature range from 50.0°C to 62.0°C, the primer pair revealed clearly distinguishable fluorescence signals up to an annealing temperature of 55°C–56°C for both assays (Figure 1). The results of serial dilutions showed a linear correlation between individual dilutions, while R-squared value (R^2) was 0.9979 and 0.9972 for methylated and unmethylated control, respectively (Figure 2). The selectivity for the U-probe was 0.027 (2.7% background), being able to detect 1 unmethylated molecule in the background of 37 methylated. The M-probe with selectivity of 0.003 (0.3% background) is able to detect 1 methylated in the background of 333 unmethylated targets.

3.2 Methylation in *GRIA4* is significantly higher in tumor tissue compared to its adjacent non-tumor tissue.

Average *GRIA4* methylation in primary tumors (N=5) and metastases (N=12) was 42.85% (ranging from 36.85% to 52.81%) and 51.04% (ranging from 3.73% to 81.0%) (Supplementary Table S3), respectively, showing no significant difference between these two groups ($p = 0.342$). Analyzing tissues of patients with primary tumor

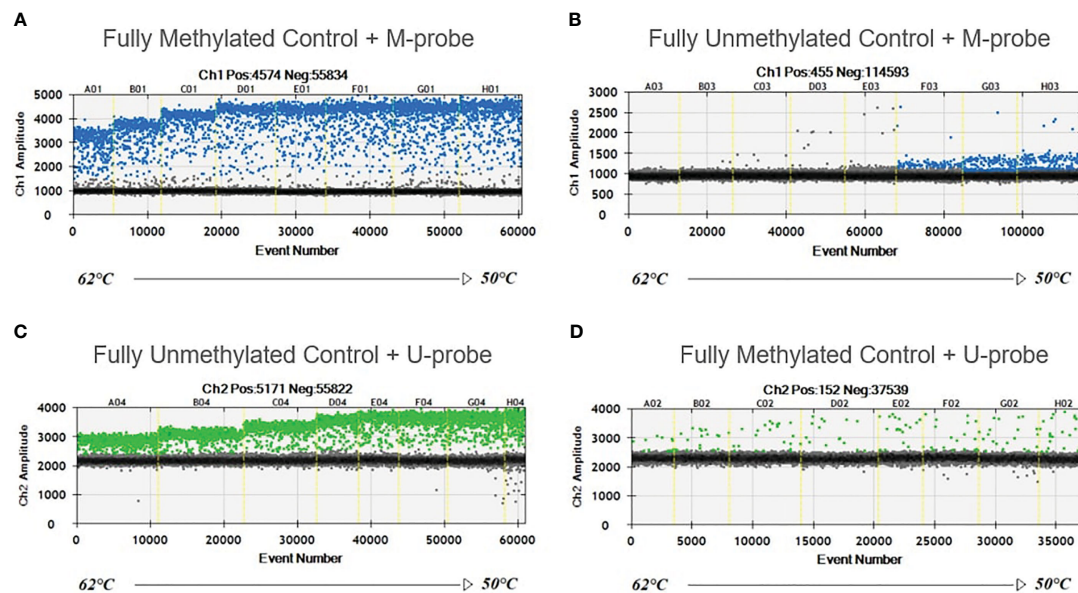


FIGURE 1

Temperature gradient for optimizing annealing temperature. Eight ddPCR reactions in temperature gradient ranging from 50°C to 62°C. Positive droplets at higher amplitude (blue or green) are methylated (A, B) or unmethylated (C, D). Negative droplets (gray) at low amplitude are without the amplification. Unmethylated control with M-probe (B) and methylated control with U-probe (D) were used to detect unspecific binding.

and patients with metastases separately, primary tumor tissue versus its adjacent non-tumor tissue (colon) showed three times average methylation gain in tumor tissue (Figure 3A). On the other hand, an even bigger difference was detected between metastatic tissue and its adjacent non-tumor tissue (liver) (Figure 3B), with 17.4 times hypermethylation decrease in liver. ROC analysis defined cutoff values to separate primary tumors (N=5) and metastasis (N=12) from non-tumor colon/rectum (N=5), as the colon/rectum is the tissue where tumors are derived from. Cutoff values were 36.85% for primary tumors (Figure 3C) and 34.81% for metastasis (Figure 3D). All primary tumors 5/5 (100%) were above this value; for metastases, 9/12 (75%) tissues samples had higher methylation than 34.81%.

3.3 Possibility to identify the tissue of origin of metastasis according to the methylation profile

Non-tumor tissue of primary tumor patients (colon/rectum), with average methylation of 14.02% (ranging from 2.09 to 24.92) (N=5) had almost five times higher *GRIA4* methylation status compared to non-tumor tissue of metastatic patients (liver), ranging from 1.44% to 4.87% (average 2.93%) (Supplementary Table S3), $p = 0.012$. When comparing the methylation status of metastatic tissue vs. non-tumor colon/rectum and liver, a smaller increase was observed in liver metastasis in combination with the colon tissue (3.6 times gain) than in liver metastasis in combination with adjacent liver tissue (17.4 times gain) (Figure 4). This finding indicates a stronger similarity between liver metastasis and colon/rectum, in contrast to the liver, given that the colon serves as the tissue of origin for the metastasis.

3.4 Increased *GRIA4* methylation status in patients' plasma compared to healthy individuals' plasma

The plasma sample of 22 patients, 6 primary tumor and 16 metastatic patients, taken before surgical removal of tumor was compared with 5 healthy volunteers' plasma. In primary tumor plasma, methylation percentage ranged between 0.70% and 1.66% (average, 1.30%); in metastatic plasma, it was 7.69% (ranging from 0.77% to 66.75%) and in healthy volunteers (N=5), values were between 0.12% and 0.87% (average, 0.47%) (Supplementary Table S3). Both primary (Figure 5A) and metastatic plasmas (Figure 5B) showed statistically significantly different mean methylation in comparison with healthy plasma, with $p = 0.039$ and $p = 0.0011$, respectively. *GRIA4* hypermethylation was, on average, 2.8 and 16.4 times higher in primary tumor and metastatic plasma than in healthy individuals.

3.5 Significant reduction of *GRIA4* methylation in post-surgery plasma of metastatic patients compared to pre-surgery plasma

Pre-surgery plasma samples showed *GRIA4* methylation ranging from 0.70% to 2.56% (average, 1.30%) in primary tumor patients; in metastatic patients, it ranged from 0.77% to 66.75% (average, 7.69%), and when comparing these two groups, hypermethylation in metastatic plasma was, on average, 5.9 times higher ($p = 0.051$). For post-surgery plasma, average *GRIA4* methylation was 0.80% (ranging from 0.22% to 1.17%) and 1.70%

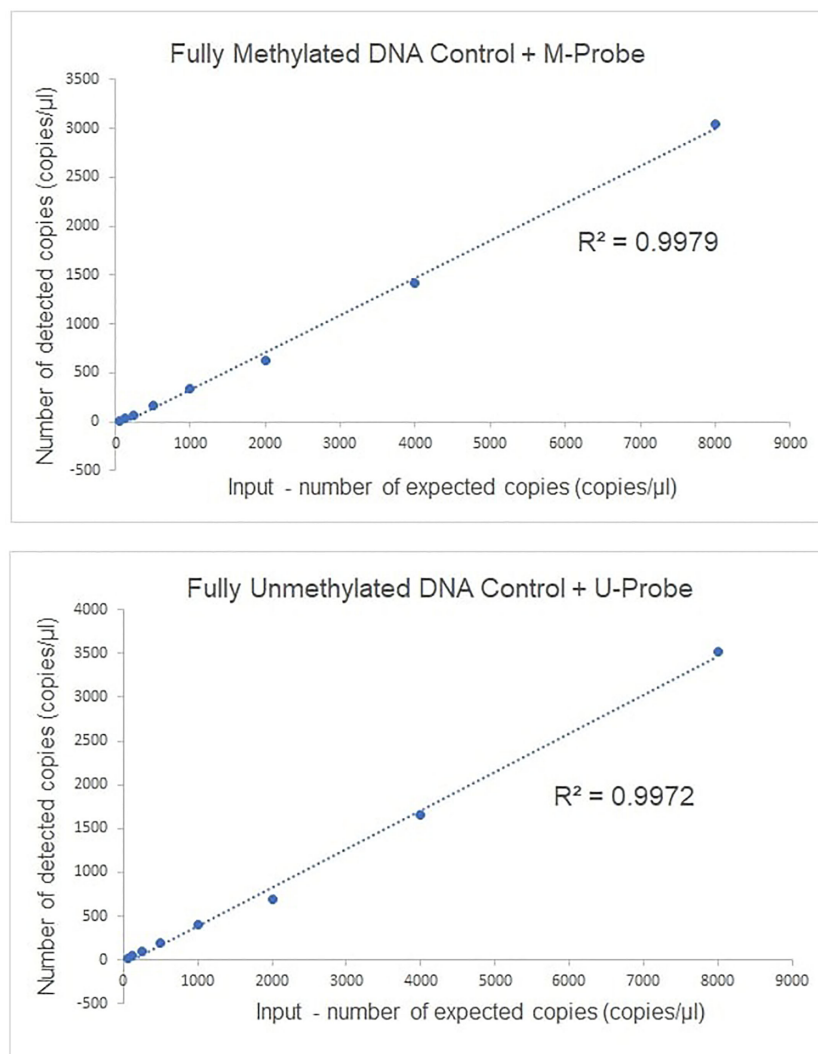


FIGURE 2

The linearity of ddPCR assays for methylated *GRIA4* detection and quantification. Linearity of ddPCR assay is shown using serially diluted EpiTech control DNAs (from 8,000, 4,000, 2,000, 1,000, 500, 250, 125, and 62.5 copies/per reaction, respectively).

(ranging from 0.28% to 10.44%) for primary and metastatic patients, respectively (Supplementary Table S3). When comparing pre- and post-surgery plasma of these two groups separately, there was an average of 1.6 times decrease in post-surgery plasma from primary tumor patients ($p=0.037$) (Figure 6A); on the other hand, metastatic patients' plasma showed, even bigger, 4.5 times decrease after surgical tumor removal ($p=0.023$) (Figure 6B).

4 Discussion

In terms of incidence, colorectal cancer is the third most common cancer worldwide, the second most prevalent in women and the third in men (1, 2). CRC development proceeds through the accumulation of genetic and epigenetic changes. DNA methylation, as the most common epigenetic modification, is an early step of colorectal carcinogenesis (50). In the last few decades, a large

number of cancer-related methylated genes have been identified in colorectal carcinoma patients (17–23).

In the current work, we focused on detecting hypermethylation status of gene *GRIA4* in CRC patients using primary tumor/metastatic tissue and complementary pre- and post-surgical plasma samples. We have decided to choose this gene based on the previous studies, where it showed high methylation status and possible detectability from different sample types, such as plasma or stool (27–29). Droplet digital PCR method was chosen based on high sensitivity and our previous experiences with this method and circulating DNA detection (51).

Our results showed significant *GRIA4* hypermethylation increase in primary tumors and liver metastases compared to their adjacent non-tumor tissues (colon, rectum, or liver). Average methylation increase was smaller in patients with primary tumor compared to metastatic patients (3× vs. 17.4× gain). Even if the metastasis is localized on the liver, it is not its

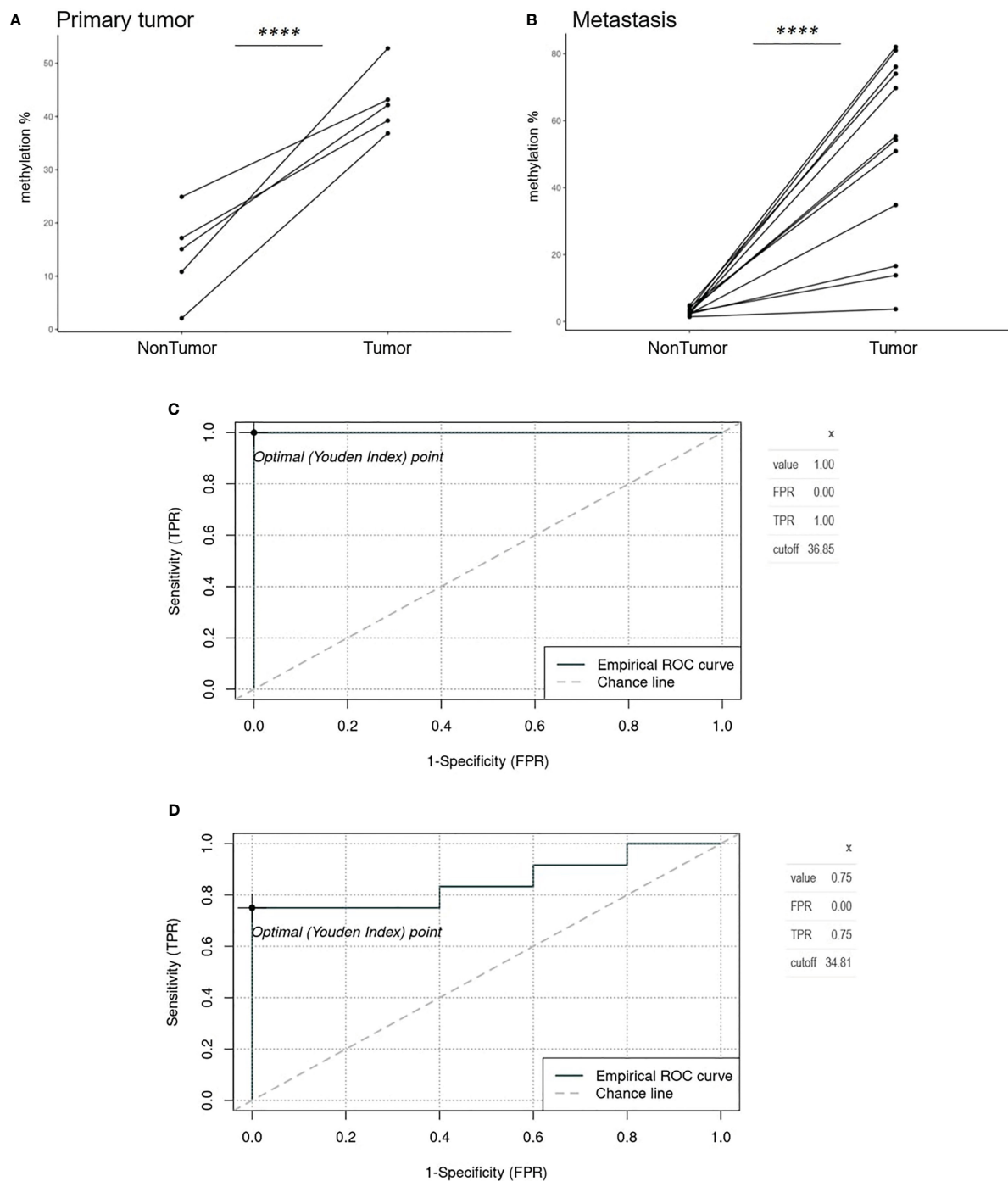


FIGURE 3

Spaghetti plots comparing tumor and non-tumor tissues of patients with primary tumor and metastasis and ROC analysis. (A) Primary tumor tissue versus adjacent non-tumor tissue (colon) ($p < 0.0001$). (B) Metastatic tissue and its adjacent non-tumor tissue (liver) ($p < 0.0001$). (C) Empirical ROC curve, Youden index, and cutoff for primary tumors ($N=5$) compared to colon non-tumor tissue ($N=5$). (D) Empirical ROC curve, Youden index, and cut-off for metastases ($N=12$) compared to colon non-tumor tissue ($N=5$). **** $p \leq 0.0001$.

tissue of origin, so we compared metastasis vs. non-tumor colon/rectum, even though the colon/rectum tissues were from other patients (P1_PT–P5_PT). A smaller gain (3.6 times) was observed between metastasis and non-tumor colon compared to metastasis and adjacent liver, indicating a higher degree of similarity between these two sample types. This similarity can be attributed to the fact that the colon is the tissue from which the metastasis originates.

Additionally, in the study by Barault et al. (27), it was *GRIA4* that was the most hypermethylated gene in non-tumor colon tissue compared to the other tested biomarkers, which confirms our findings that increased hypermethylation of this gene is present physiologically in the colon/rectum tissue. As methylation is highly tissue specific (52), it can be used to classify tumor subtypes, such as adrenocortical carcinoma (53), hepatocellular carcinoma (54), and

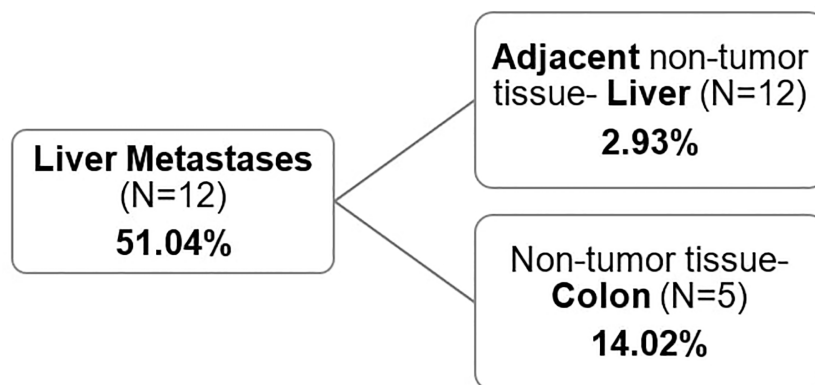


FIGURE 4

Average *GRIA4* methylation of metastasis and non-tumor tissue. Non-tumor tissue was either the adjacent non-tumor liver (N=12) or non-tumor colon/rectum from different set of patients (N=5).

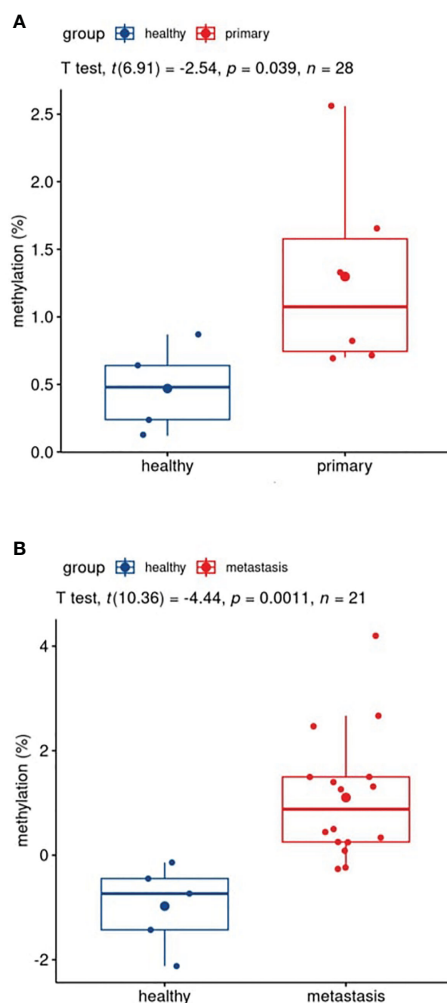


FIGURE 5

Boxplots of patients' versus healthy participants' plasma comparison. (A) Pre-surgery plasma of patients with primary tumor vs. plasma of healthy individuals; (B) pre-surgery plasma of patients with metastasis vs. plasma of healthy individuals.

cancer of unknown primary sites (55), and it can be used to identify the tissue of origin of metastasis (56). The fact that the cells at the metastatic site have similar methylation patterns suggests that they have retained some of the characteristics of the primary tumor and original tissue. It may be a useful tool to assign original site to metastasis, although additional analysis of primary tumor and metastases from same patient together with methylation status of other tissues and organs have to be performed to provide more precise data.

ROC analysis selected cutoff values to separate primary tumors and metastasis of clinically known colorectal cancer from non-tumor colon tissue. Our data suggest that all primary tumors (5/5) showed above-threshold methylation. These findings are complementary to results from previously published studies where tumors showed *GRIA4* hypermethylation in all 82 tissue samples, ranging from 18% to 97% (27), or in 99.1% (28) of 115 CRC tissues.

Liquid biopsy is a non-invasive way to obtain cancer-derived genetic material for a molecular analysis and monitor relapse or therapy response (12–16). We showed that *GRIA4* hypermethylation from the tumor was detectable in liquid biopsy plasma samples. Plasma of metastatic patients was, on average, 16.4 times higher than that of healthy individuals; in primary tumor patients, a 2.8 times gain was detected. Smaller difference in primary tumor patients can be due to the fact that the primary tumors came from the CRC patients with earlier stages, II (N=5) or III (N=1), and it is known that promoter hypermethylation correlates with tumor stage (57). Moreover, early tumor stages release less circulating DNA into the bloodstream (11, 51). However, in order to enhance future applications, it would be advantageous to conduct a comparative analysis between plasma and stool samples for the *GRIA4* methylation detection of CRC, as stool-based tests utilizing *SEPT9* methylation have demonstrated superior performance compared to plasma-based tests (58, 59). Furthermore, Vega-Benedetti and colleagues (29) proposed *GRIA4* as a potential biomarker for early CRC detection specifically from stool samples. Considering these findings, a comprehensive assessment comparing the diagnostic

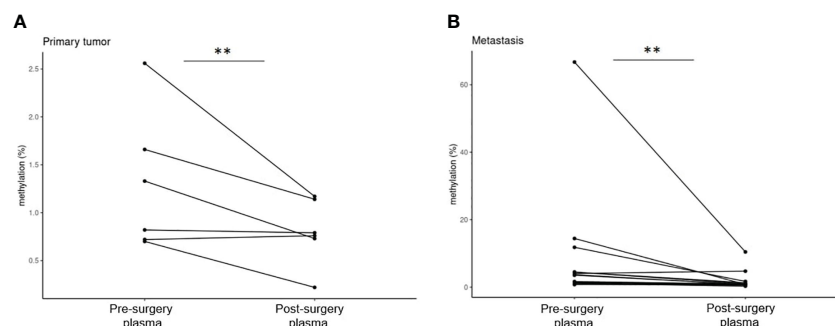


FIGURE 6

Comparison of pre- and post-surgery plasma of primary tumor and metastatic patients. (A) Pre- and post-surgery plasma of primary tumor patients ($p=0.037$); (B) pre- and post-surgery plasma of metastatic patients ($p=0.023$). $**p \leq 0.01$.

efficacy of plasma and stool samples would be valuable for advancing the field.

Subsequently, we compared *GRIA4* methylation status in plasma before and after surgical removal of primary tumor/metastasis. In both groups, a decrease in hypermethylation occurred; only three cases showed a slight increase in percentage (average, 0.25%). In the cohort of primary tumor patients, we observed a modest reduction of 1.6-fold in the *GRIA4* methylation status, which may potentially be attributed to the enrollment of individuals in the initial stages of the disease. For metastatic patients, we detected a significant decrease in hypermethylation in post-surgery plasma, by 4.5 times. Although, the current sample size employed in this study is relatively limited, necessitating the acquisition of a more extensive sample cohort to substantiate the role of hypermethylated *GRIA4* as a reliable biomarker for the early non-invasive detection of metastasis of clinically diagnosed cases of colorectal cancer. Additionally, it is important to note that there was substantial variability in methylation levels across the samples; therefore, it may be advantageous to explore the integration of complementary biomarkers in conjunction with *GRIA4*.

Significant *GRIA4* methylation decrease in post-surgery plasma indicates that this biomarker holds promise as a robust candidate for simple and cost-effective CRC detection using ddPCR, a common platform in oncology labs. Although it would be beneficial to incorporate additional biomarkers, as already mentioned above, our primary objective was to identify a specific biomarker that can contribute to the development of straightforward, single-gene tests like Epi proColonTM (24) or ColoSureTM (25). Further studies on already identified biomarkers could facilitate their progressive implementation into clinical diagnostics, as seen with the *SEPT9* or *Vimentin* genes (60–63).

5 Conclusion

This methylation-specific ddPCR assay proved to be a suitable detection method for capturing the hypermethylated *GRIA4* gene from conventional tissue as well as liquid biopsy samples. Our data suggest, that this biomarker could serve as a tool to identify colorectal cancer and its metastasis from both tissue and plasma samples, furthermore, it may aid in determining the specific tissue of origin for the metastatic lesions. We observed a significant increase in *GRIA4* methylation in the plasma of metastatic patients, with a remarkable 16.4-fold amplification, which nominates this gene as potential novel biomarker for an early non-invasive detection of metastasis of clinically known CRC, however, additional analysis of a larger sample cohort must be performed.

Data availability statement

The authors acknowledge that the data presented in this study must be deposited and made publicly available in an acceptable repository, prior to publication. Frontiers cannot accept a article that does not adhere to our open data policies. The datasets presented in this study are available on <https://data.mendeley.com/datasets/sm5n4bgtf3>.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics committee of the Jessenius Faculty of Medicine in Martin, Comenius University in Bratislava (number of approval 1863/2016). The patients/participants provided their written informed consent to participate in this study.

Author contributions

EL, TB, PP was responsible for experimental design, data analysis and wrote the manuscript, EL, TB, ET, IK, AV performed the experiments, ZL secured financial funding, was responsible for experimental design, supervised the data analysis and interpretation, and critically reviewed and revised the manuscript drafts, MG performed bioinformatic analyses, generated the figures and revised the manuscript drafts, LL, PM, EK, JM performed the histopathological evaluation and collected patient's samples. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by VEGA grants 1/0380/18 and 1/0205/20 and Comenius University Student Grant UK/25/2021 from the Ministry of Education, Science, Research, and Sport of the Slovak Republic and by The Slovak Research and Development Agency (grant no. APVV-16-0066).

References

- Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* (2015) 65:87–108. doi: 10.3322/caac.21262
- 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 Cancer J Clin* (2021) 71:209–49. doi: 10.3322/caac.21660
- de Ridder J, de Wilt JHW, Simmer F, Overbeek L, Lemmens V, Nagtegaal I. Incidence and origin of histologically confirmed liver metastases: an explorative case-study of 23,154 patients. *Oncotarget* (2016) 7:55368–76. doi: 10.18632/oncotarget.10552
- O'Reilly DA, Poston GJ. Colorectal liver metastases: current and future perspectives. *Future Oncol* (2006) 2:525–31. doi: 10.2217/14796694.2.4.525
- Hong SN. Genetic and epigenetic alterations of colorectal cancer. *Intest Res* (2018) 16:327–37. doi: 10.5217/ir.2018.16.3.327
- Kulis M, Esteller M. DNA Methylation and cancer. *Adv Genet* (2010) 70:27–56. doi: 10.1016/B978-0-12-380866-0.60002-2
- Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci USA* (1999) 96:8681–6. doi: 10.1073/pnas.96.15.8681
- Lu T, Li J. Clinical applications of urinary cell-free DNA in cancer: current insights and promising future. *Am J Cancer Res* (2017) 7:2318–32.
- Ponti G, Manfredini M, Tomasi A. Non-blood sources of cell-free DNA for cancer molecular profiling in clinical pathology and oncology. *Crit Rev Oncol Hematol* (2019) 141:36–42. doi: 10.1016/j.critrevonc.2019.06.005
- Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, et al. DNA Fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res* (2001) 61:1659–65.
- Speicher MR, Pantel K. Tumor signatures in the blood. *Nat Biotechnol* (2014) 32:441–3. doi: 10.1038/nbt.2897
- Parikh AR, Leshchiner I, Elagina L, Goyal L, Levovitz C, Siravegna G, et al. Author correction: liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers. *Nat Med* (2019) 25:1949. doi: 10.1038/s41591-019-0698-6
- Chu D, Paoletti C, Gersch C, VanDenBerg DA, Zabransky DJ, Cochran RL, et al. ESR1 mutations in circulating plasma tumor DNA from metastatic breast cancer Patients/ESR1 mutations detected in circulating tumor DNA. *Clin Cancer Res* (2016) 22:993–9. doi: 10.1158/1078-0432.CCR-15-0943
- Moss J, Magenheimer J, Neiman D, Zemmour H, Loyfer N, Korach A, et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat Commun* (2018) 9:5068. doi: 10.1038/s41467-018-07466-6
- Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* (2014) 6:224ra24. doi: 10.1126/scitranslmed.3007094
- Tie J, Kinde I, Wang Y, Wong HL, Roebert J, Christie M, et al. Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer. *Ann Oncol* (2015) 26:1715–22. doi: 10.1093/annonc/mdv177
- Petko Z, Ghiassi M, Shuber A, Gorham J, Smalley W, Washington MK, et al. Aberrantly methylated CDKN2A, MGMT, and MLH1 in colon polyps and in fecal DNA from patients with colorectal polyps. *Clin Cancer Res* (2005) 11:1203–9. doi: 10.1158/1078-0432.1203.11.3
- Psofaki V, Kalogera C, Tzambouras N, Stephanou D, Tsianos E, Seferiadis K, et al. Promoter methylation status of hMLH1, MGMT, and CDKN2A/p16 in colorectal adenomas. *World J Gastroenterol* (2010) 16:3553–60. doi: 10.3748/wjg.v16.i28.3553
- Tang D, Liu J, Wang DR, Yu HF, Li YK, Zhang JQ. Diagnostic and prognostic value of the methylation status of secreted frizzled-related protein 2 in colorectal cancer. *Clin Invest Med* (2011) 34:E88–95. doi: 10.25011/cim.v34i1.15105
- Chen WD, Han ZJ, Skoletsky J, Olson J, Sah J, Myeroff L, et al. Detection in fecal DNA of colon cancer-specific methylation of the nonexpressed vimentin gene. *J Natl Cancer Inst* (2005) 97:1124–32. doi: 10.1093/jnci/dji204
- Loh K, Chia JA, Greco S, Cozzi SJ, Buttenshaw RL, Bond CE, et al. Bone morphogenic protein 3 inactivation is an early and frequent event in colorectal cancer development. *Genes Chromosomes Cancer* (2008) 47:449–60. doi: 10.1002/gcc.20552
- Lofton-Day C, Model F, Devos T, Tetzner R, Distler J, Schuster M, et al. DNA Methylation biomarkers for blood-based colorectal cancer screening. *Clin Chem* (2008) 54:414–23. doi: 10.1373/clinchem.2007.095992
- Melotte V, Lentjes MHFM, van den Bosch SM, Hellebrekers DMEI, de Hoon JPJ, Wouters KAD, et al. N-myc downstream-regulated gene 4 (NDRG4): a candidate tumor suppressor gene and potential biomarker for colorectal cancer. *J Natl Cancer Inst* (2009) 101:916–27. doi: 10.1093/jnci/djp131
- Church TR, Wandell M, Lofton-Day C, Mongin SJ, Burger M, Payne SR, et al. Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut* (2014) 63:317–25. doi: 10.1136/gutjnl-2012-304149
- Ned RM, Melillo S, Marrone M. Fecal DNA testing for colorectal cancer screening: the ColoSure™ test. *PLoS Curr* (2011) 3:RRN1220. doi: 10.1371/currents.RRN1220
- Ahlquist DA. Multi-target stool DNA test: a new high bar for noninvasive screening. *Dig Dis Sci* (2015) 60:623–33. doi: 10.1007/s10620-014-3451-5
- Barault L, Amatu A, Siravegna G, Ponzetti A, Moran S, Cassingena A, et al. Discovery of methylated circulating DNA biomarkers for comprehensive non-invasive monitoring of treatment response in metastatic colorectal cancer. *Gut* (2018) 67:1995–2005. doi: 10.1136/gutjnl-2016-313372

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.

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/fonc.2023.1205791/full#supplementary-material>

28. Hauptman N, Jevšinek Skok D, Spasovska E, Boštjančič E, Glavač D. Genes CEP55, FOXD3, FOXF2, GNAO1, GRIA4, and KCNA5 as potential diagnostic biomarkers in colorectal cancer. *BMC Med Genomics* (2019) 12:54. doi: 10.1186/s12920-019-0501-z
29. Vega-Benedetti AF, Loi E, Moi L, Orrù S, Ziranu P, Pretta A, et al. Colorectal cancer early detection in stool samples tracing CpG islands methylation alterations affecting gene expression. *Int J Mol Sci* (2020) 21:4494. doi: 10.3390/ijms21124494
30. Sun X, Chen D, Jin Z, Chen T, Lin A, Jin H, et al. Genome-wide methylation and expression profiling identify methylation-associated genes in colorectal cancer. *Epigenomics* (2019) 12:19–36. doi: 10.2217/epi-2019-0133
31. Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem* (2011) 83:8604–10. doi: 10.1021/ac202028g
32. Cao Y, Raith MR, Griffith JF. Droplet digital PCR for simultaneous quantification of general and human-associated fecal indicators for water quality assessment. *Water Res* (2015) 70:337–49. doi: 10.1016/j.watres.2014.12.008
33. Brierley JD, Gospodarowicz MK, Wittekind C. *TNM Classification of Malignant Tumours, 8th Edition*. Chichester: John Wiley & Sons (2017), 272 p.
34. Signorelli A, Aho K, Alfons A, Anderegg N, Aragon T, Arachchige C, et al. *DescTools: tools for Descriptive Statistics*. R package version 0.99.41. (2021). Available at: <https://cran.r-project.org/package=DescTools>.
35. Kassambara A. *Ggpubr: 'ggplot2' based publication ready plots*. R package version 0.4.0.999. (2020). Available at: <https://rpkgs.datanovia.com/ggpubr>.
36. Kassambara A. *Rstatix: pipe-friendly framework for BasicStatistical tests*. R package version 0.7.0. (2021). Available at: <https://CRAN.R-project.org/package=rstatix>.
37. Khan M, Brandenburger T. *ROCit: performance assessment of Binary classifier with visualization*. R package version 2.1.1. (2020). Available at: <https://CRAN.R-project.org/package=ROCit>.
38. Lenth RV, Bolker B, Buerkner P, Gine-Vazquez I, Herve M, Jung M, et al. *Emmeans: estimated marginal means, aka least-squares Means*. R package version 1.7.5. (2022). Available at: <https://CRAN.R-project.org/package=emmeans>.
39. Lüdtke D, Bartel A, Schwemmer C, Powell C, Djalovski A, Titz J. *sjPlot: data visualization for statistics in Social Science*. R package version 2.8.12. (2013). Available at: <https://CRAN.R-project.org/package=sjPlot>.
40. Lüdtke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D. Performance: an R package for assessment, comparison and testing of statistical models. *J Open Source Software* (2021) 6(60). doi: 10.21105/joss.03139
41. Makowski D, Lüdtke D, Patil I, Thériault R, Ben-Shachar MS, Wiernik BM. Automated Results reporting as a practical tool to improve reproducibility and Methodological best practices adoption. CRAN (2023). Available at: <https://github.com/easystats/report>.
42. Ogle DH, Doll JC, Wheeler AP, Dinno A. *FSA: Simple Fisheries Stock assessment methods*. R package version 0.9.4. (2023). Available at: <https://CRAN.R-project.org/package=FSA>.
43. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. *nlme: Linear and nonlinear mixed effects models*. R package version 3.1–152 (2021). Available at: <https://CRAN.R-project.org/package=nlme>.
44. R Core Team. *R: a language and environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing (2021). Available at: <https://www.R-project.org/>.
45. Wickham H. *ggplot2: elegant graphics for data analysis*. Springer-Verlag New York (2016). Available at: <https://ggplot2.tidyverse.org>.
46. Wickham H, François R, Henry L, Müller K. *Dplyr: a grammar of Data manipulation*. R package version 1.0.10. (2022). Available at: <https://CRAN.R-project.org/package=dplyr>.
47. Xiao N. *Ggsci: scientific journal and sci-fi themed Color Palettes for 'ggplot2'*. R package version 2.9. (2018). Available at: <https://CRAN.R-project.org/package=ggsci>.
48. Zhu H, Trivison T, Tsai T, Beasley W, Xie Y, Yu C, et al. *kableExtra: construct complex table with 'kable' and Pipe syntax*. R package version 1.3.4. (2021). Available at: <https://CRAN.R-project.org/package=kableExtra>.
49. Lukacova E, Burjanivova T, Podlesniy P, Grendar M, Turyova E, Kasubova I, et al. Supplement to hypermethylated GRIA4, potential biomarker for an early non-invasive detection of metastasis of clinically known colorectal cancer. *Mendeley Data* (2023) V1. doi: 10.17632/sm5n4bgtf3.1
50. Menigatti M, Pedroni M, Verrone AM, Borghi F, Scarselli A, Benatti P, et al. O6-methylguanine-DNA methyltransferase promoter hypermethylation in colorectal carcinogenesis. *Oncol Rep* (2007) 17:1421–7. doi: 10.3892/or.17.6.1421
51. Burjanivova T, Malicherova B, Grendar M, Minarikova E, Dusenka R, Vanova B, et al. Detection of BRAFV600E mutation in melanoma patients by digital PCR of circulating DNA. *Genet Test Mol Biomarkers* (2019) 23:241–5. doi: 10.1089/gtmb.2018.0193
52. Løkk K, Modhukur V, Rajashekar B, Märtens K, Mägi R, Kolde R, et al. DNA Methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol* (2014) 15:r54. doi: 10.1186/gb-2014-15-4-r54
53. Assié G, Letouze E, Fassnacht M, Jouinot A, Luscip W, Barreau O, et al. Integrated genomic characterization of adrenocortical carcinoma. *Nat Genet* (2014) 46:607–12. doi: 10.1038/ng.2953
54. Xu RH, Wei W, Krawczyk M, Wang W, Luo H, Flagg K, et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat Mater* (2017) 16:1155–61. doi: 10.1038/nmat4997
55. Moran S, Martínez-Cardús A, Sayols S, Musulén E, Balañá C, Estival-Gonzalez A, et al. Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. *Lancet Oncol* (2016) 17:1386–95. doi: 10.1016/S1470-2045(16)30297-2
56. Liu H, Chen J, Chen H, Xia J, Wang O, Xie J, et al. Identification of the origin of brain metastases based on the relative methylation orderings of CpG sites. *Epigenetics* (2021) 16:908–16. doi: 10.1080/15592294.2020.1827720
57. Bilgrami SM, Qureshi SA, Pervez S, Abbas F. Promoter hypermethylation of tumor suppressor genes correlates with tumor grade and invasiveness in patients with urothelial bladder cancer. *Springerplus* (2014) 3:178. doi: 10.1186/2193-1801-3-178
58. Ahlquist DA, Taylor WR, Mahoney DW, Zou H, Domanico M, Thibodeau SN, et al. The stool DNA test is more accurate than the plasma septin 9 test in detecting colorectal neoplasia. *Clin Gastroenterol Hepatol* (2012) 10:272–7. doi: 10.1016/j.cgh.2011.10.008
59. Liu Y, Zhao G, Miao J, Li H, Ma Y, Liu X, et al. Performance comparison between plasma and stool methylated SEPT9 tests for detecting colorectal cancer. *Front Genet* (2020) 11:324. doi: 10.3389/fgene.2020.00324
60. Payne SR. From discovery to the clinic: the novel DNA methylation biomarker m SEPT9 for the detection of colorectal cancer in blood. *Epigenomics* (2010) 2:575–85. doi: 10.2217/epi.10.35
61. Song L, Li Y. Progress on the clinical application of the SEPT9 gene methylation assay in the past 5 years. *biomark Med* (2017) 11:415–8. doi: 10.2217/bmm-2017-0091
62. Shirahata A, Sakata M, Sakuraba K, Goto T, Mizukami H, Saito M, et al. Vimentin methylation as a marker for advanced colorectal carcinoma. *Anticancer Res* (2009) 29:279–81.
63. Shirahata A, Hibi K. Serum vimentin methylation as a potential marker for colorectal cancer. *Anticancer Res* (2014) 34:4121–5.



OPEN ACCESS

EDITED BY

Pooneh Mokarram,
Shiraz University of Medical Sciences, Iran

REVIEWED BY

Kangkang Wan,
Wuhan Ammunition Life-tech Company
Ltd., China
Huayang Wang,
Shandong University, China

*CORRESPONDENCE

Boyu Tian

✉ tianby@sysucc.org.cn

Siyu Xiao

✉ xiaosy9@mail.sysu.edu.cn

Jiahui Huang

✉ huangjh48@mail.sysu.edu.cn

RECEIVED 15 February 2023

ACCEPTED 18 July 2023

PUBLISHED 09 August 2023

CITATION

Zeng T, Huang Z, Yu X, Zheng L, Liu T,
Tian B, Xiao S and Huang J (2023)
Combining methylated SDC2 test in
stool DNA, fecal immunochemical test,
and tumor markers improves early
detection of colorectal neoplasms.
Front. Oncol. 13:1166796.
doi: 10.3389/fonc.2023.1166796

COPYRIGHT

© 2023 Zeng, Huang, Yu, Zheng, Liu, Tian,
Xiao and Huang. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/)
(CC BY). 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.

Combining methylated SDC2 test in stool DNA, fecal immunochemical test, and tumor markers improves early detection of colorectal neoplasms

Tao Zeng^{1,2}, Zhongchao Huang^{1,2}, Xufa Yu^{1,2}, Li Zheng^{1,2},
Tao Liu^{2,3}, Boyu Tian^{4,5*}, Siyu Xiao^{1,2*} and Jiahui Huang^{1,2*}

¹Department of Clinical Laboratory, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China, ²Biomedical Innovation Center, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China, ³Department of Gastroenterology, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China, ⁴Department of Clinical Laboratory, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, Guangdong, China, ⁵State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, Guangdong, China

Objective: To explore the value of testing methylated SDC2 (SDC2) in stool DNA combined with fecal immunochemical test (FIT) and serum tumor markers (TM) for the early detection of colorectal neoplasms.

Methods: A total of 533 patients, including 150 with CRC (67 with early-stage CRC), 23 with APL, 85 with non-advanced adenomas and general polyps, and 275 with benign lesions and healthy controls. SDC2 was detected by methylation-specific PCR, FIT (hemoglobin, Hb and transferrin, TF) was detected by immunoassay, and the relationships between SDC2, FIT, and clinicopathological features were analyzed. Pathological biopsy or colonoscopy were used as gold standards for diagnosis, and the diagnostic efficacy of SDC2 combined with FIT and TM in CRC and APL evaluated using receiver operating characteristic (ROC) curves.

Results: SDC2 positive rates in early-stage CRC and APL were 77.6% (38/49) and 41.2% (7/17), respectively, and combination of SDC2 with FIT increased the positive rates to 98.0% (48/49) and 82.4% (14/17). The positive rates of SDC2 combined with FIT assay in the APL and CRC groups at stages 0-IV were 82.4% (14/17), 85.7% (6/7), 100% (16/16), 100% (26/26), 97.4% (38/39), and 100% (22/22), respectively. Compared to the controls, both the CRC and APL groups showed significantly higher positive detection rates of fecal SDC2 and FIT ($\chi^2 = 114.116$, $P < 0.0001$ and $\chi^2 = 85.409$, $P < 0.0001$, respectively). Our results demonstrate a significant difference in the qualitative methods of SDC2 and FIT for the detection of colorectal neoplasms (McNemar test, $P < 0.0001$). ROC curve analysis revealed that the sensitivities of SDC2 and FIT, alone or in combination, for the detection of early CRC and APL were 69.9%, 86.3%, and

93.9%, respectively (all $P < 0.0001$). When combined with CEA, the sensitivity increased to 97.3% ($P < 0.0001$).

Conclusions: SDC2 facilitates colorectal neoplasms screening, and when combined with FIT, it enhances detection. Furthermore, the combination of SDC2 with FIT and CEA maximizes overall colorectal neoplasm detection.

KEYWORDS

SDC2, fecal immunochemical test, colorectal neoplasms, tumor markers, combined testing

1 Introduction

Colorectal cancer (CRC) is a common malignant tumor of the digestive system and one of the most frequently occurring malignant tumors in China (1). Abnormal cells carrying mutations or methylation signals are shed into the stool and can be detected, making stool a theoretically better specimen than blood for early detection of CRC (2–4). Advanced adenoma (AA) and advanced serrated polyp (ASP) are currently considered important advanced precancerous lesions (APL) of CRC, because they increase the risk of CRC disease and death (4–9). Progression from APL to CRC is thought to occur over 5–10 years, which provides a valuable window of time for disease diagnosis and intervention (10). Current guidelines recommend CRC screening methods, such as colonoscopy and fecal occult blood test (FOBT), but colonoscopy requires specialized physicians and is somewhat invasive, making it unsuitable for large-scale population screening. Therefore, there is an urgent need for non-invasive tests to improve screening rates. The fecal immunochemical test (FIT) is non-invasive, simple, highly sensitive and is now widely used in the clinic. In recent years, fecal DNA testing has developed rapidly, and fecal DNA methylation testing has undergone numerous prospective clinical trials in China, confirming its suitability for CRC screening (11–13). Some studies have shown that the specificity of fecal SDC2 methylation (SDC2) was very high, or even $> 90\%$, and had good diagnostic agreement with colonoscopy results, indicating that SDC2 is a suitable marker for CRC screening (11, 14–16); however, the clinical application of SDC2 remains somewhat controversial (17, 18). There have been few studies analyzing the relationships between clinicopathological features and SDC2 results. Moreover, comparisons of SDC2 testing with early screening for FIT and combined screening are infrequent. In this study, we analyzed the usefulness of fecal SDC2 in colorectal neoplasm screening by testing it in patients who underwent colonoscopy, and we combined this with clinical and pathological data. We also compared the results of SDC2 testing with the routine clinical application of FIT to investigate whether it could detect more early-stage CRC and APL. It is well-established that combining multiple markers for screening can improve tumor detection rates. For example, blood protein markers combined with DNA mutations can be used to detect various early-stage cancers (19). Considering the widespread use of serum tumor markers (TM) in clinical practice, we

also investigated the efficacy of combining SDC2 and FIT with TM testing for early-stage CRC and APL detection.

2 Methods

2.1 Patients and methods

This was a retrospective case-control study. Outpatients and inpatients who underwent colonoscopy and/or pathological examination at The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, from March 2019 to September 2022 were recruited to this study. A total of 533 patients were included: 150 cases of CRC, comprising 138 with stage I–IV and 12 cases of carcinoma *in situ* and intramucosal CRC; 23 cases of APL, including 21 cases of AA and 2 cases of ASP; 85 with non-AA and general polyps (mainly inflammatory polyps and hyperplastic polyps); and 275 with benign lesions and healthy controls. Among patients with CRC, 67 had early-stage disease, including 12 with stage 0, 22 with stage I, and 33 with stage II; 79 had advanced CRC (stage III and IV); and 4 cases had disease of unknown stage.

All patients were first-time outpatients or inpatients who had not received relevant treatment (including drugs and surgery), and colonoscopy or pathological findings were used as the gold standards for diagnosis, with CRC diagnosed according to the Chinese Colorectal Cancer Diagnosis and Treatment Standard (20) and CRC TNM staging according to the TNM staging system for colorectal cancer in the 8th edition of the AJCC (21).

Study exclusion criteria were as follows: 1. suffering from other systemic malignancies; 2. previous history of colorectal neoplasms, or had undergone colonoscopy resection or surgical treatment of CRC; 3. had undergone comprehensive treatment for CRC; 4. no pathological or colonoscopy findings, including cases in which polyps were found but not treated; 5. patients with incomplete clinical information and unknown diagnosis, of which eight cases with neither SDC2 nor FIT findings were also excluded; 5. cases with inadequate bowel preparation (inadequate cases); or 6. gastrointestinal mesenchymal tumors and presacral tumors.

This study was approved by the Ethics Committee of The Sixth Affiliated Hospital of Sun Yat-sen University (Ethics No. 2022ZSLYEC-508).

2.2 Instruments and methods

The instructions for specific tests are provided in the attached Instructions Summary; a brief description of each test method is also provided below.

Fecal SDC2 was detected using the human SDC2 gene methylation detection kit (methylation-specific PCR (22) method) from Creative Biosciences (Guangzhou) Co., Ltd. First, samples were extracted using magnetic beads and then treated with sulfite; following which, the methylated SDC2 gene would not be transformed, while methylated SDC2 could be amplified by specific primers, with ACTB as the internal reference gene. Samples were judged to be positive when the Ct value of the ACTB gene was ≤ 36 and that for SDC2 was ≤ 38 . The specific detection principles and steps, as well as influencing factors, were detailed in a previous report (14).

FIT is an immunoassay used to analyze fecal occult blood, including tests for hemoglobin (Hb) and transferrin (TF), using reagents and instruments from Keyu Biosciences (Zhuhai) Co., Ltd. The presence of two red lines on the test card (i.e., a quality control line and detection line) indicated a positive occult blood test. The sensitivity values of the Hb and TF tests were ≥ 100 ng/ml and ≥ 40 ng/ml, respectively, and the Hb and TF tests did not cross-react; positivity for either test was defined as a positive FIT result. Serum TM (included CEA, CA125, CA19-9, CA15-3, and AFP) were tested using Abbott Alinity and corresponding reagents and instruments, CEA results exceeding 5ng/ml are considered positive.

The kits used have obtained Chinese registration certificates and manufacturing licenses, and all operations are carried out in strict compliance with the operating manual. In addition, the specimens undergo indoor quality control before being tested.

2.3 Statistical analysis

IBM SPSS 26.0 statistical software was used for data analysis. Count data are expressed as frequencies and percentages, and the χ^2 test was used for comparison between two groups. McNemar's test was utilized to analyze the consistency and differences between the two qualitative diagnostic test methods. The sensitivity and specificity for qualitative results are calculated using a crosstab, while the sensitivity and specificity for quantitative results are calculated using logistic regression. The diagnostic efficacy of each diagnostic technique for CRC and APL was evaluated using receiver operating characteristic (ROC) curves. Mann-Whitney tests or t-tests were used for comparisons of quantitative data. All tests were two-sided. $P < 0.05$ was considered statistically significant.

3 Results

3.1 Patient characteristics

A total of 533 cases were included in the study: 173 cases of CRC and APL in the disease group, including 150 cases of CRC (67 cases

of early-stage CRC) and 23 cases of APL; and 360 disease controls and healthy controls, including 85 cases of non-AA and general polyps and 275 cases of benign lesions and healthy controls.

3.2 Stool DNA test of methylated SDC2

The positive rate of fecal SDC2 in CRC was 79.3%, including 77.6% in early-stage CRC, 82.0% in advanced CRC, and 41.2% in APL, with a gradual increasing trend of positive rate with disease severity: APL < early-stage CRC < advanced CRC (Table 1).

3.3 Consistency of fecal SDC2 with colonoscopy or pathological findings

A total of 518 cases were included in comparison of the results of SDC2 testing with those of colonoscopy or pathology findings, including 158 cases of CRC and APL. The Kappa value was 0.54 ($\chi^2 = 153.422$, $P < 0.001$), indicating that SDC2 was consistent with the results of colonoscopy with moderate concordance.

3.4 Comparison of fecal SDC2 with FIT

The positive rates of fecal SDC2 in the CRC plus APL group and the early-stage CRC plus APL group were 74.2% and 68.2%, respectively, while those for routine FIT were 89.8% and 86.4%, respectively (Table 1).

Conventional FIT demonstrated a higher positive detection rate for early-stage and advanced CRC and APL than fecal SDC2. In comparison to 180 healthy and disease controls, both the CRC and APL groups had significantly higher positive detection rates of fecal SDC2 and FIT ($\chi^2 = 114.116$, $P < 0.0001$ and $\chi^2 = 85.409$, $P < 0.0001$, respectively). Additionally, significant differences were observed in the positive detection rates of the early-stage CRC and APL groups ($\chi^2 = 69.640$, $P < 0.0001$ and $\chi^2 = 46.462$, $P < 0.0001$, respectively).

In comparison to the differences in positivity rates between different groups, we were curious about the consistency and differences in positivity rates between SDC2 and FIT qualitative methods in CRC and APL (methodological comparison). Using the McNemar test, our analysis revealed notable distinctions in positivity rates between the two methods (Difference: 20.13%, 95% CI: 13.91 - 26.35, $P < 0.0001$). We also found differences in early-stage CRC and APL groups (Difference: 21.95%, 95% CI: 14.64 - 29.26, $P < 0.0001$). Our results demonstrate a significant difference in the qualitative methods of SDC2 and FIT for the detection of colorectal neoplasms. Furthermore, the combination of these two methods enhances the positive detection rate.

3.5 Combined fecal SDC2 and FIT

The combined detection rates of fecal SDC2 and FIT were 95.3% (122/128) in CRC plus APL, and 93.9% (62/66) in early-stage CRC

TABLE 1 Clinical data from cases and controls undergoing fecal SDC2 and FIT testing.

	No.	SDC2		FIT ¹ (Hb and TF)		SDC2 and FIT ⁴	
		Positive NO.	Sensitivity (95% CI) %	Positive NO.	Sensitivity (95% CI) %	Positive NO.	Sensitivity (95% CI) %
Colorectal cancer	111	88	79.3 (70.5 - 86.4)	104	93.7 (87.4 - 97.4)	108	97.3 (92.3 - 99.4)
Early-Stage CRC	49	38	77.6 (63.4 - 88.2)	46	93.8 (83.1 - 98.7)	48	98.0 (89.1 - 99.9)
Stage 0	7	6	85.7 (42.1 - 99.6)	4	57.1 (18.4 - 90.1)	6	85.7 (42.1 - 99.6)
Stage I	16	15	93.8 (69.8 - 99.8)	16	100 (79.4 - 100.0)	16	100 (79.4 - 100.0)
Stage II	26	17	65.4 (44.3 - 82.8)	26	100 (86.8 - 100.0)	26	100 (86.8 - 100.0)
Advanced CRC	61	50	82.0 (70.0 - 90.6)	58	95.1 (86.3 - 99.0)	60	98.4 (91.2 - 100.0)
Stage III	39	30	76.9 (60.7 - 88.9)	37	94.9 (82.7 - 99.4)	38	97.4 (86.5 - 99.9)
Stage IV	22	20	90.9 (70.8 - 98.9)	21	95.5 (77.2 - 99.9)	22	100 (84.6 - 100.0)
Stage unknown	1	0	0	0	0	0	0
APL ²	17	7	41.2 (18.4 - 67.1)	11	64.7 (38.3 - 85.8)	14	82.4 (56.6 - 96.2)
Early-stage CRC and APL	66	45	68.2 (55.6 - 79.1)	57	86.4 (75.7 - 93.6)	62	93.9 (85.2 - 98.3)
CRC and APL	128	95	74.2 (65.7 - 81.5)	115	89.8 (83.3 - 94.5)	122	95.3 (90.1 - 98.3)
			Specificity (95% CI)		Specificity (95% CI)		Specificity (95% CI)
Non-advanced adenoma and general polyp	24	4	83.3 (62.6 - 95.3)	6	75.0 (53.3 - 90.2)	9	62.5 (40.6 - 81.2)
Benign lesions and negative results on colonoscopy ³	156	21	86.5 (80.2 - 91.5)	61	60.9 (52.8 - 68.6)	74	52.6 (44.4 - 60.6)

¹Any single positive result (Hb or TF) was considered a positive result. ²Including advanced adenoma and ≥ 1 cm serrated polyps. ³Benign lesions, including colitis, ulcerative colitis, Crohn's disease, intestinal tuberculosis, diverticulum, mixed hemorrhoids, anal fistula, etc. ⁴If either SDC2 or FIT has a positive result, it is considered positive. CI, Confidence interval.

plus APL (Table 1). Compared with the individual detection methods, the combined test significantly increased the positive detection rates of early-stage CRC, advanced CRC, and APL (98.0%, 98.4%, and 82.4%, respectively), suggesting that the two methods are complementary, and can detect more patients with APL and CRC when used in combination.

3.6 Associations between fecal SDC2 and FIT results and clinicopathological findings

3.6.1 CRC histological type: adenocarcinoma versus mucinous adenocarcinoma

Adenocarcinoma accounted for the majority of CRC cases included in this study, with only a small proportion being mucinous adenocarcinoma. Both SDC2 and FIT yielded higher positivity rates in adenocarcinoma than in mucinous adenocarcinoma, but their positivity rates did not differ significantly between the two types.

Additionally, the combined test improved the detection rate for the two types (Table 2; Figure 1).

3.6.2 Degree of CRC differentiation: highly and medium versus poor

The positive detection rate for fecal SDC2 or conventional FIT was low in poorly differentiated CRC, with 66.7% and 88.9%, respectively. Conversely, in medium and highly differentiated CRC, the positive detection rate for fecal SDC2 or FIT was higher, with 79.8% and 97.9%, respectively. There was no significant difference in the positivity rate of the combined SDC2 and FIT assay between medium and highly differentiated CRC compared to poorly differentiated CRC ($\chi^2 = 4.355$, $P = 0.168$).

3.6.3 Proximal versus distal CRC

Both fecal SDC2 and FIT had higher positive rates in proximal tumors than in distal tumors. There was no significant difference in the positivity rate of the combined SDC2 and FIT assay between

TABLE 2 Comparisons of fecal SDC2 and FIT positive rates with different pathological results in patients with CRC.

	No.	SDC2 Positive rate	FIT (Hb and TF) Positive rate	SDC2 and FIT ⁴ Positive rate	χ^2	P value
Histological type ¹					5.583	0.140
Adenocarcinoma	90	77.8% (70/90)	97.8% (88/90)	98.9% (89/90)		
Mucinous adenocarcinoma	7	71.4% (5/7)	85.7% (6/7)	85.7% (6/7)		
Degree of differentiation ²					4.355	0.168
Highly and Medium	94	79.8% (75/94)	97.9% (92/94)	98.9% (93/94)		
Poor	9	66.7% (6/9)	88.9% (8/9)	88.9% (8/9)		
Location ³					0.771	1.000
Proximal	22	86.4% (19/22)	95.5% (21/22)	100% (22/22)		
Distal	88	77.3% (68/88)	93.2% (82/88)	96.6% (85/88)		
TNM Stage						
T					2.807	0.156
0 - 2	28	89.3% (25/28)	85.7% (24/28)	92.9% (26/28)		
3 - 4	83	75.9% (63/83)	96.4% (80/83)	98.8% (82/83)		
N					0.343	1.000
0	55	76.4% (42/55)	92.7% (51/55)	96.4% (53/55)		
1 - 2	55	81.8% (45/55)	96.4% (53/55)	98.2% (54/55)		
M					0.509	1.000
0	88	77.3% (68/88)	94.3% (83/88)	97.7% (86/88)		
1	22	90.9% (20/22)	95.5% (21/22)	100% (22/22)		

¹Histological type: adenocarcinoma, mucinous adenocarcinoma, adenocarcinoma combined with mucinous adenocarcinoma, micropapillary carcinoma, indolent cell carcinoma, squamous carcinoma, etc. ²Highly differentiated included highly differentiated and highly-moderately differentiated. Moderately differentiated included moderately differentiated and moderately-poorly differentiated. Poorly differentiated included poorly differentiated, mucinous adenocarcinoma. ³Proximal included cecum, ascending colon, hepatic flexure, transverse colon. Distal included splenic flexure, descending colon, sigmoid colon, rectum. ⁴If either SDC2 or FIT has a positive result, it is considered positive. The chi-square test was utilized to compare differences between two groups in a qualitative test. When the expected frequency in the chi-square test is less than 5, a modified chi-square test, also known as Fisher's exact test, was used.

proximal CRC compared to distal CRC ($\chi^2 = 0.771$, $P = 1.000$). The combined fecal SDC2 and FIT test increased the positive detection rate for both proximal and distal tumors, resulting in a 100% positive rate for proximal tumors and a 96.6% positive rate for distal tumors.

3.6.4 TNM stage

The T stage is mainly related to the depth of tumor infiltration. The positive rate of SDC2 gradually decreased as the depth of infiltration increased, from 89.3% at T0-2 to 75.9% at T3-4. Conversely, the positivity rate for FIT was higher at T3-4 (96.4%) than at T0-2 (85.7%). There was no significant difference in the positivity rate of the combined SDC2 and FIT assay between T0-2 stage compared to T3-4 stage ($\chi^2 = 2.807$, $P = 0.156$). The combined test increases the positive detection rate for both T0-2 and T3-4 staging. The positive rates of SDC2 and FIT were higher in CRC with lymph node metastasis (N1-2) than in CRC without metastasis, suggesting that the detection rates of both SDC2 and FIT increase when lymph node metastasis is present. The combined test improved the detection rate of N stage, with the positive

detection rate for CRC with lymph node metastasis reaching 98.2%. SDC2 and FIT showed higher positive rates in CRC with metastases than in those without. Specifically, when the two were tested in combination, the positivity rate for FIT was 97.7% for M0 and 100% for M1.

3.7 Diagnosis efficacy of fecal SDC2, FIT, and serum CEA, alone and in combination

1. To investigate the diagnosis efficacy of relevant indicators, samples were categorized into a disease group consisting of CRC plus APL cases and a control group consisting of non-AA and general polyps, benign lesions, and healthy controls. The analysis of SDC2's diagnosis efficacy for CRC plus APL revealed a sensitivity of 75.3%, specificity of 81.4%, and AUC of 0.784 (Table 3; Figure 2). Conversely, the sensitivity, specificity, and AUC values for FIT in detecting CRC plus APL were 90.2%, 62.8%, and 0.765, respectively. These data demonstrate that SDC2 exhibits lower sensitivity, but higher specificity, than FIT. Combination of SDC2 with

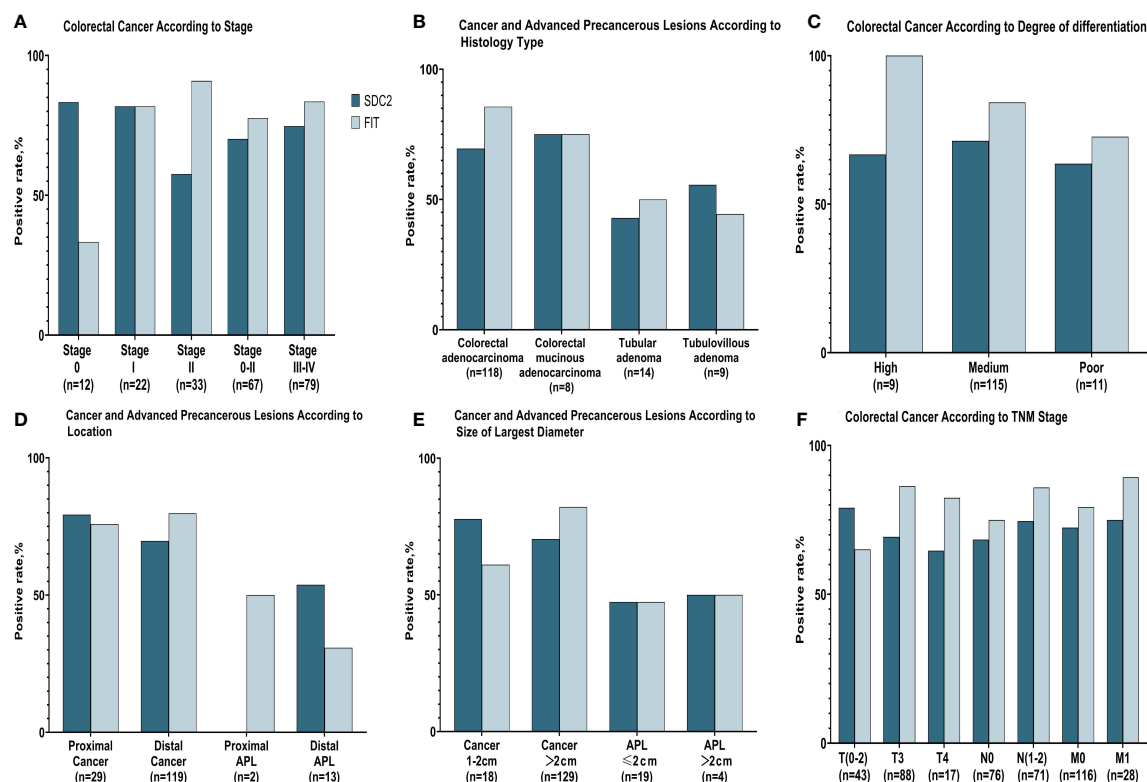


FIGURE 1

Positive rates of fecal SDC2 and FIT for detecting colorectal cancer and advanced precancerous lesions. (A) Positive rates of SDC2 and FIT for detecting different stages of CRC. (B) Positive rates of SDC2 and FIT for detecting CRC and APL with different histology types. (C) Positive rates of SDC2 and FIT for detecting CRC with different levels of differentiation. (D) Positive rates of SDC2 and FIT for detecting CRC and APL at different lesion sites. (E) Positive rates of SDC2 and FIT for detecting CRC and APL lesions of differing maximum diameters. (F) Positive rates of SDC2 and FIT for detecting CRC with different T, N, and M stages.

conventional FIT resulted in detection efficacy of 68.8 sensitivity, 95.0% specificity, and AUC 0.880. Further, the detection efficacy of SDC2 combined with conventional FIT and serum CEA generated sensitivity, specificity, and AUC values of 70.0%, 96.3%, and 0.905, respectively. The joint test initially employs logistic regression to build the prediction curve, followed by ROC curve analysis to estimate the area under the curve. The coefficients for the independent variables, SDC2 and FIT, are 2.81602 and 2.63004, respectively. The coefficient for the CEA dependent variable is 0.92039. Finally, the model's coefficient for the constant term is

-2.7346. Based on the above list of coefficients, logistic regression can be calculated with the equation: $p = e^{\beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2} / (1 + e^{\beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2})$, where β_0 , β_1 , and β_2 correspond to the constant term, SDC2 and FIT coefficients, respectively. We have created nomographs based on logistic regression, which are visually presented for the reference of clinicians (Figure 3).

2. We then assessed the effectiveness of different tests in detecting early-stage CRC plus APL, with a control group consisting of non-AA and general polyps, benign lesions, and healthy controls. Analysis of SDC2's efficacy in diagnostic early-

TABLE 3 Efficacy of fecal SDC2, FIT, and serum CEA for CRC and APL diagnosis.

	No.	Sensitivity (%)	Specificity (%)	AUC (95% CI)	P value
SDC2	518	75.3	81.4	0.784 (0.746–0.818)	< 0.0001
FIT	323	90.2	62.8	0.765 (0.715–0.810)	< 0.0001
CEA	282	50.6	77.5	0.684 (0.627–0.738)	< 0.0001
SDC2 + FIT ¹	308	68.8	95.0	0.880 (0.839–0.914)	< 0.0001
SDC2 + FIT + CEA ²	202	70.0	96.3	0.905 (0.856–0.941)	< 0.0001
SDC2 + FIT ³	308	95.3	53.9	0.746 (0.694–0.794)	< 0.0001
SDC2 + FIT + CEA ³	202	97.5	48.8	0.731 (0.665–0.791)	< 0.0001

FIT included fecal Hb and TF. The sensitivity and specificity of detecting quantitative CEA alone depend on logistic regression. ¹Use logistic regression to build prediction curves and ROC curve analysis to calculate the area under the curve. ²CEA results exceeding 5ng/ml are considered positive, use logistic regression to build prediction curves and ROC curve analysis to calculate the area under the curve. ³Result was considered positive if any one of them has a positive result.

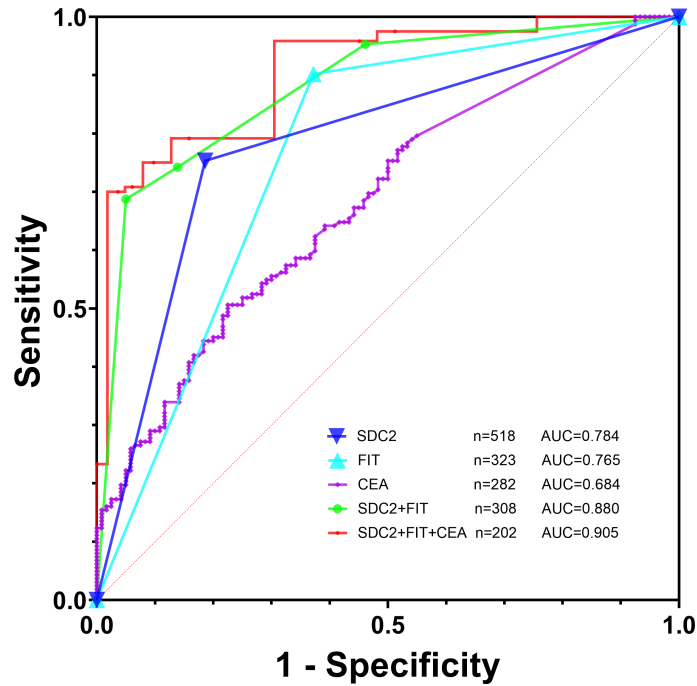


FIGURE 2
Efficacy of fecal SDC2, FIT, and tumor markers for CRC and APL diagnosis.

stage CRC plus APL showed a sensitivity, specificity, and AUC values of 69.9%, 81.4%, and 0.756, respectively (Table 4; Figure 4). Detection efficacy for FIT in early-stage CRC plus APL revealed a sensitivity of 86.3%, a specificity of 62.8%, and the AUC of 0.745, indicating that the sensitivity of SDC2 was lower than that of FIT, but with higher specificity. Among the tumor markers, CEA showed a significant difference in the detection of early colorectal neoplasms ($P = 0.0012$, Table 5). The diagnostic efficacy evaluation of SDC2 combined with FIT showed a

sensitivity of 60.6%, specificity of 95.0%, and AUC of 0.856. When combined with conventional FIT and serum CEA, the corresponding values were 66.2%, 96.3%, and 0.891, respectively. The coefficients for the independent variables, SDC2 and FIT, are 2.93287 and 2.6312, respectively. The coefficient for the CEA dependent variable is 0.17702. Finally, the model's coefficient for the constant term is -2.7346. We have also created nomographs based on logistic regression, which are visually presented for the reference of clinicians (Figure 5).

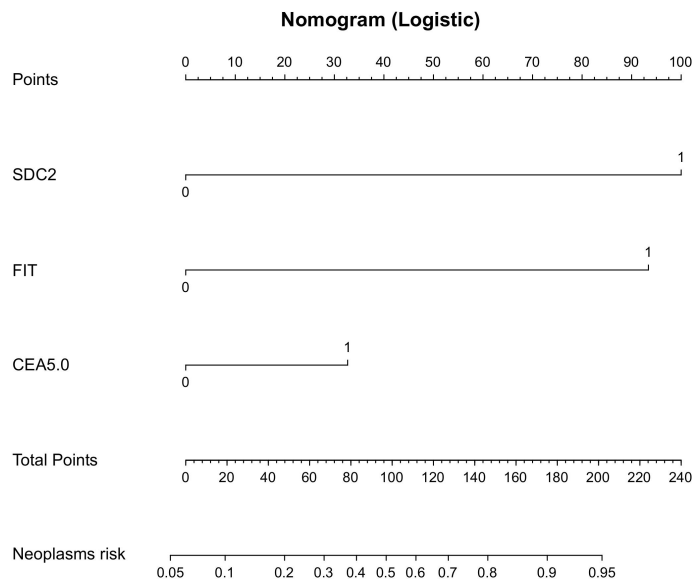


FIGURE 3
Nomogram-based prediction model for the risk of developing CRC and APL using logistic regression.

3. Result was considered positive if any one of them has a positive result, our study found that the joint detection of SDC2 and FIT can detect more cases of colorectal neoplasms, with a combined sensitivity of 95.3%. In cases of early-stage CRC and APL, the sensitivity was 93.9%. However, the joint detection of SDC2 and FIT may result in lower specificity, leading to a higher rate of false positives (Tables 3, 4). Therefore, the diagnostic efficacy of this method may be lower than that of SDC2 alone. When combined with CEA, the sensitivity of the detection method can reach its maximum, detecting up to 97.5% of colorectal neoplasms. Specifically, this method can detect 97.3% of early-stage CRC and APL cases.

4 Discussion

CRC is among the most common malignant tumors worldwide. Early screening can help to prevent and diagnose CRC, improve patient prognosis, reduce mortality, and lower the economic burden of disease on individuals (23, 24). The main detection method used is FOBT, followed by colonoscopy. FOBT is a cheap and convenient method for detecting gastrointestinal bleeding and is the most widely used screening method (25, 26). In this study, we utilized FOBTs for FIT, which could detect both hemoglobin (Hb) and transferrin (TF) without cross-reaction. Previous studies have mainly focused on Hb testing alone, while some studies suggest that TF testing can be an effective supplement for CRC and APL diagnosis. We routinely perform both types of FOBTs to effectively screen the target population (27, 28).

SDC2 methylation in stool DNA is an effective method for early detection of CRC (14, 29, 30); however, debates regarding the clinical application of fecal SDC2 are ongoing, and it has rarely been compared or combined with FIT to validate its screening efficacy. Given the effectiveness of individual target, the main use of multi-target stool DNA (4, 12, 31) or blood methylation tests (19, 32) for combined screening of colorectal neoplasms is currently in place, but their efficacy needs to be validated in large samples.

A previous multicenter study reported good diagnostic agreement between fecal SDC2 and colonoscopy, with a Kappa value of 0.84 (11). In contrast, our findings showed that the results of stool SDC2 were moderately consistent with colonoscopy (Kappa = 0.54; $P < 0.001$). This could be due to a higher number of false

positives in the non-CRC and non-APL group in our study (25/180), potential selection bias due to the relatively small sample size, or a high specificity due to the inclusion of a majority of members of the healthy population in our prospective multicenter study. Meanwhile, our results showed that fecal SDC2 and FIT have different advantages and disadvantages for CRC detection. Further, the positive detection rate of SDC2 in colorectal adenocarcinoma was 77.8% (70/90), while that of FIT in adenocarcinoma was 97.8% (88/90); and the positive detection rates of fecal SDC2 and FIT in poorly differentiated CRC were both low, at 66.7% and 88.9%, respectively, suggesting that for poorly differentiated CRC it is necessary to combine these methods with other approaches. Fecal SDC2 and FIT were able to detect more proximal tumors. The positivity rate for FIT was higher than that for SDC2 in both proximal and distal areas, with a positive detection rate of over 90% in both regions. The detection rate of fecal SDC2 decreased with increasing depth of tumor infiltration, from 89.3% at T0-2 to 75.9% at T3-4. Conversely, the detection rate of FIT was higher at T3-4 than at T0-2. The positive detection rates of fecal SDC2 and FIT were higher in patients with lymph node and distant metastasis than in those without metastasis. This suggests that positive fecal SDC2 and FIT results may be able to predict a poor prognosis for patients, although more evidence is needed to verify this hypothesis. Based on the relevant guidelines (5), FIT has high sensitivity for CRC diagnosis, but limited sensitivity for APL, which is also consistent with our results.

Our data show that the combination of fecal SDC2 and FIT improved the positive detection rates for early-stage and advanced CRC and APL to 98.0% (48/49), 98.4% (60/61), and 82.4% (14/17), respectively, suggesting that the two methods are complementary (Table 1). It is evident that the combined SDC2 and FIT test exhibited a higher detection rate for CRC and APL. Only 6 out of 128 colorectal neoplasms were not detected, indicating a remarkably high sensitivity. Given the widespread use of tumor markers in clinical practice, our data demonstrate a significant difference in CEA levels between early colorectal neoplasms and controls (Table 5). In relation to the combined SDC2, FIT, and CEA test showed a detection failure in only 3 cases out of a total of 120 colorectal neoplasms, thereby further enhancing the sensitivity. This result suggests that patients with colorectal neoplasms who test negative for CEA screening may benefit from detecting SDC2 and FIT, which may increase the positive detection rate. The data

TABLE 4 Efficacy of fecal SDC2, FIT, and serum CEA for early-stage CRC and APL diagnosis.

	No.	Sensitivity (%)	Specificity (%)	AUC (95%CI)	P value
SDC2	443	69.9%	81.4%	0.756 (0.714–0.796)	< 0.0001
FIT	253	86.3%	62.8%	0.745 (0.687–0.798)	< 0.0001
CEA	203	59.0%	60.8%	0.622 (0.551–0.689)	0.0018
SDC2 + FIT ¹	246	60.6%	95.0%	0.856 (0.806–0.897)	< 0.0001
SDC2 + FIT + CEA ¹	156	66.2%	96.3%	0.891 (0.832–0.936)	< 0.0001
SDC2 + FIT ²	246	93.9%	53.9%	0.739 (0.680–0.793)	< 0.0001
SDC2 + FIT + CEA ²	156	97.3%	48.8%	0.730 (0.654–0.798)	< 0.0001

¹Use logistic regression to build prediction curves and ROC curve analysis to calculate the area under the curve. ²Result are considered positive if any one of them has a positive result.

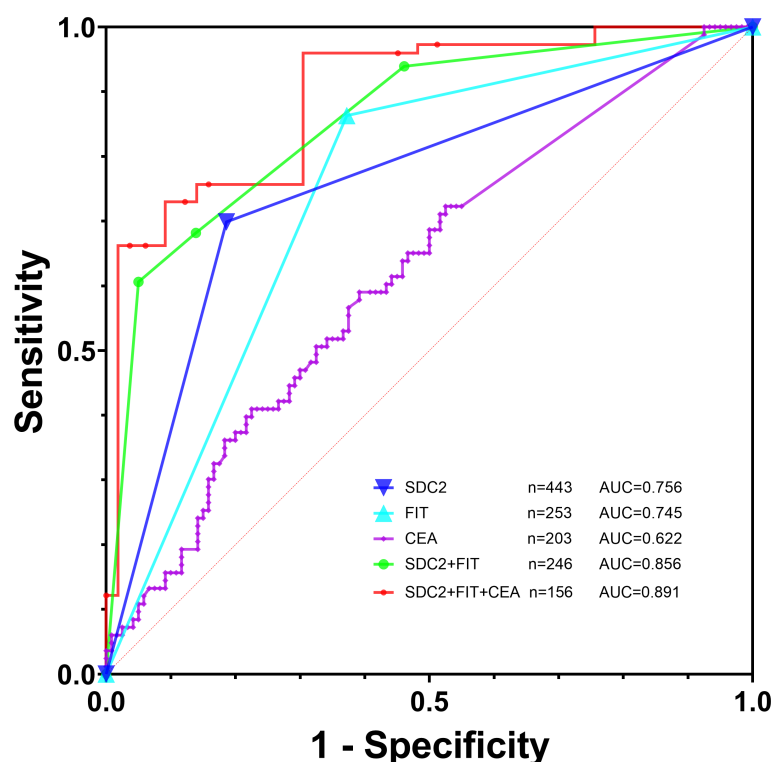


FIGURE 4
Efficacy of fecal SDC2, FIT, and tumor markers for early-stage CRC and APL diagnosis.

presented in Tables 1, 3 exhibit consistency and mutual verification. The corresponding data strongly indicate that the combination of the three tests successfully detected a maximum number of patients. The cost-effectiveness of diagnostic tests is a crucial factor to consider in clinical practice. Based on our calculations, the cost of painless colorectal microscopy and the cost of SDC2 methylation, FIT, and CEA are relatively similar (all around 1000 RMB). However, other factors such as test simplicity, patient acceptability, operability, and large-scale scalability should also be considered. Therefore, clinical practitioners should develop different screening strategies tailored to the specific needs of their patients.

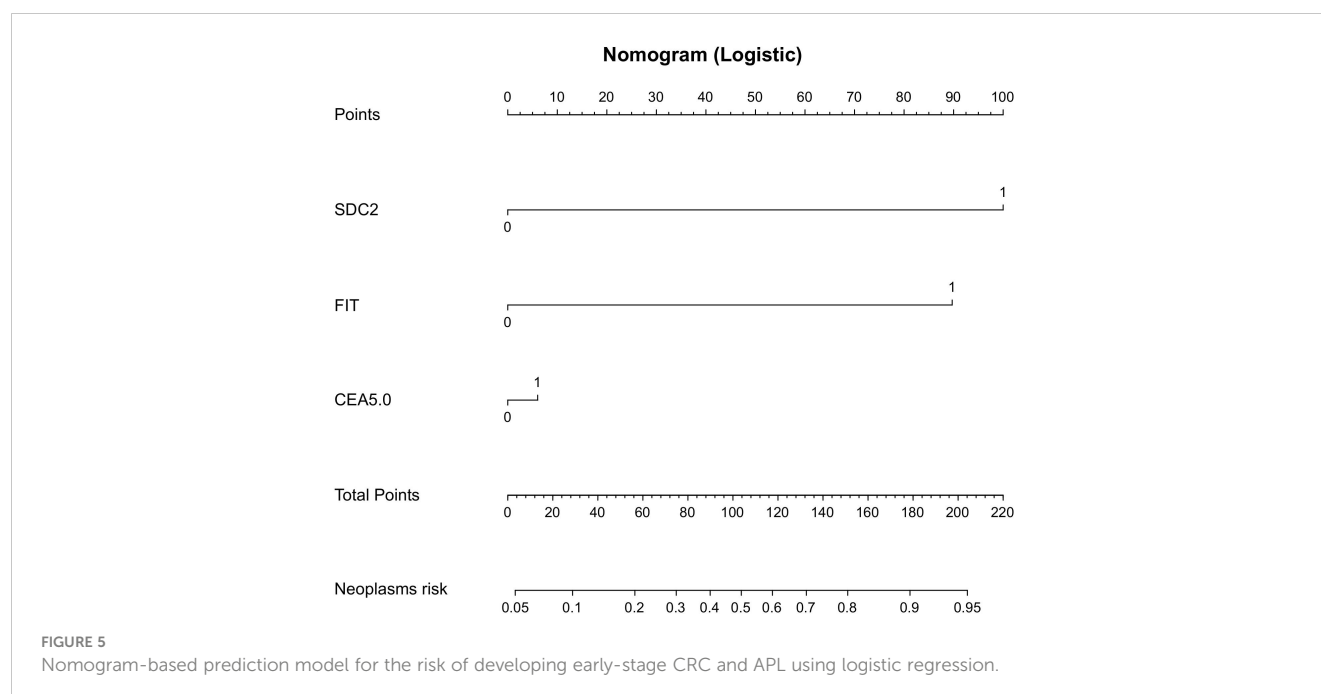
According to expert consensus (33), FIT is not currently recommended as a screening modality for APL. Biomarker combinations have been shown to have better screening efficacy than individual markers, and the combined use of fecal and blood markers can improve the sensitivity for detecting colorectal

neoplasms. Additionally, alterations in gut microbiology may also influence the development and progression of CRC (34).

Our study investigated the diagnostic effectiveness of combining SDC2, FIT, and CEA using two methods. First, we evaluated each indicator's clinical significance, alone and in combination, using logistic regression and ROC curve analysis. We found that the combined detection curve had the highest AUC at 0.891 in early-stage CRC and APL, while the AUC for a single CEA indicator was only 0.628. Second, we used a simpler approach that considered any of the indicators as positive, but this led to more false positives, mainly due to FIT's lower specificity. As a result, combining detection did not improve diagnostic efficacy and even performed worse than using SDC2 alone. Sensitivity refers to the probability of a classifier accurately predicting positive values among all positive samples. The AUC takes into account both true positive and false positive rates, making it a more comprehensive evaluation index for assessing a model's predictive performance. However, clinicians

TABLE 5 Efficacy of serum Tumor markers for early-stage CRC and APL diagnosis.

	No.	Sensitivity (%)	Specificity (%)	AUC (95%CI)	P value
CEA	193	58.0%	62.5%	0.628 (0.556–0.697)	0.0012
CA125	193	32.1%	76.8%	0.540 (0.467–0.612)	> 0.05
CA19-9	193	42.0%	67.0%	0.524(0.451–0.596)	> 0.05
CA15-3	193	55.6%	60.7%	0.575 (0.502–0.645)	> 0.05
AFP	193	95.1%	14.3%	0.516 (0.444–0.589)	> 0.05



should choose suitable methods based on their purpose and practicality when making decisions. Our test results can serve as a reference, but additional examinations such as imaging or pathology must be combined to make a comprehensive diagnosis.

5 Conclusions

Fecal SDC2 is useful for early screening of CRC and APL. Combining SDC2 with FIT (Hb and TF) can improve the positive detection rates of early and advanced CRC and APL. Additionally, combining fecal SDC2 and FIT with serum CEA has shown high detection efficacy. Using a combination of these methods could be a new approach for early screening of CRC and APL, but its effectiveness requires further validation in large sample populations.

Data availability statement

The raw data of this study are stored in the Research Data Deposit (RDD) of Sun Yat-sen University, which can be requested from the author, TZ with reasonable request.

Ethics statement

This study was approved by the Ethics Committee of The Sixth Affiliated Hospital of Sun Yat-sen University, Ethics No. 2022ZSLYEC-508. Written informed consent has been obtained from the individual(s) or their legal guardian/next of kin or from the patients/participants in this study.

Author contributions

TZ: Research design, research implementation, data collection, data analysis, article writing. ZH: Research implementation, data collection. XY: Research design, research guidance; LZ: research guidance, data collection. TL: Research design, research guidance, article review and revision. BT, SX, JH: Research design, research implementation, research guidance, data analysis and interpretation, article review and revision.

Funding

Supported by the program of Guangdong Provincial Clinical Research Center for Digestive Diseases (2020B111170004).

Acknowledgments

We would like to express our gratitude to all colleagues in the Department of Clinical Laboratory and the Guangdong Institute of Gastroenterology, The Sixth Affiliated Hospital of Sun Yat-sen University for their invaluable support in conducting this study.

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.

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.

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: Cancer J For Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
- Carethers JM. Fecal DNA testing for colorectal cancer screening. *Annu Rev Med* (2020) 71:59–69. doi: 10.1146/annurev-med-103018-123125
- Ahlquist DA, Skoletsky JE, Boynton KA, Harrington JJ, Mahoney DW, Pierceall WE, et al. Colorectal cancer screening by detection of altered human DNA in stool: feasibility of a multitarget assay panel. *Gastroenterology* (2000) 119(5):1219–27. doi: 10.1053/gast.2000.19580
- Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med* (2014) 370(14):1287–97. doi: 10.1056/NEJMoa1311194
- He J, Chen WQ, Li N, Chen HD, Du LB, Sun F, et al. China guideline for the screening, early detection and early treatment of colorectal cancer (2020, beijing). *Zhonghua Zhong Liu Za Zhi* (2021) 43(1):16–38. doi: 10.3760/cma.j.cn112152-20210105-00010
- Schreiner MA, Weiss DG, Lieberman DA. Proximal and large hyperplastic and nondysplastic serrated polyps detected by colonoscopy are associated with neoplasia. *Gastroenterology* (2010) 139(5):1497–502. doi: 10.1053/j.gastro.2010.06.074
- Li D, Liu L, Fevrier HB, Alexeeff SE, Doherty AR, Raju M, et al. Increased risk of colorectal cancer in individuals with a history of serrated polyps. *Gastroenterology* (2020) 159(2):502–11. doi: 10.1053/j.gastro.2020.04.004
- van Toledo DEFWM, Breekveldt ECH, Ijspeert JEG, van Vuuren AJ, van Kemenade F, Ramakers C, et al. Advanced serrated polyps as target of screening: detection rate and positive predictive value within a fecal immunochemical test based colorectal cancer screening population. *Endoscopy* (2022) 55(6):524–34. doi: 10.1055/a-1971-3488
- Chi Z, Lin Y, Huang J, Lv M-Y, Chen J, Chen X, et al. Risk factors for recurrence of colorectal conventional adenoma and serrated polyp. *Gastroenterol Rep* (2022) 10:goab038. doi: 10.1093/gastro/goab038
- Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet* (2019) 394(10207):1467–80. doi: 10.1016/S0140-6736(19)32319-0
- Wang J, Liu S, Wang H, Zheng L, Zhou C, Li G, et al. Robust performance of a novel stool DNA test of methylated sdc2 for colorectal cancer detection: A multicenter clinical study. *Clin Epigenet* (2020) 12(1):162. doi: 10.1186/s13148-020-00954-x
- Li YL, Guan X, Dou LZ, Liu Y, Huang HY, Huang SK, et al. [the clinical value of multi-target stool fecal immunochemical test-DNA in early screening and diagnosis for colorectal cancer]. *Zhonghua Yi Xue Za Zhi* (2022) 102(33):2607–13. doi: 10.3760/cma.j.cn112137-20220430-00974
- Xu J, Rong L, Gu F, You P, Ding H, Zhai H, et al. Asia-Pacific colorectal screening score combined with stool DNA test improves the detection rate for colorectal advanced neoplasms. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterological Assoc* (2022) 6:1627–36. doi: 10.1016/j.cgh.2022.09.002
- Niu F, Wen J, Fu X, Li C, Zhao R, Wu S, et al. Stool DNA test of methylated for the early detection of colorectal neoplasia. *Cancer Epidemiol Biomarkers Prev* (2017) 26(9):1411–9. doi: 10.1158/1055-9965.EPI-17-0153
- Ma L, Gong J, Zhao M, Kong X, Gao P, Jiang Y, et al. A novel stool methylation test for the non-invasive screening of gastric and colorectal cancer. *Front In Oncol* (2022) 12:860701. doi: 10.3389/fonc.2022.860701
- Yue C, Zhang Y, Wang Y, Zhang Z, Zhang M, Wang H, et al. The application value of syndecan-2 gene methylation for colorectal cancer diagnosis: A clinical study and meta-analyses. *Front In Med* (2022) 9:753545. doi: 10.3389/fmed.2022.753545
- Sina. *The mystery of creative biosciences's data: 50,000 people cohort study data on changanxin "Rollover" Again* (2022). Available at: <https://finance.sina.com.cn/tech/>
- 2022-05-23/doc-imizirau4283884.shtml?finpagefr=p_114 (Accessed 30 December 2022).
- guancha. *Creative biosciences is accused of misleading users with false data, how can the "Changanxin" Give people peace of mind?* (2021). Available at: <https://user.guancha.cn/main/content?id=528068> (Accessed 30 December 2022).
- Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* (2018) 359(6378):926–30. doi: 10.1126/science.aar3247
- Sun Y, Zheng S, Wan DS, Gu J, Wang JP, Zhang SZ, et al. Chinese protocol of diagnosis and treatment of colorectal cancer (2020 edition). *Zhonghua Wai Ke Za Zhi* (2020) 58(8):561–85. doi: 10.3760/cma.j.cn112139-20200518-00390
- Weiser MR. Ajcc 8th edition: colorectal cancer. *Ann Surg Oncol* (2018) 25(6):1454–5. doi: 10.1245/s10434-018-6462-1
- Herman JG, Graff JR, Myöhänen S, Nelkin BD, Baylin SB. Methylation-specific pcr: A novel pcr assay for methylation status of cpG islands. *Proc Natl Acad Sci U.S.A.* (1996) 93(18):9821–6. doi: 10.1073/pnas.93.18.9821
- Ladabaum U, Dominitz JA, Kahi C, Schoen RE. Strategies for colorectal cancer screening. *Gastroenterology* (2020) 158(2):418–32. doi: 10.1053/j.gastro.2019.06.043
- Bretthauer M, Løberg M, Wieszczyn P, Kalager M, Emilsson L, Garborg K, et al. Effect of colonoscopy screening on risks of colorectal cancer and related death. *N Engl J Med* (2022) 387(17):1547–56. doi: 10.1056/NEJMoa2208375
- Schreuders EH, Ruco A, Rabeneck L, Schoen RE, Sung JY, Young GP, et al. Colorectal cancer screening: A global overview of existing programmes. *Gut* (2015) 64(10):1637–49. doi: 10.1136/gutjnl-2014-309086
- Wolf AMD, Fontham ETH, Church TR, Flowers CR, Guerra CE, LaMonte SJ, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the american cancer society. *CA: Cancer J For Clin* (2018) 68(4):250–81. doi: 10.3322/caac.21457
- Uchida K, Matsuse R, Miyachi N, Okuda S, Tomita S, Miyoshi H, et al. Immunochemical detection of human blood in feces. *Clin Chim Acta* (1990) 189(3):267–74. doi: 10.1016/0009-8981(90)90308-f
- Chen J-G, Cai J, Wu H-L, Xu H, Zhang Y-X, Chen C, et al. Colorectal cancer screening: comparison of transferrin and immuno fecal occult blood test. *World J Gastroenterol* (2012) 18(21):2682–8. doi: 10.3748/wjg.v18.i21.2682
- Han YD, Oh TJ, Chung T-H, Jang HW, Kim YN, An S, et al. Early detection of colorectal cancer based on presence of methylated syndecan-2 (Sdc2) in stool DNA. *Clin Epigenet* (2019) 11(1):51. doi: 10.1186/s13148-019-0642-0
- Oh TJ, Oh HI, Seo YY, Jeong D, Kim C, Kang HW, et al. Feasibility of quantifying methylation in stool DNA for early detection of colorectal cancer. *Clin Epigenet* (2017) 9:126. doi: 10.1186/s13148-017-0426-3
- Imperiale TF, Kisiel JB, Itzkowitz SH, Scheu B, Duimstra EK, Statz S, et al. Specificity of the multi-target stool DNA test for colorectal cancer screening in average-risk 45–49 year-olds: A cross-sectional study. *Cancer Prev Res (Phila)* (2021) 14(4):489–96. doi: 10.1158/1940-6207.CAPR-20-0294
- Xu F, Yu S, Han J, Zong M, Tan Q, Zeng X, et al. Detection of circulating tumor DNA methylation in diagnosis of colorectal cancer. *Clin Transl Gastroenterol* (2021) 12(8):e00386. doi: 10.14309/ctg.0000000000000386
- Chan FKL, Wong MCS, Chan AT, East JE, Chiu H-M, Makharia GK, et al. Joint asian pacific association of gastroenterology (Apage)-asian pacific society of digestive endoscopy (Apsde) clinical practice guidelines on the use of non-invasive biomarkers for diagnosis of colorectal neoplasia. *Gut* (2023) 72(7):1240–54. doi: 10.1136/gutjnl-2023-329429
- Kong C, Liang L, Liu G, Du L, Yang Y, Liu J, et al. Integrated metagenomic and metabolomic analysis reveals distinct gut-microbiome-derived phenotypes in early-onset colorectal cancer. *Gut* (2023) 72(6):1129–42. doi: 10.1136/gutjnl-2022-327156



OPEN ACCESS

EDITED BY

Zora Lasabova,
Comenius University, Slovakia

REVIEWED BY

Pinyi Lu,
National Cancer Institute at Frederick
(NIH), United States
Antonella Argentiero,
National Cancer Institute Foundation
(IRCCS), Italy
Saima Wajid,
Jamia Hamdard University, India

*CORRESPONDENCE

Cong Chen

✉ chencong2006nyd@163.com

Tingting Yu

✉ tingting@njmu.edu.cn

Tao Chen

✉ taochen1381@njmu.edu.cn

[†]These authors have contributed equally to this work

RECEIVED 27 November 2022

ACCEPTED 24 July 2023

PUBLISHED 09 August 2023

CITATION

Chen W, Wang Y, Gu H, Zhang Y,
Chen C, Yu T and Chen T (2023)
Molecular characteristics, clinical
significance, and immune landscape
of extracellular matrix remodeling-
associated genes in colorectal cancer.
Front. Oncol. 13:1109181.
doi: 10.3389/fonc.2023.1109181

COPYRIGHT

© 2023 Chen, Wang, Gu, Zhang, Chen, Yu
and Chen. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). 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.

Molecular characteristics, clinical significance, and immune landscape of extracellular matrix remodeling-associated genes in colorectal cancer

Wenlong Chen^{1†}, Yiwen Wang^{1†}, Haitao Gu^{2†}, Yi Zhang¹,
Cong Chen^{3*}, Tingting Yu^{4*} and Tao Chen^{1*}

¹Department of Colorectal Surgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China, ²Department of General Surgery, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, ³Department of Neurology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China, ⁴Department of Medical Genetics, School of Basic Medical Science, Jiangsu Key Laboratory of Xenotransplantation, Nanjing Medical University, Nanjing, China

Background: Extracellular matrix (ECM) remodeling is one of the hallmark events in cancer and has been shown to be closely related to tumor immunity. Immunotherapy has evolved as an important tool to treat various cancers and improve patient prognosis. The positive response to immunotherapy relies on the unique interaction between cancer and the tumor microenvironment (TME). However, the relationship between ECM remodeling and clinical outcomes, immune cell infiltration, and immunotherapy in colorectal cancer (CRC) remains unknown.

Methods: We systematically evaluated 69 ECM remodeling-associated genes (EAGs) and comprehensively identified interactions between ECM remodeling and prognosis and the immune microenvironment in CRC patients. The EAG_score was used to quantify the subtype of ECM remodeling in patients. We then assessed their value in predicting prognosis and responding to treatment in CRC.

Results: After elaborating the molecular characteristics of ECM remodeling-related genes in CRC patients, a model consisting of two ECM remodeling-related genes (MEIS2, SLC2A3) was developed for predicting the prognosis of CRC patients. Receiver Operating Characteristic (ROC) and Kaplan-Meier (K-M) analysis verified its reliable predictive ability. Furthermore, we created a highly reliable nomogram to enhance the clinical feasibility of the EAG_score. Significant differences in TME and immune function, such as macrophages and CD8⁺ T cells, were observed between high- and low-risk CRC patients. In addition, drug sensitivity is also strongly related to EAG_score.

Conclusion: Overall, we developed a prognostic model associated with ECM remodeling, provided meaningful clinical implications for immunotherapy, and facilitated individualized treatment for CRC patients. Further studies are needed to reveal the underlying mechanisms of ECM remodeling in CRC.

KEYWORDS

extracellular matrix remodeling, colorectal cancer, prognosis, tumor microenvironment, immunotherapy

Introduction

CRC is one of the most common gastrointestinal tumors worldwide and the second leading cause of cancer-related deaths (1). Currently, TNM (tumor node metastasis) staging is the most commonly used clinical staging method to guide the treatment and management of CRC patients (2). Early stage I and II CRC can be cured by surgical resection, while the standard treatment for stage III CRC is surgical resection with adjuvant chemotherapy. There are several treatment options for metastatic CRC, including surgery, chemotherapy, radiotherapy, immunotherapy, and biologic-targeted therapy (3, 4). However, under the TNM staging criteria, due to the complexity of tumors, patients at the same TNM stage still show large differences in treatment outcomes and clinical prognosis. Therefore, it is important to find better classification methods for predicting prognosis and guiding treatment for CRC patients.

Extracellular matrix (ECM) is a collection secreted by cells to provide structural and biological support to surrounding cells and its major components include collagen, elastin, and polysaccharides (5, 6). The concept of ECM remodeling can be understood as a change in the physical to biochemical properties of the ECM, i.e., a change in the overall abundance, concentration, and structure of individual ECM components, thereby altering the three-dimensional spatial topology of the pericellular matrix, its biochemical and biophysical properties at the tissue level, and thus affecting the biological behavior of the cell. The basic processes of ECM remodeling include (1) ECM deposition, mediated by a variety of proteases such as Lysyl Oxidase (LOX) and matrix metalloproteinases family, alters the abundance and composition of ECM components, thereby altering the ECM biochemical and mechanical properties; (2) chemical modification at the post-translational level, which alters the biochemical

characteristics and structural features of the ECM; (3) protein hydrolytic degradation processes, which would release large amounts of bioactive ECM fragments and ECM binding factors and may be required to remove cellular constraints (e.g., physical barriers to migration); and (4) physical remodeling mediated by specific proteases such as procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2 (PLOD2), which affects cellular behavior by aligning ECM fibrils and stabilizing the cross-linking of ECM proteins (7, 8). More importantly, this process of matrix remodeling is accompanied by complex cell-matrix biochemical signaling and molecular communication, which has a profound influence on cell behavior. In fact, the ECM is constantly undergoing dynamic remodeling and is regulated by a variety of bioactive molecules, signaling pathways, and substances released when the ECM itself is damaged. In many solid tumors such as CRC, ECM remodeling occurs as a result of collagen crosslinking and increased stiffness, and it has been demonstrated that tumor cells can detect changes in the mechanical stress of the microenvironment in which they are exposed and thus alter cell biological behavior, such as focal adhesion assembly, changes in cell contractility, overexpression of EMT markers, and upregulation of various cellular external to internal signaling cascades such as PI3K and ERK signaling (9, 10). In addition, ECM remodeling in distant organs will create favorable conditions for tumor cell metastasis prior to the development of distant metastasis (11). Therefore, tumor progression is accompanied by dysregulation of ECM remodeling.

Although only a part of CRC patients can benefit from immunotherapy in clinical practice, its importance in the comprehensive treatment of CRC has gradually emerged (12). Immune function is dependent on the structural composition and physical properties of the ECM, and the dynamic evolution of the ECM shapes a relatively immunosuppressive environment for tumor cells, suppressing both innate and adaptive immune responses (13). The simplest explanation for this is that the increased ECM density provides a physical barrier that prevents tumor cells from interacting with immune cells during the process of ECM remodeling. In recent years, with the gradual in-depth research between ECM remodeling and tumor immunity, it has been proved that the movement and metabolism of immune cells and T cell phenotype are regulated by ECM components such as collagen, which is also directly related to the development of tumor (14). Moreover, ECM remodeling eliminates the enhancement of

Abbreviations: ECM, Extracellular matrix; TME, tumor microenvironment; CRC, colorectal cancer; EAGs, extracellular matrix associated genes; MEIS2, meis homeobox 2; SLC2A3, solute carrier family 2 member 3; TNM, tumor node metastasis; ROC, receiver operating characteristic; CAF, cancer-associated fibroblast; OS, overall survival; AJCC, American Joint Committee on Cancer; AUC, area under curve; CI, confidence interval; PD-1, programmed cell death protein-1; PD-L1, programmed cell death ligand-1; CTLA-4, cytotoxic T lymphocyte associated antigen-4; TALE, three amino-acid loop extension; AML, acute myeloid leukemia.

tumor-derived exosome diffusion and subsequent cancer-associated fibroblast (CAF) induction (15). These results suggest that the mechanism of ECM remodeling and tumor immunity is far more complex than imagined.

In this study, we grouped patients using ECM remodeling-related genes and established a prognostic model based on MEIS2 and SLC2A3. Besides, a nomogram that accurately predicted patient survival was constructed. Further, we assessed the differences between high- and low-risk groups in clinical characteristics, molecular features, immune function, and drug sensitivity. In conclusion, our work constructed a valid prognostic model and provides new insights into the immunotherapy of CRC patients.

Materials and methods

Data collection

The process of this work is shown in Figure S1. In this study, GSE39582 (16) and GSE17536 (17) were selected as training cohort ($n = 758$) from the Gene Expression Omnibus (GEO) database, 618 CRC samples from The Cancer Genome Atlas (TCGA) were applied as testing cohort. After excluding patients who lacked important clinical information such as overall survival (OS) and AJCC stage, we conducted data normalization to avoid batch effects. The gene set containing 69 genes related to ECM remodeling was obtained from the AmiGO database with the keyword “ECM remodeling” and the restriction “Homo sapiens” (18) (Table S1).

Consensus clustering analysis

Cluster analysis was performed using the “ConsensusClusterPlus” package, using agglomerative km clustering with a 1-Pearson correlation distances and resampling 80% of the samples for 10 repetitions. The optimal number of clusters was determined using the empirical cumulative distribution function plot (19).

Association between molecular patterns with the clinical characteristics and prognosis of CRC

We integrated the patients’ clinical information such as survival time, survival status, age, gender, grade, and AJCC stage. Kaplan-Meier analyses obtained by the “Survival” and “SurvMiner” packages were used to assess OS differences between different groups (20).

Relationship of molecular patterns with TME in CRC

We used the ESTIMATE algorithm to assess the Stromalscore, Immunescore, and TMEscore of CRC patients. The CIBERSORT algorithm and MCPcounter algorithm were used to calculate the

level of immune cell and stromal cell infiltration in each patient. We evaluated the association between risk scores, expression levels of genes involved in model construction, and immune cell infiltration, respectively. Then we compared subgroup differences in immune checkpoint expression and immune function. The “GSVA” package was used for GSVA analysis. We downloaded the subset `c2.cp.kegg.v7.4.symbols.gmt` to evaluate the relevant pathways and molecular mechanisms, set the minimum gene set to 5, the maximum gene set to 5000, and calculated the enrichment score of each sample in each gene set. Finally, the enrichment score matrix is obtained for further analysis.

Identification of DEGs and functional enrichment analysis

We obtained 195 differentially expressed genes (DEGs) between ECM remodeling subgroups using the “Limma” package under $|\log_{2}FC| \geq 1$ and $p < 0.05$. Then “clusterProfiler” package was used to perform Gene Ontology (GO) function and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis based on these DEGs (21).

Development of the ECM remodeling-associated prognostic EAG_Score

We constructed an EAG_Score to quantitatively assess the degree of ECM remodeling in each patient. Based on 195 DEGs, we screened 130 DEGs associated with prognosis using univariate Cox regression (uniCox) analysis and subsequently performed regression analysis using the “glmnet” package. LASSO-cox analysis was used to construct a prognostic model. the EAG_Score was calculated as: $EAG_Score = \sum (\text{gene Expression} * \text{gene coefficient})$. Based on the median EAG_Score, we divided the patients into high and low-risk groups and performed PCA analysis using the “stats” package (version 3.6.0). Specifically, we first performed z-score transformation of the patients’ gene expression profiles and further performed dimensionality reduction analysis using the `prcomp` function to obtain the reduced matrix.

Clinical significance analysis of the prognostic EAG_Score

After excluding patients with missing data, we integrated patients’ clinical information and EAG_Score for uniCox and multivariate Cox regression (multiCox) analysis; we performed ROC analysis using the `pROC` package (version 1.17.0.1) to obtain AUC (Area Under Curve) values. Specifically, AUC values and Confidence Intervals (CI) were evaluated using `CI` function of the package to obtain the final AUC results. We analyzed whether the EAG_Score model could be used as an independent prognostic predictor. In addition, we analyzed the differences in EAG_Score between the subgroups obtained by different clustering analysis methods.

Establishment of a predictive nomogram

Combined with the patients' EAG_Score and other clinicopathological features, the "rms" package was used to plot a nomogram. We conducted ROC analysis to explore the prognostic predictive power of these clinical features, particularly in predicting the patients' 1-year, 3-year, and 5-year OS. Calibration curves were used to validate the predictive accuracy of the column line graphs.

Immunohistochemical analysis

By searching the Human Protein Atlas (HPA) database, we obtained immunohistochemical staining images to determine protein expression of MEIS2 and SLC2A3 in normal and CRC tissues. HPA003256 and CAB002763 are antibodies to MEIS2 and SLC2A3, respectively.

Drug sensitivity analysis

Setting a filter condition of $p < 0.001$, we predicted patients' IC50 values for drugs commonly used in clinical practice using the "pRRophetic" package and compared them in high and low-risk groups (22).

Results

ECM remodeling related genes in CRC

First, we analyzed the mRNA expression data in the training cohort, after eliminating the batch effect (Figure S2), we identified the expression levels of 69 ECM remodeling-associated genes in tumor samples and normal samples and it was found that most of the genes were down-regulated in tumor tissue (Figure 1A). PPI networks were mapped using an online tool (<https://cn.string-db.org>) to explore the association between ECM remodeling genes. The results showed that CTNBN1, SMARCA4, MMP14, SRC, and ACTB were hub genes (confidence score = 0.900) (Figure 1B). Subsequently, we explored the prognostic value of these genes by uniCox analysis. 20 genes, including MXRA8, MXRA7, and MMP14, have been shown to be associated with the prognosis of patients (Table S3; Figure 1C). GO and KEGG analysis showed that these genes were mainly related to the regulation of tissue and ECM remodeling, cancer pathways such as liver cancer and gastric cancer (Figures 1D, E).

Generation of ECM remodeling subgroups in CRC

To further determine the relationship between ECM remodeling and CRC, we performed the clustering analysis based on ECM remodeling-associated genes and the result showed that the best clustering variable was 2. The patients were classified into EAGcluster

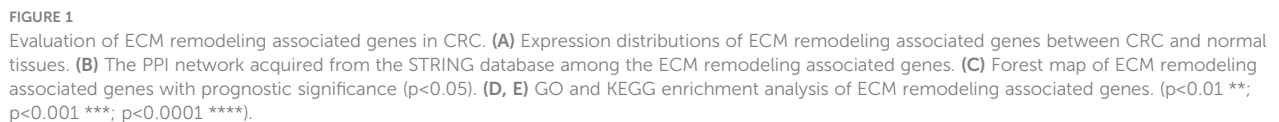
1 ($n = 347$) and EAGcluster 2 ($n = 392$) (Figure 2A). PCA analysis showed the reliability of the grouping (Figure 2B). We compared the OS between the groups and observed a significant survival difference (Figure 2C). In addition, as shown in Figure 2D, EAGcluster 1 possessed higher levels of ECM remodeling-related gene expression and was associated with a more advanced AJCC stage.

Characteristics of the TME in different subgroups

We performed GSVA analysis to compare the differences in ECM remodeling-related enrichment pathways between these subgroups, the result showed that ECM signaling exchanges (such as ECM receptor interaction, glycosaminoglycan biosynthesis), focal adhesion and multiple signaling pathways (JAK-STAT signaling pathway, MAPK signaling pathway) were enriched in EAGcluster1 (Figure 3A). The Cibersort algorithm was used to calculate the infiltration of 22 immune cells for each tumor sample (Figure 3B) and we observed significant differences between the two subgroups, T_cells_follicular_helper, Macrophages, Neutrophils were found to be relatively higher in EAGcluster 1, while T_cells_CD8, T_cells_CD4_memory_resting, T_cells_CD4_memory_activated, Tregs, and NK_cells_resting were higher in EAGcluster 2, the results of MCPcounter analysis showed that EAGcluster1 had a lower abundance of cytotoxic lymphocytes and a higher abundance of fibroblasts and endothelial cell infiltration than EAGcluster 2 (Figure 3C). Then We compared several key immune checkpoints (PD-1, PD-L1, CTLA-4) and observed that EAGcluster 1 has a higher expression level compared to EAGcluster2 (Figures 3D-F). Furthermore, to better understand the link between ECM remodeling and tumor immunity, TME scores (including StromalScore, ImmuneScore, and ESTIMATEScore) were calculated using the ESTIMATE algorithm, and the results showed that the EAGcluster 1 had a higher TME score (Figure 3G).

Identification of gene subgroups based on DEGs

To explore the potential biological activity of ECM remodeling subgroups, we obtained 195 DEGs related to ECM remodeling using the "limma" package and performed functional enrichment analysis (Tables S6, S7). GO analysis showed that DEGs were mainly enriched in biological processes such as cell adhesion, cell differentiation, and cell motility (Figure 4A). KEGG analysis showed significant enrichment in the PI3K-Akt signaling pathway, Phagosome, Cytokine-cytokine receptor interaction, etc (Figure 4B). Subsequently, we discussed the prognostic value of DEGs using uniCox analysis and obtained a total of 28 prognosis-related genes at $p < 0.001$ (Table S8). We divided patients into two geneClusters based on the expression of prognostic genes. We found that patients in geneCluster A had better OS compared to geneCluster B (Figure 4C), and geneCluster B had more abundant DEGs expression and was associated with later AJCC stage (Figure 4D). In addition, as expected, there was significant



multiCox analysis for 130 prognosis-related DEGs were conducted to establish an optimal predictive model. Finally, we obtained two genes (MEIS2 and SLC2A3) and the formula of EAG_score: risk score = $(0.167125186316921 * \text{expression of SLC2A3}) + (0.144964442480063 * \text{expression of MEIS2})$. **Figure 5A** displayed the patients' distribution in the different groups. Besides, we found EAGcluster 1 and geneCluster A had higher risk score (**Figures 5B, C**). Subsequently, patients were divided into high-risk and low-risk groups according to the

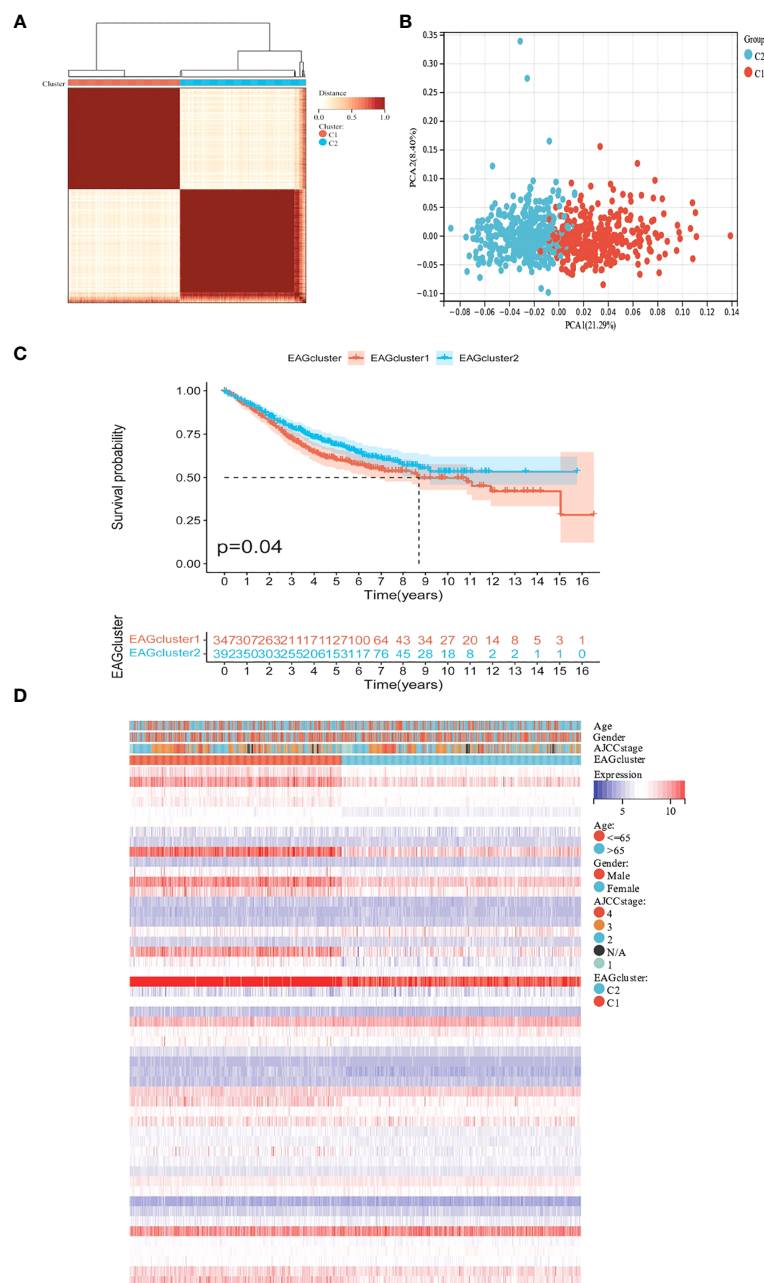


FIGURE 2

Classification of ECM remodeling associated genes subgroups and clinicopathological and biological characteristics of two distinct subtypes.

(A) Consensus analysis matrix heatmap defining two clusters ($k = 2$) and their correlation area. (B) PCA analysis showed significant differences in the transcriptome between the two subgroups. (C) Kaplan–Meier curves for OS of the EAGclusters. (D) Differences in clinicopathological characteristics and expression levels of ECM remodeling associated genes between EAGcluster1 and EAGcluster2.

median EAG_Score, Kaplan-Meier analysis indicated that low-risk patients had a better OS over high-risk patients (Figure 5D). PCA analysis showed a good distribution of patients in high and low-risk groups (Figure 5E). The AUC values of the model predicting patients' OS at 1, 3, and 5 years were 0.64 (95% CI = 0.71–0.57), 0.61 (95% CI = 0.66–0.56), and 0.60 (95% CI = 0.65–0.55), respectively (Figure 5F). Compared with other existing prognostic models for CRC, such as the ferroptosis-related genes prognostic model (AUC = 0.64, 0.64, 0.71 for 1,3,5-year OS, respectively) (23), the platelet-related prognostic model (AUC

= 0.722, 0.706, 0.689 for 1,3,5-year OS, respectively) (24), and the anoikis and immune-related genes prognostic model (AUC = 0.671, 0.634, 0.638 for 1,3,5-year OS, respectively) (25), our model also demonstrated good predictive performance. With the increase of EAG_Score, the OS of patients decreased and the mortality rate gradually increased (Figures 5G, H). Figure 5I shows the expression heatmap of MEIS2 and SLC2A3.

In addition, our prognostic model also showed good predictive power in validation cohort (Figure S3). The expressions of MEIS2 and SLC2A3 were verified by The Human Protein Atlas (Figure 6).

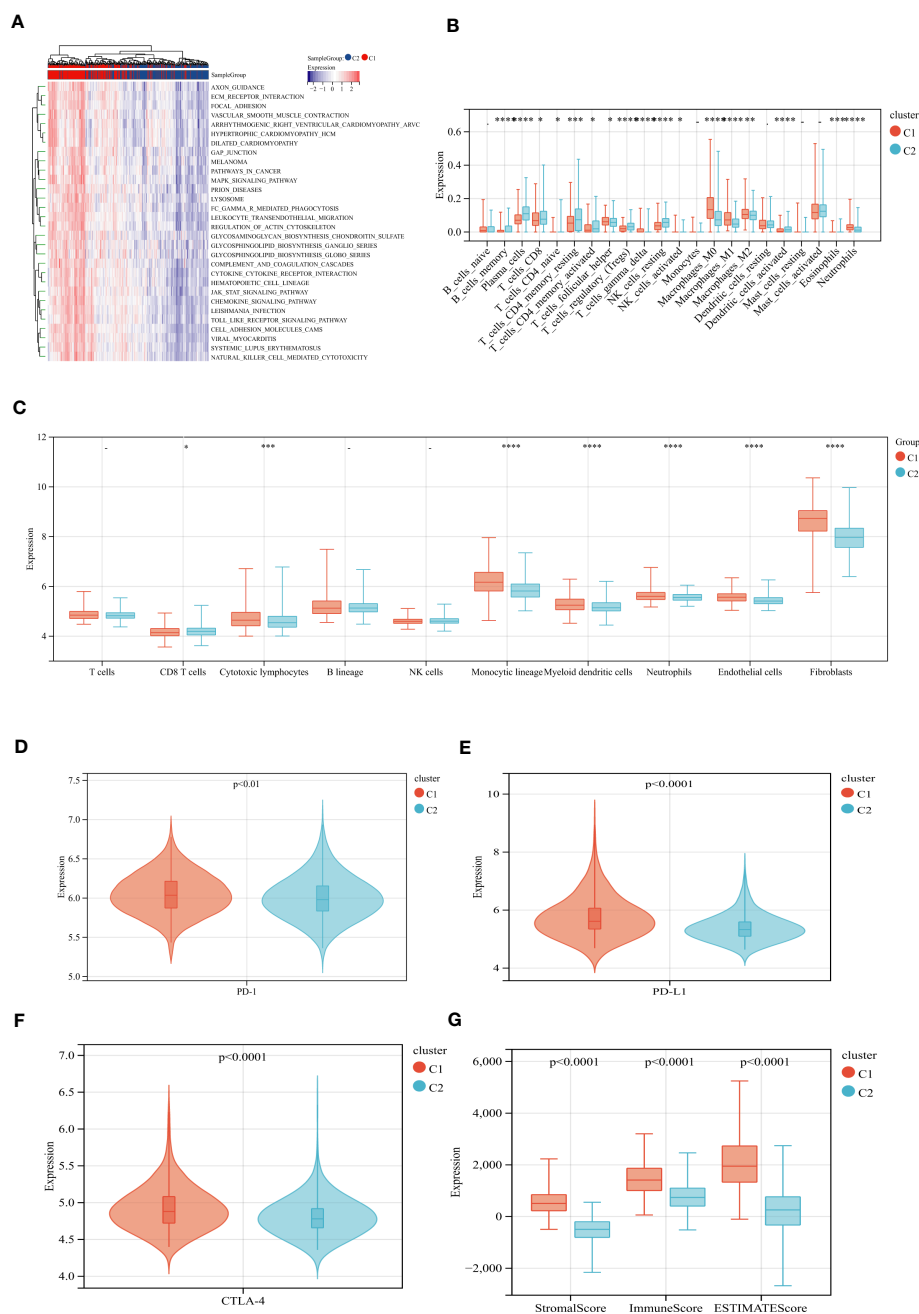


FIGURE 3

Differences in tumor immune microenvironment between EAGcluster1 and EAGcluster2. (A) GSEA of biological pathways between two distinct subgroups. (B) Abundance of 22 infiltrating immune cell types in the two subgroups. (C) Abundance of 8 infiltrating immune cell types and 2 stromal cell types in the two subgroups. (D–F) Expression levels of PD-1, PD-L1, and CTLA-4 in the two subgroups. (G) Correlations between the two subgroups and TME score. ($p < 0.05$ *; $p < 0.001$ ***; $p < 0.0001$ ****; $p > 0.05$ "-") symbol means $p > 0.05$.

Clinical correlation analysis of the prognostic EAG_Score

We performed uniCox and multiCox analyses to determine the independent prognostic value of EAG_Score (Figures 7A, B). The forest plot showed that EAG_Score could be used as an independent factor to predict the prognosis of patients. In

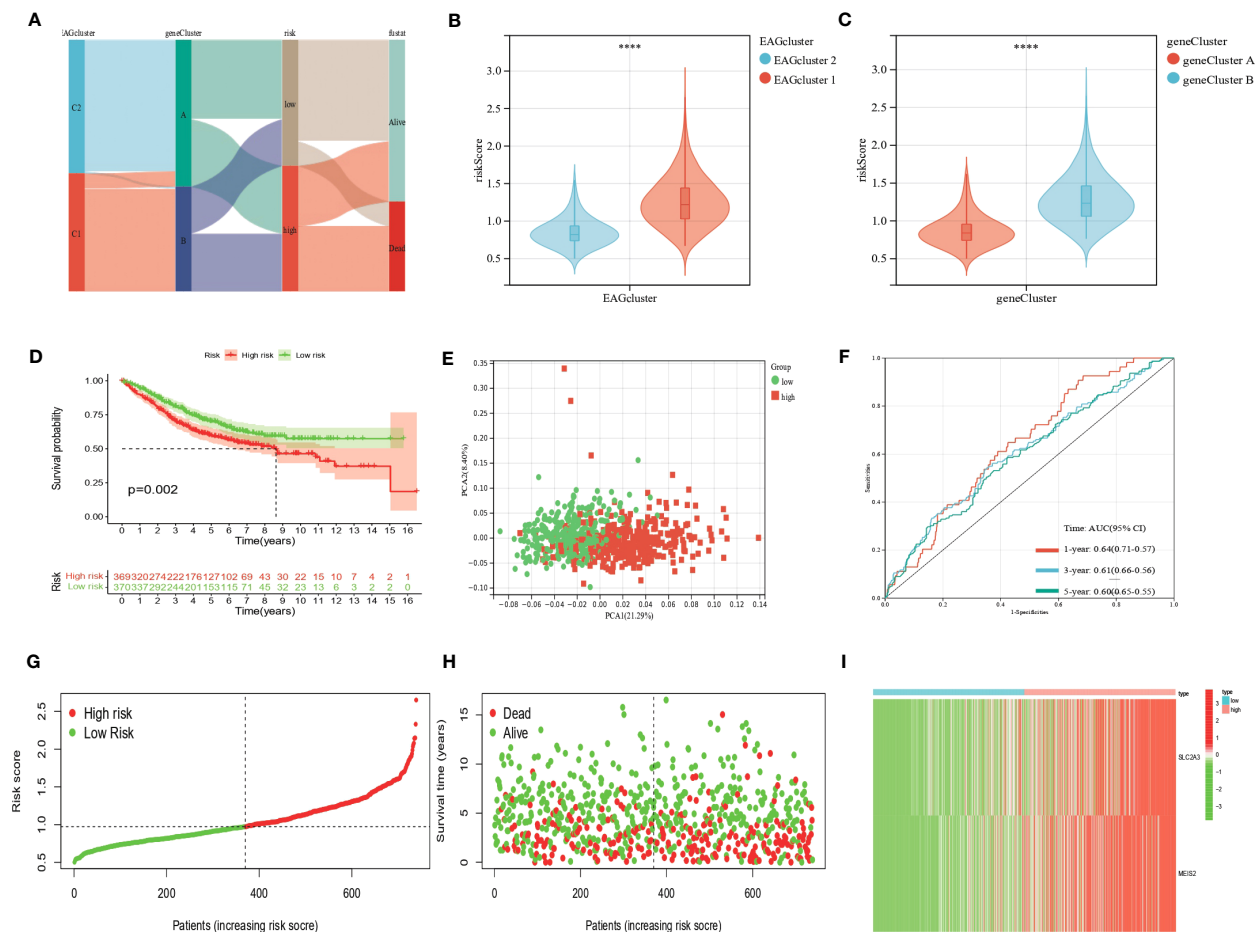
addition, to explore the relationship between EAG_Score and clinical characteristics, we discussed the correlation between clinical information such as age, gender, and AJCC stage, the results showed that later AJCC stage was associated with a higher risk score (Figure 7C). Moreover, patients with metastasis had a higher EAG_Score (Figure 7D). Overall, a higher risk score means a higher risk of metastasis and a worse prognosis.



0.84 (95%CI = 0.88-0.81) and 0.86 (95%CI = 0.89-0.82), respectively (Figure 7G).

Assessment of TME, checkpoints, and immune function in distinct groups

We calculated the correlation between immune cell abundance and EAG_Score. As shown in (Figures 8A–H), EAG_Score was positively correlated with Macrophages_M0, Macrophages_M1,



Macrophages_M2, T_cells_follicular_helper, and negatively correlated with Dendritic_cells_resting, T_cells_CD4_memory_activated, T_cells_CD4_memory_resting, T_cells_CD8. In addition, EAG_Score was associated with higher StromalScore, ImmuneScore, and ESTIMATEScore (Figure 8I). We found a significant correlation between the genes involved in the model construction and most

immune cells' infiltration levels (Figure 8J). Figure 8K demonstrated that certain immune functions including APC_co_stimulation, CCR, HLA, and T_cell_co-stimulation, differed significantly between the two distinct groups. Furthermore, we compared 35 common immune checkpoint inhibitors between high- and low-risk groups, such as PD-1, PD-L1, CTLA-4, lymphocyte-activation gene 3, and tumor necrosis

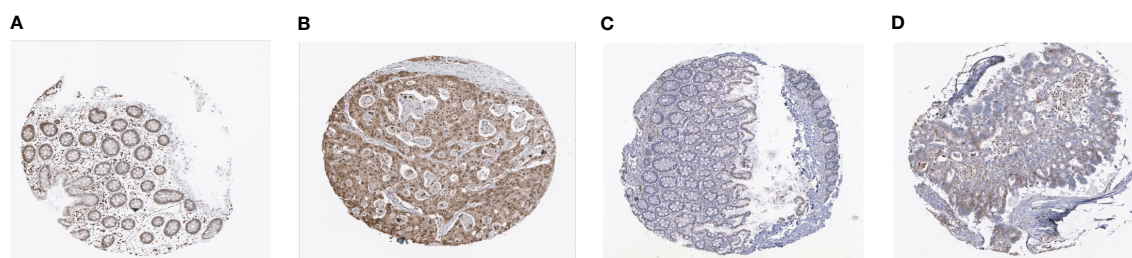


FIGURE 6
Representative immunohistochemistry images of MEIS2 and SLC2A3 in CRC tissues and normal tissues. (A, B) MEIS2 in normal tissues and CRC tissues. (C, D) SLC2A3 in normal tissues and CRC tissues.

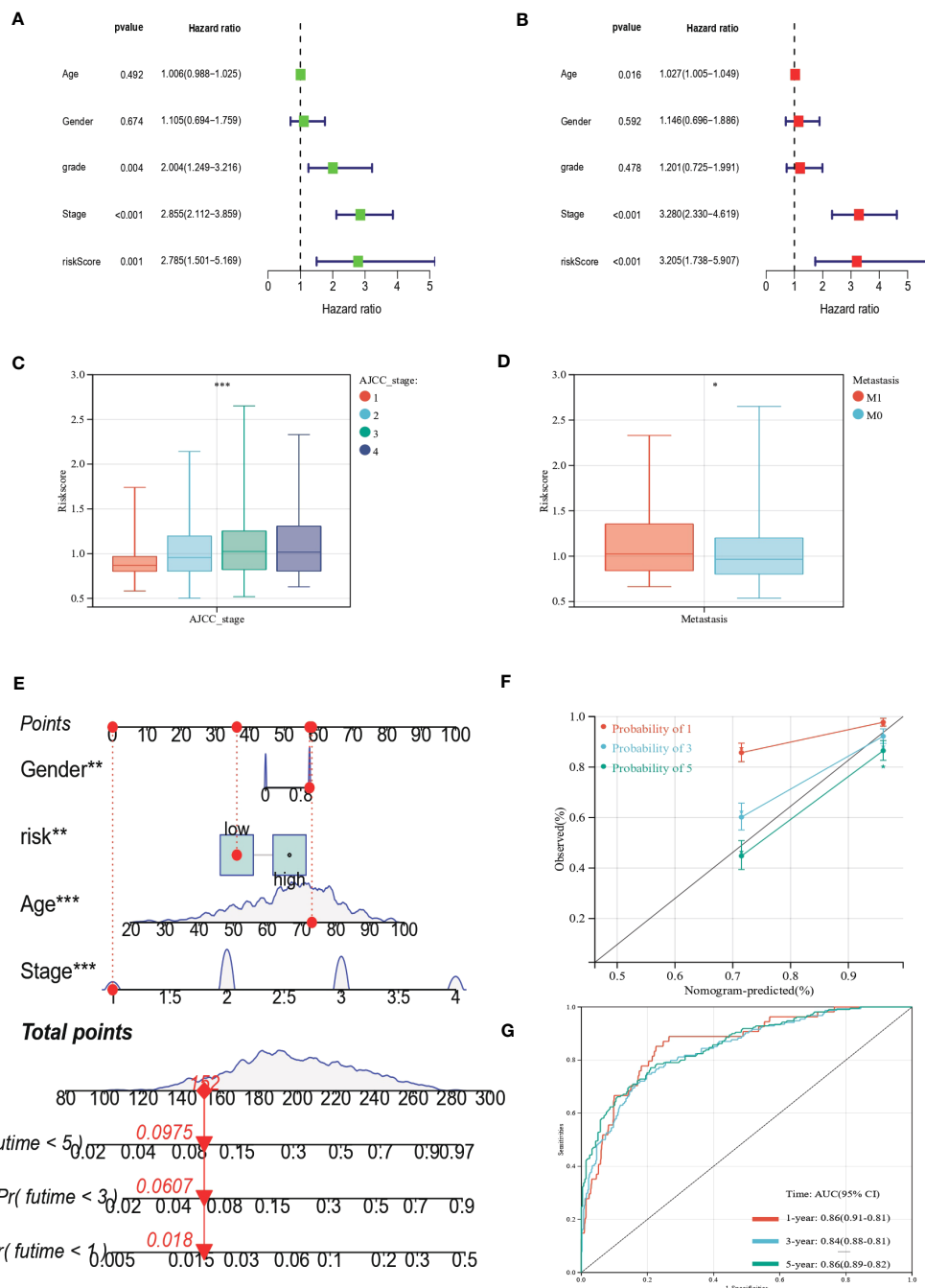


FIGURE 7

Clinical correlation analysis of the prognostic EAG_Score and establishment of the prognostic nomogram. (A, B) uniCox and multiCox analysis showed the prognostic value of the EAG_score. (C) Correlation between risk score and AJCC_stage. (D) Correlation between risk score and CRC metastasis. (E) Nomogram for predicting the 1-, 3-, and 5-year OS of CRC patients in the entire cohort. (F) Calibration curve of the prognostic nomogram. (G) ROC curves of the prognostic nomogram for 1-, 3-, and 5-year OS in CRC. ($p < 0.05$ *; $p < 0.01$ **; $p < 0.001$ ***).

factor superfamily, and they were discrepantly represented in the two risk subgroups (Figure 8L).

Drug sensitivity analysis

To assess the ability of EAG_Score in predicting clinical drug therapy sensitivity in CRC patients, we calculated the IC₅₀ values for each patient for 138 drugs using the “pRRophetic” package. We

found that patients with low EAG_Score may have positive responses to Salubrinal, Pyrimethamine, Lenalidomide, and OSI-906, while patients with high EAG_Score may have positive responses to ATRA, Cisplatin, Gemcitabine, Bleomycin Bortezomib, Docetaxel, Doxorubicin, Etoposide and some targeted drugs such as Axitinib, Dasatinib, Imatinib, Sunitinib, Nilotinib, etc (Figure 9). In conclusion, these results suggested that ECM remodeling genes were correlated with drug sensitivity.

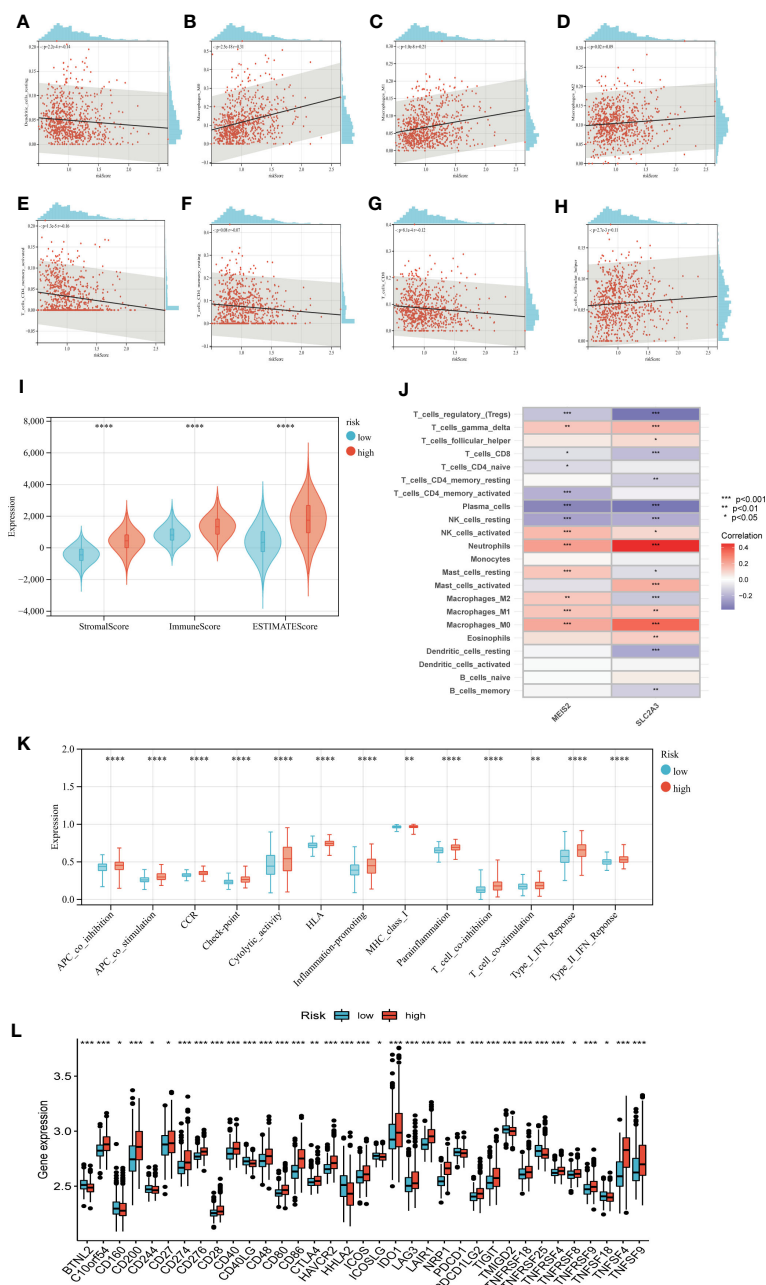


FIGURE 8

Evaluation of TME, checkpoints and immune functions between the two groups. (A–H) Correlations between EAG_score and immune cell types. (I) Correlations between EAG_score and both immune and stromal scores. (J) Correlations between the abundance of immune cells and genes involved in development of the prognostic model. (K) Assessment of differences in immune function between the two groups (L) Expression of 35 common immune checkpoints in the two groups. (p<0.05 *; p<0.01 **; p<0.001 ***).

Discussion

In cancer, the uncontrolled growth of cells remodels the ECM, this regulates cell-cell and cell-ECM interactions in turn and has profound effects on the biological behavior of cells (26). Various ECM components have pro- or anti-apoptotic effects, and dysregulation of the homeostasis of the ECM microenvironment is associated with tumor apoptotic evasion and progression (27). Clinically speaking, increased tissue stiffness can be observed in many solid tumors such as CRC, mainly due to alterations in tumor

tissue fibrosis caused by increased synthesis and cross-linking of collagen (28). Tumor cells, CAF, and immune cells play a decisive role in this process, while this change in tissue remodeling can in turn induce CAF, stromal cells to secrete various cytokines, growth factors, chemokines, and exosomes, this process provides a prerequisite for tumor cell proliferation and metastasis (29). Besides, a growing number of studies have demonstrated that there is an inextricable relationship between ECM remodeling and tumor immunity (30, 31). CDH11 promotes immunosuppression and ECM deposition to support the growth

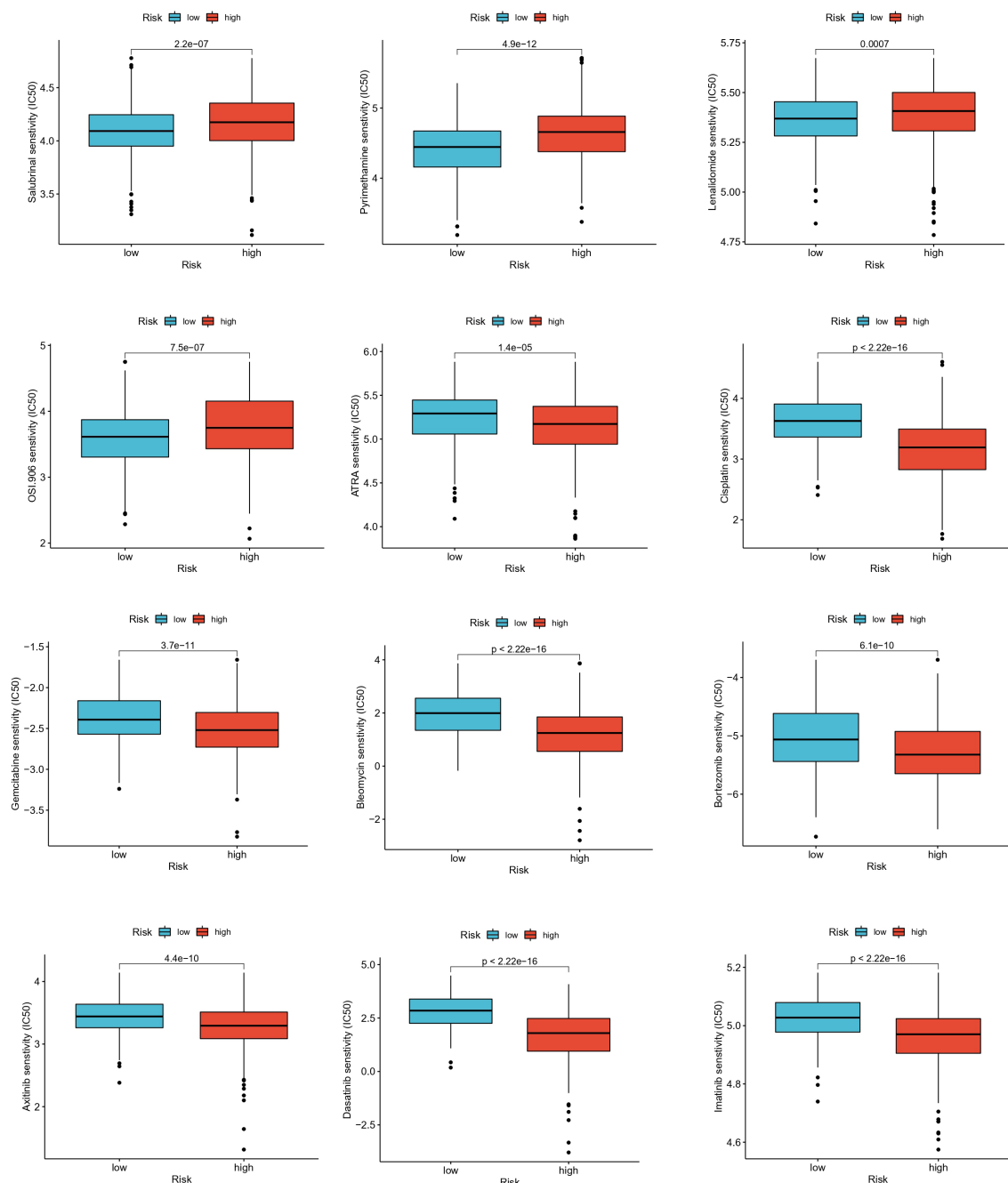


FIGURE 9
Relationships between EAG_score and drug sensitivity.

of pancreatic tumors and resistance to gemcitabine (32). Versican (VCAN), a large matrix proteoglycan with immunoregulatory activity, has been reported to prevent the interaction between hyaluronan and T cells to inhibit adhesion and migration (33), and versican-derived matrikines regulate Batf3-dendritic cell differentiation and promote CD8⁺ T-cell infiltration in CRC (34). In addition, Pirfenidone, an anti-fibrotic drug, promotes immune infiltration and enhances the efficacy of PD-L1 blockers in mouse models (35). Therefore, it is of great clinical importance to further investigate the association between ECM remodeling and tumor development and immunotherapy.

In the current research, using data from GEO dataset, we evaluated the expression of genes associated with ECM remodeling in CRC and normal tissues, and patients could be classified into two different subtypes based on ECM-related genes and DEGs, respectively. Moreover, we observed a significant difference in clinical features and prognosis, immune cell and stromal cell infiltration between the different subtypes. We developed the EAG_Score model and found that EAGcluster1 and geneClusterA had higher EAG_Score and worse prognosis, which confirmed the role of ECM remodeling in CRC prognosis. In addition, the EAG_Score was identified as an independent predictor of prognosis for CRC patients, and the ROC

curve validated its reliable predictive ability for 1-year, 3-year, and 5-year OS. Furthermore, the ECM remodeling-related prognostic nomograms showed good agreement between predicted and actual survival outcomes and a better prognostic ability than the TNM stage. In general, increases in infiltrating CD8⁺ T cells have been associated with longer OS (36), and consistent with that, we found the high-risk group showed more pronounced immunosuppressive features: with higher levels of Macrophages_M0, Macrophages_M1, Macrophages_M2 and lower levels of T_cells_CD4_memory_activated, T_cells_CD8, etc. Besides, the high-risk group had more abundant expression of ECM remodeling-related genes, and higher stromal and immune scores, suggesting that there is a relationship between ECM remodeling and tumor immunity that cannot be ignored. Considering the differences in immune function and immune checkpoint expression, we speculated that the high-risk group may have a higher sensitivity to immunotherapy. As research on the connection between the ECM and tumor immunity continues, ECM remodeling, structural plasticity, and mechanical forces are increasingly recognized as key factors in immune cell migration and spatial distribution, activation, and immune synapse formation (37–39). Therefore, targeting the ECM-mediated immunosuppressive microenvironment in combination with other systemic treatment strategies for CRC may enhance the efficacy of these therapies. Notably, capecitabine has been shown to inhibit the expression of CTLA-4 in CRC cells, which may combine immunotherapy with chemotherapy in the comprehensive treatment of CRC (40). These findings show promising prospects for targeting immune checkpoints in the integrated treatment of CRC. In recent years, although research on ECM such as targeting angiogenesis and tumor immunity has made tremendous progress (41), few clinical translations have been achieved (42), which may be partly explained by the non-specificity of drugs and complex tumor immunity (43). A systematic understanding of ECM remodeling and the complicated TME generated by stromal elements will help to identify investigational targets for the development of novel immune biomarkers and combination immunotherapy.

Meis homeobox 2 (MEIS2) belongs to TALE (three amino-acid loop extension) superfamily and is mainly involved in the Hox activity regulation by binding directly with posterior Hox proteins or indirectly with Pbx to form a homeoprotein-DNA complex, plays a crucial role in the pathogenesis of human cancer (44, 45). MEIS2 is essential for neuroblastoma cell survival and up-regulation of MEIS2 is required for the growth of AML1-ETO-positive AML (46, 47). In addition, it has been reported that MEIS2 may be associated with certain metastatic diseases (48, 49). Consistent with this, we found that mCRC patients had higher MEIS2 expression levels in our study. However, it has been also shown that high expression of MEIS2 is associated with improved prognosis of ovarian cancer (50). Therefore, further studies are needed to clarify the specific role of MEIS2 in cancer.

SLC2A3, a member of the solute carrier 2A family, encodes glucose transporter protein (51). SLC2A3 is up-regulated in a variety of tumors, such as CRC and breast cancer, and is involved in tumor progression and poor prognosis (52, 53). In cancer cells, the elevated expression of SLC2A3 helps to meet the increased glycolysis requirements and promotes the Warburg effect (54). In CRC, activation of the SLC2A3-YAP signaling pathway is a master

activator that reprograms tumor metabolism and thus promotes tumor metastasis (55). Consistent with our findings, studies have indicated changes in the ECM of SLC2A3 overexpressed cells, including upregulation of osteopontin, which has been shown to mediate cell adhesion and promote tumor metastasis (56). In addition, we found a correlation between SLC2A3 and a variety of immune cells, such as CD4⁺ and CD8⁺ T cells, macrophages, etc. Studies have suggested that SLC2A3 promotes the growth and drug resistance of gastric cancer cells by increasing the infiltration of M2 macrophages (57). Further exploration of the interaction between SLC2A3 up-regulated cancer cells and immune cells is of major importance and may provide new insights for cancer immunotherapy.

Conclusion

In short, this study provides a comprehensive description of ECM remodeling and TME, prognosis, and clinical characteristics of CRC patients, reveals the important clinical significance of ECM remodeling related genes, and provides valuable insights for individualized therapy of CRC patients. However, this study has some shortcomings. First, data related to patients receiving immunotherapy are missing in our study, external validation based on prospective and large-scale clinical trials is needed to assess the predictive power of the model in the future. Second, the interaction between the genes involved in constructing the model and immune cells in CRC also needs to be tested experimentally. Third, potential mechanisms between ECM remodeling and tumor immunity should be revealed in the future.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

Conceptualization, TC and YZ. Methodology, TC and WC. Software, WC. Validation, WC, YW, CC and HG. Formal analysis, WC. Investigation, YW. Data curation, WC, HG and YW. Writing—original draft preparation, WC and ZY. Writing—review and editing, TC, CC, TY and YZ. Visualization, WC. Supervision, YZ and TY. Project administration, TC. Funding acquisition, TC, and YZ. All authors have read and agreed to the published version of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This research was funded by the National Nature Science Foundation of China (NSFC) (Grant No. 81602431).

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.

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/fonc.2023.1109181/full#supplementary-material>

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistic, 2020. *CA Cancer J Clin* (2020) 70:7–30. doi: 10.3322/caac.21590
2. O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *J Natl Cancer Inst* (2004) 96:1420–5. doi: 10.1093/jnci/djh275
3. Benson AB, Venook AP, Al-Hawary MM, Cederquist L, Chen YJ, Ciombor KK, et al. NCCN guidelines insights: colon cancer, version 2.2018. *J Natl Compr Canc Netw* (2018) 16:359–69. doi: 10.6004/jnccn.2018.0021
4. Benson AB, Venook AP, Al-Hawary MM, Arain MA, Chen YJ, Ciombor KK, et al. NCCN guidelines insights: rectal cancer, version 6.2020. *J Natl Compr Canc Netw* (2020) 18:806–15. doi: 10.6004/jnccn.2020.0032
5. Mohan V, Das A, Sagi I. Emerging roles of ECM remodeling processes in cancer. *Semin Cancer Biol* (2020) 62:192–200. doi: 10.1016/j.semcancer.2019.09.004
6. Padhi A, Nain AS. ECM in differentiation: A review of matrix structure, composition and mechanical properties. *Ann BioMed Eng* (2020) 48:1071–89. doi: 10.1007/s10439-019-02337-7
7. Kay EJ, Koulouras G, Zanivan S. Regulation of extracellular matrix production in activated fibroblasts: roles of amino acid metabolism in collagen synthesis. *Front Oncol* (2021) 11:719922. doi: 10.3389/fonc.2021.719922
8. Rizzo G, Rubbino F, Elangovan S, Sammarco G, Lovisa S, Restelli S, et al. Dysfunctional extracellular matrix remodeling supports perianal fistulizing crohn's disease by a mechanoregulated activation of the epithelial-to-mesenchymal transition. *Cell Mol Gastroenterol Hepatol* (2023) 15:741–64. doi: 10.1016/j.jcmgh.2022.12.006
9. Ricard-Blum S, Vallet SD. Fragments generated upon extracellular matrix remodeling: Biological regulators and potential drugs. *Matrix Biol* (2019) 75–76:170–89. doi: 10.1016/j.matbio.2017.11.005
10. Girigoswami K, Saini D, Girigoswami A. Extracellular matrix remodeling and development of cancer. *Stem Cell Rev Rep* (2021) 17:739–47. doi: 10.1007/s12015-020-10070-1
11. Winkler J, Abisoye-Ogunniyan A, Metcalf KJ, Werb Z. Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nat Commun* (2020) 11:5120. doi: 10.1038/s41467-020-18794-x
12. Ganesh K, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat Rev Gastroenterol Hepatol* (2019) 16:361–75. doi: 10.1038/s41575-019-0126-x
13. Gordon-Weeks A, Yuzhalin AE. Cancer extracellular matrix proteins regulate tumour immunity. *Cancers (Basel)* (2020) 12:3331. doi: 10.3390/cancers12113331
14. Peng DH, Rodriguez BL, Diao L, Chen L, Wang J, Byers LA, et al. Collagen promotes anti-PD-1/PD-L1 resistance in cancer through LAIR1-dependent CD8 (+) T cell exhaustion. *Nat Commun* (2020) 11:4520. doi: 10.1038/s41467-020-18298-8
15. Jung WH, Yam N, Chen CC, Elawad K, Hu B, Chen Y. Force-dependent extracellular matrix remodeling by early-stage cancer cells alters diffusion and induces carcinoma-associated fibroblasts. *Biomaterials* (2020) 234:119756. doi: 10.1016/j.biomaterials.2020.119756
16. Marisa L, de Reynies 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:e1001453. doi: 10.1371/journal.pmed.1001453
17. Chen MS, Lo YH, Chen X, Williams CS, Donnelly JM, Criss ZK, et al. Growth factor-independent 1 is a tumor suppressor gene in colorectal cancer. *Mol Cancer Res* (2019) 17:697–708. doi: 10.1158/1541-7786.MCR-18-0666
18. Gene Ontology C. Gene Ontology Consortium: going forward. *Nucleic Acids Res* (2015) 43:D1049–1056. doi: 10.1093/nar/gku1179
19. Briere G, Darbo E, Thebault P, Uricaru R. Consensus clustering applied to multi-omics disease subtyping. *BMC Bioinf* (2021) 22:361. doi: 10.1186/s12859-021-04279-1
20. Rich JT, Neely JG, Paniello RC, Voelker CC, Nussenbaum B, Wang EW. A practical guide to understanding Kaplan-Meier curves. *Otolaryngol Head Neck Surg* (2010) 143:331–6. doi: 10.1016/j.otohns.2010.05.007
21. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* (2012) 16:284–7. doi: 10.1089/omi.2011.0118
22. Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PloS One* (2014) 9:e107468. doi: 10.1371/journal.pone.0107468
23. Hu D, Zhou Z, Wang J, Zhu K. Screening of ferroptosis-related genes with prognostic effect in colorectal cancer by bioinformatic analysis. *Front Mol Biosci* (2022) 9:979854. doi: 10.3389/fmolb.2022.979854
24. Wang P, Zhao W, Cao H. Development of a platelet-related prognostic model for colorectal cancer. *Front Genet* (2022) 13:904168. doi: 10.3389/fgene.2022.904168
25. Cai Z, Zhou F. A novel Anoikis and immune-related genes marked prognostic signature for colorectal cancer. *Med (Baltimore)* (2022) 101:e31127. doi: 10.1097/MD.00000000000031127
26. Oskarsson T. Extracellular matrix components in breast cancer progression and metastasis. *Breast* (2013) 22 Suppl 2:S66–72. doi: 10.1016/j.breast.2013.07.012
27. Mott JD, Werb Z. Regulation of matrix biology by matrix metalloproteinases. *Curr Opin Cell Biol* (2004) 16:558–64. doi: 10.1016/j.ceb.2004.07.010
28. Wei B, Zhou X, Liang C, Zheng X, Lei P, Fang J, et al. Human colorectal cancer progression correlates with LOX-induced ECM stiffening. *Int J Biol Sci* (2017) 13:1450–7. doi: 10.7150/ijbs.21230
29. Asif PJ, Longobardi C, Hahne M, Medema JP. The role of cancer-associated fibroblasts in cancer invasion and metastasis. *Cancers (Basel)* (2021) 13:4720. doi: 10.3390/cancers13184720
30. Kaur A, Ecker BL, Douglass SM, Kugel CH3rd, Webster MR, Almeida FV, et al. Remodeling of the collagen matrix in aging skin promotes melanoma metastasis and affects immune cell motility. *Cancer Discovery* (2019) 9:64–81. doi: 10.1158/2159-8290.CD-18-0193
31. Marozzi M, Parnigoni A, Negri A, Viola M, Vigetti D, Passi A, et al. Inflammation, extracellular matrix remodeling, and proteostasis in tumor microenvironment. *Int J Mol Sci* (2021) 22:8102. doi: 10.3390/ijms22158102
32. Peran I, Dakshanamurthy S, McCoy MD, Mavropoulos A, Allo B, Sebastian A, et al. Cadherin 11 promotes immunosuppression and extracellular matrix deposition to support growth of pancreatic tumors and resistance to gemcitabine in mice. *Gastroenterology* (2021) 160:1359–1372.e1313. doi: 10.1053/j.gastro.2020.11.044
33. Evanko SP, Potter-Perigo S, Bollyky PL, Nepom GT, Wight TN. Hyaluronan and versican in the control of human T-lymphocyte adhesion and migration. *Matrix Biol* (2012) 31:90–100. doi: 10.1016/j.matbio.2011.10.004
34. Hope C, Emmerich PB, Papadas A, Pagenkopf A, Matkowskyj KA, Van De Hey DR, et al. Versican-derived matrikines regulate batf3-dendritic cell differentiation and promote T cell infiltration in colorectal cancer. *J Immunol* (2017) 199:1933–41. doi: 10.4049/jimmunol.1700529
35. Qin W, Zou J, Huang Y, Liu C, Kang Y, Han H, et al. Pirfenidone facilitates immune infiltration and enhances the antitumor efficacy of PD-L1 blockade in mice. *Oncotarget* (2020) 9:1824631. doi: 10.1080/2162402X.2020.1824631
36. Fridman WH, Zitvogel L, Sautes-Fridman C, Kroemer G. The immune contexture in cancer prognosis and treatment. *Nat Rev Clin Oncol* (2017) 14:717–34. doi: 10.1038/nrclinonc.2017.101

37. Hynes RO. The extracellular matrix: not just pretty fibrils. *Science* (2009) 326:1216–9. doi: 10.1126/science.1176009
38. Hallmann R, Zhang X, Di Russo J, Li L, Song J, Hannocks MJ, et al. The regulation of immune cell trafficking by the extracellular matrix. *Curr Opin Cell Biol* (2015) 36:54–61. doi: 10.1016/j.ceb.2015.06.006
39. Huse M. Mechanical forces in the immune system. *Nat Rev Immunol* (2017) 17:679–90. doi: 10.1038/nri.2017.74
40. Derakhshani A, Hashemzadeh S, Asadzadeh Z, Shadbad MA, Rasibonab F, Safarpour H, et al. Cytotoxic T-lymphocyte antigen-4 in colorectal cancer: another therapeutic side of capecitabine. *Cancers (Basel)* (2021) 13:2414. doi: 10.3390/cancers13102414
41. Roma-Rodrigues C, Mendes R, Baptista PV, Fernandes AR. Targeting tumor microenvironment for cancer therapy. *Int J Mol Sci* (2019) 20:840. doi: 10.3390/ijms20040840
42. Abyaneh HS, Regenold M, McKee TD, Allen C, Gauthier MA. Towards extracellular matrix normalization for improved treatment of solid tumors. *Theranostics* (2020) 10:1960–80. doi: 10.7150/thno.39995
43. Mushtaq MU, Papadas A, Pagenkopf A, Flietner E, Morrow Z, Chaudhary SG, et al. Tumor matrix remodeling and novel immunotherapies: the promise of matrix-derived immune biomarkers. *J Immunother Cancer* (2018) 6:65. doi: 10.1186/s40425-018-0376-0
44. Geerts D, Schilderink N, Jorritsma G, Versteeg R. The role of the MEIS homeobox genes in neuroblastoma. *Cancer Lett* (2003) 197:87–92. doi: 10.1016/s0304-3835(03)00087-9
45. Delgado I, Giovinnazzo G, Temiño S, Gauthier Y, Balsalobre A, Drouin J, et al. Control of mouse limb initiation and antero-posterior patterning by Meis transcription factors. *Nat Commun* (2021) 12. doi: 10.1038/s41467-021-23373-9
46. Zha Y, Xia Y, Ding J, Choi JH, Yang L, Dong Z, et al. MEIS2 is essential for neuroblastoma cell survival and proliferation by transcriptional control of M-phase progression. *Cell Death Dis* (2014) 5:e1417. doi: 10.1038/cddis.2014.370
47. Vegi NM, Klappacher J, Oswald F, Mulaw MA, Mandoli A, Thiel VN, et al. MEIS2 is an oncogenic partner in AML1-ETO-positive AML. *Cell Rep* (2016) 16:498–507. doi: 10.1016/j.celrep.2016.05.094
48. Bhanvadia RR, VanOpstall C, Brechka H, Barashi NS, Gillard M, McAuley EM, et al. MEIS1 and MEIS2 expression and prostate cancer progression: A role for HOXB13 binding partners in metastatic disease. *Clin Cancer Res* (2018) 24:3668–80. doi: 10.1158/1078-0432.CCR-17-3673
49. Wan Z, Chai R, Yuan H, Chen B, Dong Q, Zheng B, et al. MEIS2 promotes cell migration and invasion in colorectal cancer. *Oncol Rep* (2019) 42:213–23. doi: 10.3892/or.2019.7161
50. Crijns AP, de Graeff P, Geerts D, Ten Hoor KA, Hollema H, van der Sluis T, et al. MEIS and PBX homeobox proteins in ovarian cancer. *Eur J Cancer* (2007) 43:2495–505. doi: 10.1016/j.ejca.2007.08.025
51. Cesar-Razquin A, Snijder B, Frappier-Brinton T, Isserlin R, Gyimesi G, Bai X, et al. A call for systematic research on solute carriers. *Cell* (2015) 162:478–87. doi: 10.1016/j.cell.2015.07.022
52. Onodera Y, Nam JM, Bissell MJ. Increased sugar uptake promotes oncogenesis via EPAC/RAP1 and O-GlcNAc pathways. *J Clin Invest* (2014) 124:367–84. doi: 10.1172/JCI63146
53. Tsai TH, Yang CC, Kou TC, Yang CE, Dai JZ, Chen CL, et al. Overexpression of GLUT3 promotes metastasis of triple-negative breast cancer by modulating the inflammatory tumor microenvironment. *J Cell Physiol* (2021) 236:4669–80. doi: 10.1002/jcp.30189
54. Ziegler GC, Almos P, McNeill RV, Jansch C, Lesch KP. Cellular effects and clinical implications of SLC2A3 copy number variation. *J Cell Physiol* (2020) 235:9021–36. doi: 10.1002/jcp.29753
55. Kuo CC, Ling HH, Chiang MC, Chung CH, Lee WY, Chu CY, et al. Metastatic colorectal cancer rewrites metabolic program through a glut3-YAP-dependent signaling circuit. *Theranostics* (2019) 9:2526–40. doi: 10.7150/thno.32915
56. Pietras A, Katz AM, Ekstrom EJ, Wee B, Halliday JJ, Pitter KL, et al. Osteopontin-CD44 signaling in the glioma perivascular niche enhances cancer stem cell phenotypes and promotes aggressive tumor growth. *Cell Stem Cell* (2014) 14:357–69. doi: 10.1016/j.stem.2014.01.005
57. Yao X, He Z, Qin C, Deng X, Bai L, Li G, et al. SLC2A3 promotes macrophage infiltration by glycolysis reprogramming in gastric cancer. *Cancer Cell Int* (2020) 20:503. doi: 10.1186/s12935-020-01599-9



OPEN ACCESS

EDITED BY
Nina Zidar,
University of Ljubljana, Slovenia

REVIEWED BY
Beom Jae Lee,
Korea University Guro Hospital,
Republic of Korea
Nina Hauptman,
University of Ljubljana, Slovenia

*CORRESPONDENCE
Veronika Holubekova
✉ veronika.holubekova@uniba.sk
Zora Lasabova
✉ zora.lasabova@uniba.sk

RECEIVED 15 April 2023

ACCEPTED 24 August 2023

PUBLISHED 05 October 2023

CITATION

Holubekova V, Loderer D, Grendar M,
Mikolajcik P, Kolkova Z, Turyova E,
Kudelova E, Kalman M, Marcinek J,
Miklusica J, Laca L and Lasabova Z (2023)
Differential gene expression of immunity
and inflammation genes in colorectal
cancer using targeted RNA sequencing.
Front. Oncol. 13:1206482.
doi: 10.3389/fonc.2023.1206482

COPYRIGHT

© 2023 Holubekova, Loderer, Grendar,
Mikolajcik, Kolkova, Turyova, Kudelova,
Kalman, Marcinek, Miklusica, Laca and
Lasabova. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). 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.

Differential gene expression of immunity and inflammation genes in colorectal cancer using targeted RNA sequencing

Veronika Holubekova^{1*}, Dusan Loderer¹, Marian Grendar²,
Peter Mikolajcik³, Zuzana Kolkova¹, Eva Turyova⁴,
Eva Kudelova³, Michal Kalman⁵, Juraj Marcinek⁵,
Juraj Miklusica³, Ludovit Laca³ and Zora Lasabova^{4*}

¹Laboratory of Genomics and Prenatal Diagnostics, Biomedical Center in Martin, Jessenius Faculty of Medicine, Comenius University in Bratislava, Martin, Slovakia, ²Laboratory of Bioinformatics and Biostatistics, Biomedical Center in Martin, Jessenius Faculty of Medicine, Comenius University in Bratislava, Martin, Slovakia, ³Clinic of Surgery and Transplant Center, Jessenius Faculty of Medicine in Martin, Comenius University in Bratislava, Martin University Hospital, Martin, Slovakia, ⁴Department of Molecular Biology and Genomics, Jessenius Faculty of Medicine in Martin, Comenius University in Bratislava, Martin, Slovakia, ⁵Department of Pathological Anatomy, Jessenius Faculty of Medicine, Comenius University in Bratislava, Martin University Hospital, Martin, Slovakia

Introduction: Colorectal cancer (CRC) is a heterogeneous disease caused by molecular changes, as driver mutations, gene methylations, etc., and influenced by tumor microenvironment (TME) pervaded with immune cells with both pro- and anti-tumor effects. The studying of interactions between the immune system (IS) and the TME is important for developing effective immunotherapeutic strategies for CRC. In our study, we focused on the analysis of expression profiles of inflammatory and immune-relevant genes to identify aberrant signaling pathways included in carcinogenesis, metastatic potential of tumors, and association of Kirsten rat sarcoma virus (KRAS) gene mutation.

Methods: A total of 91 patients were enrolled in the study. Using NGS, differential gene expression analysis of 11 tumor samples and 11 matching non-tumor controls was carried out by applying a targeted RNA panel for inflammation and immunity genes containing 475 target genes. The obtained data were evaluated by the CLC Genomics Workbench and R library. The significantly differentially expressed genes (DEGs) were analyzed in Reactome GSA software, and some selected DEGs were used for real-time PCR validation.

Results: After prioritization, the most significant differences in gene expression were shown by the genes *TNFRSF4*, *IRF7*, *IL6R*, *NR3C1*, *EIF2AK2*, *MIF*, *CCL5*, *TNFSF10*, *CCL20*, *CXCL11*, *RIPK2*, and *BLNK*. Validation analyses on 91 samples showed a correlation between RNA-seq data and qPCR for *TNFSF10*, *RIPK2*, and *BLNK* gene expression. The top differently regulated signaling pathways between the studied groups (cancer vs. control, metastatic vs. primary CRC and KRAS positive and negative CRC) belong to immune system, signal transduction, disease, gene expression, DNA repair, and programmed cell death.

Conclusion: Analyzed data suggest the changes at more levels of CRC carcinogenesis, including surface receptors of epithelial or immune cells, its signal transduction pathways, programmed cell death modifications, alterations in DNA repair machinery, and cell cycle control leading to uncontrolled proliferation. This study indicates only basic molecular pathways that enabled the formation of metastatic cancer stem cells and may contribute to clarifying the function of the IS in the TME of CRC. A precise identification of signaling pathways responsible for CRC may help in the selection of personalized pharmacological treatment.

KEYWORDS

immune system, inflammation, RNA sequencing, signaling pathways, colorectal cancer, KRAS mutation, differentially expressed genes (DEG)

1 Introduction

Colorectal cancer (CRC) was the main type of cancer (17.4% of all cancer cases) in all ages of male population and the second most common cancer (14.2% of all cancer cases) in all ages of female population in Slovakia in 2020 (1). It is unflattering that Slovakia had globally the highest overall mortality rate from CRC in 2020 with age-standardized rate of 21.0 per 100,000 residents (2).

The pathogenesis of CRC is a complex interplay between genetic, epigenetic, and environmental factors. Understanding of these factors and their interactions is important for developing effective prevention and treatment strategies for this disease. A family history of CRC or colorectal polyps, familial adenomatous polyposis (FAP) or Lynch syndrome, and inflammatory bowel disease, increase a risk of CRC development in the later age (3). Lifestyle factors usually influence the development of CRC in an epigenetic or environmental mode. The well-known risk factors are overweight and/or obesity and the lack of regular physical activity, alcohol consumption, and smoking. The dietary habits in the form of low intake of fiber, fruit, and vegetables, and high intake of fat and high proportion of processed meats contribute to CRC development (3, 4).

The most cases of CRCs develop from benign polyps. These polyps often contain genetic mutations that activate oncogenes or inactivate tumor suppressor genes, leading to uncontrolled cell growth and division. The pathogenesis of CRC is characterized by genetic instability mediated through chromosomal instability, microsatellite instability (MSI), and CpG island methylator phenotype (CIMP) pathways (5).

The chromosomal instability (CIN) is observed in 65%–85% of sporadic CRCs, leading to gains or losses of large or whole proportions of chromosomes and aneuploidy in number of chromosomes, genomic amplifications, and loss of heterozygosity (LOH) (6). CIN is characterized by the inactivation of tumor-suppressor genes, as tumor protein p53 (*TP53*) and adenomatous polyposis coli (*APC*), activation of oncogenes, as *KRAS* and B-raf proto-oncogene (*BRAF*), and LOH of long arm of chromosome 18. The onset of colorectal carcinogenesis is the adenomatous stage of

epithelial cells present with *APC* gene silencing and *KRAS* mutations. Subsequent inactivation of *TP53* and deletion of 18q chromosome leads to malignant transformation (7).

Microsatellite instability (MSI) presented in approximately 12%–15% of all CRCs is caused by defects in DNA mismatch repair (MMR) genes. Whereas MSI in hereditary CRC (Lynch syndrome) is caused by germline mutation in MMR genes, in the case of sporadic CRC, MSI-high (MSI-H) status is associated with CIMP-positive state resulting in hypermethylation of *mutL* homolog 1 (*MLH1*) and many other tumor-suppressor genes (8). Anyway, patients with MSI-H phenotype have better prognosis and higher response rate to immunotherapy (9, 10). On the other hand, approximately 85% of colorectal tumors are microsatellite stable (MSS) with worse prognosis and lower response to immunotherapy compared to MSI-H CRC (11, 12).

Metastatic CRC (mCRC) is the most common cause of death for CRC patients that have a poor 5-year survival fewer than 20%. Genomic profiling of somatic variants is very important for treatment decision and prediction of patient outcomes. Of the patients with mCRC carrying *KRAS/neuroblastoma RAS Viral oncogene Homolog (NRAS)/BRAF* wild-type tumors, 50% should be treated with combination of epidermal growth factor receptor (EGFR) monoclonal antibodies in combination with chemotherapy with few months extension in median survival. However, no effective targeted therapy is yet available for 35%–40% of patients with *KRAS/NRAS* mutation. Tumors carrying *BRAF* mutation V600E (5%–10% of mCRC) might be treated with a combination therapy with *BRAF* and EGFR inhibitors that may prolong the survival to 9.3 months (13).

The immune system plays a complex role in the pathogenesis and progression of colorectal cancer, with both pro- and anti-tumor effects. The studying of interactions between the IS and TME is important for developing effective immunotherapeutic strategies for CRC. In our study, we focused on the analysis of expression profiles of inflammatory and immune-relevant genes to identify aberrant signal pathways included in carcinogenesis, metastatic potential of tumors, and association of *KRAS* mutation with molecular signaling. Studying of molecular mechanisms of CRC pathogenesis and

progression, immune system deregulation, and its interaction with TME could bring new possibilities in diagnosis, treatment strategies, and identification new potential immunotherapy targets.

2 Methods

2.1 Patients, clinical tissue samples, and RNA extraction

In total, 91 CRC patients underwent resection of CRC at Clinic of General, Visceral and Transplant Surgery, University Hospital Martin. Samples were collected from cancer (n=91) and adjacent (n=71) tissues in collaboration with Department of Pathological Anatomy, University Hospital Martin. The patient characteristics are presented in [Table 1](#). The primary (n=52) and metastatic tumors (n=30) with metastases in the liver were evaluated by the experienced pathologist, and tumor tissue surgical excisions were immersed into Dulbecco's modified Eagle's medium (DMEM), penicillin/streptomycin, and 10% of fetal bovine serum and stored at 4°C. The samples treated this way were delivered to the Department of Molecular Biology and Genomics and transferred to RNA later and stored at -80°C until RNA extraction. Other sample criterion of the tissue collection above was the presence (n=34) and absence (n=49) of KRAS mutation. Tissues were homogenized, and total RNA was extracted using RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The eluted RNA was then stored at -80°C until reverse transcription reaction or RNA-seq libraries preparation.

2.2 RNA sequencing of inflammation and immunity gene expression

Using next generation RNA sequencing, a differential gene expression analysis of 11 tumor and 11 matching non-tumor tissues was carried out by applying a targeted RNA panel for inflammation and immunity genes (Qiagen, Germany) containing 475 target genes and 25 reference normalization genes. Extracted RNA was quantified by fluorometric quantitation (Qubit 3.0, Invitrogen, USA), and RNA integrity was checked by 2100 Bioanalyzer (Agilent). RNA-seq libraries for CRC and adjacent tissue were prepared using QIAseq Targeted RNA Human Inflammation & Immunity Transcriptome (Qiagen). The panel is designed for detection of DEGs for pro- or antiapoptotic genes, cytokines, chemokines, growth factors and their receptors, and transcription factors performing various functions of the immune system. cDNA libraries were prepared from 450 ng of RNA and assigned with Unique Molecular Indexes (UMIs). Quality control was performed by Bioanalyzer, and libraries were sequenced using single-end reads, 1× 150 bp on MiSeq (Illumina, USA), to a depth of 5 million reads. Particular investigation of the role of significantly deregulated DEGs has been provided by NcPath online analysis tool, which allows visualization and enrichment for non-coding RNA and KEGG signaling pathways in humans to elucidate the physiological and pathological processes in CRC (14).

2.3 Experimental validation of RNA-seq data by qPCR

Six differentially expressed genes were selected for validation using quantitative real-time PCR (qPCR). The 500 ng of sample RNA and reference RNA (Universal Human Reference RNA, Invitrogen) was reversely transcribed to cDNA using High-Capacity cDNA Reverse Transcription Kit with RNase Inhibitor (Applied Biosystems). In the less concentrated samples, the initial concentration of 250 ng was processed to reverse transcription (RT). No Enzyme Control (NEC) was also included in each series of transcription. The temperature steps of RT include incubation at 25°C for 10 min, RT at 37°C for 120 min, and enzyme inactivation at 85°C for 5 min. Samples were stored at -20°C until further use.

The initial step of relative quantification (RQ) includes 10-fold serial dilutions of reference RNA from 25 ng to 2.5pg. The six target assays, namely, receptor interacting serine/threonine kinase 2 (Applied Biosystems, RIPK2, Hs01572686_m1, FAM-MGB), interleukin 6 receptor (IL6R, Hs01075664_m1, FAM-MGB), TNF superfamily member 10 (TNFSF10, Hs00921974_m1, VIC-MGB), MHC class I polypeptide-related sequence B (MICB, Hs00792952_m1, FAM-MGB), C-C motif chemokine receptor 4 (CCR4, Hs01396342_m1, FAM-MGB), and B-cell linker (BLNK, Hs00179459_m1, VIC-MGB), and two housekeeping genes, actin beta (ACTB, Hs99999903_m1, VIC-MGB) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH, Hs99999905_m1, VIC-MGB), were tested. Each duplex reaction was run with TaqMan™ Fast Advanced Master Mix (Applied Biosystems) in the total volume of 20 µl with thermal cycling conditions, as incubation at 50°C for 2 min, polymerase activation at 95°C for 2 min, and 40 cycles with denaturation at 95°C for 3 s and annealing/extension at 60°C for 30 s in the instrument 7500 Fast Real-Time PCR System (Applied Biosystems). Standard curves were created for each tested assay from dilution series in the 7500 instrument software, and pairs of assays labeled with FAM and VIC have been chosen for duplex reactions (IL6R+BLNK, MICB+TNFSF10, RIPK2+GAPDH, and CCR4+ACTB). Data are not present in the study.

For validation of RNA sequencing by RQ, 91 tumors and 71 adjacent tissues were selected. Samples were analyzed in duplicates for each duplex reaction, and all pipetting steps were performed on BRAVO Liquid Handling Station (Agilent) to minimize subjective pipette handling bias. The results were analyzed in the instrument software with unique threshold setting and cycle threshold (Ct value) calculation.

2.4 Pathway analysis using Reactome

RNA sequencing data from immune and inflammation panel were processed in inter-run normalization to eliminate technical differences between four runs. The corrected data were analyzed in Reactome data online analysis tool. Reactome (15), as online data analysis tool, was used to map the biological pathways influenced by DEGs identified in the study. The batch-corrected RNA-seq data of four experiments (six samples per run, including tumor and adjacent tissue) have been submitted to gene expression analysis.

TABLE 1 Clinicopathological characteristics of patients included in the study.

Patient characteristics		Number of patients (n)
Average age		67.5 (SD ± 9.5)
BMI		27.9 (SD ± 4.7)
Gender	Female	39 (42.9%)
	Male	52 (57.1%)
Grade	G1	15 (16.5%)
	G2	41 (45%)
	G3	13 (14.3%)
	N/A	22 (24.2%)
T stage	T1	3 (3.3%)
	T2	15 (16.5%)
	T3	49 (53.8%)
	T4	16 (17.6%)
	N/A	8 (8.8%)
N stage	N0	39 (42.9%)
	N1	29 (31.9%)
	N2	14 (15.3%)
	N/A	9 (9.9%)
M stage	M0	41 (45%)
	M1	38 (41.8%)
	N/A	12 (13.2%)
Clinical stage	I	9 (9.9%)
	II	20 (22%)
	III	15 (16.5%)
	IV	35 (38.4%)
	N/A	12 (13.2%)
Tumor type	Primary tumor	50 (54.9%)
	Tumor with liver metastases	32 (35.2%)
	N/A	9 (9.9%)
KRAS gene mutation	present	34 (37.4%)
	absent	49 (53.8%)
	N/A	8 (8.8%)

CRC, colorectal cancer; BMI, body mass index; N/A, not applicable; T, tumor size; N, lymph nodes positive for tumor; M, metastasized cancer; KRAS, Kirsten rat sarcoma virus.

The PADOG (16) (weighted gene set analysis method that downweights genes that are present in many pathways) method has been applied to RNA-seq normalized data, recorded with annotations as cancer or control, primary or metastatic tumor or control, and KRAS mutation positive or negative. Overall, the DEGs from the RNA-seq study of 11 CRC and 11 adjacent tissues have

been enriched in 1,163 biological pathways, and 461 DEGs were identified. The selected pathways with a p-value for entity <0.05 have been considered as statistically significant.

2.5 Bioinformatics processing and data analysis

2.5.1 RNA-seq data evaluation

Sequencing data (fastq files) were imported into CLC Genomics Workbench (GW) v. 21.0.4 and processed by the RNA-seq analysis pipeline with the default settings. The report, generated by the pipeline, was utilized for obtaining information on the quality control. Transcripts per million (TPM), which are normalized for sequencing depth so their values are comparable between samples, were used in data analysis outside CLC. Differential expression analysis was performed using differential expression for RNA-seq pipeline with meta-data specifying both groups (e.g., case vs. control) and run, so that the batch correction was included into the computation of the average fold change (averaged over the “case” samples). A heatmap was created using Create Heat Map for RNA-Seq tool. In order to obtain a set of differentially expressed genes (DEGs), we used the elastic network Machine Learning algorithm. The TPM data were first subjected to batch correction by the scBatch method (17), scaled and then fed to the elastic network algorithm (18) (see, e.g., (19), for an overview of ML in RNA-seq), using R (20), ver. 4.0.5. Genes with positive variable importance were selected as DEGs. CLC was used to create heatmap for DEGs.

2.5.2 RT-PCR

Fold change (FC) was computed using the standard formula (i.e., $2^{-\Delta\Delta C_t}$). The data on FC and $\log_2(\text{FC})$ were explored in R, by means of boxplot overlaid with swarmplot. Distribution of $\log_2(\text{FC})$ was assessed by the quantile–quantile plot with the 95% confidence band constructed by bootstrap. Since the distribution of $\log_2(\text{FC})$ was either Gaussian, close to Gaussian, or symmetric, we used the Wilcoxon test to test the null hypothesis that the population median (point of symmetry) of $\log_2(\text{FC})$ is 0. p-Values were not subjected to a correction for multiple hypothesis testing. In order to compare FC obtained by RNA-seq with those from RT-PCR, the FC of RT-PCR was averaged over the “case” samples (patients in the FC of patients vs. controls; metastatic patients in the case of FC of metastasis vs. primary tumor; KRAS+ for FC of KRAS+ vs. KRAS–). A correlation between RNA-seq and qPCR data was calculated by the FC from RNA-seq, which is in the averaged form produced by CLC GW, and values were cross-plotted against the average FC from RT-PCR. A 45° line was added to the plot, to facilitate comparison of the methods.

3 Results

3.1 Evaluation of RNA-sequencing data

We analyzed the differential gene expression of immunity and inflammation genes (475 genes) in the CRC (n=11) and matching

adjacent (n=11) tissues of 11 patients. The other inclusion criterion to the study was the condition of KRAS gene that is frequently mutated in CRC. From the analyzed samples, primary CRC (prim CRC) without mutation in KRAS was present in two samples (n=2), and primary CRC with a mutation in the KRAS gene was identified in three patients (n=3). Other three samples were from CRC patients with metastases (mtsCRC) in the liver, and wild-type allele of KRAS gene (n=3) and other three mtsCRC had also mutation in KRAS gene (n=3).

After the inter-run normalization and data evaluation in CLC, DEGs were selected based on positive variable importance (Table 2), and a heatmap was generated (Figure 1). Overall, mRNA expression of 478 differentially expressed genes were found after alignment. From the designed set of DEGs, MPL, GDC_Cont, IL3, IL9, IFNW1, IFNA4, IFNA14, IFNA6, IFNA1, and IL25 expression was not found in CRC and in adjacent tissue.

A pathway analysis revealed that the most significantly expressed DEGs are enriched in cytokine (hsa-04061) and chemokine signaling pathways (hsa04062), and in inflammatory diseases (hsa05323 and hsa05417). Some DEGs are involved in pathways directly associated with diseases of the digestive tract (hsa04672 and hsa05120) or identified in cancer (hsa04668, hsa05200, and hsa05203). Moreover, some DEGs are enriched in infectious viral and bacterial diseases (hsa05163, hsa05134, hsa05171, hsa05131, hsa05133, hsa05135, hsa05132, hsa05130, hsa05162, hsa05164, and hsa05167) and in signaling pathways (hsa04620 and hsa04217) that are important in the immune defense against viral, bacterial, fungal pathogens, and parasites (21) and in mediating of inflammation by inducing type 1 interferon (IFN) production (hsa04623) (22).

The signaling pathways are listed in Supplementary Table 1 and displayed in Figure 2.

TABLE 2 Statistical analysis of batch-corrected data and feature selection revealed top 12 significant DEGs.

	Gene	Importance
1	TNFRSF4	0.915
2	IRF7	0.720
3	IL6R	0.678
4	NR3C1	0.501
5	EIF2AK2	0.434
6	MIF	0.264
7	CCL5	0.209
8	TNFSF10	0.123
9	CCL20	0.114
10	CXCL11	0.112
11	RIPK2	0.095
12	CCR4	0.005

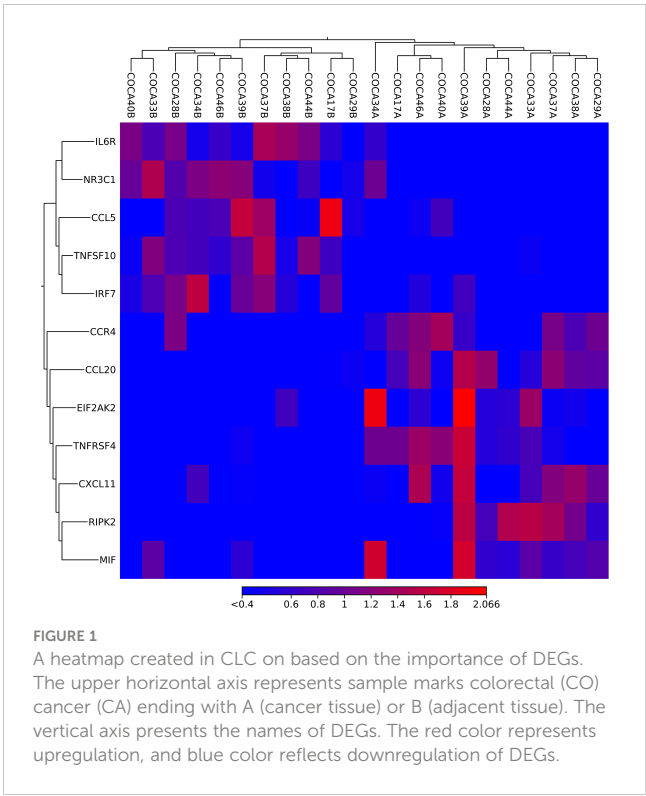
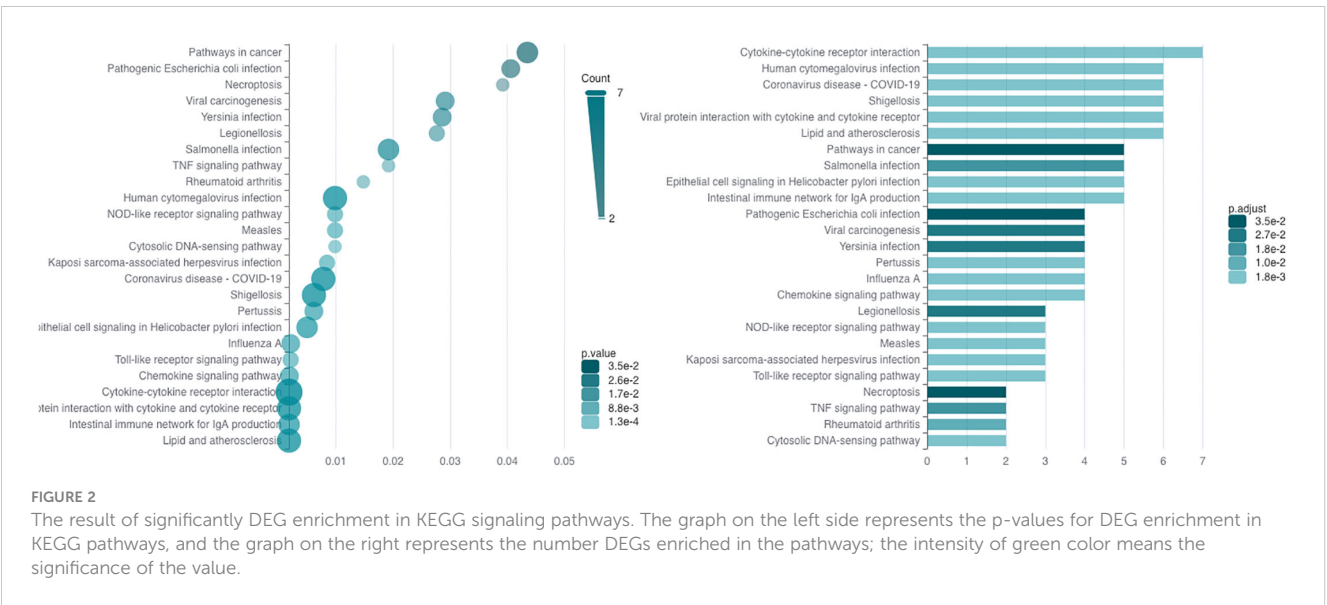


FIGURE 1 A heatmap created in CLC on based on the importance of DEGs. The upper horizontal axis represents sample marks colorectal (CO) cancer (CA) ending with A (cancer tissue) or B (adjacent tissue). The vertical axis presents the names of DEGs. The red color represents upregulation, and blue color reflects downregulation of DEGs.

3.2 Validation of DEGs on a larger cohort of CRC tissues

From the prioritized and other genes, six DEGs (BLNK, CCR4, ILR6, MICB, RIPK2, and TNFSF10) and two housekeeping genes (ACTB and GAPDH) were chosen to validate study results on a larger cohort of 162 tissue samples, including 91 CRC tissues and 71 adjacent tissues. Statistical analysis was performed to determine the fold changes between groups, as cancer/control tissues, metastatic and primary tissues, and KRAS-positive/KRAS-negative tissue. The corresponding p-values are presented in Table 3.

In comparison between cancer and adjacent control tissue, all the selected DEGs have been significantly changed in cancer CRC tissues, in exception to CCR4. The remaining DEGs, such as BLNK, IL6R, MICB, and TNFSF10, were identified with significantly decreased expression, and RIPK2 was identified with significantly increased expression (Figure 3). In the group of metastatic vs. primary CRC tissue, no significant differences in selected DEGs were found. BLNK expression was equal in metastatic cancer tissues when compared to the control tissue. DEG expression of CCR4, ILR6, and RIPK2 was increased, and the expression of MICB and TNFSF10 was decreased in metastatic tissues (Figure 4). KRAS-positive tissues revealed weakly significant upregulation of BLNK and downregulation of MICB and TNFSF10 expression when compared to KRAS-negative cancer tissue. DEG expression in KRAS-positive tissues was equal for ILR6 and decreased for CCR4 and RIKP compared to KRAS-negative FCs (Figure 5). The complete results of bioinformatical analysis of qPCR fold changes are added in Supplementary Data Sheet 1.



3.3 A correlation between qPCR and RNA-sequencing data

The next analysis was performed to see a correlation between two widely used methods for measuring RNA expression. We have calculated the relative gene expression (fold change, FC) for selected genes analyzed by qPCR. In RNA-seq data, the average FC was calculated from sample-by-sample computed FC for each gene and was correlated to the average FC of qPCR data of selected genes. To quantify potential discrepancies between RNA-seq and qPCR, the calculated gene expression fold changes of 11 RNA-seq samples were compared to the same samples analyzed in qPCR (Figure 6) and to all samples analyzed by qPCR (Figure 7). The main groups in this correlation were cancer vs. control, metastatic vs. primary cancer, and KRAS-positive vs. KRAS-negative tumors. High fold change correlation for IL6R, TNFSF10, and BLNK DEGs was found between the matching samples and all CRC tissue samples in cancer vs. control correlation. The other two groups, namely, primary or metastatic cancer and KRAS positive or negative, have a weak

correlation between RNA-seq and qPCR data for each selected DEG. The complete correlation data are attached in [Supplementary Data Sheet 2](#).

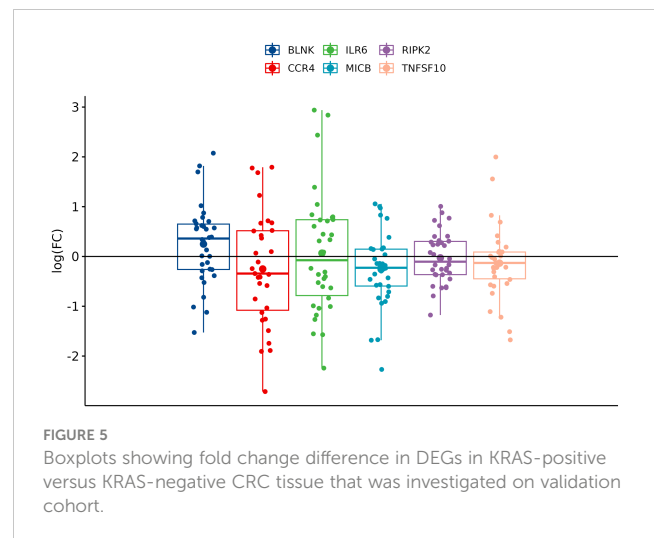
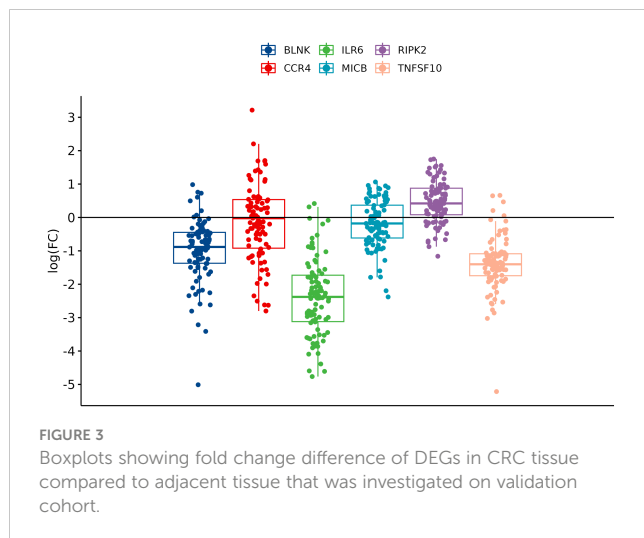
3.4 Analysis of biological pathways by Reactome v83

RNA sequencing data from immune and inflammation panel were processed in inter-run normalization to eliminate technical differences between four runs. The corrected data were analyzed in Reactome data online analysis tool. We used PADOG data analysis algorithm that is Reactome recommended as classical gene set analysis approach. Overall, the DEGs from the RNA-seq study of 11 CRC and 11 adjacent tissues have been enriched in 1,163 biological pathways, and 461 DEGs were identified. We have applied filtering of the pathways by an adjusted p-value ≤ 0.05 that have been considered as significantly regulated by DEGs.

TABLE 3 The qPCR results of selected DEG fold change analysis between the above-mentioned sample groups with corresponding p-values.

gene	Cancer vs. control tissue	Metastatic vs. primary CRC	KRAS-positive vs. KRAS-negative CRC
	p-value	p-value	p-value
BLNK	<0.001	0.517	0.055
CCR4	0.278	0.550	0.281
ILR6	<0.001	0.107	1.000
MICB	0.044	0.977	0.050
RIPK2	<0.001	0.318	0.723
TNFSF10	<0.001	0.318	0.072

CRC, colorectal cancer; KRAS, Kirsten rat sarcoma virus mutation positive; BLNK, B-cell linker; CCR4, C-C Motif Chemokine Receptor 4; ILR6, interleukin 6 (IL6) receptor complex; MICB, MHC class I polypeptide-related sequence B; RIPK2, receptor interacting serine/threonine kinase 2; TNFSF10, TNF superfamily member 10.



3.4.1 A comparison of biological signaling pathways between cancer and control adjacent tissues

This part of the study was performed to see which signaling pathways were altered in CRC tissue against adjacent control tissue. The DEGs from the RNA-seq study of 11 CRC and adjacent tissues have been enriched in 66 biological pathways (Supplementary Image 1). The top 10 up- and downregulated DEGs are presented in Table 4. In summary, the Reactome analysis identified seven DEGs (MIF, RIPK2, IRF7, IL6R, NR3C1, TNFSF10, and CCL5) that were also identified as important in RNA-seq analysis, and three of them were used in RNA-seq validation on a larger cohort.

Subsequently, we have analyzed 66 differentially regulated biological pathways related to signal transduction and overactivity of tyrosine kinases frequently occurring in cancer tissues. Presumably, the main finding in this part of the study is a gene expression–transcription, especially changes in signaling pathways associated with TP53 activity. The changes between cancer and control adjacent tissue were also identified in disease signaling pathways, DNA repair, and programmed cell death, especially in the downregulation of apoptosis and defects in cell cycle signaling

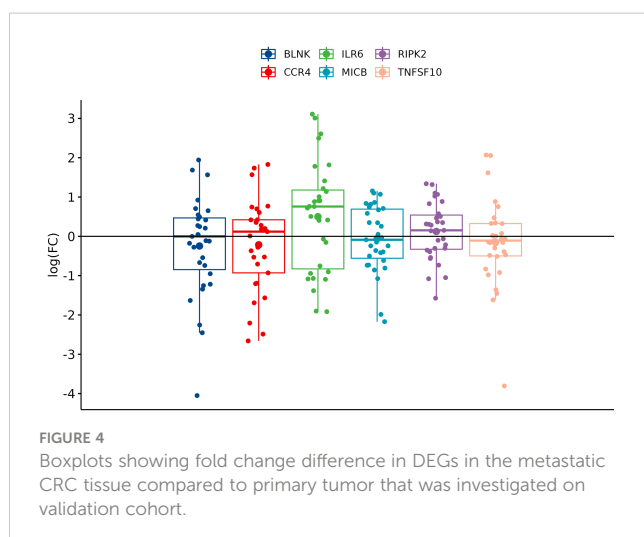
and developmental biology. Immune system signaling was related to changes in pathways activated by inflammatory cytokines and alterations in signaling of cellular responses to stimuli, incoming from external molecular and physical signals. The top differentially up- and downregulated biological pathways and all the significantly deregulated pathways with DEGs involved are listed in Supplementary Table 2.

3.4.2 A comparison of biological signaling pathways between metastatic and primary CRC tumors

“We wanted also to recognize which signaling pathways were altered in the metastatic CRC of patients with metastases in the liver compared to primary tumors. The top differentially up- and downregulated DEGs identified in Reactome v.83 are presented in Table 5.

The multi-omics analysis tool identified 88 differentially regulated biological pathways (Supplementary Image 2) related to signal transduction (ST) through cell surface receptors expressed on various cells in the CNS and the immune system triggering one or many cell responses from a single ligand binding. Disease alterations were associated with infectious diseases or aberrations in the signal transduction and immune system (IS) with its three main nodes. First is the innate IS through pathogen-associated molecular patterns (toll-like receptor (TLR) cascades) and induction of interferon-alpha/beta production. We also found changes in adaptive immune system in dysfunction of antigen processing and presentation and in cytokine signaling by interleukins (IL17). Other differently regulated pathways belong to developmental biology, gene expression, programmed cell death, cell cycle, and metabolism of carbohydrates, such as gluconeogenesis.

The connection between ST and IS signaling pathways was identified through signaling by receptor tyrosine kinases (RTKs), signaling by neurotrophic tropomyosin receptor kinases (NTRKs), and the following nuclear events (kinase and TF activation) and extracellular signal-regulated kinase (ERK)/mitogen activated protein kinase (MAPK) targets, MAPK activation, and IL17 signaling, as in the node for cytokine signaling in IS. Gene



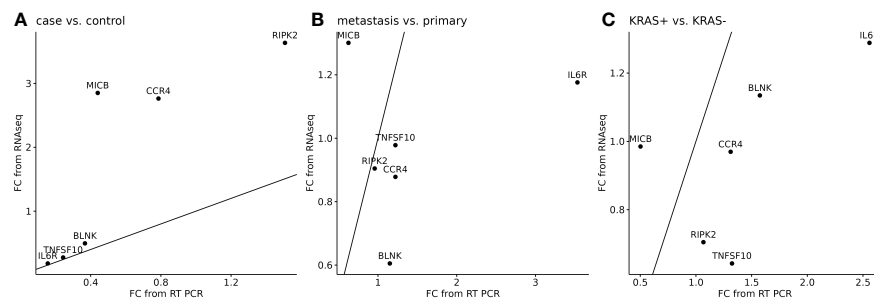


FIGURE 6

Correlation of RNA-seq data (n=11) to the matching samples analyzed by qPCR (n=11) for six selected DEGs. A separate analysis was performed to cancer or control samples (A), metastatic or primary tumor (B), and KRAS-positive or KRAS-negative tumors (C). 45° line was added to the plot to facilitate comparison of the methods.

expression (generic transcription pathway and subsequent transcriptional activity of heterotrimer SMAD2/SMAD3:SMAD4) was also linked to ST by signaling of TGF-beta receptor complex. The top up- or downregulated biological pathways and all the significantly deregulated pathways with DEGs involved are listed in [Supplementary Table 3](#).

3.4.3 A comparison of signaling biological pathways between KRAS pathogenic mutation positive and negative CRC

The next aim of the study was to investigate the influence of KRAS pathogenic mutation on the CRC tissue DEGs ([Table 6](#)) and biological signaling pathways. KRAS-positive CRC tissues compared to negative counterparts were found to be involved in 46 biological signaling pathways ([Supplementary Image 3](#)), and all the significantly deregulated pathways with DEGs involved are listed in [Supplementary Table 4](#).

The main areas of altered pathways belong to disease and alterations in some infectious disease and signal transduction disease pathways. Signal transduction pathways frequently affected in malignant transformation (NOTCH, WNT, and mTORC1-mediated signaling) and signaling by RTKs, NTRKs, and subsequent nuclear events.

Biological pathways of the immune system were affected in all main nodes. The innate IS showed an inappropriate response in recognition of antimicrobial peptides and in ion influx/efflux at the

host-pathogen interface. The adaptive IS was affected by signaling by B-cell receptor (BCR) and following downstream signaling events (as CD28 co-stimulation and CD28-dependent phosphoinositide 3-kinase (PI3K)/protein kinase B (AKT) signaling). The cytokine signaling was affected by DEGs that were found in signaling by CSF1 (M-CSF) in myeloid cells and/or signaling by interleukins, as IL12 signaling. Gene expression (transcription) was altered in RNA polymerase II transcription and generic transcription pathway and subsequent RUNX1-mediated regulation of genes. Furthermore, transcriptional regulation by the TFAP2 family of transcription factors and TP53-regulated transcription of several additional cell death genes.

Considering KRAS pathogenic mutation, signaling pathways influenced by few DEGs were found to belong to programmed cell death, developmental biology, transport of small molecules, and protein localization and mitochondrial protein import. The top up- or downregulated biological pathways are presented in [Supplementary Table 4](#).

4 Discussion

Colorectal cancer development is multistage process that allows and adapts the changes/mutations occurring during cancer progression. The pathogenesis of CRC and its genetic instability is mediated through chromosomal instability, microsatellite instability,

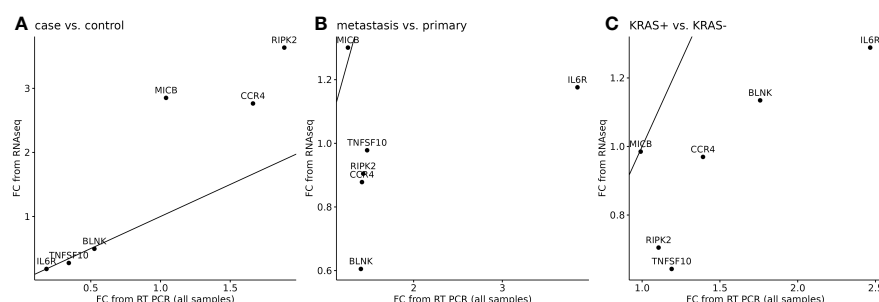


FIGURE 7

Correlation of RNA-seq data to all samples analyzed by qPCR for six selected DEGs. A separate analysis was performed to cancer tissues (n=91) or control adjacent tissues (n=71) samples (A), metastatic or primary tumor (B), and KRAS-positive or KRAS-negative tumors (C). 45° line was added to the plot to facilitate comparison of the methods.

TABLE 4 Top 10 up- and downregulated DEGs identified in CRC compared to the control tissue; logFC and p-values were estimated by the Reactome v83.

Identifier	Regulation	LogFC	p-value	Identifier	Regulation	LogFC	p-value
ELK1	Up	369.89	<0.001	IRF7	Down	-1,359.67	<0.001
CDK2	Up	560.34	<0.001	IL6R	Down	-364.34	0.003
MYC	Up	2001.07	<0.001	IFIT2	Down	-412.10	0.003
GPI	Up	6,285.67	<0.001	NFKBIA	Down	-13,101.91	0.004
MIF	Up	34,409.42	<0.001	NR3C1	Down	-842.99	0.006
RIPK2	Up	597.92	<0.001	TP53INP1	Down	-1,178.21	0.006
CD47	Up	2,780.45	<0.001	TNFSF10	Down	-2,888.22	0.008
IRF2BP1	Up	142.78	<0.001	CCL5	Down	-1,535.82	0.012
CXCL3	Up	1,612.90	<0.001	ACKR2	Down	-115.06	0.014
PRKRA	Up	1,115.71	<0.001	MAF	Down	-1,606.18	0.015

and epigenetic CIMP pathways (5). According to the CIMP state, three subclasses of genetic and epigenetic profiles have been identified, namely, CIMP-high (intense methylation of multiple genes, MSI, and BRAF mutations), CIMP-low (methylation of a limited group of genes and KRAS mutation), and CIMP-negative tissues (characterized with rare methylation and p53 mutation) (23). The regulation of several biological processes is disturbed, including cell proliferation, differentiation, angiogenesis, apoptosis and survival, and the responsible signaling pathways, such as EGFR/MAPK, Notch, PI3K, TGF- β , and Wnt signaling pathways, also do not exhibit their usual functions. Therefore, the exact identification of causal mutations and signaling pathways may trigger a novel preventive and therapeutic progress against CRC.

4.1 RNA-seq data evaluation and a brief analysis of signaling pathways responsible for CRC

The RNA-seq data were normalized, and 12 significant DEGs were selected as the most important in the cancer tissue when

compared to the adjacent tissue. These DEGs are participating in cytokine–cytokine receptor interaction pathway that is a crucial aspect of inflammation and tumor immunology for CRC (24). Cytokines are released as a response to infection, inflammation, and/or immunity, as was demonstrated in consequential pathways (hsa-04061 and hsa04062). The intestinal epithelial cells represent an important defense line in protection against pathogens (25). The changes in intestinal immune network for IgA production (hsa04672) may influence the interaction of epithelial cells to pathogenic bacteria as *Helicobacter pylori* (hsa05120) and many others (hsa05163, hsa05134, hsa05171, hsa05131, hsa05133, hsa05135, hsa05132, and hsa05130), as was confirmed by our study.

The intestinal epithelial cells recognize the colonizing microbial community by TLRs that are very important in early innate immune defense mechanisms and trigger inflammatory pathways through intracellular signaling cascades, leading to the induction of genes encoding cytokines and chemokines, involved in antimicrobial host defense (26). Analogously, this RNA-seq DEGs analysis revealed that the significantly regulated DEGs were involved in TLR signaling pathway (hsa04620). The innate

TABLE 5 Top 10 up- and downregulated DEGs identified in mCRC compared to prim CRC; LogFC and corresponding p-values were estimated by the Reactome v83.

Identifier	Regulation	LogFC	p-value	Identifier	Regulation	LogFC	p-value
CEBPB	Up	1398.57	0.007	TAP1	Down	-2,113.43	<0.001
RORA	Up	528.98	0.007	OAS1	Down	-1,425.43	<0.001
IFI30	Up	5470.27	0.009	TAPBP	Down	-1,487.27	<0.001
NFKBIA	Up	7301.55	0.013	ELK1	Down	-324.75	<0.001
LY96	Up	516.91	0.016	FADD	Down	-454.94	<0.001
MICA	Up	405.78	0.020	CCRL2	Down	-324.96	<0.001
MYD88	Up	771.50	0.027	SIGIRR	Down	-718.52	<0.001
C5	Up	380.02	0.027	PPARG	Down	-2,518.31	<0.001
HLA-DPA1	Up	5721.58	0.037	IRF2BP1	Down	-164.00	<0.001
IFITM3	Up	21922.69	0.037	PPIL2	Down	-672.46	<0.001

TABLE 6 Top 10 up- and downregulated DEGs identified in KRAS-positive compared to KRAS-negative CRC; logFC, and corresponding p-values were estimated by the Reactome v83.

Identifier	Regulation	LogFC	p-value	Identifier	Regulation	LogFC	p-value
NR4A3	Up	116.51	0.012	IL12A	Down	-26.90	0.009
HAVCR2	Up	44.47	0.014	NMI	Down	-666.91	0.043
EGR2	Up	68.78	0.014	HSPD1	Down	-49,615.67	0.057
NFATC1	Up	89.99	0.015	GZMB	Down	-1,050.27	0.087
FOSL1	Up	419.03	0.016	MIF	Down	-13,114.32	0.099
SLC11A1	Up	47.46	0.019	EGF	Down	-31.18	0.133
IL10	Up	34.80	0.024	XCL1	Down	-23.56	0.158
ITGB2	Up	191.13	0.025	HLA-G	Down	-587.70	0.183
CLEC7A	Up	94.42	0.027	AIMP1	Down	-731.05	0.196
C5AR1	Up	132.45	0.028	CD40	Down	-336.32	0.201

immune response to viral infections is mediated by the cytosolic DNA-sensing pathway (hsa04623) as is known as cGAS-STING pathway, and amendments in the pathway cascade drive inflammation-driven tumor growth (27) and participate at defense against viral infections (hsa05162, hsa05164, and hsa05167) and at virus-induced carcinogenesis (hsa05203). The changes in cytosolic DNA-sensing pathway (hsa04623) catalyzes the synthesis of cyclic dinucleotide cGAMP, which activates stimulator of IFN genes (STING) and mediates inflammation by inducing IFN1 production (IFN1) (22) and allowing innate immune response to infections, inflammation, and cancer (28). The cGAS and STING deficiency leads to eliminated or decreased level of IFN1 response to extrinsic cytosolic DNA, which may contribute to non-inflamed cancer microenvironment (28). Several colorectal human cell lines derived from adenocarcinoma have described defective or low STING pathway activity and poor stage of CRC (29).

Necroptosis (hsa04217) is another form of programmed cell death. Its mechanism is conformable to apoptosis and morphology analogous to necrosis. Necroptosis-related genes are strongly associated with TME of CRC, and patients carrying these gene changes may benefit from immunotherapy and have better prognosis of the disease (30).

The important parts of cancer research are TNF signaling pathway (hsa04668) and pathways in cancer (hsa05200). The TNF- α is a proinflammatory cytokine that often presents in the TME and is associated with chronic inflammation. Moreover, the proinflammatory NF- κ B signaling is activated by the canonical pathway triggered by TNF- α , which results in the activation of p65 that regulates inflammatory responses (31). TNF- α promotes tumor angiogenesis and accelerates tumor metastasis, but the molecular mechanisms remain unclear (32). The second signaling pathway is associated with the MAPK and p53 signaling pathways (33).

We also identified lipid and atherosclerosis (hsa05417) pathway, in which lipids are energy reservoirs that can control homeostasis, transcriptional, and enzymatic networks, and

inflammatory response and reprogramming of the lipid metabolism are two hallmarks of cancer (34). Atherosclerosis is similar to solid tumors because of the content of macrophages that participate in acidosis, anaerobic metabolism, and inflammatory process (35). The pathogenesis of rheumatoid arthritis (hsa05323) is also linked with inflammation and presence of survivin that inhibits apoptosis and contributes to persistence of autoreactive T cells and the tumor-like phenotype of fibroblast-like synoviocytes. The overexpression of survivin also affect signaling pathways, such as STAT3 and PIK3/Akt, and is involved in the severity of rheumatoid arthritis (36).

4.2 RNA-seq data verification on larger cohort of CRC patients

Six genes were selected for validation of RNA-seq data, such as IL6R, BLNK, TNFSF10, CCR4, MICB, and RIPK2. There was a correlation between RNA-seq and validation cohort for the IL6R, BLNK, and TNFSF10 DEGs when cancer and control tissues were compared. These three DEGs were also identified as top 12 DEGs. Other three DEGs had a weak correlation between the tested methods. All six DEGs had also a weak correlation between the metastatic and primary tissues and between KRAS-positive versus KRAS-negative cancers.

This study analyzed IL6R expression on a large cohort of CRC patients; we found a significantly lower expression in the cancer group and a non-significantly higher expression in the metastatic group compared to primary cancer and a slightly lower expression in the KRAS-positive group. IL6R is expressed only in cells as hepatocytes and certain leukocytes, whereas its ligand IL6 is expressed in a wide variety of cell types. IL6R transduction pathway is mediated through activated STAT3, MAPK, and PI3K activation (37). Authors found that colorectal cancer cell lines express IL6 and IL6R and co-receptor gp130 (38), and higher levels of IL6R mRNA expression were found in HER2-positive breast cancer lines and in immortalized cells derived from

nasopharyngeal epithelium with activated STAT3 (39). This study revealed significantly lower IL6R and lower STAT3 expression.

B-cell linker protein (BLNK) is crucial for B-cell receptor signaling pathway and mediates B-cell apoptosis (40). BLNK overexpression was found in approximately 70% of CRC tissues where its oncogenic activity via RAS/ERK pathway has been reported (41). BLNK overexpression was an independent risk factor for CRC recurrence (42). A reduced expression of BLNK was found in increased migration and invasion of CRC cells (43). BLNK expression on a large cohort of CRC patients was found to be significantly lower in the cancer group and equal expression in mCRC when compared to primary cancer and a higher expression in the KRAS-positive group.

Tumor necrosis factor (TNF) superfamily member 10 (TNFSF10), known as tumor necrosis factor (TNF)-related apoptosis-inducing ligand (TRAIL) or ApoEL, is able to induce cell apoptosis in various types of tumor cells (44) by receptor oligomerization and recruitment of the FADD and caspase 8 and 10 (45). TNFSF10 expression on a large cohort of CRC patients was found to be significantly lower in the cancer group compared to the control group and non-significantly lower in metastatic and KRAS-positive group. Downregulated expression of TNFSF10 was found also in transcriptomic study of CRC compared to normal data available in TCGA database (46).

MICB or MHC class I chain-related B molecule is one of the ligands of NKG2D receptor that exists in NK cells and CD8⁺ T cells. MICB is expressed by the intestinal epithelium or epithelial tumors, and their role is in the immunosurveillance and mediates antitumor response (47). MICB expression on a large cohort of CRC patients uncovered its significantly lower expression in the cancer group compared to the control group, non-significantly lower expression in the metastatic group, and weakly significantly lower expression in the KRAS-positive group. A study of MICB expression had shown a significantly high expression that was associated with tumor size and better overall survival (48). Another study revealed worse survival of patients with CRC with downregulated MICB because the tumor evades recognition by the immune system (49).

Another molecule used for the validation of RNA-seq data was the receptor-interacting serine/threonine-protein kinase-2 (RIPK2). RIPK2 was found to be involved in solid tumors and have a role in different pathways of immune and inflammatory responses (50). The RIPK2 expression on a large cohort of CRC patients was found significantly increased in the cancer group, non-significantly higher expression in the metastatic group, and non-significantly lower expression in the KRAS-positive group. High expression of RIPK2 was associated with high expression of VEGFA and increased mortality and has a potential in targeted therapy (51).

The role of CC chemokine receptor (CCR4) was identified in normal and tumor immunity and belongs to the G-protein-coupled receptor family. The binding of chemokine ligands trigger its function in human autoimmune diseases, such as atopic dermatitis, asthma, and cutaneous T-cell lymphoma (CTCL), and is expected to be a novel therapeutic target for cancer immunotherapy (52). We identified a non-significant expression of CCR4 in each group; CCR4 expression was equal in the cancer

group when compared to the control group, slightly higher in the metastatic group, and lower in the KRAS-positive group.

4.3 Reactome data analysis identification of DEGs in CRC compared to adjacent tissue

We further aimed to investigate which changes in signaling pathways could be responsible for the development of CRC. Here, we discuss an altered gene expression and signaling pathways that were significantly changed in the CRC tissue compared to the control adjacent tissue. Using *in silico* Reactome analysis, we identified the significantly upregulated genes, namely, CDK2, MYC, GPI, CD44, NOD1, UBE2N, PRKRA, and IRAK2, and the significantly downregulated gene, IRF7 gene (Supplementary Table 2).

Cyclin-dependent kinase 2 (CDK2) is a DNA damage signaling kinase, which phosphorylates proteins in many cellular processes and is hyperactivated in most cancers (53). UBE2N interacts with BRCA1, and its expression serves as a potential biomarker of response to poly(ADP-ribose) polymerase (PARP) inhibitors and other DNA-repair-targeted therapies in breast cancer (54).

Cell cycle signaling pathways importantly participate in cell cycle division. Most cancer cells have defects in G1 and G2 checkpoints, and DNA damage triggers the ATM/CHK2/p53 pathway. The weakening of G2 checkpoint leads to chromosomal instability, and dysregulated cell-cycle-related genes can be a reason for uncontrolled cell growth and proliferation that are a signs of cancer cells (55).

The *in silico* analysis of cancer cells found changes in biogenesis in miRNA and siRNA, and mainly TP53 gene expression was disturbed. The TP53 is a key tumor suppressor that regulates different cellular responses to protect against cancer development. Colorectal cancer is reported with 43% of mutations in TP53 gene, and mostly missense mutations impair wild-type p53 function (loss-of-function). Invasive and metastatic cancers may provide gain-of-function activities with more aggressive phenotype acquired by clonal evolution of cancer stem cell (56). Patients with breast cancer carrying the TP53 mutation were diagnosed with a higher glucose-6-phosphate isomerase (GPI) gene expression that indicates higher level of glycolysis in tumor cells and correlates with the degree of tumor malignancy (57). Furthermore, MYC overexpression was observed in up to 70%–80% of CRC (58). Numerous studies sustained a chemoresistance in tumor cells expressing high levels of MYC and activation of WNT signaling pathway. Gaggianesi et al. demonstrated that dual indirect targeting of CD44 and MYC in CRC stem cells, using PI3K and CDK inhibitors, reduces the survival and clonogenic activity of cancer stem cells, regardless of the mutational background (59). Similarly, MIF overexpression leads to excessive signaling through CD74 surface receptor and formation of complex with CD44 that initiates the ERK/MAPK signaling pathway. Chemokine-like function of MIF consists of the recruitment of immune cells and mediation of acute and chronic inflammatory diseases, and tumor progression and development. CRC patients with MIF

overexpression in lymph nodes had a shorter survival time after surgery (60).

PRKRA is a core component in the miRNA/siRNA biogenesis and is known as a cellular protein activator of PKR kinase (also known as EIF2AK2) in a dsRNA-independent manner in response to cellular stress. PKR kinase can then induce the expression of type I interferons (IFNs). Increased expression of PRKRA can be associated with worse survival of colorectal cancer patients (61).

Pathways associated with signal transduction were linked to the activation of TNFSF10 (alias TRAIL) and initiation of apoptosis through FADD domain complex and effector caspases (casp-3, casp-6, and casp-7) (62). TNFSF10 can also induce non-apoptotic signaling through the activation of proinflammatory pathways, including NF- κ B, PI3K/Akt, and MAPK such as JNK, ERK, and p38 (63).

Intestinal epithelial cells and immune cells recognize exogenous and endogenous stimuli through TLRs. IRAK2 is a key regulator of interleukin-1 receptor (IL1R)/TLR-mediated inflammation and is involved in NF- κ B and MAPK signaling pathways (64). In addition to TLRs, other pattern recognition molecules (PRRs) (65) are often present in cells. This study revealed overexpressed DEGs such as UBE2N, RIPK2, NOD1, and IRAK2, which points to the altered expression of the other two PRRs. NOD1 and NOD2 are bacterial sensors that trigger proinflammatory signaling mediated by receptor-interacting protein kinase 2 (RIPK2). UBE2N is a ubiquitin enzyme that eases the production of non-embedded polyubiquitin chains and serves as an activator for RIG-I on virus infection (66). The recognition of pathogenic microorganisms by PRRs induces the activation and translocation of IRF7 (interferon regulatory factor 7) to nuclear, leading to IFN-I secretion. IRF7 has opposite functions in carcinogenesis (67). This study indicates a significant downregulation of IRF7 in CRC. Otherwise, the downregulation of IRF7 expression promoted polarization of tumor-associated macrophages (TAMs), which produce anti-inflammatory factors to enhance breast cancer development by promoting immune escape, proliferation, and migration of cancer cells (68).

4.4 Reactome data analysis identification of DEGs in metastatic CRC tissue compared to primary tumor

Another part of the study was a comparison of signaling pathways in metastatic versus primary CRC. The Reactome analysis uncovered the following pathways that we discussed in a broader context of the disease. Nevertheless, the study results did not influence the treatment of patients with mCRC; patients were administrated with standard therapy prescribed by the experienced oncologist.

Using *in silico* Reactome analysis, we characterized DEGs in tumors that were able to invade the liver and form a metastasis. We found a significant downregulation of genes, such as CASP8, FADD, ELK1, ERBB2, IRAK1, SIGIRR, TAP1, TAP2, TAPBP, HLA-E, STAT3, MYC, IFNAR1, PPARG, PPIL2, and IL17RE (Supplementary Table 3).

The downregulation of FADD and CASP8 might bring the evidence of blocked apoptosis in mCRC in programmed cell death signaling that was discussed above.

Signal transduction (ST) pathways were altered in ERK/MAPK pathways where we found significantly decreased expression of ELK1 in mCRC. ELK1 is usually upregulated in cancer. Nevertheless, ELK1 expression almost disappeared in the middle stage of G1 phase and at the end of S phase (69). The analysis also uncovered a reduced expression of ERBB2 (HER2) gene that plays an essential role in the regulation of cellular proliferation, differentiation, and migration via the same signaling pathways, such as ERK/MAPK and PI3K/Akt/mTOR (70). It is hypothesized that HER2 expression is abrogated during EMT by chromatin-based epigenetic silencing of ERBB2 gene. Subsequently, tumors become resistant to HER2-targeted therapies (71). In line with previous findings, a significant downregulation of STAT3 was found in oncogenic MAPK signaling represented by RAS/RAF/MAPK cascade that is important in regulating cellular proliferation, differentiation, and survival by MAP2K mutants, such as BRAF, RAF1, and RAS gene alterations (72). Other pathways were likely responsible for the metastatic progression of CRC in the cohort, so STAT3 molecules or inhibitors would not be applicable here. Controversially, a reduced MYC expression has been found in mCRC in comparison to primary CRC. Reduced MYC expression was found in cancer cells localized in the area distant from blood vessels where TME contains limited levels of oxygen and glucose. This might be a strategy of cancer cells to survive under conditions of limited energy sources (73).

The most signaling pathways detected in this part of *in silico* analysis were associated with immune system (IS). Innate IS response in CRC tissues have increased expression of TLRs because of the presence of chronic inflammation, microbial pathogens-induced changes in metabolism, TME, and genotoxic response (74). Commensal microflora in intestinal homeostasis can modulate TLR signaling through single immunoglobulin IL-1 receptor-related molecule (SIGIRR) that is a negative regulator for TLR-IL1R signaling. Hyperactivation of TLR-IL1R-mediated Akt-mTOR signaling in SIGIRR-deficient tumors leads to cell cycle progression, loss of heterozygosity of APC, and tumor initiation (75).

TLR-IL1R signaling is also important in the coordination of the early immune response to pathogens that is mediated by the protein myeloid differentiation primary response protein 88 (MyD88) where IL-1R-associated kinase (IRAK) proteins are employed (76). Catalytically activated IRAK1 and IRAK4 with the participation of MyD88 interact with TNF-receptor-associated factor 6 (TRAF6) and drive NF- κ B and MAPK pathways that results in the production of proinflammatory cytokines (77). IRAK1 has been upregulated in many cancers and is considered as one possible target in cancer therapy. However, low expression levels of IRAK1 may cause the failure of targeted cancer therapy.

We also identified changes in cytokine signaling, especially the downregulation of interleukin 17 (IL-17) receptor E (IL17RE). IL-17RE is expressed in T_H17 cells and in the epithelial cells themselves, and its ligand is interleukin 17C (IL-17C). IL17C secretion maintains an autocrine loop in the epithelium, thereby

enhancing innate immune barriers (78). A reduction or attenuation in IL-17R induces MAPK signaling pathway in the downregulation of ERK2 expression and downstream targets (ELK1, ETS2, RSK, MNK, and PLA2). Tumors with IL-17R deletion express molecular markers for tumor growth, invasion, and metastasis (79), which was also confirmed by our study.

We also identified a decreased expression of IFNAR1 chain of the IFN1 receptor that is activated by JAK-STAT and other signaling pathways. Downregulation of IFNAR1 is often present in the malignant cells and in TME, where IFN1 pathway is suppressed (80).

Adaptive immune response was altered in class I MHC-mediated antigen processing and presentation. We identified downregulation of MHC class I, which is frequent in tumors and might be an intrinsic mechanism of acquired resistance to immunotherapy (81, 82). Subsequently, the downregulation of transporters associated with antigen processing (TAP1/TAP2 and TAPBP) is also logical, which was identified in the study. We also identified a downregulation of HLA-E expression. Studies showed that the loss of MHC class I expression is accompanied with the loss of HLA-E or HLA-G expression, mean significantly better overall, and disease-free survival for the patients (83).

Signaling pathways in gene expression were associated with transcription of several transcription factors, such as peroxisome proliferator-activated receptors (PPARs) that act as antagonists in transcription of immunity and inflammation factors (84). We identified significantly low expression of PPARG in mCRC. Authors found a lack of PPARG expression in 30% of primary CRCs that strongly correlates with promoter methylation and was found in patients with poor prognosis (85). In contrast, other authors found no difference in disease progression or survival; therefore, PPARG is not an active agent for the treatment of metastatic colorectal cancer (84). We also identified a peptidylprolyl isomerase (cyclophilin)-like 2 (PPIL2) molecule that was downregulated in mRNA biogenesis and metabolism. PPIL2 is an ubiquitin ligase and probably participates in breast cancer metastasis in animal models because its downregulation led to increased migration of human breast cancer cells (86).

The interesting revelation in mCRC was the connection between ST and IS signaling pathways mediated by receptor tyrosine kinases signaling, signaling by NTRKs and following nuclear events (kinase and TF activation) and ERK/MAPK targets, MAP kinase activation, and IL17 signaling, as in the node for cytokine signaling in IS. This molecular interaction seems to be a hot candidate in the altered molecular profile of basic tumor, which derived distant metastases when compared to primary CRC. Changes in signaling pathways can allow individual cells to detach from the original tumor and spread to other parts of the body.

4.5 Reactome data analysis identification of DEGs in KRAS-positive versus KRAS-negative CRC tissue

The data analysis of KRAS-positive versus KRAS-negative CRC tissues has not revealed any significantly up- or downregulated

DEGs. Therefore, we present the top up- and downregulated pathways according to FDR below 0.05 (Supplementary Table 4).

KRAS gene is highly mutated in up to 50% of colorectal cancer. The activation of KRAS proteins can be performed by growth factors, receptor tyrosine kinase (RTK), chemokines, and Ca^{2+} ions. The downstream signaling pathways of activated KRAS protein are the RAF-MEK-ERK (MAPK) signaling pathway, PI3K-Akt-mTOR signaling pathway, and other signaling pathways (87). KRAS mutation has a significant influence on the progression and treatment of colorectal cancer (88). KRAS-positive tumors had alterations in two main categories of the disease signaling pathways. Upregulation in pathways is an anti-inflammatory response favoring *Leishmania* parasite infection and mitigation of host antiviral defense response. These pathways can point that T-cell response and antigen presentation were reduced in KRAS-positive tissues, as described by Liu et al. (89).

The alterations in signal transduction pathways confirmed the role of KRAS in signaling by Notch, Wnt, mTOR, and NTRKs that function as a network in CRC stem cells (90). The major downstream effectors in KRAS oncogenic signaling is the TGF- β pathway that has been identified as the top pathway enriched in invasive and metastatic tumors (91).

The Reactome analysis identified a wide involvement of the immune system in cancer tissues carrying KRAS mutations. TME is present with inflammation, high levels of inflammatory cytokines and chemokines, and is infiltrated with multiple immune cells (92). We identified upregulation in C-type lectin receptor (CLR) signaling, as dectin-1 (CLEC7A) that increases production of cytokines and chemokines through activation of NF- κ B via caspases and MAPK, and nuclear factor of activated T cells (NFAT) pathways and induces reactive oxygen species (ROS) production. The detailed process is described in the review (93). Dectin-1 expression in TAMs promoted their suppression of anti-tumor immunity and toleration of tumor cells in pancreatic ductal adenocarcinoma, making Dectin-1 an interesting target for immunotherapy (94).

Adaptive immunity, with its components B- and T-cells, is involved in pathogen clearing. T-cells are activated by antigen recognition by T-cell receptor and by costimulatory molecule, such as CD28 receptors, and may activate PI3K/Akt signaling that promotes cytokine transcription, survival, cell-cycle entry, and growth (95). Signaling by B-cell receptor (BCR) and downstream phosphorylation of ITAMs lead to NFAT activation by calcineurin, IP3, and/or PIP3, and NF- κ B is activated via PKC, Ras is activated via RasGRP, and Akt is activated via PDK1 (95). Cytokine signaling regulates and mediates immunity, inflammation, and hematopoiesis. Reactome analysis also identified changes in signaling by CSF1 (macrophage colony stimulating factor, M-CSF) in myeloid cells and/or signaling by interleukins, such as IL12 signaling. Tumors frequently contain TAMs, which contribute to their development and progression and are useful in antitumoral therapy (96).

DEGs enriched in signaling pathways of gene expression (transcription) revealed RUNX1-mediated regulation of cell differentiation genes, including keratinocytes, myeloid and megakaryotic progenitors, and regulatory T and B lymphocytes (97).

The roles of transcription factors with RUNX in EMT and tumor progression should be more studied (98, 99). Transcriptional regulation by the TFAP2 family of transcription factors was found to be reduced in high-grade colorectal adenocarcinomas (100). Next, we found alterations in TP53 regulation of transcription of cell death genes with uncertain role in apoptosis. Authors found that patients with a high expression of TP53 in KRAS-positive CRC had a poor prognosis of the disease (101). An upregulation in the FOXO-mediated transcription of cell death genes was also identified in the study. FOXO transcription factors participate in cell proliferation and in cell apoptosis including the control of autophagy, metabolism, inflammation, and differentiation (102).

The alterations in transport of small molecules through transporters, such as ATP-powered pumps, ion channels, and transporters, represent the imbalance in the transport of bile salts, organic acids, metal ions, and amine compounds, and subsequent organic cation transport and/or organic cation/anion/zwitterion transport, and metal ion SLC transporters, and protein localization and mitochondrial protein import. It was proven that KRAS mutations in CRC cells correlate with the higher amino acid uptake when compared to KRAS wild-type colorectal cells (103).

5 Conclusion

The understanding of pathology and altered gene expression of CRC would help to identify the key point in treatment selection, monitoring, and prevention of disease relapse. Our data may partially contribute to clarifying the function of the immune system in the TME of CRC.

The difference between CRC and control adjacent tissue identified the main altered pathways in CRC. A significant participation of tumor-associated macrophages was seen in TME. The initiation of the inflammatory response likely plays the role in altered gene expression of TP53, and the accumulation of other gene mutations probably affects subsequent errors in DNA repair and cell cycle pathways. The *in silico* analysis revealed the changes at the level of surface receptors of epithelial or immune cells through which they interact with the colonizing microbiome. The subsequent reaction of macrophages supports the immune response and inflammation; the permanent production of proinflammatory cytokines and chemokines leads to further changes in cellular pathways. The emergence of mutations and the selection of tumor clones with blocked apoptosis allow mistakes in the cell cycle and accumulation of mutations and uncontrolled proliferation of tumor cells.

Analysis of metastatic tumor tissues compared to the primary tumor showed persistent changes in the expression of cellular surface receptors, chronic inflammation, and the presence of proinflammatory cytokines (e.g., IL-17) and changes in downstream signaling pathways (TGF- β and MAPK signaling). The main finding was the influence of the immune system to signal transduction pathways that may be linked to the activity of tumor-associated macrophages. Unfortunately, this study does not analyze secondary metastatic tumors. Therefore, the study does not

reflect the clonal evolution and associated signaling pathways that lead to detachment of specific tumor cell. We indicate only basic molecular pathways that enabled the formation of metastatic cancer stem cells.

KRAS mutations also affected signal transduction pathways, particularly Notch1 and Wnt signaling pathways, the innate and adaptive immune system, and cytokine production. The gene expression level was mainly affected by RUNX1-mediated regulation of cellular differentiation and by regulation of TP53 transcription.

A precise identification of signaling pathways responsible for CRC may help in the selection of personalized pharmacological treatment because many targeted therapies to specific cancers have been developed.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: <https://data.mendeley.com/datasets/xjxkym6vr/1>.

Ethics statement

The studies involving humans were approved by Ethics committee in the Jessenius Faculty of Medicine in Martin, Comenius University in Bratislava, number 1863/2016. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

VH performed signaling pathways analysis and wrote the manuscript; DL performed RNAseq experiments; MG performed bioinformatic analyses and reviewed the manuscript; ZK and VH performed qPCR experiments; PM, JMi, and EK collected patient's samples; ET assisted the experiments; MK and JMa performed histopathological evaluation; LL reviewed and revised the manuscript; ZL designed the study, ensured the financial funding, supervised the data analysis and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from The Ministry of Education, Science, Research and Sport of the Slovak Republic (grant VEGA no. 1/0269/22) and by The Slovak Research and Development Agency (grant APVV no. 16-0066 and no. 21-0448), and by the European cooperation in science and technology no. COST CA17118.

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.

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/fonc.2023.1206482/full#supplementary-material>

References

- 703-Slovakia-fact-sheets.pdf. Available at: <https://gco.iarc.fr/today/data/factsheets/populations/703-Slovakia-fact-sheets.pdf>.
- Holubekova V, Loderer D, Grendar M, Mikolajcik P, Kolkova Z, Turyova E, et al. Colorectal cancer statistics (2023). WCRF International. Available at: <https://www.wcrf.org/cancer-trends/colorectal-cancer-statistics/>.
- Colorectal cancer: epidemiology, risk factors, and protective factors (2023). Available at: <https://www.uptodate.com/contents/colorectal-cancer-epidemiology-risk-factors-and-protective-factors>.
- Puzzono M, Mannucci A, Grannò S, Zuppardo RA, Galli A, Danese S, et al. The role of diet and lifestyle in early-onset colorectal cancer: A systematic review. *Cancers* (2021) 13(23):5933. doi: 10.3390/cancers13235933
- Pino MS, Chung DC. THE CHROMOSOMAL INSTABILITY PATHWAY IN COLON CANCER. *Gastroenterology* (2010) 138(6):2059–72. doi: 10.1053/j.gastro.2009.12.065
- Wang W, Kandimalla R, Huang H, Zhu L, Li Y, Gao F, et al. Molecular subtyping of colorectal cancer: recent progress, new challenges and emerging opportunities. *Semin Cancer Biol* (2019) 55:37–52. doi: 10.1016/j.semcancer.2018.05.002
- Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* (1990) 61(5):759–67. doi: 10.1016/0092-8674(90)90186-I
- Ogino S, Nosho K, Kirkner GJ, Kawasaki T, Meyerhardt JA, Loda M, et al. CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut* (2009) 58(1):90–6. doi: 10.1136/gut.2008.155473
- Vilar E, Gruber SB. Microsatellite instability in colorectal cancer—the stable evidence. *Nat Rev Clin Oncol* (2010) 7(3):153–62. doi: 10.1038/nrclinonc.2009.237
- Chang L, Chang M, Chang HM, Chang F. Microsatellite instability: A predictive biomarker for cancer immunotherapy. *Appl Immunohistochem Mol Morphol* (2018) 26(2):e15–21. doi: 10.1097/PAL.0000000000000575
- Advani SM, Swartz MD, Loree J, Davis JS, Sarsashek AM, Lam M, et al. Epidemiology and molecular-pathologic characteristics of CpG island methylator phenotype (CIMP) in colorectal cancer. *Clin Colorectal Cancer* (2021) 20(2):137–147.e1. doi: 10.1016/j.clcc.2020.09.007
- Fu J, Jin X, Chen W, Chen Z, Wu P, Xiao W, et al. Identification of the molecular characteristics associated with microsatellite status of colorectal cancer patients for the clinical application of immunotherapy. *Front Pharmacol* (2023) 14:1083449. doi: 10.3389/fphar.2023.1083449
- Biller LH, Schrag D. Diagnosis and treatment of metastatic colorectal cancer: A review. *JAMA* (2021) 325(7):669–85. doi: 10.1001/jama.2021.0106
- Li Z, Zhang Y, Fang J, Xu Z, Zhang H, Mao M, et al. NcPath: a novel platform for visualization and enrichment analysis of human non-coding RNA and KEGG signaling pathways. *Bioinformatics* (2023) 39(1):btac812. doi: 10.1093/bioinformatics/btac812
- Griss J, Viteri G, Sidiropoulos K, Nguyen V, Fabregat A, Hermjakob H. ReactomeGSA - efficient multi-omics comparative pathway analysis. *Mol Cell Proteomics MCP* (2020) 19(12):2115–24. doi: 10.1074/mcp.TIR120.002155
- Tarca AL, Draghici S, Bhatti G, Romero R. Down-weighting overlapping genes improves gene set analysis. *BMC Bioinf* (2012) 13:136. doi: 10.1186/1471-2105-13-136
- Fei T, Yu T. scBatch: batch-effect correction of RNA-seq data through sample distance matrix adjustment. *Bioinformatics* (2020) 36(10):3115–23. doi: 10.1093/bioinformatics/btaa097
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* (2010) 33(1):1–22. doi: 10.18637/jss.v033.i01
- Liñares-Blanco J, Pazos A, Fernandez-Lozano C. Machine learning analysis of TCGA cancer data. *PeerJ Comput Sci* (2021) 7:e584. doi: 10.7717/peerj-cs.584
- Holubekova V, Loderer D, Grendar M, Mikolajcik P, Kolkova Z, Turyova E, et al. *a language and environment for statistical computing*. Available at: <https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing>.
- McClure R, Massari P. TLR-dependent human mucosal epithelial cell responses to microbial pathogens. *Front Immunol* (2014) 5:386. doi: 10.3389/fimmu.2014.00386
- Suter MA, Tan NY, Thiam CH, Khatoo M, MacAry PA, Angeli V, et al. cGAS-STING cytosolic DNA sensing pathway is suppressed by JAK2-STAT3 in tumor cells. *Sci Rep* (2021) 11(1):7243. doi: 10.1038/s41598-021-86644-x
- Serra RW, Fang M, Park SM, Hutchinson L, Green MR. A KRAS-directed transcriptional silencing pathway that mediates the CpG island methylator phenotype. *eLife* (2014) 3:e02313. doi: 10.7554/eLife.02313
- Dong C, Wang X, Xu H, Zhan X, Ren H, Liu Z, et al. Identification of a cytokine-cytokine receptor interaction gene signature for predicting clinical outcomes in patients with colorectal cancer (2017).
- Pracht K, Wittner J, Kagerer F, Jäck HM, Schuh W. The intestine: A highly dynamic microenvironment for IgA plasma cells. *Front Immunol* (2023) 14:1114348. doi: 10.3389/fimmu.2023.1114348
- Li TT, Ogino S, Qian ZR. Toll-like receptor signaling in colorectal cancer: Carcinogenesis to cancer therapy. *World J Gastroenterol WJG* (2014) 20(47):17699–708. doi: 10.3748/wjg.v20.i47.17699
- Kwon J, Bakhoun SF. The cytosolic DNA-sensing cGAS-STING pathway in cancer. *Cancer Discov* (2020) 10(1):26–39. doi: 10.1158/2159-8290.CD-19-0761
- Jiang M, Chen P, Wang L, Li W, Chen B, Liu Y, et al. cGAS-STING, an important pathway in cancer immunotherapy. *J Hematol Oncol Hematol Oncol* (2020) 13(1):81. doi: 10.1186/s13045-020-00916-z
- Xia T, Konno H, Ahn J, Barber GN. Dereglulation of STING signaling in colorectal carcinoma constrains DNA damage responses and correlates with tumorigenesis. *Clin Rep* (2016) 14(2):282–97. doi: 10.1016/j.celrep.2015.12.029
- Peng X, Xu Z, Guo Y, Zhu Y. Necroptosis-related genes associated with immune activity and prognosis of colorectal cancer. *Front Genet* (2022). 13:909245. doi: 10.3389/fgene.2022.909245
- Kobelt D, Zhang C, Clayton-Lucey IA, Glauben R, Voss C, Siegmund B, et al. Pro-inflammatory TNF- α and IFN- γ Promote tumor growth and metastasis via induction of MACC1. *Front Immunol* (2020) 11:980. doi: 10.3389/fimmu.2020.00980
- Zhao P, Zhang Z. TNF- α promotes colon cancer cell migration and invasion by upregulating TROP-2. *Oncol Lett* (2018) 15(3):3820–7. doi: 10.3892/ol.2018.7735
- Zhu H, Ji Y, Li W, Wu M. Identification of key pathways and genes in colorectal cancer to predict the prognosis based on mRNA interaction network. *Oncol Lett* (2019) 18(4):3778–86. doi: 10.3892/ol.2019.10698
- Niculae AM, Dobre M, Herlea V, Vasilescu F, Ceafalan LC, Trandafir B, et al. Lipid handling protein gene expression in colorectal cancer: CD36 and targeting miRNAs. *Life Basel Switz* (2022) 12(12):2127. doi: 10.3390/life12122127
- Lee-Rueckert M, Lappalainen J, Kovanen PT, Escola-Gil JC. Lipid-laden macrophages and inflammation in atherosclerosis and cancer: an integrative view. *Front Cardiovasc Med* (2022) 9:777822. doi: 10.3389/fcvm.2022.777822
- Morand S, Staats H, Creeden JF, Iqbal A, Kahaleh B, Stanbery L, et al. Molecular mechanisms underlying rheumatoid arthritis and cancer development and treatment. *Future Oncol* (2020) 16(9):483–95. doi: 10.2217/fon-2019-0722
- Scheller J, Garbers C, Rose-John S. Interleukin-6: From basic biology to selective blockade of pro-inflammatory activities. *Semin Immunol* (2014) 26(1):2–12. doi: 10.1016/j.smim.2013.11.002
- Holmer R, Wätzig GH, Tiwari S, Rose-John S, Kalthoff H. Interleukin-6 trans-signaling increases the expression of carcinoembryonic antigen-related cell adhesion

molecules 5 and 6 in colorectal cancer cells. *BMC Cancer*. (2015) 15(1):975. doi: 10.1186/s12885-015-1950-1

39. Zhang G, Tsang CM, Deng W, Yip YL, Lui VWY, Wong SCC, et al. Enhanced IL-6/IL-6R signaling promotes growth and Malignant properties in EBV-infected premalignant and cancerous nasopharyngeal epithelial cells. *PLoS One* (2013) 8(5): e62284. doi: 10.1371/journal.pone.0062284
40. Imoto N, Hayakawa F, Kurahashi S, Morishita T, Kojima Y, Yasuda T, et al. B cell linker protein (BLNK) is a selective target of repression by PAX5-PML protein in the differentiation block that leads to the development of acute lymphoblastic leukemia. *J Biol Chem* (2016) 291(9):4723–31. doi: 10.1074/jbc.M115.637835
41. Grassilli E, Pisano F, Cialdella A, Bonomo S, Missaglia C, Cerrito MG, et al. A novel oncogenic BTK isoform is overexpressed in colon cancers and required for RAS-mediated transformation. *Oncogene* (2016) 35(33):4368–78. doi: 10.1038/onc.2015.504
42. Lee JH, Lee JH, Ahn BK, Paik SS, Lee KH. Prognostic value of B-cell linker protein in colorectal cancer. *Pathol - Res Pract* (2020) 216(3):152821. doi: 10.1016/j.prp.2020.152821
43. Xu J, Meng Q, Li X, Yang H, Xu J, Gao N, et al. Long Noncoding RNA MIR17HG Promotes Colorectal Cancer Progression via miR-17-5p. *Cancer Res* (2019) 79(19):4882–95. doi: 10.1158/0008-5472.CAN-18-3880
44. Lin YC, Richburg JH. Characterization of the role of tumor necrosis factor apoptosis inducing ligand (TRAIL) in spermatogenesis through the evaluation of trail gene-deficient mice. *PLoS One* (2014) 9(4):e93926. doi: 10.1371/journal.pone.0093926
45. Keding V, Muller S, Gronemeyer H. Targeted expression of tumor necrosis factor-related apoptosis-inducing ligand TRAIL in skin protects mice against chemical carcinogenesis. *Mol Cancer*. (2011) 10(1):34. doi: 10.1186/1476-4598-10-34
46. Zhao H, Huang C, Luo Y, Yao X, Hu Y, Wang M, et al. A correlation study of prognostic risk prediction for colorectal cancer based on autophagy signature genes. *Front Oncol* (2021) 11:595099. doi: 10.3389/fonc.2021.595099
47. Diefenbach A, Hsia JK, Hsiung MYB, Raulet DH. A novel ligand for the NKG2D receptor activates NK cells and macrophages and induces tumor immunity. *Eur J Immunol* (2003) 33(2):381–91. doi: 10.1002/immu.200310012
48. Feng Q, Yu S, Mao Y, Ji M, Wei Y, He G, et al. High MICB expression as a biomarker for good prognosis of colorectal cancer. *J Cancer Res Clin Oncol* (2020) 146(6):1405–13. doi: 10.1007/s00432-020-03159-0
49. Ferrari de Andrade L, Tay RE, Pan D, Luoma AM, Ito Y, Badrinath S, et al. Antibody-mediated inhibition of MICA and MICB shedding promotes NK cell-driven tumor immunity. *Science* (2018) 359(6383):1537–42. doi: 10.1126/science.aao0505
50. Chin AI, Dempsey PW, Bruhn K, Miller JF, Xu Y, Cheng G. Involvement of receptor-interacting protein 2 in innate and adaptive immune responses. *Nature* (2002) 416(6877):190–4. doi: 10.1038/416190a
51. Jaafar RF, Ibrahim Z, Ataya K, Hassanieh J, Ard N, Faraj W. Receptor-interacting serine/threonine-protein kinase-2 as a potential prognostic factor in colorectal cancer. *Medicina (Mex)*. (2021) 57(7):709. doi: 10.3390/medicina57070709
52. Wu Xs, Lonsdorf AS, Hwang ST. Cutaneous T cell lymphoma: roles for chemokines and chemokine receptors. *J Invest Dermatol* (2009) 129(5):1115–9. doi: 10.1038/jid.2009.45
53. Volkart PA, Bitencourt-Ferreira G, Souto AA, de Azevedo WF. Cyclin-dependent kinase 2 in cellular senescence and cancer. A structural and functional review. *Curr Drug Targets* (2019). 20(7):716–26. doi: 10.2174/1389450120666181204165344
54. Kim M, Park J, Bouhadou M, Kim K, Rojc A, Modak M, et al. A protein interaction landscape of breast cancer. *Science* (2021) 374(6563):eab3066. doi: 10.1126/science.abf3066
55. Zhang Z, Chen J, Zhu S, Zhu D, Xu J, He G. Construction and validation of a cell cycle-related robust prognostic signature in colon cancer. *Front Cell Dev Biol* (2020) 8:611222. doi: 10.3389/fcell.2020.611222
56. Liebl MC, Hofmann TG. The role of p53 signaling in colorectal cancer. *Cancers* (2021) 13(9):2125. doi: 10.3390/cancers13092125
57. Harami-Papp H, Pongor LS, Munkácsy G, Horváth G, Nagy ÁM, Ambrus A, et al. TP53 mutation hits energy metabolism and increases glycolysis in breast cancer. *Oncotarget* (2016) 7(41):67183–95. doi: 10.18632/oncotarget.11594
58. He WL, Weng XT, Wang JL, Lin YK, Liu TW, Zhou QY, et al. Association between c-myc and colorectal cancer prognosis: A meta-analysis. *Front Physiol* (2018) 9:1549. doi: 10.3389/fphys.2018.01549
59. Gaggiani M, Mangiapane LR, Modica C, Pantina VD, Porcelli G, Di Franco S, et al. Dual inhibition of myc transcription and PI3K activity effectively targets colorectal cancer stem cells. *Cancers* (2022) 14(3):673. doi: 10.3390/cancers14030673
60. Olsson L, Lindmark G, Hammarström ML, Hammarström S, Sitohy B. Evaluating macrophage migration inhibitory factor 1 expression as a prognostic biomarker in colon cancer. *Tumour Biol J Int Soc Oncodevelopmental Biol Med* (2020) 42(6):1010428320924524. doi: 10.1177/1010428320924524
61. Mullany LE, Herrick JS, Wolff RK, Slattery ML. Single nucleotide polymorphisms within MicroRNAs, MicroRNA targets, and MicroRNA biogenesis genes and their impact on colorectal cancer survival. *Genes Chromosomes Cancer* (2017) 56(4):285–95. doi: 10.1002/gcc.22434
62. Cardoso Alves L, Corazza N, Micheau O, Krebs P. The multifaceted role of TRAIL signaling in cancer and immunity. *FEBS J* (2021) 288(19):5530–54. doi: 10.1111/febs.15637
63. Azijli K, Weyhenmeyer B, Peters GJ, de Jong S, Kruyt FAE. Non-canonical kinase signaling by the death ligand TRAIL in cancer cells: discord in the death receptor family. *Cell Death Differ* (2013) 20(7):858–68. doi: 10.1038/cdd.2013.28
64. Zhang X, Mosser D. Macrophage activation by endogenous danger signals. *J Pathol* (2008) 214(2):161–78. doi: 10.1002/path.2284
65. Chen N, Xia P, Li S, Zhang T, Wang TT, Zhu J. RNA sensors of the innate immune system and their detection of pathogens. *IUBMB Life* (2017) 69(5):297–304. doi: 10.1002/iub.1625
66. Shi Y, Yuan B, Zhu W, Zhang R, Li L, Hao X, et al. Ube2D3 and Ube2N are essential for RIG-I-mediated MAVS aggregation in antiviral innate immunity. *Nat Commun* (2017) 8:15138. doi: 10.1038/ncomms15138
67. Qing F, Liu Z. Interferon regulatory factor 7 in inflammation, cancer and infection. *Front Immunol* (2023) 14:1190841. doi: 10.3389/fimmu.2023.1190841
68. Tu D, Dou J, Wang M, Zhuang H, Zhang X. M2 macrophages contribute to cell proliferation and migration of breast cancer. *Cell Biol Int* (2021) 45(4):831–8. doi: 10.1002/cbin.11528
69. Ahmad A, Hayat A. Expression of oncogenes ELK1 and ELK3 in cancer. *Ann Colorectal Cancer Res*. (2019) 1(1):1001–6.
70. Meric-Bernstam F, Hurwitz H, Kanwal Pratap Singh R, McWilliams RR, Fakih M, VanderWalde A, et al. Pertuzumab and trastuzumab for HER2-amplified metastatic colorectal cancer: an updated report from MyPathway, a multicentre, open-label, phase 2a multiple basket study. *Lancet Oncol* (2019) 20(4):518–30. doi: 10.1016/S1470-2045(18)30904-5
71. Nami B, Ghanaeian A, Black C, Wang Z. Epigenetic silencing of HER2 expression during epithelial-mesenchymal transition leads to trastuzumab resistance in breast cancer. *Life Basel Switz*. (2021) 11(9):868. doi: 10.3390/life11090868
72. Dillon M, Lopez A, Lin E, Sales D, Perets R, Jain P. Progress on ras/MAPK signaling research and targeting in blood and solid cancers. *Cancers* (2021) 13(20):5059. doi: 10.3390/cancers13205059
73. Okuyama H, Endo H, Akashika T, Kato K, Inoue M. Downregulation of c-MYC Protein Levels Contributes to Cancer Cell Survival under Dual Deficiency of Oxygen and Glucose. *Cancer Res* (2010) 70(24):10213–23. doi: 10.1158/0008-5472.CAN-10-2720
74. Khan AA, Khan Z, Warnakulasuriya S. Cancer-associated toll-like receptor modulation and insinuation in infection susceptibility: association or coincidence? *Ann Oncol* (2016) 27(6):984–97. doi: 10.1093/annonc/mdw053
75. Xiao H, Yin W, Khan MA, Gulen MF, Zhou H, Sham HP, et al. Loss of single immunoglobulin interlukin-1 receptor-related molecule leads to enhanced colonic polyposis in Apcmin mice. *Gastroenterology* (2010) 139(2):574–85. doi: 10.1053/j.gastro.2010.04.043
76. Pereira M, Gazzinelli RT. Regulation of innate immune signaling by IRAK proteins. *Front Immunol* (2023) 14:1133354. doi: 10.3389/fimmu.2023.1133354
77. Zou Y, Cai Y, Lu D, Zhou Y, Yao Q, Zhang S. MicroRNA-146a-5p attenuates liver fibrosis by suppressing profibrogenic effects of TGFβ1 and lipopolysaccharide. *Cell Signal* (2017) 39:1–8. doi: 10.1016/j.cellsig.2017.07.016
78. Nies JF, Panzer U. IL-17C/IL-17RE: emergence of a unique axis in TH17 biology. *Front Immunol* (2020) 11:341. doi: 10.3389/fimmu.2020.00341
79. Yan C, Huang WY, Boudreau J, Mayavannan A, Cheng Z, Wang J. IL-17R deletion predicts high-grade colorectal cancer and poor clinical outcomes. *Int J Cancer* (2019) 145(2):548–58. doi: 10.1002/ijc.32122
80. Odnokoz O, Yu P, Peck AR, Sun Y, Kovatich AJ, Hooke JA, et al. Malignant cell-specific pro-tumorigenic role of type I interferon receptor in breast cancers. *Cancer Biol Ther* (2020) 21(7):629–36. doi: 10.1080/15384047.2020.1750297
81. Taylor BC, Balko JM. Mechanisms of MHC-I downregulation and role in immunotherapy response. *Front Immunol* (2022) 13:844866. doi: 10.3389/fimmu.2022.844866
82. DhatChinamoorthy K, Colbert JD, Rock KL. Cancer immune evasion through loss of MHC class I antigen presentation. *Front Immunol* (2021) 12:636568. doi: 10.3389/fimmu.2021.636568
83. Zeestraten ECM, Reimers MS, Saadatmand S, Dekker JWT, Liefers GJ, van den Elsen PJ, et al. Combined analysis of HLA class I, HLA-E and HLA-G predicts prognosis in colon cancer patients. *Br J Cancer* (2014) 110(2):459–68. doi: 10.1038/bjc.2013.696
84. Villa ALP, Parra RS, Feitosa MR, de Camargo HP, MaChado VF, Tirapelli DP da C, et al. PPARG expression in colorectal cancer and its association with staging and clinical evolution 1. *Acta Cirurgica Bras* (2020) 35(7):e202000708. doi: 10.1590/s0102-865020200070000008
85. Sabatino L, Fucci A, Pancione M, Carafa V, Nebbioso A, Pistore C, et al. UHRF1 coordinates peroxisome proliferator activated receptor gamma (PPARG) epigenetic silencing and mediates colorectal cancer progression. *Oncogene* (2012) 31(49):5061–72. doi: 10.1038/onc.2012.3
86. Jia Z, Wang M, Li S, Li X, Bai XY, Xu Z, et al. U-box ubiquitin ligase PPI2 suppresses breast cancer invasion and metastasis by altering cell morphology and promoting SNAI1 ubiquitination and degradation. *Cell Death Dis* (2018) 9(2):1–14. doi: 10.1038/s41419-017-0094-4
87. Huang L, Guo Z, Wang F, Fu L. KRAS mutation: from undruggable to druggable in cancer. *Signal Transduct Target Ther* (2021) 6:386. doi: 10.1038/s41392-021-00780-4

88. Meng M, Zhong K, Jiang T, Liu Z, Kwan HY, Su T. The current understanding on the impact of KRAS on colorectal cancer. *BioMed Pharmacother* (2021) 140:111717. doi: 10.1016/j.biopha.2021.111717
89. Liu H, Liang Z, Cheng S, Huang L, Li W, Zhou C, et al. Mutant KRAS drives immune evasion by sensitizing cytotoxic T-cells to activation-induced cell death in colorectal cancer. *Adv Sci* (2023) 10(6):2203757. doi: 10.1002/advs.202203757
90. Silva VR, Santos L de S, Dias RB, Quadros CA, Bezerra DP. Emerging agents that target signaling pathways to eradicate colorectal cancer stem cells. *Cancer Commun* (2021) 41(12):1275–313. doi: 10.1002/cac2.12235
91. Boutin AT, Liao WT, Wang M, Hwang SS, Karpinetz TV, Cheung H, et al. Oncogenic Kras drives invasion and maintains metastases in colorectal cancer. *Genes Dev* (2017) 31(4):370–82. doi: 10.1101/gad.293449.116
92. Landskron G, de la Fuente M, Thuwajit P, Thuwajit C, Hermoso MA. Chronic inflammation and cytokines in the tumor microenvironment. *J Immunol Res* (2014) 2014:149185. doi: 10.1155/2014/149185
93. Hangai S, Kimura Y, Taniguchi T, Yanai H. Signal-transducing innate receptors in tumor immunity. *Cancer Sci* (2021) 112(7):2578–91. doi: 10.1111/cas.14848
94. Daley D, Mani VR, Mohan N, Akkad N, Ochi A, Heindel DW, et al. Dectin-1 activation on macrophages by galectin-9 promotes pancreatic carcinoma and peritumoral immune-tolerance. *Nat Med* (2017) 23(5):556–67. doi: 10.1038/nm.4314
95. Cantrell D. Signaling in lymphocyte activation. *Cold Spring Harb Perspect Biol* (2015) 7(6):a018788. doi: 10.1101/cshperspect.a018788
96. Dwyer AR, Greenland EL, Pixley FJ. Promotion of tumor invasion by tumor-associated macrophages: the role of CSF-1-activated phosphatidylinositol 3 kinase and Src family kinase motility signaling. *Cancers* (2017) 9(6):68. doi: 10.3390/cancers9060068
97. Cheng H, Zheng Z, Cheng T. New paradigms on hematopoietic stem cell differentiation. *Protein Cell* (2020) 11(1):34–44. doi: 10.1007/s13238-019-0633-0
98. Guo Y, Luo J, Zou H, Liu C, Deng L, Li P. Context-dependent transcriptional regulations of YAP/TAZ in cancer. *Cancer Lett* (2022) 527:164–73. doi: 10.1016/j.canlet.2021.12.019
99. Shao DD, Xue W, Krall EB, Bhutkar A, Piccioni F, Wang X, et al. KRAS and YAP1 converge to regulate EMT and tumor survival. *Cell* (2014) 158(1):171–84. doi: 10.1016/j.cell.2014.06.004
100. Ropponen K, Kellokoski J, Pirinen R, Moisio K, Eskelinen M, Alhava E, et al. Expression of transcription factor AP-2 in colorectal adenomas and adenocarcinomas; comparison of immunohistochemistry and *in situ* hybridisation. *J Clin Pathol* (2001) 54(7):533–8. doi: 10.1136/jcp.54.7.533
101. Wang L, Lin S, Yang C, Cai S, Li W. Effect of KRAS mutations and p53 expression on the postoperative prognosis of patients with colorectal cancer. *Mol Genet Genomic Med* (2022) 10(7):e1905. doi: 10.1002/mgg3.1905
102. Farhan M, Silva M, Li S, Yan F, Fang J, Peng T, et al. The role of FOXOs and autophagy in cancer and metastasis—Implications in therapeutic development. *Med Res Rev* (2020) 40(6):2089–113. doi: 10.1002/med.21695
103. Kandasamy P, Zlobec I, Nydegger DT, Pujol-Giménez J, Bhardwaj R, Shirasawa S, et al. Oncogenic KRAS mutations enhance amino acid uptake by colorectal cancer cells via the hippo signaling effector YAP1. *Mol Oncol* (2021) 15(10):2782–800. doi: 10.1002/1878-0261.12999

Frontiers in Oncology

Advances knowledge of carcinogenesis and tumor progression for better treatment and management

The third most-cited oncology journal, which highlights research in carcinogenesis and tumor progression, bridging the gap between basic research and applications to improve diagnosis, therapeutics and management strategies.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

