







RESEARCH ARTICLE

Clinical and molecular assessment of an onco-immune signature with prognostic significance in patients with colorectal cancer

Pankaj Ahluwalia¹  | Ashis K. Mondal¹  | Meenakshi Ahluwalia²  |
Nikhil S. Sahajpal¹ | Kimya Jones¹ | Yasmeen Jilani¹ | Gagandeep K. Gahlay³  |
Amanda Barrett¹ | Vamsi Kota⁴  | Aryn M. Rojiani⁵  | Ravindra Kolhe¹ 

¹Department of Pathology, Medical College of Georgia at Augusta University, Augusta, Georgia, USA

²Department of Neurosurgery, Augusta University, Augusta, Georgia, USA

³Department of Molecular Biology and Biochemistry, Guru Nanak Dev University, Amritsar, India

⁴Department of Medicine, Medical College of Georgia at Augusta University, Augusta, Georgia, USA

⁵Department of Pathology, Penn State College of Medicine, Hershey, USA

Correspondence

Ravindra Kolhe, Medical College of Georgia, Augusta University, 1120 15th Street, Augusta, GA 30912, USA.
Email: rkolhe@augusta.edu

Funding information

The funding for this study was provided by RK.

Abstract

Understanding the complex tumor microenvironment is key to the development of personalized therapies for the treatment of cancer including colorectal cancer (CRC). In the past decade, significant advances in the field of immunotherapy have changed the paradigm of cancer treatment. Despite significant improvements, tumor heterogeneity and lack of appropriate classification tools for CRC have prevented accurate risk stratification and identification of a wider patient population that may potentially benefit from targeted therapies. To identify novel signatures for accurate prognostication of CRC, we quantified gene expression of 12 immune-related genes using a medium-throughput NanoString quantification platform in 93 CRC patients. Multivariate prognostic analysis identified a combined four-gene prognostic signature (*TGFB1*, *PTK2*, *RORC*, and *SOCS1*) (HR: 1.76, 95% CI: 1.05–2.95, * $p < 0.02$). The survival trend was captured in an independent gene expression data set: GSE17536 (177 patients; HR: 3.31, 95% CI: 1.99–5.55, * $p < 0.01$) and GSE14333 (226 patients; HR: 2.47, 95% CI: 1.35–4.53, * $p < 0.01$). Further, gene set enrichment analysis of the TCGA data set associated higher prognostic scores with epithelial–mesenchymal transition (EMT) and inflammatory pathways. Comparatively, a lower prognostic score was correlated with oxidative phosphorylation and MYC and E2F targets. Analysis of immune parameters identified infiltration of T-reg cells, CD8⁺ T cells, M2 macrophages, and B cells in high-risk patient groups along with upregulation of immune exhaustion genes. This molecular study has identified a novel prognostic gene signature with clinical utility in CRC. Therefore, along with prognostic features, characterization of immune cell infiltrates and immunosuppression provides

actionable information that should be considered while employing personalized medicine.

KEYWORDS

colon, colorectal cancer, gene signature, immune, personalized medicine, preventive, prognostic genes

1 | INTRODUCTION

Colorectal cancer is the most common digestive cancer with approximately 53,000 annual deaths in 2020 in the United States.¹ Although surgery is curative for 15%–20% of eligible CRC patients, recurrence is the unfortunate outcome for most of this resected patients.² Inherent molecular heterogeneity of CRC tumors leading to differential susceptibility to chemotherapy and immunotherapy are significant barriers to reducing the overall mortality. There is a need to identify prognostic and predictive tools that can stratify patients based on mortality risk and response to therapies, respectively. Prognostic gene signatures with underlying immune perturbations may assist in the stratification of patients for emerging personalized therapies such as NK cell-based therapies or a combination of chemotherapy and immunotherapy.^{3–5}

Recently, gene expression-based classification has helped in the characterization of several cancers with varying degrees of validation.^{6–8} In CRC, one of the most robust classifiers is based on consensus molecular subtypes (CMSs) but its application is limited due to its dependence on >500 genes and the underlying complexity of the method.⁹ There is a need to identify cost-effective clinical assays for the prognosis of CRC patients.¹⁰ Recent advances in RNA-sequencing technologies and microarray have been widely utilized to identify prognostic gene signatures. Gene signatures based on immune and lipid mediator pathways have recently been found to be prognostic in CRC patients.^{11,12} Although several gene expression biomarkers have been identified, they lack clinical validation. Thus, there is a need to identify gene signatures with prognostic significance and associated pathway perturbations that can stratify patients based on survival and assist in the application of personalized therapeutics. In this study, we have identified a novel four-gene signature with prognostic and clinical utility in CRC. The expression of 12 immune-related genes was quantified in FFPE tissues of CRC patients and was analyzed in the context of public data sets. Network analysis and gene set enrichment analysis of colorectal cancer patients identified significant perturbations

of homeostatic functions in the high-risk group. Deconvolution analyses revealed higher infiltration of Neutrophils, B cells, and macrophages in high-risk patients. Further, there was higher infiltration of CD8⁺ T cells but these tumors were also found to be enriched in immune exhaustion genes.

2 | MATERIALS AND METHODS

2.1 | Design of gene panel and data acquisition

We accessed scientific literature using PubMed to identify 12 genes with an immunological role in colorectal cancer (Table 2). The TCGA-COAD data set was analyzed for genetic mutation, transcriptome, and other clinical variations.^{13,14} The mutation data were visualized through the “maftools” package in the R program and Gene set cancer analysis.^{14,15} For external validation, gene microarray data and corresponding clinical information of verifying cohorts GSE17536 and GSE14333 were downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>).^{16–18}

2.2 | Patient samples

The study included 250 FFPE blocks of colorectal cancer patients with the protocol approved by the Institutional Review Board (IRB-HAC # 611298) from the Medical College of Georgia at Augusta University. As there were multiple blocks of each patient, only blocks with high-tumor content as assessed by a pathologist were included in this study. The samples with incomplete documentation, lack of tumor tissue in blocks, failure of RNA isolation, or degradation of RNA were excluded from this study.

2.3 | RNA isolation from FFPE blocks

Five micrometers thick, parallel sections were generated from FFPE tissues. H&E staining was performed using

a standard protocol and was examined for tumor-rich regions (>50% tumor) by a board-certified pathologist (RK). Macrodissection was performed to isolate RNA from selected regions. Total RNA was isolated through miRNEasy FFPE kit (Qiagen) using a standard protocol. The RNA was further quantified using a Nanodrop spectrophotometer (NanoDrop ND-1000, NanoDrop Technologies).

2.4 | Quantification of mRNA using NanoString platform

The quantification of mRNA molecules was performed using a digital quantification instrument (NanoString Technologies Inc.). nCounter PlexSet technology is a digital quantification system, which identifies and quantifies specific RNA molecules using a target-specific oligonucleotide probe pair. The design and construction of probes were performed by NanoString and Integrated DNA Technologies, Inc. (IDT). The target sequences, Probe A, and Probe B sequences are present in Table S1. PlexSet consists of a unique fluorescently coded barcode linked to reporter tags and a biotinylated universal capture tag. The specificity of the signal is achieved through a unique barcoded signature. The fluorescence by reporter tags is processed during subsequent data capture and analysis. The universal capture tag anchors specific RNA molecules to the streptavidin-coated lane on the nCounter instrument.¹⁹ The data collection was performed on the nCounter Digital Analyzer (DA). The field of view (FOV) setting for DA was set at 280, as previously noted.²⁰ A total of 300 ng of total RNA was used as an input for this analysis. The raw gene expression counts were further processed and normalized using nCounter software.

2.5 | Identification of prognostic gene signature

Univariate and multivariate Cox proportional hazard models were used to analyze clinicopathological variables. For every patient, the prognostic score was calculated by multiplying the expression value of a gene with its corresponding Cox proportion regression coefficient (prognostic score = \sum Cox regression coefficient of Gene_{*i*} × expression value of gene Gene_{*i*}). Patients were stratified into low- and high-risk groups according to the median risk score, and survival was assessed using the Kaplan–Meier method and the log-rank test. The hazard ratios (HRs) and log-rank *p* values were based on overall survival (OS). The combined prognostic gene signature

was validated in two independent data sets GSE17536 (*n* = 177) and GSE14333 (*n* = 226).

2.6 | Identification of differentially expressed pathways

DESeq2 package in R was utilized to identify differentially expressed genes (DEGs) (<http://bioconductor.org/packages/release/bioc/html/DESeq2.html>) between high-risk and low-risk patients in the TCGA data set. The results were graphed as a volcano plot using the “enhancedvolcano” package (<http://bioconductor.org/packages/EnhancedVolcano.html>). The Cytoscape software (<http://www.cytoscape.org/>) was used for the construction and visualization of the pathway hubs.²¹

2.7 | Gene set enrichment analysis

GSEA software downloaded from the Broad Institute (<http://www.broadinstitute.org/gsea>) was utilized to identify gene set variations in two risk groups. In this analysis, the top 25 and bottom 25 patients were compared to identify the most significantly perturbed pathways in the TCGA data set. The genes were preranked based on fold-change calculated using the DESeq2 algorithm in R and the target gene set was “H: hallmark gene sets” with the number of permutations set at 1000. Enrichment statistic was set to “weighted” and “Signal2Noise” metric was used for ranking genes.

2.8 | Immune cell infiltration and exhaustion analysis

In this study, the immune cell infiltration landscape of high- and low-risk patients was analyzed using QuantiSeq’s computational pipeline.²² The *z* scores of exhaustion genes were downloaded for TCGA-COAD cBioportal.¹³

2.9 | Statistical analysis

The normalization of raw NanoString data was performed using the nCounter software (NanoString Technologies Inc.). Briefly, the geometric mean of the negative and positive control was used to normalize the expression across all samples. Further normalization was performed using six internal control genes (*ABCF1*, *GUSB*, *HPRT1*, *LDHA*, *POLR1B*, and *RPLO*). Hierarchical clustering of correlation between immune cells and risk

groups was generated using Ward's methods. Kaplan-Meier analysis and a log-rank test was used to compare the survival distribution using the GEPIA portal (TCGA data set) and JMP-Pro for internal and GEO data sets. During the initial screening, the prognostic gene pairs were iteratively tested for prognostic significance, and H.R and *p* values were derived from the GEPIA portal. Cell plots were generated using JMP-Pro. In GSEA analysis, the normalized gene enrichment score with >2 value was considered strong. All the statistical analyses were performed using R (version 4.0, R Foundation for Statistical Computing, Vienna, Austria) (<http://www.R-project.org/>), JMP-Pro (version 15.0.0, SAS Institute), and GraphPad Prism (version 9, GraphPad Software). *p* values with ≤ 0.05 were considered statistically significant.

3 | RESULTS

3.1 | Clinical features of CRC patients

The clinicopathological features of the CRC patients included in this study are summarized in Table 1. The clinical variable included age, stage, grade, sex, vital status, metastasis, and chemotherapy. There were several socioeconomic parameters such as ethnicity, alcohol consumption, tobacco consumption, and history of cancer. Clinically, in this data set, most of the patients were >68 years of age ($n = 65$, 69.8%), stage III + IV ($n = 48$, 51.6%), and grade I + II ($n = 61$, 65.5%), and only a subset of patients received chemotherapy ($n = 31$, 33%). Among environmental variables, most patients did not smoke ($n = 59$, 63.4%) or drink alcohol ($n = 72$, 78.2%).

3.2 | Development and assessment of immune gene panel

The genes included in this study were identified from the literature and are known to play a role in the tumor microenvironment (TME) (Table 2). To further characterize their location in the cells of TME, the compartmental distribution of genes was performed using the "genecards" web portal. The distribution of three genes *TGFB1*, *PTK2*, and *BCL2L1* showed distribution in >3 compartments (Table 2). At the clinical level, the genomic analysis identified the highest gene mutations in *PTK2* and *STAT1* with 25% and 24% of TCGA-COAD patients (Figure 1A). Further, in comparison with other major cancers, the gene set showed a higher mutation frequency (Figure 1B). Copy Number Variation analysis identified the highest heterozygous amplification in *PTK2* and *BCL2L* genes

TABLE 1 Clinicopathological characteristics of CRC patients

Clinicopathological features		Total = 93 patients	
Age	>68 years	65	69.89%
	<68 years	28	30.10%
Stage	I + II	45	48.30%
	III + IV	48	51.60%
Grade	I + II	61	65.50%
	III	32	34.40%
Sex	Female	53	56.98%
	Male	40	43.01%
Vital status	Alive	33	35.43%
	Dead	60	64.51%
Metastasis	Metastasis	37	40.21%
	No metastasis	55	59.78%
Ethnicity	African American	42	46.66%
	Caucasian	48	53.33%
Alcohol consumption	Alcohol used	20	21.73%
	No alcohol use	72	78.26%
Tobacco consumption	None	59	63.44%
	Smoked	34	36.55%
Cancer history	History of cancer	38	47.50%
	No history of cancer	42	52.50%
Chemotherapy	Chemotherapy administered	31	33.33%
	No chemotherapy	62	66.66%

(Figure 1C). To further characterize the interaction of these genes, protein network analysis was performed. It identified close interactions between the genes included in this panel with major pathways related to cancer, cell surface receptor signaling, and immune system processes (Figure 1D). To further characterize the distribution of these genes in cancer, its expression was compared with the normal tissue using PCA (Figure 1E). The variance pattern resulted in a distinct separation of COAD tumors from normal tissue.

3.3 | Survival analysis

The survival analysis was performed to identify genes with good and bad prognoses in CRC patients. Iterative screening of gene combinations was performed to identify significant genes (Tables S2 and S3). The combined two-gene signature (*TGFB1* and *PTK2*) correlated with poor prognosis as its expression was found to be associated with poor overall survival in both internal and TCGA data sets (Figure 2A,B). Further, the combined

TABLE 2 Biological roles of immune genes included in this study

Gene	Entrez ID	Confidence value (cell compartment)	Role in tumorigenesis
TGFB1 (transforming growth factor beta 1)	7040	5 (PM, ECM, N, GA)	<i>TGFB1</i> plays a critical role in later stages of cancer invasion and metastasis as it promotes features such as EMT ²³
PTK2 (protein tyrosine kinase 2)	5747	5 (PM, CS, N, CY)	Elevated levels of <i>PTK2</i> were found to be associated with cancer progression or metastasis ²⁴
RORC (RAR-related orphan receptor C)	6097	5 (N)	Th17 expression of <i>RORC</i> plays a dual role in tumorigenesis ²⁵
SOCS1 (suppressor of cytokine signaling 1)	8651	5 (CS, N)	The expression of <i>SOCS1</i> was found to decrease with advanced stages and grade ²⁶
BCL2L1 (BCL2-like 1)	598	5 (CS, M, N, CY)	The expression of <i>BCL2L1</i> is higher in colorectal tumors compared to normal ²⁷
CASP3 (Caspase 3)	836	5 (N, CY)	Cleaved caspase-3 has been associated with aggressive cancer ²⁸
CCL17 (C-C motif chemokine ligand 17)	6361	5 (E)	<i>CCL17</i> leads to the migration of human cancer cells ²⁹
IDO1 (indoleamine 2,3-dioxygenase 1)	6059	5 (CY)	<i>IDO1</i> expression promotes cancer cell survival and proliferation ³⁰
IDO2 (indoleamine 2,3-dioxygenase 2)	169355	4 (CY)	<i>IDO2</i> plays a role in progression of cancer ³¹
IFNA1 (interferon alpha 1)	3439	5 (E)	Interferon therapy perturbs metastasis of colorectal cancer ³²
IFNG (interferon gamma)	3458	5 (E)	Interferon gamma deficiency leads to colorectal cancer progression ³³
STAT1 (signal transducer and activator of transcription 1)	6772	5 (N, CY)	<i>STAT1</i> can act as a potential biomarker for early stage colon cancer ³⁴

Note: The cell compartment confidence value ranged from 0 (absence of any evidence) to 5 (highest confidence). The cellular localization of these genes was incorporated from Genecards (<https://www.genecards.org/>).

Abbreviations: CS, cytoskeleton; CY, cytosol; E, extracellular; ECM, extracellular matrix; GA, golgi apparatus; M, mitochondria; N, nucleus; PM, plasma membrane.

gene signature of *RORC* and *SOCS1* showed a good prognosis in the internal and TCGA data set (Figure 2C,D). To further evaluate the prognostic distribution, a Cox proportional model was developed to identify the prognostic value of the combined four-gene signature. In a univariate analysis of the internal data set, the high-risk group was associated with worse survival (HR: 1.76, 95% CI: 1.05–2.95, $*p < 0.02$) (Table 3). Other significant variables included higher age (HR: 1.8, 95% CI: 0.98–2.95, $*p < 0.02$) and higher grade (HR: 2.09, 95% CI: 1.25–3.50, $*p < 0.02$). The multivariate model based on an internal data set maintained significance in the classification of high-risk patients (HR: 1.74, 95% CI: 1.02–2.98, $*p < 0.04$) (Table 3). Further, hazard assessment of the independent gene expression data set validated the prognostic distribution of the four-gene signature. GSE17536 (177 patients; HR: 3.31, 95% CI: 1.99–5.55, $*p < 0.01$) and GSE14333 (226 patients, HR: 2.47, 95% CI: 1.35–4.53, $*p < 0.01$) captured a similar trend of prognostication (Table 4). In the internal data

set, k-mean clustering identified two clusters of patients with a higher four-gene signature expression, showing worse survival (Figure 3A,B). The KM analysis based on the median-cutoff of prognostic score classified patients with mortality risk with significance in internal and external data sets (Figure 3C–E).

3.4 | Differential expression of genes

Differential expression analyses between high- and low-risk patient groups were performed using DESeq2 analysis (Table S4). Among 20,501 protein-coding genes, 561 genes were upregulated >2-fold in high-risk patients compared with the lower-risk group. In the low-risk group, a total of 93 genes were upregulated >2-fold compared with the high-risk group. A volcano plot of differentially expressed genes between high-risk and low-risk patients is depicted in Figure 4A. The enriched gene ontology terms in the low-risk group included pathways such as

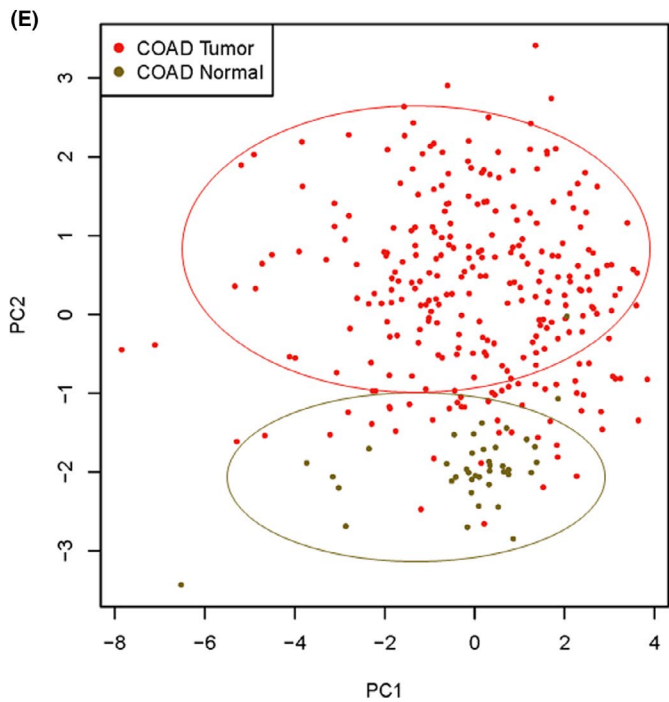
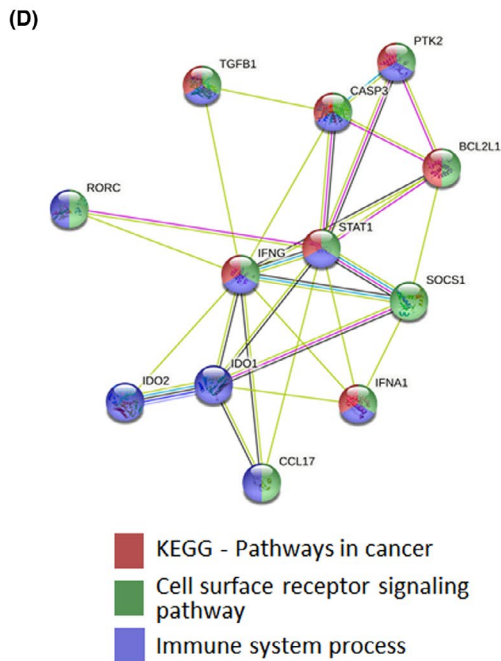
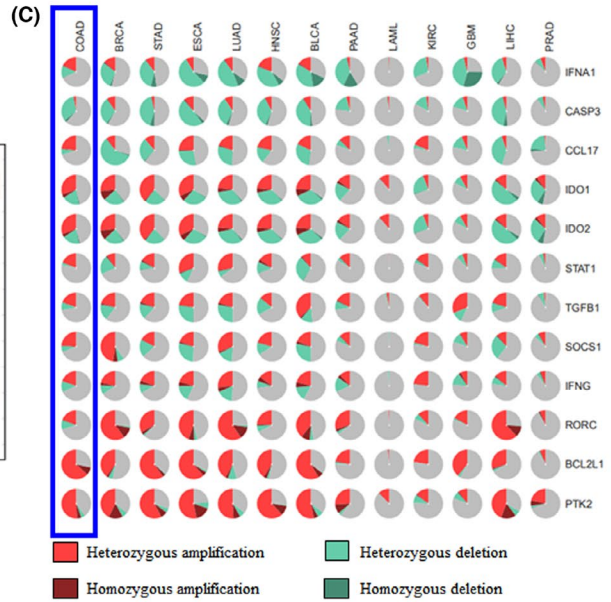
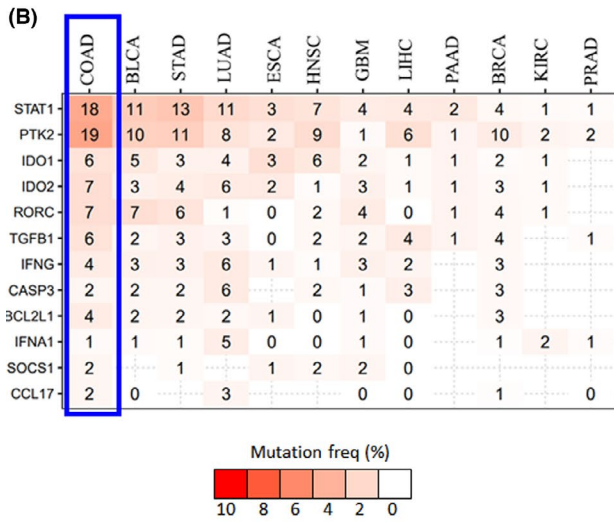
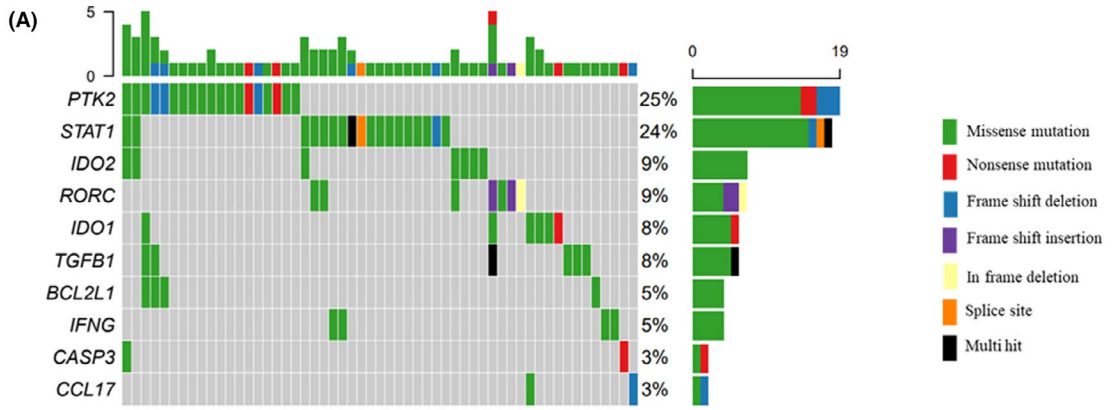


FIGURE 1 Mutation profiling of immune gene panel in TCGA-COAD. (A) Waterfall plot showing the landscape of mutations in COAD samples. The tumor mutation burden is depicted as a bar plot above the legend. Different colors depict specific mutation types. (B) SNV (single nucleotide variation) percentage heatmap depicting higher frequency in COAD data set, (C) comparative pie-plot summary of CNV percentage in major cancer, (D) network analysis of immune genes (E) PCA analysis of 12 genes in normal versus tumor tissue of TCGA-COAD CRC samples

FIGURE 2 Survival analysis. (A) The combined prognostic score of *TGFB1* and *PTK2* in the internal data set and (B) TCGA data set. (C) The combined prognostic score of *RORC* and *SOCS1* in the internal data set and (D) TCGA data set

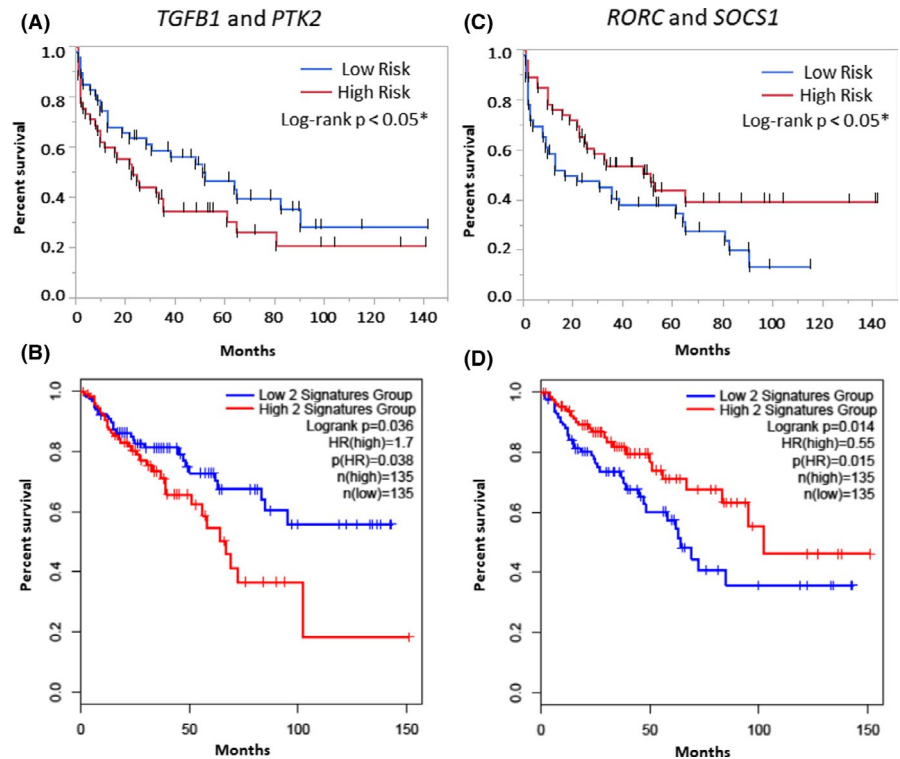


TABLE 3 Univariate and multivariate Cox proportional hazard analysis of the internal data set

Variable	Univariate			Multivariate		
	Hazard ratio	95% CI	p value	Hazard ratio	95% CI	p value
Combined four-gene signature (high, low)	1.76	1.05–2.95	0.02*	1.74	1.02–2.98	0.04*
Age (>68 years, <68 years)	1.8	0.98–3.30	0.05	1.72	0.90–3.26	0.09
Stage (I + II, III + IV)	0.86	0.51–1.44	0.51	1.12	0.62–2.00	0.7
Sex (male, female)	1	0.60–1.68	0.97	1.06	0.63–1.78	0.82
Chemotherapy (yes, no)	0.94	0.54–1.62	0.83	1.05	0.30–1.10	0.09
Grade (I + II, III)	2.09	1.25–3.50	0.004*			
Metastasis (metastasis, no metastasis)	0.82	0.48–1.41	0.48			
Ethnicity (African American, Caucasian)	1.42	0.85–2.37	0.17			
Alcohol consumption (yes, no)	1.12	0.61–2.08	0.71			
Tobacco consumption (yes, no)	0.79	0.46–1.34	0.38			
Cancer history (yes, no)	0.59	0.33–1.05	0.07			

*p < 0.05

translation, RNA processing, translation, and mitochondrial translocation (Figure 4B). The enriched pathways in the high-risk group were immune-related pathways such as T cell migration, complement and defense response,

and acute-phase response pathways (Figure 4C). Further, GSEA identified inflammatory gene expression and EMT pathways compared with MYC and E2F target genes enrichment (Figure 4D).

TABLE 4 Validation of prognostic scores in independent GEO data set

Variable	GSE 17536			GSE 14333		
	Hazard ratio	95% CI	p value	Hazard ratio	95% CI	p value
Combined four-gene signature (high, low)	3.31	1.99–5.55	0.01*	2.47	1.35–4.53	0.01*
Age (>50 years, <50 years)	0.77	0.36–1.61	0.49	0.68	0.33–1.41	0.33
Sex (male, female)	1.1	0.69–1.75	0.67	0.9	0.62–1.92	0.73
Chemotherapy (yes, no)				1.89	1.08–3.29	0.02*
Stage (III + IV, I + II)	4.22	2.3–7.46	0.01*			
Grade (III, II + I)	2.19	1.25–3.82	0.01*			
Ethnicity (African American, Caucasian)	2.26	0.97–5.25	0.05*			

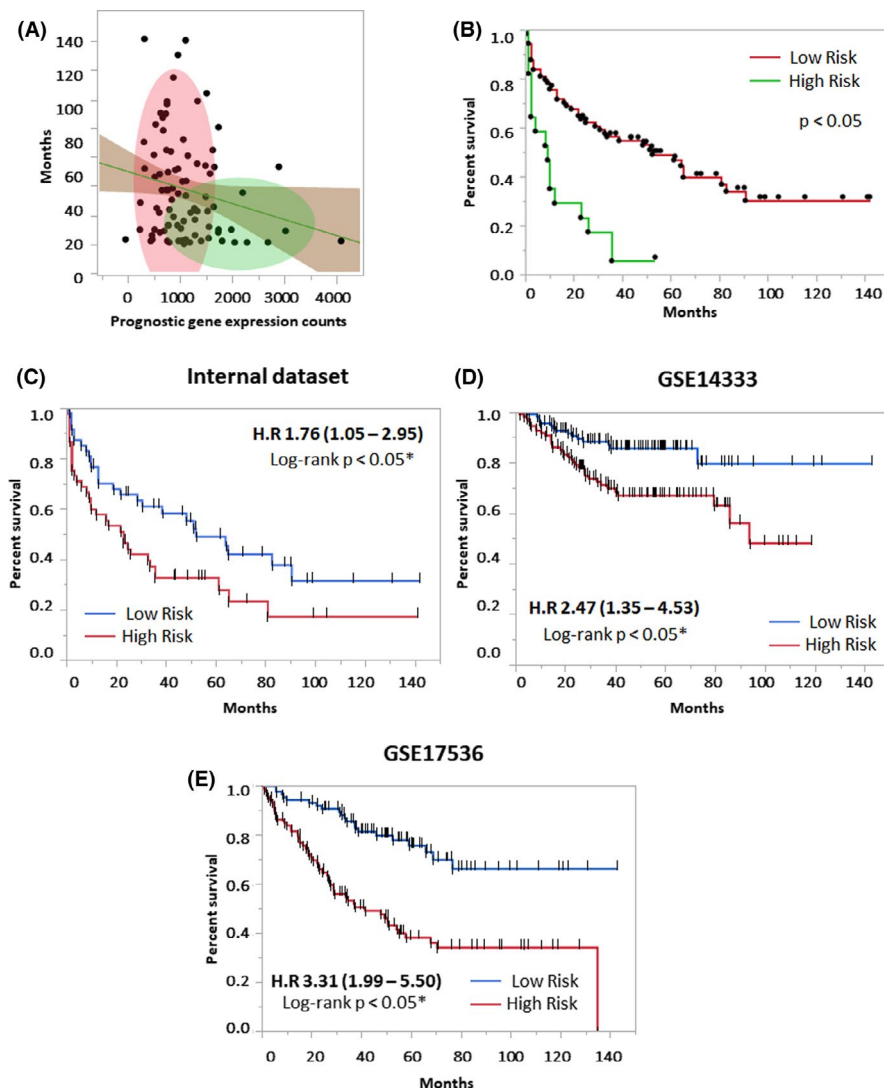
* $p \leq 0.05$ 

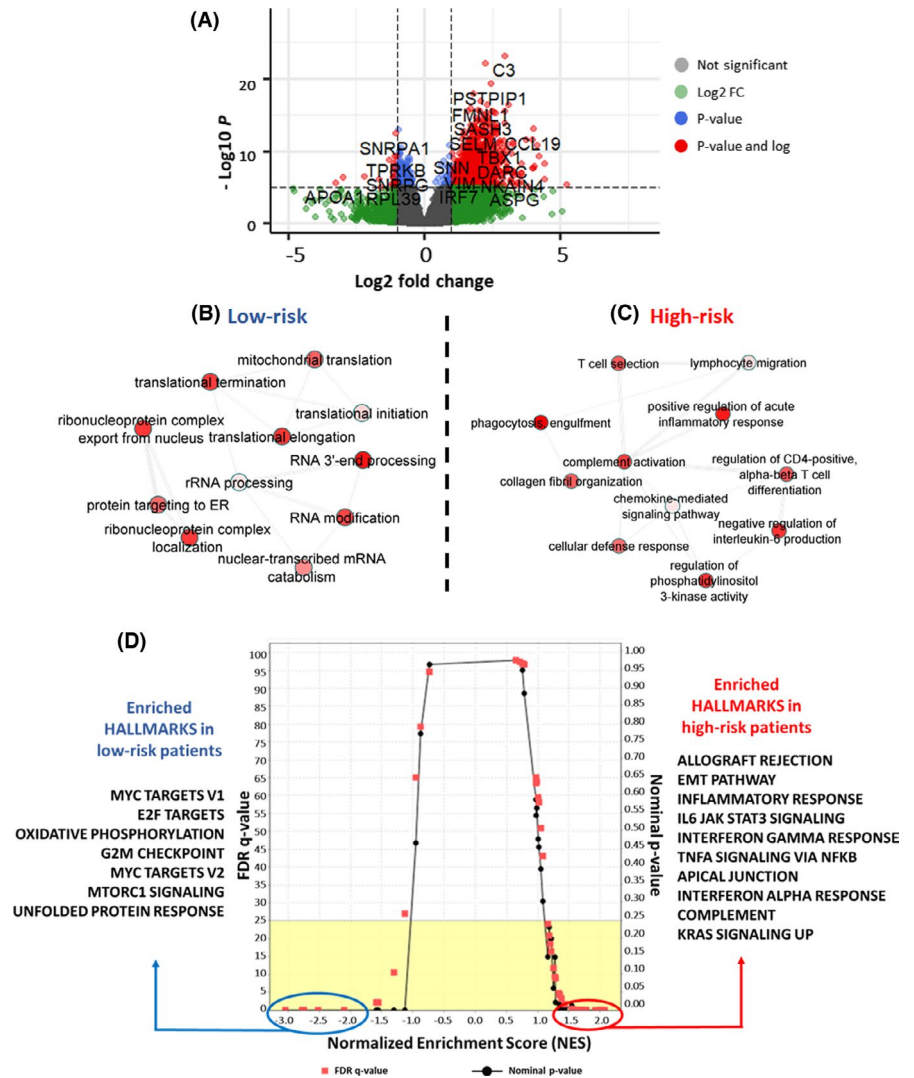
FIGURE 3 Combined four-gene prognostic assessment in internal and external data sets. (A) k-mean clustering of gene expression and overall survival in internal data sets; (B) survival difference between higher and lower gene expression clusters (C) KM estimate based on the four-gene prognostic score in the internal data set (D) external GEO data set GSE14333, and (E) external GEO data set GSE17536

3.5 | Gene set enrichment analysis

To analyze gene perturbation in two risk clusters, GSEA results were analyzed through normalized enrichment scores (NESs). The HALLMARK pathways with the

highest NES in the high-risk group were allograft rejection (NES = 2.07, FDR \leq 0.001), epithelial–mesenchymal transition (NES = 2.03, FDR \leq 0.001), and inflammatory response (NES = 2.01, FDR \leq 0.001) (Figure 5A–C). Interestingly, patients in the low-risk category showed

FIGURE 4 (A) Volcano plot depicting differential expression of genes in high-risk and low-risk CRC patients in TCGA data set. (B) GO term enrichment in the low-risk group. (C) GO term enrichment in the high-risk group and (D) differential enrichment of hallmarks in cancer in two risk groups



enrichment in MYC targets (NES = -3.05, FDR ≤ 0.001), E2F targets (NES = -2.78, FDR ≤ 0.001), and oxidative phosphorylation (NES = -2.75, FDR ≤ 0.001) (Figure 5D–F). Cell plot analysis captured the variation of gene expression in the two risk groups (Figure 5G).

3.6 | Infiltration of immune cells and exhaustion analysis

To analyze the distribution of immune cells in these two risk groups, the Quantiseq algorithm was utilized. Patients at high risk showed an abundance of B cells, CD8⁺ T cells, and T-reg cells (Figure 6A). Among macrophages, M2 macrophages showed a significant association with the high-risk group (Figure 6B). To further evaluate the immunosuppressive microenvironment, the expression of 10 immune gene exhaustion genes was quantified in both the high-risk and low-risk groups. These genes were mostly involved in the negative regulation of T cells (Figure 6C). Interestingly, patients in the high-risk group showed

higher expression (higher z-score) of all the immune exhaustion genes (Figure 6D).

4 | DISCUSSION

CRC is a complex disease with a heterogeneous tumor microenvironment that plays a critical role in disease progression.³⁵ The advancements in transcriptomics have led to the identification of several gene expression-based classification methods for colorectal cancer.^{36–39} Recently, an inflammatory gene module of 14 genes has been identified as prognostic biomarkers for colorectal cancer.⁴⁰ Despite several advances, there is limited progress in the validation and characterization of biological networks in high-risk populations. In this study, we have identified a four-gene prognostic signature (*TGFB1*, *PTK2*, *RORC*, and *SOCS1*) using FFPE tissue and assessed its distribution in public data sets. TGFB1 is a ligand secreted by several cells in the tumor microenvironment and plays a critical role in lung carcinoma.⁴¹ Its activity

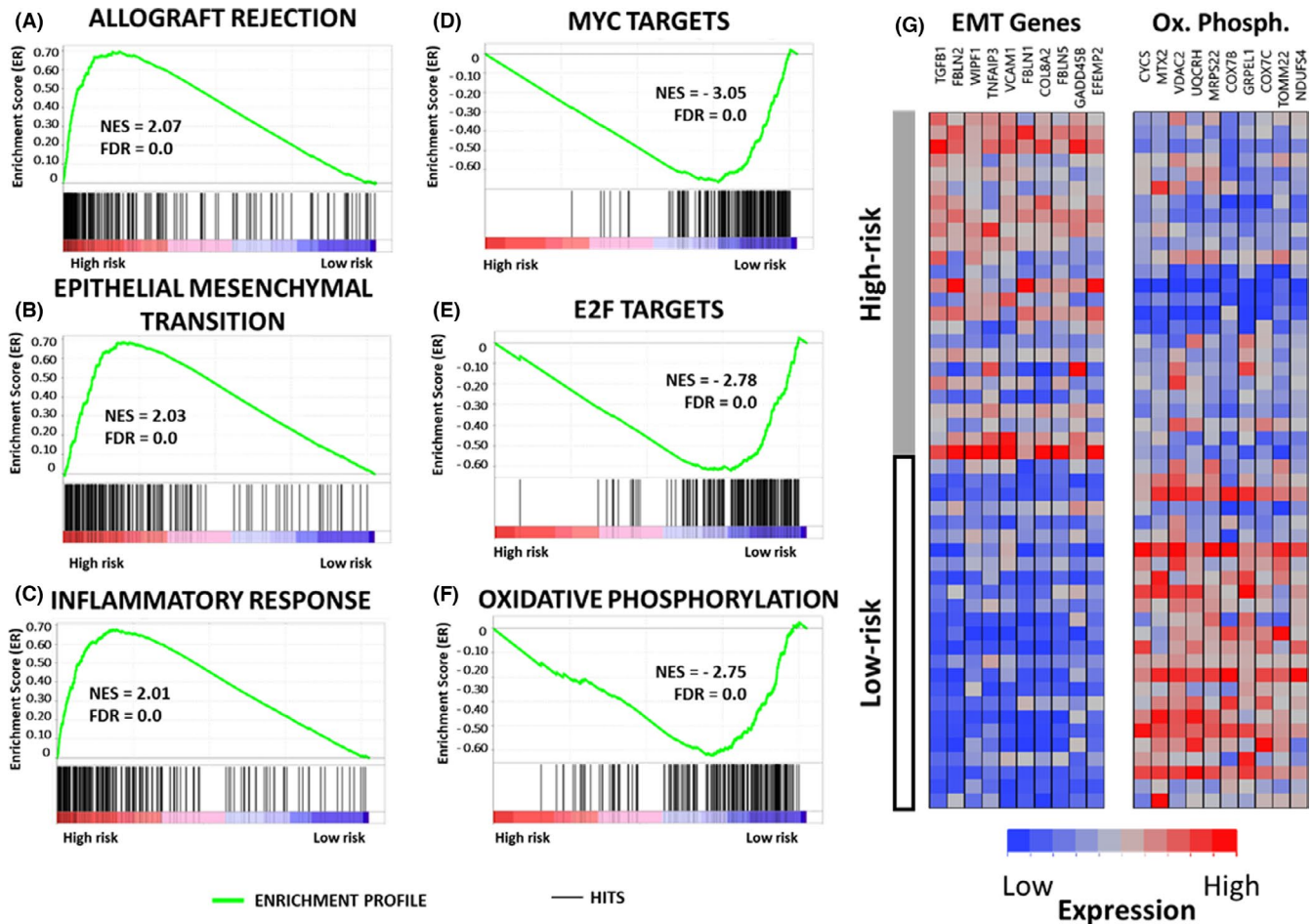


FIGURE 5 Gene set enrichment analysis based on risk stratification in TCGA data set. In the high-risk group, enriched pathways included (a) allograft rejection, (b) epithelial–mesenchymal transition, and (c) inflammatory response. In the low-risk group, (d) MYC targets, (e) E2F targets, and (f) oxidative phosphorylation were enriched. (g) Comparative analysis of the high- and low-risk groups, depicting differential expression of genes in key pathways

affects cell proliferation, differentiation, metastasis, cell adhesion, and resistance to drugs.^{42,43} Further, TGF β expression in the tumor microenvironment fuels transformation of normal fibroblasts into cancer-associated fibroblasts (CAFs) in CRC.⁴⁴ Also, the TGF β pathway is a key player in EMT, tumor invasion, metastasis to the liver, and angiogenesis.⁴³ In a recent study, the TGF β -activating-like gene phenotype showed worse survival compared with the TGF β -inactivating-like signature.⁴⁵ In a recent study on the serum of CRC patients, TGF β 1 levels were found to be higher in untreated patients compared with patients under chemotherapeutic treatment. In this study, TGF β 1 signaling was found to be critical for metastasis and stromal cell independence.⁴⁶ A second gene, protein tyrosine kinase 2 (PTK2) is a member of the nonreceptor tyrosine kinase and regulates cell proliferation, migration, and invasion.⁴⁷ Higher PTK2 expression is associated with cancer progression, metastasis, and poor overall survival.⁴⁸ In a network analysis of different genes, PTK2 was identified as an inflammation-related

gene signature in CRC.⁴⁰ Further, PTK2 activation confers adaptive resistance to chemotherapy in CRC.⁴⁹ The third gene in our set, RORC is a member of the nuclear orphan receptor family and is involved in cell growth, metastasis, and resistance to chemotherapy in various tumors. In bladder cancer, higher RORC expression led to suppression of cell growth, glucose metabolism, and negatively regulated growth of PD-L1.⁵⁰ Further, it has been shown that the tissue microenvironment plays a critical role in tumor-suppressive effects of immune cells.⁵¹ SOCS1 negatively regulates JAK–STAT signaling pathways and is involved in the inhibition of inflammation. In an earlier study, SOCS1 was identified as a tumor suppressor in certain colorectal cancer patients.⁵² SOCS1 has also been shown to have a role in tumor progression in CRC.⁵³ It has been previously reported that SOCS1-based suppression of Src activity led to downregulation of the EMT pathway.⁵⁴ Thus, the function of SOCS1 can be either tumor-suppressive or oncogenic and is dependent on tissue *milieu*.⁵⁵

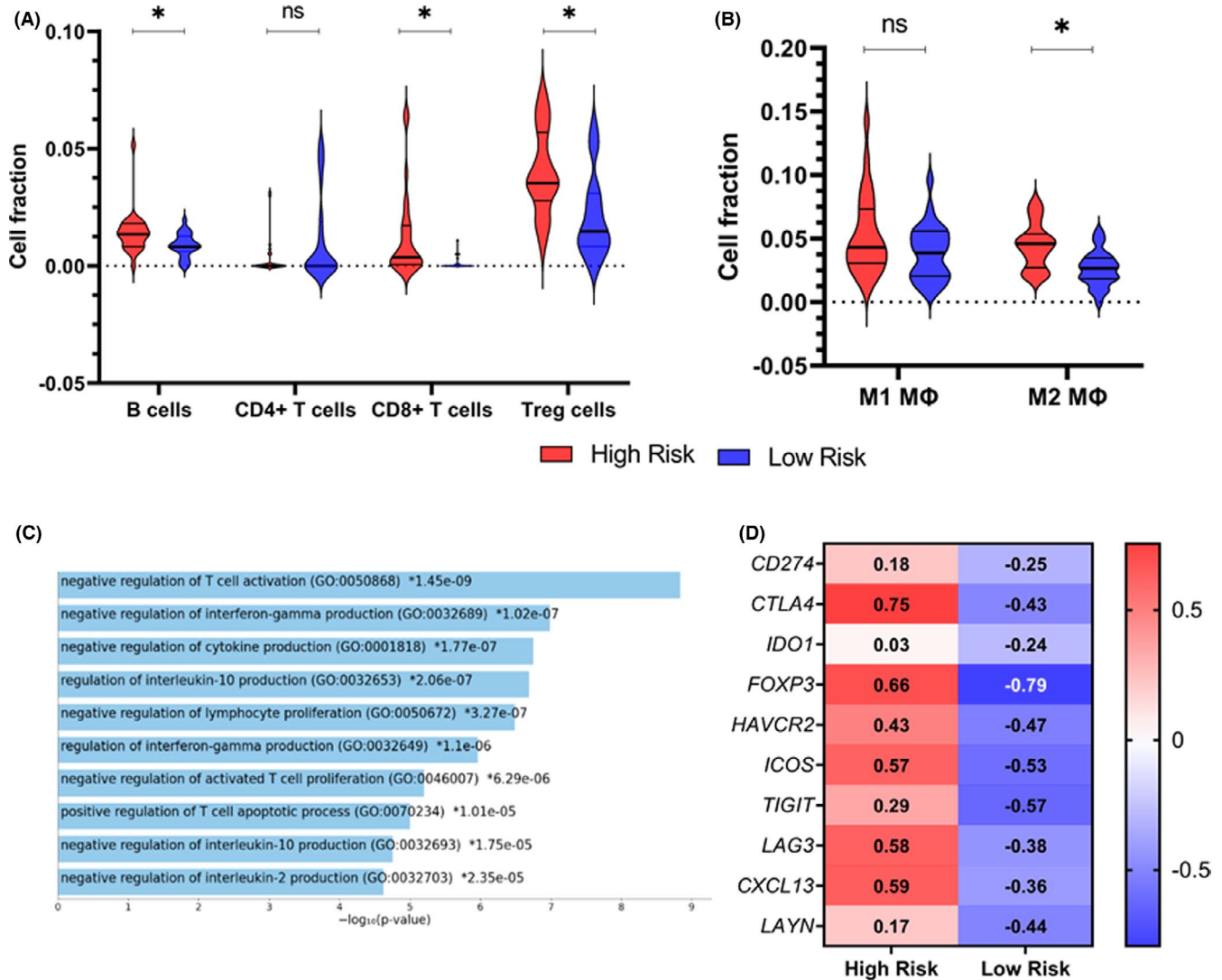


FIGURE 6 (A) Differential infiltration of immune cells in high-risk and low-risk groups of TCGA data set. (B) Infiltration of macrophages. (C) Pathway clustering of exhaustion genes. (D) z-score of exhaustion genes in the two risk groups

In this study, network analysis identified significant upregulation of immune-related and EMT pathways in high-risk patients. EMT progression has been shown to play a central role in tumor initiation and metastasis of CRC.⁵⁶ In low-risk CRC patients, cancer progression was found to depend upon oxidative phosphorylation and MYC and E2F targets. In CRC tumors, OXPHOS has been shown to promote tumor progression, metastasis, and resistance to drugs.⁵⁷ Further, our study identified higher infiltration of T-regulatory cells and M2 macrophages in the high-risk group. Previous analysis in CRC patients has associated infiltration of T-regulatory cells and M2 macrophages with poor prognosis.⁵⁸ Additionally, we identified higher infiltration of B cells, CD8+ T cells, and immune exhaustion genes in high-risk patients. Several studies have documented the upregulation of immune exhaustion in CRC tumors.^{59,60}

In a recent study on metastatic CRC mouse model, TGFβ blockade led to susceptibility to anti-PD1-PD-L1 therapy.⁶¹ Although anti-TGFβ has shown improved efficacy in PD-1/PD-L1, its clinical assessment remains under study.⁶² Similarly, to counter the effects of PTK2 overexpression, PTK2 inhibitors are being explored in several clinical trials.^{63,64} On the other hand, RORC was included in an immune panel that can act as a predictive biomarker for determining the efficacy of anti-PD-1 treatment in NSCLC patients.⁶⁵ In this study, we have correlated gene expression of inflammatory, EMT pathways, and immune exhaustion genes with high-risk patients that can provide insights for stratification of patients for personalized therapies. This strategy can also improve the efficacy of current treatment strategies and can lead to better outcomes for CRC patients.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTION

Pankaj Ahluwalia and Ravindra Kolhe conceived, designed, and wrote the first draft. Pankaj Ahluwalia, Ashis K. Mondal, Meenakshi Ahluwalia, Nikhil S. Sahajpal, Kimya Jones, and Yasmeen Jilani performed the experiments. Meenakshi Ahluwalia, Gagandeep K. Gahlay, Amanda Barrett, Vamsi Kota, and Aryn M. Rojiani helped in the manuscript and data review. Ravindra Kolhe supervised and provided financial resources for the study. All authors approved the final version of the manuscript.

ETHICS APPROVAL

The study was approved by the Institutional Review Board (IRB-HAC # 611298) of Augusta University.

DATA AVAILABILITY STATEMENT

The data sets used and/or analyzed in the current study are available from the corresponding author on reasonable request.

ORCID

Pankaj Ahluwalia  <https://orcid.org/0000-0003-0060-6152>

Ashis K. Mondal  <https://orcid.org/0000-0003-3826-9489>

Meenakshi Ahluwalia  <https://orcid.org/0000-0003-3009-7344>

Gagandeep K. Gahlay  <https://orcid.org/0000-0002-0058-2827>

Vamsi Kota  <https://orcid.org/0000-0002-5290-9289>

Aryn M. Rojiani  <https://orcid.org/0000-0001-5621-3724>

Ravindra Kolhe  <https://orcid.org/0000-0002-8283-2403>

REFERENCES

1. Siegel RL, Miller KD, Goding Sauer A, et al. Colorectal cancer statistics, 2020. *CA Cancer J Clin.* 2020;70(3):145-164. doi:10.3322/caac.21601
2. Nordlinger B, Sorbye H, Glimelius B, et al.; EORTC Gastro-Intestinal Tract Cancer Group, Cancer Research UK, Arbeitsgruppe Lebermetastasen und-tumoren in der Chirurgischen Arbeitsgemeinschaft Onkologie (ALM-CAO), Australasian Gastro-Intestinal Trials Group (AGITG), Fédération Francophone de Cancérologie Digestive (FFCD). Perioperative FOLFOX4 chemotherapy and surgery versus surgery alone for resectable liver metastases from colorectal cancer (EORTC 40983): long-term results of a randomised, controlled, phase 3 trial. *Lancet Oncol.* 2013;14(12):1208-1215. doi:10.1016/S1470-2045(13)70447-9
3. Turin I, Delfanti S, Ferulli F, et al. In vitro killing of colorectal carcinoma cells by autologous activated NK cells is boosted by anti-epidermal growth factor receptor-induced ADCC regardless of RAS mutation status. *J Immunother.* 2018;41(4):190-200. doi:10.1097/CJI.0000000000000205
4. Lanuza PM, Viguera A, Oliván S, et al. Activated human primary NK cells efficiently kill colorectal cancer cells in 3D spheroid cultures irrespectively of the level of PD-L1 expression. *Onco Targets Ther.* 2018;7(4):e1395123. doi:10.1080/2162402X.2017.1395123
5. Yan Y, Kumar AB, Finnes H, et al. Combining immune checkpoint inhibitors with conventional cancer therapy. *Front Immunol.* 2018;9:1739. doi:10.3389/fimmu.2018.01739
6. Ahluwalia P, Ahluwalia M, Mondal AK, et al. Prognostic and therapeutic implications of extracellular matrix associated gene signature in renal clear cell carcinoma. *Sci Rep.* 2021;11(1):7561. doi:10.1038/s41598-021-86888-7
7. Wang Z, Embaye KS, Yang Q, et al. Establishment and validation of a prognostic signature for lung adenocarcinoma based on metabolism-related genes. *Cancer Cell Int.* 2021;21(1):219. doi:10.1186/s12935-021-01915-x
8. Zhang Z, Chen C, Fang Y, et al. Development of a prognostic signature for esophageal cancer based on nine immune related genes. *BMC Cancer.* 2021;21(1):113. doi:10.1186/s12885-021-07813-9
9. Buechler SA, Stephens MT, Hummon AB, et al. ColoType: a forty gene signature for consensus molecular subtyping of colorectal cancer tumors using whole-genome assay or targeted RNA-sequencing. *Sci Rep.* 2020;10(1):12123. doi:10.1038/s41598-020-69083-y
10. Cai D, Wang W, Zhong M-E, et al. Deep learning to identify a gene signature associated with molecular subtypes that predicts prognosis in colorectal cancer. *J Clin Oncol.* 2021;39(15):e15520-e15520.
11. Wang J, Yu S, Chen G, et al. A novel prognostic signature of immune-related genes for patients with colorectal cancer. *J Cell Mol Med.* 2020;24(15):8491-8504. doi:10.1111/jcmm.15443
12. Gharib E, Nasrinasrabadi P, Zali MR. Development and validation of a lipogenic genes panel for diagnosis and recurrence of colorectal cancer. *PLoS One.* 2020;15(3):e0229864. doi:10.1371/journal.pone.0229864
13. Gao J, Aksoy BA, Dogrusoz U, 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
14. Liu CJ, Hu FF, Xia MX, Han L, Zhang Q, Guo AY. GSCALite: a web server for gene set cancer analysis. *Bioinformatics.* 2018;34(21):3771-3772. doi:10.1093/bioinformatics/bty411
15. 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):1747-1756. doi:10.1101/gr.239244.118
16. Smith JJ, Deane NG, Wu F, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology.* 2010;138(3):958-968. doi:10.1053/j.gastro.2009.11.005
17. Barrett T, Troup DB, Wilhite SE, et al. NCBI GEO: mining tens of millions of expression profiles--database and tools update. *Nucleic Acids Res.* 2007;35(Database issue):D760-D765. doi:10.1093/nar/gkl887

18. Jorissen RN, Gibbs P, Christie M, et al. Metastasis-associated gene expression changes predict poor outcomes in patients with dukes stage B and C colorectal cancer. *Clin Cancer Res*. 2009;15(24):7642-7651. doi:10.1158/1078-0432.CCR-09-1431
19. Kulkarni MM. Digital multiplexed gene expression analysis using the NanoString nCounter system. *Curr Protoc Mol Biol*. 2011;Chapter 25(94.1):25B-10. doi:10.1002/0471142727.mb25b10s94
20. Veldman-Jones MH, Brant R, Rooney C, et al. Evaluating robustness and sensitivity of the NanoString technologies nCounter platform to enable multiplexed gene expression analysis of clinical samples. *Cancer Res*. 2015;75(13):2587-2593. doi:10.1158/0008-5472.CAN-15-0262
21. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498-2504. doi:10.1101/gr.1239303
22. Finotello F, Mayer C, Plattner C, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med*. 2019;11(1):34. doi:10.1186/s13073-019-0638-6
23. Batista KP, De Pina KAR, Ramos AA, Vega IF, Saiz A, Alvarez Vega MA. The role of contextual signal TGF-beta1 inducer of epithelial mesenchymal transition in metastatic lung adenocarcinoma patients with brain metastases: an update on its pathological significance and therapeutic potential. *Contemp Oncol*. 2019;23(4):187-194. doi:10.5114/wo.2019.91543
24. Tong X, Tanino R, Sun R, et al. Protein tyrosine kinase 2: a novel therapeutic target to overcome acquired EGFR-TKI resistance in non-small cell lung cancer. *Respir Res*. 2019;20(1):270. doi:10.1186/s12931-019-1244-2
25. De Simone V, Pallone F, Monteleone G, Stolfi C. Role of TH17 cytokines in the control of colorectal cancer. *Onco Targets Ther*. 2013;2(12):e26617. doi:10.4161/onci.26617
26. David M, Naudin C, Letourneur M, et al. Suppressor of cytokine signaling 1 modulates invasion and metastatic potential of colorectal cancer cells. *Mol Oncol*. 2014;8(5):942-955. doi:10.1016/j.molonc.2014.03.014
27. Sillars-Hardebol AH, Carvalho B, Belien JA, et al. BCL2L1 has a functional role in colorectal cancer and its protein expression is associated with chromosome 20q gain. *J Pathol*. 2012;226(3):442-450. doi:10.1002/path.2983
28. Hu Q, Peng J, Liu W, et al. Elevated cleaved caspase-3 is associated with shortened overall survival in several cancer types. *Int J Clin Exp Pathol*. 2014;7(8):5057-5070.
29. Al-haidari AA, Syk I, Jirstrom K, Thorlacius H. CCR4 mediates CCL17 (TARC)-induced migration of human colon cancer cells via RhoA/rho-kinase signaling. *Int J Color Dis*. 2013;28(11):1479-1487. doi:10.1007/s00384-013-1712-y
30. Bishnupuri KS, Alvarado DM, Khouri AN, et al. IDO1 and kynurenine pathway metabolites activate PI3K-Akt signaling in the neoplastic colon epithelium to promote cancer cell proliferation and inhibit apoptosis. *Cancer Res*. 2019;79(6):1138-1150. doi:10.1158/0008-5472.CAN-18-0668
31. Li P, Xu W, Liu F, et al. The emerging roles of IDO2 in cancer and its potential as a therapeutic target. *Biomed Pharmacother*. 2021;137:111295. doi:10.1016/j.biopha.2021.111295
32. Di Franco S, Turdo A, Todaro M, Stassi G. Role of type I and II interferons in colorectal cancer and melanoma. *Front Immunol*. 2017;8:878. doi:10.3389/fimmu.2017.00878
33. Wang L, Wang Y, Song Z, Chu J, Qu X. Deficiency of interferon-gamma or its receptor promotes colorectal cancer development. *J Interf Cytokine Res*. 2015;35(4):273-280. doi:10.1089/jir.2014.0132
34. Tanaka A, Zhou Y, Ogawa M, et al. STAT1 as a potential prognosis marker for poor outcomes of early stage colorectal cancer with microsatellite instability. *PLoS One*. 2020;15(4):e0229252. doi:10.1371/journal.pone.0229252
35. Kasprzak A. The role of tumor microenvironment cells in colorectal cancer (CRC) cachexia. *Int J Mol Sci*. 2021;22(4):1565. doi:10.3390/ijms22041565
36. Ahluwalia P, Mondal AK, Bloomer C, et al. Identification and clinical validation of a novel 4 gene-signature with prognostic utility in colorectal cancer. *Int J Mol Sci*. 2019;20(15):3818. doi:10.3390/ijms20153818
37. Ahluwalia P, Kolhe R, Gahlay GK. The clinical relevance of gene expression based prognostic signatures in colorectal cancer. *Biochim Biophys Acta Rev Cancer*. 2021;1875(2):188513. doi:10.1016/j.bbcan.2021.188513
38. Wang L, Jiang X, Zhang X, Shu P. Prognostic implications of an autophagy-based signature in colorectal cancer. *Medicine*. 2021;100(13):e25148. doi:10.1097/MD.0000000000025148
39. Qian Y, Wei J, Lu W, et al. Prognostic risk model of immune-related genes in colorectal cancer. *Front Genet*. 2021;12:619611. doi:10.3389/fgene.2021.619611
40. Jiang H, Dong L, Gong F, et al. Inflammatory genes are novel prognostic biomarkers for colorectal cancer. *Int J Mol Med*. 2018;42(1):368-380. doi:10.3892/ijmm.2018.3631
41. Pellatt AJ, Mullany LE, Herrick JS, et al. The TGFbeta-signaling pathway and colorectal cancer: associations between dysregulated genes and miRNAs. *J Transl Med*. 2018;16(1):191. doi:10.1186/s12967-018-1566-8
42. Soleimani A, Khazaei M, Ferns GA, Ryzhikov M, Avan A, Hassanian SM. Role of TGF-beta signaling regulatory microRNAs in the pathogenesis of colorectal cancer. *J Cell Physiol*. 2019;234:14574-14580. doi:10.1002/jcp.28169
43. Chruscik A, Gopalan V, Lam AK. The clinical and biological roles of transforming growth factor beta in colon cancer stem cells: a systematic review. *Eur J Cell Biol*. 2018;97(1):15-22. doi:10.1016/j.ejcb.2017.11.001
44. Chen W, Chen Y, Su J, et al. CaMKII mediates TGFbeta1-induced fibroblasts activation and its cross talk with colon cancer cells. *Dig Dis Sci*. 2021;67:134-145. doi:10.1007/s10620-021-06847-0
45. Schlicker A, Ellappalayam A, Beumer IJ, et al. Investigating the concordance in molecular subtypes of primary colorectal tumors and their matched synchronous liver metastasis. *Int J Cancer*. 2020;147(8):2303-2315. doi:10.1002/ijc.33003
46. Chiavarina B, Costanza B, Ronca R, et al. Metastatic colorectal cancer cells maintain the TGFbeta program and use TGFBI to fuel angiogenesis. *Theranostics*. 2021;11(4):1626-1640. doi:10.7150/thno.51507
47. Aboubakar Nana F, Lecocq M, Ladjemi MZ, et al. Therapeutic potential of focal adhesion kinase inhibition in small cell lung cancer. *Mol Cancer Ther*. 2019;18(1):17-27. doi:10.1158/1535-7163.MCT-18-0328
48. Zhao J, Guan JL. Signal transduction by focal adhesion kinase in cancer. *Cancer Metastasis Rev*. 2009;28(1-2):35-49. doi:10.1007/s10555-008-9165-4

49. Taylor KN, Schlaepfer DD. Adaptive resistance to chemotherapy: a multi-FAK-torial linkage. *Mol Cancer Ther.* 2018;17(4):719-723. doi:10.1158/1535-7163.MCT-17-1177
50. Cao D, Qi Z, Pang Y, et al. Retinoic acid-related orphan receptor C regulates proliferation, glycolysis, and chemoresistance via the PD-L1/ITGB6/STAT3 signaling Axis in bladder cancer. *Cancer Res.* 2019;79(10):2604-2618. doi:10.1158/0008-5472.CAN-18-3842
51. Nussbaum K, Burkhard SH, Ohs I, et al. Tissue microenvironment dictates the fate and tumor-suppressive function of type 3 ILCs. *J Exp Med.* 2017;214(8):2331-2347. doi:10.1084/jem.20162031
52. Fujitake S, Hibi K, Okochi O, et al. Aberrant methylation of SOCS-1 was observed in younger colorectal cancer patients. *J Gastroenterol.* 2004;39(2):120-124. doi:10.1007/s00535-003-1262-0
53. Tobelaim WS, Beaurivage C, Champagne A, et al. Tumour-promoting role of SOCS1 in colorectal cancer cells. *Sci Rep.* 2015;5:14301. doi:10.1038/srep14301
54. Jung SH, Kim SM, Lee CE. Mechanism of suppressors of cytokine signaling 1 inhibition of epithelial-mesenchymal transition signaling through ROS regulation in colon cancer cells: suppression of Src leading to thioredoxin up-regulation. *Oncotarget.* 2016;7(38):62559-62571. doi:10.18632/oncotarget.11537
55. Beaurivage C, Champagne A, Tobelaim WS, Pomerleau V, Menendez A, Saucier C. SOCS1 in cancer: an oncogene and a tumor suppressor. *Cytokine.* 2016;82:87-94. doi:10.1016/j.cyto.2016.01.005
56. Shan Z, Wu W, Yan X, et al. A novel epithelial-mesenchymal transition molecular signature predicts the oncological outcomes in colorectal cancer. *J Cell Mol Med.* 2021;25(7):3194-3204. doi:10.1111/jcmm.16387
57. Huang Q, Chen Z, Cheng P, et al. LYRM2 directly regulates complex I activity to support tumor growth in colorectal cancer by oxidative phosphorylation. *Cancer Lett.* 2019;455:36-47. doi:10.1016/j.canlet.2019.04.021
58. Xu X, Ma J, Yu G, Qiu Q, Zhang W, Cao F. Effective predictor of colorectal cancer survival based on exclusive expression pattern among different immune cell infiltration. *J Histochem Cytochem.* 2021;69(4):271-286. doi:10.1369/0022155421991938
59. Liang R, Zhu X, Lan T, et al. TIGIT promotes CD8(+)T cells exhaustion and predicts poor prognosis of colorectal cancer. *Cancer Immunol Immunother.* 2021;70:2781-2793. doi:10.1007/s00262-021-02886-8
60. Saleh R, Taha RZ, Toor SM, et al. Expression of immune checkpoints and T cell exhaustion markers in early and advanced stages of colorectal cancer. *Cancer Immunol Immunother.* 2020;69(10):1989-1999. doi:10.1007/s00262-020-02593-w
61. Tauriello DVF, Palomo-Ponce S, Stork D, et al. TGFbeta drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature.* 2018;554(7693):538-543. doi:10.1038/nature25492
62. Lim YW, Coles GL, Sandhu SK, Johnson DS, Adler AS, Stone EL. Single-cell transcriptomics reveals the effect of PD-L1/TGF-beta blockade on the tumor microenvironment. *BMC Biol.* 2021;19(1):107. doi:10.1186/s12915-021-01034-z
63. Lee BY, Timpson P, Horvath LG, Daly RJ. FAK signaling in human cancer as a target for therapeutics. *Pharmacol Ther.* 2015;146:132-149. doi:10.1016/j.pharmthera.2014.10.001
64. de Jonge MJA, Steeghs N, Lolkema MP, et al. Phase I study of BI 853520, an inhibitor of focal adhesion kinase, in patients with advanced or metastatic nonhematologic malignancies. *Target Oncol.* 2019;14(1):43-55. doi:10.1007/s11523-018-00617-1
65. Chen B, Yang M, Li K, et al. Immune-related genes and gene sets for predicting the response to anti-programmed death 1 therapy in patients with primary or metastatic non-small cell lung cancer. *Oncol Lett.* 2021;22(1):540. doi:10.3892/ol.2021.12801

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Ahluwalia P, Mondal AK, Ahluwalia M, et al. Clinical and molecular assessment of an onco-immune signature with prognostic significance in patients with colorectal cancer. *Cancer Med.* 2022;11:1573–1586. doi: [10.1002/cam4.4568](https://doi.org/10.1002/cam4.4568)