BMC Cancer

RESEARCH Open Access

Identifying distinct prognostic and predictive contributions of tumor epithelium versus tumor microenvironment in colorectal cancer

Mingli Yang[1*], Michael V. Nebozhyn[2], Michael J. Schell[3], Nishant Gandhi[4], Lance Pflieger[5], Andrey Loboda[2], W. Jack Pledger[6,7], Ramani Soundararajan[1], Michelle Maurin[1], Heiman Wang[1], Jetsen Rodriguez Silva[1], Ashley Alden[1], Domenico Coppola[8,9], Andrew Elliott[4], George Sledge[4], Moh'd Khushman[10], Emil Lou[11], Sanjay Goel[12] and Timothy J. Yeatman[1,6,7*]

Abstract

Background Accumulating evidence has suggested that cancer progression and therapeutic response depend on both tumor epithelium (EPI) and tumor microenvironment (TME). However, the dependency of clinical outcomes on the tumor EPI vs. the TME has neither been clearly defined nor quantified.

Methods We classified 2373 colorectal cancer (CRC) tumors into the consensus molecular subtypes (CMS1-4) and generated the 10-gene TME^S and the 10-gene EPI^S signatures as the serendipitous derivatives of the most (positively vs. negatively) correlated genes of a highly-prognostic, ~500-gene signature we previously identified. Distinct TME vs. EPI cellular features of the signature genes were identified by CIBERSORT deconvolution and validated by scRNASEQ in an independent public dataset.

Results The TME^S signature was strongly associated with the immune/stromal TME-rich CMS1/CMS4 subtypes that portended worse survival, whereas the EPI^S signature was predominantly related to the TME-poor, epithelial CMS2/CMS3 classes that portended better survival. Multivariable Cox regression analysis against 29 TME-related signatures revealed that the TME^S signature was the most strikingly impacted by the "Cancer-associated fibroblasts" signature (HR: 10.87 vs. 0.13, both $P<0.0001$). Moreover, the TME^S score was strongly correlated with EMT, SRC activation and MEK inhibitor resistance in 2373 CRC tumors (Spearman $r=0.727$, 0.802, 0.824, respectively), which was validated in two independent CRC datasets ($n=626$ and $n=566$). By contrast, the EPI^S score was the dominant force in associating with longer progression free survival in cetuximab-treated metastatic CRC patients derived from two independent clinical trials (Logrank trend $P=0.0005/n=80$; $P=0.0013/n=44$). This finding was further validated in a large real-world clinico-genomics dataset with EGFR inhibitor therapy, which demonstrated that higher EPI^S scores were associated with increased overall survival (EGFRi, Logrank trend $P<0.0001/n=2343$) and time on treatment (cetuximab, $P=0.003/n=953$; panitumumab, $P<0.0001/n=1307$).

*Correspondence:
Mingli Yang
mingliyang@usf.edu
Timothy J. Yeatman
tyeatman@usf.edu

Full list of author information is available at the end of the article

Conclusions Here we identified a pair of new, distinct 10-gene signatures (the EPI^S vs. the TME^S) capable of distinguishing the cellular contribution of the tumor EPI vs. the TME in determining CRC prognosis and therapeutic outcomes. With targeted approaches emerging to address both tumor epithelial cells and the TME, the EPI^S vs. TME^S signature scores may have a novel biomarker role to permit optimization of CRC therapy by identifying sensitive vs. resistant subpopulations.

Keywords Colorectal cancer, Consensus molecular subtype, Clinical outcome, Tumor epithelium, Tumor microenvironment, Gene expression signature, EGFR inhibitor, Cetuximab, Panitumumab, MEK inhibitor

Background

It has been suggested that tumorigenesis and therapeutic response may depend not only on tumor epithelium (EPI), but also on the tumor microenvironment (TME), composed of a variety of non-cancerous immune and stromal cells [1–4]. Cancer progression and metastasis are thought to result from complex interactions between tumor cells and the TME [1–4]. Colorectal cancer (CRC) is a highly heterogeneous disease that has diverse genetic, molecular and clinical features associated with metastasis, prognosis and therapeutic outcomes [5, 6]. While both the biology of the CRC tumor and its associated TME are both likely contributory to therapeutic clinical outcomes, the cellular contribution of the tumor epithelial cell *versus* the resident TME towards drug sensitivity and resistance has neither been clearly defined nor quantified.

Heterogenous CRC has been classified into four distinct consensus molecular subtypes (CMS): CMS1 (MSI, immune), characterized by microsatellite instability (MSI) and immune activation, with worse survival after relapse (SAR); CMS2 (canonical), epithelial, WNT /MYC signaling; CMS3 (metabolic), epithelial, dysregulated metabolism; and CMS4 (mesenchymal), stromal infiltration, angiogenesis and TGFβ activation in association with worse overall survival (OS) and worse relapse free survival (RFS) [7]. The CMS1-4 subtypes have been applied to immune-classify CRC: CMS1, immune activated; CMS2, immune desert; CMS3, immune excluded; CMS4, immune inflamed [8, 9]. In this study, we classified 2373 human CRC tumors into the CMS1-4 subtypes with distinct, variable TME cellular features defined by CIBERSORT deconvolution analysis of bulk gene expression data. scRNASEQ derived from an independent dataset documented the precise cellular origin of signature transcripts. Here we present evidence clearly demonstrating and quantifying the distinct cellular contributions of the EPI vs. the TME in determining CRC prognosis and therapeutic outcomes. Moreover, these analyses have resulted in the generation of a pair of new, distinct, *predictive* 10-gene signature scores (the TME^S score vs. the EPI^S score)---biomarkers capable of quantifying the dependency of clinical outcomes on tumor epithelial cells vs. the TME, which may ultimately help optimize therapeutic strategies for CRC patients.

Methods

A flowchart of this study is shown in Supplementary Figure S1, which illustrates generation and analysis of the 10-gene TME^S versus the 10-gene EPI^S signature scores in the Merck-Moffitt CRC dataset and various other CRC datasets.

Merck-Moffitt CRC dataset ($n = 2373$)

The cohort of 2373 colorectal adenocarcinoma tumors, including 1571 *primary* lesions and 802 *metastatic* lesions (see Table 1), were obtained from adult patients treated through the Total Cancer Care initiative, which was created by the H. Lee Moffitt Cancer Center (Tampa, FL) [10]. All tumor tissues were collected by macrodisssection and snap freezing in liquid nitrogen within 15–20 min of extirpation. The Affymetrix gene expression data were obtained as described by the Merck group as a part of the large reference datasets representing > 25 different cancers to develop consensus signatures for pembrolizumab response [11, 12]. This study used only retrospective and de-identified clinical data for various bioinformatic analyses. Of note, we previously analyzed a subset of 468 tumors with global gene expression data, MSI status, and targeted gene sequencing of 1321 cancer-related genes including key driver genes (*BRAF*, *KRAS*, *APC* and *TP53*) [6]. Here we also performed correlation analyses of the TME^S score vs. the EPI^S signature score with these key driver genes in the 468 tumors.

Ethics statement

All the experiment protocol for involving human data was in accordance with the guidelines of national/international/institutional or Declaration of Helsinki. The tissue and clinical data were collected under the approval of the Institutional Review Board (IRB) of Moffitt Cancer Center with the informed written consent obtained from participating patients (the IRB No. is MCC14690).

CARIS LS real-world clinico-genomics dataset with longitudinal outcomes

Whole transcriptome sequencing (WTS)

Formalin fixed paraffin embedded (FFPE) CRC specimens (from 11,369 non-EGFRi and 2,343 EGFRi treated patients) underwent pathology review to measure percent tumor content and tumor size; a minimum of 10%

Table 1 Baseline characteristics of Moffitt-Merck 2373 CRC tumors

Characteristic	Total (*n* = 2373) (100%)	CMS1 (*n* = 329) (14%)	CMS2 (*n* = 689) (29%)	CMS3 (*n* = 357) (15%)	CMS4 (*n* = 745) (31%)	CMS-NA (*n* = 253) (11%)	Chi-Square *p*-Value
Age (yr)							
Mean	66	70	65	68	64	66	
Median	67	72	66	69	64	68	
Range	(21–88)	(21–87)	(24–88)	(26–87)	(28–87)	(31–87)	
Age subgroups (yr)	2373 (100%)	329 (100%)	689 (100%)	357 (100%)	745 (100%)	253 (100%)	
< 50	268 (11%)	25 (8%)	74 (11%)	29 (8%) ↓	**112 (15%)** ↑↑↑	28 (11%)	< 0.0001
50–59	452 (19%)	**44 (13%)** ↓↓	147 (21%)	63 (18%)	154 (20%)	44 (17%)	V = 0.101
60–69	664 (28%)	**68 (21%)** ↓↓	202 (29%)	98 (27%)	229 (31%) ↑	67 (26%)	
≥ 70	989 (42%)	**192 (58%)** ↑↑↑	266 (39%)	**167 (47%)** ↑	**250 (34%)** ↓↓↓	114 (45%)	
Sex	2372 (100%)	329 (100%)	688 (100%)	357 (100%)	745 (100%)	253 (100%)	
Male	1243 (52%)	**130 (40%)** ↓↓↓	**401 (58%)** ↑↑↑	175 (49%)	402 (54%)	135 (53%)	< 0.0001
Female	1129 (48%)	**199 (60%)** ↑↑↑	**287 (42%)** ↓↓↓	182 (51%)	343 (46%)	118 (47%)	V = 0.120
Race	2076 (100%)	290 (100%)	595 (100%)	304 (100%)	666 (100%)	221 (100%)	
Asian	12 (1%)	1 (0%)	3 (1%)	1 (0%)	4 (1%)	3 (1%)	.084[a]
Black or African American	142 (7%)	**12 (4%)** ↓	44 (7%)	23 (8%)	54 (8%)	9 (4%)	V = .063[a]
White	1900 (92%)	**273 (94%)** ↑	542 (91%)	278 (91%)	599 (90%)	208 (94%)	
Other	22 (1%)	4 (1%)	6 (1%)	2 (1%)	9 (1%)	1 (0%)	
Ethnicity	1772 (100%)	245 (100%)	513 (100%)	255 (100%)	573 (100%)	186 (100%)	
Hispanic or Latino	79 (4%)	10 (4%)	29 (6%)	15 (6%)	18 (3%)	7 (4%)	0.23
Not Hispanic or Latino	1693 (96%)	235 (96%)	484 (94%)	240 (94%)	555 (97%)	179 (96%)	V = 0.056
Stage at diagnosis	2156 (100%)	306 (100%)	640 (100%)	337 (100%)	641 (100%)	232 (100%)	
1 (0[b])	278 (13%)	40 (13%)	85 (13%)	**86 (26%)** ↑↑↑	**43 (7%)** ↓↓↓	24 (10%)	< 0.0001
2	491 (23%)	**111 (36%)** ↑↑↑	130 (20%)	79 (23%)	138 (22%)	**33 (14%)** ↓↓↓	V = 0.158
3	579 (27%)	73 (24%)	145 (23%) ↓↓	**107 (32%)** ↑	187 (29%)	67 (29%)	
4	808 (37%)	**82 (27%)** ↓↓↓	**280 (44%)** ↑↑↑	**65 (19%)** ↓↓↓	**273 (43%)** ↑↑	**108 (47%)** ↑↑	
Microsatellite status	2373 (100%)	329 (100%)	689 (100%)	357 (100%)	745 (100%)	253 (100%)	
MSI	237 (10%)	**165 (50%)** ↑↑↑	**0 (0%)** ↓↓↓	41 (11%)	**12 (2%)** ↓↓↓	19 (8%)	< 0.0001
MSS	2136 (90%)	**164 (50%)** ↓↓↓	**689 (100%)** ↑↑↑	316 (89%)	**733 (98%)** ↑↑↑	234 (92%)	V = 0.554
Specimen type	2373 (100%)	329 (100%)	689 (100%)	357 (100%)	745 (100%)	253 (100%)	
Primary	1571 (66%)	**256 (78%)** ↑↑↑	**412 (60%)** ↓↓↓	**318 (89%)** ↑↑↑	**448 (60%)** ↓↓↓	**137 (54%)** ↓↓↓	< 0.0001
Metastatic	802 (34%)	**73 (22%)** ↓↓↓	**277 (40%)** ↑↑↑	**39 (11%)** ↓↓↓	**297 (40%)** ↑↑↑	**116 (46%)** ↑↑↑	V = 0.247
Sidedness	1526 (100%)	243 (100%)	410 (100%)	296 (100%)	435 (100%)	142 (100%)	
Left	826 (54%)	**54 (22%)**↓↓↓	**300 (73%)** ↑↑↑	**132 (45%)** ↓↓↓	**275 (63%)** ↑↑↑	**65 (46%)** ↓	< 0.0001
Right	700 (46%)	**189 (78%)** ↑↑↑	**110 (27%)**↓↓↓	**164 (55%)** ↑↑↑	**160 (37%)** ↓↓↓	**77 (54%)** ↑	V = 0.352

Notes[a]comparing Black with White; [b]including 6 Stage 0 cases; the CMS subtypes were generated by the CMScaller

↑(↓) for $0.01 < p < 0.05$; ↑↑(↓↓) for $0.001 < p < 0.01$; ↑↑↑(↓↓↓) for $p < 0.001$ based on the standardized residual; Cramer's V (last column) shows the strength of the association

of tumor content in the area for microdissection was required to enable enrichment and extraction of tumor-specific RNA. The Qiagen RNA FFPE tissue extraction kit was used for RNA extraction, and the RNA quality and quantity were determined using the Agilent TapeStation. Biotinylated RNA baits were hybridized to the synthesized and purified cDNA targets, and the bait-target complexes were amplified in a post-capture PCR reaction. The Illumina NovaSeq 6500 was used to sequence the whole transcriptome from tumor samples to an average of 60 M reads. Raw data were demultiplexed by Illumina Dragen BioIT accelerator, trimmed, counted, removed of PCR-duplicates, and aligned to human reference genome hg19 by STAR aligner. TPM (Transcripts Per Million Molecules) were generated using the Salmon expression pipeline.

The TME^S/EPI^S genes were log transformed and z-scored. The TME^S/EPI^S signature scores were calculated as the average z-scores for all the TME^S/EPI^S genes for a given specimen.

Clinical outcomes

Real-world overall survival (OS) information was obtained from insurance claims data using Kaplan-Meier estimates. OS/EGFRi-OS was calculated from the time of tissue biopsy/EGFRi to last contact while time on EGFRi treatment (TOT) was calculated from the initiation of Cetuximab/Panitumumab to its termination

(Cetuximab-TOT/Panitumumab-TOT). *P* values were calculated using the log-rank test for trend.

Ethics statement

This study was conducted in accordance with the guidelines of the Declaration of Helsinki, Belmont Report, and U.S. Common Rule. In keeping with 45 CFR 46.101(b)(4) [], this study was performed using retrospective and de-identified clinical data. This study was thus considered institutional review board exempt, and no patient consent was necessary.

Pelka et al. scRNASEQ CRC dataset (*n* = 62)[13]

Recently, Pelka et al. used scRNASEQ to transcriptionally profile 371,223 cells from colorectal tumors and adjacent normal tissues of 62 CRC patients [13]. Here we accessed this public available scRNASEQ dataset on Human Colon Cancer Atlas (c295) - Single Cell Portal (broadinstitute.org) to analyze and visualize the single cell expression of the two distinct 10-gene signatures in epithelial tumor cells vs. immune/stromal cells.

Marisa et al. CRC dataset (*n* = 566)[14] and TCGA CRC dataset (*n* = 626)[5]

These two public CRC datasets were accessed as described for our previous validation analyses [15, 16]. Here 19 normal samples were removed from the Marisa dataset (*n* = 585), whereas 51 normal samples were removed from the TCGA dataset (*n* = 677). We used the Marisa (*n* = 566) and TCGA (*n* = 626) CRC datasets to independently validate the distinct gene expression/score/CMS correlations in relation to CRC outcomes and resistance to targeted therapy. In addition, we also performed correlation analyses of the TME^S score vs. the EPI^S signature score with *BRAF(V600E)* and *KRAS* mutations reported in the Marisa et al. CRC dataset [14].

Merck (PN04, *n* = 44)[17] and BMS (Khambata-Ford, *n* = 80)[18] (cetuximab) clinical trial datasets

The PN04 dataset had 44 WT *KRAS* CRC samples selected from the control arm (cetuximab + irinotecan) of a Merck prospective clinical trial (MK0646) [17], whereas the public available Khambata-Ford dataset was from a BMS trial including 80 cetuximab-treated CRC patient samples among which 43 had WT *KRAS* [18]. Here we used their Affymetrix data to examine the ability of the 10-gene EPI^S (vs. 10-gene TME^S) signature score in predicting progression free survival (PFS) in cetuximab-treated metastatic patients.

Gene expression signature scores

The gene expression signature scores were calculated for each tumor as we previously described [15, 16, 19, 20]. Briefly, a score was computed for each of the signatures as the arithmetic mean of all probesets corresponding to gene symbols present in the corresponding gene signature. Lists of the signature genes used here include previously described ΔPC1.EMT and EMT scores [20], SRC activation [15, 21], 13-gene MEKi resistance and 18-gene MEK pathway activation [15, 22]. In addition, we also computed the gene signature scores derived from the iCMS2 and iCMS3 signatures reported by Joanito et al. [23] and the 29 TME-related functional gene expression signatures (Fges) reported by Bagaev et al. [24], respectively.

Statistical methods

CMS classification

the Merck-Moffitt (*n* = 2373), TCGA (*n* = 626) and Marisa (*n* = 566) CRC tumors were classified by the CMScaller reported by Eide et al. [25] as previously described [15]. The CMS1*, CMS2*, CMS3* and CMS4* scores were generated for each of these CRC tumors as previously described [15]. The CMS1-4* scores were designated to measure the propensity of a tumor to fall into CMS1, CMS2, CMS2 and CMS4 classes, respectively [15]. Moreover, the Merck-Moffitt CRC tumors were also classified by the random forest (rf) and single-sample predictor (ssp.), the original CMS classifiers reported by Guinney et al. [7], respectively. Of note, the CMScaller is a cancer cell-intrinsic CMS classifier that could recapitulate the CMS subtypes in both in vitro and in vivo models and also performed well in primary tumors [25]. Our analysis in the Merck-Moffitt tumors also showed that the CMS subtypes were well correlated among the three CMS classifiers (CMScaller, the rf and the ssp.) (as shown later). Since the CMScaller could classify significantly more tumors into the CMS1-4 subtypes (*n* = 2120) than the rf (*n* = 1646) and the ssp. (*n* = 1766), the CMS subtypes generated by the CMScaller were primarily presented in this study.

PDS classification

the Merck-Moffitt CRC tumors (*n* = 2373) were also classified by the pathway-derived subtypes (PDS1-3) as reported by Malla et al. [26].

CIBERSORT is a method for enumeration of cell subsets to characterize cell composition of complex tissues from their gene expression profiles [27]. We performed a CIBERSORT deconvolution analysis of the global expression data on the Merck-Moffitt 2373 CRC tumors similarly as reported [27].

Statistical analyses, including Kaplan Meier (KM) survival, Spearman correlation, heatmap, Welch's *t* Test, Cramer's V Test, and CMS and PDS classifications were performed using GraphPad Prism 10.0.2, SAS 9.4 and R version 4.3. The statistical tests with an $\alpha = 0.05$ were chosen as the significance level. In addition, for the Welch *t*

test in comparison among multiple subgroups, adjusted *p* values after adjustments for multi-comparisons by Holm-Bonferroni method [28] were indicated. All tests were two-sided unless indicated otherwise.

Results

The highly prognostic ΔPC1.EMT score was strongly associated with CMS1 and CMS4 subtypes independent of age, sex, race, stages, MSI, metastasis and sidedness in 2373 CRCs

We first classified Merck-Moffitt 2373 CRC tumors by the CMS subtypes using the CMScaller [25]. Among 2373 tumor samples analyzed, CMS1-4 subtypes could be determined in 2120 samples whereas 253 tumors were not applicable to be classified into any CMS1-4 subtypes (CMS-NA). The frequencies of CMS subtypes were as follows: CMS1 ($n=329$; 14%), CMS2 ($n=689$, 29%), CMS3 ($n=357$; 15%), and CMS4 ($n=745$; 31%) as well as CMS-NA ($n=253$; 11%). Various baseline phenotypic characteristics and related statistics among CMS subtypes are listed in Table 1. Here the *p* values based on standardized residual were used to compare the subgroups of the baseline phenotypic characteristics in each CMS subtype. The Cramer's V (last column) shows the strength of their association with the CMS classes. *It is noteworthy that the baseline characteristics of CMS1 and CMS3 tumors were generally shared, as were the CMS2 and CMS4 tumors.* Notably, among patients identified with "Race and Ethnicity" information, 7% of patients were "Black or African American" and appeared to have lower CMS1 tumors (4%, $P<0.05$).

We previously reported an epithelial-to-mesenchymal transition (EMT)-based gene expression signature score that was highly-correlated to the PC1 (the first principal component) signature score in human CRC tumors [19]. We then generated a composite signature score (ΔPC1.EMT) by subtracting the EMT score from the PC1 score [20]. As a result, the ΔPC1.EMT score demonstrated a dramatically improved capacity to predict metastasis and clinical outcomes over its parent scores (either PC1 or EMT) and ten other published scores [20]. However, the biology underneath this composite score is still poorly understood. To better understand the biology underpinning this complex ~ 500 gene score, here we re-analyzed the ΔPC1.EMT score in the context of various baseline characteristics and the CMS classes. Here the Welch's *t* test was used for score comparison. Higher ΔPC1.EMT scores were significantly ($P<0.0001$) associated with higher stages (IV > III > II > I) tumors (including both primary and metastatic) and metastatic tumors themselves, and to a lesser extent, with MSI tumors (Fig. 1A-C). The Kaplan Meier (KM) survival analysis was also performed by ΔPC1.EMT score quartiles, which confirmed the score to be highly prognostic (Fig. 1A-C). Moreover, the ΔPC1.EMT scores were remarkably higher in the CMS1/CMS4 than CMS2/CMS3 classes when compared across all stages, individual stages (I, II, III, IV), primary and metastatic tumors, or MSI and MSS tumors (Fig. 1D-F). In addition, among the subgroups of age, sex, race and sidedness, higher scores appeared to associate with younger patients of < 50 year (all stages considered) (see Additional File 1-Fig. S2). Again, strong associations with the CMS1/CMS4 (vs. CMS2/CMS3) subtypes were observed in all the subgroups tested (Fig. S2). Taken together, the highly prognostic ΔPC1.EMT score was strongly associated with CMS1 and CMS4 subtypes independent of age, sex, race, stages, MSI, metastasis and sidedness in 2373 CRCs. *This strongly suggests that the biology underpinning the ΔPC1.EMT score may be highly related to the distinct cellular features of the CMS1/CMS4 vs. CMS2/CMS3 classes.*

While ΔPC1.EMT's top 10 positively-correlated (POS) genes were predominantly associated with the CMS1/CMS4 subtypes that portended worse survival, its top 10 negatively-correlated (NEG) genes were strongly related to the CMS2/CMS3 subtypes that portended better survival

The composite construction of the ~ 500-gene ΔPC1.EMT signature score (a combination of 125 PC1-UP genes, 120 PC1-DOWN genes, 150 EMT-UP genes and 162 EMT-DOWN genes) [20], made it difficult to dissect its complex biology. Thus, we elected to further interrogate ΔPC1.EMT's most-correlated genes with known biological functions. From an analysis of six independent CRC datasets, we previously conscripted 10 top-ranked, *positively*-correlated (POS) genes (*CD109*, *AHNAK2*, *GAS1*, *PRKCDBP (i.e. CAVIN3)*, *MEIS2*, *NXN*, *GFPT2*, *PMP22*, *WWTR1*, *PTRF (i.e. CAVIN1)*) (adj $P<0.0001$) and 10 top-ranked, *negatively*-correlated (NEG) genes (*CDX1*, *CDX2*, *C10orf99*, *DDC*, *GPA33*, *FAM84A (i.e. LRATD1)*, *NR1I2*, *MYB*, *C2orf89 (i.e.TRABD2A)*, *EPHB2*) (adj $P<0.0001$) [20]. Of note, these 20 top genes are mostly related to cancer and metastasis [20]. Remarkably, all top 10 POS genes had significantly higher gene expression in CMS1 and CMS4 tumors (Welch's t test $P<0.0001$), whereas each of the 10 top NEG genes had significantly higher gene expression in CMS2 and CMS3 tumors ($P<0.0001$) (Fig. 2A, B). This prompted us to generate a 10-gene POS *signature score* and a 10-gene NEG *signature score* (Fig. 2C).

In addition, we also generated the CMS1-4 subtypes by the random forest (rf) and single-sample predictor (ssp.), the original CMS classifiers reported by Guinney et al. [7], respectively (Additional File 1-Fig. S3). Notably, the degree of Spearman correlation between the CMScaller and the rf or between the CMScaller and the ssp. were the same as between the two original classifiers (rf and ssp.) (Fig. S3A), validating the use of the CMScaller in

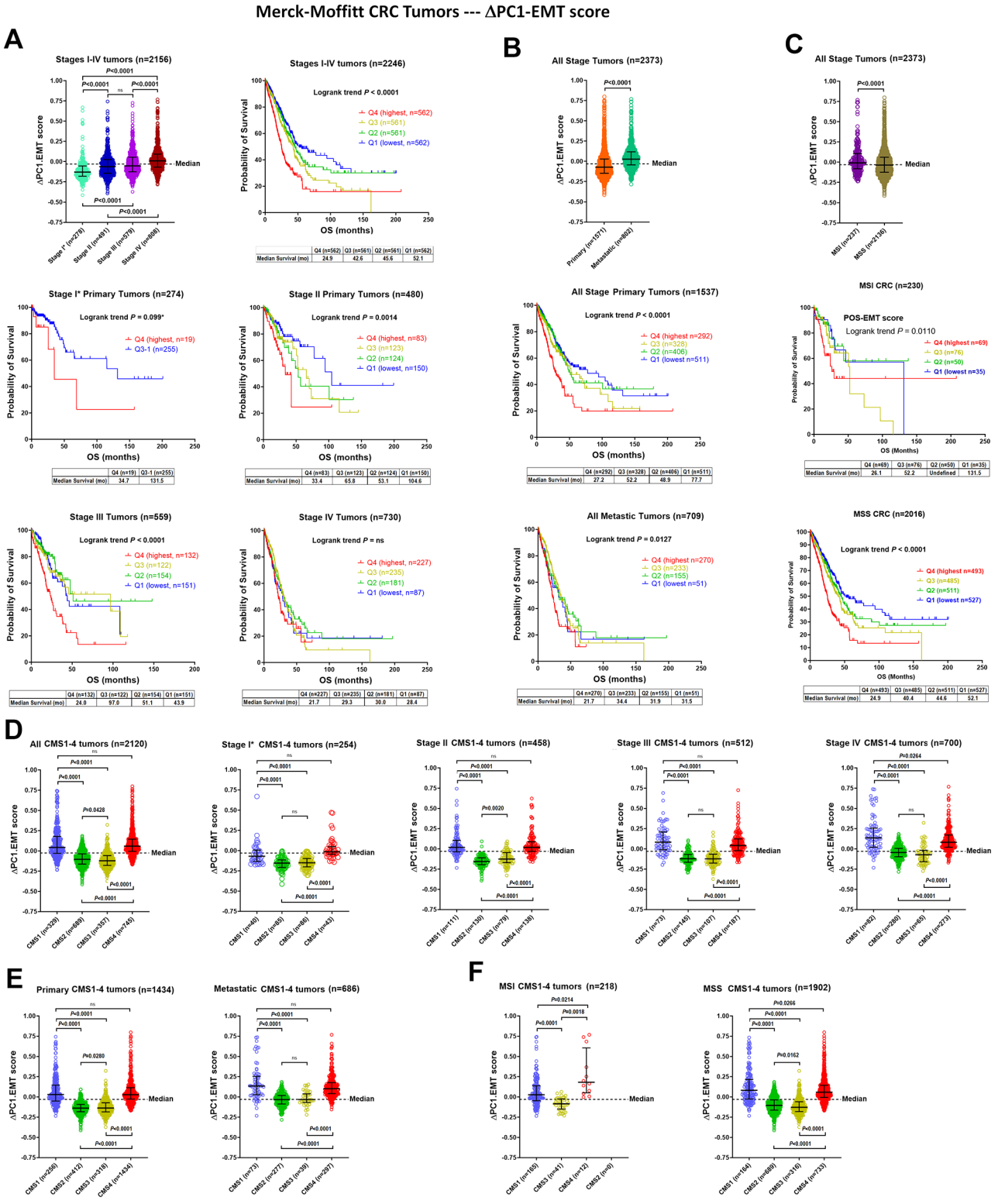


Fig. 1 The highly prognostic ΔPC1.EMT score was strongly associated with both CMS1 and CMS4 subtypes independent of stages, primary/metastatic tumor types, and MSI/MSS status in the Merck-Moffitt CRCs. **A**-**C**. The ΔPC1.EMT score comparison and Kaplan-Meier (KM) survival analyses by the score quartiles (Q1-Q4) in Merck-Moffitt CRCs that had corresponding overall survival (OS) data. **D**-**F**. The ΔPC1.EMT score comparison among the CMS1 vs. CMS2 vs. CMS 3 vs. CMS4 subtypes. Here the CMS1-4 subtypes were generated by CMScaller. **A**,**D**. all stages and individual stages (Stage I* (including 6 Stage 0 cases), Stage II, Stage III and Stage IV); **B**,**E**. primary and metastatic tumors; **C**,**F**. MSI and MSS tumors. Bars represent Median with interquartile range. Adjusted *P* values are shown for two-tailed Welch's t test after adjustments for multi-comparisons by Holm-Bonferroni method [28]. See Table 1 for detailed description of the baseline phenotypic characteristics

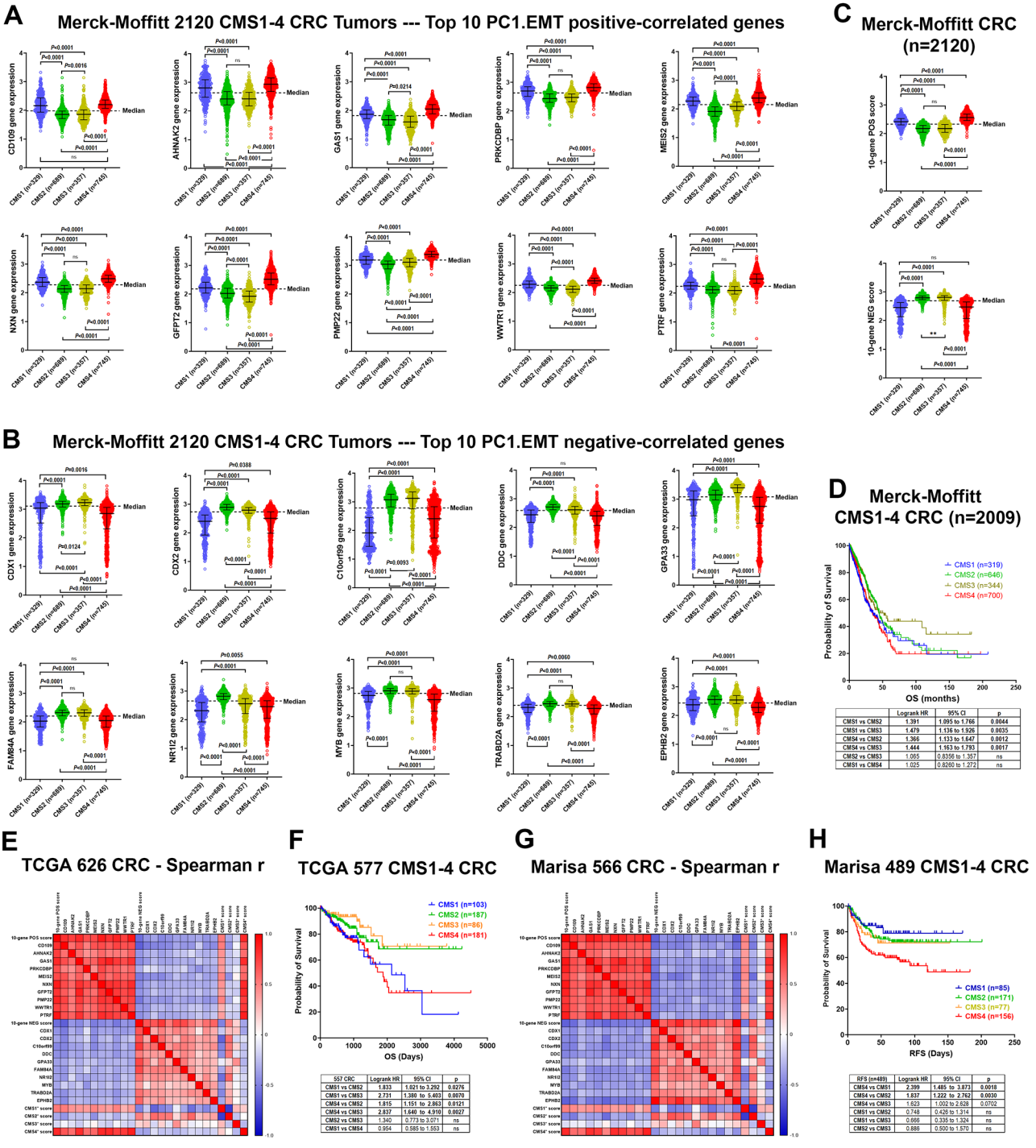


Fig. 2 The ΔPC1.EMT's top 10 positively-correlated (POS) genes (vs. top 10 negatively-correlated (NEG) genes) were strongly correlated with the CMS1 and CMS4 (vs. the CMS2 and CMS3) subtypes that portended worse survival. Comparison of gene expression of (**A**) the ΔPC1.EMT's top 10 positively-correlated (POS) genes and (**B**) top 10 negatively-correlated (NEG) genes across the CMS1-4 subtypes. (**C**) Comparison of the 10-gene POS and 10-gene NEG signature scores across the CMS1-4 subtypes. Here the CMS1-4 subtypes were generated by CMScaller. Bars represent Median with interquartile range. Adjusted *P* values are shown for two-tailed Welch's t test after adjustments for multi-comparisons by Holm-Bonferroni method [28]. Spearman correlation of the 10 POS genes/score and 10 NEG genes/score with the CMS1*, CMS2*, CMS3* and CMS4* scores in (**E**) the TCGA ($n=626$) and (**G**) the Marisa ($n=566$) CRC datasets, respectively. The CMS1*, CMS2*, CMS3* and CMS4* scores were designated to measure the propensity of a tumor to fall into CMS1, CMS2, CMS2 and CMS4 classes, respectively. The Kaplan-Meier (KM) survival analyses among CMS1, CMS2, CMS3 and CMS4 subtypes in (**D**) Merck-Moffitt ($n=2009$, OS), (**F**) TCGA ($n=577$, OS) and (**H**) Marisa ($n=489$, RFS) tumors, respectively

the primary tumors. Similar distinct association of the 10-gene POS signature score vs. the 10-gene NEG signature score with the CMS1-4 subtypes was also obtained by the rf and ssp. classifiers (Fig. S3B, C). Moreover, these distinct associations were independently validated in the TCGA ($n = 626$) [5] and Marisa ($n = 566$) [14] CRC datasets (Fig. 2E, G).

Furthermore, the KM analysis among the CMS1-4 classes revealed that both CMS1 and CMS4 classes portended worse OS in the Merck-Moffitt tumors (Fig. 2D (the CMScaller), Fig. S3D (the rf)). More distinct OS differences between CMS1/CMS4 vs. CMS2/CMS3 tumors were noted in the TCGA dataset (Fig. 2F) whereas worse RFS was only seen for CMS4 in the Marisa dataset (Fig. 2H). *These data suggest that the 10 POS (vs. the 10 NEG) signature genes/score were strongly associated with the CMS1/CMS4 (rather than CMS2/CMS3) subtypes, which portended worse survival.*

CIBERSORT Deconvolution cell type analysis confirmed that the POS vs. the NEG signature genes/scores were differentially associated with variable immune/stromal cellular features

To investigate whether the POS vs. the NEG genes/scores might capture the distinct cellular features associated with the CMS subtypes, we performed CIBERSORT deconvolution analysis in Merck-Moffitt 2373 CRC tumors (Fig. 3). A few key observations were noted:

(1) For the majority of CIBERSORT-defined immune cell populations, similar correlation patterns were seen between CMS1 and CMS4, or between CMS2 and CMS3. While CMS1*/CMS4* scores had negative correlations, the CMS2*/CMS3* scores had positive or no correlations with the cell scores of "B cells naïve", "Plasma cells", "T cells CD4 naïve", "T cells CD4 memory resting", "T cells CD4 memory activated", "T cells follicular helper", "T cells regulatory (Tregs)", "NK cells resting", "Mast cells resting", "Mast cells activated" and "Eosinophils". In addition, while the CMS2*/CMS3* scores had negative correlations with "Monocytes", "Macrophages M2", "Dendritic cells resting", and "mMDSC", CMS1*/CMS4* scores had positive correlations with them.

(2) Distinct CMS-associated correlations were also observed. (i) CMS1 was positively correlated with immune "active" cell populations including "B cells memory", "T cells CD8", "T cells gamma delta", "NK cells activated", "Macrophages M1" and "Dendritic cells activated". (ii) CMS2 and CMS3 tended to correlate with immune "inactive" cells including "B cells naïve", "NK cells resting" and "Macrophages M0". (iii) CMS4 was strongly correlated with "stromal" and "tumor" cells.

(3) The POS genes/score tended to correlate with the CMS1/CMS4-related immune/stromal cell populations. On the other hand, the NEG genes/score tended to correlate with the CMS2/CMS3-related immune inactive/poor cell populations.

(4) The 10-gene POS score was correlated stronger with "CMS4" than "CMS1", suggesting more preferential association with stromal-related immune-suppressive features.

Analysis of an independent ScRNASEQ dataset validated that the 10 POS genes were principally TME-associated genes and the 10 NEG genes were predominantly EPI-associated genes

To determine the *precise* cellular origin of the gene transcripts for the POS/NEG signatures, we assessed their expression levels at single cell resolution using a public colon cancer scRNAseq dataset (Pelka et al. $n = 62$) [13] (Fig. 4A-E). Strikingly, all 10 NEG signature genes were consistently and predominantly expressed in the epithelial tumor cells and epithelial normal mucosa, whereas each of the 10 POS signature genes were preferentially expressed in various stromal/immune TME cells (Fig. 4D, E). For example, all of the 10 POS genes were strongly expressed in the fibroblast cells, whereas 9 of the 10 POS genes (except *GFPT2*) had abundant expression in endothelial cells. Of note, *PMP22* showed strong expression in the macrophages. Thus, the 10 POS genes were preferentially expressed by the stromal (and to a lesser extent, immune) TME. In striking contrast, each of the 10 NEG genes was predominantly expressed in epithelial (EPI) tumor cells. *The distinct EPI vs. TME expression patterns led us to rename the NEG signature as "EPI^S" and the POS signature as "TME^S".*

The TME^S vs. the EPI^S genes, and their respective 10-gene signature scores, were distinctly correlated with BRAF and APC mutations

We previously carried out targeted DNA sequencing of 1321 cancer-related genes (including key driver genes such as *APC*, *TP53*, *KRAS* and *BRAF*) and MSI analysis on a set of 468 clinically-characterized, sporadic, colorectal tumors [6]. Here we assessed correlation of key driver mutations with the TME^S vs. the EPI^S genes/scores on the 468 tumors (Additional File 1-Fig. S4A). The TME^S score/genes were positively correlated with *BRAF* (V600E) and MSI and negatively correlated with *APC* truncating mutations, whereas the opposite was seen for the EPI^S score/genes. Of note, only the *BRAF*(V600E) and *APC* truncating mutations were considered as functional mutations for these two driver genes [6]. The distinct correlation with the *BRAF vs. APC* mutations was clearly illustrated by score comparison analyses against the prevalence of these mutations (Fig. S4B, C). Notably, the CMS correlations were seen for *BRAF (V600E)* with CMS1* score and for *APC* mutations with CMS2* score (Fig. S4A).

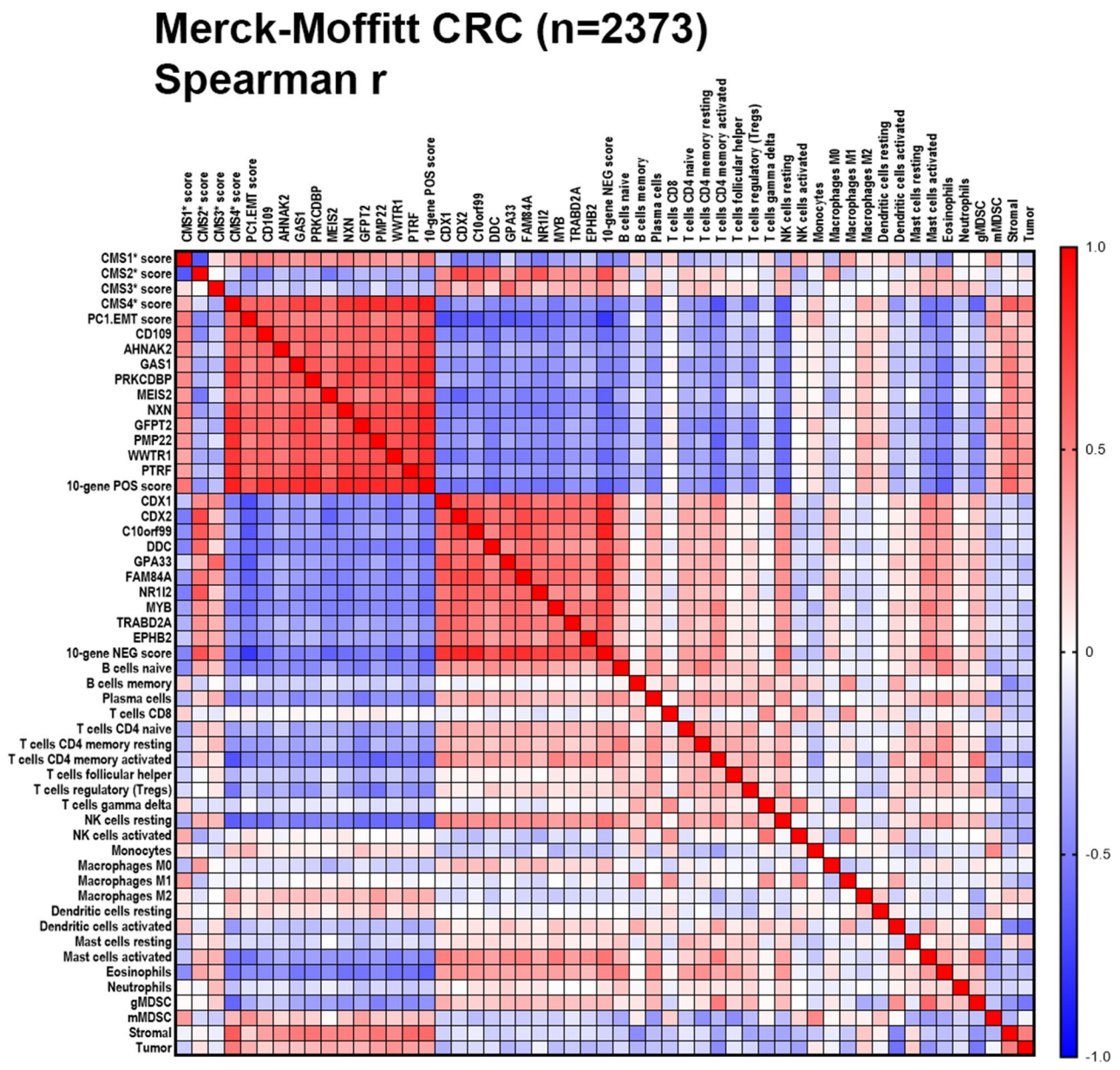


Fig. 3 CIBERSORT deconvolution and spatial transcriptomics revealed that the 10 POS genes/score vs. the 10 NEG genes/score were distinctly associated with the epithelial tumor vs. immune/stromal TME cellular features. Spearman correlation of the 10 POS genes/score and 10 NEG genes/score with the 20 CIBERSORT scores and the CMS1*, CMS2*, CMS3* and CMS4* scores in the Merck-Moffitt CRC tumors ($n=2373$). The 20 CIBERSORT cell scores were derived from the deconvolution analysis of Affymetrix gene expression data to measure the abundances of various immune cell populations as well as those of "stromal" cells and "tumor" cells, respectively

The TME^S and EPI^S genes/scores were associated with distinct prognostic contributions

We next performed the KM survival analyses to assess if the TME^S vs. the EPI^S genes/scores might be prognostic in the Merck-Moffitt 2246 tumors that had available OS data. Remarkably, higher expression of 9 of the 10 TME^S genes portended worse OS, whereas higher expression of all the 10 EPI^S genes were associated with better OS (Fig. 5A, B). Moreover, higher TME^S scores and lower EPI^S scores were highly prognostic in all stages, Stage I-III primary and MSS tumors, and to a lesser extent, predicted worse OS in MSI and metastatic tumors (Fig. 5C-G). These data suggest the TME^S vs. the EPI^S genes/scores were associated with distinct prognostic contributions. Of note, the 20-gene TME^S-EPI^S composite score (generated by subtracting the EPI^S score from the TME^S score) did not further improve prognostic prediction.

In addition, we also performed the KM survival analyses on the 10-gene TME^S score and the 10-gene EPI^S score in the Marisa CRC tumors that had corresponding

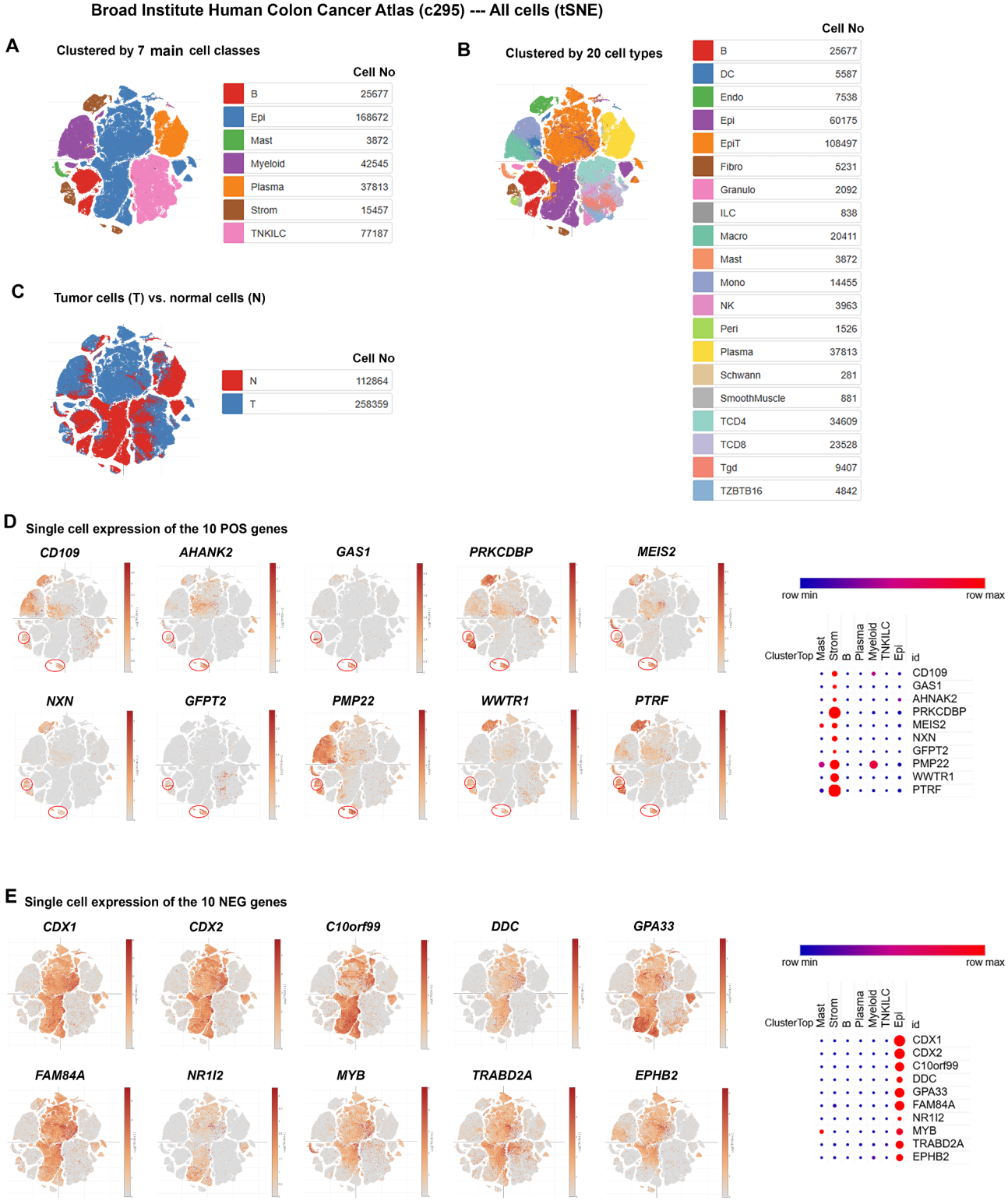


Fig. 4 Single cell expression analysis using an independent public scRNASEQ dataset validated that the 10 POS genes were principally TME-associated genes and the 10 NEG genes were predominantly EPI-associated genes. The scRNASEQ dataset ($n=62$) was reported to transcriptionally profiled 371,223 cells (258,359/112,864 cells, T/N) from colorectal tumors and adjacent normal tissues of 62 CRC patients [13]. The 10 POS genes and the 10 NEG genes were analyzed using the data of all cells on the Human Colon Cancer Atlas (c295)– Single Cell Portal (broadinstitute.org). The t-distributed stochastic neighbor embedding (t-SNE) of cell subsets are illustrated for (**A**) clustered by 7 major cell classes, (**B**) clustered by 20 cell subclasses and (**C**) clustered by tumor vs. normal cells. The single cell expression is shown for each of (**D**) the 10 POS genes and (**E**) the 10 NEG genes and both t-SNE and dot plots are shown. The POS gene expression in the tumor fibroblasts are highlighted by red circles in B, D

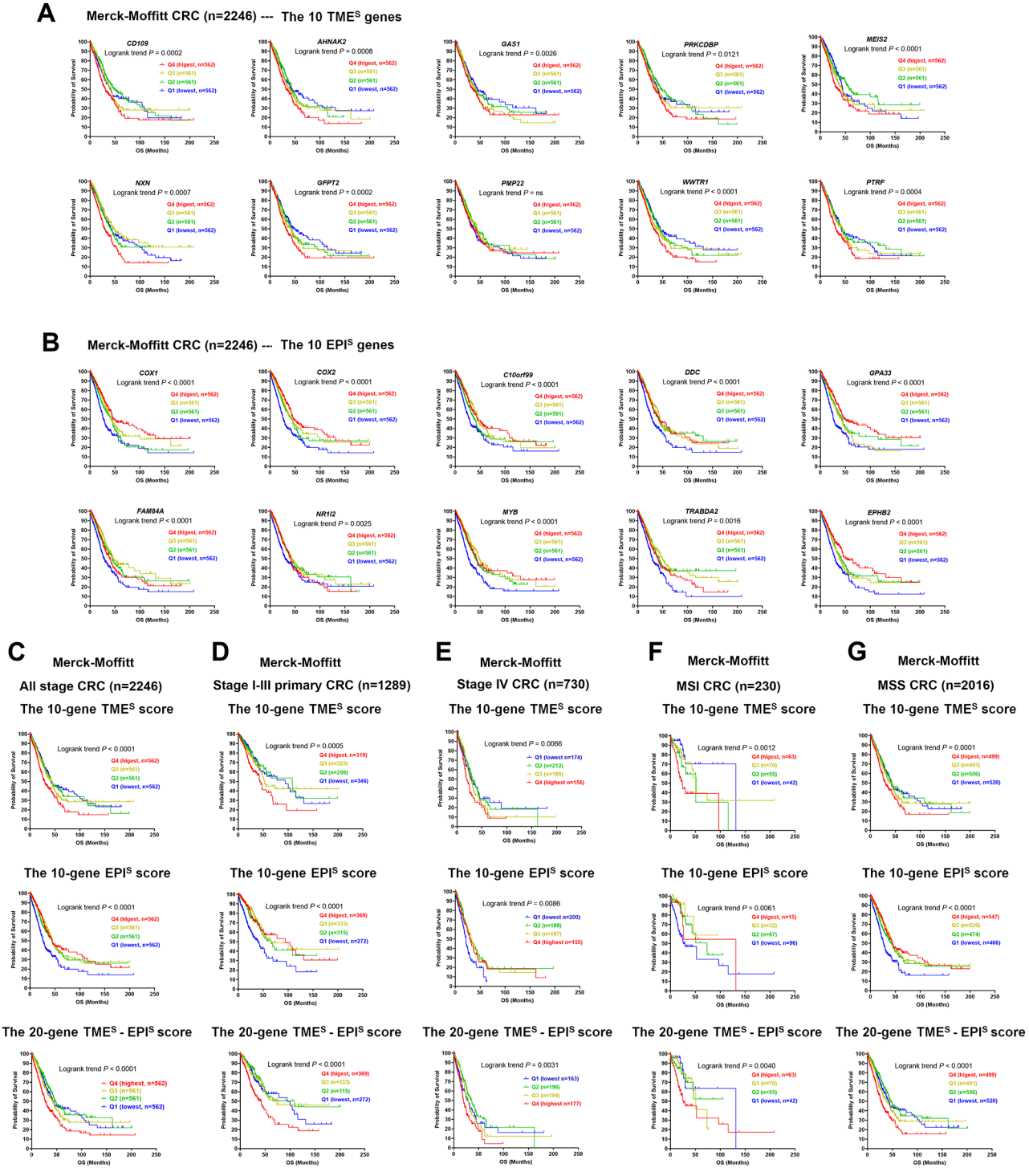


Fig. 5 The TME^S vs. the EPI^S genes and their respective signature scores portended distinct prognostic outcomes (worse vs. better OS). The Kaplan-Meier (KM) survival analyses by the gene expression quartiles (Q1-Q4) of (**A**) the 10 TME^S genes and (**B**) the 10 EPI^S genes in the Merck-Moffitt CRC tumors that had corresponding OS data ($n = 2246$). Notably, 9 of the 10 TME^S genes portended worse OS, whereas all the 10 EPI^S genes portended better OS. The KM analyses by the score quartiles (Q1-Q4) of the 10-gene TME^S score, the 10-gene EPI^S score and their 20-gene TME^S-EPI^S composite score in (**C**) all stage tumors ($n = 2246$), (**D**) Stage I-III primary tumors ($n = 1289$), (**E**) Stage IV tumors ($n = 730$), (**F**) MSI tumors ($n = 230$), (**G**) MSS tumors ($n = 2016$), respectively

RFS data ($n = 557$) (Additional File 1-Fig. S5). While the logrank trend p values were not significant, the plots show a trend for higher TME^S scores (Fig. S5A, higher quartiles Q4,Q3 vs. lower quartiles Q2,Q1) and lower EPI^S scores (Fig. S5B, the lowest quartile Q1 vs. other quartiles Q2-4) to be associated with worse RFS.

Multivariable Cox regression analyses against the 29 known TME-signatures revealed that while the TME^S and the EPI^S were independent prognostic signatures, the TME^S score was strikingly impacted by the "Cancer-associated fibroblasts" signature

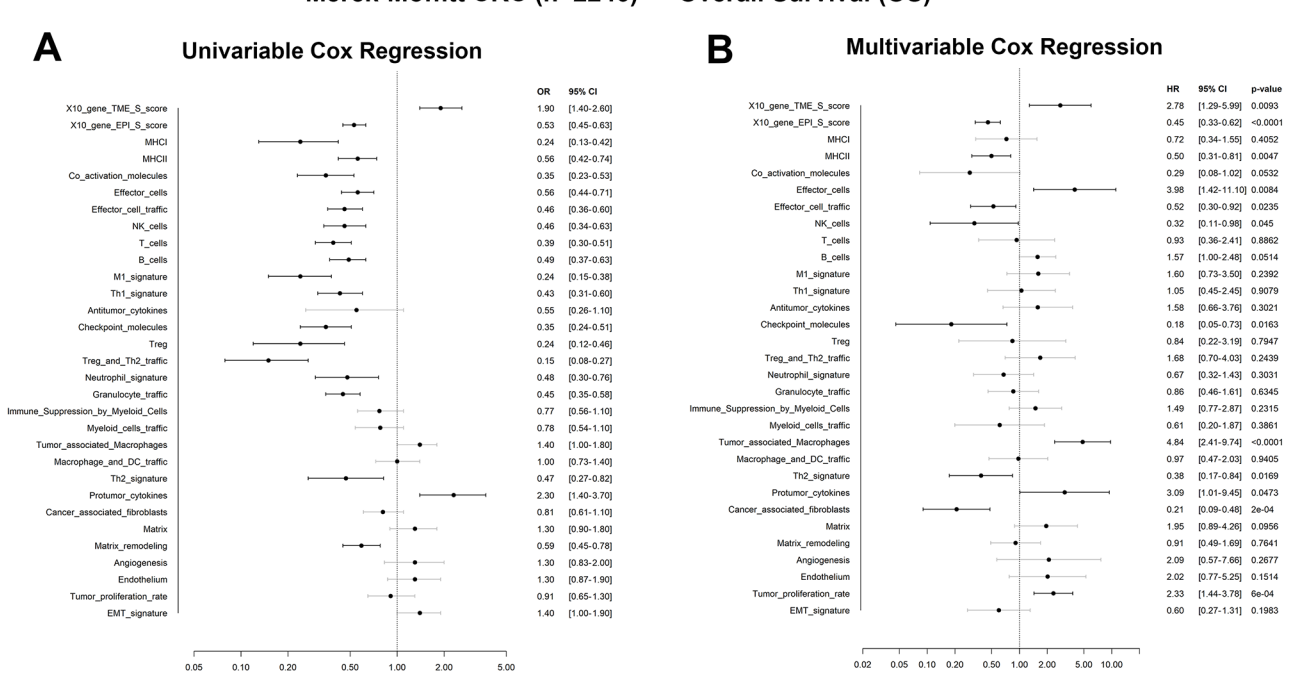
Bagaev et al. recently established a list of 29 TME gene expression signatures (Fges) representing the major functional components and immune, stromal and other cellular populations of a tumor for a pan-cancer TME subtyping analysis [24]. To determine whether the 10-gene TME^S and the 10-gene EPI^S signatures were independent prognostic signatures, we performed a multivariable Cox regression analysis against the 29 TME-Fges in the Merck-Moffitt CRC dataset (Fig. 6). Here we also performed a univariable Cox regression analysis of these signatures as a reference (Fig. 6A). Both the TME^S and the EPI^S were shown to be *independent prognostic signatures* when analyzed along with the 29 TME-Fges (Fig. 6B). Moreover, bivariable Cox analyses of the TME^S signature confirmed the TME^S signature as an independent prognostic signature when competing against each of the 29 Fges, respectively (Fig. 6C). Remarkably, the TME^S signature was strikingly impacted by the "Cancer-associated fibroblasts" (CAF) signature (HR: 10.87 [6.64–17.81] for TME^S vs. 0.13 [0.09–0.21] for CAF, both $P < 0.0001$) when compared to their univariable Cox OR (1.90 [1.40–2.60] for TME^S and 0.81 [0.61–1.10] for CAF). This is supported by the single cell gene expression data showing that all the 10 genes were predominantly expressed in stromal/fibroblast cell types (Fig. 4D). In addition, to a lesser degree, the TME^S signature was also considerably impacted by a number of other Fges (e.g. HR: 3.38 [2.46–4.65] for TME^S vs. 0.16 [0.10–0.25] for "Co-activation molecules", both $P < 0.0001$; HR: 3.36 [2.44–4.64] for TME^S vs. 0.18 [0.12–0.27] for "Checkpoint molecules", both $P < 0.0001$; HR: 3.61 [2.09–6.24], $P < 0.0001$ for TME^S vs. 0.45 [0.25–0.81] for "EMT-signature", $P = 0.0074$). These data suggest that the 10-gene TME^S signature could capture various stromal and immune TME components. These data are in agreement with the strong association of the TME^S signature with both immune CMS1 and stromal CMS4 subtypes that portended worse survival (Fig. 2).

The TME^S and the EPI^S scores were distinctly correlated with the iCMS3 signature score

Joanito et al. recently identified two intrinsic (epithelial) subtypes (iCMS2 and iCMS3) that refined CMS classification [23]. The iCMS3 subtype, that comprised MSI tumors and also one-third of MSS tumors, portended worse prognosis (vs. the iCMS2 subtype) [23]. Notably, the iCMS3 associated with the CMS4 had the worst prognosis [23]. Here we performed a heatmap correlation analysis on the iCMS2/iCMS3 signatures vs. the CMS subtypes in the Merck-Moffitt tumors ($n = 2373$) (Additional File 1-Fig. S6). The iCMS2 up genes were primarily correlated with the CMS2 subtype, supporting the most prominent feature of the iCMS2 [23] (Fig. S6A). However, the iCMS2 down genes, the iCMS3 up and down genes had less clear, "mixed" correlations with CMS1/CMS3/CMS4 (Fig. S6A). Further signature score correlation analysis shows that the 10-gene TMS^S score and the 10-gene EPI^S score had no correlation with the iCMS2 (up– down) signature score (Fig. S6B). By contrast, the TMS^S score and the EPI^S score were distinctly correlated with the iCMS3 (up– down) signature score (Spearman $r = 0.57$ for the TMS^S vs. $r = -0.77$ for the EPI^S) (Fig. S6B). Since the iCMS3 was associated with worse prognosis [23], this supports the distinct association of the TMS^S (worse) vs. the EPI^S (better) with CRC prognosis.

The TME^S and the EPI^S scores were distinctly associated with the PDS1-3 subtypes

Malla et al. recently performed pathway level subtyping of CRC and defined three pathway-derived subtypes (PDS1-3) [26]: PDS1—canonical/LGR5$^+$ stem-rich and highly proliferative, with good prognosis; PDS2—regenerative/ANXA1$^+$ stem-rich, with elevated stromal and immune TME lineages; PDS3—a slow-cycling subset of CMS2 tumors with reduced stem populations and increased differentiated lineages and with the worst prognosis in locally advance disease [26]. Here we classified the Merck-Moffitt tumors ($n = 2373$) into the PDS1 ($n = 817$), the PDS2 ($n = 608$) and the PDS3 ($n = 368$) as well as the mixed subtype ($n = 580$). We next performed the Welch's t test to compare that the 10-gene TMS^S score and the 10-gene EPI^S score among the PDS1-3 subtypes, respectively (Additional File 1-Fig. S7A). Significant score differences were seen for both the TMS^S and the EPI^S but with opposite directionalities (PDS2 > PDS3 > PDS1, all $P < 0.0001$ for the TMS^S vs. PDS2 < PDS3 < PDS1, all $P < 0.0001$ for the EPI^S). Notably, the highest TMS^S score vs. the lowest EPI^S score in the PDS2 "TME" subtype [26] agrees with the distinct association of the two scores with the TME. In addition, the lowest TMS^S score vs. the highest EPI^S score in the PDS1 (good prognosis) subtype [26] is also consistent with the distinct association of the two scores with CRC prognosis (see Fig. 5C-G). Of note,



C

Bivariable Cox Regression

1	HR	95% CI	p-value
x10_gene_TME_S_score	2.02	1.51-2.71	<0.0001
MHCI	0.21	0.12-0.37	<0.0001

2	HR	95% CI	p-value
x10_gene_TME_S_score	3.05	2.21-4.20	<0.0001
MHCII	0.35	0.26-0.48	<0.0001

3	HR	95% CI	p-value
x10_gene_TME_S_score	3.38	2.46-4.65	<0.0001
Co_activation_molecules	0.16	0.10-0.25	<0.0001

4	HR	95% CI	p-value
x10_gene_TME_S_score	2.36	1.75-3.19	<0.0001
Effector_cells	0.48	0.38-0.61	<0.0001

5	HR	95% CI	p-value
x10_gene_TME_S_score	2.52	1.88-3.38	<0.0001
Effector_cell_traffic	0.38	0.29-0.49	<0.0001

6	HR	95% CI	p-value
x10_gene_TME_S_score	2.29	1.69-3.09	<0.0001
NK_cells	0.39	0.28-0.53	<0.0001

7	HR	95% CI	p-value
x10_gene_TME_S_score	2.46	1.82-3.31	<0.0001
T_cells	0.33	0.25-0.43	<0.0001

8	HR	95% CI	p-value
x10_gene_TME_S_score	2.21	1.64-2.98	<0.0001
B cells	0.44	0.33-0.57	<0.0001

9	HR	95% CI	p-value
x10_gene_TME_S_score	2.07	1.55-2.76	<0.0001
M1_signature	0.21	0.13-0.34	<0.0001

10	HR	95% CI	p-value
x10_gene_TME_S_score	2.33	1.73-3.15	<0.0001
Th1_signature	0.36	0.26-0.50	<0.0001

11	HR	95% CI	p-value
x10_gene_TME_S_score	1.98	1.48-2.66	<0.0001
Antitumor_cytokines	0.45	0.21-0.95	0.0367

12	HR	95% CI	p-value
x10_gene_TME_S_score	3.36	2.24-4.64	<0.0001
Checkpoint_molecules	0.18	0.12-0.27	<0.0001

13	HR	95% CI	p-value
x10_gene_TME_S_score	2.9	2.11-3.99	<0.0001
Treg	0.09	0.05-0.19	<0.0001

14	HR	95% CI	p-value
x10_gene_TME_S_score	2.1	1.56-2.82	<0.0001
Treg_and_Th2_traffic	0.13	0.07-0.24	<0.0001

15	HR	95% CI	p-value
x10_gene_TME_S_score	2.27	1.66-3.09	<0.0001
Neutrophil_signature	0.35	0.22-0.57	<0.0001

16	HR	95% CI	p-value
x10_gene_TME_S_score	2.31	1.72-3.11	<0.0001
Granulocyte_traffic	0.39	0.30-0.51	<0.0001

17	HR	95% CI	p-value
x10_gene_TME_S_score	2.57	1.85-3.56	<0.0001
Immune_Supression_by_myeloid_cells	0.49	0.34-0.70	0.0001

18	HR	95% CI	p-value
x10_gene_TME_S_score	3.06	2.15-4.35	<0.0001
Myeloid_cells_traffic	0.36	0.24-0.56	<0.0001

19	HR	95% CI	p-value
x10_gene_TME_S_score	2.02	1.40-2.91	0.0002
Tumor_associated macrophages	0.92	0.64-1.31	0.6414

20	HR	95% CI	p-value
x10_gene_TME_S_score	2.77	1.92-3.99	<0.0001
Macrophage_and_DC_traffic	0.51	0.34-0.44	0.0013

21	HR	95% CI	p-value
x10_gene_TME_S_score	2.44	1.78-3.33	<0.0001
Th2_signature	0.27	0.15-0.49	<0.0001

22	HR	95% CI	p-value
x10_gene_TME_S_score	1.81	1.15-2.85	0.0098
Protumor_cytokines	1.12	0.55-2.30	0.7482

23	HR	95% CI	p-value
x10_gene_TME_S_score	10.87	6.64-17.81	<0.0001
Cancer_associated_fibroblasts	0.13	0.09-0.21	<0.0001

24	HR	95% CI	p-value
x10_gene_TME_S_score	2.83	1.88-4.26	<0.0001
Matrix	0.53	0.33-0.85	0.0083

25	HR	95% CI	p-value
x10_gene_TME_S_score	2.94	2.15-4.03	<0.0001
Matrix_remodeling	0.37	0.27-0.50	<0.0001

26	HR	95% CI	p-value
x10_gene_TME_S_score	2.85	1.89-4.29	<0.0001
Angiogensis	0.43	0.23-0.79	0.0072

27	HR	95% CI	p-value
x10_gene_TME_S_score	2.67	1.79-3.99	<0.0001
Endothelium	0.53	0.31-0.90	0.0185

28	HR	95% CI	p-value
x10_gene_TME_S_score	2.12	1.52-2.95	<0.0001
Tumor_proliferation_rate	1.29	0.89-1.88	0.1768

29	HR	95% CI	p-value
x10_gene_TME_S_score	3.61	2.09-6.24	<0.0001
EMT_signature	0.45	0.25-0.81	0.0074

Fig. 6 Cox proportional hazards regression analyses of the 10-gene TME^S score and the 10-gene EPI^S score, along with 29 TME-functional gene expression signatures (Fges), in the Merck-Moffitt CRC tumors that had corresponding OS data ($n = 2246$). **A**. Univariable Cox regression analysis. **B**. Multivariable Cox regression analysis. **C**. Bivariable Cox regression analyses of the 10-gene TME^S score against each of the 29 TME-Fges, respectively (total 29 pairs). OR, odd ratio; HR, hazard ratio. Here X10_gene_TME_S_score represents the 10-gene TME^S score and X10_gene_EPI_S_score represents the 10-gene EPI^S score

the KM survival analysis of the PDS1-3 subtypes in the Merck-Moffitt tumors shows that while not significant, the PDS1 tended to have better OS (Fig. S7B).

The TME^S score was strongly correlated with EMT, SRC activation and MEK inhibitor resistance

Over half of CRC tumors are driven by oncogenic RAS/RAF mutations, suggesting that CRC is hard-wired to RAS/RAF/MEK signaling [5, 6, 29]. *KRAS/BRAF*-mutated CRC has been targeted but therapeutic inhibitors are largely ineffective, likely due to complex resistance mechanisms that are not well understood [22, 30, 31]. We previously reported the SRC activation signature score was strongly correlated with the 13-gene MEKi-resistance signature score (that predicts drug resistance caused by "bypass" proliferation/survival signaling pathways) [15, 22]. We now identified that the TME^S signature score was highly correlated with the EMT, SRC activation and 13-gene MEKi-resistance signature scores in Merck-Moffitt 2373 CRC tumors (Spearman $r = 0.727$, 0.802, 0.824, respectively, all $P < 10^{-320}$, see Fig. 7A). By contrast, the EPI^S signature score was, to a lesser statistical degree, negatively-correlated with these scores (Fig. 7B). Notably, CMS4 (and to a lesser extent, CMS1) tumors were strongly associated with higher scores of EMT, SRC activation and MEKi-resistance, whereas CMS2 and CMS3 tumors were highly associated with their lower scores (Fig. 7A, B). Moreover, these findings were validated in the Merck-Moffitt Stage IV tumors (Additional File 1-Fig. S8) and the Marisa CRC dataset ($n = 566$) (Fig. 7C, D). Furthermore, the same distinct correlation patterns were also seen for all the TME^S vs. the EPI^S genes in multiple datasets (Additional File 1-Fig. S9-S11), suggesting a contributive role of the individual genes.

The EPI^S score was associated with longer progression free survival in cetuximab-treated metastatic CRC patients derived from two independent clinical trials

We next examined the potential of the EPI^S score vs. the TME^S score to predict outcomes in metastatic CRC patients treated with cetuximab in two independent clinical trial datasets (BMS (Khambata-Ford) [18], Merck (PN04) [17]). Higher EPI^S scores were highly associated with longer progression free survival (PFS) in the BMS dataset with all patients regardless of *KRAS* mutation status) ($n = 80$, Logrank trend $P = 0.0005$, median survival (days): 103.5 (Q4, highest) vs. 53 (Q1, lowest)) and in the subset of WT *KRAS* patients ($n = 43$, $P = 0.0062$, median survival (days): 135.0 (Q4) vs. 49.5 (Q1)) (Fig. 8A, B). The TME^S score, by itself, was not predictive of CTX sensitivity, and it did not add any predictive value to the EPI^S score when combined with it (20-gene EPI^S /TME^S score) (Fig. 8A, B). Of note, higher EPI^S scores were significantly associated with patients with complete/partial responses (CR/PR) and stable disease (SD) (Fig. 8C). The same predictive value was also demonstrated in the Merck-PN04 dataset with WT *KRAS* patients ($n = 44$, $P = 0.0013$, median survival (days): 208.0 (Q4) vs. 75.5 (Q1)) (Fig. 8D).

Establishing the association of EPI^S and TME^S scores with OS, EGFRi-OS and EGFRi-TOT in a large real world clinic-genomics dataset

Finally, using a large real-world dataset, we investigated the association of the EPI^S score and TME^S scores with clinical outcomes. The KM analysis showed that the EPI^S score modestly but very significantly (Logrank trend $P < 0.0001$) predicted better OS whereas the TME^S score was barely prognostic ($P = 0.0153$) on all CRC patients excluding those treated with EGFRi ($n = 11369$, Fig. 9A, E). We next investigated the association between the EPI^S and TME^S scores with survival from the initiation of EGFRi (OS) and time on treatment (TOT) as a surrogate for PFS in patients treated with cetuximab (CTX) or panitumumab (PMB). These treatments were delivered in all lines of therapy. We found that higher EPI^S scores were associated with increased OS ($n = 2343$, $P < 0.0001$) and TOT with EGFRi therapy (CTX, $n = 953$, $P = 0.003$; PMB, $n = 1307$, $P < 0.0001$) (Fig. 9B-D). On the other hand, higher TME^S scores were significantly, but to a lesser extent, associated with decreased OS ($P < 0.0001$), CTX-TOT ($P = 0.004$) and PMB-TOT ($P = 0.0242$) (Fig. 9F-H).

Discussion

Cancer progression and therapeutic response depend on both tumor epithelium (EPI) and their surrounding immune/stromal TME [1, 32, 33]. However, the differential contribution of cell type and relative geography of the EPI vs. the TME to CRC prognosis and therapeutic outcomes is not yet clearly defined. Here we report an analysis of 2373 human colorectal cancers, validated by multiple independent CRC datasets (Additional File 1-Fig. S1 for study flowchart), leading to development of a companion pair of distinct 10-gene signatures (TME^S vs. EPI^S) scores capable of distinguishing the cellular origin and quantitative contribution of epithelial tumor vs. the TME to both prognostic and therapeutic outcomes.

We deciphered the biology underpinning the complex ΔPC1.EMT score. The ΔPC1.EMT score and its most correlated genes (10 POS genes vs. 10 NEG genes) had distinct correlations with the *TME-rich* CMS1 and CMS4 (associated with worse survival) vs. the *TME-poor* (epithelial) CMS2 and CMS3 subtypes (associated with better survival) (Figs. 1 and 2, Additional File 1-Fig. S2). Notably, higher ΔPC1.EMT scores appeared to associate with younger patients of < 50 year (all stages considered) (Fig. S2). It is noteworthy that the CRC incidence

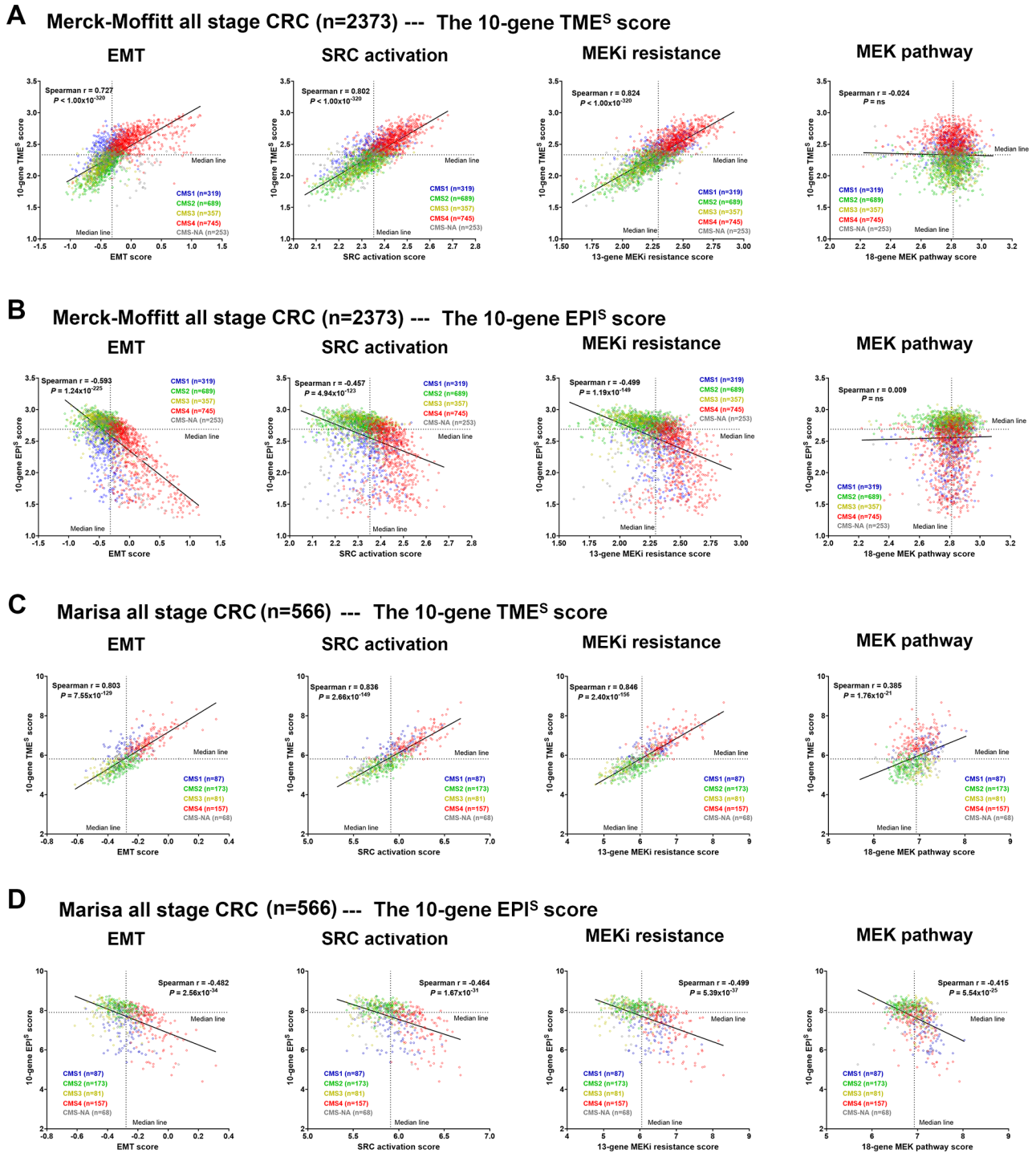


Fig. 7 The 10-gene TME^S signature score (vs. the 10-gene EPI^S signature score) was strongly correlated with the EMT, SRC activation and MEK inhibitor (MEKi) resistance signature scores in multiple CRC datasets. Scatter plots of (**A**) the 10-gene TME^S score and (**B**) the 10-gene EPI^S score *versus* the EMT, SRC activation, 13-gene MEKi resistance and 18-gene MEK pathway activation signature scores, respectively, in the Merck-Moffitt CRC tumors ($n=2373$). Spearman r and *P* values are shown. The CMS1-4 subtypes are indicated by red (CMS4) vs. orange (CMS3) vs. green (CMS2) vs. blue (CMS1) colors. (**C**,**D**) Similar scatter plots were made in an independent (Marisa) CRC dataset ($n=566$). Here the CMS1-4 subtypes were generated by CMScaller. Scatter plots were also made in the Merck-Moffitt Stage IV tumors ($n=808$) (see Additional File 1-Fig S8). Spearman correlation heatmaps of the individual TME^S and EPI^S genes with the EMT, SRC activation, 13-gene MEIk resistance and 18-gene MEK pathway activation signature scores in the Merck-Moffitt ($n=2373$), TCGA ($n=626$) and Marisa ($n=566$) CRC datasets are shown in Additional File 1-Figs S9-11, respectively

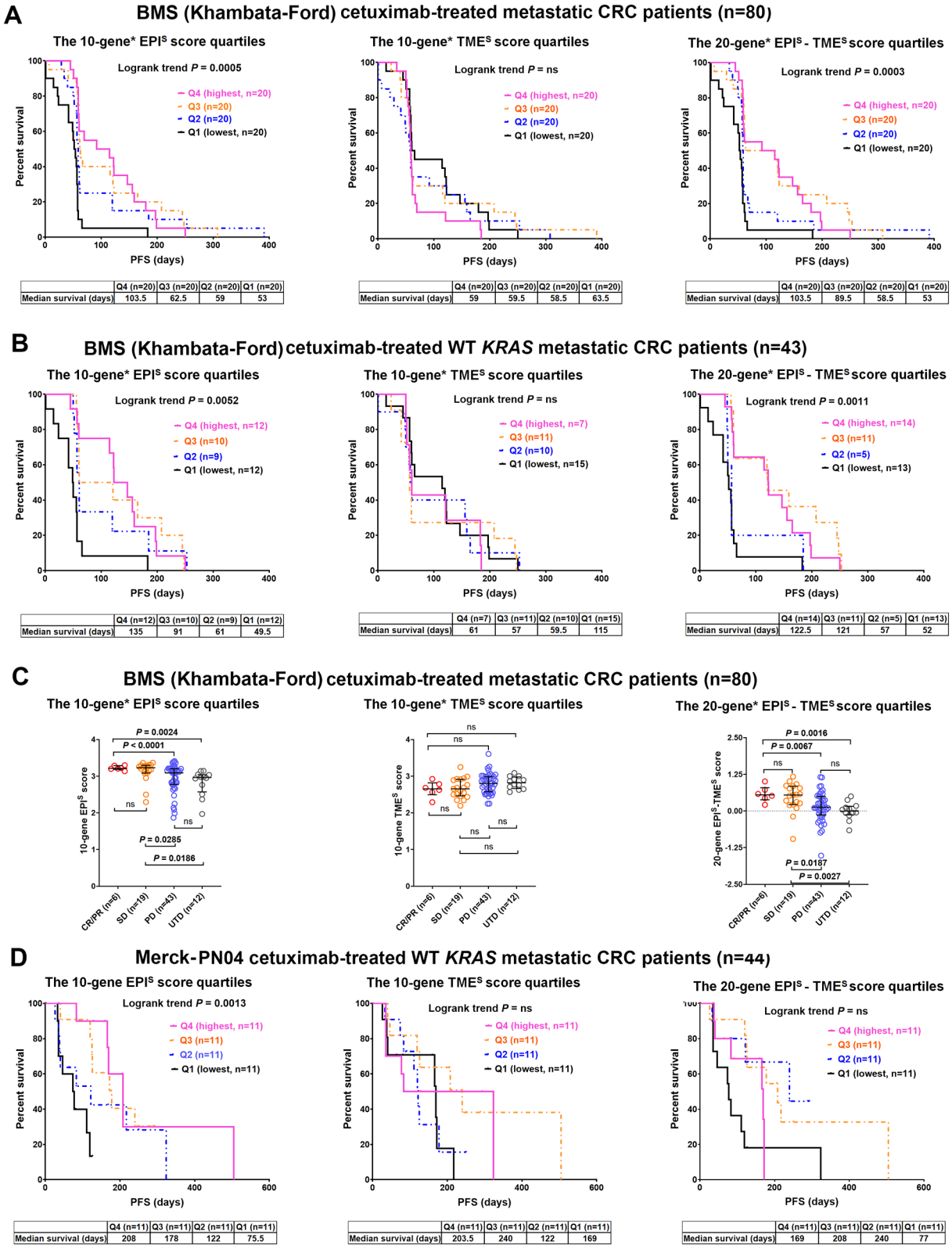


Fig. 8 (See legend on next page.)

(See figure on previous page.)

Fig. 8 Retrospective analysis of two independent clinical trial datasets showed that the 10-gene EPI^S signature score significantly predicted longer progression free survival (PFS) in cetuximab (CTX)-treated metastatic CRC patients. The Kaplan-Meier (KM) survival analyses of the 10-gene EPI^S, the 10-gene TME^S, and 20-gene EPI^S-TME^S signature scores in (**A**) the Khambata-Ford CTX-treated metastatic CRC patients ($n = 80$, all cases regardless of *KRAS* mutation status) [18] and (**B**) a subset of WT *KRAS* patients ($n = 43$) as well as (**D**) Merck-PN04 CTX-treated WT *KRAS* metastatic CRC patients [17]. **C**. Comparison of the 10-gene EPI^S, the 10-gene TME^S, or 20-gene TME^S-EPI^S scores among Khambata-Ford CR/PR (complete response/partial response), SD (stable disease), PD (progressed disease) and UTD (undetermined) patients. Bars represent Median with interquartile range. Unadjusted *P* values are for two-tailed Welch's t test; ns — not significant. Of note, those significant unadjusted *p* values remained significant after adjustments for multi-comparisons by Holm-Bonferroni method [28]. Note: *3 EPI^S genes (*C10orf99*, *FAM84A*, *TRABD2A*) and 1 TME^S gene (*CD109*) did not have probe values in the Khambata-Ford dataset so these 4 genes were excluded in the analyses

was increasing in young patients, which was higher for African Americans as reported by Ashktorab et al. [34]. Our analysis also shows that the "Black or African American" patients had lower CMS1 tumors (4%, $P < 0.05$) (see Table 1) but more CRC classification analyses are needed for African Americans [35]. The differential association of the 10 POS genes/score vs. the 10 NEG genes/score with variable immune/stromal TME cellular features was demonstrated by CIBERSORT deconvolution analysis (Fig. 3). Next, scRNASEQ analyses (Fig. 4) both confirmed that the 10 POS genes, that were strongly correlated with the CMS1/CMS4 subtypes, represented TME associated genes (TME^S). In striking contrast, the 10 NEG genes, that were strongly correlated with the CMS2/CMS3 subtypes, represented EPI associated genes (EPI^S).

Our data revealed that the TME^S genes/score predicted worse OS (Fig. 5). This is supported by the observation that all the 10 TME^S genes were principally expressed in the tumor stroma, primarily in the tumor fibroblasts (Fig. 4B, D). The tumor stroma can account for up to > 50% of tumor mass of CRC and its extent is predictive of worse survival [7, 36–38]. Cancer associated fibroblasts, the dominant cellular components of tumor stroma, were reported to promote tumor cell invasion and metastasis [32, 39]. The multivariable Cox regression analyses against the 29 TME gene expression signatures [24] in the Merck-Moffitt tumors show the HR of the TME^S score was the most strikingly impacted by the "Cancer associated fibroblasts" signature (Fig. 6C). The TME induces EMT in various types of tumor cells [40]. EMT is aberrantly activated in cancer as a major mechanism promoting invasion and metastasis and is also known to contribute to drug resistance [40]. We found that the 10 TME^S genes and their composite score were strongly correlated with the EMT score and the "mesenchymal" CMS4 subtype that portended worse survival [7] (Figs. 2 and 7, Additional File 1-Fig. S8-11).

Intriguingly, the 10 TME^S individual genes and their signature score were also strongly associated with the CMS1 (MSI, immune) subtype (Figs. 2 and 3, Fig. S2). The multivariable Cox analyses also show that the HR of the TME^S score was also considerably impacted by various "immune" signatures including the "Co-activation molecules" and "Checkpoint molecules" signatures that were associated with better OS (Fig. 6). These data suggest that the 10-gene TME^S signature could capture various stromal and immune TME components. The "MSI, immune" CMS1 subtype was associated with response to checkpoint inhibitor therapy (e.g. pembrolizumab) [9, 41]. However, despite its positive association with immunotherapy, the CMS1 subtype portended worse prognosis (OS) in the Merck-Moffitt CRC tumors ($n = 2246$) (Fig. 2). The association of CMS1 with worse prognosis was also recently reported by Chowdhury et al. when analyzed in all CRC ($n = 24{,}939$), primary or local tumors ($n = 14{,}153$) and distant metastatic tumors ($n = 10{,}776$) [41]. Why the CMS1 tumors were associated with worse prognosis is largely not clear [41]. The CMS1 tumors are less metastatic (see Table 1) and confer the best prognosis in early-stage CRCs but portend very poor outcomes at advanced stages [42, 43]. Moreover, *BRAF*(V600) mutation was the most frequently present in CMS1 [7]. *BRAF*(V600E) was reported as a strong negative prognostic marker of CRC [20, 44, 45] but the biology is poorly understood. Notably, we found that the TME^S genes and score were positively correlated with *BRAF* (V600) (Additional File 1-Fig. S4), suggesting a potential biological role of the TME in *BRAF* (V600)-mediated poor prognosis. Moreover, the TME^S score was strongly related to the stromal/fibroblast TME (Figs. 4 and 6) and the fibrotic subtypes were reported to associate with the non-responders in immunotherapy [24], suggesting that the TME^S score may not have association with immunotherapy response. On the other hand, our data revealed that the TME^S genes/score were strongly associated with SRC activation and MEK inhibitor resistance and (Fig. 7, Additional File 1-Fig. S8-11). Of note, the SRC oncogene is a non-receptor tyrosine kinase that mediates cell proliferation, survival, invasion and motility [46, 47] and is regulated by the TME [48].

By contrast, the 10-gene EPI^S genes and their composite score predicted better OS (Fig. 5). The EPI^S score was associated with longer PFS in cetuximab-treated metastatic CRC patients in two clinical trial datasets (Fig. 8). Furthermore, we independently validated the EPI^S score as mildly prognostic (better OS) but highly predictive of EGFRi outcomes (better OS and longer TOT) using a large real world CarisLS dataset (Fig. 9). Notably, the TME^S score was also prognostic and predictive (worse outcomes) but to a lesser degree (Fig. 9). In addition, the

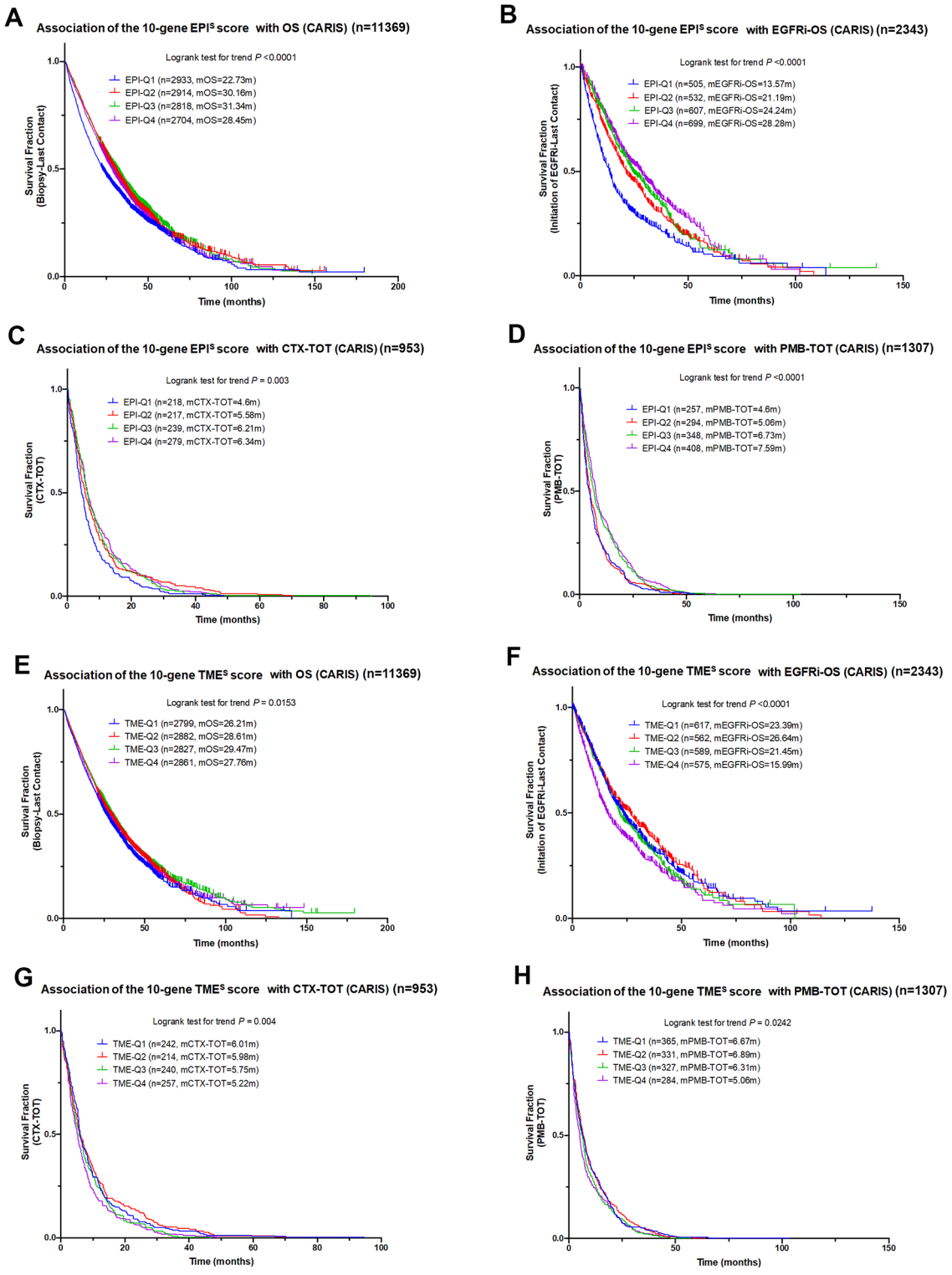


Fig. 9 The 10-gene EPI^S (vs. the 10-gene TME^S) signature scores were independently validated as mildly prognostic, but highly predictive of EGFRi outcomes using a large real world CarisLS CRC dataset. The Kaplan-Meier (KM) survival analyses of (**A**-**D**) the 10-gene EPI^S and (**E**-**H**) the 10-gene TME^S score quartiles (Q4, highest; Q3; Q2; Q1, lowest) were performed in the CarisLS (**A**,**E**) all non-EGFRi CRC tumors ($n = 11369$, OS), and (**B**,**F**) EGFRi-treated ($n = 2343$, OS), (**C**,**G**) cetuximab (CTX)-treated ($n = 953$, TOT) and (**D**,**H**) panitumumab (PMB)-treated ($n = 1307$, TOT) tumors, respectively. EGFR inhibitor (EGFRi) therapies include both CTX and PMB. OS — overall survival (OS); TOT — time on treatment

EPI^S score was significantly correlated with *APC*-truncating mutations (Additional File 1-Fig. S4) that may predict cetuximab sensitivity [16]. These data suggest clinical potential for this 10-gene EPI^S signature as a predictive biomarker to improve EGFRi outcome.

In addition to the CMS classification, other new CRC classifications have been recently developed as well. For example, Joanito et al. identified two intrinsic (epithelial) subtypes (iCMS2 and iCMS3) [23], whereas Malla et al. defined three pathway-derived subtypes (PDS1-3) [26]. Our additional analyses in the Merck-Moffitt CRC tumors revealed that 10-gene TME^S score vs. the 10-gene EPI^S score were distinctly associated with the iCMS3 signature and with three PDS subtypes (Additional File 1-Fig. S6, Fig. S7). These data also support the association of the 10-gene TME^S vs. the 10-gene EPI^S with distinct prognostic contributions.

Remarkably, nine of the 10 TME^S genes and all the 10 EPI^S genes were prognostic (Fig. 5), suggesting individual functional importance. The biological roles of many of these genes have been reported [20]. For example, *CD109* (a TME^S gene) was identified as a metastasis-associated protein marker [49]. Reduced expression of *EPHB2* (an EPI^S gene) was associated with metastasis [50] and its overexpression induced EMT [51]. How the biological functions of the TME^S vs. EPI^S genes are related to their distinct contributions to CRC outcomes remains to be investigated.

Limitation of the study

We serendipitously identified 10-gene TME^S and 10-gene EPI^S signature scores that may have the potential to quantify the cellular contribution of the TME vs. the EPI to therapeutic outcomes. However, this study was by nature a retrospective study. Thus, additional technical and clinical validation analyses of these two newly identified signature scores, especially in prospective studies, will be needed for developing clinical applications.

Conclusions

This study has generated a pair of new, distinct 10-gene signature scores (the TME^S vs. the EPI^S) that identify differential cellular contributions of the TME and the tumor epithelium to therapeutic outcomes. While the TME^S score was strongly correlated with MEKi resistance, the EPI^S score was highly *predictive* of EGFRi outcomes as demonstrated in clinical trial and large real-world datasets. With targeted approaches emerging to address both the TME and the tumor EPI, the distinct 10-gene signature scores (the TME^S vs. the EPI^S) may have a novel biomarker role to permit optimization of CRC therapy by identifying resistant vs. sensitive subpopulations.

Abbreviations

CMS	Consensus Molecular Subtype
CRC	Colorectal Cancer
CR/PR	Complete/Partial Responses
CTX	Cetuximab
CTX-S	Cetuximab Sensitivity
EGFRi	EGFR Inhibitor
EMT	Epithelial to Mesenchymal Transition
EPI	Epithelium
EPI^S	the 10-gene EPI^S Signature Score
FFPE	Formalin Fixed Paraffin Embedded
KM survival	Kaplan Meier survival analysis
MEKi	MEK inhibitor
MSI	Microsatellite Instability
MSS	Microsatellite Stable
NEG	Negatively Correlated
OS	Overall Survival
PFS	Progression Free Survival
RFS	Relapse Free Survival
PC1	The First Principal Component
PD	Progressed Disease
PMB	Panitumumab
POS	Positively Correlated
SAR	Survival After Relapse
scRNASEQ	single cell RNA Sequencing
SD	Stable Disease
SRCi	SRC inhibitor
TGCA	The Cancer Genome Atlas Program
TME	Tumor Microenvironment
TME^S	the 10-gene TME^S signature score
TOT	Time on Treatment
UTD	Undetermined
WT	Wild Type
WTS	Whole Transcriptome Sequencing

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-13829-2.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

M.Y. and T.J.Y. conceived the hypothesis and designed the projects; M.Y. wrote the main manuscript text; T.J.Y acquired funding; M.Y., M.V.N., M.J.S., N.G., L.P., A.L., W.J.P., R.S., M.M., H.W., J.R.S., A.A., D.C., A.E., G.S., M.K., E.L., S.G. and T.J.Y. analyzed and interpreted data; M.Y., M.V.N., M.J.S., N.G., L.P., A.L., W.J.P., R.S., M.M., H.W. J.R.S., A.A., D.C., A.E., G.S., M.K., E.L., S.G. and T.J.Y. helped revise the manuscript; All authors have reviewed and approved the manuscript.

Funding

This study received support from the Tampa General Hospital Foundation and NIH UH3CA227955 (TJY), R21CA256372 (TJY) and R21CA255312 (TJY).

Data availability

Affymetrix gene expression datasets of the Merck-Moffitt CRC (***n***=2373) and Merck-PN04 CRC (***n***=44) are owned by Merck whose data can be accessed by qualified researchers as described(refs 11, 12). The CarisLS's de-identified RNASEQ data are owned by the company. Qualified researchers can apply for access to these data by contacting CarisLS (N.G.), submitting a brief proposal and signing a data usage agreement. Other data are either publicly available or are available from the corresponding authors (M.Y., T.J.Y.) upon reasonable request.

Declarations

Ethics approval and consent to participate

1. Merck-Moffitt CRCs: All the experiment protocol for involving human data was in accordance with the guidelines of national/international/institutional or Declaration of Helsinki. The tissue and clinical data were collected under the approval of the Institutional Review Board (IRB) of Moffitt Cancer Center with the informed written consent obtained from participating patients. *The approved IRB No. is MCC14690.2. CARIS LS real-world CRC dataset*: This study was conducted in accordance with the guidelines of the Declaration of Helsinki, Belmont Report, and U.S. Common Rule. Real-world clinical outcome information was obtained from insurance claims data. In keeping with 45 CFR 46.101(b) [4], this study was performed using retrospective and de-identified clinical data. *This study was thus consideredinstitutional review board (IRB) exempt*, *and no patient consent was necessary.*

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Surgery, University of South Florida, 560 Channelside Drive, Tampa, FL 33602, USA
[2]Merck Research Laboratories, 33 Avenue Louis Pasteur, Boston, MA 02115, USA
[3]Department of Biostatistics and Bioinformatics, Moffitt Cancer Center & Research Institute, 12902 Magnolia Drive, Tampa, FL 33612, USA
[4]Medical Affairs, Caris Life Sciences, 4610 S 44th Pl, Phoenix, AZ 85040, USA
[5]Phenome Health, 401 Terry Ave N, Seattle, WA 98109, USA
[6]Department of Molecular Medicine, University of South Florida, 12901 Bruce B. Downs Boulevard, Tampa, FL 33612, USA
[7]Tampa General Hospital Cancer Institute, 1 Tampa General Circle, Tampa, FL 33606, USA
[8]Department of Pathology, Florida Digestive Health Specialists, 10920 Technology Ter, Lakewood Ranch, FL 34202, USA
[9]Department of Pathology, Moffitt Cancer Center & Research Institute, 12902 Magnolia Drive, Tampa, FL 33612, USA
[10]Division of Medical Oncology, Department of Medicine, Washington University, 4590 Nash Way, St. Louis, MO 63110, USA
[11]Division of Hematology, Oncology and Transplantation, Department of Medicine and Masonic Cancer Center, University of Minnesota, 420 Delaware Street SE, Minneapolis, MN 55455, USA
[12]Department of Medical Oncology, Rutgers Cancer Institute of New Jersey, Robert Wood Johnson School of Medicine, 195 Little Albany Street, New Brunswick, NJ 08903, USA

Received: 17 September 2024 / Accepted: 27 February 2025
Published online: 12 March 2025

References

1. Bissell MJ, Radisky D. Putting tumours in context. Nat Rev Cancer. 2001;1(1):46–54.
2. Bruni D, Angell HK, Galon J. The immune contexture and immunoscore in cancer prognosis and therapeutic efficacy. Nat Rev Cancer. 2020;20(11):662–80.
3. Cascio S, Chandler C, Zhang L, Sinno S, Gao B, Onkar S, et al. Cancer-associated MSC drive tumor immune exclusion and resistance to immunotherapy, which can be overcome by Hedgehog Inhibition. Sci Adv. 2021;7(46):eabi5790.
4. Galindo-Pumarino C, Collado M, Herrera M, Pena C. Tumor Microenvironment in Metastatic Colorectal Cancer: The Arbitrator in Patients' Outcome. Cancers (Basel). 2021;13(5).
5. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487(7407):330–7.
6. Schell MJ, Yang M, Teer JK, Lo FY, Madan A, Coppola D, et al. A multigene mutation classification of 468 colorectal cancers reveals a prognostic role for APC. Nat Commun. 2016;7:11743.
7. Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21(11):1350–6.
8. Picard E, Verschoor CP, Ma GW, Pawelec G. Relationships between immune landscapes, genetic subtypes and responses to immunotherapy in colorectal Cancer. Front Immunol. 2020;11:369.
9. Roelands J, Kuppen PJK, Vermeulen L, Maccalli C, Decock J, Wang E et al. Immunogenomic classification of colorectal Cancer and therapeutic implications. Int J Mol Sci. 2017;18(10).
10. Fenstermacher DA, Wenham RM, Rollison DE, Dalton WS. Implementing personalized medicine in a cancer center. Cancer J. 2011;17(6):528–36.
11. Cristescu R, Nebozhyn M, Zhang C, Albright A, Kobie J, Huang L, et al. Transcriptomic determinants of response to pembrolizumab monotherapy across solid tumor types. Clin Cancer Res. 2022;28(8):1680–9.
12. Ayers M, Nebozhyn M, Cristescu R, McClanahan TK, Perini R, Rubin E, et al. Molecular profiling of cohorts of tumor samples to guide clinical development of pembrolizumab as monotherapy. Clin Cancer Res. 2019;25(5):1564–73.
13. Pelka K, Hofree M, Chen JH, Sarkizova S, Pirl JD, Jorgji V, et al. Spatially organized multicellular immune hubs in human colorectal cancer. Cell. 2021;184(18):4734–e5220.
14. 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(5):e1001453.
15. Yang M, Davis TB, Pflieger L, Nebozhyn MV, Loboda A, Wang H, et al. An integrative gene expression signature analysis identifies CMS4 KRAS-mutated colorectal cancers sensitive to combined MEK and SRC targeted therapy. BMC Cancer. 2022;22(1):256.
16. Yang M, Schell MJ, Loboda A, Nebozhyn M, Li J, Teer JK, et al. Repurposing EGFR inhibitor utility in colorectal Cancer in mutant APC and TP53 subpopulations. Cancer Epidemiol Biomarkers Prev. 2019;28(7):1141–52.
17. Sclafani F, Kim TY, Cunningham D, Kim TW, Tabernero J, Schmoll HJ, et al. A randomized phase II/III study of dalotuzumab in combination with cetuximab and Irinotecan in chemorefractory, KRAS Wild-Type, metastatic colorectal Cancer. J Natl Cancer Inst. 2015;107(12):djv258.
18. Khambata-Ford S, Garrett CR, Meropol NJ, Basik M, Harbison CT, Wu S, et al. Expression of Epiregulin and Amphiregulin and K-ras mutation status predict disease control in metastatic colorectal cancer patients treated with cetuximab. J Clin Oncol. 2007;25(22):3230–7.
19. Loboda A, Nebozhyn MV, Watters JW, Buser CA, Shaw PM, Huang PS, et al. EMT is the dominant program in human colon cancer. BMC Med Genomics. 2011;4:9.
20. Schell MJ, Yang M, Missiaglia E, Delorenzi M, Soneson C, Yue B, et al. A composite gene expression signature optimizes prediction of colorectal Cancer metastasis and outcome. Clin Cancer Res. 2016;22(3):734–45.
21. Broecker F, Hardt C, Herwig R, Timmermann B, Kerick M, Wunderlich A, et al. Transcriptional signature induced by a metastasis-promoting c-Src mutant in a human breast cell line. FEBS J. 2016;283(9):1669–88.
22. Dry JR, Pavey S, Pratilas CA, Harbron C, Runswick S, Hodgson D, et al. Transcriptional pathway signatures predict MEK addiction and response to selumetinib (AZD6244). Cancer Res. 2010;70(6):2264–73.
23. Joanito I, Wirapati P, Zhao N, Nawaz Z, Yeo G, Lee F, et al. Single-cell and bulk transcriptome sequencing identifies two epithelial tumor cell States and refines the consensus molecular classification of colorectal cancer. Nat Genet. 2022;54(7):963–75.
24. Bagaev A, Kotlov N, Nomie K, Svekolkin V, Gafurov A, Isaeva O, et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. Cancer Cell. 2021;39(6):845–65. e7.
25. Eide PW, Bruun J, Lothe RA, Sveen A. CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. Sci Rep. 2017;7(1):16618.
26. Malla SB, Byrne RM, Lafarge MW, Corry SM, Fisher NC, Tsantoulis PK, et al. Pathway level subtyping identifies a slow-cycling biological phenotype associated with poor clinical outcomes in colorectal cancer. Nat Genet. 2024;56(3):458–72.
27. 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.
28. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat. 1979;6(2):65–70.
29. Schubbert S, Shannon K, Bollag G. Hyperactive Ras in developmental disorders and cancer. Nat Rev Cancer. 2007;7(4):295–308.

30. Caunt CJ, Sale MJ, Smith PD, Cook SJ. MEK1 and MEK2 inhibitors and cancer therapy: the long and winding road. Nat Rev Cancer. 2015;15(10):577–92.
31. Yaeger R, Yao Z, Hyman DM, Hechtman JF, Vakiani E, Zhao H, et al. Mechanisms of acquired resistance to BRAF V600E Inhibition in Colon cancers converge on RAF dimerization and are sensitive to its Inhibition. Cancer Res. 2017;77(23):6513–23.
32. Bhowmick NA, Neilson EG, Moses HL. Stromal fibroblasts in cancer initiation and progression. Nature. 2004;432(7015):332–7.
33. Quante M, Varga J, Wang TC, Greten FR. The Gastrointestinal tumor microenvironment. Gastroenterology. 2013;145(1):63–78.
34. Ashktorab H, Vilmenay K, Brim H, Laiyemo AO, Kibreab A, Nouraie M. Colorectal Cancer in young African Americans: is it time to revisit guidelines and prevention?? Dig Dis Sci. 2016;61(10):3026–30.
35. Ashktorab H, Brim H. Colorectal cancer subtyping. Nat Rev Cancer. 2022;22(2):68–9.
36. Mesker WE, Junggeburt JM, Szuhai K, de Heer P, Morreau H, Tanke HJ, et al. The carcinoma-stromal ratio of colon carcinoma is an independent factor for survival compared to lymph node status and tumor stage. Cell Oncol. 2007;29(5):387–98.
37. Herrera M, Islam AB, Herrera A, Martin P, Garcia V, Silva J, et al. Functional heterogeneity of cancer-associated fibroblasts from human colon tumors shows specific prognostic gene expression signature. Clin Cancer Res. 2013;19(21):5914–26.
38. Calon A, Lonardo E, Berenguer-Llergo A, Espinet E, Hernando-Momblona X, Iglesias M, et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. Nat Genet. 2015;47(4):320–9.
39. Paauwe M, Schoonderwoerd MJA, Helderman R, Harryvan TJ, Groenewoud A, van Pelt GW, et al. Endoglin expression on Cancer-Associated fibroblasts regulates invasion and stimulates colorectal Cancer metastasis. Clin Cancer Res. 2018;24(24):6331–44.
40. Dongre A, Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. Nat Rev Mol Cell Biol. 2019;20(2):69–84.
41. Chowdhury S, Xiu J, Ribeiro JR, Nicolaides T, Zhang J, Korn WM, et al. Consensus molecular subtyping of metastatic colorectal cancer expands biomarker-directed therapeutic benefit for patients with CMS1 and CMS2 tumors. Br J Cancer. 2024;131(8):1328–39.
42. Lenz HJ, Ou FS, Venook AP, Hochster HS, Niedzwiecki D, Goldberg RM, et al. Impact of consensus molecular subtype on survival in patients with metastatic colorectal cancer: results from CALGB/SWOG 80405 (Alliance). J Clin Oncol. 2019;37(22):1876–85.
43. Okita A, Takahashi S, Ouchi K, Inoue M, Watanabe M, Endo M, et al. Consensus molecular subtypes classification of colorectal cancer as a predictive factor for chemotherapeutic efficacy against metastatic colorectal cancer. Oncotarget. 2018;9(27):18698–711.
44. Farina-Sarasqueta A, van Lijnschoten G, Moerland E, Creemers GJ, Lemmens VE, Rutten HJ, et al. The BRAF V600E mutation is an independent prognostic factor for survival in stage II and stage III colon cancer patients. Annals Oncology: Official J Eur Soc Med Oncol / ESMO. 2010;21(12):2396–402.
45. Roth AD, Tejpar S, Delorenzi M, Yan P, Fiocca R, Klingbiel D, et al. Prognostic role of KRAS and BRAF in stage II and III resected colon cancer: results of the translational study on the PETACC-3, EORTC 40993, SAKK 60– 00 trial. J Clin Oncol. 2010;28(3):466–74.
46. Frame MC. Src in cancer: deregulation and consequences for cell behaviour. Biochim Biophys Acta. 2002;1602(2):114–30.
47. Yeatman TJ. A renaissance for SRC. Nat Rev Cancer. 2004;4(6):470–80.
48. Caner A, Asik E, Ozpolat B. SRC signaling in Cancer and tumor microenvironment. Adv Exp Med Biol. 2021;1270:57–71.
49. Karhemo PR, Ravela S, Laakso M, Ritamo I, Tatti O, Makinen S, et al. An optimized isolation of biotinylated cell surface proteins reveals novel players in cancer metastasis. J Proteom. 2012;77:87–100.
50. Yu G, Gao Y, Ni C, Chen Y, Pan J, Wang X, et al. Reduced expression of EphB2 is significantly associated with nodal metastasis in Chinese patients with gastric cancer. J Cancer Res Clin Oncol. 2011;137(1):73–80.
51. Gao Q, Liu W, Cai J, Li M, Gao Y, Lin W, et al. EphB2 promotes cervical cancer progression by inducing epithelial-mesenchymal transition. Hum Pathol. 2014;45(2):372–81.

Publisher's note