Additional File 1 ─ Supplementary Figures S1-S11

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

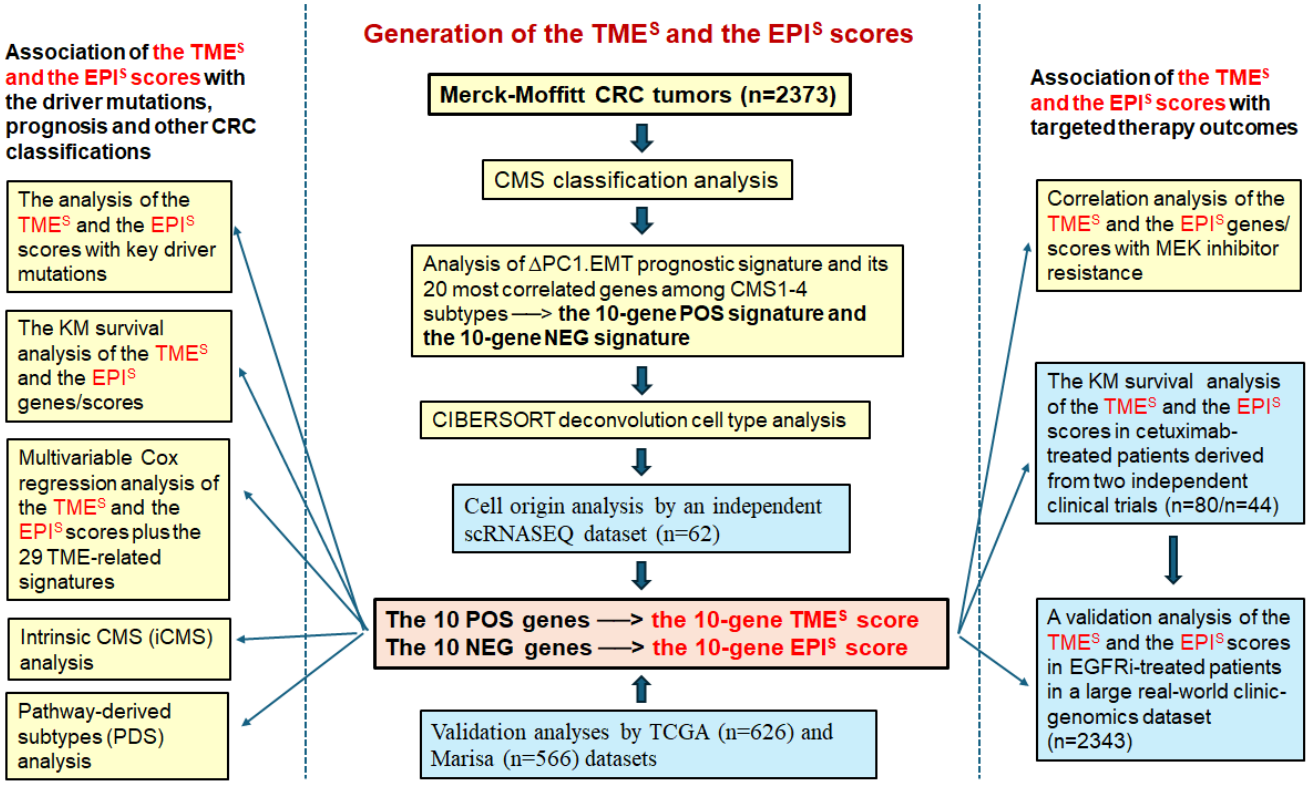
Mingli Yang*, Michael V. Nebozhyn, Michael J. Schell, Nishant Gandhi, Lance Pflieger, Andrey Loboda, W. Jack Pledger, Ramani Soundararajan, Michelle Maurin, Heiman Wang, Jetsen Rodriguez Silva, Ashley Alden, Domenico Coppola, Andrew Elliott, George Sledge, Moh'd Khushman, Emil Lou, Sanjay Goel, Timothy J Yeatman*

***Corresponding authors:**

Mingli Yang, PhD, Department of Surgery, University of South Florida, 560 Channelside Drive, Tampa, FL, 33602, (USA), Tel: +1 813-974-2632, email: mingliyang@usf.edu

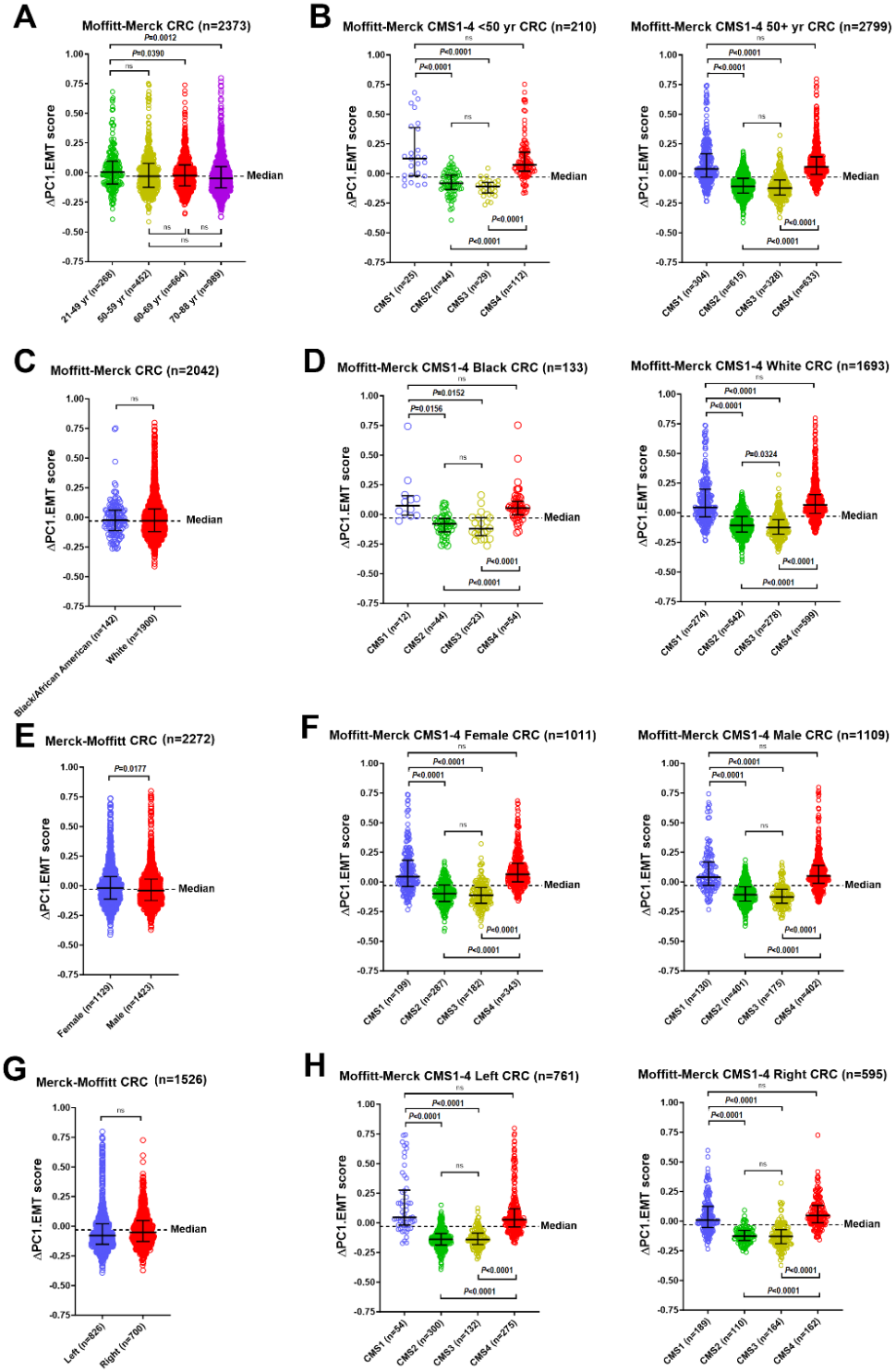
Timothy J. Yeatman, MD, Department of Surgery, University of South Florida, 560 Channelside Drive, Tampa, FL, 33602, (USA), Tel: +1 813-454-9129, email: tyeatman@usf.edu

Fig. S1



Supplementary Fig. S1. A flowchart of the study illustrating generation and analysis of the 10-gene TME^S and the 10-gene EPI^S signature scores using the Merck-Moffitt CRC dataset and various other CRC datasets. The studies using the Merck-Moffitt dataset are highlighted by light yellow color, whereas the studies using other independent datasets are highlighted by light blue color.

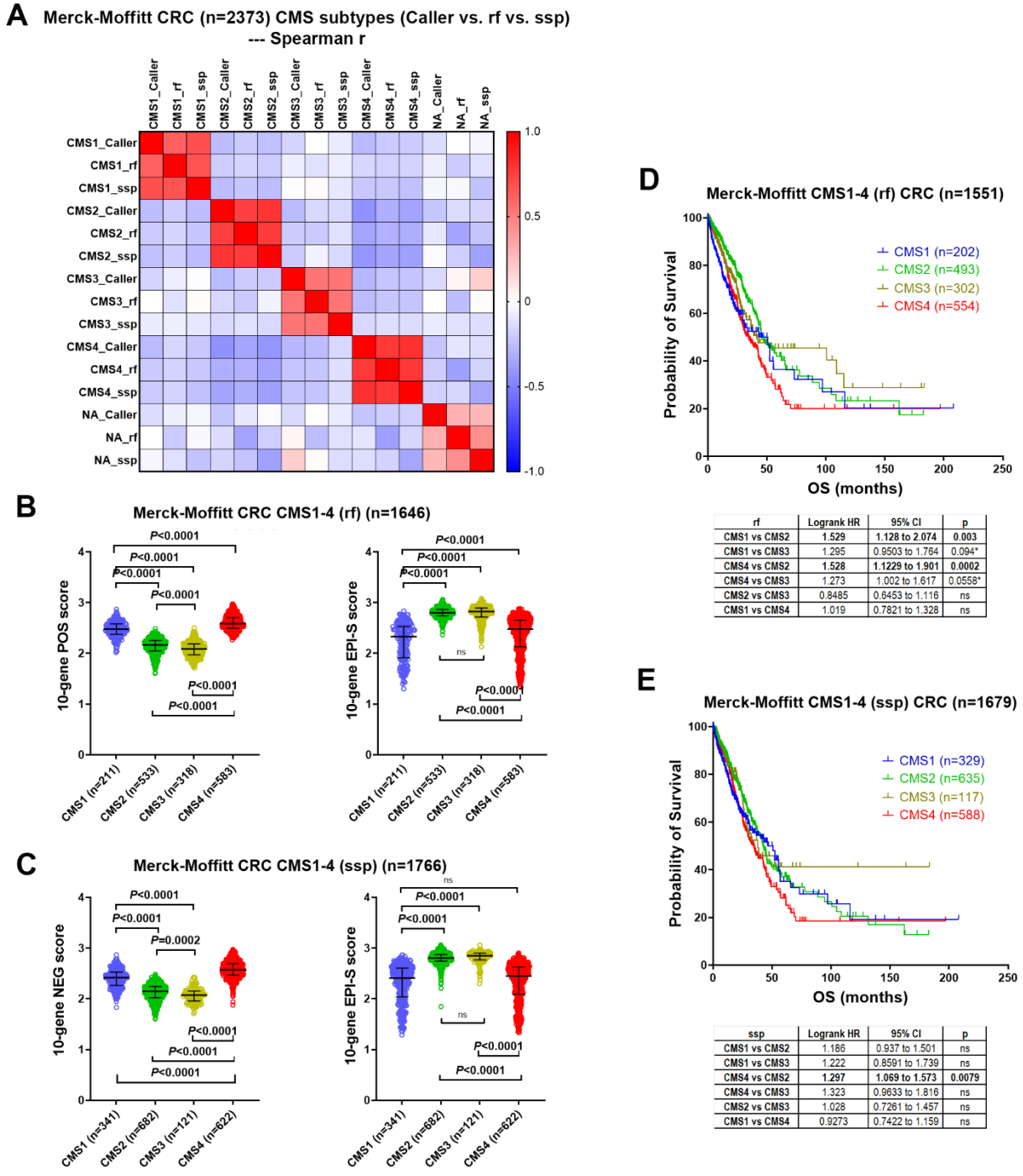
Fig. S2



Supplementary Fig. S2. The ΔPC1.EMT score was strongly associated with both CMS1 and CMS4 subtypes independent of age, race, sex and sidedness in the Merck-Moffitt CRCs. The

ΔPC1.EMT score comparison among the subgroups of age (**A**), race (**C**), sex (**E**) and sidedness (**G**). The ΔPC1.EMT score comparison among the CMS1 vs. CMS2 vs. CMS 3 vs. CMS4 subtypes in the patients of (**B**) < 50 yr and > 50+ yr, (**D**) African American/Black and White, (**F**) female and male, and (**H**) left- and right-sided, respectively. 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(1) (1). See **Tables 1** for detailed description of the baseline phenotypic characteristics.

Fig. S3



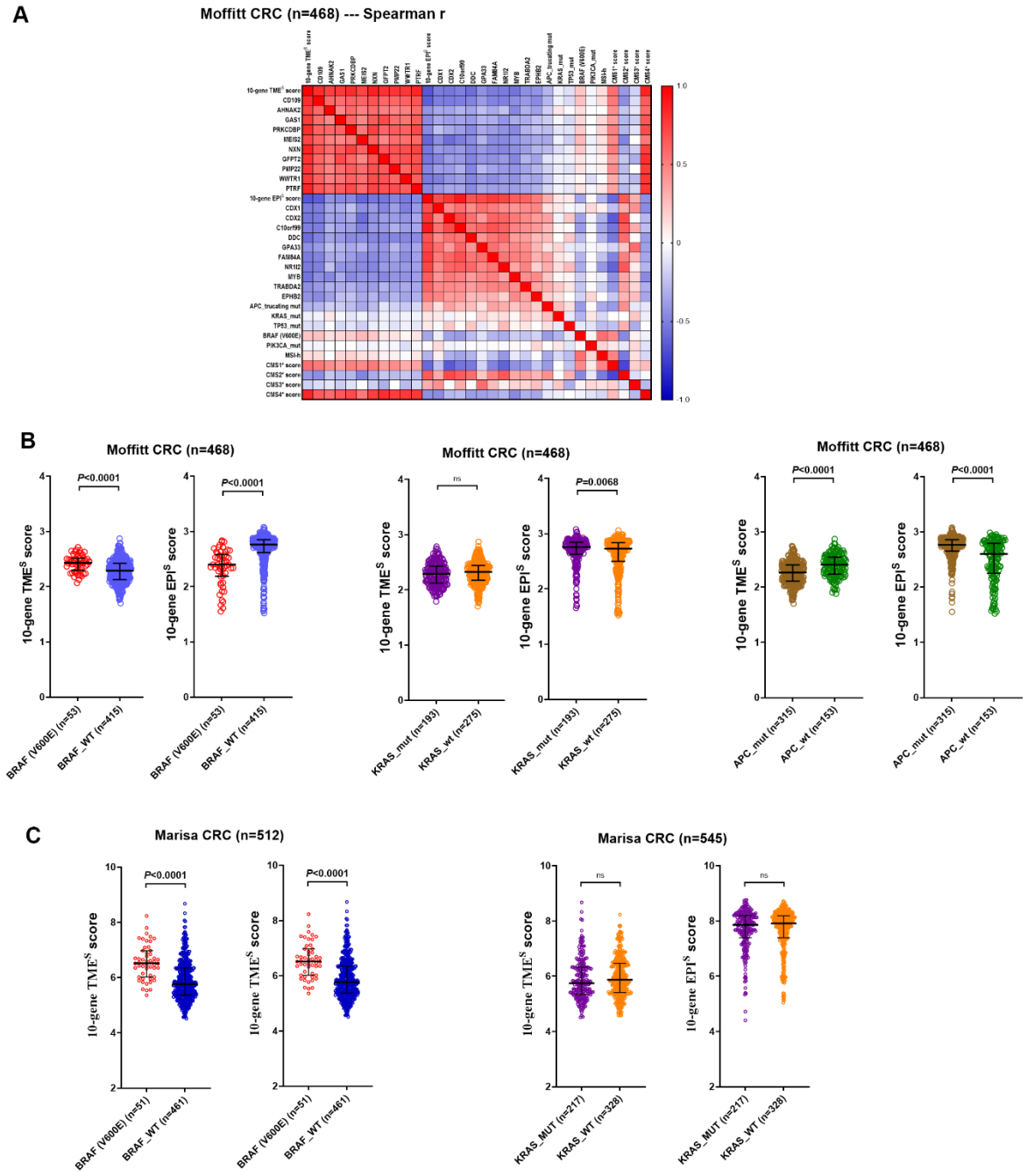
rf	Logrank HR	95% CI	p
CMS1 vs CMS2	1.529	1.128 to 2.074	0.003
CMS1 vs CMS3	1.295	0.9503 to 1.764	0.094*
CMS4 vs CMS2	1.528	1.1229 to 1.901	0.0002
CMS4 vs CMS3	1.273	1.002 to 1.617	0.0558*
CMS2 vs CMS3	0.8485	0.6453 to 1.116	ns
CMS1 vs CMS4	1.019	0.7821 to 1.328	ns

ssp	Logrank HR	95% CI	p
CMS1 vs CMS2	1.186	0.937 to 1.501	ns
CMS1 vs CMS3	1.222	0.8591 to 1.739	ns
CMS4 vs CMS2	1.297	1.069 to 1.573	0.0079
CMS4 vs CMS3	1.323	0.9633 to 1.816	ns
CMS2 vs CMS3	1.028	0.7261 to 1.457	ns
CMS1 vs CMS4	0.9273	0.7422 to 1.159	ns

Supplementary Fig. S3. A. Spearman correlation of the CMS subtypes generated by the CMScaller, the random forest (rf) and the single-cell predictor (ssp) in the Merck-Moffitt CRC tumors (n=2373). Comparison of the 10-gene POS and 10-gene NEG signature scores across the

CMS1-4 subtypes generated by the rf (**B**) and by the ssp (**C**). 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 (1). The Kaplan-Meier (KM) survival analyses among CMS1-4 subtypes generated by the rf (**D**) and by the ssp (**E**).

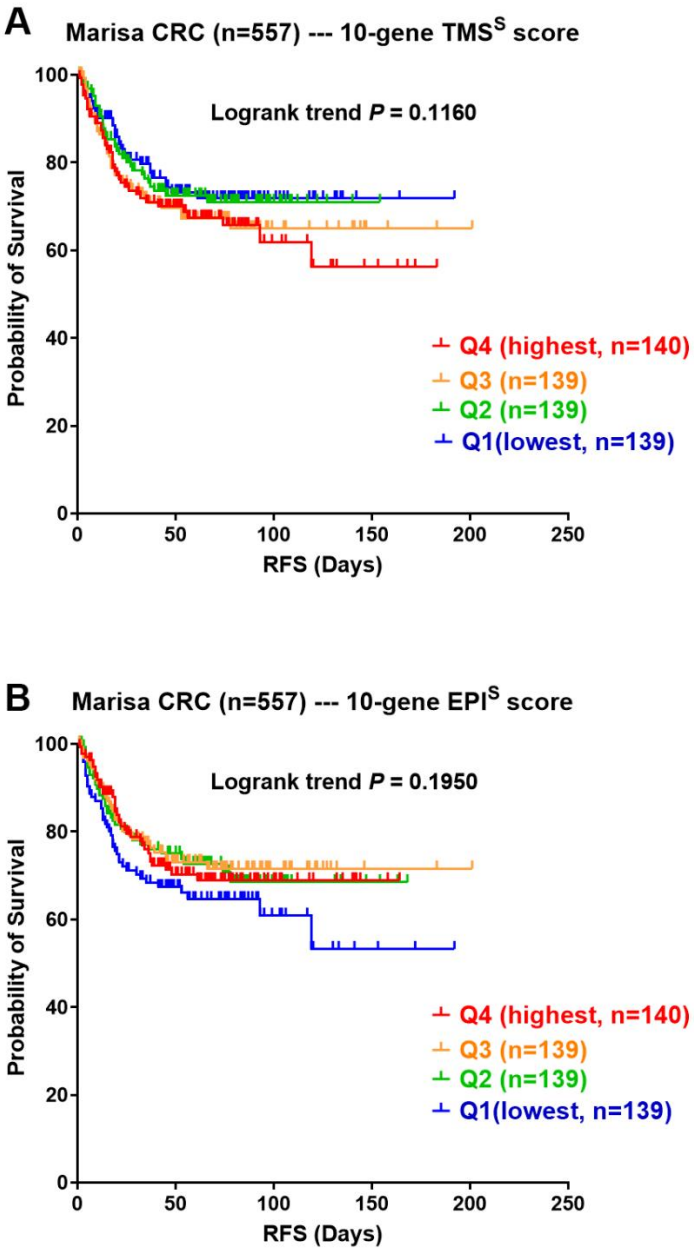
Fig. S4



Supplementary Fig. S4. The TME^S genes/score vs. the EPI^S genes/score were distinctly correlated with *BRAF* (V600E) and *APC* truncating mutations. (**A**) Spearman correlation of the 10 TME^S and the 10 EPI^S genes, and their respective 10-gene signature scores with *APC*

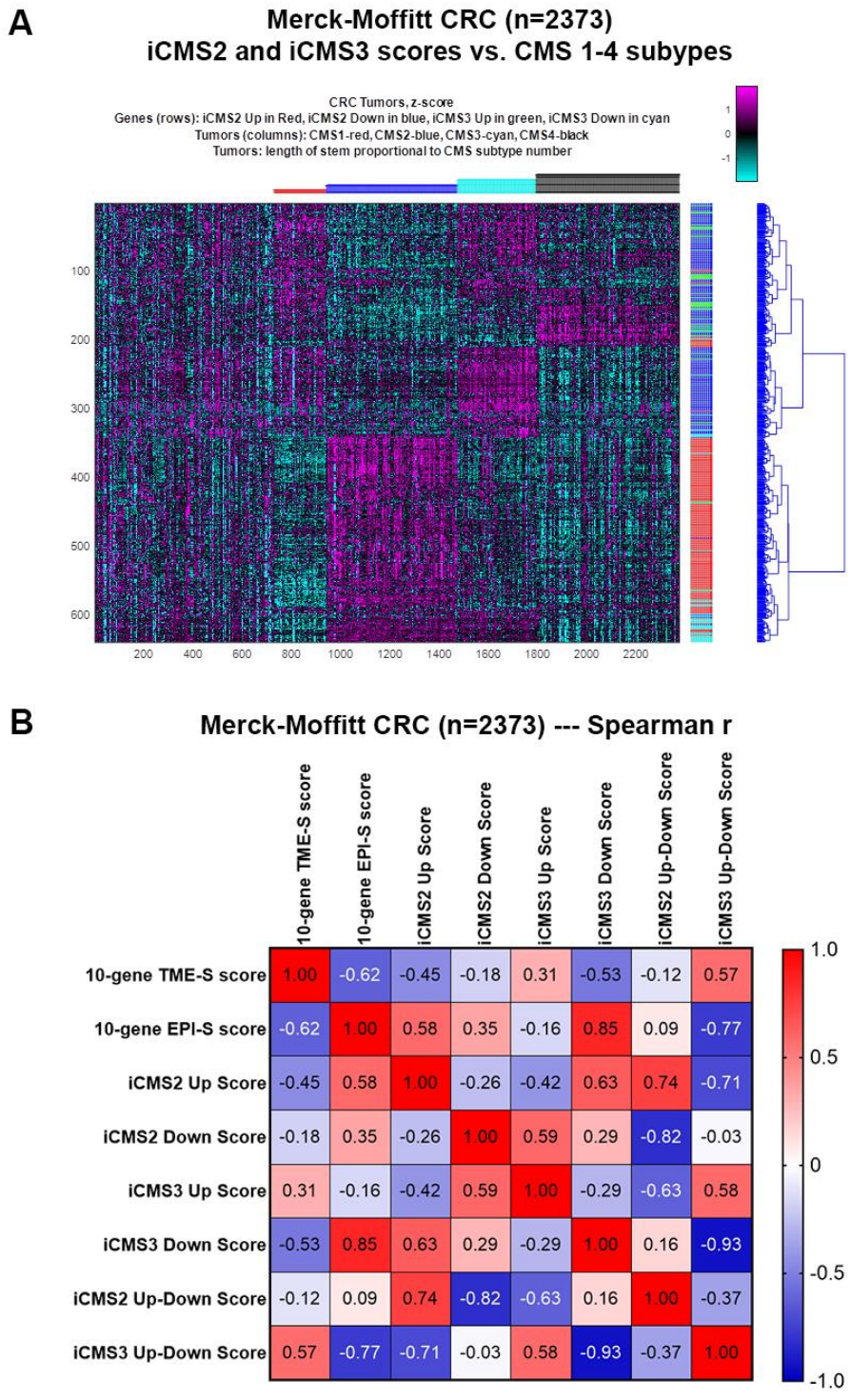
(truncating), *KRAS*, *TP53*, *BRAF* (V600E) and *PIK3CA* mutations, and with MSI-H as well as with the CMS1*, CMS2*, CMS3* and CMS4* scores in the Moffitt 468 CRC tumors. Of note, only the *BRAF*(V600E) and *APC* truncating mutations were considered as functional mutations for these two driver genes(2). The 10-gene TME^S score and most of the TME^S genes were significantly correlated with *BRAF* (V600E) and MSI and anti-correlated with *APC* truncating mutations. By contrast, the 10-gene EPI^S score and most of the EPI^S genes were significantly correlated with *APC* truncating mutations and anti-correlated with *BRAF* (V600E) and MSI. (**B**) Comparison of the 10-gene TME^S score and the 10-gene EPI^S score between *BRAF* (V600E) and *BRAF* WT (i.e. without *BRAF* (V600E)) or between *KRAS*-mutations and *KRAS* WT or between *APC* truncating mutations and *APC* WT (i.e. without *APC* truncating mutations) in the Moffitt 468 tumors. (**C**) Comparison of the TME^S and the EPI^S scores between *BRAF* (V600E) and *BRAF* WT or between *KRAS*-mutations and *KRAS* WT in an independent (Marisa) CRC dataset that had *BRAF* and *KRAS* mutation data. Bars represent Median with interquartile range. *P* values are for two-tailed Welch's t test.

Fig. S5



Supplementary Fig. S5. The Kaplan-Meier (KM) survival analyses by the score quartiles (Q1-Q4) of (**A**) the 10-gene TME^S signature score and (**B**) the 10-gene EPI^S signature score in the Marisa CRC tumors that had corresponding RFS data (n=557).

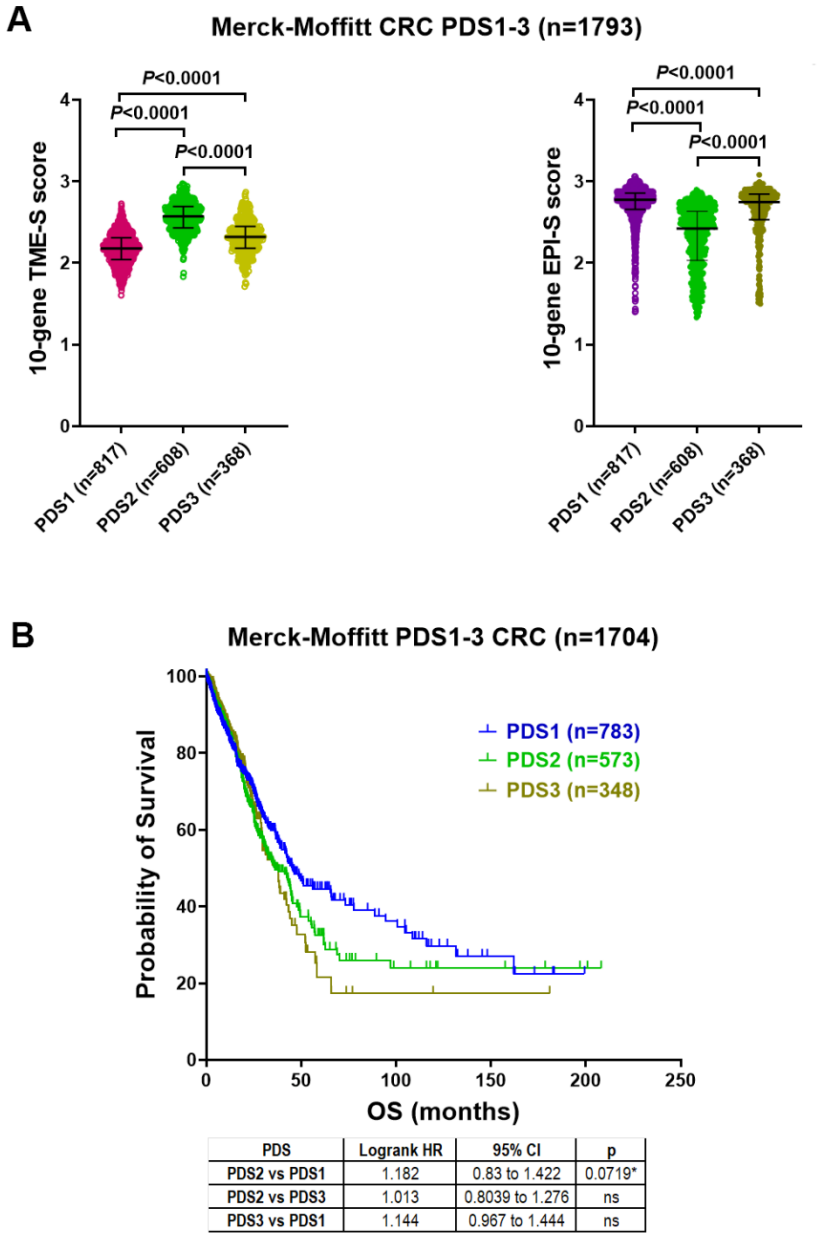
Fig S6



Supplementary Fig. S6. A. Heatmap correlation analysis on the iCMS2/iCMS3 signatures *vs.* the CMS subtypes in the Merck-Moffitt tumors (n=2373). **B.** Spearman correlation of the 10-gene

TME^S score and the 10-gene EPI^S score with the iCMS2 and iCMS3's up and down gene signature scores as well as their "up -down" signature scores in the Merck-Moffitt tumors (n=2373).

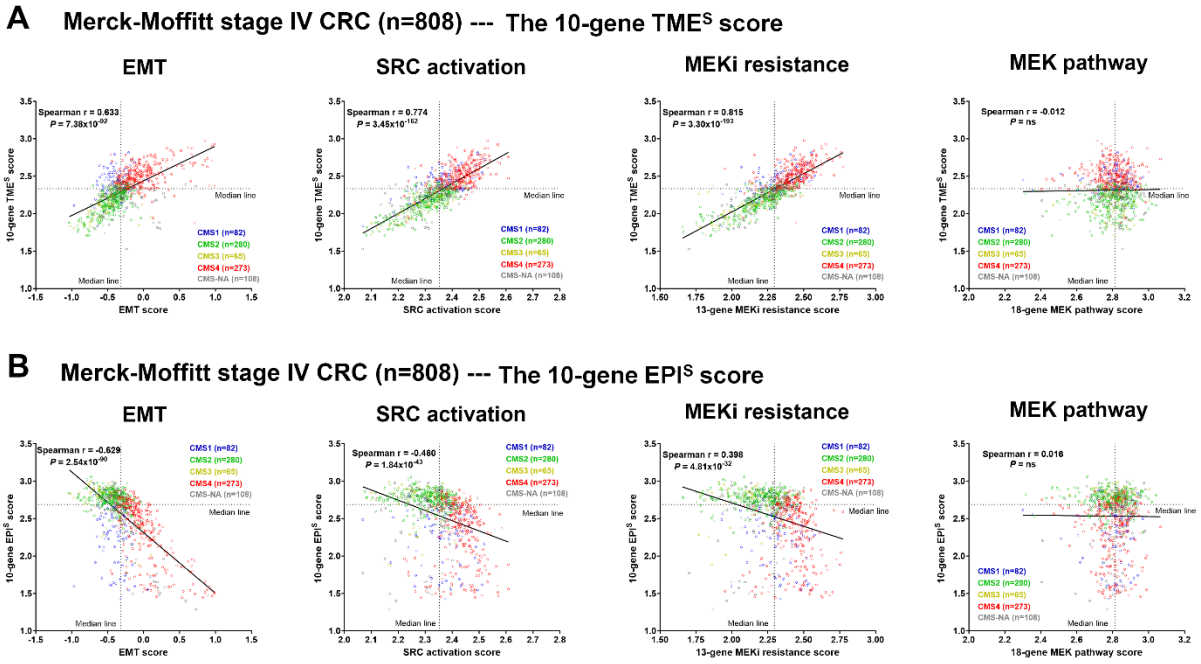
Fig S7



PDS	Logrank HR	95% CI	p
PDS2 vs PDS1	1.182	0.83 to 1.422	0.0719*
PDS2 vs PDS3	1.013	0.8039 to 1.276	ns
PDS3 vs PDS1	1.144	0.967 to 1.444	ns

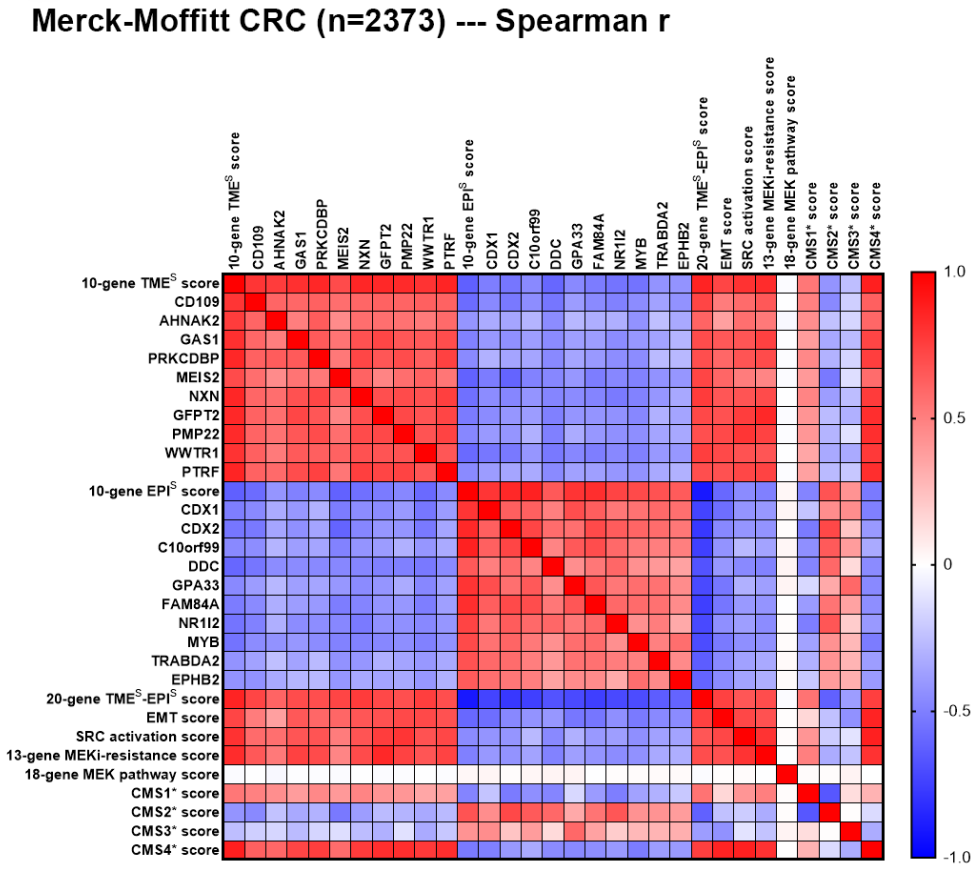
Supplementary Fig. S7. A. Comparison of the 10-gene TME^{S} and 10-gene EPI^{S} signature scores across the PDS1-3 subtypes in the Merck-Moffitt CRC tumors. Adjusted P values are shown for two-tailed Welch's t test after adjustments for multi-comparisons by Holm-Bonferroni method (1). **B**. The Kaplan-Meier (KM) survival analyses among PDS1-3 subtypes.

Fig. S8



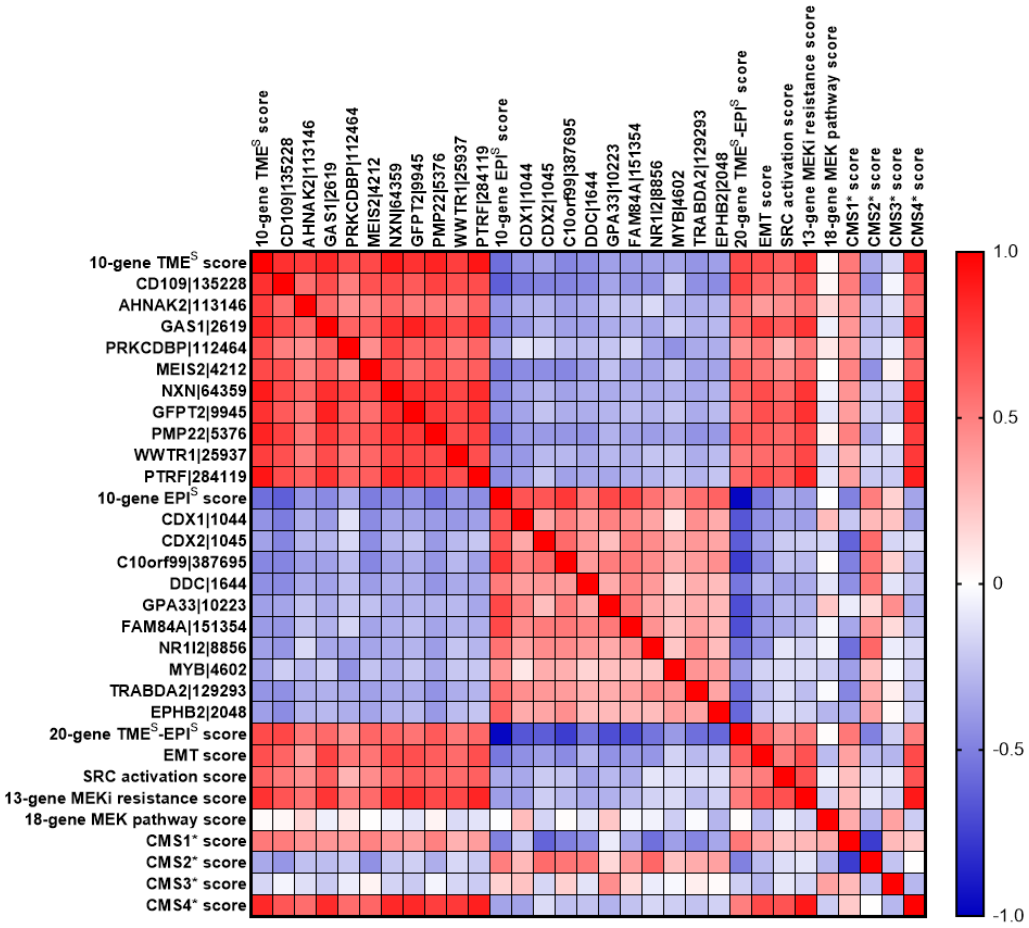
Supplementary Fig. S8. The 10-gene TME^S signature score (*vs.* the 10-gene EPI^S signature score) was strongly correlated with the EMT, SRC activation and MEK inhibitor resistance signature scores in the Merck-Moffitt Stage IV CRC tumors (n=808). Scatter plots of (**A**) the 10-gene TME^S score and (**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. Spearman r and *P* values are shown. The CMS1-4 subtypes are indicated by colors (CMS4 vs. CMS3 vs. CMS2 vs. CMS1 vs. CMS-NA). Here the CMS1-4, NA subtypes were generated by CMScaller.

Fig. S9



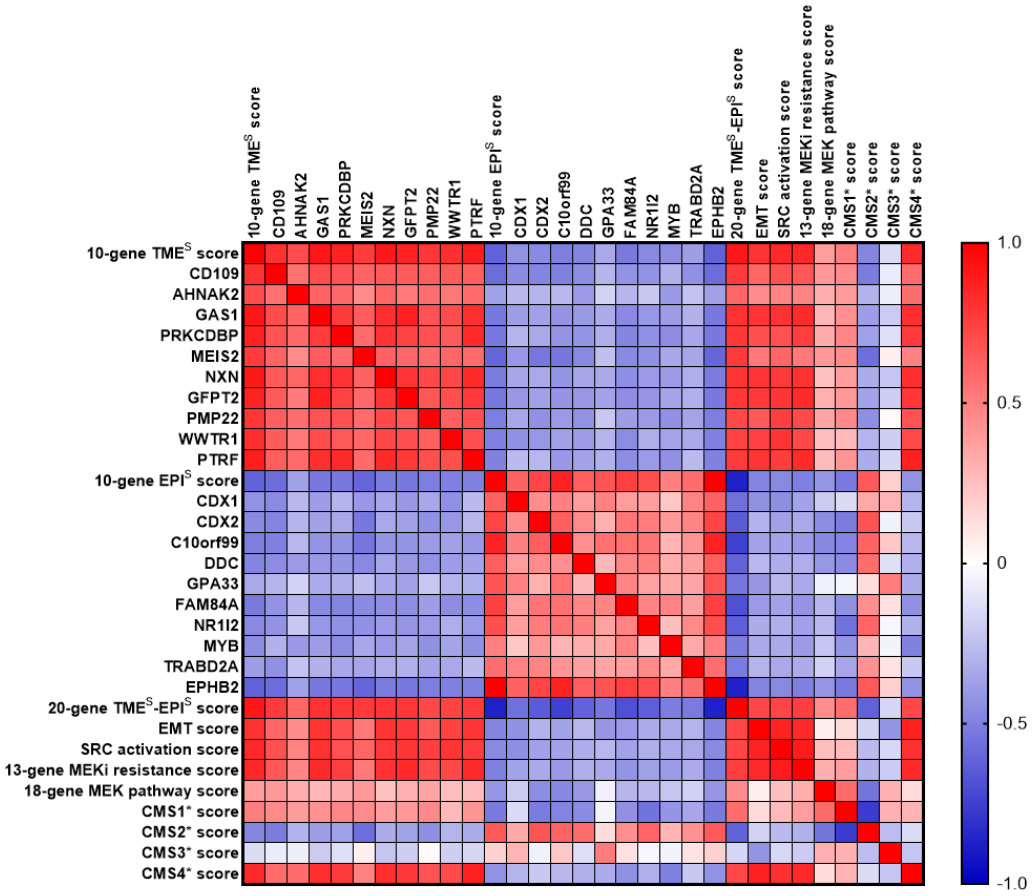
Supplementary Fig. S9. The 10 TME^S genes (*vs.* the 10 EPI^S genes) were strongly correlated with the EMT, SRC activation, 13-gene MEK inhibitor resistance signature scores in the Merck-Moffitt CRC tumors (n=2373)**.** Spearman correlation heatmap among the expression of individual genes and the EMT, SRC activation, 13-gene MEKi resistance and 18-gene MEK pathway activation signature scores is shown.

Fig. S10



Supplementary Fig. S10. The 10 TME^S genes (*vs.* the 10 EPI^S genes) were strongly correlated with the EMT, SRC activation, 13-gene MEK inhibitor resistance signature scores in an independent validation (TCGA) CRC dataset (n=626)**.** Spearman correlation heatmap among the expression of individual genes and the EMT, SRC activation, 13-gene MEKi resistance and 18-gene MEK pathway activation signature scores is shown.

Fig. S11



Supplementary Fig. S11. The 10 TME^S genes (*vs.* the 10 EPI^S genes) were strongly correlated with the EMT, SRC activation, 13-gene MEK inhibitor resistance signature scores in another independent validation (Marisa) CRC dataset (n=566)**.** Spearman correlation heatmap among the expression of individual genes and the EMT, SRC activation, 13-gene MEKi resistance and 18-gene MEK pathway activation signature scores is shown.

1. Holm S. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics. 1979;6(2):65–70.
2. 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.