Translational Oncology 40 (2024) 101845

Contents lists available at ScienceDirect

Translational Oncology

journal homepage: www.elsevier.com/locate/tranon

Characterizing heterogeneity along EMT and metabolic axes in colorectal cancer reveals underlying consensus molecular subtype-specific trends

Manas Sehgal [a,#], Soundharya Ramu [a,#], Joel Markus Vaz [a,b], Yogheshwer Raja Ganapathy [c], Srinath Muralidharan [a], Sankalpa Venkatraghavan [c], Mohit Kumar Jolly [a,*]

[a] Department of Bioengineering, Indian Institute of Science, Bangalore 560012, India
[b] School of Biological Sciences, Georgia Institute of Technology, Atlanta 30332, United States
[c] Department of Biotechnology, Indian Institute of Technology, Chennai 600036, India

ARTICLE INFO

Keywords:
Phenotypic heterogeneity
EMT
Colorectal cancer
Metabolic plasticity
Meta-analysis

ABSTRACT

Colorectal cancer (CRC) is highly heterogeneous with variable survival outcomes and therapeutic vulnerabilities. A commonly used classification system in CRC is the Consensus Molecular Subtypes (CMS) based on gene expression patterns. However, how these CMS categories connect to axes of phenotypic plasticity and heterogeneity remains unclear. Here, in our analysis of CMS-specific TCGA data and 101 bulk transcriptomic datasets, we found the epithelial phenotype score to be consistently positively correlated with scores of glycolysis, OXPHOS and FAO pathways, while PD-L1 activity scores positively correlated with mesenchymal phenotype scoring, revealing possible interconnections among plasticity axes. Single-cell RNA-sequencing analysis of patient samples revealed that that CMS2 and CMS3 subtype samples were relatively more epithelial as compared to CMS1 and CMS4. CMS1 revealed two subpopulations: one close to CMS4 (more mesenchymal) and the other closer to CMS2 or CMS3 (more epithelial), indicating a partial EMT-like behavior. Consistent observations were made in single-cell analysis of metabolic axes and PD-L1 activity scores. Together, our results quantify the patterns of two functional interconnected axes of phenotypic heterogeneity – EMT and metabolic reprogramming – in a CMS-specific manner in CRC.

Introduction

Colorectal cancer (CRC) is a multifaceted malignancy that arises from the epithelial lining of the colon or rectum [1]. It is the third most prevalent cancer worldwide [2]. Although surgical resection is the primary method of treatment for localized tumors, non-resectable tumors present substantial clinical challenges [3]. Targeted treatments such as bevacizumab (that prevents angiogenesis by blocking vascular endothelial growth factor (VEGF)), cetuximab and panitumumab (that block EGFR), pembrolizumab (that blocks the immune checkpoint PD-1) and vemurafenib (that inactivates BRAF V600 kinase) are frequently used [4,5]. The inherent diversity in molecular and phenotypic heterogeneity of tumors leads to variable responses to targeted therapies and survival outcomes [6,7,8].

To characterize this diversity, the Consensus Molecular Subtypes (CMS) system, a classification of CRC tumors based on the transcriptomic profile [9] was proposed. Based on gene expression profiles, the 'CMSclassifier' divided the CRC tumors into four subtypes, using a random forest algorithm based on gene expression signals from tumor's immune and stromal compartments: a) Microsatellite instability-immune, or CMS1 tumors, that show significant immune activation and genomic instability, b) 'Canonical' CMS2 tumors that exhibit WNT and MYC signalling pathway activation, c) 'Metabolic' CMS3 tumors that show Epithelial-mesenchymal transition (EMT) features and metabolic dysregulation, and d) 'Mesenchymal' CMS4 tumours that display significant stromal infiltration, angiogenesis, and mesenchymal characteristics [10–12]. However, CMSclassifier algorithm sometimes failed to correctly identify CMS4-mesenchymal sub--population in cell lines, patient-derived organoids, and xenografts. Thus, CMSCaller, a more robust classifier developed lately, is being increasingly adopted to deliver a more accurate subtype classification utilizing multiple sources of transcriptomic data [13].

CMS classification provides valuable therapeutic insights. For instance, clinical studies have found that the CMS1 patients who

* Corresponding author.
E-mail address: mkjolly@iisc.ac.in (M.K. Jolly).
# These authors contributed equally to this work.

https://doi.org/10.1016/j.tranon.2023.101845
Received 6 August 2023; Received in revised form 1 November 2023; Accepted 21 November 2023
Available online 28 November 2023

received bevacizumab had significantly higher overall survival and progression free survival rates compared to CMS1 patients who received cetuximab [14,15] whereas for CMS4 tumors, irinotecan (IRI)-based chemotherapy outperforms oxaliplatin (OX)-based chemotherapy [16]. Since each subtype responds differently to therapies, CMS classification offers more personalized strategies for therapeutic interventions, thus improving treatment outcomes.

Recently, another method is being increasingly used to quantify the extent of heterogeneity in tumor cells using single-cell transcriptomic data (scRNA-seq) - Shannon Entropy. Entropy is derived from information theory and is helpful in providing a quantitative measure of the diversity and distribution of gene expression within cell populations [17,18]. Moreover, Shannon entropy can be used to identify key genes driving cellular heterogeneity. Genes with high entropy are often associated with the dynamics of lineage specification and cell-fate decisions. Their expression patterns indicate transcriptional plasticity and cell-state transitions, providing valuable insights into underlying cellular dynamics driven by regulatory networks [19,20]. Thus, investigating the entropy of tumor cells, along with their functional attributes, is important to fully understand the underlying biological variability and vulnerabilities in CRC.

Here, we analyze 101 bulk transcriptomic datasets, along with patient tumor samples in colorectal cancer from The Cancer Genome Atlas (TCGA) and single-cell RNA sequencing data, to evaluate the degree of variation among CMS subtypes across metabolic reprogramming and EMT axes. Further, we analyzed the consequences of associations between these axes for disease prognosis. Our results show that the epithelial phenotype score was positively correlated with scores of glycolysis, OXPHOS and FAO pathways, while mesenchymal scores showed CMS subtype-specific associations with metabolic axes. PD-L1 activity scores consistently correlated positively with mesenchymal signature ones and negatively with epithelial signature ones, across CMS categories. Finally, we observed that CMS2 and CMS3 were more epithelial as compared to CMS1 and CMS4. Interestingly, single-cell RNA-seq data revealed two subpopulations in CMS1: one close to CMS4 (more mesenchymal) and the other closer to CMS2 or CMS3 (more epithelial), indicating a partial EMT-like behavior. Together, our results quantify the patterns of epithelial-mesenchymal heterogeneity and interplay between EMT and metabolic plasticity in different CMS subtypes in CRC.

Methods

Software and datasets

Computational and statistical analyses were conducted using R (version 4.3.0) and Python (version 3.9). Microarray datasets were retrieved from National Center for Biotechnology Information's Gene expression omnibus (NCBI GEO) using the 'GEOquery' R package. Processed RNA sequencing and single-cell RNA sequencing data were also obtained directly from individual datasets from the NCBI GEO database (**Table S1**). TCGA datasets were obtained using UCSC Xena tools for COAD_READ.

Pre-processing of datasets

Pre-processing of microarray datasets was conducted to obtain the gene-wise expression from the probe-wise expression matrix using respective annotation files for the mapping of probes to genes. In case multiple probes were mapped to a single gene, the mean expression of all mapped probes was utilized to obtain the final values for those genes.

Raw counts obtained for RNA and single-cell RNA sequencing data were normalized for gene length and transformed to transcripts per million (TPM) values. They were then log2 normalized to acquire the final expression data.

For single-cell RNA sequencing (scRNA-seq) datasets, MAGIC (version 2.0.3) [21] imputation algorithm was utilized to recover noisy and sparse single-cell data using diffusion geometry. To map individual reads to corresponding genes, relevant platform annotation files were utilized.

CMS Classification

CMS classification for colorectal cancer samples and tumor cells was carried out using 'CMScaller' [13]. 'CMScaller' uses the nearest template prediction algorithm to assign the CMS (CMS1-CMS4) to each sample with a prediction distance with a corresponding p-value. The predictions with insignificant p-values ($p > 0.05$) are not assigned any subtype. The CMS template genes were also used as signatures for ssGSEA scoring to identify the enrichment of the CMS-specific signatures in the sample.

Gene signature scoring

To quantify the enrichment of epithelial and mesenchymal signatures independently, ssGSEA (single sample gene set enrichment analysis) was performed on epithelial (for Epi scores) and mesenchymal (for Mes scores) gene signatures [22] separately using GSEAPY python library for bulk RNA sequencing and microarray datasets. Normalized enrichment score (NES) for these genesets was obtained for further analysis. A higher NES score corresponds to enrichment of that particular phenotype for the given sample. Similarly, the gene signatures for hallmark EMT, hallmark Fatty Acid Oxidation, and hallmark Glycolysis were obtained from molecular signatures database MSigDB and the respective scores were calculated [23]. PD-L1 signature was curated as reported earlier [24], wherein the top correlated genes (Spearman correlation coefficient > 0.5 and $p < 0.01$) with PD-L1 levels in at least any 13 out of 27 cancer types were considered (**Table S2**).

The activity scores for metabolic pathways, PD-L1, and E/M signatures for single-cell RNA sequencing datasets were computed using AUCell (version 1.18.1) [25] from the 'Bioconductor' package [26] in R package with default parameters.

Survival analysis

Survival data were obtained from the TCGA cohort of patients for colorectal cancer. The samples were categorized into CMS high, and CMS low groups based on median of the respective CMS scores for each CMS subtype. Kaplan-Meier analysis was performed using the R package 'survival.' A log-rank test was used to compute the *p*-values. The reported hazard ratio (HR) and confidence interval (95% CI) were determined using Cox regression using the 'coxph' function.

Additionally, colorectal cancer samples were also split into High (+) and Low (-) expression subgroups based on the median for different gene signatures (Epi, Mes, FAO, glycolysis, OXPHOS, and PD-L1) scores. The effect of the simultaneous enrichment on survival was calculated for all gene signatures in a pairwise manner. HR and p-values were depicted in forest plots created using the 'ggforest' function from the 'survminer' package.

Entropy Calculation

Cell-wise entropy values were calculated for each geneset using the following formula [27]:

$$Shannon\ entropy\ =\ -\sum P(f)ln(P(f))$$

Where P(*f*) is the ratio of the normalized expression value of a particular gene to the sum of all gene expression values for a cell. The entropy values were normalized for the number of genes in each gene set to maintain comparability.

Code availability

Codes used in this study are available at https://github.com/Soundharya-R/CRC

Data availability

Publicly available transcriptomics datasets from NCBI GEO (**Table S1**), TCGA cohorts from UCSC Xena and single-cell RNA sequencing data (GSE132465, GSE144375) were analyzed in this study.

Results

Associations between EMT, metabolism axes and immune evasion and their associations with patient survival

To assess the associations between multiple axes of plasticity governing CRC heterogeneity, we began by investigating three important aspects - cancer cell metabolism, immune evasion, and EMT. We investigated the associations among these three axes in 101 bulk transcriptomic datasets with CRC samples (**Table S1**) by calculating ssGSEA scores of associated gene signatures. We observed that among the 43 cases where the epithelial phenotype correlated significantly ($p < 0.05$)

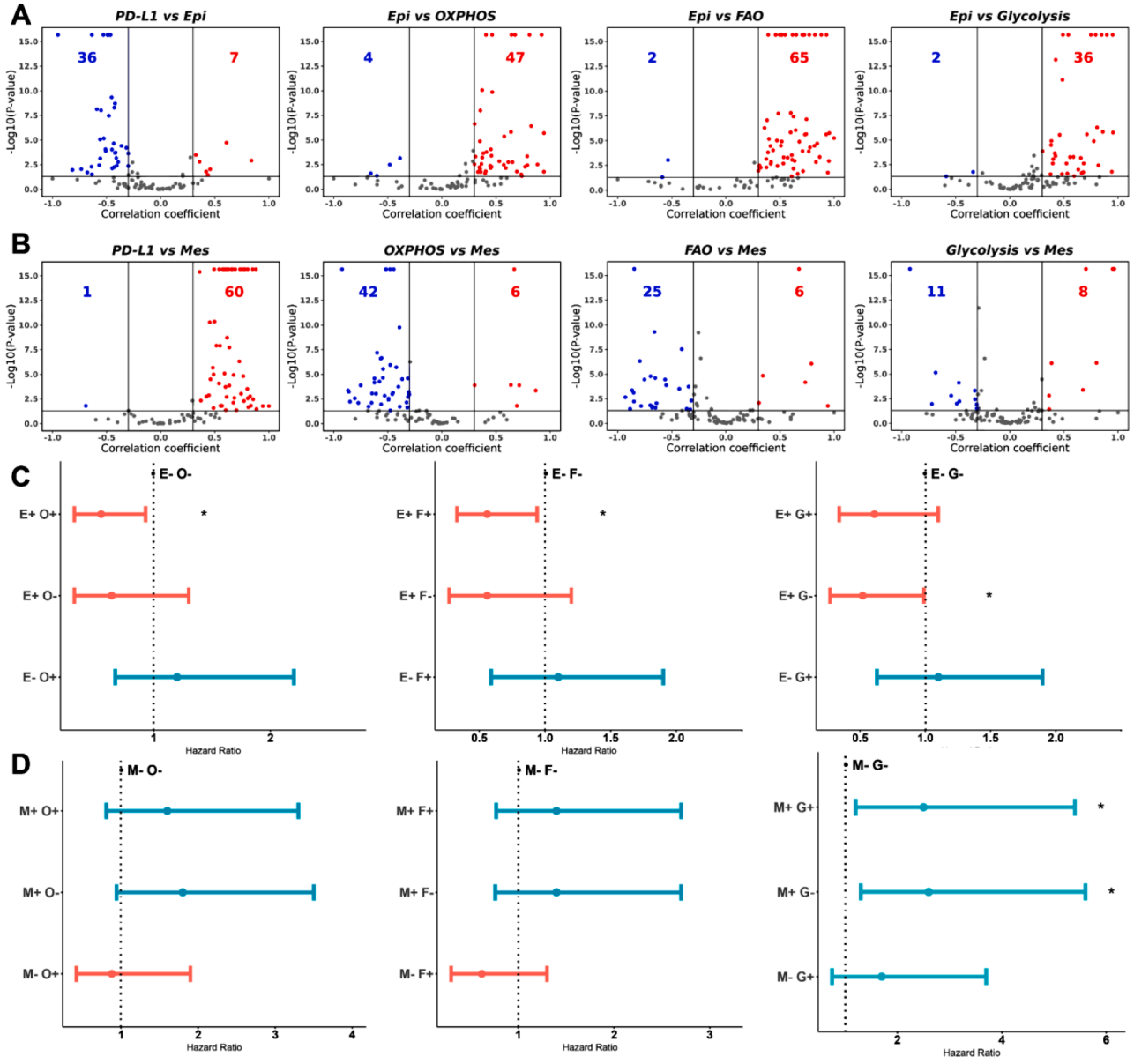

Fig. 1. Relationship between EMT, metabolism, and PD-L1 signatures in CRC bulk-level transcriptomes and differences in survival probabilities for their pairwise concurrent enrichment. A) Volcano plots depicting Spearman correlation coefficient (x-axis) and -log10(p-values) (y-axis) for PD-L1 vs. Epi scores (left), Epi vs. OXPHOS (middle-left), Epi vs. FAO (middle-right) and Epi vs Glycolysis scores (right). Boundaries for significant correlation are set at $R > \pm 0.3$ and $p < 0.05$. Red data points indicate datasets for which the association is significantly positive, blue for negative, and gray for insignificant correlation. **B)** Same as A) but for PD-L1 vs. Mes (left), OXPHOS vs. Mes (middle-left), FAO vs. Mes (middle-right), and Glycolysis vs. Mes scores (right). **C)** Forest plots depicting mean hazard ratios (HR) $\pm$ 95% confidence intervals and corresponding p-values ('*' for $p < 0.05$) for overall survival associated with simultaneous enrichment of epithelial and OXPHOS (left), epithelial and FAO (middle) and epithelial and Glycolysis signatures (right). (+) and (-) subgroups are based on median values. Mean HR values > 1 are shown in blue, while those < 1 are shown in red. **D)** same as C) but for mesenchymal and OXPHOS (left), mesenchymal and FAO (middle), and mesenchymal and Glycolysis signatures (right).

with the PD-L1 pathway, the correlation coefficient was negative in 36 cases, indicating that epithelial phenotype is largely negatively correlated with immune evasion (Fig 1A). Conversely, mesenchymal phenotype predominantly correlated positively with the PD-L1 signature (60 out of 61 cases). Epithelial and mesenchymal signatures showed a dominantly negative correlation with each other (53 out of 58 datasets), as expected (**Fig S1**). Similar antagonistic trends were seen for epithelial vs. mesenchymal signatures with the metabolic axes - while the epithelial phenotype correlated positively with oxidative phosphorylation (OXPHOS) (47 out of 51 datasets), and fatty acid oxidation (FAO) (65 out of 67 datasets), mesenchymal phenotype enrichment was negatively correlated with OXPHOS (42 out of 48 datasets) and FAO axes (25 out of 31 datasets) (Fig 1B). The trend with the glycolysis signatures was not as strong, yet antagonistic for the epithelial vs. mesenchymal signature ssGSEA scores. These trends are reminiscent of our pan-cancer analysis showing that FAO and OXPHOS were negatively correlated with EMT [28]. However, we had noticed an association between glycolysis with partial or full EMT, which is not as strongly recapitulated in CRC data. Such context-specific differences may emerge from the heterogeneity in CRC samples from a CMS context.

To understand the clinical impact of the enrichment of these diverse signatures/pathways (epithelial, mesenchymal, FAO, OXPHOS, glycolysis and PD-L1 pathway) individually and in combination, we plotted Kaplan-Meier graphs for overall survival data on the TCGA colorectal cancer patient cohort. Our evaluation of pairwise comparisons revealed that upregulation of epithelial signature along with higher expression of FAO and OXPHOS genes (E+F+ and E+O+) had significantly better overall survival probability compared to Epi-Low/FAO low group (E-F-) and Epi-Low/OXPHOS-low group (E-O-) respectively. No such trend was observed for glycolysis signature, though (Fig 1C). In contrast, the combination of high expression of glycolysis pathway genes with the mesenchymal phenotype (G+M+) has significantly worse implications for survival probability with reference (G-M-) (Fig 1D), but no trend was noticed for co-enrichment of mesenchymal with FAO or OXPHOS. Additionally, the PD-L1 signature, despite being positively correlated with mesenchymal signatures, is associated with better survival probability in CRC (**Fig S1**). This analysis revealed the strong association of epithelial phenotype with a better survival response, whereas the presence of mesenchymal phenotype with a worse prognosis, as also noted in trends for epithelial and mesenchymal signatures alone (**Fig S1C**). Further, the metabolic pathways, such as FAO and OXPHOS, that correlated positively with epithelial phenotype, had a similar effect on the hazard ratio (HR). On the contrary, the glycolysis signature and PD-L1 signature, although more strongly positively associated with epithelial phenotype and mesenchymal phenotype respectively, tend to affect the hazard ratios in opposite directions (Fig 1D, S1B). Together, these results suggest an interplay between the different interconnected axes of cellular plasticity (EMT, metabolic switching) in mediating patient survival [29].

Heterogeneity among Consensus Molecular Subtypes (CMS) and their associations with EMT and different axes of metabolism and immune evasion

The intra-tumor heterogeneity in CRC patients can be better studied by subdividing cancer samples into subtypes based on their shared features and differences at the molecular and cellular levels. The subtypes differ widely in their transcriptomic and genomic profiles while also showing differences in prevalence in parts of the colon [6].

To begin, we performed Kaplan Meier survival analysis to determine the clinical outcome for the enrichment of different CMS using patient data from TCGA database. The CMS scores were calculated with ssGSEA using the subtype-specific template signatures used by 'CMSCaller' function. We hypothesized that the samples with higher CMS2 and CMS3 scores would have better survival probabilities than those with higher CMS4 scores since the latter has upregulation of EMT pathway, correlating to poor patient outcomes [30].

The Kaplan Meier plots compared CMS-high versus CMS-low expression for the four subtypes, and we observed the hazard ratios for samples with high CMS1 and 4 scores were greater than 1 (i.e. enrichment of CMS1 or CMS4 associated with worse survival), and hazard ratios for samples with higher CMS2 and 3 scores were lower than 1. (i.e., enrichment of CMS2 or CMS3 associated with better survival) (Fig 2A). However, the trend was statistically significant only for CMS2 and CMS4. Similar analysis was done to test the enrichment of the different CMS signatures on progression free interval (PFI) and disease-free interval (DFI). The enrichment of CMS4 subtype was found to associate with worse PFI significantly, but no clear trends were noticed for DFI (**Fig S2**). Similarly, low epithelial and/or high mesenchymal scores associated with worse DFI and/or PFI (**Fig S3**).

To further understand the variability among the CMS subtypes, we looked at the correlation between well-studied signatures pertaining to metabolism, immune evasion, and EMT, in TCGA data available for CRC, in a CMS-specific manner (Fig 2). We observed that when CRC samples are not segregated by CMS, the common trends seen for the bulk dataset meta-analysis (Fig 1) hold true, i.e., a positive correlation between epithelial scores with FAO, OXPHOS and glycolysis scores, a positive correlation between mesenchymal and PD-L1 scores, and a negative correlation for Epi scores vs. Mes scores, and a negative correlation of Epi scores with PD-L1 scores (**Fig S4**). However, as we delved into examining these trends at the CMS subtype level, we saw CMS subtype-specific differences. For instance, in the case of glycolysis scores versus epithelial and mesenchymal scores (Fig 2E), the variability in the correlation coefficient values seen (both positive and negative values across 4 CMS subtypes) seem to explain their seemingly counterintuitive association with respect to epithelial and mesenchymal scores seen in the bulk data analysis noted earlier (Fig 1A-B**)**. We further noticed that epithelial and mesenchymal programs were negatively correlated even at individual subtype-level and PD-L1 signature scores associated positively with mesenchymal ones across the CMS subtypes. While, in CMS3 samples, the PD-L1 and Epi scores did not show any association, in the rest of the subtypes they were negatively correlated (Fig 2B). OXPHOS was shown to be positively linked with Epi and negatively associated with Mes in CMS1, CMS2 and CMS4 but these associations were found to be in the reverse direction in the CMS3 samples (Fig 2C). Similarly, in CMS3 and CMS4, FAO and Mes are correlated positively while in CMS1 and CMS2, they are negatively associated. However, FAO was associated positively with epithelial phenotype in all four subtypes to a similar extent (Fig 2D). Such differences in the relationship between these two key axes of plasticity (EMT, metabolic switching) at subtype level could serve as a distinguishing functional role of different CMS categories.

Different CMS subtype samples have varied status of epithelial-mesenchymal plasticity

Given the observed role of epithelial vs. mesenchymal phenotypes in determining patient survival, we assessed the extent of epithelial/mesenchymal, metabolic pathway and PD-L1 enrichment across CMS subtypes. We chose five datasets with samples corresponding to each subtype (GSE196576, GSE161158, GSE96528, GSE14333, GSE14095). The samples in each dataset were classified into the respective subtypes using the 'CMSCaller' package. Here, we consistently observed that the Epi score was the highest in CMS3 subtype, followed by CMS2, whereas the Mes scores were comparable for these two subtypes. Further, we saw that CMS4 samples had the highest Mes score and lowest Epi score among all and was followed by CMS1 samples (Fig 3A). Consistently, CMS1 and CMS4 -the two more mesenchymal subtypes – had higher levels of PD-L1 enrichment. Also, all the metabolic signatures - Glycolysis, FAO and OXPHOS - consistently showed lower activity in the CMS4 subtype (**Fig S5A-B**), reinforcing our observations from meta-analysis of EMT with metabolic axes (Fig 1A-B).

Next, we obtained two datasets, one with TGFβ treated samples

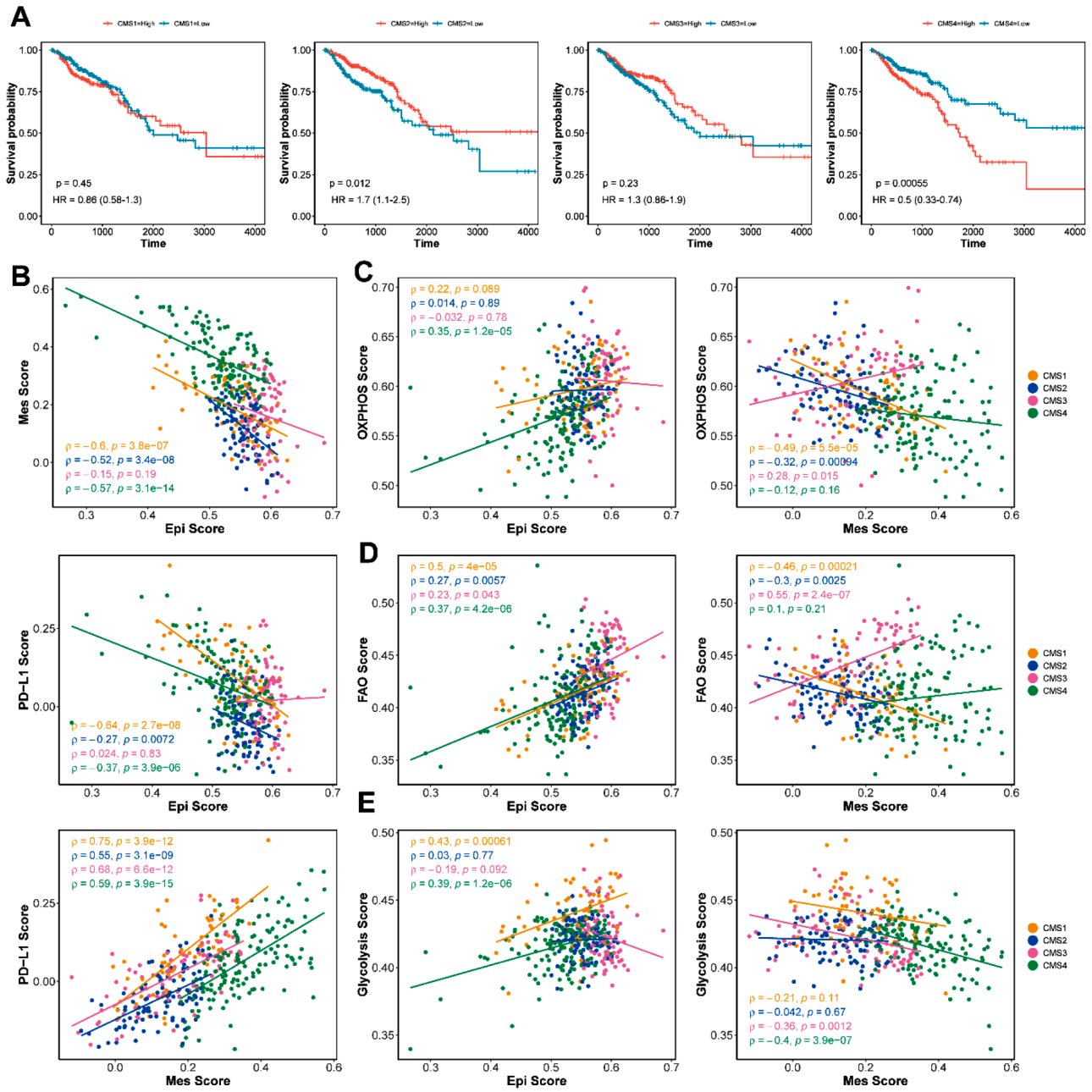

Fig. 2. CMS-specific differences in survival probabilities and associations between EMT, metabolism, and PD-L1 in TCGA colorectal cancer samples. A) Kaplan-Meier curves showing differences in survival probabilities for CMS1-high (red) and CMS1-low (blue) (left), CMS2-high and low (middle-left), CMS3-high and low (middle-right) and CMS4-high and low (right). Reported p-values are based on a log-rank test and indicate differences in survival between the subgroups. Mean hazard ratios (HR) ± 95% confidence intervals (95% CI) are shown. **B)** Scatter plot illustrating subtype-wise Epi (x-axis) and Mes (y-axis) scores (top), Epi and PD-L1 scores (middle), and Mes and PD-L1 scores (bottom) for TCGA colorectal cancer patient samples. Pearson's correlation coefficient 'ρ' and p-values for subtype-wise correlation are shown. Blue data points are for cells classified as CMS1, red for CMS2, green for CMS3, purple for CMS4. **C)** same as **B)** but for Epi and OXPHOS scores (left) and Mes and OXPHOS scores (right), **D)** Epi and FAO scores (left) and Mes and FAO scores (right), and **E)** Epi and Glycolysis scores (left) and Mes and Glycolysis scores (right).

(GSE137779) and the other having samples with overexpression of EMT-inducing transcription factor SNAI1 over time (GSE115716) to gauge the extent of influence of EMT on the CMS sub-classification score. We compared the changes in CMS3 and CMS4 when samples were treated with TGFβ or SNAI1. In SNAIL induction experiment, the LS174T colorectal carcinoma cells were made to express a dox-inducible SNAI1 transgene, that was incubated at 0, 3, 6, 24, 48, 72 or 96 hours with 0.1 μg/ml Doxycycline. Expectedly, we observed that SNAI1 overexpression causes a decrease in Epi scores and an increase in Mes scores. Intriguingly, we observed that SNAI1 overexpression led to higher CMS4 enrichment scores but reduced CMS3 scores (Fig 3D), indicating that SNAI1, a potent EMT inducer, possibly plays a role in controlling CMS plasticity. On the other hand, when SW480 colorectal cells were treated with 10 ng/ml TGF-β1, Epi scores decreased, whereas Mes scores do not show the expected increase. While projecting the same set of samples on the CMS3/CMS4 axes, we saw that TGFβ treatment seemed to increase both CMS3 and CMS4 scores, and PD-L1 activity, but decreased FAO and OXPHOS (Fig 3B-C). These examples illustrate that the mode of EMT induction dictates the extent of EMT observed as well as corresponding changes in their molecular subtyping. Another reason for this difference

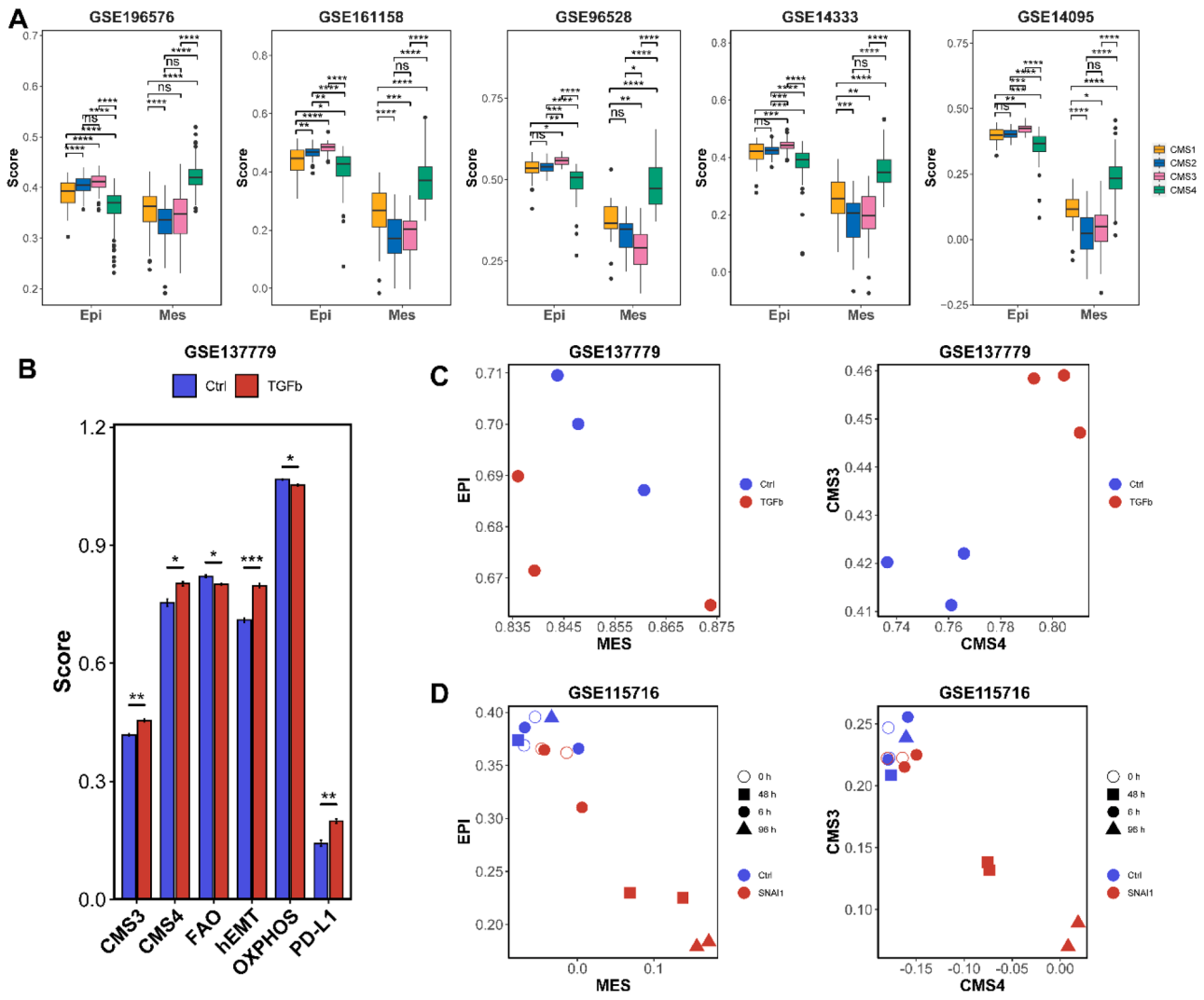

Fig. 3. Relationship between CMS and EMT induction in bulk datasets. A) Boxplots showing differences in epithelial and mesenchymal scores in a subtype-specific manner. *, **, ***, **** denote $p < 0.05$, 0.01, 0.001, 0.0001, respectively. Plot titles denote NCBI-GEO dataset IDs. **B)** Bar plots showing ssGSEA scores of CMS3, CMS4 genes, FAO, hallmark EMT, OXPHOS, and PD-L1 gene sets in control (blue) vs. TGFβ-treated samples (red). **C)** Scatterplot with Epi (y-axis) and Mes (x-axis) scores (left) and CMS3 (y-axis) and CMS4 (x-axis) scores of control and TGFβ-treated samples of GSE137779. **D)** Same as **C)** but for control samples and samples with SNA1 overexpression (GSE115716).

can be non-EMT associated changes driven by TGFβ in cellular response.

Single-cell RNA-sequencing analysis reveals CMS subtype-specific patterns of epithelial-mesenchymal heterogeneity

Our bulk-level CMS-specific investigation highlighted associations between CMS subtyping and EMT. We examined these associations at individual cell level through single-cell RNA-sequencing (scRNA-seq) datasets GSE132465 and GSE144375 [31]. The scRNA-seq data were filtered for tumor cells, and 'CMSCaller' was used to assign the appropriate CMS subtype. Only the statistically significant predictions in the context of CMS assignment were used for further analysis. We noticed all the four CMS subtypes to be well-represented in these two scRNA-seq datasets (Fig 4A**, i-ii**). First, we observed that the antagonism between epithelial and mesenchymal axes is maintained even at the single-cell level (Fig 4B**, ii-ii**) across cells belonging to all four CMS subtypes. Interestingly, we also noticed a spectrum of epithelial-mesenchymal states in this two-dimensional projection: while cells classified to belong to CMS4 cells localized in (high mesenchymal, low epithelial) area, the ones classified as CMS2 and CMS3 were centered around (low mesenchymal, high epithelial) area. The cells categorized as CMS1 occupied intermediary position, indicating a hybrid E/M phenotype with the simultaneous enrichment of both epithelial and mesenchymal characteristics.

To assess how the epithelial /mesenchymal, metabolic pathway and PD-L1 scores are distributed across the cells, we plotted the kernel density estimates of AUCell scores in a CMS subtype-wise manner. We noticed that the Epi density distribution for CMS4 shows a narrow peak centered around low Epi scores, whereas the CMS2 and CMS3 scores have wider distributions centered around higher Epi scores (Fig 4C**, i-ii, left)**. On the other hand, the CMS2 and CMS3 samples have low Mes scores with less variability (narrower peaks centered at low Mes scores), while CMS4 scores show higher average Mes scores as well as more heterogeneity in them (Fig 4C**, i-ii, right**). Interestingly, in both these datasets, CMS1 subtype distinctly showed two subpopulations in terms of their epithelial and mesenchymal scores (Fig 4C**, i-ii**). In CMS1 subtype, this bimodality was also seen in levels of PD-L1 activity and the three metabolic axes in CMS1; however, other subtypes had homogeneously low PD-L1 activity scores. Moreover, as expected, CMS4 subtype had minimal metabolic activity (**Fig S6**). Together, our results support high phenotypic variability along multiple axes in the CMS1 subtype.

In both the scRNA-seq datasets, CMS specific associations of

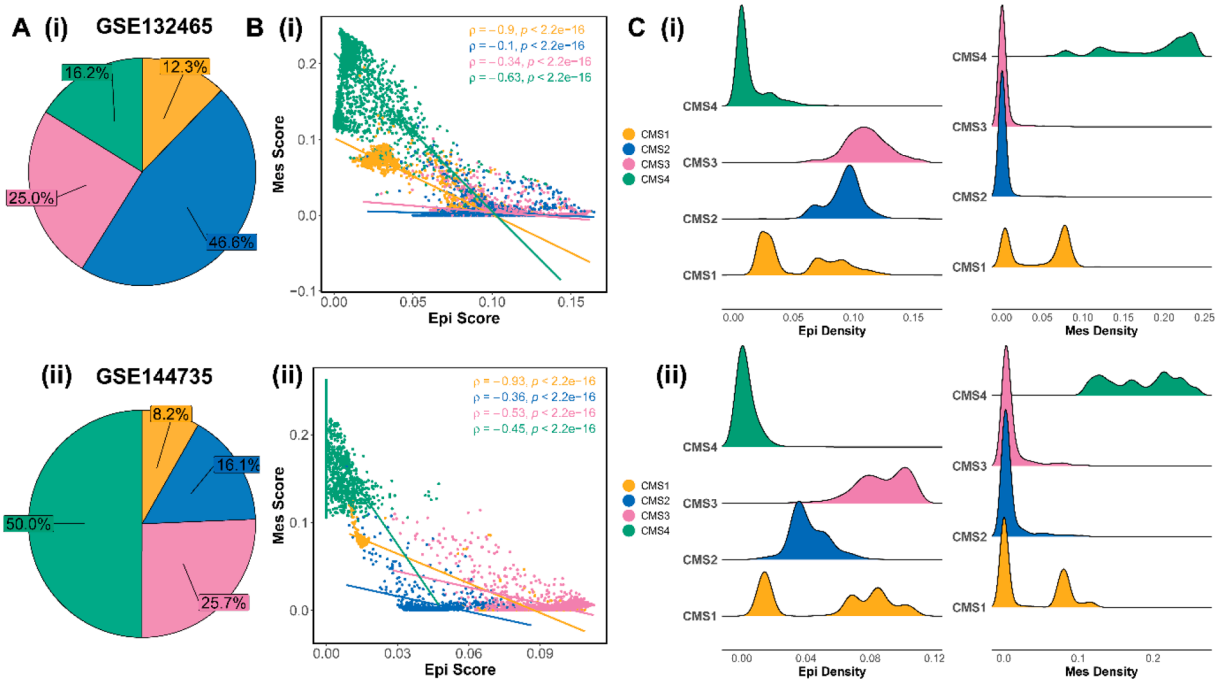

Fig. 4. Heterogeneity among CMS subtypes along the epithelial-mesenchymal axis. A) i) Pie chart illustrating the percentage of cells classified as CMS1 (yellow), CMS2 (blue), CMS3 (pink), and CMS4 (green) in GSE132465. **B) i)** Scatter plot depicting the subtype-wise association between Epi (x-axis) and Mes scores (y-axis). Pearson's correlation coefficient 'ρ' and p-values for subtype-wise correlation are shown. **C) i)** Subtype-wise kernel density estimate plots for AUCell scores of Epi (left) and Mes (right) signatures across cells in GSE132465. **A ii), B ii)**, and **C ii)** are the same as **A i), B i), C i)** respectively, but for GSE144735.

epithelial and mesenchymal programs with metabolic axes and PD-L1 were largely consistent with our earlier subtype-specific trend seen in bulk (TCGA) data. Simultaneously considering the associations between these axes of plasticity help in distinguishing similar CMS groups such as CMS1 and CMS4 (***S7-S8***). For instance, In CMS1 cells, PDL1 and Epi scores are strongly negatively correlated, but not in other subtypes (**Fig**

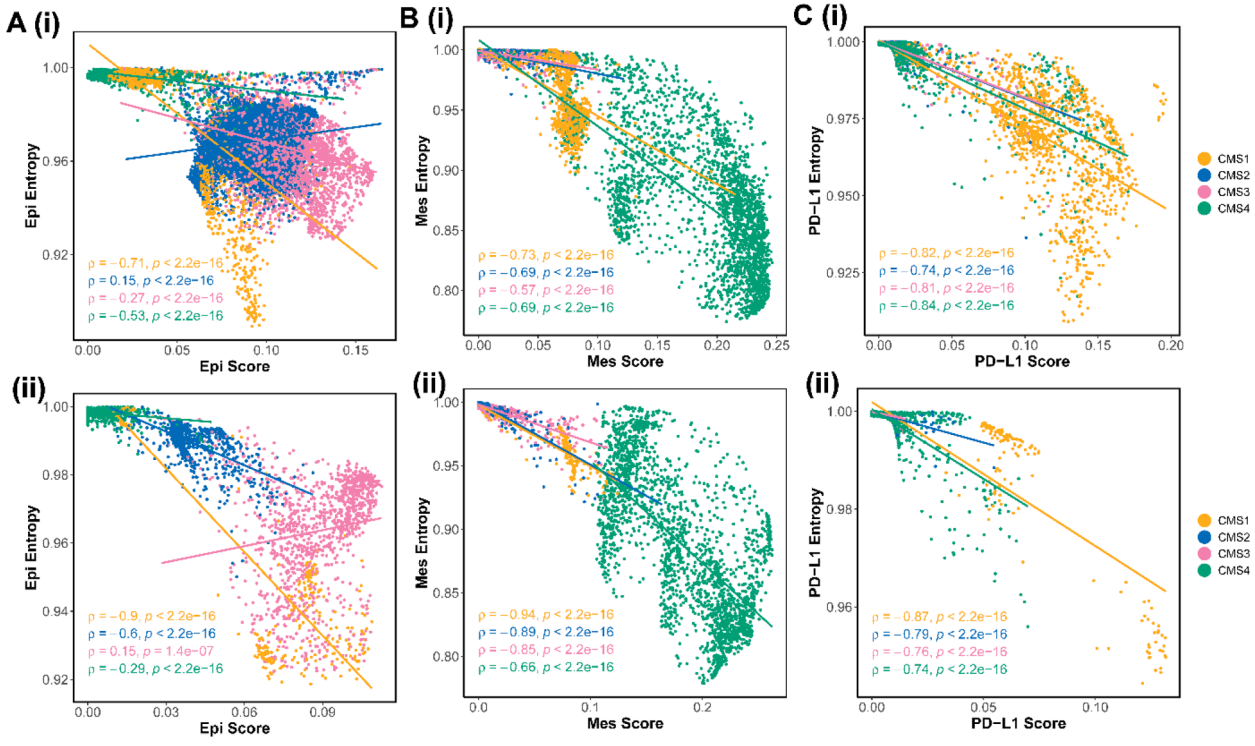

Fig. 5. CMS-specific associations between cell-wise AUCell scores and entropy values of Epi and Mes genesets at a single-cell resolution. A) i) Scatter plot depicting the association between Epi scores (x-axis) and Epi entropy values (y-axis) for GSE132465. Same as **A i)** but for **B) i)** Mes score and entropy, **C) i)** PD-L1 score and entropy. **A ii), B ii)**, and **C ii)** are the same as **A i), B i)**, and **C i)** respectively, but for GSE144735. Pearson's correlation coefficient 'ρ' and p-values for subtype-wise correlation are shown. Blue data points are for cells classified as CMS1, red for CMS2, green for CMS3, purple for CMS4.

S8D, i).

Next, we wanted to quantify the amount of heterogeneity seen across these signatures in a subtype-specific manner. We used Shannon Entropy to calculate the variability among certain genes involved in a particular pathway across different sub-populations of CRC tumor cells [32,33]. Higher entropy scores correspond to more variability in that axis for a particular cell. Cell-wise entropy values and activity scores largely showed a negative association with each other for all gene sets (Fig 5, **S9**). A possible explanation for this trend can be that once a cell acquires a particular phenotype, the genes involved in that pathway are coordinately being upregulated (or downregulated) and therefore have uniform high (or low) expression levels, reducing the underlying variability. Thus, it was unsurprising to notice the entropy for epithelial signature in CMS2 and CMS3 subtypes decreased with an increase in epithelial scores (Fig 5A**, i-ii**), potentially because those two subtypes are more epithelial relative to CMS1 and CMS4. Similarly, for CMS4, the most mesenchymal subtype, the increase in Mes scores associated with a decrease in entropy for mesenchymal signature in a cell (Fig 5B**, i-ii**). Further, in CMS1, the subtype enriched in immune activation, an increase in PD-L1 signature scores correlated with a consistent decrease in entropy of corresponding signature (Fig 5C**, i-ii**). However, the entropy of metabolic signatures did not show any CMS-specific trend, while they were also negatively correlated consistently with the corresponding signature scores (**Fig S9**).

Discussion

Colorectal cancer is one of the most heterogeneous cancers characterized by intra and inter-tumoral heterogeneity. Thus, the classification of CRC into four CMS sub-types – while a helpful metric – does not entirely depict the heterogeneity in CRC. Here, we evaluated how some key axes that drive tumor progression and metastasis in carcinomas (immune evasion, metabolic reprogramming, and EMT) to understand their heterogeneity within CMS subtypes. EMT is not a binary switch but a spectrum of states including many hybrid epithelial/mesenchymal ones [34,35] that can co-express both epithelial (E) and mesenchymal (M) markers and show mixed functional traits. Hybrid E/M cells have been shown to be positively associated with immune evasion in other cancers [36,28]. Here, we observed a positive correlation between PD-L1 signature and mesenchymal state, however, across the CMS subtypes, PD-L1 signature score was not always negatively associated with epithelial scores. Thus, the positive association of PD-L1 with both epithelial and mesenchymal signature furthers the hypothesis about hybrid E/M cells possessing these immune-evasive traits [37,24].

The spectrum of states along the EMT spectrum provides cancer cells stem-like traits, thereby facilitating aggressive tumor progression [38]. Cell populations with higher plasticity along the E/M axis tend to be more metastatic and pose significant hurdles for treatment [39,40]. In this context, our analysis of CRC samples treated with EMT-inducers, SNAI1 and TGFβ revealed context-specific changes in the CMS state of samples with a concurrent alteration in their E/M state. While SNAI1, a highly specific EMT-TF reduces the epithelial and CMS3 enrichment and increases Mes and CMS4 scores, while TGFβ downregulates genes driving epithelial phenotype and increases enrichment of both CMS3 and CMS4 genes, possibly due to its role in mediating other axes of plasticity such as metabolic state [41]. The results indicate the possibility of phenotypic switching within subtypes by inducing EMT; however further experimental validation in perhaps a larger cohort would be needed to assess its clinical impact. Our results showing CMS4 to be most mesenchymal are consistent with earlier observations about higher methylation of miR-200 family in CMS4 cell lines and tumors [42]. Further, transcriptomes of TGFβ treatment of CRC organoids resemble the CMS4 signature seen in human tumors [43].

Our results for scRNA-seq and bulk RNA-seq analysis are largely self-consistent across the CMS subtypes, such as antagonism between epithelial and mesenchymal programs, or the association of those programs with metabolic axes and PD-L1 signature enrichment scores. The different modalities of associations observed between these axes may explain the CMS subtype-specific observations of patient survival and/or sensitivity to various therapeutics and remains a key focus for our future work. A key point that our scRNA-seq analysis reveals is that CMS2 and CMS3 are relatively most epithelial, while CMS4 and CMS1 being more mesenchymal. This categorization is reminiscent of multi-omics profiling of CRC cell lines suggesting that CMS2 and CMS3 ones are more colon-like, while CMS and CMS4 ones are more undifferentiated and had higher expression of genes associated with EMT and TGFβ signaling [44]. Similarly, in TCGA CRC data, most patients from $ZEB1^{hi}$ group belonged to CMS4 subtype, while the $ZEB1^{lo}$ group was mainly composed of CMS2 and CMS3 tumors [45]. CMS4 subtype expression also correlates well with the signature of EpCAM ^{lo} sub-population in HCT116 and SW480 cells [46]. Overall, our observations highlight how the four CMS subtypes are positioned differently along the interconnected axes of phenotypic plasticity – EMT, metabolic activity, and immune-evasion – thus suggesting CMS-specific functional maps of CRC heterogeneity and pinpointing varied therapeutic vulnerabilities of different CMS subtypes.

Besides EMT, metabolic reprogramming is another key axis of cancer cell plasticity. Recent studies have shown that along with the classical Warburg effect observed in cancer cells, some cancers, including cervical and breast cancer, predominantly use OXPHOS as a primary energy source [47,48]. In fact, Colorectal cancer (CRC) cells have a higher OXPHOS rate compared to normal colon cells [49]. Through the mechanism known as Reverse Warburg effect, elements of the tumor microenvironment, such as cancer-associated fibroblasts (CAF), can regulate the OXPHOS-glycolysis metabolic switch in cancers [50]. Thus, varying microenvironments may possibly explain our counterintuitive association of glycolysis with epithelial and mesenchymal scores. Metabolism-based characterization of CRC samples has also been attempted recently [51]. Future efforts enabling more accurate classification of CRC patients into different subgroups with specified vulnerability can integrate such efforts being made to unravel phenotypic heterogeneity among multiple interconnected axes of plasticity.

CRediT authorship contribution statement

Manas Sehgal: Investigation, Data curation, Writing – original draft. **Soundharya Ramu:** Investigation, Data curation, Writing – original draft. **Joel Markus Vaz:** Investigation. **Yogheshwer Raja Ganapathy:** Investigation. **Srinath Muralidharan:** Investigation. **Sankalpa Venkatraghavan:** Investigation. **Mohit Kumar Jolly:** Funding acquisition, Supervision, Conceptualization.

Declaration of Competing Interest

The authors declare no conflict of interest.

Funding

This work was supported by Ramanujan Fellowship (SB/S2/RJN-049/2018) awarded to MKJ by Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2023.101845.

References

[1] J Ferlay, M Colombet, I Soerjomataram, et al., Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods, Int. J. Cancer 144 (2019) 1941–1953, https://doi.org/10.1002/IJC.31937.

[2] AR de Abreu, K Op de Beeck, P Laurent-Puig, et al., The position of circulating tumor DNA in the clinical management of colorectal cancer, Cancers 15 (2023) 1284, https://doi.org/10.3390/CANCERS15041284.

[3] G Sun, Y Li, Y Peng, et al., Impact of the preoperative prognostic nutritional index on postoperative and survival outcomes in colorectal cancer patients who underwent primary tumor resection: a systematic review and meta-analysis, Int. J. Colorectal Dis. 34 (2019) 681–689, https://doi.org/10.1007/S00384-019-03241-1.

[4] M El Bali, J Bakkach, M Bennani Mechita, Colorectal cancer: from genetic landscape to targeted therapy, J. Oncol. 2021 (2021), https://doi.org/10.1155/2021/9918116.

[5] S Piawah, AP Venook, Targeted therapy for colorectal cancer metastases: a review of current methods of molecularly targeted therapy and the use of tumor biomarkers in the treatment of metastatic colorectal cancer, Cancer 125 (2019) 4139–4147, https://doi.org/10.1002/CNCR.32163.

[6] S Chowdhury, M Hofree, K Lin, et al., Implications of intratumor heterogeneity on consensus molecular subtype (Cms) in colorectal cancer, Cancers (Basel) 13 (2021) 4923, https://doi.org/10.3390/CANCERS13194923/S1.

[7] R Deshmukh, M Prajapati, RK Harwansh, A review on emerging targeted therapies for the management of metastatic colorectal cancers, Med. Oncol. 40 (2023) 159, https://doi.org/10.1007/S12032-023-02020-X.

[8] YH Xie, YX Chen, JY Fang, Comprehensive review of targeted therapy for colorectal cancer, Signal Transduct. Targeted Therapy 5 (2020) 22, https://doi.org/10.1038/s41392-020-0116-z.

[9] J Guinney, R Dienstmann, X Wang, et al., The consensus molecular subtypes of colorectal cancer, Nature Med. 21 (11) (2015) 1350–1356, https://doi.org/10.1038/nm.3967.

[10] PW Eide, SH Moosavi, IA Eilertsen, et al., Metastatic heterogeneity of the consensus molecular subtypes of colorectal cancer, NPJ Genomic Med 6 (2021) 59, https://doi.org/10.1038/s41525-021-00223-7.

[11] E Fessler, JP Medema, Colorectal Cancer Subtypes: Developmental Origin and Microenvironmental Regulation, Trends Cancer 2 (2016) 505–518, https://doi.org/10.1016/J.TRECAN.2016.07.008.

[12] DG Menter, JS Davis, BM Broom, et al., Back to the colorectal cancer consensus molecular subtype future, Curr. Gastroenterol. Rep. 21 (2019) 5, https://doi.org/10.1007/S11894-019-0674-9.

[13] PW Eide, J Bruun, RA Lothe, A Sveen, CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models, Sci. Rep. 7 (1) (2017) 1–8, https://doi.org/10.1038/s41598-017-16747-x.

[14] M Rebersek, Consensus molecular subtypes (CMS) in metastatic colorectal cancer - personalized medicine decision, Radiol. Oncol. 54 (2020) 272–277, https://doi.org/10.2478/RAON-2020-0031.

[15] S Stintzing, P Wirapati, HJ Lenz, et al., Consensus molecular subgroups (CMS) of colorectal cancer (CRC) and first-line efficacy of FOLFIRI plus cetuximab or bevacizumab in the FIRE3 (AIO KRK-0306) trial, Ann. Oncol. 30 (2019) 1796–1803, https://doi.org/10.1093/ANNONC/MDZ387.

[16] A Okita, S Takahashi, K Ouchi, et al., Consensus molecular subtypes classification of colorectal cancer as a predictive factor for chemotherapeutic efficacy against metastatic colorectal cancer, Oncotarget 9 (2018) 18698–18711, https://doi.org/10.18632/ONCOTARGET.24617.

[17] O Gandrillon, M Gaillard, T Espinasse, et al., Entropy as a measure of variability and stemness in single-cell transcriptomics, Curr. Opin. Syst. Biol. 27 (2021), 100348, https://doi.org/10.1016/J.COISB.2021.05.009.

[18] PE García-Nieto, B Wang, HB Fraser, Transcriptome diversity is a systematic source of variation in RNA-sequencing data, PLoS Comput. Biol. 18 (2022), e1009939, https://doi.org/10.1371/JOURNAL.PCBI.1009939.

[19] R Berretta, P Moscato, Cancer biomarker discovery: the entropic hallmark, PLoS One 5 (2010) e12262, https://doi.org/10.1371/JOURNAL.PONE.0012262.

[20] P.V. Kharchenko, The triumphs and limitations of computational methods for scRNA-seq, Nature Methods 18 (7) (2021) 723–732, https://doi.org/10.1038/s41592-021-01171-x.

[21] D van Dijk, R Sharma, J Nainys, et al., Recovering gene interactions from single-cell data using data diffusion, Cell 174 (2018) 716–729, https://doi.org/10.1016/j.cell.2018.05.061.

[22] TZ Tan, QH Miow, Y Miki, et al., Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients, EMBO Mol. Med. 6 (2014) 1279–1293, https://doi.org/10.15252/EMMM.201404208.

[23] A Liberzon, A Subramanian, R Pinchback, et al., Molecular signatures database (MSigDB) 3.0, Bioinformatics 27 (2011) 1739, https://doi.org/10.1093/BIOINFORMATICS/BTR260.

[24] S Sahoo, SP Nayak, K Hari, et al., Immunosuppressive traits of the hybrid epithelial/mesenchymal phenotype, Front. Immunol. 12 (2021), 797261, https://doi.org/10.3389/FIMMU.2021.797261.

[25] S Aibar, CB González-Blas, T Moerman, et al., SCENIC: single-cell regulatory network inference and clustering, Nature Methods 14 (11) (2017) 1083–1086, https://doi.org/10.1038/nmeth.4463.

[26] RC Gentleman, VJ Carey, DM Bates, et al., Bioconductor: open software development for computational biology and bioinformatics, Genome Biol. 5 (2004) R80, https://doi.org/10.1186/GB-2004-5-10-R80.

[27] S Kannan, M Farid, BL Lin, et al., Transcriptomic entropy benchmarks stem cell-derived cardiomyocyte maturation against endogenous tissue at single cell level, PLoS Comput. Biol. 17 (2021), e1009305, https://doi.org/10.1371/JOURNAL.PCBI.1009305.

[28] S Muralidharan, M Sehgal, R Soundharya, et al., PD-L1 activity is associated with partial EMT and metabolic reprogramming in carcinomas, Curr. Oncol. 29 (2022) 8285–8301, https://doi.org/10.3390/CURRONCOL29110654.

[29] D Jia, JH Park, H Kaur, et al., Towards decoding the coupled decision-making of metabolism and epithelial-to-mesenchymal transition in cancer, British J. Cancer 124 (12) (2021) 1902–1911, https://doi.org/10.1038/s41416-021-01385-y.

[30] T Matsuyama, T Ishikawa, N Takahashi, et al., Transcriptomic expression profiling identifies ITGBL1, an epithelial to mesenchymal transition (EMT)-associated gene, is a promising recurrence prediction biomarker in colorectal cancer, Mol. Cancer 18 (2019) 19, https://doi.org/10.1186/S12943-019-0945-Y.

[31] HO Lee, Y Hong, HE Etlioglu, et al., Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer, Nat. Genet. 52 (2020) 594–603, https://doi.org/10.1038/S41588-020-0636-Z.

[32] AJ Conforte, JA Tuszynski, Silva FAB da, N Carels, Signaling complexity measured by shannon entropy and its application in personalized medicine, Front. Genet. 10 (2019), https://doi.org/10.3389/FGENE.2019.00930/FULL.

[33] A Karolak, S Branciamore, JS McCune, et al., Concepts and applications of information theory to immuno-oncology, Trends Cancer 7 (2021) 335–346, https://doi.org/10.1016/j.trecan.2020.12.013.

[34] MK Jolly, RJ Murphy, S Bhatia, et al., Measuring and modelling the epithelial-mesenchymal hybrid state in cancer: clinical implications, Cells Tissues Organs 211 (2022) 110–133, https://doi.org/10.1159/000515289.

[35] AR Subbalakshmi, B Ashraf, MK Jolly, Biophysical and biochemical attributes of hybrid epithelial/mesenchymal phenotypes, Phys. Biol. 19 (2022), https://doi.org/10.1088/1478-3975/AC482C.

[36] A Dongre, M Rashidian, F Reinhardt, et al., Epithelial-to-Mesenchymal transition contributes to immunosuppression in breast carcinomas, Cancer Res. 77 (2017) 3982–3989, https://doi.org/10.1158/0008-5472.CAN-16-3292.

[37] A Dongre, M Rashidian, EN Eaton, et al., Direct and indirect regulators of epithelial–mesenchymal transition– mediated immunosuppression in breast carcinomas, Cancer Discov. 11 (2021) 1286–1305, https://doi.org/10.1158/2159-8290.CD-20-0603.

[38] SA Mani, W Guo, M.J Liao, et al., The epithelial-mesenchymal transition generates cells with properties of stem cells, Cell 133 (2008) 704–715, https://doi.org/10.1016/j.cell.2008.03.027.

[39] MK Jolly, M Boareto, B Huang, et al., Implications of the hybrid epithelial/mesenchymal phenotype in metastasis, Front. Oncol. 5 (2015), 153176, https://doi.org/10.3389/FONC.2015.00155.

[40] MK Jolly, JA Somarelli, M Sheth, et al., Hybrid epithelial/mesenchymal phenotypes promote metastasis and therapy resistance across carcinomas, Pharmacol. Ther. 194 (2019) 161–184, https://doi.org/10.1016/J.PHARMTHERA.2018.09.007.

[41] X Shi, J Yang, S Deng, et al., TGF-β signaling in the tumor metabolic microenvironment and targeted therapies, J. Hematol. Oncol. 15 (1) (2022) 1–39, https://doi.org/10.1186/S13045-022-01349-6.

[42] E Fessler, M Jansen, E De Sousa, F Melo, et al., A multidimensional network approach reveals microRNAs as determinants of the mesenchymal colorectal cancer subtype, Oncogene 35 (2016) 6026–6037, https://doi.org/10.1038/ONC.2016.134.

[43] M Flum, S Dicks, YH Teng, et al., Canonical TGFβ signaling induces collective invasion in colorectal carcinogenesis through a Snail1- and Zeb1-independent partial EMT, Oncogene 41 (10) (2022) 1492–1506, https://doi.org/10.1038/s41388-022-02190-4.

[44] KCG Berg, PW Eide, IA Eilertsen, et al., Multi-omics of 34 colorectal cancer cell lines - a resource for biomedical studies, Mol. Cancer 16 (2017) 116, https://doi.org/10.1186/S12943-017-0691-Y.

[45] T Xu, M Verhagen, R Joosten, et al., Alternative splicing downstream of EMT enhances phenotypic plasticity and malignant behavior in colon cancer, eLife 11 (2022), https://doi.org/10.7554/ELIFE.82006.

[46] A Sacchetti, M Teeuwssen, M Verhagen, et al., Phenotypic plasticity underlies local invasion and distant metastasis in colon cancer, eLife 10 (2021), https://doi.org/10.7554/ELIFE.61461.

[47] I Hernández-Reséndiz, A Román-Rosales, E García-Villa, et al., Dual regulation of energy metabolism by p53 in human cervix and breast cancer cells, Biochim. Biophys. Acta 1853 (2015) 3266–3278, https://doi.org/10.1016/J.BBAMCR.2015.09.033.

[48] S Rodríguez-Enríquez, L Carreño-Fuentes, JC Gallardo-Pérez, et al., Oxidative phosphorylation is impaired by prolonged hypoxia in breast and possibly in cervix carcinoma, Int. J. Biochem. Cell Biol. 42 (2010) 1744–1751, https://doi.org/10.1016/J.BIOCEL.2010.07.010.

[49] A Kaldma, A Klepinin, V Chekulayev, et al., An in situ study of bioenergetic properties of human colorectal cancer: the regulation of mitochondrial respiration and distribution of flux control among the components of ATP synthasome, Int. J. Biochem. Cell Biol. 55 (2014) 171–186, https://doi.org/10.1016/J.BIOCEL.2014.09.004.

[50] Bonuccelli G, Whitaker-Menezes D, Castello-Cros R, et al (2010) The reverse Warburg Effect: Glycolysis inhibitors prevent the tumor promoting effects of caveolin-1 deficient cancer associated fibroblasts. 9:1960–1971. https://doi.org/10.4161/CC.9.10.11601.

[51] M Zhang, HZ Wang, RY Peng, et al., Metabolism-Associated molecular classification of colorectal cancer, Front. Oncol. 10 (2020), https://doi.org/10.3389/FONC.2020.602498/FULL.