**Analysis of immune subtypes across the epithelial-mesenchymal plasticity spectrum**

Priyanka Chakraborty[1], Emily Chen[2], Isabelle McMullens[2], Andrew J. Armstrong[2,3,4], Mohit Kumar Jolly[1,*], and Jason A. Somarelli[2, 3,*]

[1]Centre for BioSystems Science and Engineering, Indian Institute of Science, Bangalore, 560012; [2]Department of Medicine, Durham, NC; [3]Duke Cancer Institute Center for Prostate and Urologic Cancer, Durham, NC; [4]Department of Pharmacology and Cancer Biology, Duke University, Durham, NC

*authors contributed equally; address correspondence to: mkjolly@iisc.ac.in; jason.somarelli@duke.edu

**Keywords:** Gene expression signature, immune subtypes, hybrid epithelial/ mesenchymal, tumor heterogeneity, immune invasion

**Abstract**

Epithelial-mesenchymal plasticity plays a critical role in many solid tumor types as a mediator of metastatic dissemination and treatment resistance. In addition, there is also a growing appreciation that the epithelial/mesenchymal status of a tumor plays a role in immune evasion and immune suppression. A deeper understanding of the immunological features of different tumor types has been facilitated by the availability of large gene expression datasets and the development of methods to deconvolute bulk RNA-Seq data. These resources have generated powerful new ways of characterizing tumors, including classification of immune subtypes based on differential expression of immunological genes. In the present work, we combine scoring algorithms to quantify epithelial-mesenchymal plasticity with immune subtype analysis to understand the relationship between epithelial plasticity and immune subtype across cancers. We find heterogeneity of epithelial-mesenchymal transition (EMT) status both within and between cancer types, with greater heterogeneity in the expression of EMT-related factors than of MET-related factors. We also find that specific immune subtypes have associated EMT scores and differential expression of immune checkpoint markers.

## Introduction

Epithelial-to-mesenchymal transition (EMT) is a cellular process in which epithelial cells lose epithelial characteristics, such as apical-basal polarity and tight cell-cell adhesions, and gain mesenchymal features, such as anterior-posterior polarization, focal cell contacts, and enhanced motility and invasiveness (Jolly et al. 2017; Nieto et al. 2016; J. Yang et al. 2020). Initially characterized in the field of embryology as a feature of normal development, EMT is now also known to play fundamental roles in multiple cellular processes, such as wound healing, fibrosis, and cancer (Stone et al. 2016). The regulation of EMT is complex and is controlled by the combined action of core EMT transcription factors (EMT-TFs, such as Snail/Slug, Twist1, and Zeb1/2) (Nieto et al. 2016; J. Yang et al. 2020), epithelial factors (such as GRHL2 (Somarelli et al. 2016; Gao et al. 2015) and OVOL1/2 (Roca et al. 2013)), post-transcriptional regulation, including microRNAs (Paterson et al. 2008; Gregory et al. 2008) and alternative splicing (Jolly et al. 2018; Bebee et al. 2015; Warzecha et al. 2009), epigenetic modifications (Skrypek et al. 2017), and post-translational regulation (Serrano-Gomez, Maziveyi, and Alahari 2016). In addition to these cell-intrinsic mechanisms, features of the tumor microenvironment, such as hypoxia (Saxena and Jolly 2019; Choi et al. 2017; Cannito et al. 2008; M.-H. Yang et al. 2008) and interactions between cancer cells and stromal (Y. Wang et al. 2021; L. Wang et al. 2018; M.-Q. Gao et al. 2010; Giannoni et al. 2010) and immune cells (Fedele and Melisi 2020; Romeo et al. 2019; Toh et al. 2011) also promote EMT.

Early work on EMT in cancer focused on EMT as a driver of metastatic dissemination. The reversion of cancer cells from a post-EMT-like state back to an epithelial-like phenotype by mesenchymal-to-epithelial transition (MET) was shown to be important for metastatic colonization subsequent to dissemination and seeding (Esposito et al. 2019;Somarelli et al. 2016; Stankic et al. 2013; Tsai et al. 2012; Ocaña et al. 2012). In addition to metastatic potential, however, EMT has also been shown to function in several other key cancer processes, including tumor-propagating/stemness-like phenotypes (Pasani, Sahoo, and Jolly 2020; Guen et al. 2017; Li et al. 2012; D. Guo et al. 2012; Wellner et al. 2009) and treatment resistance (Fischer et al. 2015; Zheng et al. 2015). Consistent with the role of EMT in metastatic dissemination and therapy resistance, the number and EMT phenotype of circulating tumor cells (CTCs) varies depending on a patient's treatment response status, with treatment-refractory patients having

more overall CTCs with a higher proportion of mesenchymal-type CTCs, and treatment responders having fewer overall CTCs with a higher proportion of epithelial-type CTCs (Yu et al. 2013).

Although early studies on EMT and MET viewed these phenotypic transitions as binary states, as this field of study has matured the simplified view of EMT/MET dynamics and roles in cancer progression has been shown to be more complex than originally described (Nieto et al. 2016; Biswas, Jolly, and Ghosh 2019). EMT, in some contexts, has now been shown to be dispensable for metastatic dissemination (Nieto et al. 2016; Zheng et al. 2015; Fischer et al. 2015). Similarly, MET has also been demonstrated to be dispensable for metastatic colonization, with two distinct paths to metastatic colonization (Somarelli et al. 2016; Jolly et al. 2017) and predicted by (Brabletz 2012). Likewise, while the importance of these processes is more complex, the dynamics of this gene regulatory program are also more nuanced than originally appreciated. Rather than a binary switch between states, EMT/MET is now viewed as a spectrum along which cells can have various hybrid E/M phenotypes (Nieto et al. 2016; Biswas, Jolly, and Ghosh 2019; Jolly et al. 2017). In fact, hybrid E/M phenotypes may be marks of increased cancer aggressiveness and metastatic ability (George et al. 2017). This has been observed not only in cancer cell lines, but also in clinical CTC samples (George et al. 2017; Armstrong et al. 2011; Yu et al. 2013).

Just as EMT promotes chemoresistance in multiple cancer types, the EMT/MET status of cancers has also been linked to resistance to novel immunotherapies. Immune checkpoint inhibitors have revolutionized the treatment landscape and vastly improved patient outcomes of several cancer subtypes, including non-small cell lung cancer (NSCLC) (Brahmer et al. 2018), melanoma (Weiss, Wolchok, and Sznol 2019), and renal cell carcinoma (RCC) (Weiss, Wolchok, and Sznol 2019; Rini et al. 2019). However, resistance to these agents is common, and there are several cancer subtypes that do not respond well to immune checkpoint blockade.

One possible explanation for poor immunotherapy response may be the relationship between the EMT status of a tumor and the expression of immune checkpoint molecules (Cao, Zhang, Kamimura, et al. 2011). For example, mesenchymal-like cancer cells have been shown to be capable of immunosuppression via interactions with stromal immune cells in the tumor

microenvironment. In particular, Snail, a core EMT-TF, upregulates expression of CXCL2, a major mediator of myeloid-derived suppressor cell (MDSC) infiltration (Taki et al. 2018). MDSCs are major drivers of immunosuppression in the tumor microenvironment. MDSC infiltration renders cancer cells less susceptible to attack by CD8+ T cells and NK cells and leads to an increased ratio of immunosuppressive CD4+Foxp3+ Treg-like cells, which facilitates tumor growth (Qian et al. 2017). Snail1 and vimentin, a mesenchymal cytoskeletal marker, are positively correlated with programmed death-ligand 1 (PD-L1) scores (Kim et al. 2016). Furthermore, PD-L1 transcription is regulated by Snail1 and the miR200/ZEB1 axis (L. Chen et al. 2014). In contrast, E-cadherin (an epithelial marker) is negatively correlated with PD-L1 scores and epithelial-like cancer cells have higher numbers of infiltrating M1 macrophages and CD8+ T cells in the tumor microenvironment, which allows for greater susceptibility to immune checkpoint inhibitors (Dongre et al. 2017). It has also been proposed that the regulation between EMT and PD-L1 is actually bidirectional (Dongre et al. 2017; Jiang and Zhan 2020). This is supported by the observation that PD-L1 upregulates the EMT-TF, Twist (Cao, Zhang, Ritprajak, et al. 2011). In addition to this bidirectional regulation, there may be common inducers of EMT and immune evasion, including chronic inflammation, hypoxia, and metabolic reprogramming.

Our understanding of EMT, immune evasion, and other facets of cancer biology has been greatly aided by large, publicly-available gene expression datasets (Ding et al. 2018; Cancer Genome Atlas Research Network 2011). Analyses of these datasets have allowed researchers to characterize cancer types to an unprecedented level of detail. The coupling of these large genomics data sets with innovative algorithms that can uncouple bulk RNA-Seq data has also enabled inference of immune subtypes based on differential expression of immune-related genes within tumors (Eddy et al. 2020; B. Chen et al. 2018; Aran, Hu, and Butte 2017). These immune subtypes have prognostic value and can be used for survival stratification. Differences in outcome exist within and between cancers when stratified by immune subtypes.

In this study, we combined three novel EMT scoring metrics (Chakraborty et al. 2020) with immune subtype analysis from the iATLAS algorithm (Thorsson et al. 2018; Eddy et al. 2020) to understand the relationships between EMT and immune signature within and between cancer types. Our analyses demonstrate heterogeneity in both E/M phenotype and immune signature within a single cancer type. By comparing known drivers of epithelial and mesenchymal lineages

across cancers, we reveal consistency in gene expression of epithelial factors across epithelial tumors, but cancer type-specific expression of a subset of mesenchymal drivers, suggesting that epithelial-derived cancers have convergence of key epithelial driver genes while mesenchymal-derived cancers may be driven by heterogeneous expression of one or more EMT-TFs. These drivers of E/M phenotype are also associated with distinct immune subtypes, with enrichment for specific immune subtypes across the EMT spectrum, illustrating the relationships between E/M status, immune subtype, with potential implications for patient response to immunotherapy.

## Results

### The EMT status of cancers is heterogeneous within and between cancer types

The EMT status of specific tumor types has been quantified using a variety of previously-established signatures, most of which use a limited set of molecules or functional traits and/or individual algorithms to measure the extent of EMT on a continuum (Mandal et al. 2016; Schliekelman et al. 2015; Ruscetti et al. 2016; Bhatia et al. 2019; S. Tripathi et al. 2020; Devaraj and Bose 2019; Puram et al. 2017). Here, to provide a robust comparison of EMT-like status across cancer types, we calculated the EMT scores with available RNA-Seq data from the Cancer Cell Encyclopedia (CCLE) using three distinct EMT scoring algorithms (George et al. 2017; Tan et al. 2014; Byers et al. 2013). Each of these three metrics – KS, MLR and GS76 - score the extent of EMT on a continuum, based on the expression of EMT-specific genes identified by various groups. These three methods use different gene lists and scoring methods: the GS76 method uses a weighted sum of expression levels of 76 genes, the KS method compares the cumulative distribution functions of epithelial and mesenchymal signatures, and the MLR method uses a multinomial logistic regression to calculate the probability of a sample to belong to varying EMT categories. KS and MLR score samples on a scale of [-1, 1] and [0, 2], respectively, while the GS76 metric has no pre-defined scale. Higher MLR or KS scores represent more mesenchymal samples while this is the inverse for GS76 scores (lower = more mesenchymal). Thus, KS and MLR scores of samples in a given dataset correlate positively with

one another, and both KS and MLR correlate negatively with GS76 scores, as observed across multiple datasets (Chakraborty et al. 2020).

Consistent with their lineages of origin, carcinoma cell lines, such as colorectal, breast, stomach and prostate lines have lower median KS and MLR scores and higher GS76 scores as compared to mesenchymally-derived cancer cell lines, such as sarcoma, melanomas and glioma (**Figures 1A, S1A, S2A**). It is also noteworthy that the variance in EMT scores is higher in cell lines of epithelial lineages as compared to mesenchymally-derived cell lines (**Figures 1A, S1A, S2A**).

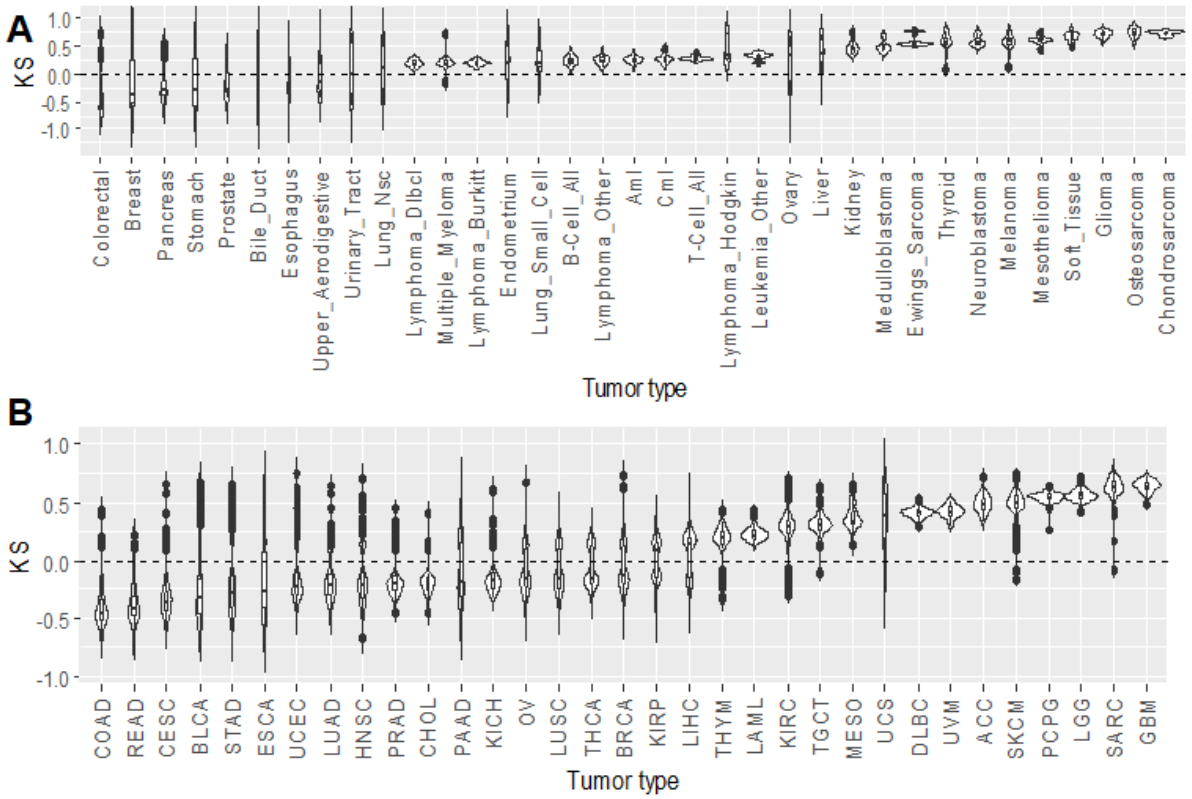



**Figure 1: Carcinomas have more epithelial EMT scores**. The KS algorithm scores samples on a scale of [-1, 1], with more positive scores representing more mesenchymal samples. **(A)** KS algorithm applied to RNA-Seq data from the Cancer Cell Encyclopedia (CCLE). **(B)** KS algorithm applied to RNA-Seq data from the Cancer Genome Atlas (TCGA); COAD = Colon adenocarcinoma, READ = Rectum adenocarcinoma, CESC = Cervical squamous cell carcinoma and endocervical adenocarcinoma, BLCA = Bladder urothelial carcinoma, STAD = stomach adenocarcinoma, ESCA = esophageal carcinoma, UCEC = uterine corpus endometrial carcinoma, LUAD = lung adenocarcinoma, HNSC = head and neck squamous cell carcinoma, PRAD = prostate adenocarcinoma, CHOL = cholangiocarcinoma, PAAD = pancreatic adenocarcinoma, KICH = kidney chromophobe, OV = ovarian serous cystadenocarcinoma, LUSC = lung squamous cell carcinoma, THCA = thyroid carcinoma, BRCA = breast invasive carcinoma, KIRP = kidney renal papillary cell carcinoma, LIHC = liver hepatocellular carcinoma, THYM = thymoma,

LAML = acute myeloid leukemia, KIRC = kidney renal clear cell carcinoma, TGCT = testicular germ cell tumors, MESO = mesothelioma, UCS = uterine carcinosarcoma, DLBC = lymphoid neoplasm diffuse large B-cell lymphoma, UVM = uveal melanoma, SKCM = skin cutaneous melanoma, PCPG = pheochromocytoma and paraganglioma, LGG = brain lower grade glioma, SARC = sarcoma, GBM = glioblastoma multiforme.

We next investigated the distribution of EMT scores across tumor types from TCGA. Similar to the CCLE data, epithelial tumors (adenocarcinomas of the colon, rectum, stomach, prostate, etc.) have lower mean value of KS and MLR scores and a higher mean value GS76 score as compared to tumors derived from mesenchymal lineages (glioblastoma, glioma, sarcoma) (**Figures 1B, S1B, S2B**). Also consistent with CCLE data, mesenchymally-derived tumors have a lower variance in scores whereas carcinomas display higher variability in EMT score (**Figures 1B, S1B, S2B**), suggesting less heterogeneity in the EMT status of mesenchymally-derived cancers. To account for any potential bias in EMT score calculation that may be due to excess stromal contamination we applied a previously-reported combination of five tumor purity prediction methods (Aran, Sirota, and Butte 2015). EMT score was weakly correlated with tumor purity (**Figure S3A**). The ESTIMATE algorithm was the most highly correlated to EMT score across cancers, with correlations of -0.47, -0.23, and 0.17 to KS, MLR, and GS76, respectively (**Figure S3A**). Other metrics of tumor purity, such as immunohistochemistry and the LUMP, ABSOLUTE, and CPE algorithms, had a maximum correlation of -0.23 (CPE vs. KS; **Figure S3A**). We also re-analyzed TCGA samples in the 50th and 75th percentiles of tumor purity. While samples with higher tumor purity tended to have lower variance, samples with higher purity were not markedly different in EMT score compared to all samples in a given cancer type (**Figure S3B**). Together, these analyses suggest that, while single samples may be influenced to a moderate extent by tumor purity, EMT score for a given cancer type is driven predominantly by tumor cell lineage and not a consequence of high stromal contamination.

**EMT-TFs display heterogeneous patterns across cancers while MET factors are more highly correlated**

Given the heterogeneity in EMT scores within and between carcinomas, we asked if this heterogeneity was correlated with differences in EMT-TFs and MET-associated factors. To address this question we assessed the levels of a panel of five well-studied EMT-inducing or

EMT-associated factors – SNAI1, SNAI2, ZEB1, ZEB2, and TWIST1 (Subbalakshmi et al. 2021; J. Yang et al. 2004; Drápela et al. 2020; Vandewalle 2005) – and five MET-inducing or MET-associated factors – GRHL2, OVOL1, OVOL2, ESRP1, and ESRP2 (Saxena et al. 2020; Cieply et al. 2012; Jolly et al. 2018; Jeong et al. 2017; Fici et al. 2017) – across tumor types. From this analysis we noted the emergence of two large clusters: one set of tumors comprised predominantly of carcinomas, with higher levels of MET factors and low levels of EMT factors (top cluster, **Figure 2A**), and another set of tumors with the opposite trend, comprised mostly of mesenchymally-derived cancers (glioblastoma, glioma, and sarcoma). We also noted a distinct difference in the relationship between MET factors and EMT factors in these groups: In the carcinoma subset MET factors are all highly expressed, while this consistency across EMT factors is not observed in the predominantly-mesenchymal cancers. Instead, mesenchymal tumors are characterized by upregulation of one predominant EMT-factor, such as SNAI2 for uveal melanoma and TWIST1 for sarcoma (**Figure 2A**). Conversely, while the mesenchymal-like cluster display consistently low levels of MET factors, the epithelial-like cluster displays high expression of EMT factors in several cases. Principal component analysis indicated that expression of MET factors is the predominant contributor to the formation of the epithelial-like and mesenchymal-like clusters, while the EMT factors alone are unable to segregate clusters by epithelial/mesenchymal lineage (**Figure S4**).

To further analyse this trend quantitatively, we calculated the pairwise correlations between expression values of all 10 EMT and MET factors across a subset of tumor types. Consistent with our qualitative observations above, the five MET factors are significantly positively correlated with each other, while the EMT and MET factors are significantly negatively correlated. However, unlike the MET factors, the correlation among different EMT factors is less consistent and often not statistically significant (**Figure 2B**). We also quantified the mean of the correlation coefficients for all pairwise correlations between any two EMT factors or for any two MET factors. These values are shown as x- and y-axes on a scatter plot, where each dot represents a tumor type. The majority of the tumors are above the x=y line, signifying that the average correlation between any two MET factors is greater than that between two EMT factors (**Figure 2C**). Similarly, visualization of the variance of the pairwise correlation coefficients for EMT and MET factors revealed a lower variability in MET factors as compared to EMT factors (**Figure S5**). These trends are also observed in CCLE cancer types (**Figures S6, S7**). Together,

these analyses suggest that more heterogeneity in gene expression exists for EMT factors as compared to MET factors.

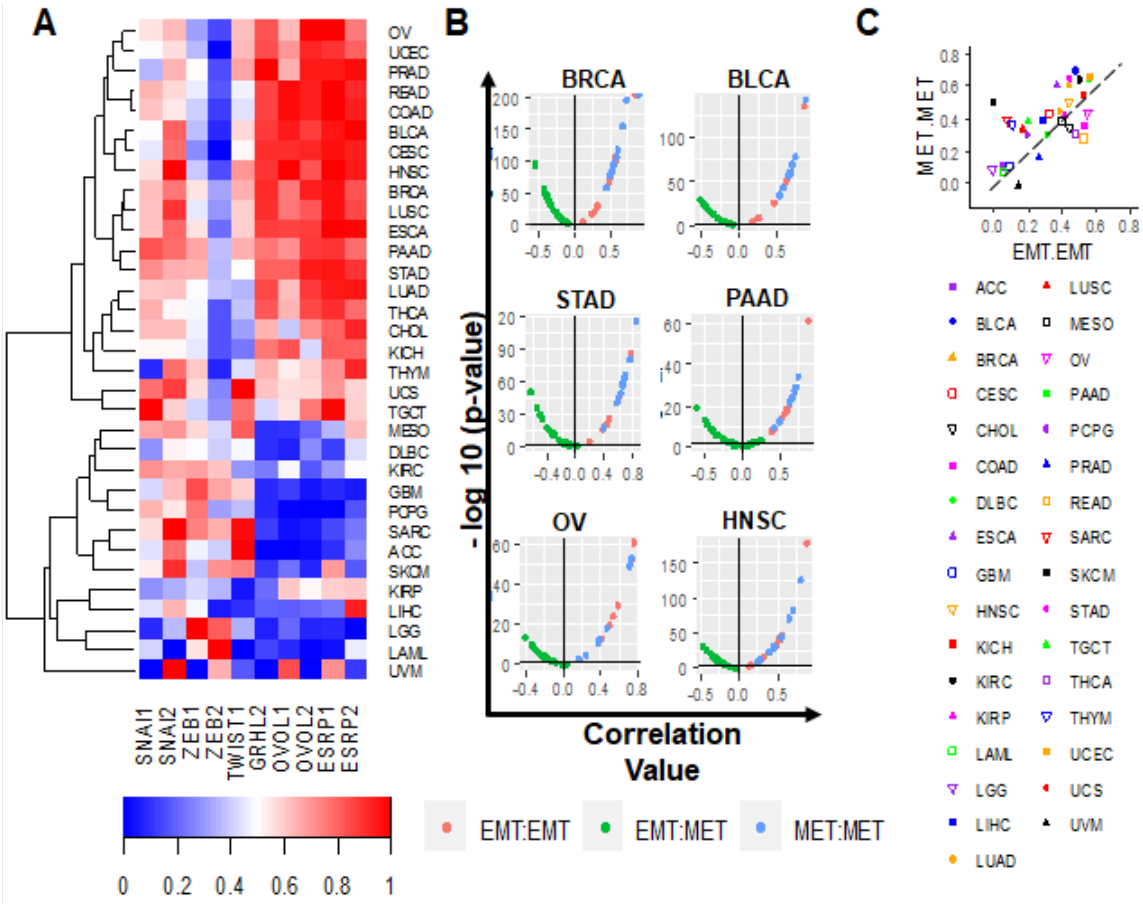



**Figure 2: EMT factors are more heterogeneous across cancers than MET factors.** EMT and MET marker expression across TCGA tumor types; COAD = Colon adenocarcinoma, READ = Rectum adenocarcinoma, CESC = Cervical squamous cell carcinoma and endocervical adenocarcinoma, BLCA = Bladder urothelial carcinoma, STAD = stomach adenocarcinoma, ESCA = esophageal carcinoma, UCEC = uterine corpus endometrial carcinoma, LUAD = lung adenocarcinoma, HNSC = head and neck squamous cell carcinoma, PRAD = prostate adenocarcinoma, CHOL = cholangiocarcinoma, PAAD = pancreatic adenocarcinoma, KICH = kidney chromophobe, OV = ovarian serous cystadenocarcinoma, LUSC = lung squamous cell carcinoma, THCA = thyroid carcinoma, BRCA = breast invasive carcinoma, KIRP = kidney renal papillary cell carcinoma, LIHC = liver hepatocellular carcinoma, THYM = thymoma, LAML = acute myeloid leukemia, KIRC = kidney renal clear cell carcinoma, TGCT = testicular germ cell tumors, MESO = mesothelioma, UCS = uterine carcinosarcoma, DLBC = lymphoid neoplasm diffuse large B-cell lymphoma, UVM = uveal melanoma, SKCM = skin cutaneous melanoma, PCPG = pheochromocytoma and paraganglioma, LGG = brain lower grade glioma, SARC = sarcoma, GBM = glioblastoma multiforme **(A)** Normalized EMT and MET marker expression across all TCGA tumor types. **(B)** Pairwise correlations between expression values of all EMT-EMT, MET-MET, and EMT-MET factor pairs across a subset of TCGA tumor types. **(C)** Plot of mean pairwise correlation

coefficients for all EMT-EMT factor pairs (x-axis) versus mean pairwise correlation coefficients for all MET-MET factor pairs (y-axis) across all TCGA tumor types.

## EMT score is associated with specific immune subtypes in cancer

Activation of a partial or complete EMT has been associated with immune evasion across multiple cancers (S. C. Tripathi et al. 2016; Terry et al. 2017). A recent approach characterized immune tumor microenvironment across 33 cancer types in TCGA and defined six major immune subtypes spanning cancer types and molecular subtypes – C1 (wound healing), C2 (IFN-γ dominant), C3 (inflammatory), C4 (lymphocyte depleted), C5 (immunologically quiet), C6 (TGF-β dominant) (Thorsson et al. 2018). Given the relationship between EMT and immune checkpoint molecules (Terry et al. 2017; Kim et al. 2016; L. Chen et al. 2014) as well as immunomodulatory cytokines, we sought to understand the relationship between EMT status and immune subtype within and across cancers. To do this, we calculated EMT scores for all TCGA samples for which immune subtypes have been assigned. Overall, cancers with the C1 wound healing subtype and C2 IFN-γ dominant subtype tended to have more epithelial-like scores, while all other subtypes were more mesenchymal (**Figures 3A, S8A, C**). In particular, the C5 quiescence subtype is the most mesenchymal subtype (**Figure 3A-B, S8**). This observation is consistent with its sample composition, as C5 contained almost exclusively low-grade glioma samples (Thorsson et al. 2018). A leave-one-out analysis further demonstrated these trends, with the wound healing (C1) and IFN-γ (C2) signatures more epithelial than all other cancer samples, and the other immune signatures more mesenchymal than other signatures/samples (**Figures 3B, S8B, D**). We applied Singscore, a recently-developed rank-based single sample method to quantify the epithelial and mesenchymal scores separately (Foroutan et al. 2018) (**Figure 3C**). Singscore is a non-parametric method that calculates scores for ensembles of gene sets corresponding to epithelial and mesenchymal phenotypes. Using this method, the C5 samples showed high mesenchymal scores and very low epithelial scores. Interestingly, the mesenchymal scores for all six immune subtypes do not vary much while the epithelial score is the lowest for C5 (**Figure 3D**). The C6 TGF-β dominant immune subtype shows a high score for both the axes, which may suggest that cancers with this immune signature would have a more hybrid epithelial/mesenchymal (E/M) phenotype (**Figure 3D**).

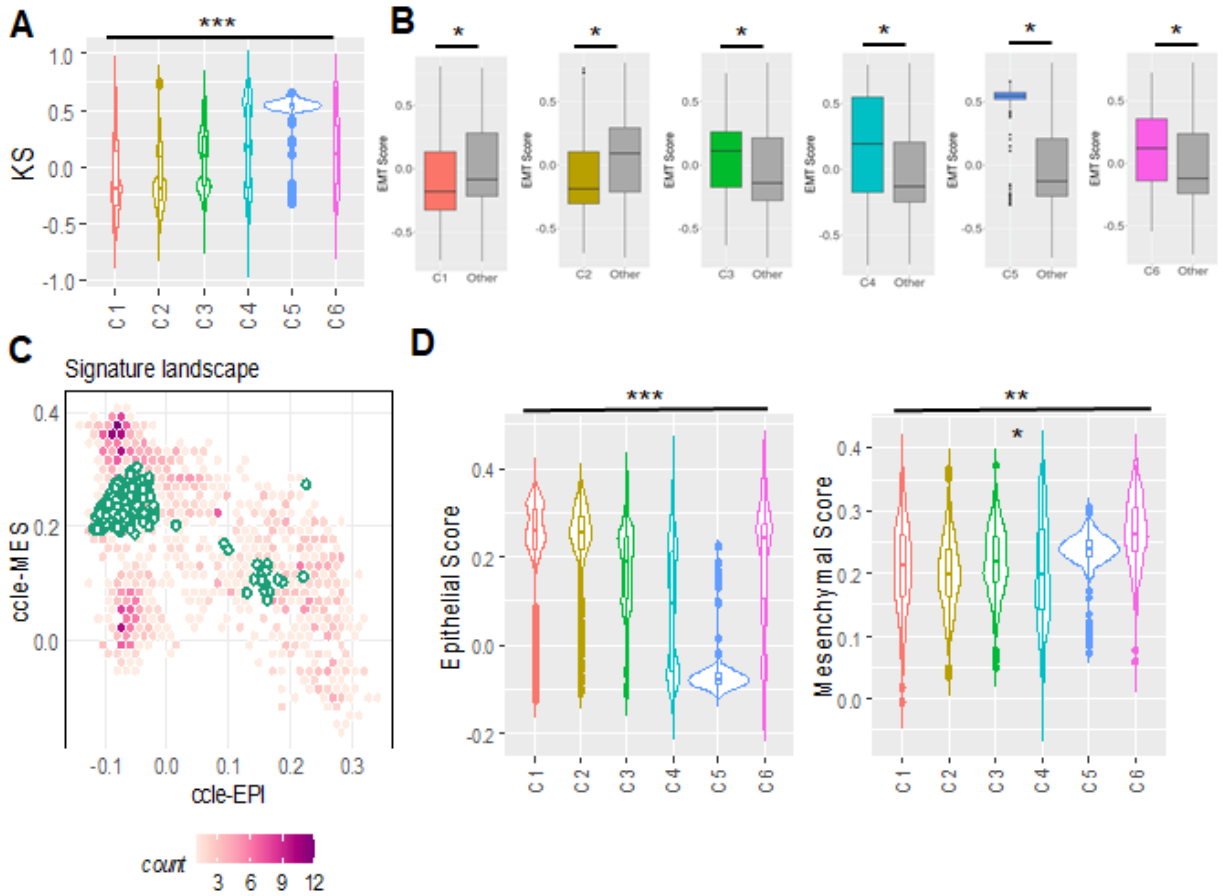



**Figure 3: Immune subtypes are associated with EMT scores.** EMT scores across TCGA immune subtypes; C1 = wound healing, C2 = IFN-γ dominant, C3 = inflammatory, C4 = lymphocyte depleted, C5 = immunologically quiet, C6 = TGF-β dominant **(A)** Plot of calculated KS score across all immune subtypes **(B)** Leave-one-out-analysis: pairwise comparison of each cancer immune subtype's KS score to the KS scores of all other immune subtypes **(C)** Application of Singscore to C5 immune subtype samples to quantify epithelial and mesenchymal scores separately. **(D)** Singscore epithelial and mesenchymal scores across immune subtypes.

Given the marked heterogeneity in EMT scores and immune subtypes, we sought to understand whether particular cancer types are enriched in specific EMT scores and immune subtypes. Analysis of the upper and lower quartiles of EMT score across immune subtypes revealed distinct cancer types within each of the immune subtypes. For example, tumors in the upper quartile of the C1 wound healing subtype are enriched in colorectal cancer specimens as compared to the lower quartile of C1 (**Figure S9A**). Conversely, the composition of cancer types between upper quartile of EMT scores within the IFNγ-dominant C2 subtype are spread across

multiple cancer types of diverse lineages, including, but not limited to, breast cancer, head and neck cancer, sarcomas, melanomas, and testicular germ cell tumors (**Figure S9B**). Likewise, the cancer types in upper and lower quartiles of EMT scores within the inflammatory (C3) subtype also differ substantially. More epithelial tumors within the C3 subtype (inflammatory) are enriched in renal clear cell carcinomas as compared to the lower quartile of C3 tumors, which is comprised of lung, prostate, breast, and thyroid cancers (**Figure S9A, B**).

To further understand how EMT and MET factors may be associated with specific immune signatures we performed separate analyses of EMT factors for each immune subtype. EMT factors Snail, Slug and Twist are most highly expressed in the C6 TGF-β dominant immune signature, with ZEB1 and ZEB2 most upregulated in the C5 immunologically quiescent signature (**Figure 4A-E**). Other immune subtypes have low to moderate levels of EMT factors, with wide distributions in expression across these subtypes. MET factors are predominantly expressed in C1-C3 signatures, with a bimodal distribution of expression in C4, low expression in C5, and relatively high expression in C6 (**Figure 4F-J**). Such co-expression of EMT and MET factors in C6 samples suggests that samples in the TGF-β dominant C6 immune subtype correspond to a more hybrid epithelial/ mesenchymal phenotype(s).

We also calculated the pairwise correlations between the EMT scores from the different scoring metrics (GS76, KS, MLR) and expression of EMT and MET factors from samples in each immune subtype. Across all immune signatures, these pairwise correlations showed consistent trends: GS76 scores correlated positively with levels of MET-factors and negatively with those of EMT-factors; KS and MLR scores followed the inverse trend (**Figure S10**). Among the EMT factors, ZEB1 and ZEB2 had the highest correlations with all EMT scoring metrics (**Fig S10**). The upregulation of ZEB1 and ZEB2 in C5, and the strong correlations of ZEB1 and ZEB with EMT scoring metrics further supports observations about the key roles of these two EMT-TFs in maintenance of a mesenchymal phenotype (Addison et al. 2021).

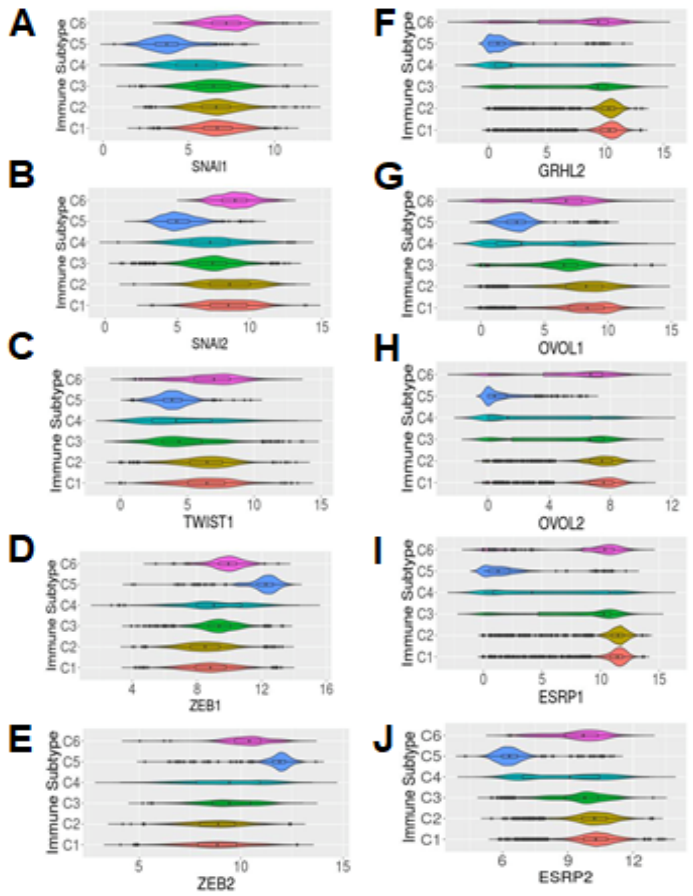



**Figure 4: EMT and MET factor gene expression varies across immune subtypes. (A-E)** Expression of EMT factors across immune subtypes; **(F-J)** Expression of MET factors acorss immune subtypes; **(A)** SNAI1, **(B)** SNAI2, **(C)** TWIST1, **(D)** ZEB1, **(E)** ZEB2, **(F)** GRHL2, **(G)** OVOL1, **(H)** OVOL2, **(I)** ESRP1, and **(J)** ESRP2 expression.

## Immune checkpoint markers across immune subtypes

We next investigated the levels of different immune checkpoint markers across the immune subtypes – CTLA-4 (cytotoxic T lymphocyte antigen-4), CD274 (PD-L1; programmed death-ligand 1) (Wei, Anang, and Sharma 2019), LAG3 (lymphocyte activation gene 3) (Solinas et al. 2019), CD276 (B7-H3) (S. Yang, Wei, and Zhao 2020), CD47 (cluster of differentiation 47) (Liu et al. 2015), and HAVCR2 (TIM-3) and its ligand LGASL9 (galectin 9) (Holderried et al. 2019) (**Figure 5A-G**). While expression of most immune checkpoint molecules varies across all immune subtypes, the immunologically quiescent C5 immune subtype has the lowest levels of CTLA-4, CD274, and CD276, but with HAVCR2 and CD47 expression similar to other immune

subtypes (**Figure 5**). The TGF-β-dominant subtype (C6) displays elevated HAVCR2 and LGALS9 (**Figure 5**). The expression levels of these immune checkpoint markers are positively correlated with one another as well as with the single-sample GSEA scores for EMT and partial EMT (Puram et al. 2017) signatures (**Figure 5H**). These results suggest that common signalling pathways implicated in EMT may be associated with changes in levels of various immune checkpoint markers.

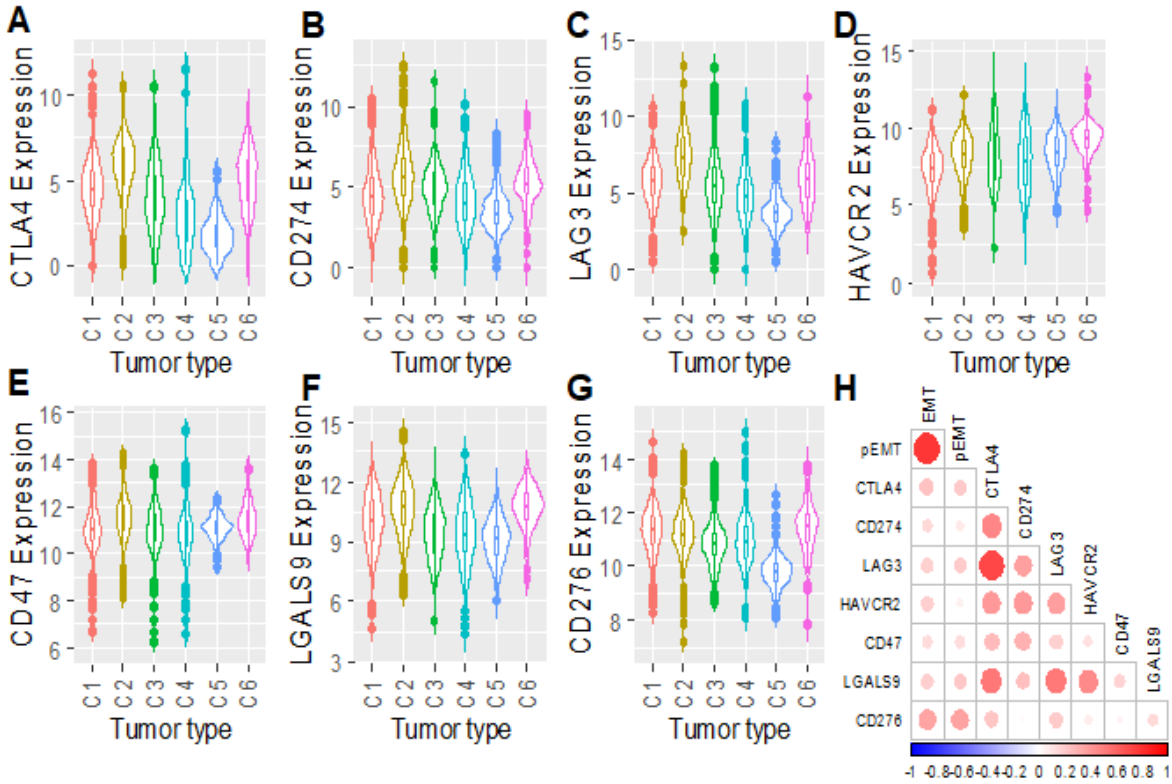



**Figure 5:** Immune checkpoint expression varies across immune subtypes. **(A)** CTLA4, **(B)** CD274, **(C)** LAG3, **(D)** HAVCR2, **(E)** CD47, **(F)** LGALS9, and **(G)** CD276 expression across immune subtypes. **(H)** Correlation matrix of EMT and pseudo-EMT scores with immune checkpoint markers.

## Discussion

Applying three distinct EMT scoring algorithms – KS, MLR, and GS76 – we characterized the diversity and heterogeneity of EMT status both within and across cancer types. These analyses revealed a higher variance in EMT score among cancers of an epithelial lineage as compared to mesenchymally-derived tumors. Similarly, analysis of EMT- and MET-associated factors revealed more heterogeneity in gene expression for EMT factors as compared to MET factors. While these data suggest that the EMT program is more heterogeneous than the MET program across cancers, the reasons for this are not clear. One possible explanation for this may be that the selected MET factors, such as GRHL2, OVOL1, and OVOL2 are indicative of epithelial lineages (Jolly et al. 2016; Hong et al. 2015; Lee et al. 2014), while the selected EMT factors may be more unique to a specific cellular lineage. It is possible that greater heterogeneity in the MET program would perhaps be more accurately reflected by analysing tissue-specific cadherins and keratins that mark specific lineages (Wahl and Spike 2017).

Another possible explanation may be that the MET program represents a more fixed, derived phenotype while the EMT program is more plastic in nature. This notion is supported by the observations that the EMT program is coupled to cancer stemness-like pathways (Mani et al. 2008; W. Guo et al. 2012; Wellner et al. 2009). Although it is possible that EMT scores may be skewed by differences in tumor:stromal ratios across samples and cancer types (Aran, Sirota, and Butte 2015; L. Wang et al. 2018), our analyses revealed low correlations between EMT score and tumor purity scores (Aran, Sirota, and Butte 2015), suggesting that the EMT scores were not substantially skewed by high stromal content. In addition, the scoring metrics were also consistent when applied to both TCGA and CCLE data, suggesting that these scoring metrics can be applied to both potentially-heterogeneous cancer samples and more homogeneous cancer cell lines. The remarkable heterogeneity in EMT and MET scores and EMT/MET factors across cancers underscores the concept of EMT/MET dynamics as a spectrum of phenotypic states (Nieto et al. 2016; J. Yang et al. 2020).

The present work sheds further light on our collective understanding of the potential cross-talk between cancer cells and immune subsets across the EMT spectrum. Both EMT scores and known EMT/MET factors were associated with certain immune subtypes. For example, cancers

with C1 (wound healing) and C2 (IFN-γ dominant) subtypes tended to have more epithelial scores with high expression of MET factors, C5 (quiescent) was dominated by a high mesenchymal score with upregulation of ZEB1 and ZEB2, and C6 (TGF-β dominant) was enriched for a more hybrid epithelial/mesenchymal score. Consistent with this, the TGF-β dominant subtype shows upregulation of ZEB1 and ZEB2 as well as GRHL2, ESRP1, and ESRP2 (**Figure 4**). The hybrid E/M state plays an important role in therapeutic resistance and formation of highly tumorigenic CTC clusters (Jolly et al. 2016; Jolly 2015). TGF-β has been shown to activate ZEB1 through DNA methylation of the ZEB1 repressive microRNA, miR-200 (Gregory et al. 2011). Our prior work indicated that both GRHL2 and miR200s are necessary to induce MET in sarcomas(Somarelli et al. 2016), and it is possible that the TGF-β-driven subtype requires additional signals to drive the phenotype toward a more complete EMT. One limitation of this study is that the analysis of TCGA data was from almost exclusively primary tumor sites rather than metastases. Metastatic samples comprise just 3.4% of all TCGA solid tumor samples (https://portal.gdc.cancer.gov/). Given the considerable differences in EMT biology between primary tumors and metastases (Tsai et al. 2012; Ocaña et al. 2012; Bonnomet et al. 2012; Somarelli et al. 2016), it is likely that the biology of the immune subsets within metastatic microenvironments also differs.

Analysis of seven immune checkpoint molecules – CTLA-4, CD274/PD-L1, LAG3, CD276/B7-H3, CD47, HAVCR2/TIM-3, and LGAL29/galectin-9 – across the immune subtypes revealed the highest levels among C2 (IFN-γ dominant) and C6 (TGF-β dominant) subtypes and lowest in the C5 (quiescent) subtype. IFN-γ is a potent inducer of PD-L1 expression (Zou, Wolchok, and Chen 2016), which may explain the high levels of immune checkpoint molecules in the C2 subtype despite its more epithelial-like score. In addition, epithelial-like cancer cells are known to have higher numbers of infiltrating M1 macrophages and CD8+ T cells in the tumor microenvironment (Dongre et al. 2017), which is consistent with the observation that the C2 (IFN-γ dominant) subtype had the highest M1/M2 macrophage polarization and strong CD8 signal among all six subtypes (Thorsson et al. 2018). The TGF-β dominant subtype (C6) displayed increased levels of HAVCR2 (also known as Tim-3), and LGALS2 (galectin-9). Interestingly, HAVCR2 expression is upregulated by TGF-β (Yan et al. 2015; Wiener et al. 2007), and galectin-9 promotes TGF-β-induced signalling and subsequent conversion of CD4+CD25- T cells into regulatory T-cells(Lv et al. 2013; Wu et al. 2014). These observations

suggest that patients enriched for the C6 subtype may benefit from TGF-β inhibitors in combination with HAVCR2 or galectin-9 suppression.

## Methods

All analyses were performed in R (version 3.4.4), and data was plotted using the ggplot2 package. Heatmaps were plotted using the gplots R package.

**TCGA datasets**: TCGA gene expression datasets were obtained from https://xenabrowser.net/datapages/

**CCLE dataset**: CCLE gene expression data was downloaded from https://portals.broadinstitute.org/ccle/data

**EMT scoring:** Three different EMT scoring methods – KS, MLR, GS76 – were used to score samples separately in all the datasets as previously described (Chakraborty et al. 2020).

**ssGSEA analysis:** ssGSEA analysis for various different gene sets were performed using GSVA R Bioconductor package with "ssgsea" option for method argument (Barbie et al. 2009).

**Statistical analysis:** All the pairwise comparison significance was tested using student's t-test and the multiple group comparisons significance was tested using ANOVA.

**Min-max standardization**: The gene expression values and EMT scores were standardized in the range of 0 to 1 as following:

Where, $X_{scaled}$ is the min-max standardized value of a gene $X$.

**Principal component analysis:** Principal component analysis was used to visualize the gene expression data of multiple variables (5 EMT and/or 5 MET factors). multidimensional gene expression and simulation data. To determine the correlation between variables and the representation of variables by the principal components, a correlation circle with squared cosines was plotted.

**Figure legends**

**Figure 1: Carcinomas have more epithelial EMT scores**. The KS algorithm scores samples on a scale of [-1, 1], with more positive scores representing more mesenchymal samples. **(A)** KS algorithm applied to RNA-Seq data from the Cancer Cell Encyclopedia (CCLE). **(B)** KS algorithm applied to RNA-Seq data from the Cancer Genome Atlas (TCGA); COAD = Colon adenocarcinoma, READ = Rectum adenocarcinoma, CESC = Cervical squamous cell carcinoma and endocervical adenocarcinoma, BLCA = Bladder urothelial carcinoma, STAD = stomach adenocarcinoma, ESCA = esophageal carcinoma, UCEC = uterine corpus endometrial carcinoma, LUAD = lung adenocarcinoma, HNSC = head and neck squamous cell carcinoma, PRAD = prostate adenocarcinoma, CHOL = cholangiocarcinoma, PAAD = pancreatic adenocarcinoma, KICH = kidney chromophobe, OV = ovarian serous cystadenocarcinoma, LUSC = lung squamous cell carcinoma, THCA = thyroid carcinoma, BRCA = breast invasive carcinoma, KIRP = kidney renal papillary cell carcinoma, LIHC = liver hepatocellular carcinoma, THYM = thymoma, LAML = acute myeloid leukemia, KIRC = kidney renal clear cell carcinoma, TGCT = testicular germ cell tumors, MESO = mesothelioma, UCS = uterine carcinosarcoma, DLBC = lymphoid neoplasm diffuse large B-cell lymphoma, UVM = uveal melanoma, SKCM = skin cutaneous melanoma, PCPG = pheochromocytoma and paraganglioma, LGG = brain lower grade glioma, SARC = sarcoma, GBM = glioblastoma multiforme.

**Figure 2: EMT factors are more heterogeneous across cancers than MET factors.** EMT and MET marker expression across TCGA tumor types; COAD = Colon adenocarcinoma, READ = Rectum adenocarcinoma, CESC = Cervical squamous cell carcinoma and endocervical adenocarcinoma, BLCA = Bladder urothelial carcinoma, STAD = stomach adenocarcinoma, ESCA = esophageal carcinoma, UCEC = uterine corpus endometrial carcinoma, LUAD = lung adenocarcinoma, HNSC = head and neck squamous cell carcinoma, PRAD = prostate adenocarcinoma, CHOL = cholangiocarcinoma, PAAD = pancreatic adenocarcinoma, KICH = kidney chromophobe, OV = ovarian serous cystadenocarcinoma, LUSC = lung squamous cell carcinoma, THCA = thyroid carcinoma, BRCA = breast invasive carcinoma, KIRP = kidney renal papillary cell carcinoma, LIHC = liver hepatocellular carcinoma, THYM = thymoma, LAML = acute myeloid leukemia, KIRC = kidney renal clear cell carcinoma, TGCT = testicular

germ cell tumors, MESO = mesothelioma, UCS = uterine carcinosarcoma, DLBC = lymphoid neoplasm diffuse large B-cell lymphoma, UVM = uveal melanoma, SKCM = skin cutaneous melanoma, PCPG = pheochromocytoma and paraganglioma, LGG = brain lower grade glioma, SARC = sarcoma, GBM = glioblastoma multiforme **(A)** Normalized EMT and MET marker expression across all TCGA tumor types. **(B)** Pairwise correlations between expression values of all EMT-EMT, MET-MET, and EMT-MET factor pairs across a subset of TCGA tumor types. **(C)** Plot of mean pairwise correlation coefficients for all EMT-EMT factor pairs (x-axis) versus mean pairwise correlation coefficients for all MET-MET factor pairs (y-axis) across all TCGA tumor types.

**Figure 3: Immune subtypes are associated with EMT scores.** EMT scores across TCGA immune subtypes; C1 = wound healing, C2 = IFN-γ dominant, C3 = inflammatory, C4 = lymphocyte depleted, C5 = immunologically quiet, C6 = TGF-β dominant **(A)** Plot of calculated KS score across all immune subtypes **(B)** Leave-one-out-analysis: pairwise comparison of each cancer immune subtype's KS score to the KS scores of all other immune subtypes **(C)** Application of Singscore to C5 immune subtype samples to quantify epithelial and mesenchymal scores separately. **(D)** Singscore epithelial and mesenchymal scores across immune subtypes

**Figure 4: EMT and MET factor gene expression varies across immune subtypes. (A-E)** Expression of EMT factors across immune subtypes; **(F-J)** Expression of MET factors acorss immune subtypes; **(A)** SNAI1, **(B)** SNAI2, **(C)** TWIST1, **(D)** ZEB1, **(E)** ZEB2, **(F)** GRHL2, **(G)** OVOL1, **(H)** OVOL2, **(I)** ESRP1, and **(J)** ESRP2 expression.

**Figure 5:** Immune checkpoint expression varies across immune subtypes. **(A)** CTLA4, **(B)** CD274, **(C)** LAG3, **(D)** HAVCR2, **(E)** CD47, **(F)** LGALS9, and **(G)** CD276 expression across immune subtypes. **(H)** Correlation matrix of EMT and pseudo-EMT scores with immune checkpoint markers.

**Acknowledgments**

JAS is supported by the Department of Defense (W81XWH-18-1-0189). AJA and JAS acknowledge funding support from NCI 1R01CA233585-02. The results published here are based, in part, upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

**References**

Addison, Joseph B., Maria A. Voronkova, James H. Fugett, Chen-Chung Lin, Nathaniel C. Linville, Brandon Trinh, Ryan H. Livengood, et al. 2021. "Functional Hierarchy and Cooperation of EMT Master Transcription Factors in Breast Cancer Metastasis." *Molecular Cancer Research: MCR*, January. https://doi.org/10.1158/1541-7786.MCR-20-0532.

Aran, Dvir, Zicheng Hu, and Atul J. Butte. 2017. "xCell: Digitally Portraying the Tissue Cellular Heterogeneity Landscape." *Genome Biology* 18 (1): 220.

Aran, Dvir, Marina Sirota, and Atul J. Butte. 2015. "Systematic Pan-Cancer Analysis of Tumour Purity." *Nature Communications* 6 (December): 8971.

Armstrong, Andrew J., Matthew S. Marengo, Sebastian Oltean, Gabor Kemeny, Rhonda L. Bitting, James D. Turnbull, Christina I. Herold, Paul K. Marcom, Daniel J. George, and Mariano A. Garcia-Blanco. 2011. "Circulating Tumor Cells from Patients with Advanced Prostate and Breast Cancer Display Both Epithelial and Mesenchymal Markers." *Molecular Cancer Research: MCR* 9 (8): 997–1007.

Barbie, David A., Pablo Tamayo, Jesse S. Boehm, So Young Kim, Susan E. Moody, Ian F. Dunn, Anna C. Schinzel, et al. 2009. "Systematic RNA Interference Reveals That Oncogenic KRAS-Driven Cancers Require TBK1." *Nature*. https://doi.org/10.1038/nature08460.

Bebee, Thomas W., Juw Won Park, Katherine I. Sheridan, Claude C. Warzecha, Benjamin W. Cieply, Alex M. Rohacek, Yi Xing, and Russ P. Carstens. 2015. "The Splicing Regulators Esrp1 and Esrp2 Direct an Epithelial Splicing Program Essential for Mammalian Development." *eLife* 4 (September). https://doi.org/10.7554/eLife.08954.

Bhatia, Sugandha, James Monkman, Tony Blick, Cletus Pinto, Mark Waltham, Shivashankar H. Nagaraj, and Erik W. Thompson. 2019. "Interrogation of Phenotypic Plasticity between Epithelial and Mesenchymal States in Breast Cancer." *Journal of Clinical Medicine Research* 8 (6). https://doi.org/10.3390/jcm8060893.

Biswas, Kuheli, Mohit Kumar Jolly, and Anandamohan Ghosh. 2019. "Stability and Mean Residence Times for Hybrid Epithelial/mesenchymal Phenotype." *Physical Biology* 16 (2): 025003.

Bonnomet, A., L. Syne, A. Brysse, E. Feyereisen, E. W. Thompson, A. Noël, J-M Foidart, P. Birembaut, M. Polette, and C. Gilles. 2012. "A Dynamic in Vivo Model of Epithelial-to-Mesenchymal Transitions in Circulating Tumor Cells and Metastases of Breast Cancer." *Oncogene* 31 (33): 3741–53.

Brabletz, Thomas. 2012. "To Differentiate or Not--Routes towards Metastasis." *Nature Reviews. Cancer* 12 (6): 425–36.

Brahmer, Julie R., Ramaswamy Govindan, Robert A. Anders, Scott J. Antonia, Sarah Sagorsky, Marianne J. Davies, Steven M. Dubinett, et al. 2018. "The Society for Immunotherapy of Cancer Consensus Statement on Immunotherapy for the Treatment of Non-Small Cell Lung Cancer (NSCLC)." *Journal for Immunotherapy of Cancer* 6 (1): 75.

Byers, Lauren Averett, Lixia Diao, Jing Wang, Pierre Saintigny, Luc Girard, Michael Peyton, Li Shen, et al. 2013. "An Epithelial-Mesenchymal Transition Gene Signature Predicts Resistance to EGFR and PI3K Inhibitors and Identifies Axl as a Therapeutic Target for Overcoming EGFR Inhibitor Resistance." *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 19 (1): 279–90.

Cancer Genome Atlas Research Network. 2011. "Integrated Genomic Analyses of Ovarian Carcinoma." *Nature* 474 (7353): 609–15.

Cannito, Stefania, Erica Novo, Alessandra Compagnone, Lorenzo Valfrè di Bonzo, Chiara Busletta, Elena Zamara, Claudia Paternostro, et al. 2008. "Redox Mechanisms Switch on Hypoxia-Dependent Epithelial-Mesenchymal Transition in Cancer Cells." *Carcinogenesis* 29 (12): 2267–78.

Cao, Yujia, Lu Zhang, Yosuke Kamimura, Patcharee Ritprajak, Masaaki Hashiguchi, Sachiko Hirose, and Miyuki Azuma. 2011. "B7-H1 Overexpression Regulates Epithelial-Mesenchymal Transition and Accelerates Carcinogenesis in Skin." *Cancer Research* 71 (4): 1235–43.

Cao, Yujia, Lu Zhang, Pacharee Ritprajak, Fumihiko Tsushima, Pornpan Youngnak-Piboonratanakit, Yosuke Kamimura, Masaaki Hashiguchi, and Miyuki Azuma. 2011. "Immunoregulatory Molecule B7-H1 (CD274) Contributes to Skin Carcinogenesis." *Cancer Research* 71 (14): 4737–41.

Chakraborty, Priyanka, Jason T. George, Shubham Tripathi, Herbert Levine, and Mohit Kumar Jolly. 2020. "Comparative Study of Transcriptomics-Based Scoring Metrics for the Epithelial-Hybrid-Mesenchymal Spectrum." *Frontiers in Bioengineering and Biotechnology* 8 (March): 220.

Chen, Binbin, Michael S. Khodadoust, Chih Long Liu, Aaron M. Newman, and Ash A. Alizadeh. 2018. "Profiling Tumor Infiltrating Immune Cells with CIBERSORT." *Methods in Molecular Biology* 1711: 243–59.

Chen, Limo, Don L. Gibbons, Sangeeta Goswami, Maria Angelica Cortez, Young-Ho Ahn, Lauren A. Byers, Xuejun Zhang, et al. 2014. "Metastasis Is Regulated via microRNA-200/ZEB1 Axis Control of Tumour Cell PD-L1 Expression and Intratumoral Immunosuppression." *Nature Communications* 5 (October): 5241.

Choi, Bae-Jung, Sin-Aye Park, Sung-Young Lee, Young Nam Cha, and Young-Joon Surh. 2017. "Hypoxia Induces Epithelial-Mesenchymal Transition in Colorectal Cancer Cells through Ubiquitin-Specific Protease 47-Mediated Stabilization of Snail: A Potential Role of Sox9." *Scientific Reports* 7 (1): 15918.

Cieply, B., P. Riley, P. M. Pifer, J. Widmeyer, and J. B. Addison. 2012. "Suppression of the Epithelial–mesenchymal Transition by Grainyhead-like-2." *Cancer Research*. https://cancerres.aacrjournals.org/content/72/9/2440.short.

Devaraj, Vimalathithan, and Biplab Bose. 2019. "Morphological State Transition Dynamics in EGF-Induced Epithelial to Mesenchymal Transition." *Journal of Clinical Medicine Research* 8 (7). https://doi.org/10.3390/jcm8070911.

Ding, Li, Matthew H. Bailey, Eduard Porta-Pardo, Vesteinn Thorsson, Antonio Colaprico, Denis Bertrand, David L. Gibbs, et al. 2018. "Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics." *Cell* 173 (2): 305–20.e10.

Dongre, Anushka, Mohammad Rashidian, Ferenc Reinhardt, Aaron Bagnato, Zuzana Keckesova, Hidde L. Ploegh, and Robert A. Weinberg. 2017. "Epithelial-to-Mesenchymal Transition Contributes to Immunosuppression in Breast Carcinomas." *Cancer Research* 77 (15): 3982–89.

Drápela, Stanislav, Jan Bouchal, Mohit Kumar Jolly, Zoran Culig, and Karel Souček. 2020. "ZEB1: A Critical Regulator of Cell Plasticity, DNA Damage Response, and Therapy Resistance." *Frontiers in Molecular Biosciences* 7 (March): 36.

Eddy, James A., Vésteinn Thorsson, Andrew E. Lamb, David L. Gibbs, Carolina Heimann, Jia Xin Yu, Verena Chung, et al. 2020. "CRI iAtlas: An Interactive Portal for Immuno-Oncology Research." *F1000Research* 9 (August): 1028.

Esposito, Mark, Nandini Mondal, Todd M. Greco, Yong Wei, Chiara Spadazzi, Song-Chang Lin, Hanqiu Zheng, et al. 2019. "Bone Vascular Niche E-Selectin Induces Mesenchymal-Epithelial Transition

and Wnt Activation in Cancer Cells to Promote Bone Metastasis." *Nature Cell Biology* 21 (5): 627–39.

Fedele, Vita, and Davide Melisi. 2020. "Permissive State of EMT: The Role of Immune Cell Compartment." *Frontiers in Oncology* 10 (April): 587.

Fici, Pietro, Giulia Gallerani, Anne-Pierre Morel, Laura Mercatali, Toni Ibrahim, Emanuela Scarpi, Dino Amadori, Alain Puisieux, Michel Rigaud, and Francesco Fabbri. 2017. "Splicing Factor Ratio as an Index of Epithelial-Mesenchymal Transition and Tumor Aggressiveness in Breast Cancer." *Oncotarget* 8 (2): 2423–36.

Fischer, Kari R., Anna Durrans, Sharrell Lee, Jianting Sheng, Fuhai Li, Stephen T. C. Wong, Hyejin Choi, et al. 2015. "Epithelial-to-Mesenchymal Transition Is Not Required for Lung Metastasis but Contributes to Chemoresistance." *Nature* 527 (7579): 472–76.

Foroutan, Momeneh, Dharmesh D. Bhuva, Ruqian Lyu, Kristy Horan, Joseph Cursons, and Melissa J. Davis. 2018. "Single Sample Scoring of Molecular Phenotypes." *BMC Bioinformatics* 19 (1): 404.

Gao, Ming-Qing, Baek Gil Kim, Suki Kang, Yoon Pyo Choi, Hangran Park, Kyu Sub Kang, and Nam Hoon Cho. 2010. "Stromal Fibroblasts from the Interface Zone of Human Breast Carcinomas Induce an Epithelial-Mesenchymal Transition-like State in Breast Cancer Cells in Vitro." *Journal of Cell Science* 123 (Pt 20): 3507–14.

Gao, Xia, Aman S. Bali, Scott H. Randell, and Brigid L. M. Hogan. 2015. "GRHL2 Coordinates Regeneration of a Polarized Mucociliary Epithelium from Basal Stem Cells." *The Journal of Cell Biology* 211 (3): 669–82.

George, Jason T., Mohit Kumar Jolly, Shengnan Xu, Jason A. Somarelli, and Herbert Levine. 2017. "Survival Outcomes in Cancer Patients Predicted by a Partial EMT Gene Expression Scoring Metric." *Cancer Research* 77 (22): 6415–28.

Giannoni, Elisa, Francesca Bianchini, Lorenzo Masieri, Sergio Serni, Eugenio Torre, Lido Calorini, and Paola Chiarugi. 2010. "Reciprocal Activation of Prostate Cancer Cells and Cancer-Associated Fibroblasts Stimulates Epithelial-Mesenchymal Transition and Cancer Stemness." *Cancer Research* 70 (17): 6945–56.

Gregory, Philip A., Cameron P. Bracken, Andrew G. Bert, and Gregory J. Goodall. 2008. "MicroRNAs as Regulators of Epithelial-Mesenchymal Transition." *Cell Cycle* 7 (20): 3112–18.

Gregory, Philip A., Cameron P. Bracken, Eric Smith, Andrew G. Bert, Josephine A. Wright, Suraya Roslan, Melanie Morris, et al. 2011. "An Autocrine TGF-beta/ZEB/miR-200 Signaling Network Regulates Establishment and Maintenance of Epithelial-Mesenchymal Transition." *Molecular Biology of the Cell* 22 (10): 1686–98.

Guen, Vincent J., Tony E. Chavarria, Cornelia Kröger, Xin Ye, Robert A. Weinberg, and Jacqueline A. Lees. 2017. "EMT Programs Promote Basal Mammary Stem Cell and Tumor-Initiating Cell Stemness by Inducing Primary Ciliogenesis and Hedgehog Signaling." *Proceedings of the National Academy of Sciences of the United States of America* 114 (49): E10532–39.

Guo, Dan, Ben-Ling Xu, Xu-Hua Zhang, and Ming-Min Dong. 2012. "Cancer Stem-like Side Population Cells in the Human Nasopharyngeal Carcinoma Cell Line Cne-2 Possess Epithelial Mesenchymal Transition Properties in Association with Metastasis." *Oncology Reports* 28 (1): 241–47.

Guo, Wenjun, Zuzana Keckesova, Joana Liu Donaher, Tsukasa Shibue, Verena Tischler, Ferenc Reinhardt, Shalev Itzkovitz, et al. 2012. "Slug and Sox9 Cooperatively Determine the Mammary Stem Cell State." *Cell* 148 (5): 1015–28.

Holderried, Tobias A. W., Luka de Vos, Emma Grace Bawden, Timo J. Vogt, Joern Dietrich, Romina Zarbl, Friedrich Bootz, et al. 2019. "Molecular and Immune Correlates of TIM-3 (HAVCR2) and Galectin 9 (LGALS9) mRNA Expression and DNA Methylation in Melanoma." *Clinical Epigenetics*. https://doi.org/10.1186/s13148-019-0752-8.

Hong, Tian, Kazuhide Watanabe, Catherine Ha Ta, Alvaro Villarreal-Ponce, Qing Nie, and Xing Dai. 2015. "An Ovol2-Zeb1 Mutual Inhibitory Circuit Governs Bidirectional and Multi-Step Transition between Epithelial and Mesenchymal States." *PLoS Computational Biology* 11 (11): e1004569.

Jeong, H. M., J. Han, S. H. Lee, H. J. Park, H. J. Lee, J. S. Choi, Y. M. Lee, Y. L. Choi, Y. K. Shin, and

M. J. Kwon. 2017. “ESRP1 Is Overexpressed in Ovarian Cancer and Promotes Switching from Mesenchymal to Epithelial Phenotype in Ovarian Cancer cellsThis Article Has Been Corrected since Advance Online Publication and an Erratum Is Also Printed in This Issue.” *Oncogenesis* 6 (10): e389–e389.

Jiang, Yuanyuan, and Hanxiang Zhan. 2020. “Communication between EMT and PD-L1 Signaling: New Insights into Tumor Immune Evasion.” *Cancer Letters* 468 (January): 72–81.

Jolly, Mohit Kumar. 2015. “Implications of the Hybrid Epithelial/Mesenchymal Phenotype in Metastasis.” *Frontiers in Oncology*. https://doi.org/10.3389/fonc.2015.00155.

Jolly, Mohit Kumar, Bogdan-Tiberius Preca, Satyendra C. Tripathi, Dongya Jia, Jason T. George, Samir M. Hanash, Thomas Brabletz, Marc P. Stemmler, Jochen Maurer, and Herbert Levine. 2018. “Interconnected Feedback Loops among ESRP1, HAS2, and CD44 Regulate Epithelial-Mesenchymal Plasticity in Cancer.” *APL Bioengineering* 2 (3): 031908.

Jolly, Mohit Kumar, Satyendra C. Tripathi, Dongya Jia, Steven M. Mooney, Muge Celiktas, Samir M. Hanash, Sendurai A. Mani, Kenneth J. Pienta, Eshel Ben-Jacob, and Herbert Levine. 2016. “Stability of the Hybrid Epithelial/mesenchymal Phenotype.” *Oncotarget* 7 (19): 27067–84.

Jolly, Mohit Kumar, Kathryn E. Ware, Shivee Gilja, Jason A. Somarelli, and Herbert Levine. 2017. “EMT and MET: Necessary or Permissive for Metastasis?” *Molecular Oncology* 11 (7): 755–69.

Kim, Sehui, Jaemoon Koh, Moon-Young Kim, Dohee Kwon, Heounjeong Go, Young A. Kim, Yoon Kyung Jeon, and Doo Hyun Chung. 2016. “PD-L1 Expression Is Associated with Epithelial-to-Mesenchymal Transition in Adenocarcinoma of the Lung.” *Human Pathology* 58 (December): 7–14.

Lee, Briana, Alvaro Villarreal-Ponce, Magid Fallahi, Jeremy Ovadia, Peng Sun, Qian-Chun Yu, Seiji Ito, Satrajit Sinha, Qing Nie, and Xing Dai. 2014. “Transcriptional Mechanisms Link Epithelial Plasticity to Adhesion and Differentiation of Epidermal Progenitor Cells.” *Developmental Cell* 29 (1): 47–58.

Liu, Xiaojuan, Yang Pu, Kyle Cron, Liufu Deng, Justin Kline, William A. Frazier, Hairong Xu, Hua Peng, Yang-Xin Fu, and Meng Michelle Xu. 2015. “CD47 Blockade Triggers T Cell–mediated Destruction of Immunogenic Tumors.” *Nature Medicine*. https://doi.org/10.1038/nm.3931.

Li, Yijing, Lei Wang, Loretta Pappan, Amy Galliher-Beckley, and Jishu Shi. 2012. “IL-1β Promotes Stemness and Invasiveness of Colon Cancer Cells through Zeb1 Activation.” *Molecular Cancer* 11 (November): 87.

Lv, Kun, Yingying Zhang, Mengying Zhang, Min Zhong, and Qifeng Suo. 2013. “Galectin-9 Promotes TGF-β1-Dependent Induction of Regulatory T Cells via the TGF-β/Smad Signaling Pathway.” *Molecular Medicine Reports* 7 (1): 205–10.

Mandal, Mousumi, Biswajoy Ghosh, Anji Anura, Pabitra Mitra, Tanmaya Pathak, and Jyotirmoy Chatterjee. 2016. “Modeling Continuum of Epithelial Mesenchymal Transition Plasticity.” *Integrative Biology: Quantitative Biosciences from Nano to Macro* 8 (2): 167–76.

Mani, Sendurai A., Wenjun Guo, Mai-Jing Liao, Elinor Ng Eaton, Ayyakkannu Ayyanan, Alicia Y. Zhou, Mary Brooks, et al. 2008. “The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells.” *Cell* 133 (4): 704–15.

Nieto, M. Angela, Ruby Yun-Ju Huang, Rebecca A. Jackson, and Jean Paul Thiery. 2016. “EMT: 2016.” *Cell* 166 (1): 21–45.

Ocaña, Oscar H., Rebeca Córcoles, Angels Fabra, Gema Moreno-Bueno, Hervé Acloque, Sonia Vega, Alejandro Barrallo-Gimeno, Amparo Cano, and M. Angela Nieto. 2012. “Metastatic Colonization Requires the Repression of the Epithelial-Mesenchymal Transition Inducer Prrx1.” *Cancer Cell* 22 (6): 709–24.

Pasani, Satwik, Sarthak Sahoo, and Mohit Kumar Jolly. 2020. “Hybrid E/M Phenotype(s) and Stemness: A Mechanistic Connection Embedded in Network Topology.” *Journal of Clinical Medicine Research* 10 (1). https://doi.org/10.3390/jcm10010060.

Paterson, Emily L., Natasha Kolesnikoff, Philip A. Gregory, Andrew G. Bert, Yeesim Khew-Goodall, and Gregory J. Goodall. 2008. “The microRNA-200 Family Regulates Epithelial to Mesenchymal

Transition." *TheScientificWorldJournal* 8 (September): 901–4.
Puram, Sidharth V., Itay Tirosh, Anuraag S. Parikh, Anoop P. Patel, Keren Yizhak, Shawn Gillespie, Christopher Rodman, et al. 2017. "Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer." *Cell* 171 (7): 1611–24.e24.
Qian, Li, Yang Liu, Shaoqing Wang, Weijuan Gong, Xiaoqin Jia, Lu Liu, Feng Ye, et al. 2017. "NKG2D Ligand RAE1ε Induces Generation and Enhances the Inhibitor Function of Myeloid-Derived Suppressor Cells in Mice." *Journal of Cellular and Molecular Medicine* 21 (9): 2046–54.
Rini, Brian I., Dena Battle, Robert A. Figlin, Daniel J. George, Hans Hammers, Tom Hutson, Eric Jonasch, et al. 2019. "The Society for Immunotherapy of Cancer Consensus Statement on Immunotherapy for the Treatment of Advanced Renal Cell Carcinoma (RCC)." *Journal for Immunotherapy of Cancer* 7 (1): 354.
Roca, Hernan, James Hernandez, Savannah Weidner, Richard C. McEachin, David Fuller, Sudha Sud, Taibriana Schumann, et al. 2013. "Transcription Factors OVOL1 and OVOL2 Induce the Mesenchymal to Epithelial Transition in Human Cancer." *PloS One* 8 (10): e76773.
Romeo, Elisabetta, Carmelo Antonio Caserta, Cristiano Rumio, and Fabrizio Marcucci. 2019. "The Vicious Cross-Talk between Tumor Cells with an EMT Phenotype and Cells of the Immune System." *Cells* 8 (5). https://doi.org/10.3390/cells8050460.
Ruscetti, M., E. L. Dadashian, W. Guo, B. Quach, D. J. Mulholland, J. W. Park, L. M. Tran, et al. 2016. "HDAC Inhibition Impedes Epithelial-Mesenchymal Plasticity and Suppresses Metastatic, Castration-Resistant Prostate Cancer." *Oncogene* 35 (29): 3781–95.
Saxena, Kritika, and Mohit Kumar Jolly. 2019. "Acute vs. Chronic vs. Cyclic Hypoxia: Their Differential Dynamics, Molecular Mechanisms, and Effects on Tumor Progression." *Biomolecules* 9 (8). https://doi.org/10.3390/biom9080339.
Saxena, Kritika, Syamanthak Srikrishnan, Toni Celia-Terrassa, and Mohit Kumar Jolly. 2020. "OVOL1/2: Drivers of Epithelial Differentiation in Development, Disease, and Reprogramming." *Cells, Tissues, Organs*, September, 1–10.
Schliekelman, Mark J., Ayumu Taguchi, Jun Zhu, Xudong Dai, Jaime Rodriguez, Muge Celiktas, Qing Zhang, et al. 2015. "Molecular Portraits of Epithelial, Mesenchymal, and Hybrid States in Lung Adenocarcinoma and Their Relevance to Survival." *Cancer Research* 75 (9): 1789–1800.
Serrano-Gomez, Silvia Juliana, Mazvita Maziveyi, and Suresh K. Alahari. 2016. "Regulation of Epithelial-Mesenchymal Transition through Epigenetic and Post-Translational Modifications." *Molecular Cancer* 15 (February): 18.
Skrypek, Nicolas, Steven Goossens, Eva De Smedt, Niels Vandamme, and Geert Berx. 2017. "Epithelial-to-Mesenchymal Transition: Epigenetic Reprogramming Driving Cellular Plasticity." *Trends in Genetics: TIG* 33 (12): 943–59.
Solinas, Cinzia, Edoardo Migliori, Pushpamali De Silva, and Karen Willard-Gallo. 2019. "LAG3: The Biological Processes That Motivate Targeting This Immune Checkpoint Molecule in Human Cancer." *Cancers* 11 (8). https://doi.org/10.3390/cancers11081213.
Somarelli, J. A., D. Schaeffer, M. S. Marengo, T. Bepler, D. Rouse, K. E. Ware, A. J. Hish, et al. 2016. "Distinct Routes to Metastasis: Plasticity-Dependent and Plasticity-Independent Pathways." *Oncogene* 35 (33): 4302–11.
Somarelli, Jason A., Samantha Shetler, Mohit K. Jolly, Xueyang Wang, Suzanne Bartholf Dewitt, Alexander J. Hish, Shivee Gilja, et al. 2016. "Mesenchymal-Epithelial Transition in Sarcomas Is Controlled by the Combinatorial Expression of MicroRNA 200s and GRHL2." *Molecular and Cellular Biology* 36 (19): 2503–13.
Stankic, Marko, Svetlana Pavlovic, Yvette Chin, Edi Brogi, David Padua, Larry Norton, Joan Massagué, and Robert Benezra. 2013. "TGF-β-Id1 Signaling Opposes Twist1 and Promotes Metastatic Colonization via a Mesenchymal-to-Epithelial Transition." *Cell Reports* 5 (5): 1228–42.
Stone, Rivka C., Irena Pastar, Nkemcho Ojeh, Vivien Chen, Sophia Liu, Karen I. Garzon, and Marjana Tomic-Canic. 2016. "Epithelial-Mesenchymal Transition in Tissue Repair and Fibrosis." *Cell and Tissue Research* 365 (3): 495–506.

Subbalakshmi, Ayalur R., Sarthak Sahoo, Kuheli Biswas, and Mohit Kumar Jolly. 2021. "A Computational Systems Biology Approach Identifies SLUG as a Mediator of Partial Epithelial-Mesenchymal Transition (EMT)." *Cells, Tissues, Organs*, February, 1–14.

Taki, Mana, Kaoru Abiko, Tsukasa Baba, Junzo Hamanishi, Ken Yamaguchi, Ryusuke Murakami, Koji Yamanoi, et al. 2018. "Snail Promotes Ovarian Cancer Progression by Recruiting Myeloid-Derived Suppressor Cells via CXCR2 Ligand Upregulation." *Nature Communications* 9 (1): 1685.

Tan, Tuan Zea, Qing Hao Miow, Yoshio Miki, Tetsuo Noda, Seiichi Mori, Ruby Yun-Ju Huang, and Jean Paul Thiery. 2014. "Epithelial-Mesenchymal Transition Spectrum Quantification and Its Efficacy in Deciphering Survival and Drug Responses of Cancer Patients." *EMBO Molecular Medicine* 6 (10): 1279–93.

Terry, Stéphane, Pierre Savagner, Sandra Ortiz-Cuaran, Linda Mahjoubi, Pierre Saintigny, Jean-Paul Thiery, and Salem Chouaib. 2017. "New Insights into the Role of EMT in Tumor Immune Escape." *Molecular Oncology* 11 (7): 824–46.

Thorsson, Vésteinn, David L. Gibbs, Scott D. Brown, Denise Wolf, Dante S. Bortone, Tai-Hsien Ou Yang, Eduard Porta-Pardo, et al. 2018. "The Immune Landscape of Cancer." *Immunity* 48 (4): 812–30.e14.

Toh, Benjamin, Xiaojie Wang, Jo Keeble, Wen Jing Sim, Karen Khoo, Wing-Cheong Wong, Masashi Kato, Armelle Prevost-Blondel, Jean-Paul Thiery, and Jean-Pierre Abastado. 2011. "Mesenchymal Transition and Dissemination of Cancer Cells Is Driven by Myeloid-Derived Suppressor Cells Infiltrating the Primary Tumor." *PLoS Biology* 9 (9): e1001162.

Tripathi, Satyendra C., Haley L. Peters, Ayumu Taguchi, Hiroyuki Katayama, Hong Wang, Amin Momin, Mohit Kumar Jolly, et al. 2016. "Immunoproteasome Deficiency Is a Feature of Non-Small Cell Lung Cancer with a Mesenchymal Phenotype and Is Associated with a Poor Outcome." *Proceedings of the National Academy of Sciences of the United States of America* 113 (11): E1555–64.

Tripathi, Shubham, Priyanka Chakraborty, Herbert Levine, and Mohit Kumar Jolly. 2020. "A Mechanism for Epithelial-Mesenchymal Heterogeneity in a Population of Cancer Cells." *PLoS Computational Biology* 16 (2): e1007619.

Tsai, Jeff H., Joana Liu Donaher, Danielle A. Murphy, Sandra Chau, and Jing Yang. 2012. "Spatiotemporal Regulation of Epithelial-Mesenchymal Transition Is Essential for Squamous Cell Carcinoma Metastasis." *Cancer Cell* 22 (6): 725–36.

Vandewalle, C. 2005. "SIP1/ZEB2 Induces EMT by Repressing Genes of Different Epithelial Cell-Cell Junctions." *Nucleic Acids Research*. https://doi.org/10.1093/nar/gki965.

Wahl, Geoffrey M., and Benjamin T. Spike. 2017. "Cell State Plasticity, Stem Cells, EMT, and the Generation of Intra-Tumoral Heterogeneity." *NPJ Breast Cancer* 3 (April): 14.

Wang, Li, Abdel Saci, Peter M. Szabo, Scott D. Chasalow, Mireia Castillo-Martin, Josep Domingo-Domenech, Arlene Siefker-Radtke, et al. 2018. "EMT- and Stroma-Related Gene Expression and Resistance to PD-1 Blockade in Urothelial Cancer." *Nature Communications* 9 (1): 3503.

Wang, Yingyan, Wen Lan, Mingxin Xu, Jing Song, Jun Mao, Chunyan Li, Xiaohui Du, et al. 2021. "Cancer-Associated Fibroblast-Derived SDF-1 Induces Epithelial-Mesenchymal Transition of Lung Adenocarcinoma via CXCR4/β-catenin/PPARδ Signalling." *Cell Death & Disease* 12 (2): 214.

Warzecha, Claude C., Trey K. Sato, Behnam Nabet, John B. Hogenesch, and Russ P. Carstens. 2009. "ESRP1 and ESRP2 Are Epithelial Cell-Type-Specific Regulators of FGFR2 Splicing." *Molecular Cell* 33 (5): 591–601.

Wei, S. C., Naas Anang, and R. Sharma. 2019. "Combination anti–CTLA-4 plus anti–PD-1 Checkpoint Blockade Utilizes Cellular Mechanisms Partially Distinct from Monotherapies." *Proceedings of the*. https://www.pnas.org/content/116/45/22699.short.

Weiss, Sarah A., Jedd D. Wolchok, and Mario Sznol. 2019. "Immunotherapy of Melanoma: Facts and Hopes." *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 25 (17): 5191–5201.

Wellner, Ulrich, Jörg Schubert, Ulrike C. Burk, Otto Schmalhofer, Feng Zhu, Annika Sonntag, Bettina

Waldvogel, et al. 2009. “The EMT-Activator ZEB1 Promotes Tumorigenicity by Repressing Stemness-Inhibiting microRNAs.” *Nature Cell Biology* 11 (12): 1487–95.

Wiener, Zoltan, Barbara Kohalmi, Peter Pocza, Judit Jeager, Gergely Tolgyesi, Sara Toth, Eva Gorbe, Zoltan Papp, and Andras Falus. 2007. “TIM-3 Is Expressed in Melanoma Cells and Is Upregulated in TGF-Beta Stimulated Mast Cells.” *The Journal of Investigative Dermatology* 127 (4): 906–14.

Wu, Chuan, Theresa Thalhamer, Rafael F. Franca, Sheng Xiao, Chao Wang, Chie Hotta, Chen Zhu, Mitsuomi Hirashima, Ana C. Anderson, and Vijay K. Kuchroo. 2014. “Galectin-9-CD44 Interaction Enhances Stability and Function of Adaptive Regulatory T Cells.” *Immunity* 41 (2): 270–82.

Yang, Jing, Parker Antin, Geert Berx, Cédric Blanpain, Thomas Brabletz, Marianne Bronner, Kyra Campbell, et al. 2020. “Guidelines and Definitions for Research on Epithelial-Mesenchymal Transition.” *Nature Reviews. Molecular Cell Biology* 21 (6): 341–52.

Yang, Jing, Sendurai A. Mani, Joana Liu Donaher, Sridhar Ramaswamy, Raphael A. Itzykson, Christophe Come, Pierre Savagner, Inna Gitelman, Andrea Richardson, and Robert A. Weinberg. 2004. “Twist, a Master Regulator of Morphogenesis, Plays an Essential Role in Tumor Metastasis.” *Cell* 117 (7): 927–39.

Yang, Muh-Hwa, Min-Zu Wu, Shih-Hwa Chiou, Po-Min Chen, Shyue-Yih Chang, Chung-Ji Liu, Shu-Chun Teng, and Kou-Juey Wu. 2008. “Direct Regulation of TWIST by HIF-1alpha Promotes Metastasis.” *Nature Cell Biology* 10 (3): 295–305.

Yang, Shuo, Wei Wei, and Qi Zhao. 2020. “B7-H3, a Checkpoint Molecule, as a Target for Cancer Immunotherapy.” *International Journal of Biological Sciences*. https://doi.org/10.7150/ijbs.41105.

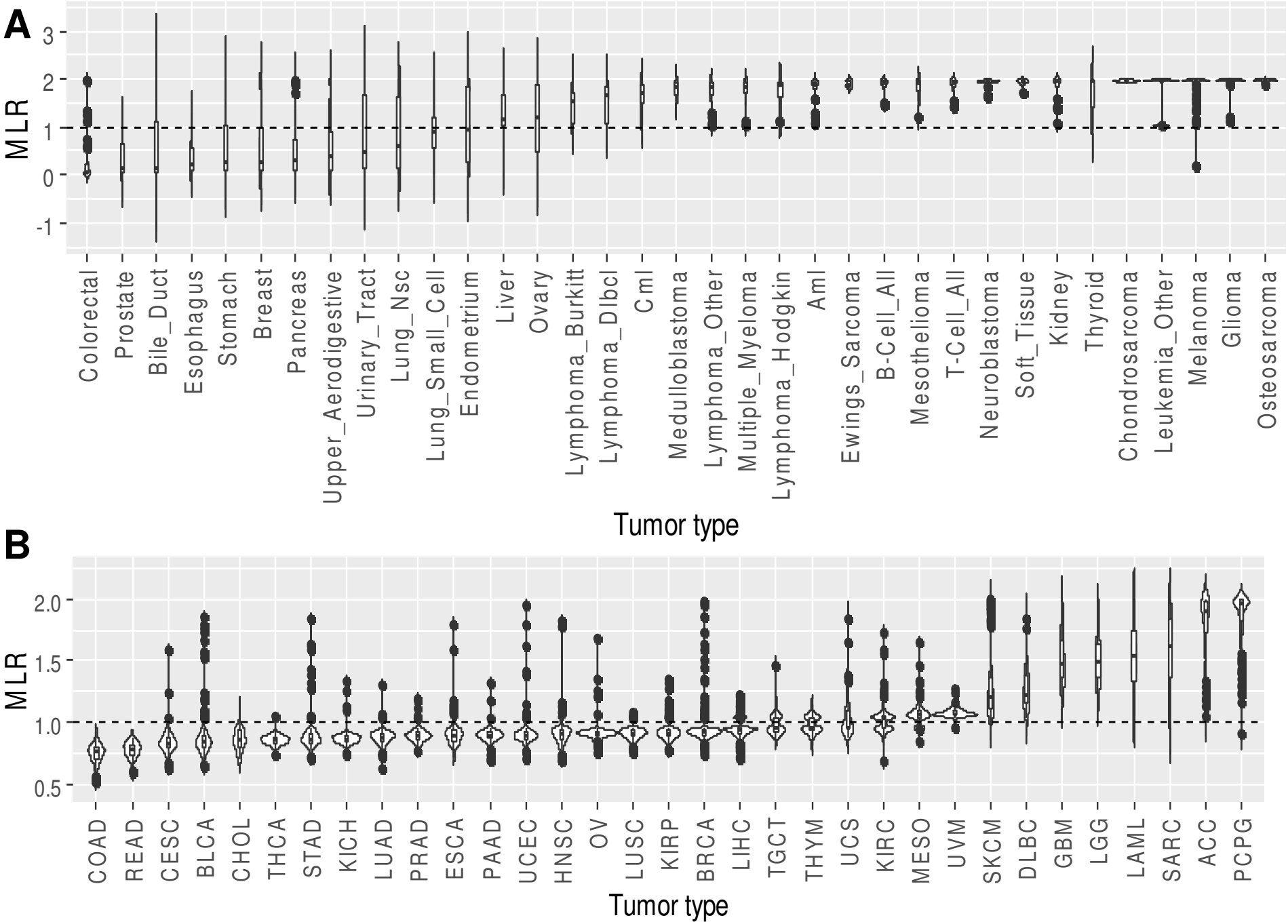
Yan, Wenjiang, Xiao Liu, Hongxin Ma, Hualin Zhang, Xiaojia Song, Lifen Gao, Xiaohong Liang, and Chunhong Ma. 2015. “Tim-3 Fosters HCC Development by Enhancing TGF-β-Mediated Alternative Activation of Macrophages.” *Gut* 64 (10): 1593–1604.

Yu, Min, Aditya Bardia, Ben S. Wittner, Shannon L. Stott, Malgorzata E. Smas, David T. Ting, Steven J. Isakoff, et al. 2013. “Circulating Breast Tumor Cells Exhibit Dynamic Changes in Epithelial and Mesenchymal Composition.” *Science* 339 (6119): 580–84.

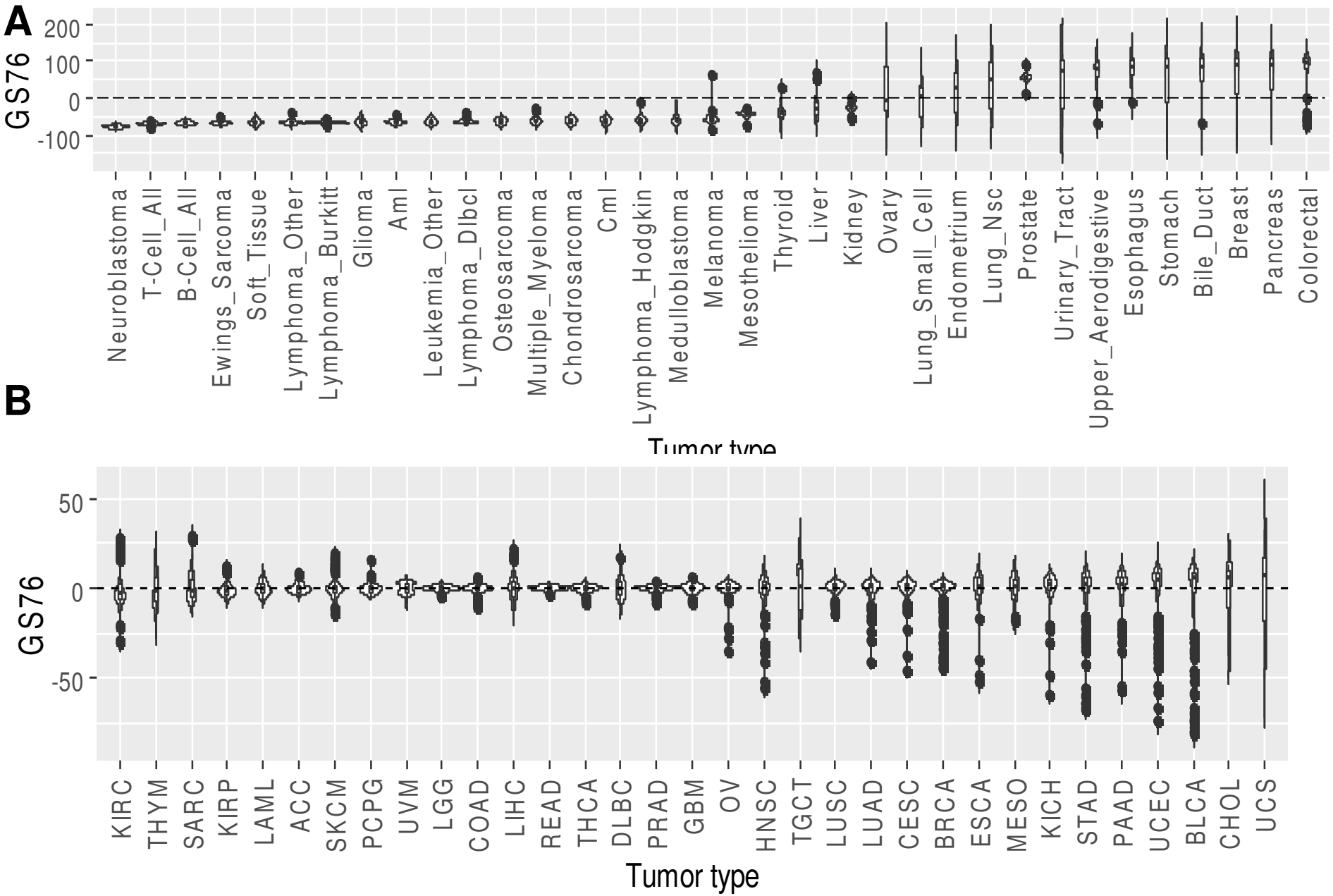
Zheng, Xiaofeng, Julienne L. Carstens, Jiha Kim, Matthew Scheible, Judith Kaye, Hikaru Sugimoto, Chia-Chin Wu, Valerie S. LeBleu, and Raghu Kalluri. 2015. “Epithelial-to-Mesenchymal Transition Is Dispensable for Metastasis but Induces Chemoresistance in Pancreatic Cancer.” *Nature* 527 (7579): 525–30.

Zou, Weiping, Jedd D. Wolchok, and Lieping Chen. 2016. “PD-L1 (B7-H1) and PD-1 Pathway Blockade for Cancer Therapy: Mechanisms, Response Biomarkers, and Combinations.” *Science Translational Medicine* 8 (328): 328rv4.

**Supplementary Figure 1**

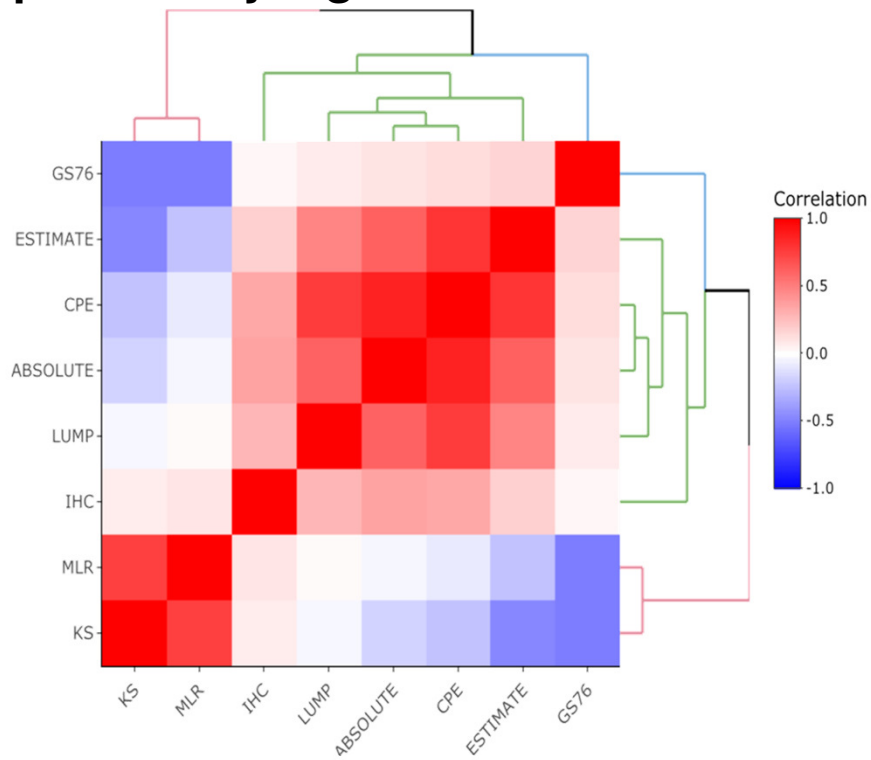


## Supplementary Figure 2

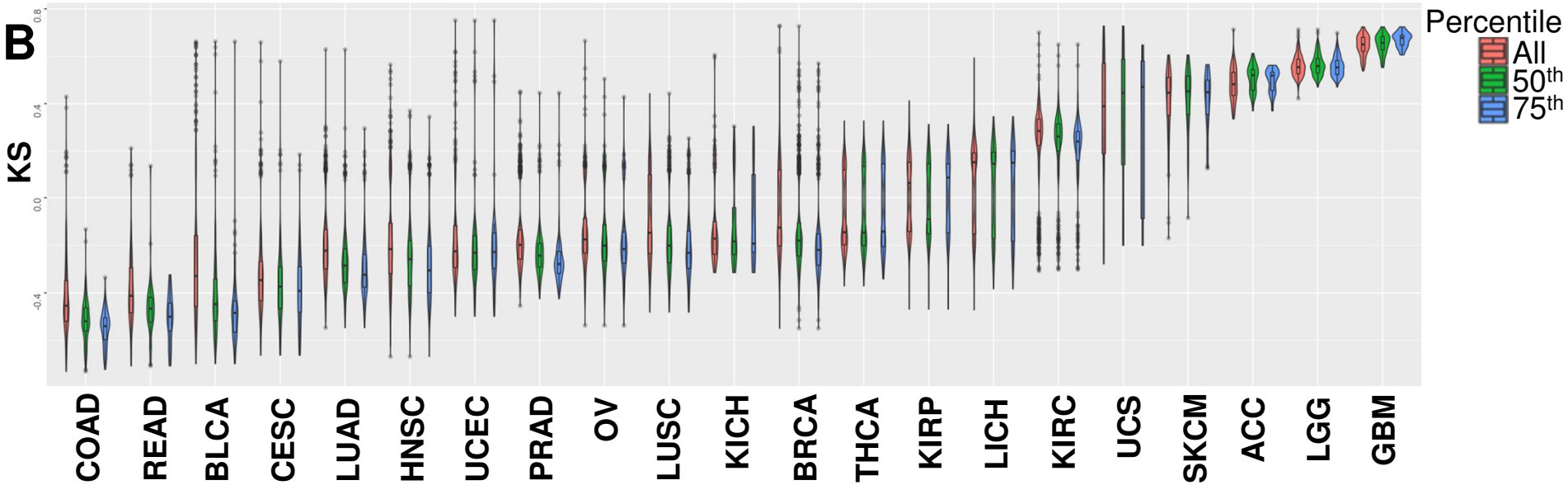


# Supplementary Figure 3

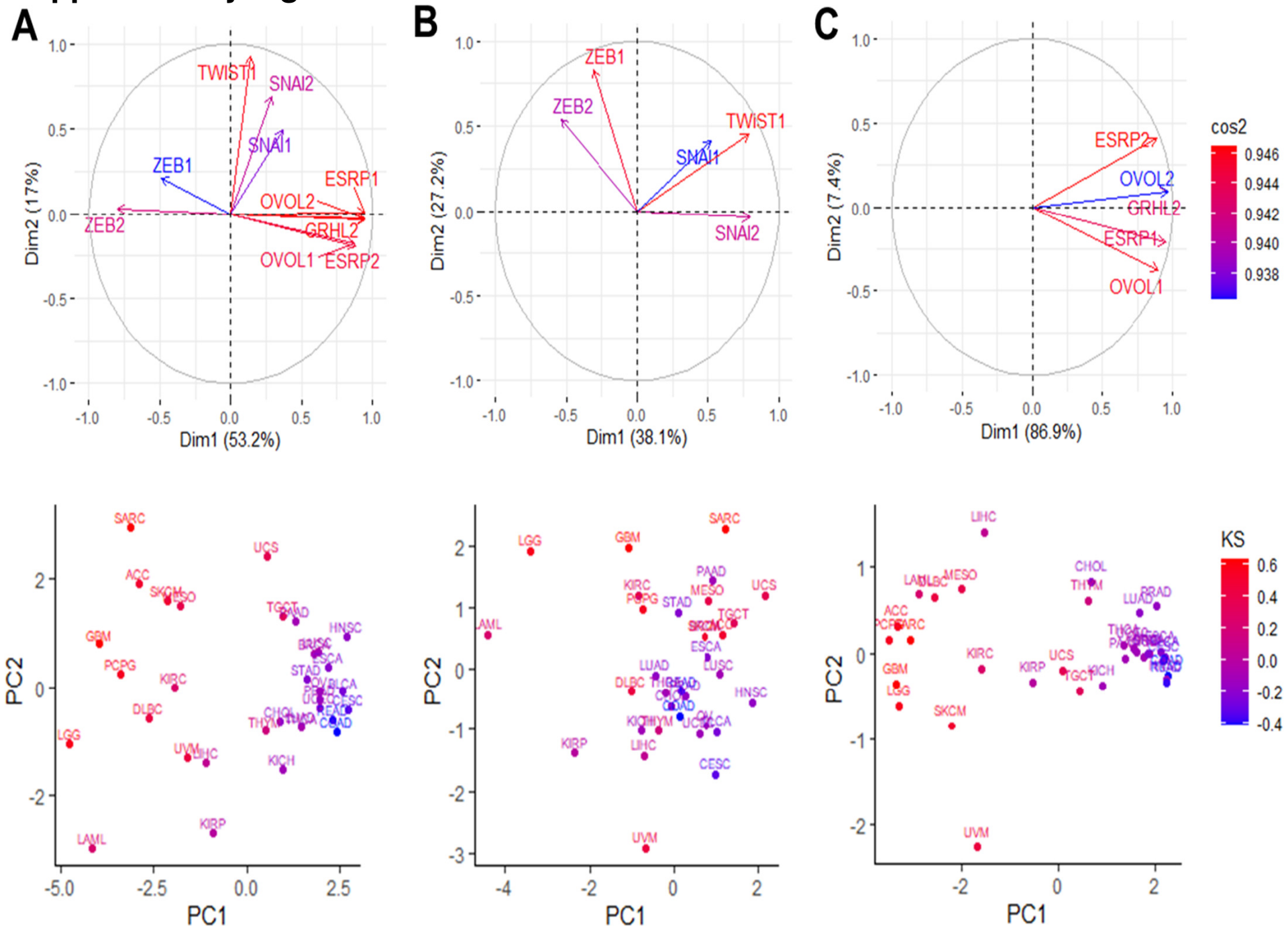
**A**



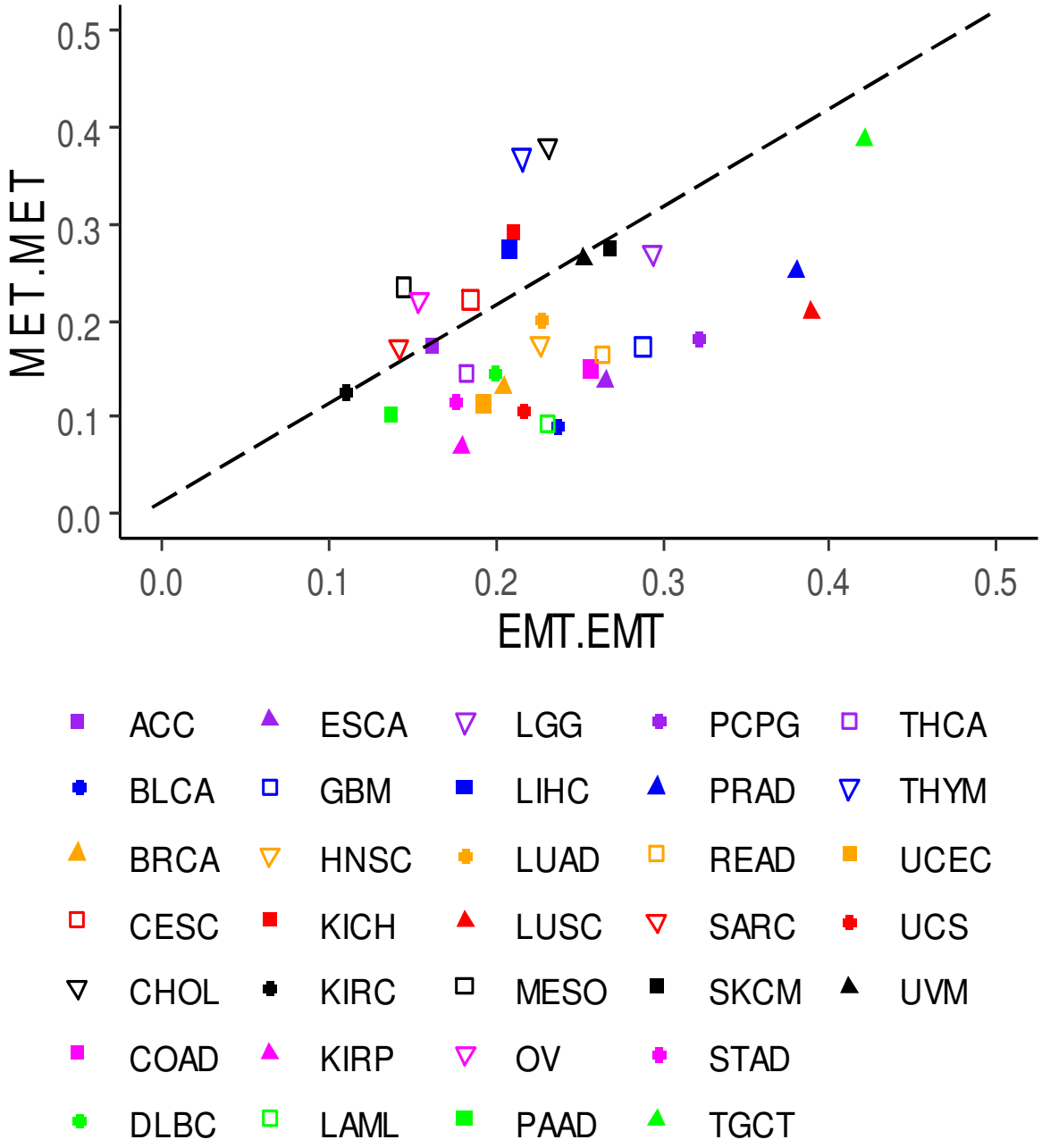

**B**



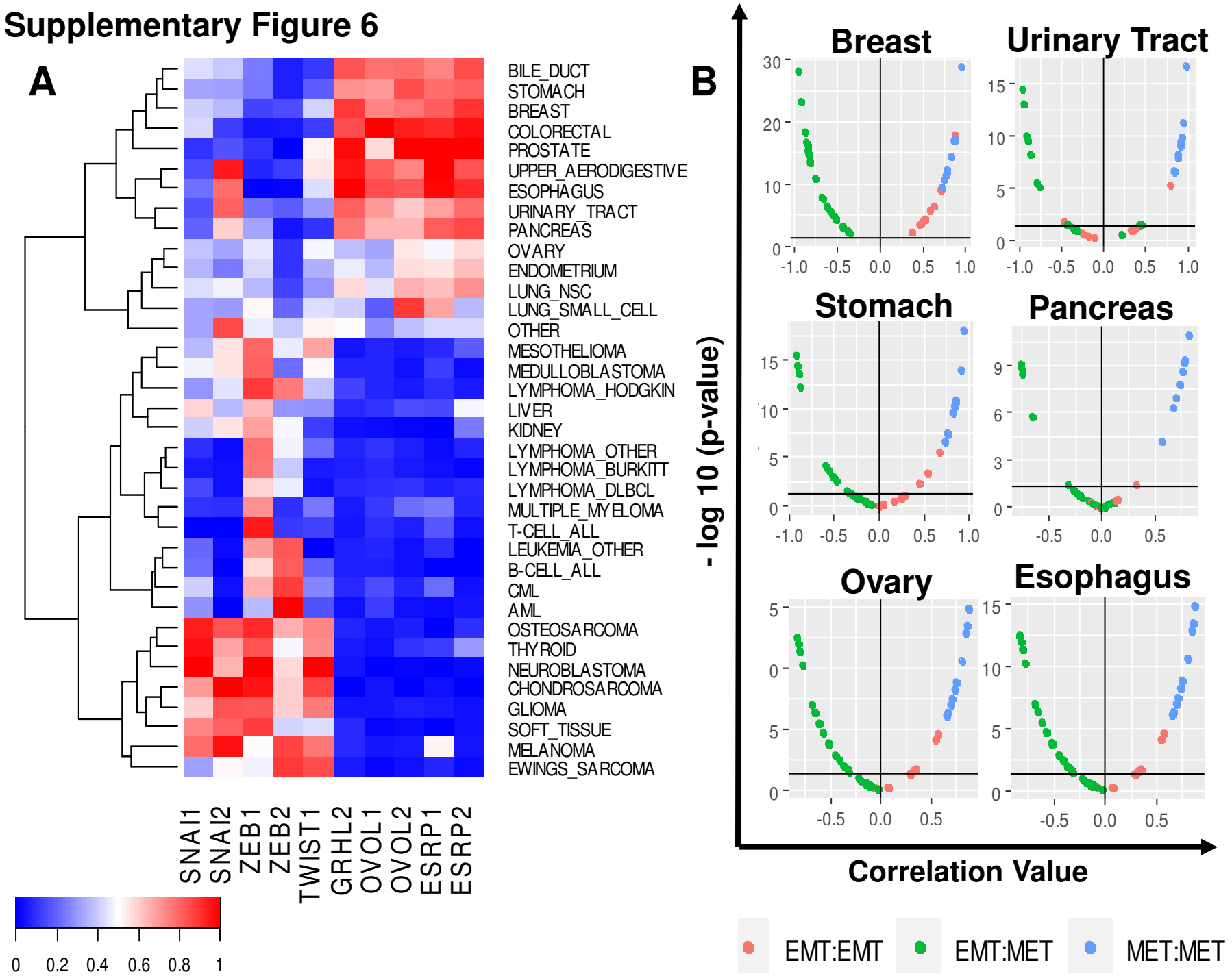
## Supplementary Figure 4



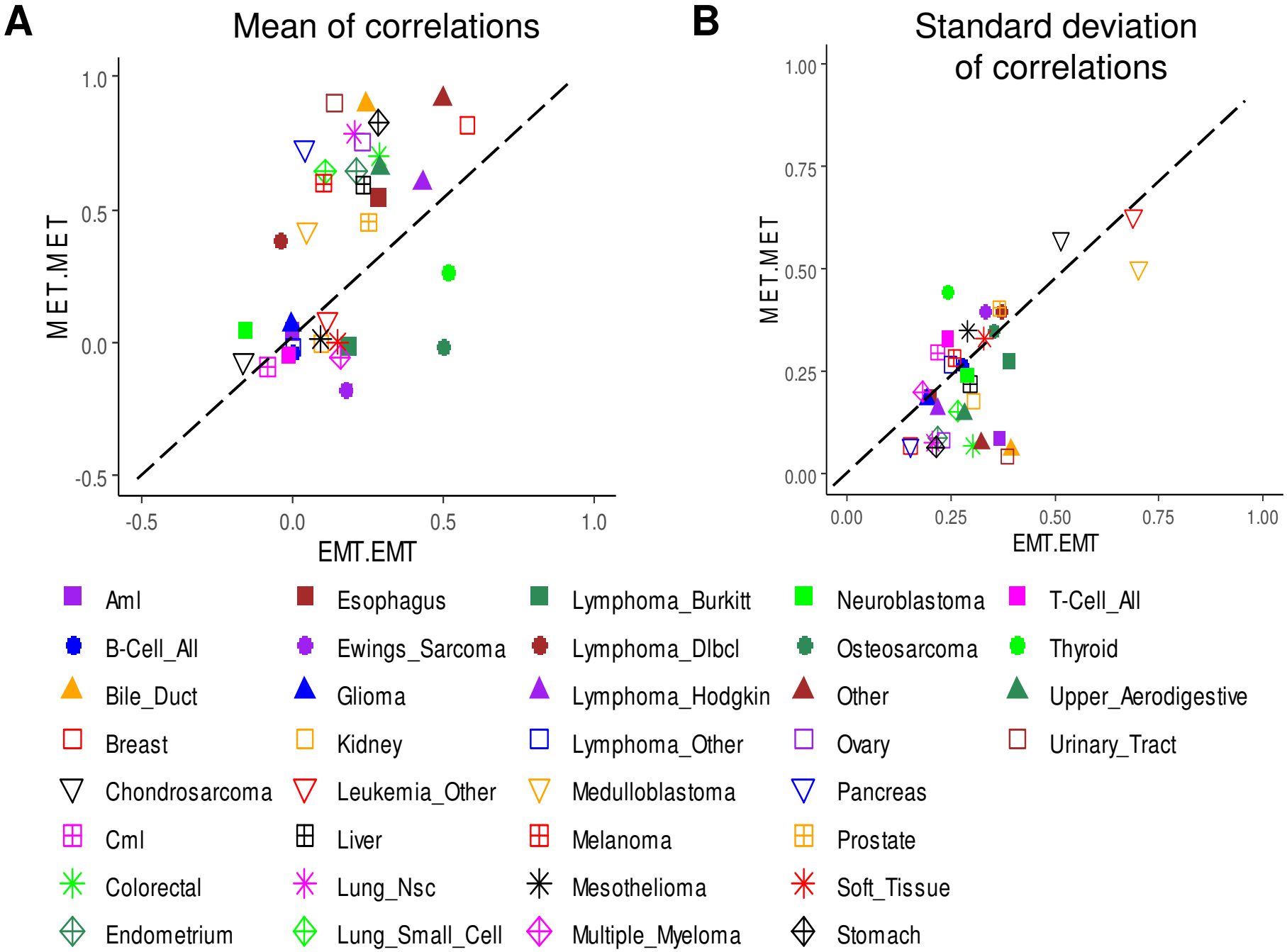
**Supplementary Figure 5**



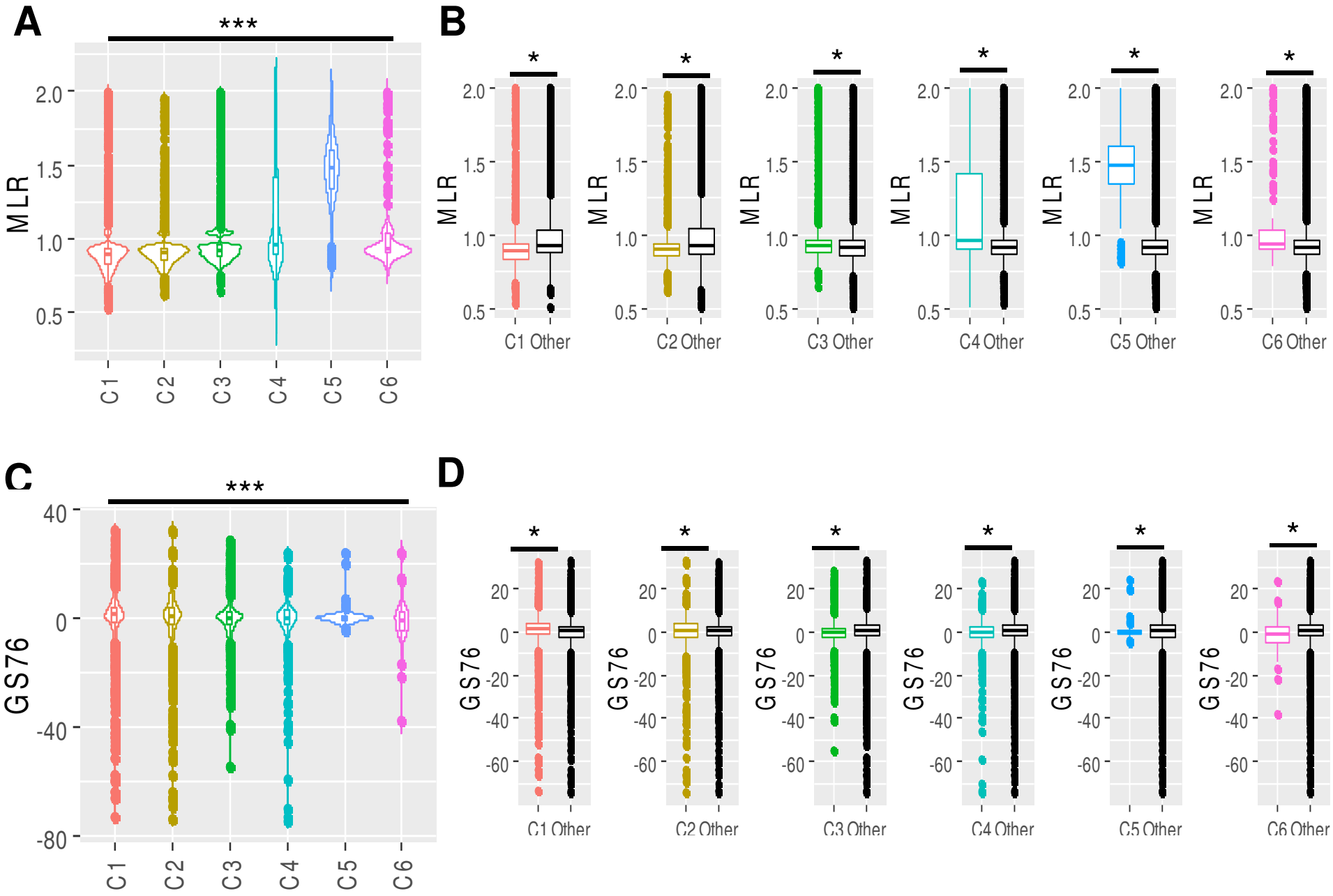
# Supplementary Figure 6



**Supplementary Figure 7**



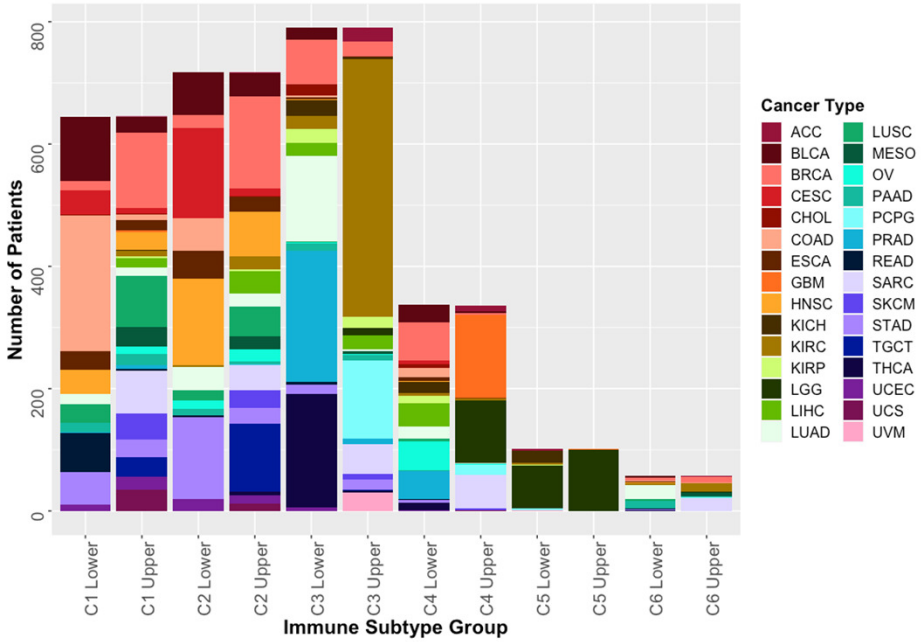
# Supplementary Figure 8



# Supplementary Figure 9

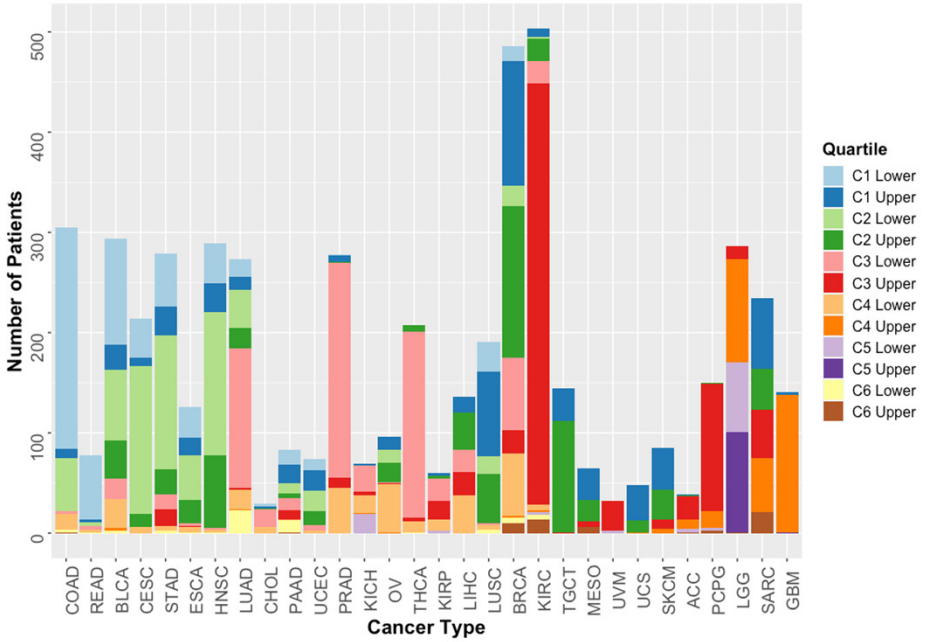
**A**

**Cancer type from upper and lower quartiles of EMT score**

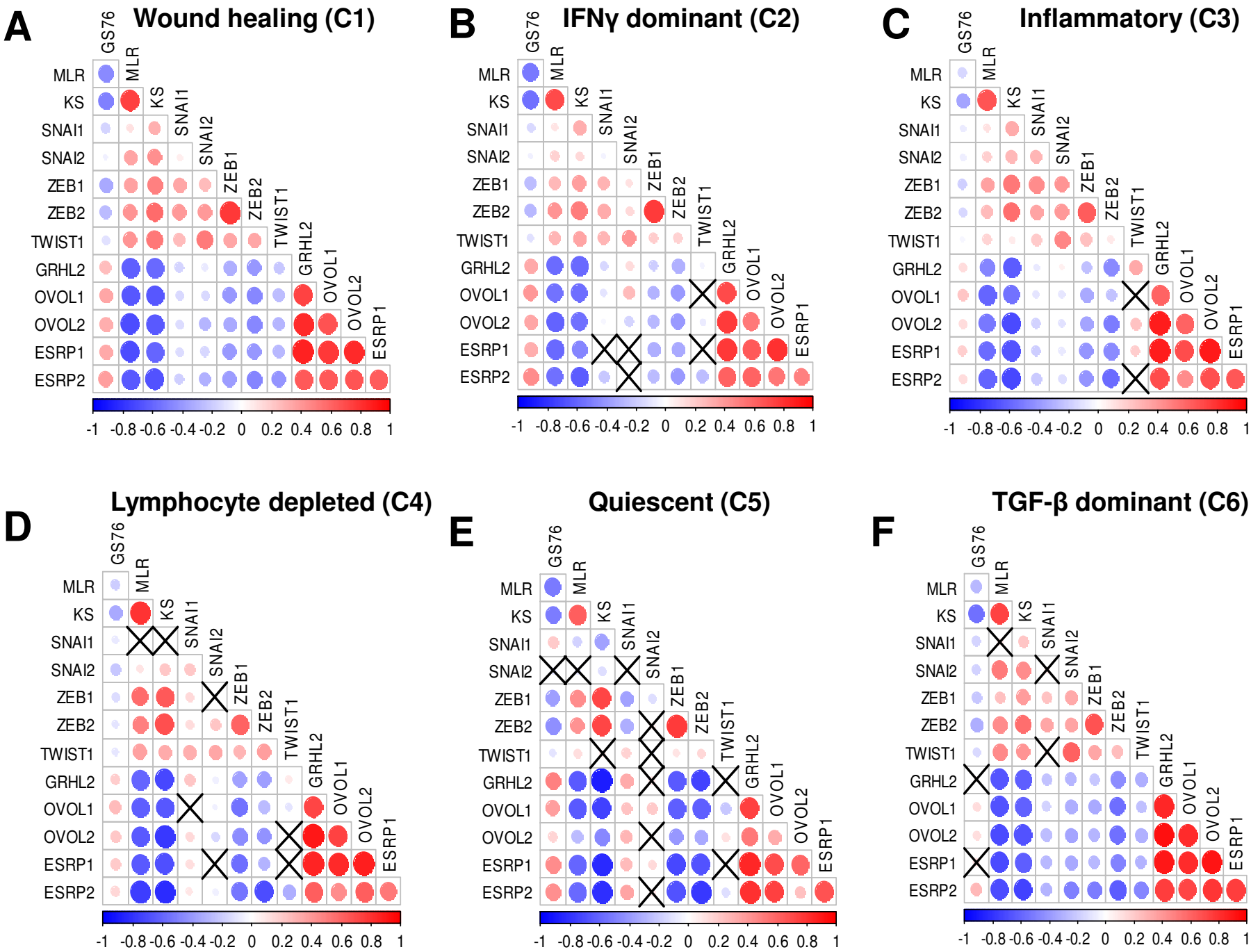



**B**

**Immune subtype distributions among upper and lower quartiles of EMT score**



# Supplementary Figure 10



**Supplementary figure legends**

**Figure S1: EMT scores across cancer types based on the MLR scoring metric.** The MLR algorithm scores samples on a scale of [-1, 1], with more positive scores representing more mesenchymal samples. **(A)** MLR algorithm applied to RNA-Seq data from the Cancer Cell Encyclopedia (CCLE). **(B)** MLR algorithm applied to RNA-Seq data from the Cancer Genome Atlas (TCGA); COAD = Colon adenocarcinoma, READ = Rectum adenocarcinoma, CESC = Cervical squamous cell carcinoma and endocervical adenocarcinoma, BLCA = Bladder urothelial carcinoma, STAD = stomach adenocarcinoma, ESCA = esophageal carcinoma, UCEC = uterine corpus endometrial carcinoma, LUAD = lung adenocarcinoma, HNSC = head and neck squamous cell carcinoma, PRAD = prostate adenocarcinoma, CHOL = cholangiocarcinoma, PAAD = pancreatic adenocarcinoma, KICH = kidney chromophobe, OV = ovarian serous cystadenocarcinoma, LUSC = lung squamous cell carcinoma, THCA = thyroid carcinoma, BRCA = breast invasive carcinoma, KIRP = kidney renal papillary cell carcinoma, LIHC = liver hepatocellular carcinoma, THYM = thymoma, LAML = acute myeloid leukemia, KIRC = kidney renal clear cell carcinoma, TGCT = testicular germ cell tumors, MESO = mesothelioma, UCS = uterine carcinosarcoma, DLBC = lymphoid neoplasm diffuse large B-cell lymphoma, UVM = uveal melanoma, SKCM = skin cutaneous melanoma, PCPG = pheochromocytoma and paraganglioma, LGG = brain lower grade glioma, SARC = sarcoma, GBM = glioblastoma multiforme.

**Figure S2: EMT scores across cancer types based on the 76GS scoring metric.** There is no pre-defined scale for 76GS algorithm. More negative scores represent more mesenchymal samples. **(A)** 76GS algorithm applied to RNA-Seq data from the Cancer Cell Encyclopedia (CCLE). **(B)** 76GS algorithm applied to RNA-Seq data from the Cancer Genome Atlas (TCGA); COAD = Colon adenocarcinoma, READ = Rectum adenocarcinoma, CESC = Cervical squamous cell carcinoma and endocervical adenocarcinoma, BLCA = Bladder urothelial carcinoma, STAD = stomach adenocarcinoma, ESCA = esophageal carcinoma, UCEC = uterine corpus endometrial carcinoma, LUAD = lung adenocarcinoma, HNSC = head and neck squamous cell carcinoma, PRAD = prostate adenocarcinoma, CHOL = cholangiocarcinoma, PAAD = pancreatic adenocarcinoma, KICH = kidney chromophobe, OV = ovarian serous cystadenocarcinoma, LUSC = lung squamous cell carcinoma, THCA = thyroid carcinoma, BRCA = breast invasive carcinoma, KIRP = kidney renal papillary cell carcinoma, LIHC = liver hepatocellular carcinoma, THYM = thymoma, LAML = acute myeloid leukemia, KIRC = kidney renal clear cell carcinoma, TGCT = testicular germ cell tumors, MESO = mesothelioma, UCS = uterine carcinosarcoma, DLBC = lymphoid neoplasm diffuse large B-cell lymphoma, UVM = uveal melanoma, SKCM = skin cutaneous melanoma, PCPG = pheochromocytoma and paraganglioma, LGG = brain lower grade glioma, SARC = sarcoma, GBM = glioblastoma multiforme.

**Figure S3: Correlation between EMT score and tumor purity. (A)** Four metrics of tumor purity, including immunohistochemistry (IHC) and the LUMP, ABSOLUTE, and CPE

algorithms, correlated to differing degrees with EMT scores calculated by the KS, MLR, and GS76 algorithms. (B) KS-based EMT scores with all samples, top 50$^{th}$ percentile of tumor purity score, and top 75$^{th}$ percentile of tumor purity score.

**Figure S4: MET factors are the predominant determinant of epithelial-like and mesenchymal-like clusters.** Principal component analysis (PCA) of EMT and MET marker expression values. The blue to red color scale represents the contribution of each gene to the segregation of tumor types into epithelial-like and mesenchymal-like clusters. **(A)** PCA of all five EMT and MET markers **(B)** PCA of only EMT markers **(C)** PCA of only MET markers.

**Figure S5: Standard deviation of the correlation of EMT and MET factor expression.** Standard deviation of pairwise correlation coefficients between EMT-EMT markers (x-axis) and MET-MET markers (y-axis) across TCGA tumor types

**Figure S6: EMT factors are heterogeneously expressed across cancers.** EMT and MET marker expression across CCLE tumor types **(A)** Normalized marker expression across tumor types **(B)** Pairwise correlations between expression values of all EMT-EMT, MET-MET, and EMT-MET factor pairs across a subset of tumor types.

**Figure S7: Variability in EMT and MET factor expression. (A)** Plot of mean pairwise correlation coefficients for all EMT-EMT factor pairs(x-axis) versus mean pairwise correlation coefficients for all MET-MET factor pairs (y-axis) across all CCLE tumor types. **(B)** Standard deviation of pairwise correlation coefficients between EMT-EMT markers (x-axis) and MET-MET markers (y-axis) across CCLE tumor types.

**Figure S8: Immune subtypes are associated with EMT scores across scoring algorithms.** EMT scores across TCGA immune subtypes; C1 = wound healing, C2 = IFN-γ dominant, C3 = inflammatory, C4 = lymphocyte depleted, C5 = immunologically quiet, C6 = TGF-β dominant **(A)** Plot of calculated MLR scores across all immune subtypes **(B)** Leave-one-out-analysis: pairwise comparison of each cancer immune subtype's MLR score to the MLR scores of all other immune subtypes **(C)** Plot of calculated 76GS scores across all immune subtypes **(D)** Leave-one-out-analysis: pairwise comparison of each cancer immune subtype's 76GS score to the 76GS scores of all other immune subtypes

**Figure S9: Extremes of EMT score are enriched for specific cancer types across immune subtypes. (A)** Analysis of TCGA cancer type distribution across the upper and lower EMT score quartiles of all immune subtype groups; COAD = Colon adenocarcinoma, READ = Rectum adenocarcinoma, CESC = Cervical squamous cell carcinoma and endocervical adenocarcinoma, BLCA = Bladder urothelial carcinoma, STAD = stomach adenocarcinoma, ESCA = esophageal carcinoma, UCEC = uterine corpus endometrial carcinoma, LUAD = lung adenocarcinoma, HNSC = head and neck squamous cell carcinoma, PRAD = prostate adenocarcinoma, CHOL = cholangiocarcinoma, PAAD = pancreatic adenocarcinoma, KICH = kidney chromophobe, OV = ovarian serous cystadenocarcinoma, LUSC = lung squamous cell carcinoma, THCA = thyroid

carcinoma, BRCA = breast invasive carcinoma, KIRP = kidney renal papillary cell carcinoma, LIHC = liver hepatocellular carcinoma, THYM = thymoma, LAML = acute myeloid leukemia, KIRC = kidney renal clear cell carcinoma, TGCT = testicular germ cell tumors, MESO = mesothelioma, UCS = uterine carcinosarcoma, DLBC = lymphoid neoplasm diffuse large B-cell lymphoma, UVM = uveal melanoma, SKCM = skin cutaneous melanoma, PCPG = pheochromocytoma and paraganglioma, LGG = brain lower grade glioma, SARC = sarcoma, GBM = glioblastoma multiforme **(B)** Analysis of the upper and lower EMT score quartiles of immune subtype across TCGA cancer types

**Figure S10: EMT scores correlate with EMT and MET factor gene expression.** Correlation between EMT factors, MET factors, and EMT scores across all immune subtypes (A-F)