CANCER

Transcriptional census of epithelial-mesenchymal plasticity in cancer

David P. Cook[1,2]* and Barbara C. Vanderhyden[1,2]

Epithelial-mesenchymal plasticity (EMP) contributes to tumor progression, promoting therapy resistance and immune cell evasion. Definitive molecular features of this plasticity have largely remained elusive due to the limited scale of most studies. Leveraging single-cell RNA sequencing data from 266 tumors spanning eight different cancer types, we identify expression patterns associated with intratumoral EMP. Integrative analysis of these programs confirmed a high degree of diversity among tumors. These diverse programs are associated with combinations of various common regulatory mechanisms initiated from cues within the tumor microenvironment. We show that inferring regulatory features can inform effective therapeutics to restrict EMP.

INTRODUCTION

Epithelial-mesenchymal plasticity (EMP) refers to the ability of cells to interconvert between epithelial and mesenchymal phenotypes, dynamically adopting mixed features of these states in response to signals in the cells' microenvironment (1). Throughout a tissue, cells with phenotypes spanning an epithelial/mesenchymal (E/M) continuum can be observed, emerging in response to specific features of their local environment. At the leading edge of tumors, for example, epithelial architecture becomes progressively disorganized, and the cancer cells express higher levels of mesenchymal-associated genes (2, 3). In cancer, this plasticity has been broadly associated with promoting metastasis, chemoresistance, and immunosuppression (4).

Given its supposed impact on tumor progression and treatment, understanding the molecular mechanisms that drive EMP and developing therapeutic strategies to modulate it have been a priority for years (1, 5–7). Identifying molecular determinants of EMP has largely focused on studying dynamics associated with the epithelial-mesenchymal transition (EMT) induced in experimental settings, through the addition of exogenous cytokines [e.g., transforming growth factor–β (TGF-β1)] or genetic manipulation. Over the past two decades, however, it has become increasingly clear that molecular features of the EMT are highly context specific (1, 8–12). The reliability of even the most canonical EMT genes (e.g., SNAI1, SNAI2, CDH1, CDH2, and VIM) has become unclear, and the reliance on specific genes as molecular markers of EMP has led to controversy (13–17). As a result, recent guidelines from "the EMT International Association" suggest that the primary criteria for defining EMP should focus on changes to cellular properties (e.g., loss of cell-cell junctions and enhanced migratory capacity) (1). Given the increasing number of genes becoming associated with EMP in the literature, this recommendation will be helpful in avoiding erroneous conclusions based on the expression of a small number of genes. In certain settings, however, assessing cellular properties is not particularly reasonable. Many studies depend on retrospective analysis of samples (e.g., data generated from tumor samples), and it may not be feasible to faithfully recapitulate EMP ex vivo. Comprehensive interrogation into the molecular properties of EMP and, importantly, the diversity of features across contexts is critical to enabling reliable interpretation of data from samples that are not amenable to phenotypic assessment. Also, embracing the diversity of EMP gene expression programs will allow for an updated conceptual model that may help explain its involvement in tumor progression and how it may be addressed therapeutically.

The advent of single-cell transcriptomics (scRNA-seq) has enabled the identification of coordinated gene expression patterns within individual cells. Developments to increase the throughput of these assays have allowed for the parallel measurement of gene expression in thousands of single cells, sampling the phenotypic diversity among a population within a tissue (18, 19). Under the assumption that intratumoral EMP is reasonably prevalent, single-cell genomics holds the promise of revealing its intrinsic molecular characteristics. Supporting this assumption, studies applying these methods have independently identified heterogeneous expression of EMT-associated genes in a variety of cancers, including squamous cell carcinoma (SCC) (3, 20, 21), pancreatic ductal adenocarcinoma (PDAC) (22), colorectal (23), and lung (24). Detailed exploration of these programs has been limited, however, and little has been done to compare features of these programs between individual tumors and cancers. Kinker et al. (25) recently assessed sources of heterogeneity in the expression profiles of monolayer-cultured cancer cell lines and identified three recurrent programs consistent with EMP, but the similarity of these programs to those that emerge in solid tumors is unclear.

Here, we leveraged scRNA-seq data from 266 tumor samples spanning eight cancer types to identify coordinated expression programs consistent with intratumoral EMP. While the overall composition of these programs was highly variable, we derived a set of well-conserved genes from these programs that can serve as a general EMP signature. We used this signature to query the pan-cancer data from The Cancer Genome Atlas (TCGA) and found that EMP was associated with reduced progression-free intervals (PFIs) and changes in immune cell proportions within the tumor microenvironment (TME). Inferences of regulatory mechanisms contributing to these EMP programs suggest that the diversity of these programs can arise from common underlying regulatory mechanisms, including ubiquitous activation of TGF-β1/nuclear factor κB (NFκB)/tumor necrosis factor (TNF) signaling but that some programs have notably high mitogen-activated protein kinase (MAPK) or signal transducers and activators of transcription (STAT)/hypoxia signaling

[1]Cancer Therapeutics Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada. [2]Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON, Canada.
*Corresponding author. Email: david.cook@uottawa.ca (D.P.C.)

activity. We integrated transcriptomic data from kinase inhibitor screens to demonstrate that these common regulatory mechanisms present promising targets for therapeutic restriction of EMP.

RESULTS

A pan-cancer census of EMT-associated gene expression

We first collected droplet-based scRNA-seq data from 17 studies of eight different cancer types, including colorectal (*26–28*), gastric (*29*), lung (*24, 28, 30–32*), nasopharyngeal (*33*), SCC (*20*), ovarian (*28, 34*), PDAC (*35*), and breast (*28, 36, 37*). After removing samples with fewer than 100 malignant cells, the data comprise expression profiles of 223,501 cancer cells from 266 tumor samples (Fig. 1A and figs. S1 and S2).

To assess intratumoral heterogeneity of EMT-associated gene expression, we used rank-based scoring of each cell for its relative expression of genes contained in 11 different EMT gene sets, including several curated sets [the "Epithelial-mesenchymal transition" Gene Ontology (GO) Term, MSigDB Hallmark gene set (*38*), CancerSEA (*39*), dbEMT1.0 (*40*), and Tan cancer-specific EMT gene set (*41*)] and others derived from individual studies [Taube (*9*), Puram (*3*), Cook (*8*), and Kinker I, II, III (*25*)]. The composition of these gene sets is highly variable, with more than two-thirds of the genes being present in only a single gene set (fig. S3A). While it is possible that these diverse gene sets could represent different subsets of a common underlying biological process, we correlated the gene set scores across all cells and found that this is not uniformly true. Several gene sets were poorly correlated with the others, suggesting that they may only reflect EMT programs under specific contexts. Other gene sets were well correlated despite differences in their composition,

suggesting that they likely reflect a common expression program (fig. S3B).

Gene set scores can provide biological insight into gene expression patterns, but they can be influenced by uninteresting features of the data. Specifically, they can be inflated by high expression of a small proportion of the set's genes that are not necessarily determinants of the queried biological process. Variation in scores across a population of cells may also reflect random fluctuations of the set's genes and not necessarily a coordinated activation of the process. Matrix factorization approaches have been applied to scRNA-seq data to learn coordinated expression programs heterogeneously expressed across a population. By learning these programs from the data itself, reliance on previously defined gene sets is restricted to only the interpretation of the programs.

We next sought to explore heterogeneously expressed programs from malignant populations from each of the 266 tumors independently. To identify these gene expression programs contributing to intratumoral heterogeneity, we performed multiresolution archetypal analysis on each tumor sample using the ACTIONet algorithm (*42*). This method identifies a set of extremal phenotypes (archetypes) from the distribution of gene expression profiles and represents each cell as a mixture of these archetype programs (Fig. 1B). This revealed multiple programs in each sample, including expected sources of variation, such as cell cycle activity. To identify those associated with EMP, we correlated the cellular activity of each archetype program learned by ACTIONet with the previous EMT gene set scores. We defined EMP programs by hierarchically clustering the archetype programs to identify those well correlated with EMT gene set scores (Fig. 1, C and D, and fig. S4A). Of the 266 tumors, 245 had identifiable EMP programs, suggesting that
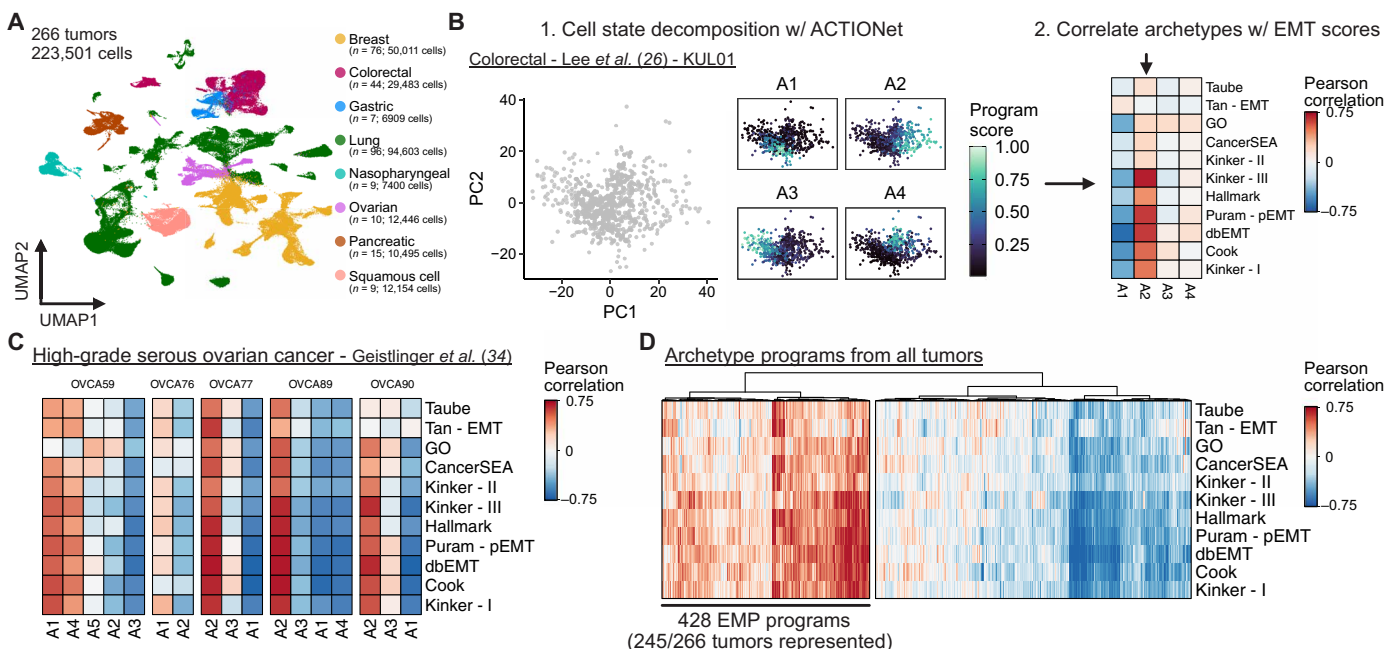

Fig. 1. Using archetypal analysis to learn EMP-associated gene expression programs. (**A**) UMAP embedding of malignant cell scRNA-seq data from 266 tumors and 17 different studies. (**B**) Schematic representation of the analysis strategy to identify sample-specific EMP programs using archetypal analysis and correlating archetype scores with EMT gene set scores. (**C**) Pearson correlation coefficients of archetype scores with EMT gene set scores of individual cells from five high-grade serous ovarian tumors. (**D**) Hierarchically clustered heatmap of Pearson correlation coefficients of all archetype scores with EMT gene set scores from the 266 tumors analyzed.

intratumoral EMP is a ubiquitous feature of solid tumors and readily captured in scRNA-seq experiments.

Defining a conserved signature associated with EMP

To define the specific genes contributing to EMP in each tumor, we identified genes whose expression is associated with each archetype program learned from ACTIONet. We have previously shown that transcriptional responses of experimentally induced EMTs are highly context specific, but it was unclear whether the same diversity existed in vivo (*8*). Of the 4187 genes up-regulated in at least one EMP program, the vast majority were associated with a small number of samples. Assessing the frequency at which each gene was differentially expressed, we assembled a general EMP signature from genes differentially expressed in 20 or more EMP programs (Fig. 2A). We further refined this signature by removing genes that were also down-regulated in more than 10 EMP programs, resulting in an EMP signature of 328 genes. We assessed these genes to determine whether the association of their expression with EMP was cancer type specific and found that the majority (251 of 329) showed no specific bias (analysis of variance, $P > 0.05$). Many of the 78 with cancer-specific patterns differed only in the magnitude of their association (fig. S4B).

Genes down-regulated upon the activation of EMP programs were even less consistent, with many down-regulated in only a small number of programs (fig. S4C). There was no overwhelming evidence of a suppressed epithelial phenotype, which may further support the growing evidence of hybrid E/M phenotypes being highly prevalent in cancer (*3, 43–45*). Several repressed genes were

consistent, however, with activation of a mesenchymal phenotype, including suppression of *STMN1* (*46*) and *PGC* (*47*), along with reduced expression of the proliferation markers *TOP2A* and *MKI67* (*48*).

Of the 328 genes positively associated with EMP, no individual gene was a perfect indicator of its activity, and in many EMP programs, they show little-to-no activity (Fig. 2B). As a collective, however, the signature represents a set of the most consistent markers of intratumoral EMP. They also enrich for GO terms consistent with a mesenchymal phenotype, including cell motility, regulation of cell adhesion, and response to wounding (Fig. 2C). Many of the canonical EMP genes are not included in this signature, including *CDH2*, *VIM*, *SNAI1*, *SNAI2*, and *ZEB1*, although these genes did have variable associations with EMP programs (fig. S5). The signature did, however, include 38 of the 200 EMT Hallmark genes, and many that have previously been implicated in the EMT, including various transcription factors (*SOX4*, *KLF4/5/6/10*, and *JUND*), integrins (*ITGB1/4* and *ITGA2*), secreted factors (*VEGFA*, *IL32*, *CXCL1*, and *CXCL8*), ECM components (*FN1* and *COL17A1*), and membrane proteins (*CD55* and *CD59*), and more (Fig. 2B and table S1).

The stability and distribution of phenotypes along an EM continuum has gained recent attention, with the relevance of hybrid phenotypes being contrasted to fully epithelial or mesenchymal cells (*49*). Using gene set scores of the conserved signature as a relative measure of the cells' EMP, we found that average scores were heterogeneous between samples (Fig. 2D and fig. S6). Intratumoral heterogeneity of scores was higher than expected by chance in all samples ($P < 0.01$, permutation test with random gene sets) (fig. S6).
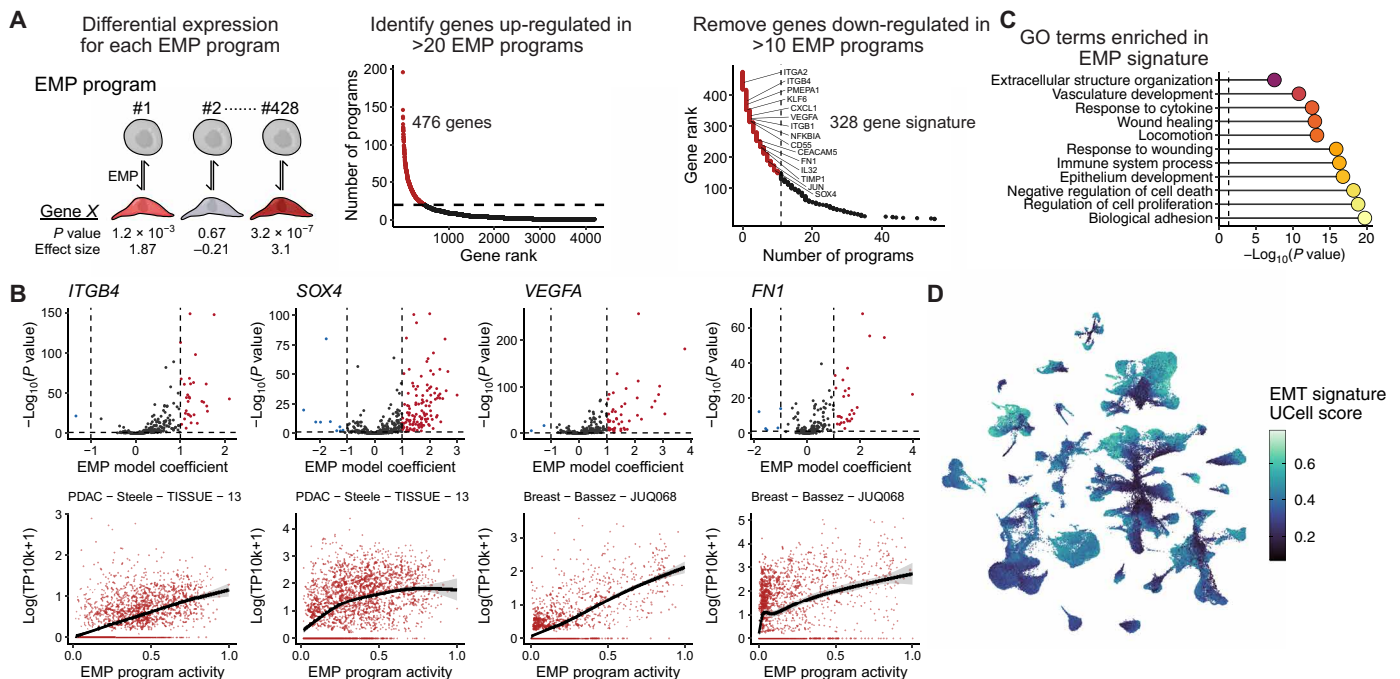

Fig. 2. Defining a conserved EMP gene signature. (**A**) Schematic showing the strategy for defining a signature of genes most frequently associated with the 428 identified EMP signatures. (**B**) EMP-associated GO terms significantly enriched in the 289 conserved EMP genes. *P* values were calculated using Fisher's exact tests and adjusted using the Benjamini-Hochberg method. (**C**) Examples of EMP associated genes. Top plots show the distribution of effect sizes (model coefficients) for each gene with all EMP programs. Dashed line represents the mean value. The horizontal dashed line represents a Benjamini-Hochberg adjusted *P* value of 0.05. (**D**) UMAP embedding of malignant cells from the 266 tumors analyzed, colored by a gene set score for the signature of 289 EMP-associated genes.

We also note that most tumors do not have clear multimodal distributions that would be consistent with a model where various states along the phenotypic continuum have elevated stability. Rather, cells span the continuum, forming a distribution with most cells occupying intermediate states and tails spreading to more extreme phenotypes.

A refined, malignant cell–specific EMP signature is associated with poor patient prognosis

Because of their inherent similarities, the ability to distinguish fibroblast and EMP-specific expression patterns has been a challenge. Many EMT gene sets contain genes highly expressed in fibroblast populations, and as a result, "mesenchymal" features of tumors defined from bulk RNA-seq data have been found to often be associated with fibroblast content of the tumor rather than cancer cell plasticity (3, 50, 51). The choice of specific markers used to assess EMP in studies has also led to controversy (1). This confusion has made it challenging to draw conclusions about the involvement of EMP in tumor progression and clinical outcomes. Recently, Tyler and Tirosh (52) identified malignant cell–specific EMP signatures by using scRNA-seq data to assess the specificity of EMT gene sets and genes that correlate with canonical EMT markers. As we have established a conserved EMP signature, we similarly used expression profiles from nonmalignant cells to refine a cancer cell–specific signature.

For each tumor sample, we calculated a cell type specificity score for each of the 328 conserved EMP genes and averaged these scores across tumors to get an overview of how specific the markers were to cancer cells (Fig. 3A). Of the 328 genes, 128 were highly specific to cancer cells, whereas the remaining 200 were also expressed in fibroblasts, macrophages, and/or T cells. A signature of cancer cell–specific EMP genes could be valuable for generating EMP activity scores in scRNA-seq data, so we established a refined signature comprising the 128 genes highly specific to cancer cells (table S2). To confirm the specificity of this signature in data that were not used to generate the signature, we scored cells from 24 pancreatic tumors (53) and found heterogeneous activity among cancer cell populations, with minimal activity in nonmalignant cell types (Fig. 3B).

Given this relatively high specificity, the signature could be used as a measure of mesenchymal properties in bulk RNA-seq data without being confounded by expression from stromal cell types. We used the pan-cancer RNA-seq data from TCGA (54) to calculate a relative EMP score for all tumors. Modeling patients' PFI as a function of this signature activity, tumor type, stage, age at diagnosis, and tumor purity, we found that EMP activity was associated with a reduced PFI (Cox hazard ratio: 2.65; $P = 7.18 \times 10^{-5}$) (Fig. 3, C and D). We then modeled estimates of immune cell proportions for all TCGA samples (55) as a function of this EMP score while controlling for the same covariates as above and found an association with not only higher proportions of inflammatory cell types (activated mast and dendritic cells) but also more immunosuppressive regulatory T cells and fewer effector cell types, including CD8 T cells and naive CD4 T cells (Fig. 3E).

Reconstructing EMP regulatory networks

While unifying molecular signatures are appealing, appreciating the diversity of EMP programs is critical as it likely contributes to functional nuances of the phenotype. These programs may also have varying regulatory dependencies that would warrant different
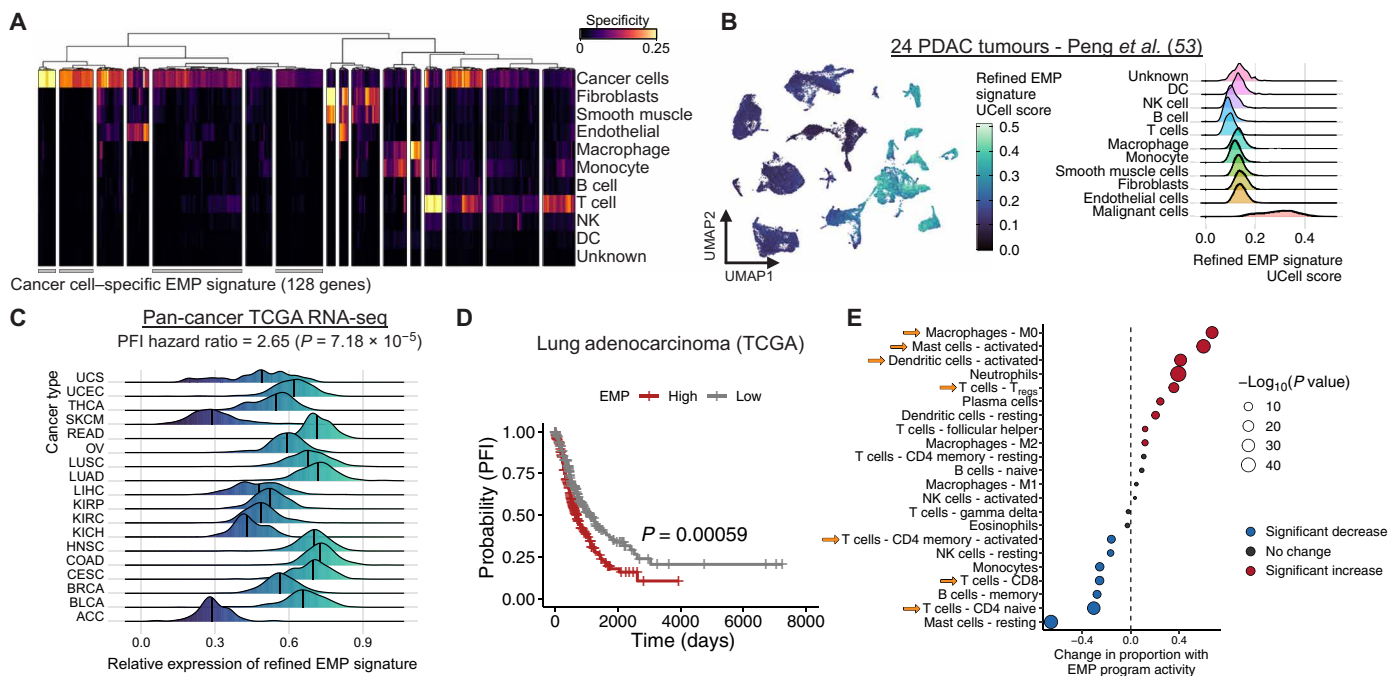

Fig. 3. EMP is associated with worse progression-free survival and an immunosuppressive TME. (**A**) Clustered heatmap of cell type specificity scores for each of the 328 EMP-associated genes. Clusters of genes with high cancer cell specificity were defined as a cancer cell–specific EMP signature. (**B**) UMAP embedding (left) and density plot (right) showing the distribution of gene set UCell scores for the cancer cell–specific EMP signature in all cell types from 24 pancreatic tumors. (**C**) PFI hazard ratio and relative expression (average Z score) of the cancer-specific EMP signature in the TCGA's pan-cancer bulk RNA-seq cohort. The hazard ratio and P value are derived from a Cox proportional hazards model, including variables for the refined EMP signature's expression, cancer type, stage, age, and tumor purity. (**D**) Changes in immune cell proportion estimates as a function of EMP signature scores from the TCGA's pan-cancer cohort. Cell type proportions were modeled using a linear model including the same variables as in (C). The change in proportion reflects the model coefficient for the EMP expression variable. DC, dendritic cell; T_{regs}, regulatory T cells.

therapeutic approaches. Variation in cell state emerges from complex differences in the cells' microenvironment that ultimately converge on signal transduction pathways and transcription factor networks. We next sought to use computational approaches to infer how regulation at these levels contribute to the patient-specific EMP programs we learned from scRNA-seq data.

Since the datasets we have used included matched gene expression profiles of stromal cells for all tumors, we used NicheNet (56) to infer cell-cell communication to identify factors from the TME that could contribute to each tumor's specific EMP program (Fig. 4A). Several cytokines were associated with EMP in many tumors, including TGF-β1, interleukin-6, posttranscriptional gene silencing 2 (PTGS2), interferon-γ (IFN-γ), and others (Fig. 4B). Consistent with previous literature, macrophages and fibroblasts frequently expressed many factors contributing to EMP (Fig. 4C) (3, 4). The majority of stromal cell types contribute implicated factors, such as PDGFB (Platelet-derived growth factor) primarily from endothelial cells; INHBA, FN1, and FGF7 from fibroblasts; IFN-γ and XCL1 from natural killer (NK) cells; and TNF, matrix metallopeptidase 9, and PTGS2 from myeloid cells. Malignant cells also express various ligands predicted to promote EMP, such as VEGFA and bone morphogenetic protein 7, suggesting that they may establish self-regulatory signaling within the TME (Fig. 4C).

To assess intracellular signaling activity associated with EMP, we used PROGENy (57)—a model of consensus signatures of signaling pathways—to calculate a relative activity score for 14 different signaling pathways in each cell. We then modeled EMP program activity as a function of these signaling activity scores (Fig. 4D). EMP programs were consistently associated with elevated TGF-β, NFκB, and TNFα signaling, but some had distinctly high levels of either MAPK and VEGF signaling or hypoxia and p53 signaling (Fig. 4, E and F). This suggests that hypoxia-associated EMP may have distinct features from plasticity coordinated by MAPK/extracellular signal–regulated kinase (ERK) signaling. This is consistent with findings that treatment of breast cancer cells with epidermal growth factor (EGF) induces an EMT with distinct characteristics from those following a hypoxia-induced EMT (58). These programs likely emerge in response to different interactions with the TME, as we found several ligands preferentially associated with each, such the ERK-activating ITGA4 and HBEGF (Herapin-binding EGF-like growth factor) with MAPK/VEGF programs (Fig. 4G).

Last, as this regulatory signaling converges on transcription factor networks to modulate gene expression, we used SCENIC (59) to construct gene regulatory networks for each tumor sample and then assessed how transcription factor activity changes in each EMP program. The most frequently associated transcription factors include many noncanonical factors, such as AP-1 (JUNB and FOSB), SOX4, KLF2/4/6, and STAT1 (Fig. 5A). Fewer transcription factors had consistent deactivation upon activation of EMP programs; however, several factors were repressed fairly commonly, including MYBL2, E2F1, and HES6 (Fig. 5A).

Pharmacologic restriction of EMP

EMP has been implicated in both chemoresistance and immune cell evasion (4). Given this, we predict that nonlethal restriction of EMP could be a promising therapeutic approach to sensitize tumors to orthogonal treatments and elicit immune cell killing. Diversity of EMP programs could introduce challenges for effectively preventing these cell dynamics, but the dependence of EMP on factors from

the cells' microenvironment suggests that the diversity likely arises from combinatorial effects from the relatively limited number of signal transduction pathways. Therefore, we suspect that diverse EMP programs may be susceptible to common pathway perturbations, and rational treatments could be devised by inferring signaling activity associated with EMP in a given tumor.

To begin to test this prediction, we first explored the MIX-Seq dataset comprising scRNA-seq profiles of more than 100 cancer cell lines treated with various drugs, including the MAPK kinase (MEK) inhibitor trametinib (60). We used ACTIONet to define cell line–specific EMP programs from untreated expression profiles, identifying high-confidence EMP programs in 46 of the 99 lines we assessed (lines with >100 cells; fig. S7). Many others had programs that correlated well with individual EMT gene sets and may represent EMP, but out of caution, we did not annotate them as such. We then inferred changes in signaling pathway activity associated with all archetype programs and found that, like in the tumor samples, TGF-β/NFκB/TNFα activity was consistently higher in EMP programs, but programs could be distinguished by high MAPK or hypoxia/p53 signaling (Fig. 6A).

To determine whether MEK inhibition preferentially limits MAPK-associated EMP, we used the MIX-Seq dataset to assess the effects of trametinib on all EMP programs. As expected, trametinib impaired EMP in the programs inferred to be associated with high MAPK signaling (Fig. 6, A and B). Kinase inhibition can promote a mesenchymal phenotype in breast cancer cells under hypoxic conditions (58), and we observe that this is generalizable across the diverse cancer lines included in the MIX-seq dataset, with trametinib enhancing EMP activity in programs associated with hypoxia/p53 signaling (Fig. 6, A and B). Given the consistent association of EMP with TGF-β1 and NFκB signaling, we would also predict that inhibition of these pathways could restrict EMP. While TGF-β1 and NFκB inhibitors were not included in the MIX-Seq dataset, we have previously published scRNA-seq data of four cancer cell lines (A549, DU145, MCF7, and OVCA420) treated with the TGF-βR1 inhibitor LY364947. Performing the same analysis on these data, we found that inhibition of TGF-βR1 in cells cultured in control conditions led to repression of most EMP programs, proportional to the inferred level of TGF-β1 activity associated with each line's intrinsic EMP (Fig. 6C and fig. S7B). Together, this supports that EMP programs with highly diverse molecular features may still have common dependencies, and effective strategies for restricting EMP can be found by inferring these regulatory mechanisms.

DISCUSSION

EMP has long been appreciated as a prominent source of intratumoral heterogeneity that ultimately promotes tumor progression and hinders effective treatment. While select molecular patterns associated with EMP have been known for decades, variability in their involvement across contexts has raised confusion, and as a result, high-level conceptual models to understand the molecular basis of this plasticity have been lacking. Here, we used scRNA-seq data of 266 tumor samples to identify gene expression programs consistent with EMP, compared features of these programs between tumor samples, inferred regulatory networks driving them, and explored therapeutic options to limit their activity. We have found that, consistent with experimental studies, intratumoral EMP–associated expression is highly variable. Unique genetic aberrations of each tumor likely
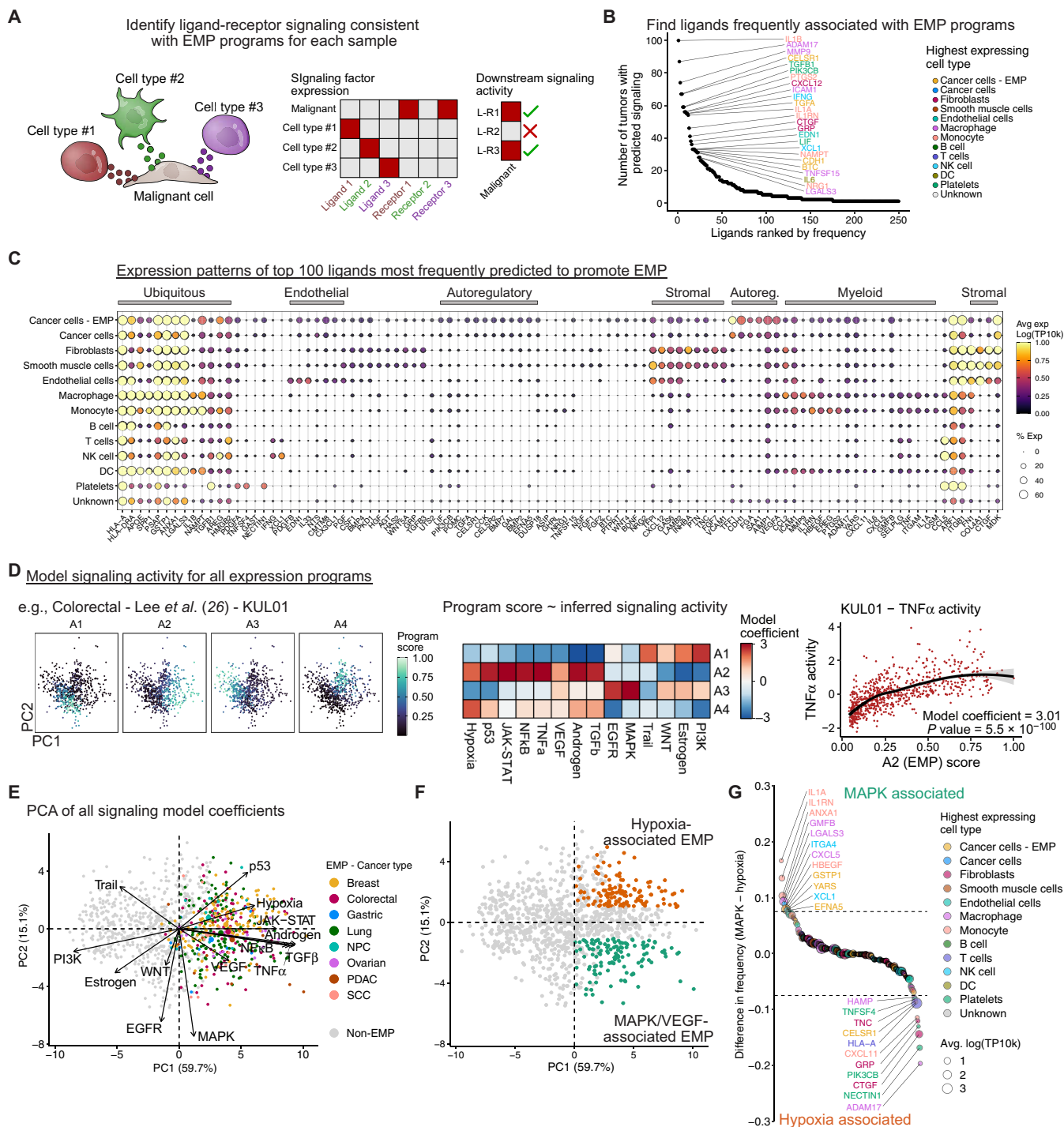
Fig. 4. Signaling networks regulating EMP in cancer. (**A**) Schematic showing the identification of ligands that may contribute to EMP. Expression patterns of ligands and their cognate receptors are assessed across all cell types. Putative regulatory ligands are identified if malignant cells express a matching receptor and if the sample's EMP program is consistent with downstream signaling related to that ligand. (**B**) Potential ligands from the TME regulating EMP. Ligands were ranked by the number of tumors in which they were predicted to promote EMP programs. (**C**) Expression of the top 100 ligands in cell types of the TME. The gene order is based on hierarchical clustering and manually annotated. Expression values (log-transformed and scaled UMI counts) were averaged across all 266 tumors. (**D**) Schematic showing the inference of signaling pathway activity associated with EMP programs. Signaling activity scores are inferred for each individual cell using PROGENy, and each archetype score for a given tumor is then modeled as a function of signaling activity using a linear model. (**E**) Principal components analysis (PCA) of the signaling model coefficients for all archetype programs. Each point represents an archetype program with EMP programs colored by the cancer type they are associated with. Loading vectors for each of the 14 signaling pathways are shown. (**F**) Sample as (E) but colored to define hypoxia/p53 (orange) and MAPK/VEGF-associated (green) EMP programs. These were defined as programs having PC1 > 0 and either PC2 > 1 (hypoxia) or PC2 ≤ 1 (MAPK/VEGF). (**G**) Ligands with preferential associations with MAPK/VEGF- or hypoxia/p53-associated EMP. Values represent the difference in the proportion of programs in which each ligand was implicated.
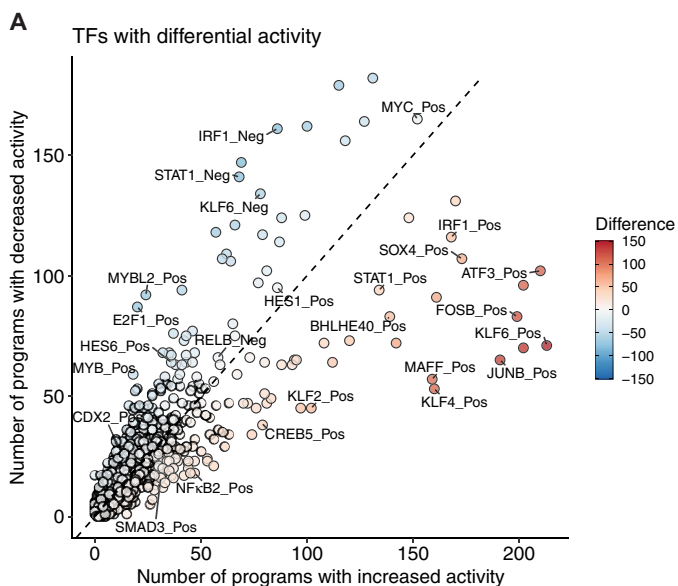
A

Fig. 5. Transcription factor regulons associated with EMP. (**A**) Transcription factors predicted to have differential activity in EMP programs. Each point represents a group of coexpressed genes coregulated by a given transcription factor (a regulon).

contribute to this diversity but are challenging to assess. Experimental strategies to promote an EMT in both normal and cancer cell lines with distinct inducers can elicit unique EMT responses, suggesting that the diversity cannot solely be due to genetic variation (8, 61). It is likely that this diversity is the product of multiple regulatory variables, including genetic variation, the epigenetic profile moulded by the cell's developmental history, spatial proximity to stromal cell types, and physical properties of the TME.

EMP programs identified in each sample were consistently associated with higher expression of genes associated with mesenchymal traits (e.g., motility and cellular rearrangement), yet there was no consistent reduction of epithelial genes upon activation of these programs. This supports studies demonstrating the presence of hybrid phenotypes in cancer with distinct functional traits (43–45). It is possible that reduction of epithelial traits is achieved through internalization of epithelial membrane proteins rather than transcription repression, as has been seen with hybrid E/M phenotypes in pancreatic cancer (49). However, the generalizability of posttranslational regulation to the broader scope of phenotypically relevant epithelial proteins is still unclear. This could, however, reconcile the presence of various epithelial membrane proteins in EMP programs, perhaps acting as a feedback mechanism to counter this regulation.

We acknowledge that our ability to define EMP programs may be limited by the sparsity and sensitivity constraints of scRNA-seq assays. It is very possible that some phenotypically relevant expression patterns were not reliably detected in our analysis for technical reasons. For example, we found that some of the canonical EMT transcription factors (e.g., *SNAI1*, *SNAI2*, and *ZEB1*) were poorly detected, and it is unclear whether this is a feature of these samples or a technical limitation. However, other approaches have supported that the involvement of canonical transcription factors is inconsistent (11), and in some settings, their expression could be attributed to the presence of cancer-associated fibroblasts rather than EMP (52). Moreover, given the relevance of posttranscriptional regulation of

EMP (49), it is feasible that some of the most conserved features of EMP reside beyond cells' expression profiles. Advances in morphological and protein-based assays (e.g., high-content imaging, imaging mass cytometry, etc.) will allow for the discovery of additional features of EMP.

A general observation worth noting is that the E/M phenotypes observed within a single sample seem to span a continuous, often monomodal distribution. There has been much speculation about the stability of phenotypes along an E/M continuum and the presence of discrete stable states (62). Modeling frameworks suggest their existence (63, 64), but it is not clear that they are compatible with the distributions of phenotypes observed in these samples: The dominant E/M state between tumors is variable, the extent of phenotypic variation within individual samples is highly variable, and very few samples show the multimodality that should coincide with varying stability along a phenotypic continuum. Although, considering these findings, it is intriguing to question whether it is more relevant to a tumor's biology that its malignant cells are more mesenchymal on average or whether they span a wider range of E/M phenotypes.

We identified a signature of 328 genes that were commonly associated with EMP programs and a refined signature of 128 genes with high specificity to malignant cells, which could be useful for quantifying EMP in gene expression data, as we did here with TCGA data. In no way do we argue that this signature represents the most biologically relevant components of EMP gene expression programs. Enrichment of GO terms related to mesenchymal functions suggests that they likely contribute to the phenotype, but we found a total of 4187 genes associated with EMP across the various samples. Even under highly controlled experimental conditions, EMT responses are vastly context specific (8, 10, 61). The phenotypic changes associated with EMP are likely an emergent property of these many changes, and variation in expression programs may provide phenotypic nuances that we are unaware of. Hence, rather than try to reduce EMP to some consistent phenomenon, we think it is important to recognize context specificity and begin building conceptual models and experimental designs to understand its importance.

The ability to therapeutically restrict plasticity in tumors could greatly improve the efficacy of existing treatment options. One could imagine several strategies for accomplishing this: interfering with the cues that initiate plasticity, therapeutically impairing transcriptional regulatory mechanisms (e.g., inhibiting histone modifying proteins), or targeting effector proteins associated with a given cell state (e.g., neutralizing immunosuppressive cytokines released from mesenchymal cells). The latter may be challenging due to the diversity of EMP phenotypes but is perhaps the most direct strategy for preventing undesirable features of mesenchymal cells. In this study, we have begun to explore strategies for blocking stimulatory pathways, restricting diverse programs with common signaling cues. We computationally inferred that EMP programs were consistently associated with TGF-β and NFκB signaling, whereas some had notably high levels of either MAPK or hypoxia/p53 signaling. Leveraging transcriptomic data from various drug screens, we confirmed that targeting these active pathways led to reduced activity of sample-specific EMP programs despite the diversity of the programs themselves. While this diversity introduces challenges for effective therapeutics, this observation suggests that these diverse programs may have common dependencies that can be inferred from their molecular features and exploited therapeutically.
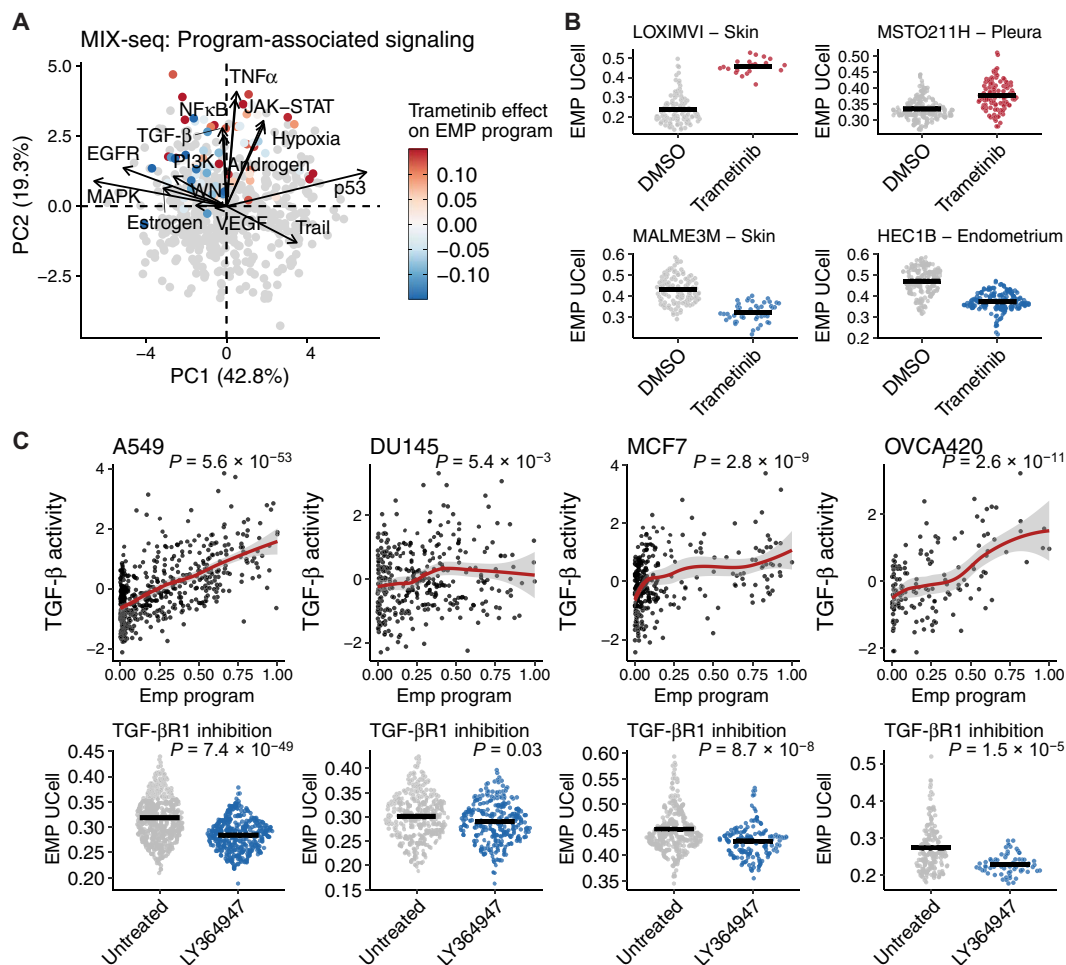
Fig. 6. Regulatory predictions can infer strategies to therapeutically restrict EMP. (**A**) PCA of signaling model coefficients for all archetype programs of the 99 cell lines in the MIX-seq dataset. Gray dots represent programs not associated with EMP. EMP programs are colored by the effect trametinib has on its activity. (**B**) Examples of cell lines whose specific EMP program is enhanced (top) or limited (bottom) by trametinib. DMSO, dimethyl sulfoxide. (**C**) Effect of LY364947-mediated TGF-βR1 inhibition on sample-specific EMP programs in A549, DU145, MCF7, and OVCA420 cell lines. *P* values were calculated from linear models for each condition and were all corrected with the Benjamini-Hochberg method.

In summary, we have used scRNA-seq data from 266 tumors spanning eight cancer types to identify molecular features associated with EMP. We find that EMP is a ubiquitous source of intratumoral heterogeneity but—consistent with previous findings—is highly context specific. We identify a cancer cell–specific signature of the most common genes positively associated with EMP and demonstrate its utility as a general-use gene set, using the TCGA pan-cancer RNA-seq data to associate EMP with worse PFIs and a more immunosuppressive TME. We use computational approaches to infer regulatory features of EMP across hundreds of samples and highlight that diversity may emerge from common regulatory mechanisms that can be inferred and used to rationalize therapeutic strategies.

METHODS
Processing and annotating scRNA-seq data
A summary of scRNA-seq datasets used in this study are described in table S3. To avoid comparing data collected from vastly different technologies, only droplet-based scRNA-seq data were used in the

analysis. Raw UMI (Unique Molecular Identifier) count matrices and cell metadata were collected from the various sources (table S3). Several datasets included matched normal tissue samples. In these cases, we removed the normal samples and only proceeded with the tumor samples.

Initial quality control was performed independently for each dataset using the R package Seurat v4.0 (*65*). Cells with fewer than 200 detected genes were removed, and only genes detected in more than three cells were included in the analysis. Data points representing potential cell doublets within individual samples were removed using scDblFinder v1.6.0 using default parameters (https://github.com/plger/scDblFinder). Cells with a high percentage (>20%) of mitochondrial transcripts were also removed. However, for several datasets, the original authors prefiltered the data at a lower threshold, and for the breast cancer samples from Wu *et al.* (*36*), the original authors had filtered cells with >30% due to the higher average values, but data quality was not apparently affected. In these cases, the original filtering thresholds were used. Table S1 summarizes these thresholds.

UMI counts were normalized with standard library size scaling and log transformation. The data were then processed with principal

components analysis (PCA), a nearest neighbor graph was generated on the first 30 PCs, and the data were clustered at a fairly low resolution (FindClusters, resolution = 0.2) using the Louvain algorithm. For visualizations presented in the figures, UMAP (Uniform manifold approximation and projection) embeddings were generated from the first 30 PCs.

The identity of Individual cells for each sample was annotated using SingleR v1.6.1 (66) and the Human Primary Cell Atlas from the R package celldex v1.2.0. Only expected cell types from the TME were included as references in the annotation. This includes fibroblasts, monocytes, endothelial cells, T cells, NK cells, B cells, macrophages, smooth muscle cells, dendritic cells, platelets, and epithelial cells. Following automated annotation, malignant epithelial cells were identified by inferring copy number variations across the annotated cell types with inferCNV. Epithelial populations with copy number aberrations were considered malignant cells in downstream analysis. While mesenchymal markers may be expressed in fibroblasts, we found that global gene expression profiles were sufficiently different such that cells annotated as fibroblasts formed distinct clusters, well separated from malignant cells.

To determine the quality of the annotation, we assessed the purity of annotations within clusters of cells with similar gene expression profiles. We found strong concordance between clustering patterns and annotations (fig. S1). Similarly, we find that annotations match expression patterns of canonical cell type markers. In some cases, clusters were annotated as two closely related cell types (e.g., fibroblasts and smooth muscle cells). We found that these mixed annotations could be fairly uniformly distributed within a cluster and were not distributed among groups of cells that could be separated by higher-resolution clustering. Although we suspect that these cells are truly a single cell type/state, we used the automated annotations for downstream analysis to ensure consistency. This issue did not occur for epithelial populations, so it does not affect the majority of the analysis.

Identifying latent EMP expression programs with archetypal analysis

To ensure that we had a sufficient number of cells to identify heterogeneous expression patterns within samples, we removed samples with <100 annotated malignant cells from the analysis. Multiresolution archetypal analysis (42) was performed independently of the cancer cells from all 266 tumors using ACTIONet v2.1.7 (42) to decompose cells' gene expression profiles into a small set of latent expression programs that are heterogeneously expressed throughout the population. Reduced kernel matrices were first computed with the reduce.ace() function implemented in the R package ACTIONet with the parameter reduced_dim=20. Given that each population represented a single cell type, ACTIONet was then run with the k_max=10 option to reduce the maximum depth of decompositions and with min_cells_per_arch=5 to prevent archetypes driven by a small number of cells.

Resulting archetype footprints (program activities) were correlated with gene set scores for 11 EMT gene sets. Clustering of the Pearson correlation coefficients allowed us to define EMP-associated programs as clusters with high correlation values. We then used linear models to identify genes whose expression is associated with all archetype programs. To only recover reliable changes, differential expression was limited to the top 2000 variable genes with a minimum detection frequency of 5% in each sample's malignant cell population. Normalized (scaled and log-transformed) counts were used for differential expression, modeling each gene's expression as a function of program activity, and program-associated genes were defined as those with a Benjamini-Hochberg–corrected P value of <0.05 and a model coefficient (effect size) of >1.

Gene set scoring

Gene set scoring was performed using the R package UCell v1.0.0 (67). UCell scores are based on the Mann-Whitney U statistic, which evaluates the rank of each query genes' expression level in individual cells. Because it is rank-based, the scores are independent of the cellular composition of the dataset and are interpretable as the relative ranking of the gene set within the cell's transcriptome. For gene set scoring of the EMP signature, we also performed 100 permutations of random gene sets of an equivalent size to define the distribution of scores expected chance.

Cell type specificity scoring

Specificity scores were calculated for EMP signature genes using the R package genesorteR v0.4.3 (68). Specificity scores represent the exclusivity of a gene to a given cluster and extent to which it is expressed (proportion of cells with detection greater than population median levels). Values range from 0 to 1, with 1 corresponding to genes that are exclusive to a given cluster and detected in all cells.

Pan-cancer TCGA analysis

Bulk RNA-seq profiles and clinical outcomes were accessed from https://gdc.cancer.gov/node/905, tumor purity estimates were acquired from Aran *et al.* (69), and immune cell proportion estimates were accessed from Thorsson *et al.* (70).

RNA-seq count data for each sample were normalized to counts per million and log-transformed. EMP signature scores were calculated for each sample by computing the average gene-level Z score of the genes from the 128 cancer cell–specific genes. The association of EMP signature activity with PFI was assessed using a Cox proportional hazards model, including tumor type, purity, stage, and age along with continuous EMP scores as covariates in the model. Changes in immune cell proportions were assessed by independently using linear models to model each immune cell type's predicted proportion as a function EMP signature score, including tumor type as a covariate to account for differences simply attributable to cancer type. We used the Benjamini-Hochberg method to correct P values to account for multiple comparisons.

Inferring EMP-associated signaling activity

The R package PROGENy v1.11.2 (57, 71, 72) was used to infer the activity of 14 signaling pathways in each individual cell. For each tumor sample, normalized expression values were used to calculate pathway activity with the progeny() function with the top=500 parameter set to use the top 500 genes of each pathway in PROGENy's model for the activity calculation. We used simple linear models to identify changes in signaling pathway activity with archetype program activity.

Cell communication inference

The R package nichenetr (NicheNet; v1.0.0) (56) was used to infer cell communication within the TME that could contribute to a sample's specific EMP program. Given the diversity of EMP programs, this was performed on all 266 tumors independently. For each sample, cancer cells annotated by ACTIONet as the EMP-associated archetype

were defined as the "receiver" population, and all cell types were considered as potential "sender" cells. The "gene set of interest" was defined as that sample's specific EMP program, and expressed genes were defined as those with a detection rate of at least 5%. For each sample, we considered top ligands as the top 10 ligands inferred to promote expression of the EMP-associated genes.

Assessing effects of small-molecule inhibitors on EMP

ScRNA-seq data from epithelial cancer cell lines in control and drug-treated culture conditions were acquired from McFarland *et al.* (*60*) and Cook and Vanderhyden (*8*). Initial data processing was performed identically to tumor samples, and cell lines with fewer than 100 measured cells were removed. EMP programs were defined for each cell line using ACTIONet with only data from control cultured cells. Genes associated with each program were defined and used as a sample-specific EMP gene set for scoring both control and drug-treated cells. We then used a linear model to compare the effects of MEK and TGF-βR inhibition on EMP in each line.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abi7640

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. J. Yang, P. Antin, G. Berx, C. Blanpain, T. Brabletz, M. Bronner, K. Campbell, A. Cano, J. Casanova, G. Christofori, S. Dedhar, R. Derynck, H. L. Ford, J. Fuxe, A. Garcia de Herreros, G. J. Goodall, A.-K. Hadjantonakis, R. J. Y. Huang, R. Kalcheim, R. Kalluri, Y. Kang, Y. Khew-Goodall, H. Levine, J. Liu, G. D. Longmore, S. A. Mani, J. Massagué, R. Mayor, D. McClay, K. E. Mostov, D. F. Newgreen, M. A. Nieto, A. Puisieux, R. Runyan, P. Savagner, B. Stanger, M. P. Stemmler, Y. Takahashi, M. Takeichi, E. Theveneau, J. P. Thiery, E. W. Thompson, R. A. Weinberg, E. D. Williams, J. Xing, B. P. Zhou, G. Sheng; EMT International Association (TEMTIA), Guidelines and definitions for research on epithelial-mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* **21**, 341–352 (2020).

2. H. Gabbert, R. Wagner, R. Moll, C. D. Gerharz, Tumor dedifferentiation: An important step in tumor invasion. *Clin. Exp. Metastasis* **3**, 257–279 (1985).

3. S. V. Puram, I. Tirosh, A. S. Parikh, A. P. Patel, K. Yizhak, S. Gillespie, C. Rodman, C. L. Luo, E. A. Mroz, K. S. Emerick, D. G. Deschler, M. A. Varvares, R. Mylvaganam, O. Rozenblatt-Rosen, J. W. Rocco, W. C. Faquin, D. T. Lin, A. Regev, B. E. Bernstein, Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624.e24 (2017).

4. A. Dongre, R. A. Weinberg, New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat. Rev. Mol. Cell Biol.* **20**, 69–84 (2019).

5. L. A. Horn, K. Fousek, C. Palena, Tumor plasticity and resistance to immunotherapy. *Trends Cancer Res.* **6**, 432–441 (2020).

6. V. Ramesh, T. Brabletz, P. Ceppi, Targeting EMT in cancer with repurposed metabolic inhibitors. *Trends Cancer* **6**, 942–950 (2020).

7. S. Bhatia, P. Wang, A. Toh, E. W. Thompson, New insights into the role of phenotypic plasticity and EMT in driving cancer progression. *Front. Mol. Biosci.* **7**, 71 (2020).

8. D. P. Cook, B. C. Vanderhyden, Context specificity of the EMT transcriptional response. *Nat. Commun.* **11**, 2142 (2020).

9. J. H. Taube, J. I. Herschkowitz, K. Komurov, A. Y. Zhou, S. Gupta, J. Yang, K. Hartwell, T. T. Onder, P. B. Gupta, K. W. Evans, B. G. Hollier, P. T. Ram, E. S. Lander, J. M. Rosen, R. A. Weinberg, S. A. Mani, Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15449–15454 (2010).

10. P. Peixoto, A. Etcheverry, M. Aubry, A. Missey, C. Lachat, J. Perrard, E. Hendrick, R. Delage-Mourroux, J. Mosser, C. Borg, J.-P. Feugeas, M. Herfs, M. Boyer-Guittaut, E. Hervouet, EMT is associated with an epigenetic signature of ECM remodeling genes. *Cell Death Dis.* **10**, 205 (2019).

11. M. P. Stemmler, R. L. Eccles, S. Brabletz, T. Brabletz, Non-redundant functions of EMT transcription factors. *Nat. Cell Biol.* **21**, 102–112 (2019).

12. E. D. Williams, D. Gao, A. Redfern, E. W. Thompson, Controversies around epithelial-mesenchymal plasticity in cancer metastasis. *Nat. Rev. Cancer* **19**, 716–732 (2019).

13. K. R. Fischer, A. Durrans, S. Lee, J. Sheng, F. Li, S. T. C. Wong, H. Choi, T. El Rayes, S. Ryu, J. Troeger, R. F. Schwabe, L. T. Vahdat, N. K. Altorki, V. Mittal, D. Gao, Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472–476 (2015).

14. X. Zheng, J. L. Carstens, J. Kim, M. Scheible, J. Kaye, H. Sugimoto, C.-C. Wu, V. S. LeBleu, R. Kalluri, Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**, 525–530 (2015).

15. N. M. Aiello, T. Brabletz, Y. Kang, M. A. Nieto, R. A. Weinberg, B. Z. Stanger, Upholding a role for EMT in pancreatic cancer metastasis. *Nature* **547**, E7–E8 (2017).

16. X. Ye, T. Brabletz, Y. Kang, G. D. Longmore, M. A. Nieto, B. Z. Stanger, J. Yang, R. A. Weinberg, Upholding a role for EMT in breast cancer metastasis. *Nature* **547**, E1–E3 (2017).

17. K. R. Fischer, N. K. Altorki, V. Mittal, D. Gao, Fischer et al. reply. *Nature* **547**, E5–E6 (2017).

18. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).

19. A. M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz, M. W. Kirschner, Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).

20. A. L. Ji, A. J. Rubin, K. Thrane, S. Jiang, D. L. Reynolds, R. M. Meyers, M. G. Guo, B. M. George, A. Mollbrink, J. Bergenstråhle, L. Larsson, Y. Bai, B. Zhu, A. Bhaduri, J. M. Meyers, X. Rovira-Clavé, S. T. Hollmig, S. Z. Aasi, G. P. Nolan, J. Lundeberg, P. A. Khavari, Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* **182**, 1661–1662 (2020).

21. A. Sharma, E. Y. Cao, V. Kumar, X. Zhang, H. S. Leong, A. M. L. Wong, N. Ramakrishnan, M. Hakimullah, H. M. V. Teo, F. T. Chong, S. Chia, M. T. Thangavelu, X. L. Kwang, R. Gupta, J. R. Clark, G. Periyasamy, N. Gopalakrishna Iyer, R. DasGupta, Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy. *Nat. Commun.* **9**, 4931 (2018).

22. S. Raghavan, P. S. Winter, A. W. Navia, H. L. Williams, A. DenAdel, R. L. Kalekar, J. Galvez-Reyes, K. E. Lowder, N. Mulugeta, M. S. Raghavan, A. A. Borah, S. A. Väyrynen, A. D. Costa, R. W. S. Ng, J. Wang, E. Reilly, D. Y. Ragon, L. K. Brais, A. M. Jaeger, L. F. Spurr, Y. Y. Li, A. D. Cherniack, I. Wakiro, A. Rotem, B. E. Johnson, J. M. McFarland, E. T. Sicinska, T. E. Jacks, T. E. Clancy, K. Perez, D. A. Rubinson, K. Ng, J. M. Cleary, L. Crawford, S. R. Manalis, J. A. Nowak, B. M. Wolpin, W. C. Hahn, A. J. Aguirre, A. K. Shalek, Transcriptional subtype-specific microenvironmental crosstalk and tumor cell plasticity in metastatic pancreatic cancer. doi:10.1101/2020.08.25.256214.

23. K. Ganesh, H. Basnet, Y. Kaygusuz, A. M. Laughney, L. He, R. Sharma, K. P. O'Rourke, V. P. Reuter, Y.-H. Huang, M. Turkekul, E. Emrah, I. Masilionis, K. Manova-Todorova, M. R. Weiser, L. B. Saltz, J. Garcia-Aguilar, R. Koche, S. W. Lowe, D. Pe'er, J. Shia, J. Massagué, L1CAM defines the regenerative origin of metastasis-initiating cells in colorectal cancer. *Nat. Cancer.* **1**, 28–45 (2020).

24. A. M. Laughney, J. Hu, N. R. Campbell, S. F. Bakhoum, M. Setty, V.-P. Lavallée, Y. Xie, I. Masilionis, A. J. Carr, S. Kottapalli, V. Allaj, M. Mattar, N. Rekhtman, J. B. Xavier, L. Mazutis, J. T. Poirier, C. M. Rudin, D. Pe'er, J. Massagué, Regenerative lineages and immune-mediated pruning in lung cancer metastasis. *Nat. Med.* **26**, 259–269 (2020).

25. G. S. Kinker, A. C. Greenwald, R. Tal, Z. Orlova, M. S. Cuoco, J. M. McFarland, A. Warren, C. Rodman, J. A. Roth, S. A. Bender, B. Kumar, J. W. Rocco, P. A. C. M. Fernandes, C. C. Mader, H. Keren-Shaul, A. Plotnikov, H. Barr, A. Tsherniak, O. Rozenblatt-Rosen, V. Krizhanovsky, S. V. Puram, A. Regev, I. Tirosh, Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat. Genet.* **52**, 1208–1218 (2020).

26. H.-O. Lee, Y. Hong, H. E. Etlioglu, Y. B. Cho, V. Pomella, B. Van den Bosch, J. Vanhecke, S. Verbandt, H. Hong, J.-W. Min, N. Kim, H. H. Eum, J. Qian, B. Boeckx, D. Lambrechts, P. Tsantoulis, G. De Hertogh, W. Chung, T. Lee, M. An, H.-T. Shin, J.-G. Joung, M.-H. Jung, G. Ko, P. Wirapati, S. H. Kim, H. C. Kim, S. H. Yun, I. B. H. Tan, B. Ranjan, W. Y. Lee, T.-Y. Kim, J. K. Choi, Y.-J. Kim, S. Prabhakar, S. Tejpar, W.-Y. Park, Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat. Genet.* **52**, 594–603 (2020).

27. F. Uhlitz, P. Bischoff, S. Peidli, A. Sieber, B. Obermayer, E. Blanc, A. Trinks, M. Lüthen, Y. Ruchiy, T. Sell, S. Mamlouk, R. Arsie, T.-T. Wei, K. Klotz-Noack, R. F. Schwarz, B. Sawitzki, C. Kamphues, D. Beule, M. Landthaler, C. Sers, D. Horst, N. Blüthgen, M. Morkel, Mitogen-activated protein kinase activity drives cell trajectories in colorectal cancer. *EMBO Mol. Med.* **13**, e14123 (2021).

28. J. Qian, S. Olbrecht, B. Boeckx, H. Vos, D. Laoui, E. Etlioglu, E. Wauters, V. Pomella, S. Verbandt, P. Busschaert, A. Bassez, A. Franken, M. V. Bempt, J. Xiong, B. Weynand, Y. van Herck, A. Antoranz, F. M. Bosisio, B. Thienpont, G. Floris, I. Vergote, A. Smeets, S. Tejpar, D. Lambrechts, A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Res.* **30**, 745–762 (2020).

29. A. Sathe, S. M. Grimes, B. T. Lau, J. Chen, C. Suarez, R. J. Huang, G. Poultsides, H. P. Ji, Single-cell genomic characterization reveals the cellular reprogramming of the gastric tumor microenvironment. *Clin. Cancer Res.* **26**, 2640–2653 (2020).

30. D. Lambrechts, E. Wauters, B. Boeckx, S. Aibar, D. Nittner, O. Burton, A. Bassez, H. Decaluwé, A. Pircher, K. Van den Eynde, B. Weynand, E. Verbeken, P. De Leyn, A. Liston,

J. Vansteenkiste, P. Carmeliet, S. Aerts, B. Thienpont, Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* **24**, 1277–1289 (2018).

31. N. Kim, H. K. Kim, K. Lee, Y. Hong, J. H. Cho, J. W. Choi, J.-I. Lee, Y.-L. Suh, B. M. Ku, H. H. Eum, S. Choi, Y.-L. Choi, J.-G. Joung, W.-Y. Park, H. A. Jung, J.-M. Sun, S.-H. Lee, J. S. Ahn, K. Park, M.-J. Ahn, H.-O. Lee, Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat. Commun.* **11**, 2285 (2020).

32. F. Wu, J. Fan, Y. He, A. Xiong, J. Yu, Y. Li, Y. Zhang, W. Zhao, F. Zhou, W. Li, J. Zhang, X. Zhang, M. Qiao, G. Gao, S. Chen, X. Chen, X. Li, L. Hou, C. Wu, C. Su, S. Ren, M. Odenthal, R. Buettner, N. Fang, C. Zhou, Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat. Commun.* **12**, 2540 (2021).

33. Y.-P. Chen, J.-H. Yin, W.-F. Li, H.-J. Li, D.-P. Chen, C.-J. Zhang, J.-W. Lv, Y.-Q. Wang, X.-M. Li, J.-Y. Li, P.-P. Zhang, Y.-Q. Li, Q.-M. He, X.-J. Yang, Y. Lei, L.-L. Tang, G.-Q. Zhou, Y.-P. Mao, C. Wei, K.-X. Xiong, H.-B. Zhang, S.-D. Zhu, Y. Hou, Y. Sun, M. Dean, I. Amit, K. Wu, D.-M. Kuang, G.-B. Li, N. Liu, J. Ma, Single-cell transcriptomics reveals regulators underlying immune cell diversity and immune subtypes associated with prognosis in nasopharyngeal carcinoma. *Cell Res.* **30**, 1024–1042 (2020).

34. L. Geistlinger, S. Oh, M. Ramos, L. Schiffer, R. S. LaRue, C. M. Henzler, S. A. Munro, C. Daughters, A. C. Nelson, B. J. Winterhoff, Z. Chang, S. Talukdar, M. Shetty, S. A. Mullany, M. Morgan, G. Parmigiani, M. J. Birrer, L.-X. Qin, M. Riester, T. K. Starr, L. Waldron, Multiomic analysis of subtype evolution and heterogeneity in high-grade serous ovarian carcinoma. *Cancer Res.* **80**, 4335–4345 (2020).

35. N. G. Steele, E. S. Carpenter, S. B. Kemp, V. R. Sirihorachai, S. The, L. Delrosario, J. Lazarus, E.-A. D. Amir, V. Gunchick, C. Espinoza, S. Bell, L. Harris, F. Lima, V. Irizarry-Negron, D. Paglia, J. Macchia, A. K. Y. Chu, H. Schofield, E.-J. Wamsteker, R. Kwon, A. Schulman, A. Prabhu, R. Law, A. Sondhi, J. Yu, A. Patel, K. Donahue, H. Nathan, C. Cho, M. A. Anderson, cV. Sahai, C. A. Lyssiotis, W. Zou, B. L. Allen, A. Rao, H. C. Crawford, F. Bednar, T. L. Frankel, M. P. di Magliano, Multimodal mapping of the tumor and peripheral blood immune landscape in human pancreatic cancer. *Nat. Cancer.* **1**, 1097–1112 (2020).

36. S. Z. Wu, D. L. Roden, C. Wang, H. Holliday, K. Harvey, A. S. Cazet, K. J. Murphy, B. Pereira, G. Al-Eryani, N. Bartonicek, R. Hou, J. R. Torpy, S. Junankar, C.-L. Chan, C. E. Lam, M. N. Hui, L. Gluch, J. Beith, A. Parker, E. Robbins, D. Segara, C. Mak, C. Cooper, S. Warrier, A. Forrest, J. Powell, S. O'Toole, T. R. Cox, P. Timpson, E. Lim, X. S. Liu, A. Swarbrick, Stromal cell diversity associated with immune evasion in human triple-negative breast cancer. *EMBO J.* **39**, e104063 (2020).

37. A. Bassez, H. Vos, L. Van Dyck, G. Floris, I. Arijs, C. Desmedt, B. Boeckx, M. Vanden Bempt, I. Nevelsteen, K. Lambein, K. Punie, P. Neven, A. D. Garg, H. Wildiers, J. Qian, A. Smeets, D. Lambrechts, A single-cell map of intratumoral changes during anti-PD1 treatment of patients with breast cancer. *Nat. Med.* **27**, 820–832 (2021).

38. A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, P. Tamayo, The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).

39. H. Yuan, M. Yan, G. Zhang, W. Liu, C. Deng, G. Liao, L. Xu, T. Luo, H. Yan, Z. Long, A. Shi, T. Zhao, Y. Xiao, X. Li, CancerSEA: A cancer single-cell state atlas. *Nucleic Acids Res.* **47**, D900–D908 (2019).

40. M. Zhao, L. Kong, Y. Liu, H. Qu, dbEMT: An epithelial-mesenchymal transition associated gene resource. *Sci. Rep.* **5**, 11459 (2015).

41. T. Z. Tan, Q. H. Miow, Y. Miki, T. Noda, S. Mori, R. Y.-J. Huang, J. P. Thiery, Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol. Med.* **6**, 1279–1293 (2014).

42. S. Mohammadi, J. Davila-Velderrain, M. Kellis, A multiresolution framework to characterize single-cell state landscapes. *Nat. Commun.* **11**, 5399 (2020).

43. I. Pastushenko, A. Brisebarre, A. Sifrim, M. Fioramonti, T. Revenco, S. Boumahdi, A. Van Keymeulen, D. Brown, V. Moers, S. Lemaire, S. De Clercq, E. Minguijón, C. Balsat, Y. Sokolow, C. Dubois, F. De Cock, S. Scozzaro, F. Sopena, A. Lanas, N. D'Haene, I. Salmon, J.-C. Marine, T. Voet, P. A. Sotiropoulou, C. Blanpain, Identification of the tumour transition states occurring during EMT. *Nature* **556**, 463–468 (2018).

44. C. Kröger, A. Afeyan, J. Mraz, E. N. Eaton, F. Reinhardt, Y. L. Khodor, P. Thiru, B. Bierie, X. Ye, C. B. Burge, R. A. Weinberg, Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 7353–7362 (2019).

45. I. Pastushenko, C. Blanpain, EMT transition states during tumor progression and metastasis. *Trends Cell Biol.* **29**, 212–226 (2019).

46. K. Williams, R. Ghosh, P. V. Giridhar, G. Gu, T. Case, S. M. Belcher, S. Kasper, Inhibition of stathmin1 accelerates the metastatic process. *Cancer Res.* **72**, 5407–5417 (2012).

47. M. A. B. Rosales, D. Y. Shu, J. Iacovelli, M. Saint-Geniez, Loss of PGC-1α in RPE induces mesenchymal transition and promotes retinal degeneration. *Life Sci. Alliance* **2**, e201800212 (2019).

48. J. H. Tsai, J. Yang, Epithelial-mesenchymal plasticity in carcinoma metastasis. *Genes Dev.* **27**, 2192–2206 (2013).

49. N. M. Aiello, R. Maddipati, R. J. Norgard, D. Balli, J. Li, S. Yuan, T. Yamazoe, T. Black, A. Sahmoud, E. E. Furth, D. Bar-Sagi, B. Z. Stanger, EMT subtype influences epithelial plasticity and mode of cell migration. *Dev. Cell.* **45**, 681–695.e4 (2018).

50. B. Izar, I. Tirosh, E. H. Stover, I. Wakiro, M. S. Cuoco, I. Alter, C. Rodman, R. Leeson, M.-J. Su, P. Shah, M. Iwanicki, S. R. Walker, A. Kanodia, J. C. Melms, S. Mei, J.-R. Lin, C. B. M. Porter, M. Slyper, J. Waldman, L. Jerby-Arnon, O. Ashenberg, T. J. Brinker, C. Mills, M. Rogava, S. Vigneau, P. K. Sorger, L. A. Garraway, P. A. Konstantinopoulos, J. F. Liu, U. Matulonis, B. E. Johnson, O. Rozenblatt-Rosen, A. Rotem, A. Regev, A single-cell landscape of high-grade serous ovarian cancer. *Nat. Med.* **26**, 1271–1279 (2020).

51. C. Isella, A. Terrasi, S. E. Bellomo, C. Petti, G. Galatola, A. Muratore, A. Mellano, R. Senetta, A. Cassenti, C. Sonetto, G. Inghirami, L. Trusolino, Z. Fekete, M. De Ridder, P. Cassoni, G. Storme, A. Bertotti, E. Medico, Stromal contribution to the colorectal cancer transcriptome. *Nat. Genet.* **47**, 312–319 (2015).

52. M. Tyler, I. Tirosh, Decoupling epithelial-mesenchymal transitions from stromal profiles by integrative expression analysis. *Nat. Commun.* **12**, 2592 (2021).

53. J. Peng, B.-F. Sun, C.-Y. Chen, J.-Y. Zhou, Y.-S. Chen, H. Chen, L. Liu, D. Huang, J. Jiang, G.-S. Cui, Y. Yang, W. Wang, D. Guo, M. Dai, J. Guo, T. Zhang, Q. Liao, Y. Liu, Y.-L. Zhao, D.-L. Han, Y. Zhao, Y.-G. Yang, W. Wu, Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* **29**, 725–738 (2019).

54. K. A. Hoadley, C. Yau, T. Hinoue, D. M. Wolf, A. J. Lazar, E. Drill, R. Shen, A. M. Taylor, A. D. Cherniack, V. Thorsson, R. Akbani, R. Bowlby, C. K. Wong, M. Wiznerowicz, F. Sanchez-Vega, A. G. Robertson, B. G. Schneider, M. S. Lawrence, H. Noushmehr, T. M. Malta; Cancer Genome Atlas Network, J. M. Stuart, C. C. Benz, P. W. Laird, Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* **173**, 291–304.e6 (2018).

55. V. Thorsson, D. L. Gibbs, S. D. Brown, D. Wolf, D. S. Bortone, T.-H. O. Yang, E. Porta-Pardo, G. F. Gao, C. L. Plaisier, J. A. Eddy, E. Ziv, A. C. Culhane, E. O. Paull, I. K. A. Sivakumar, A. J. Gentles, R. Malhotra, F. Farshidfar, A. Colaprico, J. S. Parker, L. E. Mose, N. S. Vo, J. Liu, Y. Liu, J. Rader, V. Dhankani, S. M. Reynolds, R. Bowlby, A. Califano, A. D. Cherniack, D. Anastassiou, D. Bedognetti, Y. Mokrab, A. M. Newman, A. Rao, K. Chen, A. Krasnitz, H. Hu, T. M. Malta, H. Noushmehr, C. S. Pedamallu, S. Bullman, A. I. Ojesina, A. Lamb, W. Zhou, H. Shen, T. K. Choueiri, J. N. Weinstein, J. Guinney, J. Saltz, R. A. Holt, C. S. Rabkin; Cancer Genome Atlas Research Network, A. J. Lazar, J. S. Serody, E. G. Demicco, M. L. Disis, B. G. Vincent, I. Shmulevich, The immune landscape of cancer. *Immunity* **48**, 812–830.e14 (2018).

56. R. Browaeys, W. Saelens, Y. Saeys, NicheNet: Modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).

57. M. Schubert, B. Klinger, M. Klünemann, A. Sieber, F. Uhlitz, S. Sauer, M. J. Garnett, N. Blüthgen, J. Saez-Rodriguez, Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat. Commun.* **9**, 20 (2018).

58. J. Cursons, K.-J. Leuchowius, M. Waltham, E. Tomaskovic-Crook, M. Foroutan, C. P. Bracken, A. Redfern, E. J. Crampin, I. Street, M. J. Davis, E. W. Thompson, Stimulus-dependent differences in signalling regulate epithelial-mesenchymal plasticity and change the effects of drugs in breast cancer cell lines. *Cell Commun. Signal.* **13**, 26 (2015).

59. S. Aibar, C. B. González-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J.-C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters, S. Aerts, SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).

60. J. M. McFarland, B. R. Paolella, A. Warren, K. Geiger-Schuller, T. Shibue, M. Rothberg, O. Kuksenko, W. N. Colgan, A. Jones, E. Chambers, D. Dionne, S. Bender, B. M. Wolpin, M. Ghandi, I. Tirosh, O. Rozenblatt-Rosen, J. A. Roth, T. R. Golub, A. Regev, A. J. Aguirre, F. Vazquez, A. Tsherniak, Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action. *Nat. Commun.* **11**, 4296 (2020).

61. J. L. McFaline-Figueroa, A. J. Hill, X. Qiu, D. Jackson, J. Shendure, C. Trapnell, A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. *Nat. Genet.* **51**, 1389–1398 (2019).

62. M. A. Nieto, R. Y.-J. Huang, R. A. Jackson, J. P. Thiery, EMT: 2016. *Cell* **166**, 21–45 (2016).

63. S. Zadran, R. Arumugam, H. Herschman, M. E. Phelps, R. D. Levine, Surprisal analysis characterizes the free energy time course of cancer cells undergoing epithelial-to-mesenchymal transition. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13235–13240 (2014).

64. D. Jia, M. K. Jolly, S. C. Tripathi, P. Den Hollander, B. Huang, M. Lu, M. Celiktas, E. Ramirez-Peña, E. Ben-Jacob, J. N. Onuchic, S. M. Hanash, S. A. Mani, H. Levine, Distinguishing mechanisms underlying EMT tristability. *Cancer Converg.* **1**, 2 (2017).

65. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zagar, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. B. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *bioRxiv*, 2020.10.12.335331, (2020).

66. D. Aran, A. P. Looney, L. Liu, E. Wu, V. Fong, A. Hsu, S. Chak, R. P. Naikawadi, P. J. Wolters, A. R. Abate, A. J. Butte, M. Bhattacharya, Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).

67. M. Andreatta, S. J. Carmona, UCell: Robust and scalable single-cell gene signature scoring. *Comput. Struct. Biotechnol. J.* **19**, 3796–3798 (2021).

68. M. M. Ibrahim, R. Kramann, genesorteR: Feature ranking in clustered single cell data. *bioRxiv*, 676379 (2019).

69. D. Aran, M. Sirota, A. J. Butte, Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **6**, 8971 (2015).

70. V. Thorsson, D. L. Gibbs, S. D. Brown, D. Wolf, D. S. Bortone, T.-H. O. Yang, E. Porta-Pardo, G. F. Gao, C. L. Plaisier, J. A. Eddy, E. Ziv, A. C. Culhane, E. O. Paull, I. K. A. Sivakumar, A. J. Gentles, R. Malhotra, F. Farshidfar, A. Colaprico, J. S. Parker, L. E. Mose, N. S. Vo, J. Liu, Y. Liu, J. Rader, V. Dhankani, S. M. Reynolds, R. Bowlby, A. Califano, A. D. Cherniack, D. Anastassiou, D. Bedognetti, Y. Mokrab, A. M. Newman, A. Rao, K. Chen, A. Krasnitz, H. Hu, T. M. Malta, H. Noushmehr, C. S. Pedamallu, S. Bullman, A. I. Ojesina, A. Lamb, W. Zhou, H. Shen, T. K. Choueiri, J. N. Weinstein, J. Guinney, J. Saltz, R. A. Holt, C. S. Rabkin; Cancer Genome Atlas Research Network, A. J. Lazar, J. S. Serody, E. G. Demicco, M. L. Disis, B. G. Vincent, I. Shmulevich, The immune landscape of cancer. *Immunity* **51**, 411–412 (2019).

71. C. H. Holland, B. Szalai, J. Saez-Rodriguez, Transfer of regulatory knowledge from human to mouse for functional genomics analysis. *Biochim. Biophys. Acta Gene Regul. Mech.* **1863**, 194431 (2020).

72. C. H. Holland, J. Tanevski, J. Perales-Patón, J. Gleixner, M. P. Kumar, E. Mereu, B. A. Joughin, O. Stegle, D. A. Lauffenburger, H. Heyn, B. Szalai, J. Saez-Rodriguez, Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol.* **21**, 36 (2020).