Modeling the dynamics of EMT reveals genes associated with pan-cancer intermediate states and plasticity

MeiLu McDermott[1], Riddhee Mehta[1], Evanthia T. Roussos Torres[2] & Adam L. MacLean[1,*]

[1]Department of Quantitative and Computational Biology, Dornsife College of Letters, Arts and Sciences, University of Southern California, Los Angeles, CA 90089, USA
[2]Department of Medicine, Division of Medical Oncology, Keck School of Medicine, Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90033, USA
[*]Correspondence: macleana@usc.edu (A.L.M.)

Abstract

Epithelial-mesenchymal transition (EMT) is a developmental cell state transition co-opted by cancer to drive metastasis and resistance. Stable EMT intermediate states play a particularly important role in cell state plasticity and confer metastatic potential. To explore the dynamics of EMT and identify marker genes of highly metastatic intermediate cells, we analyzed EMT across multiple tumor types and stimuli via mathematical modeling with single-cell RNA-sequencing (scRNA-seq) data. We identified pan-cancer genes consistently expressed or upregulated in EMT intermediate states, most of which were not previously annotated as markers of EMT. Using Bayesian parameter inference, we fit a simple mathematical model to scRNA-seq data, revealing tumor-specific transition rates. This mathematical model offers a framework to quantify EMT progression. A consensus analysis of differential intermediate expression, regulation, and model-derived dynamics identified marker genes associated with persistence of the intermediate EMT state. *SFN* and *NRG1* emerged as genes with the strongest evidence for their role influencing intermediate EMT dynamics. Through analysis of an independent cell line, we verified the role of *SFN* as a marker intermediate EMT transition. Modeling and inference of genes associated with EMT dynamics offer means to find biomarkers and to identify therapeutic approaches to harness or reverse tumor-promoting cell state transitions driven by EMT.

Modeling the dynamics of EMT reveals genes associated with pan-cancer intermediate states and plasticity

1 INTRODUCTION

Cell state transitions are phenotypic changes in the state of a cell, primarily driven by transcriptional programs. Such phenotypic transitions underlie development, regeneration, and cancer. Our ability to interrogate cell state transitions and their consequences has dramatically increased with advances in single-cell genomics (Rood et al., 2022). We can dissect the timing of key events as cells change state (Qiu, Martin, et al., 2024) and identify transient or intermediate states (MacLean, Hong, et al., 2018). Efforts to produce a comprehensive catalogue of cell states are underway (Regev et al., 2017), yet large gaps in our understanding remain: both regarding cell states and even more regarding the transitions they undertake. We do not have satisfactory explanations of what are the initiating factors of a cell state transition, nor what is the relationship between the dynamics of cell phenotypic change and the transcriptional dynamics acting within the cell.

The epithelial-to-mesenchymal transition (EMT), during which epithelial cells become mesenchymal or mesenchymal-like (Thiery, 2003), is an exemplary cell state transition. EMT is necessary during development and wound healing and is co-opted by cancer, where it is a crucial component of metastasis. Understanding EMT is thus imperative to slowing or preventing metastasis, the leading cause of death from cancer (Dillekås et al., 2019). Classical conceptions of EMT characterize a binary process, with cells being either completely epithelial or mesenchymal (Thiery, 2003). However, experimental and theoretical studies have demonstrated the existence of EMT intermediate states (Bracken et al., 2008; Deshmukh et al., 2021; Hong, Watanabe, et al., 2015; Hong and Xing, 2024; Sha, Haensel, et al., 2019; Xing et al., 2019). Pan-cancer studies of intermediate EMT states have revealed insight into transcriptomic signatures underlying EMT transformation (Tagliazucchi et al., 2023). The intermediate state displays partial EMT phenotypes, with characteristics of both epithelial and mesenchymal states, and may also be called partial EMT, hybrid EMT, or an E/M state (Nieto et al., 2016). EMT intermediate states are closely tied with the concept of epithelial-mesenchymal plasticity (EMP): dynamic, bidirectional transitions through multiple EMT states.

EMT intermediate states are found in both non-malignant EMT and cancer (Nieto et al., 2016; Shaw et al., 2016). The relevance of targeting these states is compelling: EMT intermediate states have been associated with circulating tumor cells (Ruscetti et al., 2015; Ting et al., 2014; Yu et al., 2013) and metastasis (Hendrix et al., 1997), perhaps even more potently than mesenchymal cells alone (Jolly et al., 2015; Simeonov et al., 2021). We focus here on stable EMT intermediate states: biologically, this refers to cells in a state that can be isolated and persist under sufficient conditions; mathematically, stability is defined via the Lyapunov exponents of a dynamical system (Jost et al., 2015). EMT intermediate states have been described as "metastable" in the literature, which in this case refers to stable cell states with small basins of attraction. EMT intermediate state cells may be hard to observe in part due to their rarity (small population sizes or small basins of attraction) or their location (existing at the margins rather than throughout a tissue [Leggett et al., 2021]), although they are not necessarily a minority of cells in a sample.

Mathematical models of EMT have predicted and identified intermediate states, using transcriptional networks that can successfully capture both the steady states of the system and its dynamic properties (Chaffer et al., 2016; Hong, Watanabe, et al., 2015; Lu et al., 2013; Medici et al., 2008; Tian et al., 2013; Zhang, Tian, et al., 2014). These transcription models

of EMT, typically regulated by transforming growth factor-beta (TGF-β), primarily focus on a core network with transcription factors ZEB, SNAIL, and OVOL, and micro-RNAs miR200 and miR34. Although greater attention has been paid to the transcriptional dynamics, there has also been mathematical modeling of the cell population dynamics during EMT (reviewed in Tripathi et al., 2021).

Integrating single-cell genomics with mathematical models offers means to infer dynamic properties from high-dimensional systems (Cho et al., 2022; Wu et al., 2023). EMT, with its relatively straightforward trajectory (non-branching, non-cyclical), lends itself well to analysis via trajectory inference (pseudotime) (Saelens et al., 2019), albeit not taking into account the spatial components of the cell fate decisions which can be decidedly more complex (MacLean, Smith, et al., 2017). Trajectory inference coupled with mathematical modeling has led to insight into the initiation and timing of EMT (Sha, Wang, Zhou, et al., 2020). Despite limitations in inferring Markovian cell dynamics from single-cell data (Weinreb et al., 2018), experimental methodology such as metabolic labeling (Qiu, Zhang, et al., 2022) or lineage tracing (Simeonov et al., 2021) can overcome these challenges. Here we take an alternative approach to inferring the population dynamics model directly from data (Fischer et al., 2019; Weinreb et al., 2018), and (in keeping with the observation that cell state transition dynamics are non-Markovian [Stumpf et al., 2017]) we propose a population model of EMT cell state transitions *a priori*. We subsequently learn rates of cell state transition for each individual sample via Bayesian parameter inference of the cell dynamics over pseudotime.

Here we use single-cell RNA sequencing (scRNA-seq) data to fit mathematical models of EMT population dynamics across various tumor types and stimuli. Parameter inference across these different conditions reveals shared and distinct properties of the routes of EMT. We identify shared genes associated with EMT intermediate states across tumor types via differential expression and differential RNA velocity analyses. By comparing intermediate state genes with inferred EMT parameters, we identify genes associated with EMT dynamics – that is, genes that speed up or slow down EMT. We confirm top predictions by an independent analysis of EMT in a new cell type, demonstrating how these methods offer novel means to identify biomarkers or potential targets during cell state transitions.

2 METHODS

2.1 Single-cell data acquisition and processing

2.1.1 Data sources

In this study, we conducted an integrated analysis of several single-cell RNA sequencing (scRNA-seq) datasets in the public domain. We included datasets from Pastushenko et al., 2018 (GEO accession GSE110357); van Dijk et al., 2018 (GSE114397); Cook and Vanderhyden, 2020 (GSE147405); and Panchy et al., 2022 (GSE213753). For data from Cook and Vanderhyden (2020), samples collected after the removal of the EMT stimulus were not included. For data from Panchy et al. (2022), unstimulated cells were not included.

2.1.2 Sequence alignment

Data from Cook and Vanderhyden (2020) were re-aligned to obtain spliced and unspliced read counts for RNA velocity analysis below. Raw sequence files (accession SRP253729) were downloaded from the NIH Sequence Read Archive using the SRA Toolkit (Leinonen et al., 2011) and converted from SRA to FASTQ files using `fasterq-dump`. Python package cutadapt was used to trim the barcode sequences to 26 base pairs (Martin, 2011). The splici (spliced+intron) index was constructed using the GRCh38 human reference genome with Python package salmon (Srivastava et al., 2019). Sequence pseudoalignment was performed with salmon alevin-fry. Barcode demultiplexing was carried out using the R package MULTIseq (McGinnis et al., 2019). Contaminant cells in the OVCA420 samples were removed as noted by the original authors.

2.1.3 scRNA-seq data preprocessing and normalization

All scRNA-seq data were processed and analyzed using Scanpy (Wolf et al., 2018). Cells with fewer than 200 genes and genes expressed in fewer than three cells were filtered out. Cells with high mitochondrial percentages or disproportionately high total read counts were excluded based on dataset-specific cutoffs (Supp. Table 1). In HMLE samples stimulated with TGF-β, cells with disproportionately high ribosomal percentages were filtered out (<1% of cells). Counts were normalized to 10,000 and $log(x+1)$ transformed. Batch correction for samples from Cook and Vanderhyden (2020) was performed using ComBat in Scanpy (Johnson et al., 2007). Cell cycle effects, which significantly impacted clustering by EMT state identity, were regressed out (Satija et al., 2015; Tirosh et al., 2016), similar to the original analyses. Additional preprocessing included regressing out total counts and percent mitochondrial counts per cell, scaling counts to uniform variance, and selecting highly variable genes for downstream analysis.

2.1.4 Cell clustering and scoring by EMT status

Principal component analysis (PCA) was performed, and the top 15 components were used to construct a nearest neighbor graph. Based on this graph, cell clustering was conducted using the Leiden algorithm (Traag et al., 2019) with dataset-specific resolutions (Supp. Table 1). Differentially expressed genes for each cell cluster were identified using Wilcoxon rank-sum test with Benjamini-Hochberg correction. Cell clusters were visualized in two dimensions using UMAP and PHATE (McInnes et al., 2020; Moon et al., 2019).

To infer the EMT status of single cells based on a set of EMT marker genes, an "EMTscore" was created using the UCell scoring method (Andreatta et al., 2021) with the Hallmark EMT gene set from the Molecular Signatures Database (MSigDB) (Liberzon et al., 2015; Subramanian et al., 2005). UCell calculates single-cell gene expression scores from a gene set using a rank-based approach, which we found to effectively quantify EMT across disparate tumor types and experimental conditions. Genes were input into UCell as filtered and normalized counts.

2.1.5 Identifying shared EMT intermediate state genes

Genes were included in the intermediate state analysis if they were differentially expressed (DE) in an intermediate state with a Benjamini-Hochberg adjusted Wilcoxon rank-sum p-value of $p <$

0.01, up to a maximum of 500 genes per sample. To account for the complexities of comparing gene expression across different datasets and conditions (e.g. batch effects, instrumentation, sequencing depth), we calculated $\log_2$ fold change ($\log_2$FC) values of intermediate state genes ins Scanpy, i.e. according to (following the notation of (Moses et al., 2023)):

$$\log_2 \text{FC of gene } g = \log_2 \left(\exp \left(\frac{1}{n_1} \sum_{i \in G_1} Y_{ig} \right) - 1 + \epsilon \right) - \log_2 \left(\exp \left(\frac{1}{n_2} \sum_{i \in G_2} Y_{ig} \right) - 1 + \epsilon \right)$$

where G_1 is the focal group of cells of size n_1 cells, G_2 is the comparison group with n_2 cells, and Y_{ig} denotes the log-normalized counts of gene g in cell i. The pseudocount $\epsilon = 10^{-9}$ is added to avoid division by zero (Moses et al., 2023). Genes were selected as intermediate state-associated if they met the following criteria: i) a $\log_2$FC ≥ 0.58 (1.5-fold change) in at least five samples, and ii) at least two of these samples were from experiments not performed on HMLE cells.

2.2 Trajectory inference & EMT subpopulation dynamics

Diffusion pseudotime (DPT) was used for trajectory analysis (Haghverdi et al., 2016). Root nodes were chosen as the epithelial cells with extreme coordinates on a diffusion map. Pseudotime was calculated five times with different epithelial root nodes, and the median values were assigned to each cell, with the standard deviation indicating pseudotime variation. This approach minimized the impact of root node selection on pseudotime calculation. Pseudotime values range from 0 (epithelial) to 1 (mesenchymal). This range was divided into fifteen bins (twelve for Pastushenko et al. (2018) due to fewer cells) and cell counts were calculated for each cluster (epithelial, intermediate, and mesenchymal) for each bin. The counts per bin were converted into cell population proportions.

2.3 RNA velocity analysis

RNA velocity analysis was conducted in Python using the package scVelo in dynamical mode on highly variable genes (Bergen et al., 2020). Each sample was analyzed individually. Differential velocity (DV) was assessed using the `rank_dynamical_genes` function on clusters. Genes with a DV score above 0.25 were retained as DV genes. To ensure monotonic transitions, genes with Spearman correlation coefficients below 0.5 were excluded. Additionally, DV genes with poor dynamical model fits were filtered out. Ultimately, we retained DV genes that were upregulated in the majority of cancer samples, designating them as shared upregulated velocity genes across EMT.

2.4 A mathematical model of EMT dynamics

We developed a mathematical model of the dynamics of EMT described by ordinary differential equations (ODEs). Specifically, we sought to describe the cell state transitions during EMT, from the epithelial (E) to intermediate (I) state or states, and then to the mesenchymal (M) state. While EMT systems may also exhibit direct transitions ($E \rightarrow M$) and reverse transitions,

our data specifically investigate forward EMT and do not exhibit strong evidence for direct transitions.

The population dynamics of E, I, and M are described by:

$$\frac{dE}{dt} = -k_1 EI$$
$$\frac{dI}{dt} = k_1 EI - k_2 IM$$
$$\frac{dM}{dt} = k_2 IM$$

where k_1 denotes the transition rate from E to I, and k_2 denotes the transition rate from I to M. We consider second-order transitions, meaning both the initial and final states influence the transition rate to the final state. In cases where more than one intermediate state exists, the model can be extended using the same framework (Supp. Fig. 7).

2.5 Parameter inference of cell population dynamics over pseudotime

We sought to infer the rates of EMT using Bayesian parameter inference with the Turing.jl package in Julia (Bezanson et al., 2017; Ge et al., 2018; Rackauckas et al., 2017). The input data for each model consists of the cell state dynamics over pseudotime. To focus on relevant dynamics, we excluded periods where all cells remained in the epithelial state. Time points along pseudotime were normalized to a range of $t \in [0, 10]$, facilitating direct comparison of EMT trajectories across samples. For each sample with one intermediate state, we fit three parameters: k_1, k_2, and the observational noise parameter σ. For the *in vivo* sample with two intermediate states, we fit four parameters: k_1, k_2, k_3, and σ.

Letting f represent the numerical solution to the ODE model and y_0 the initial conditions, we performed parameter inference as follows:

$$\theta_{k_i} \sim \mathcal{N}(4, 1)$$
$$\sigma \sim \text{Inv-Gamma}(3, 1)$$
$$\widehat{y}(t) = f(y_0, t; \theta)$$
$$y(t) \sim \mathcal{N}(\widehat{y}(t), \sigma)$$

where $\theta = (\theta_{k_i}, \sigma)$ gives the prior parameter distribution, and $y(t)$ defines the likelihood function in terms of ODE model simulation ($\widehat{y}(t)$) for transition rate parameters θ_{k_i} and noise parameter σ.

The posterior parameter distribution was estimated via Markov chain Monte Carlo (MCMC) simulations using the No-U-Turn Sampler (NUTS) (Hoffman et al., 2014). MCMC chains were each run for 1,000 iterations following 250 warmup iterations to ensure convergence. Fitted trajectories were visualized by solving the model using 300 joint parameter sets of k_n, randomly selected from the posterior distribution for each sample, and plotting the mean and standard deviation of the resulting trajectories.

2.6 Comparative analysis of EMT intermediate state-associated genes

We identified genes associated with EMT transition rates by analyzing correlations between model-inferred posterior parameters and gene expression. For each transition rate parameter k_n, we used its maximum *a posteriori* value for each sample and examined pairwise correlations with the $\log_2$FC expression of 145 genes, each present in at least 5 samples with an intermediate state $\log_2\text{FC} \geq 0.2$. Genes with a Spearman's rank correlation coefficient of $\rho > |0.6|$ ($p < 0.05$) were considered associated with, and potential influencers of, transitions into or out of EMT states.

To specifically identify genes linked to EMT intermediate state dynamics, we focused on genes positively correlated with k_1 (faster $E \rightarrow I$) and negatively correlated with k_2 (slower $I \rightarrow M$). Genes meeting both correlation criteria were included, as well as those showing either correlation as well as differential expression or differential velocity in the intermediate state. Cellular location annotations were performed using DAVID (Huang et al., 2009; Sherman et al., 2022) and PANTHER (Thomas et al., 2022).

3 RESULTS

3.1 Single-cell analysis of EMT across cancer types & stimuli identifies a spectrum of EMT states

To characterize trajectories across a spectrum of EMT, we studied twelve scRNA-seq datasets across five cancer types. Cells were processed and clustered to identify cell states. We found evidence for three cell states in each of the *in vitro* cell populations and four states in the *in vivo* mouse skin squamous cell carcinoma (SCC) sample (Supp. Fig. 1 & 2). Clusters were labeled based on EMT markers from the literature, including Hallmark EMT genes from the Molecular Signatures Database (MSigDB) (Liberzon et al., 2015) and epithelial cell genes from PanglaoDB (Franzén et al., 2019). Distinct clusters representing epithelial and mesenchymal cell types were identified in each dataset, although the relative sizes of these clusters varied widely (Fig. 1A). In all datasets, at least one cluster expressing combinations of epithelial and mesenchymal marker genes was identified as an intermediate state. Certain samples from Cook and Vanderhyden (2020) that did not exhibit a clear EMT were excluded from further analyses (Supp. Fig. 3 & 4). This is in agreement with Cook and Vanderhyden (2020), who also found that certain conditions did not permit a full EMT within the experimental timeframe.

EMT scores were assigned to single cells across all datasets (Fig. 1B). Single cells were each assigned an EMTscore via UCell (Andreatta et al., 2021) using MSigDB Hallmark EMT genes. Each sample exhibited a range of EMTscore, reproducible by replicate and varying considerably by cell type and stimulus. Notably, not only the variance but also the start and end points vary by cell type, highlighting differences not only in EMT but also in the "epithelialness" of different cell types. Samples excluded from analysis due to lack of/incomplete EMT, as identified by marker gene expression, exhibited little to no variation in EMTscore (Supp. Fig. 4), confirming the lack of cell state transition under the tested conditions. The *in vivo* EMT in mouse SCC exhibited the largest range of EMTscore by a wide margin, highlighting the increase in heterogeneity among single cells during a spontaneous, unstimulated, environment-dependent EMT. Since an additional intermediate state was identified in this dataset, in line with previous

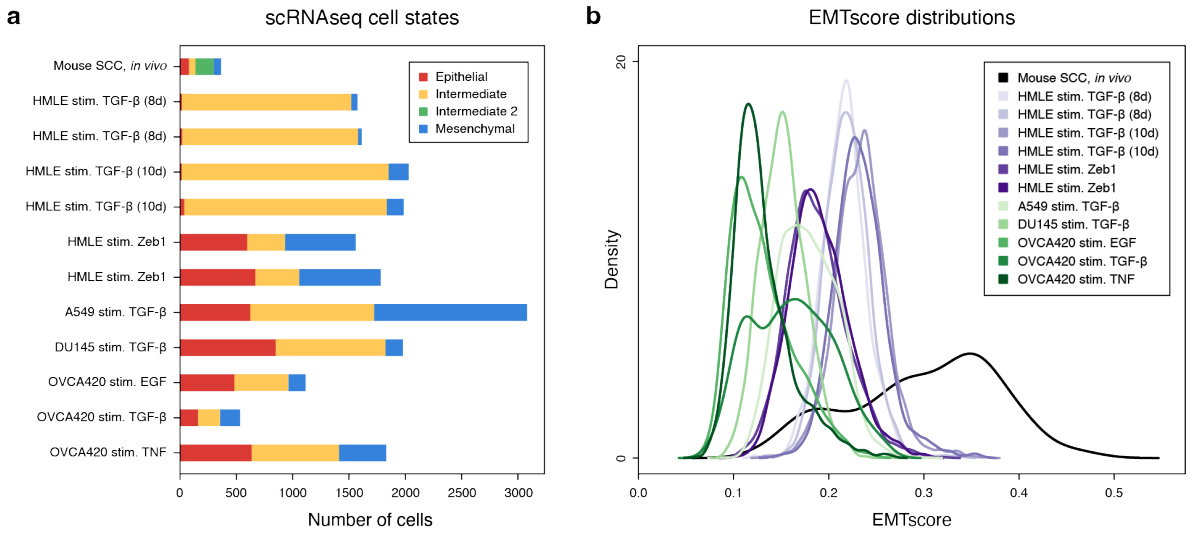

Figure 1: ***a.*** *Cell counts of EMT states per cancer sample, including intermediate states. Cell states were identified via clustering and gene expression.* ***b.*** *Kernel density estimates of the EMTscore distributions for each scRNAseq sample. Single-cell EMT scores were assigned via Hallmark EMT genes from MSigDB.*

work (Pastushenko et al., 2018; Sha, Wang, Bocci, et al., 2021), the data suggest that both the number of attractor states and the size of their basins of attraction are larger for cells in their natural environment than cell line-derived models stimulated *in vitro*.

3.2 Shared marker genes of intermediate EMT states are associated with extracellular function

EMT can proceed along many paths (Hong and Xing, 2024), and both cell/treatment-specific and consensus EMT pathways are important to study in different contexts. Here, we focus on the shared properties of EMT cell state transitions. To study intermediate state gene expression across an EMT spectrum, we performed differential gene expression and differential RNA velocity analysis across intermediate states in different cell populations (Fig. 2A). We identified differentially expressed genes for intermediate states in each sample (2,396 genes total) and examined shared intermediate state-specific genes, defined as those upregulated in an intermediate state relative to epithelial/mesenchymal states. Using a $\log_2$ fold change ($\log_2$FC) threshold of +0.58 (1.5-fold change) in at least five samples, we identified 32 genes shared among EMT intermediate states (Supp. Fig. 5).

Among the 32 genes shared across EMT intermediate states, most were not found in canonical EMT or epithelial gene sets (Fig. 2B). Two predicted intermediate state genes, *ITGB4* and *SFN*, are annotated as epithelial genes in PanglaoDB (Franzén et al., 2019), although the literature on these genes is complicated: Integrin β4 (*ITGB4*) (Fig. 2D) was initially identified in epithelial cells and tumors (Biffo et al., 1997) but has also been linked to promoting EMT in hepatocellular and pancreatic carcinoma (Li et al., 2017; Masugi et al., 2015). *ITGB4* pairs with another intermediate state gene, integrin α6 (*ITGA6*) (Fig. 2E), to form the α6β4 complex, which is implicated in promoting EMT characteristics in hepatocellular carcinoma cells (Zheng

Modeling the dynamics of EMT reveals genes associated with pan-cancer intermediate states and plasticity

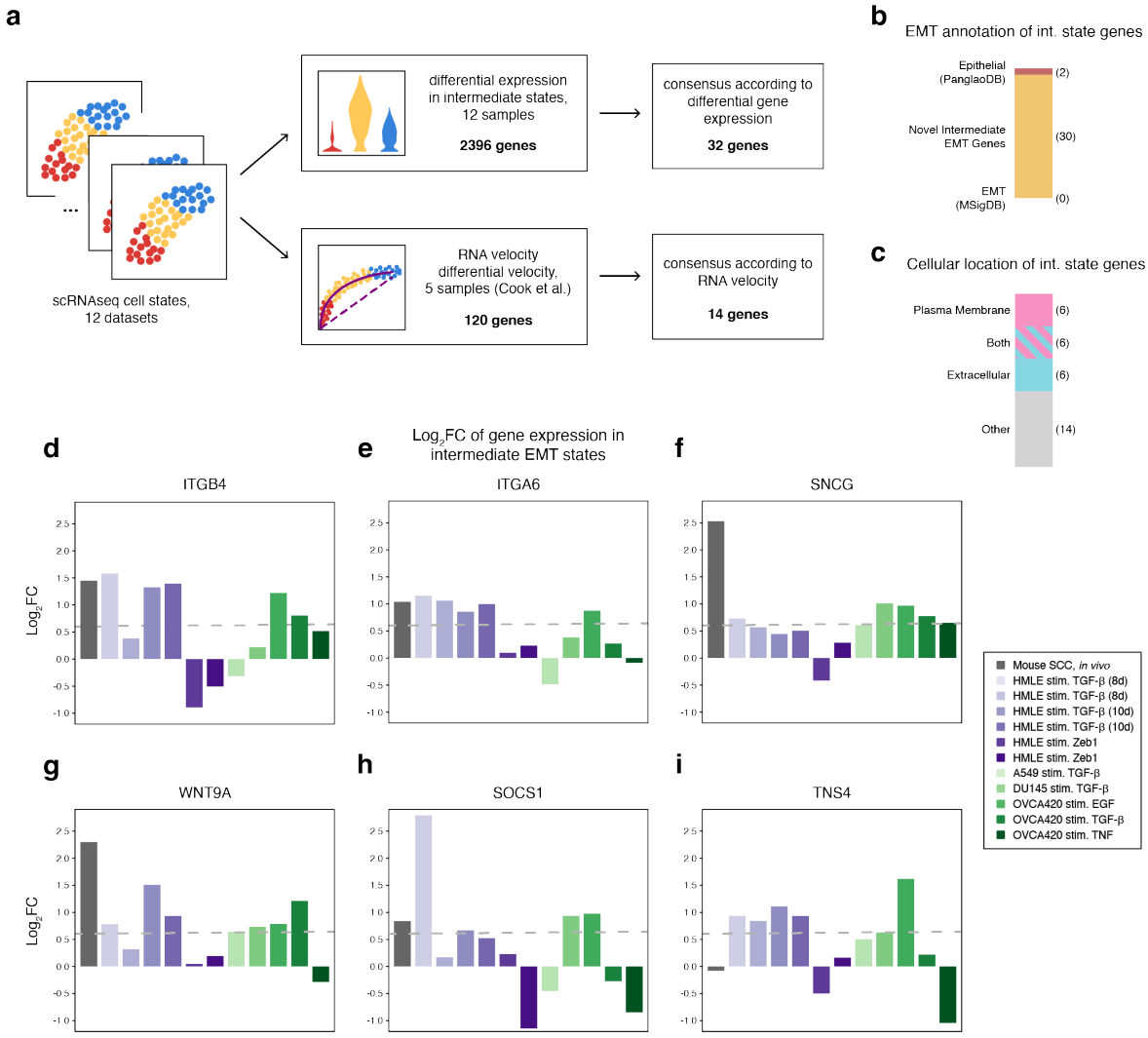

Figure 2: ***a.*** *Data analysis pipeline: pan-cancer intermediate EMT marker genes were identified using differential expression and compared against genes differentially regulated via RNA velocity.* ***b.*** *Annotation of predicted intermediate EMT genes by canonical epithelial/mesenchymal gene sets.* ***c.*** *Annotation of predicted intermediate EMT genes by cellular location.* ***d-i.*** *Comparison of expression of predicted intermediate EMT genes by log_2FC (in the intermediate state) across samples. Dashed line represents 1.5-fold change ($log_2FC > 0.58$).*

et al., 2022). Stratifin (*SFN*) (Supp. Fig. 5) is annotated as epithelial (named for its role in the stratification of epithelial cells [Leffers et al., 1993]) but is also linked to cell migration and EMT markers in cervical and hepatocellular carcinoma (Hu et al., 2019; Ye et al., 2023; Zhao et al., 2023). The apparent contradictory roles of both *ITGB4* and *SFN* as marking for both epithelial and mesenchymal states can be reconciled if these genes are in fact markers of an intermediate EMT state, as predicted by our analysis.

A majority of predicted intermediate EMT marker genes encode proteins localized in the extracellular space, on the plasma membrane or as secreted signaling factors (Fig. 2C). Gamma-synuclein (*SNCG*), upregulated across multiple cell lines (Fig. 2F), is found in the extracellular exosome. It plays a role in suppressing mesenchymal markers including *CDH2* (N-cadherin) and *VIM* (Ni et al., 2021) while promoting cancer cell migration (Liu et al., 2022; Takemura et al., 2021; Zhuang et al., 2015). Other notable upregulated genes include *WNT9A*, *IL4R*, and *IL6R*. Wnt-9a (*WNT9A*) (Fig. 2G) is a secreted protein in the canonical Wnt/β-catenin signaling pathway that is implicated in partial EMT by mediating cell adhesion (Basu et al., 2018). *IL4R* and *IL6R* (Supp. Fig. 5) are interleukin cell surface receptors, with their cytokines IL4, IL13, and IL6 associated with EMT promotion (Cao et al., 2016; Chen et al., 2018; Sun et al., 2020). Interestingly, *SOCS1* (suppressor of cytokine signaling 1) (Fig. 2H) is a negative regulator of IL6 yet conversely has been found to promote EMT (Berzaghi et al., 2017), highlighting the bidirectional signaling at play during the establishment of intermediate EMT states.

Tensin 4 (*TNS4*) (Fig. 2I) is involved in focal adhesion & integrin interaction and promotes EMT and cell motility (Katz et al., 2007; Thorpe et al., 2017). Tubulointerstitial nephritis antigen-like 1 (*TINAGL1*) (Supp. Fig. 5) encodes another secreted protein that binds directly to certain integrins, and it is found to both promote and inhibit metastasis in different cancers *in vivo* (Shan et al., 2021; Shen et al., 2019). Both *TNS4* and *TINAGL1* interact with epidermal growth factor receptor *EGFR*, yet their effects are contradictory: *TNS4* reduces *EGFR* degradation (Hong, Shih, et al., 2013), while *TINAGL1* binds directly to *EGFR* and suppresses *EGFR* signalling (Shen et al., 2019). These opposing interactions may again reflect the dynamic balance necessary to sustain the intermediate EMT state.

Overall, many genes associated with the intermediate EMT state exhibit conflicting roles in the literature, including *ITGB4*, *SFN*, *IL4R* & *IL6R* with *SOCS1*, and *TINAGL1* with *TNS4*. These genes can contribute both to the promotion and inhibition of EMT as well as the balance between epithelial and mesenchymal states. This duality underscores the dynamic nature of EMT and the importance of intermediate states.

3.3 Differential regulation via RNA velocity reveals EM plasticity genes in EMT intermediate states

To investigate dynamically regulated genes during EMT, we performed differential RNA velocity across EMT cell states (Bergen et al., 2020; La Manno et al., 2018). Fourteen genes had differential velocity (DV) in the intermediate state in at least three of the five Cook and Vanderhyden (2020) samples, which includes cells from lung, prostate, and ovarian tumors (Supp. Fig. 6). Of the 14 DV genes, all but one encode proteins located extracellularly or in the plasma membrane (Fig. 3A). Several of these genes are involved in focal adhesion, including integrins *ITGA2* and *ITGB4*, laminins *LAMC2* and *LAMB3*, collagen *COL4A2*, and plasma membrane caveolae component *CAV1*. Eleven of the DV genes have annotated signal

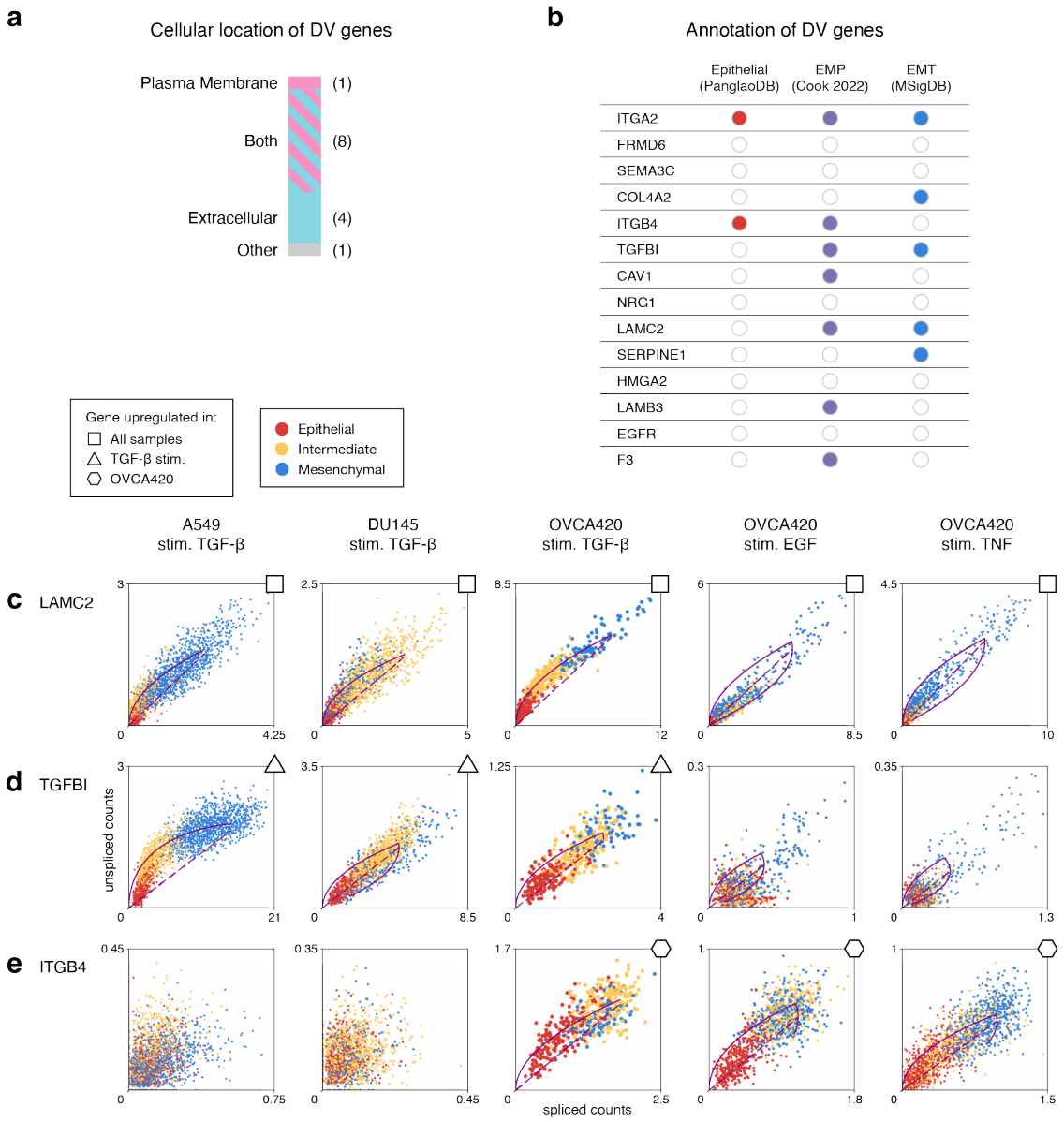

Figure 3: ***a.*** *Annotation of cellular locations for genes differentially regulated in the intermediate state, identified through differential velocity (DV) analysis.* ***b.*** *Annotation of the EMT properties of DV genes by comparison with three EMT marker gene sources. EMP: epithelial-mesenchymal plasticity.* ***c.*** *Examples of DV genes upregulated in intermediate EMT states across different conditions. Solid line represents the dynamical model fit; dashed line represents the inferred steady state. LAMC2 is upregulated across samples from different cell lines & stimuli.* ***d.*** *ITGB4 is upregulated only with a specific stimulus: TGF-β.* ***e.*** *ITGB4 is upregulated only in a specific tumor type: ovarian cells OVCA420.*

peptide sequences, underscoring their designation as secretory/membrane proteins (Sherman et al., 2022; The UniProt Consortium, 2023).

DV genes showed greater overlap with canonical EMT gene sets than the intermediate state marker genes we identified (Fig. 3B). This is expected, as genes actively upregulated in EMT intermediate states are more likely to overlap with mesenchymal markers. Epithelial-mesenchymal plasticity (EMP), i.e. bidirectional cell state transitions between epithelial and mesenchymal phenotypes (Cook and Vanderhyden, 2022), is also characteristic of the DV genes identified. This overlap supports EMP conceptually: capricious cells require dynamic changes in gene expression to change state.

Comparison of DV genes across samples revealed a variety of responses: some genes were shared across different cell types and conditions, while others were specific to certain conditions. Genes upregulated regardless of cell type or stimuli included *LAMC2* (Fig. 3C), *FRMD6*, and *SERPINE1* (Supp. Fig. 6). In contrast, and perhaps unsurprisingly, TGF-β-induced protein *TGFBI* (Fig. 3D) was upregulated in various cell types only when stimulated by TGF-β. A similar pattern was observed for *COL4A2* (Supp. Fig. 6). Genes upregulated by multiple stimuli in one cell type, human ovarian OVCA420 cells, included *ITGB4* (Fig. 3E), *CAV1*, *HMGA2*, *F3*, and *LAMB3* (Supp. Fig. 6). Overall, RNA velocity analysis elucidates gene regulation during EMT. Most differentially regulated genes are specific to a stimuli or cell line, fewer are conserved across conditions. There is substantial overlap between actively regulated genes during EMT and those linked to EMP, highlighting the role of dynamic transitions between cell states during EMT.

3.4 Mathematical modeling & parameter inference quantifies EMT population dynamics

Gene expression is not static: life arises from dynamics. To study the dynamics of EMT in more depth, we developed a mathematical model describing cell state transitions during EMT (Fig. 4A). The model is characterized by rate parameters for transitions between epithelial (E), intermediate (I), and mesenchymal (M) states, such as $E \rightarrow I$ at rate k_1 (Fig. 4B; Supp. Fig. 7A). These rate parameters were fit to scRNA-seq data, characterizing cell state transitions during EMT across pseudotime. Multiple pseudotime trajectories were calculated for each sample, rooted by different epithelial cells, to estimate the mean & variance in pseudotime based on root node selection. Cell state proportions across pseudotime, representing cell population dynamics during EMT, were fitted to the model.

EMT dynamics for each dataset were fit using Bayesian parameter inference (Fig. 4C; Supp. Fig. 7; Supp. Table 2). Differences in EMT dynamics were observed across different datasets, both by cell type and by stimulus. For instance, the intermediate state persisted longer in HMLE cells compared to A549 or OVCA420 cells. Analysis of the parameter posterior distributions for each fitted EMT trajectory revealed similarities and differences in EMT dynamics (Fig. 4D). Dividing the posterior space into three approximate regions: $k_1 \approx k_2$ (similar transition rates across EMT); $k_1 > k_2$ (faster transition rates for $E \rightarrow I$ than $I \rightarrow M$); and $k_1 < k_2$ (faster transition rates for $I \rightarrow M$ than $E \rightarrow I$) highlights how both cell type and stimulus can strongly impact EMT dynamics. For example, OVCA420 cells exhibited $k_1 > k_2$ dynamics regardless of stimulus, where $k_1 > k_2$ implies a larger/more stable intermediate state. In contrast, HMLE cells exhibited $k_1 > k_2$ dynamics for TGF-β stimulation but $k_1 < k_2$ for ZEB1 stimulation, indicating

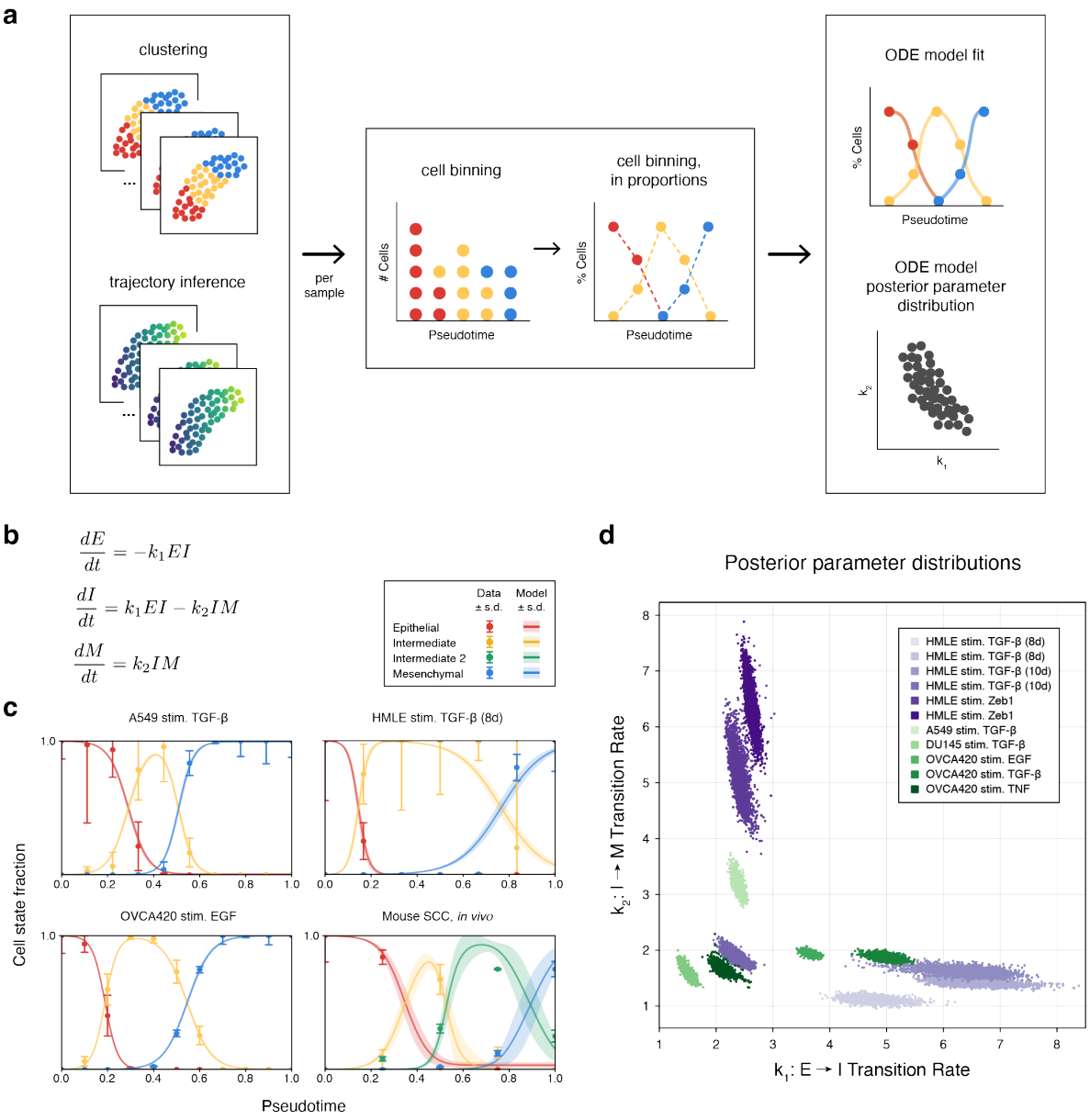

Figure 4: ***a.*** *Workflow to infer dynamic EMT transition rates from scRNA-seq data. For each sample, clustering and trajectory inference information was processed to quantify cell states over pseudotime. A mathematical model was then fit to each sample to infer parameter posterior distributions.* ***b.*** *Mathematical model representing transitions from epithelial (E) to intermediate (I) to mesenchymal (M) state cells. k_1 is the transition rate $E \rightarrow I$; k_2 is the transition rate $I \rightarrow M$. Additional intermediate states can be seamlessly added (Supp. Fig. 7A).* ***c.*** *Model fits following parameter inference: data vs. trajectory simulations, with simulation parameters sampled from the posterior of each model.* ***d.*** *Posterior parameter distributions of the model for each fitted sample.*

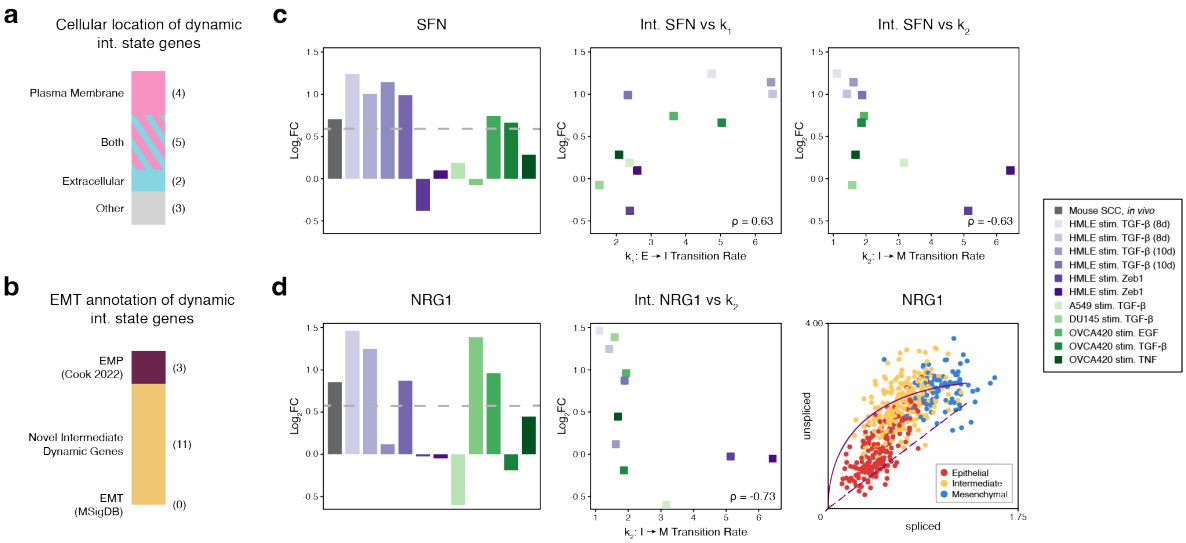

Figure 5: ***a.*** *Annotation of predicted dynamic intermediate state genes by cellular location.* ***b.*** *Annotation of predicted dynamic intermediate state genes by EMT marker gene sources.* ***c.*** *SFN is a predicted dynamic intermediate state gene, with its differential expression in multiple EMT intermediate states and significant correlations with model parameters k_1 and k_2.* ***d.*** *NRG1 is another predicted dynamic intermediate state gene, with its differential expression in multiple EMT intermediate states, significant negative correlation with model parameter k_2, and differential regulation in EMT intermediate states via RNA velocity.*

that the persistence/stability of the HMLE intermediate state depends on the stimulating factor.

An inverse proportion relationship is evident across cell types/stimuli and within a sample; this concordance is notable since more generally different types of parameter covariation can exist (Wu et al., 2023). This analysis highlights how EMT intermediate persistence and stability depend on the intrinsic properties of the EMT experiment, with different carcinomas exhibiting greater or lesser sensitivity to EMT-inducing factors and thus affecting EMT progression.

3.5 Consensus analysis predicts that *SFN* and *NRG1* influence intermediate cell state dynamics during EMT

To identify genes influencing intermediate EMT dynamics, we studied associations between intermediate EMT genes and fitted parameters of the mathematical model. A gene's positive correlation with k_1 indicates faster transition $E \rightarrow I$, while a negative correlation with k_2 means a slower transition $I \rightarrow M$; either correlation suggests that the gene is associated with a more persistent intermediate state. Genes with significant Spearman's correlation were compared with differential expression and differential velocity genes in intermediate states, and those supported by multiple lines of evidence were consolidated into a consensus gene list of 14 genes (Supp. Table 3; Supp. Fig. 8). The majority of intermediate EMT dynamics genes were located at the plasma membrane or in the extracellular region (Fig. 5A). Of the 14 predicted intermediate EMT dynamics genes, three were identified in a prior EMP study (Cook and Vanderhyden, 2022) (Fig. 5B), consistent with the conceptual overlap between intermediate EMT dynamics and EMP. Notably, there is no overlap between intermediate EMT dynamics

genes and those from hallmark EMT (mesenchymal) genes, demonstrating that our proposed gene set is novel and distinct from previous EMT gene sets.

Two predicted EMT dynamics genes had the strongest support (three lines of evidence each; Supp. Table 3). *NRG1* was the only gene identified in all three analyses, while *SFN* was the only gene with intermediate EMT differential expression and significant correlations with both k_1 and k_2 transition rates. Stratifin (*SFN*) was positively correlated with k_1 ($E \rightarrow I$) and negatively correlated with k_2 ($I \rightarrow M$) across cancer samples (Fig. 5C), suggesting that it stabilizes the intermediate EMT state. Although RNA velocity for *SFN* was not captured due to insufficient counts, it was differentially expressed in intermediate states. Neuregulin 1 (*NRG1*) was negatively correlated with k_2 ($I \rightarrow M$), suggesting it slows the exit from the intermediate state (Fig. 5D), and *NRG1* was also significant in intermediate EMT differential expression and velocity (Supp. Fig. 6).

Consensus gene analysis predicts that *SFN* promotes transitions from an epithelial state to the metastatic intermediate EMT state. This prediction helps to reconcile literature, which reports both epithelial and pro-EMT roles for *SFN*. Named for its expression in stratified epithelial cells (Leffers et al., 1993), *SFN* can be secreted and is found in extracellular vesicles (Hou et al., 2022). Recombinant *SFN* treatment has been shown to significantly enhance extracellular matrix degradation in human dermal fibroblasts *in vitro* (Ghaffari et al., 2006). Despite its epithelial association, *SFN* knockdown in *in vitro* models has led to reduced mesenchymal marker expression in cervical cancer cells (Hu et al., 2019) and decreased cell migration in other carcinomas (Kim, Kim, et al., 2022; Ye et al., 2023; Zhao et al., 2023). *In vivo*, *SFN* knockdown suppressed tumor formation and metastasis in lung adenocarcinoma models (Shiba-Ishii et al., 2015). Clinically, *SFN* is linked to poor prognosis, including advanced tumor stages in lung adenocarcinoma and hepatocellular carcinoma (Kim, Shiba-Ishii, et al., 2018; Ye et al., 2023), as well as lower survival rates in pancreatic ductal adenocarcinoma (Robin et al., 2020) and head and neck squamous cell carcinoma (Chung et al., 2006). Our findings suggest that *SFN* promotes intermediate EMT dynamics, potentially explaining its dual role in epithelial cells while facilitating EMT.

Consensus gene analysis also identified *NRG1* as playing a pivotal role in intermediate EMT state dynamics, as the sole gene that was significant in intermediate expression, regulation, and modeled dynamics. A member of the epidermal growth factor (EGF) family (The UniProt Consortium, 2023), *NRG1* activates *ERBB2* (*HER2*) and *ERBB3* (*HER3*) (Miano et al., 2022). *NRG1* isoforms can be found in the plasma membrane or secreted (Esper et al., 2006), and it binds integrins including *ITGA6*:*ITGB4* and *ITGAV*:*ITGB3* (Ieguchi et al., 2010). *In vivo*, *NRG1* suppression reduces tumor growth and metastasis in hepatocellular carcinoma (Shi et al., 2018). Clinically, *NRG1* overexpression correlated with poor outcomes, including lymph node metastasis, in gastric cancer (Yun et al., 2018). Notably, *NRG1* has been found to promote partial EMT in cultured patient *HER2*-positive breast cancer (Guardia et al., 2021). While *NRG1* has been mostly described to drive EMT in epithelial cells, *NRG1* stimulation on mesenchymal cells that already underwent EMT has been shown to instead induce epithelial gene expression in esophageal adenocarcinoma (Ebbing et al., 2017). Taken together, our analyses along with literature suggest that *NRG1* is a marker of highly plastic intermediate state cells during EMT.

3.6 *SFN* is a marker of intermediate state EMT in independently analyzed MCF10A cells

To assess predicted intermediate EMT genes, we analyzed a dataset of EMT under different experimental conditions and in a different cell line: the dose-dependent TGF-β stimulation of MCF10A breast cells (Panchy et al., 2022). Similar to previous analyses, scRNA-seq data was clustered, and canonical markers were used to identify epithelial, intermediate, and mesenchymal states (Fig. 6A). Differential expression by cell state showed strong agreement with our predictions, with 11 of the top 25 intermediate state genes in this sample overlapping with our predicted intermediate EMT genes (Fig. 6B), notably including *SFN*. These results highlight that shared EMT intermediate state features can be found across diverse biological and experimental conditions, with independent evidence corroborating one of the top genes associated with intermediate EMT.

To further evaluate EMT dynamics in these MCF10A cells, we fit the mathematical model to this dataset (Fig. 6C). The posterior parameter distribution lies in the region where $k_1 \geq k_2$, consistent with EMT dynamics induced by TGF-β in other cell types (Fig. 6D). Across different cancer types, we see that mammary (MCF10A and HMLE) and ovarian (OVCA420) cells stimulated with TGF-β generally exhibit $k_1 > k_2$ dynamics, favoring stabilization of the intermediate state. In contrast, lung (A549) and prostate (DU145) cells stimulated with TGF-β show balanced rates of entry and exit from the intermediate state, with $k_1 \approx k_2$. The similarity in transition dynamics between mammary and ovarian cells is notable, given the shared genetic and microenvironmental factors during oncogenesis and tumor progression (Roskelley et al., 2002).

4 DISCUSSION

Here, we characterized intermediate EMT states and identified genes involved in dynamic transitions between states. Multiple lines of evidence suggest EMT intermediate states are the most cancer stem-like and exhibit the highest metastatic potential (Aiello et al., 2018; Brown et al., 2022; Pastushenko et al., 2018; Puram et al., 2017; Zhang, Donaher, et al., 2022). Our analysis predicted intermediate state genes in agreement with recent work, such as *ITGB4* and *LAMB3* (Cheng et al., 2024), as well as novel EMT intermediate genes, such as *SFN* and *NRG1*. While there are many paths of EMT, our comparison across different cell types and stimuli revealed common markers for intermediate states and highlighted the role many of these genes have in extracellular remodeling.

EMT is heterogeneous (Yang et al., 2020). Multiple transcription factors can initiate EMT (Nieto et al., 2016) and act in complex and nonlinear ways, both alone (Hartmann et al., 2024) or in combination (Sanford et al., 2020). Additional factors contributing to EMT complexity, including subtypes and intermediate states, are hysteresis during the reverse mesenchymal-epithelial transition, differences in cell types or stimuli, and state transitions driven by intrinsic or extrinsic noise. Whereas EMT is most frequently modeled via gene regulatory networks, here we modeled the population dynamics to study cell state heterogeneity and its effects on EMT path variation. In doing so, we assumed a monostable landscape, whereas in reality multiple stable steady states exist (Hong, Watanabe, et al., 2015). Some of the gene expression heterogeneity underlying these multiple states is likely collapsed by this approach, but in

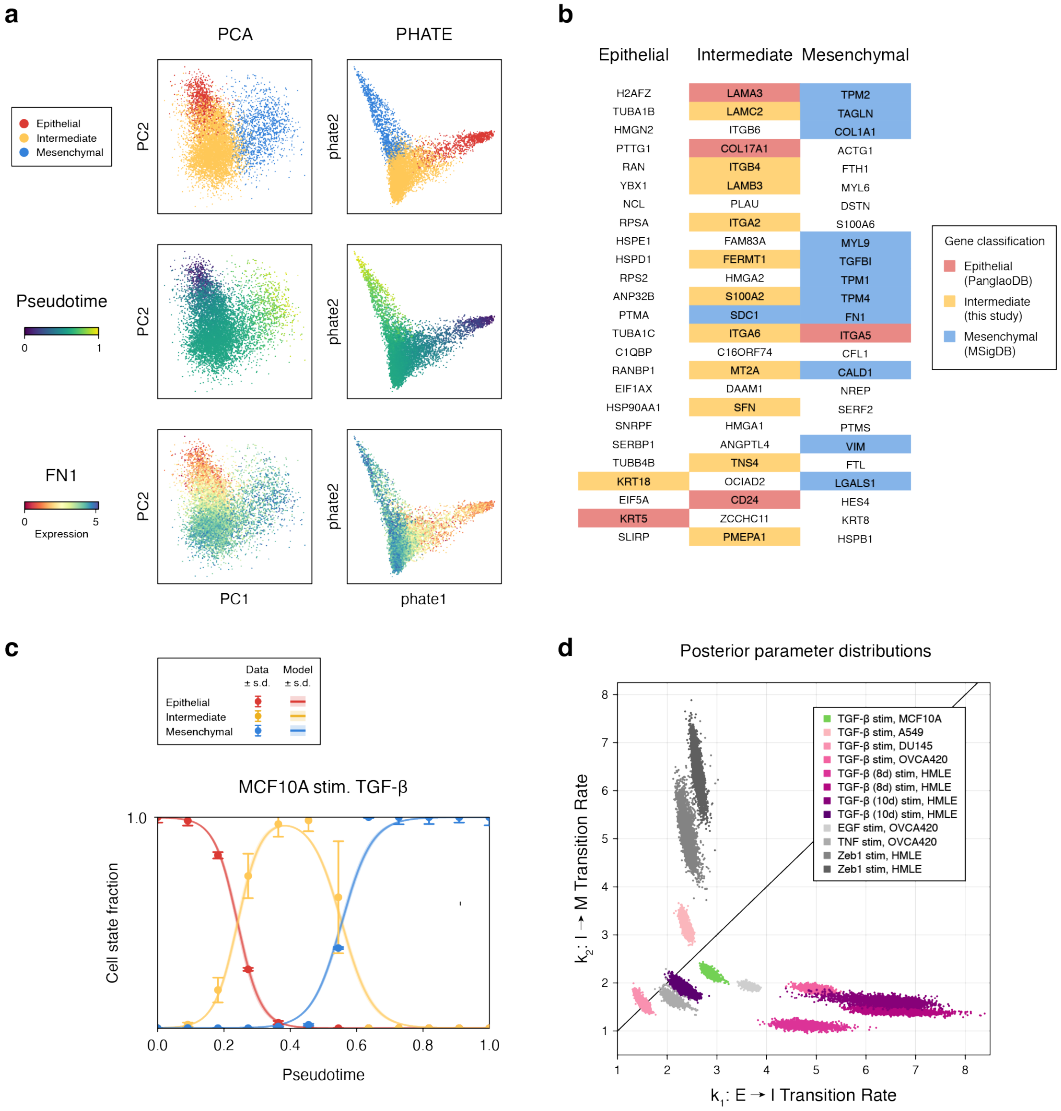

Figure 6: ***a.*** *MCF10A cells were analyzed separately and exhibit a linear trajectory across EMT states.* ***b.*** *For each MCF10A cell state, the top 25 differentially expressed genes colored by gene set annotation.* ***c.*** *Model fit following parameter inference: data vs. trajectory simulations, with simulation parameters sampled from the posterior distribution.* ***d.*** *Comparison of posterior parameter distributions, with the MCF10A sample highlighted in green. Other distributions are replicated from Fig. 4D, shown in different colors here to highlight stimulation by TGF-β.*

doing so we can identify consensus genes marking for properties of EMT states across different conditions. Our model can be adapted in the future to consider multiple intermediate states and more complex (e.g. convergent/divergent) EMT paths.

Summarizing complex data across conditions to find consensus requires simplifying assumptions. To compare gene expression across datasets we used log-fold changes and rank-based comparisons, similar to other recent work (Theodoris et al., 2023). Doing so relies on the accurate quantification of cell states, which is not guaranteed, and can obscure single-cell resolution information by taking pseudo-bulk measurements. While we sought to standardize data analysis pipelines as far as possible, scRNA-seq data analysis relies on certain parameter choices. While clustering cells we sought fewer clusters (lower resolution) where supported, to reduce overfitting cell states. Trajectory inference relies on accurate choice of root cells and the sufficiency of the similarity metric used. RNA velocity analysis is limited by the ratio of spliced to unspliced counts, typically around 75-85% spliced to 15–25% unspliced (La Manno et al., 2018). This abundance limitation affects genes with low or no unspliced counts, such as *SFN* in our study, where RNA velocity analysis could not be performed due to a lack of unspliced counts. This abundance limitation could be addressed by experimental methods targeting dynamics, such as RNA metabolic labeling (Qiu, Zhang, et al., 2022).

Mathematical modeling and parameter inference with single-cell data allow us to investigate the genes and pathways associated with dynamic transitions between states rather than the cell states themselves – transitions which are strongly relevant to epithelial-mesenchymal plasticity (Williams et al., 2019). EMP, exemplary of cell state plasticity, has been shown to play decisive roles in tumorigenesis and cancer progression (Househam et al., 2022; Street et al., 2023). This property can assist tumors in developing powerful 'generalist' phenotypes as they evolve (Cook and Wrana, 2022). The mathematical model with which we study EMT population dynamics is phenomenological: capturing the rates of entry/exit between EMT states without transcriptional information or feedback signaling. It does not incorporate additional complexities such as reverse transitions or stochasticity. We have used external information from the biological properties of EMT to construct our mathematical model, and not obtained it purely from the cell dynamics observed in the data (Weinreb et al., 2018). Nonetheless, to compare relative transition rates, a simple three-compartment model seems reasonable to describe most conditions analyzed and fits both the inferred cell states (clusters) and the pseudotemporal dynamics during EMT. In the future, combining the cell population dynamics with a transcriptional EMT network (Hong, Watanabe, et al., 2015) to investigate the role of cell-cell communication (Rommelfanger et al., 2021) on the population dynamics of EMT could lead to additional insight – though additional data may be required for the transcriptional dynamics of such a model to avoid double dipping (Neufeld et al., 2024).

Canonical EMT states are defined by morphological features: epithelial cells adhere to each other with apical-basal polarity; mesenchymal cells are spindle-shaped, migratory, and lack cell-cell adhesion (Thiery et al., 2009). These morphological/adhesive properties cannot be fully captured by sequencing data alone. Moreover, multiple EMT gene lists (typically focusing on mesenchymal traits) have been proposed, with varying levels of agreement (Cook and Vanderhyden, 2022; Kinker et al., 2020; Puram et al., 2017; Tan et al., 2014; Yang et al., 2020). This variability in consensus genes also applies to epithelial genes, which can show tissue-specific heterogeneity. No single gene list can do justice to the heterogeneous paths of EMT, yet as we have shown, distinctive dynamic properties of EMT intermediate states can be

captured by marker genes.

Genes predicted here as candidate markers of intermediate state EMT genes may serve as biomarkers of cells likely to metastasize and could be tested as predictors of clinical progression. In addition, such genes may mark for high-risk tumor cells prone to metastasis or recurrence, given the high metastatic potential of EMT intermediate state cells (Brown et al., 2022; Lüönd et al., 2021; Simeonov et al., 2021; Yu et al., 2013). More broadly, this study has shed new light on the plasticity of the EMT landscape and how it shapes the cell state transitions underlying cancer metastasis.

AUTHOR STATEMENTS

Author Contributions

M.M.: Conceptualization, software, methodology, investigation, formal analysis, writing—original draft, writing—reviewing & editing. **R.M.**: Investigation, writing—reviewing & editing. **E.T.R.T.**: Investigation, supervision, writing—reviewing & editing. **A.L.M.**: Conceptualization, software, methodology, investigation, funding acquisition, supervision, writing—original draft, writing—reviewing & editing.

Funding Statement

This work was supported by the National Institutes of Health (R35GM143019) and the National Science Foundation (DMS2045327) (to A.L.M.).

Conflict of Interest Statement

The authors declare no competing interests.

Data Availability Statement

All code and data analysis associated with this study are released under an MIT license, available on GitHub: `https://github.com/maclean-lab/dynamicEMT-genes`. All raw sequencing data used in this study are publicly available on the Gene Expression Omnibus (GEO); see Methods for accession numbers.

References

Aiello, Nicole M., Ravikanth Maddipati, Robert J. Norgard, David Balli, Jinyang Li, Salina Yuan, Taiji Yamazoe, Taylor Black, Amine Sahmoud, Emma E. Furth, Dafna Bar-Sagi, and Ben Z. Stanger (2018). "EMT subtype influences epithelial plasticity and mode of cell migration". *Developmental Cell* 45.6. ISSN: 1534-5807. DOI: 10.1016/j.devcel.2018.05.027.

Andreatta, Massimo and Santiago J. Carmona (2021). "UCell: Robust and scalable single-cell gene signature scoring". *Computational and Structural Biotechnology Journal* 19. ISSN: 2001-0370. DOI: 10.1016/j.csbj.2021.06.043.

Basu, Sayon, Sanith Cheriyamundath, and Avri Ben-Ze'ev (2018). "Cell–cell adhesion: linking Wnt/β-catenin signaling with partial EMT and stemness traits in tumorigenesis". *F1000Research* 7. ISSN: 2046-1402. DOI: 10.12688/f1000research.15782.1.

Bergen, Volker, Marius Lange, Stefan Peidli, F. Alexander Wolf, and Fabian J. Theis (2020). "Generalizing RNA velocity to transient cell states through dynamical modeling". *Nature Biotechnology* 38.12. ISSN: 1546-1696. DOI: 10.1038/s41587-020-0591-3.

Berzaghi, R., V. S. C. Maia, F. V. Pereira, F. M. Melo, M. S. Guedes, C. S. T. Origassa, J. B. Scutti, A. L. Matsuo, N. O. S. Câmara, E. G. Rodrigues, and L. R. Travassos (2017). "SOCS1 favors the epithelial-mesenchymal transition in melanoma, promotes tumor progression and prevents antitumor immunity by PD-L1 expression". *Scientific Reports* 7.1. ISSN: 2045-2322. DOI: 10.1038/srep40585.

Bezanson, Jeff, Alan Edelman, Stefan Karpinski, and Viral B. Shah (2017). "Julia: A Fresh Approach to Numerical Computing". *SIAM Review*. DOI: https://doi.org/10.1137/141000671.

Biffo, Stefano, Francesca Sanvito, Silvana Costa, Laura Preve, Raffaella Pignatelli, Laura Spinardi, and Pier Carlo Marchisio (1997). "Isolation of a Novel β4 Integrin-binding Protein (p27BBP) Highly Expressed in Epithelial Cells". *Journal of Biological Chemistry* 272.48. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.272.48.30314.

Bracken, Cameron P., Philip A. Gregory, Natasha Kolesnikoff, Andrew G. Bert, Jun Wang, M. Frances Shannon, and Gregory J. Goodall (2008). "A Double-Negative Feedback Loop between ZEB1-SIP1 and the microRNA-200 Family Regulates Epithelial-Mesenchymal Transition". *Cancer Research* 68.19. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-08-1942.

Brown, Meredith S., Behnaz Abdollahi, Owen M. Wilkins, Hanxu Lu, Priyanka Chakraborty, Nevena B. Ognjenovic, Kristen E. Muller, Mohit Kumar Jolly, Brock C. Christensen, Saeed Hassanpour, and Diwakar R. Pattabiraman (2022). "Phenotypic heterogeneity driven by plasticity of the intermediate EMT state governs disease progression and metastasis in breast cancer". *Science Advances* 8.31. DOI: 10.1126/sciadv.abj8002.

Cao, Hui, Jing Zhang, Hong Liu, Ledong Wan, Honghe Zhang, Qiong Huang, Enping Xu, and Maode Lai (2016). "IL-13/STAT6 signaling plays a critical role in the epithelial-mesenchymal transition of colorectal cancer cells". *Oncotarget* 7.38. ISSN: 1949-2553. DOI: 10.18632/oncotarget.11282.

Chaffer, Christine L., Beatriz P. San Juan, Elgene Lim, and Robert A. Weinberg (2016). "EMT, cell plasticity and metastasis". *Cancer and Metastasis Reviews* 35.4. ISSN: 1573-7233. DOI: 10.1007/s10555-016-9648-7.

Chen, Jiaoe, Chaoju Gong, Huiqin Mao, Zhaoyun Li, Zejun Fang, Qiang Chen, Min Lin, Xiang Jiang, Yanyan Hu, Wei Wang, Xiaomin Zhang, Xianjun Chen, and Hongzhang Li (2018). "E2F1/SP3/STAT6 axis is required for IL-4-induced epithelial-mesenchymal transition of colorectal cancer cells". *International Journal of Oncology* 53.2. ISSN: 1791-2423. DOI: 10.3892/ijo.2018.4429.

Cheng, Yu-Chen, Yun Zhang, Shubham Tripathi, B. V. Harshavardhan, Mohit Kumar Jolly, Geoffrey Schiebinger, Herbert Levine, Thomas O. McDonald, and Franziska Michor (2024). "Reconstruction of single-cell lineage trajectories and identification of diversity in fates during the epithelial-to-mesenchymal transition". *Proceedings of the National Academy of Sciences* 121.32. DOI: 10.1073/pnas.2406842121.

Cho, Heyrim, Ya-Huei Kuo, and Russell C. Rockne (2022). "Comparison of cell state models derived from single-cell RNA sequencing data: graph versus multi-dimensional space". *Mathematical Biosciences and Engineering* 19.8. ISSN: 1551-0018. DOI: 10.3934/mbe.2022395.

Chung, Christine H., Joel S. Parker, Kim Ely, Jesse Carter, Yajun Yi, Barbara A. Murphy, K. Kian Ang, Adel K. El-Naggar, Adam M. Zanation, Anthony J. Cmelak, Shawn Levy, Robbert J. Slebos, and Wendell G. Yarbrough (2006). "Gene Expression Profiles Identify Epithelial-to-Mesenchymal Transition and Activation of Nuclear Factor-κB Signaling as Characteristics of a High-risk Head and Neck Squamous Cell Carcinoma". *Cancer Research* 66.16. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-06-1213.

Cook, David P. and Barbara C. Vanderhyden (2020). "Context specificity of the EMT transcriptional response". *Nature Communications* 11.1. ISSN: 2041-1723. DOI: 10.1038/s41467-020-16066-2.

Cook, David P. and Barbara C. Vanderhyden (2022). "Transcriptional census of epithelial-mesenchymal plasticity in cancer". *Science Advances* 8.1. DOI: 10.1126/sciadv.abi7640.

Cook, David P. and Jeffrey L. Wrana (2022). "A specialist-generalist framework for epithelial-mesenchymal plasticity in cancer". *Trends in Cancer* 8.5. ISSN: 2405-8033. DOI: 10.1016/j.trecan.2022.01.014.

Deshmukh, Abhijeet P., Suhas V. Vasaikar, Katarzyna Tomczak, Shubham Tripathi, Petra den Hollander, Emre Arslan, Priyanka Chakraborty, Rama Soundararajan, Mohit Kumar Jolly, Kunal Rai, Herbert Levine, and Sendurai A. Mani (2021). "Identification of EMT signaling cross-talk and gene regulatory networks by single-cell RNA sequencing". *Proceedings of the National Academy of Sciences* 118.19. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2102050118.

Dillekås, Hanna, Michael S. Rogers, and Oddbjørn Straume (2019). "Are 90% of deaths from cancer caused by metastases?" *Cancer Medicine* 8.12. ISSN: 2045-7634. DOI: 10.1002/cam4.2474.

Ebbing, Eva A., Anne Steins, Evelyn Fessler, Phylicia Stathi, Willem Joost Lesterhuis, Kausilia K. Krishnadath, Louis Vermeulen, Jan Paul Medema, Maarten F. Bijlsma, and Hanneke W. M. van Laarhoven (2017). "Esophageal Adenocarcinoma Cells and Xenograft Tumors Exposed to Erb-b2 Receptor Tyrosine Kinase 2 and 3 Inhibitors Activate Transforming Growth Factor Beta Signaling, Which Induces Epithelial to Mesenchymal Transition". *Gastroenterology* 153.1. ISSN: 0016-5085, 1528-0012. DOI: 10.1053/j.gastro.2017.03.004.

Esper, Raymond M., Mark S. Pankonin, and Jeffrey A. Loeb (2006). "Neuregulins: versatile growth and differentiation factors in nervous system development and human disease". *Brain Research Reviews* 51.2. ISSN: 0165-0173. DOI: 10.1016/j.brainresrev.2005.11.006.

Fischer, David S., Anna K. Fiedler, Eric M. Kernfeld, Ryan M. J. Genga, Aimée Bastidas-Ponce, Mostafa Bakhti, Heiko Lickert, Jan Hasenauer, Rene Maehr, and Fabian J. Theis (2019). "Inferring population dynamics from single-cell RNA-sequencing time series data". *Nature Biotechnology* 37.4. ISSN: 1546-1696. DOI: 10.1038/s41587-019-0088-0.

Franzén, Oscar, Li-Ming Gan, and Johan L. M. Björkegren (2019). "PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data". *Database: The Journal of Biological Databases and Curation* 2019. ISSN: 1758-0463. DOI: 10.1093/database/baz046.

Ge, Hong, Kai Xu, and Zoubin Ghahramani (2018). "Turing: A Language for Flexible Probabilistic Inference". *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, pp. 1682–1690. URL: https://proceedings.mlr.press/v84/ge18b.html.

Ghaffari, Abdi, Yunyaun Li, Ali Karami, Mazyar Ghaffari, Edward E. Tredget, and Aziz Ghahary (2006). "Fibroblast extracellular matrix gene expression in response to keratinocyte-releasable stratifin". *Journal of Cellular Biochemistry* 98.2. ISSN: 1097-4644. DOI: 10.1002/jcb.20782.

Guardia, Cristina et al. (2021). "Preclinical and Clinical Characterization of Fibroblast-derived Neuregulin-1 on Trastuzumab and Pertuzumab Activity in HER2-positive Breast Cancer". *Clinical Cancer Research* 27.18. ISSN: 1078-0432. DOI: 10.1158/1078-0432.CCR-20-2915.

Haghverdi, Laleh, Maren Büttner, F. Alexander Wolf, Florian Buettner, and Fabian J. Theis (2016). "Diffusion pseudotime robustly reconstructs lineage branching". *Nature Methods* 13.10. ISSN: 1548-7105. DOI: 10.1038/nmeth.3971.

Hartmann, Laura, Panajot Kristofori, Congxin Li, Kolja Becker, Lorenz Hexemer, Stefan Bohn, Sonja Lenhardt, Sylvia Weiss, Björn Voss, Alexander Loewer, and Stefan Legewie (2024). "Transcriptional regulators ensuring specific gene expression and decision making at high TGFβ doses". *bioRxiv*. DOI: 10.1101/2024.04.23.590740.

Hendrix, M. J., E. A. Seftor, R. E. Seftor, and K. T. Trevor (1997). "Experimental co-expression of vimentin and keratin intermediate filaments in human breast cancer cells results in phenotypic interconversion and increased invasive behavior". *The American Journal of Pathology* 150.2, pp. 483–495. ISSN: 0002-9440.

Hoffman, Matthew D and Andrew Gelman (2014). "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo". *Journal of Machine Learning Research*, p. 31.

Hong, Shiao-Ya, Yi-Ping Shih, Tianhong Li, Kermit L. Carraway, and Su Hao Lo (2013). "CTEN prolongs signaling by EGFR through reducing its ligand-induced degradation". *Cancer Research* 73.16. ISSN: 1538-7445. DOI: 10.1158/0008-5472.CAN-12-4441.

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. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004569.

Hong, Tian and Jianhua Xing (2024). "Data- and theory-driven approaches for understanding paths of epithelial–mesenchymal transition". *Genesis* 62.2. ISSN: 1526-968X. DOI: 10.1002/dvg.23591.

Hou, Wenyun, Meng Pan, Yi Xiao, and Wei Ge (2022). "Serum Extracellular Vesicle Stratifin Is a Biomarker of Perineural Invasion in Patients With Colorectal Cancer and Predicts Worse Prognosis". *Frontiers in Oncology* 12. ISSN: 2234-943X. DOI: 10.3389/fonc.2022.912584.

Househam, Jacob et al. (2022). "Phenotypic plasticity and genetic control in colorectal cancer evolution". *Nature* 611.7937. ISSN: 0028-0836. DOI: 10.1038/s41586-022-05311-x.

Hu, Yi, Yan Ma, Jie Liu, Yanlin Cai, Mengmeng Zhang, and Xiaoling Fang (2019). "LINC01128 expedites cervical cancer progression by regulating miR-383-5p/SFN axis". *BMC Cancer* 19.1. ISSN: 1471-2407. DOI: 10.1186/s12885-019-6326-5.

Huang, Da Wei, Brad T. Sherman, and Richard A. Lempicki (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources". *Nature Protocols* 4.1. ISSN: 1750-2799. DOI: 10.1038/nprot.2008.211.

Ieguchi, Katsuaki, Masaaki Fujita, Zi Ma, Parastoo Davari, Yukimasa Taniguchi, Kiyotoshi Sekiguchi, Bobby Wang, Yoko K. Takada, and Yoshikazu Takada (2010). "Direct binding of the EGF-like domain of neuregulin-1 to integrins (αvβ3 and α6β4) is involved in neuregulin-1/ErbB signaling". *The Journal of Biological Chemistry* 285.41. ISSN: 1083-351X. DOI: 10.1074/jbc.M110.113878.

Johnson, W. Evan, Cheng Li, and Ariel Rabinovic (2007). "Adjusting batch effects in microarray expression data using empirical Bayes methods". *Biostatistics* 8.1. ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxj037.

Jolly, Mohit Kumar, Marcelo Boareto, Bin Huang, Dongya Jia, Mingyang Lu, Eshel Ben-Jacob, José N. Onuchic, and Herbert Levine (2015). "Implications of the Hybrid Epithelial/Mesenchymal Phenotype in Metastasis". *Frontiers in Oncology* 5. ISSN: 2234-943X. DOI: 10.3389/fonc.2015.00155.

Jost, Michael, Gabriele Pannocchia, and Martin Mönnigmann (2015). "Online constraint removal: Accelerating MPC with a Lyapunov function". *Automatica* 57. ISSN: 0005-1098. DOI: 10.1016/j.automatica.2015.04.014.

Katz, Menachem et al. (2007). "A reciprocal tensin-3–cten switch mediates EGF-driven mammary cell migration". *Nature Cell Biology* 9.8. ISSN: 1476-4679. DOI: 10.1038/ncb1622.

Kim, Ji Young, Mi-Jeong Kim, Ji Su Lee, Juhee Son, Duk-Hwan Kim, Joo Sang Lee, Soo-Kyung Jeong, Eunyoung Chun, and Ki-Young Lee (2022). "Stratifin (SFN) regulates lung cancer progression via nucleating the Vps34-BECN1-TRAF6 complex for autophagy induction". *Clinical and Translational Medicine* 12.6. ISSN: 2001-1326. DOI: 10.1002/ctm2.896.

Kim, Yunjung, Aya Shiba-Ishii, Tomoki Nakagawa, Shun-ichiro Iemura, Tohru Natsume, Noriyuki Nakano, Ryota Matsuoka, Shingo Sakashita, SangJoon Lee, Atsushi Kawaguchi, Yukio Sato, and Masayuki Noguchi (2018). "Stratifin regulates stabilization of receptor tyrosine kinases via interaction with ubiquitin-specific protease 8 in lung adenocarcinoma". *Oncogene* 37.40. ISSN: 1476-5594. DOI: 10.1038/s41388-018-0342-9.

Kinker, Gabriela S. et al. (2020). "Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity". *Nature Genetics* 52.11. ISSN: 1061-4036, 1546-1718. DOI: 10.1038/s41588-020-00726-6.

La Manno, Gioele et al. (2018). "RNA velocity of single cells". *Nature* 560.7719. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-018-0414-6.

Leffers, H, P Madsen, H H Rasmussen, B Honoré, A H Andersen, E Walbum, J Vandekerckhove, and J E Celis (1993). "Molecular cloning and expression of the transformation sensitive epithelial marker stratifin. A member of a protein family that has been involved in the protein kinase C signalling pathway". *Journal of Molecular Biology* 231.4. ISSN: 1089-8638. DOI: 10.1006/jmbi.1993.1346.

Leggett, Susan E., Alex M. Hruska, Ming Guo, and Ian Y. Wong (2021). "The epithelial-mesenchymal transition and the cytoskeleton in bioengineered systems". *Cell Communication and Signaling* 19.1. ISSN: 1478-811X. DOI: 10.1186/s12964-021-00713-2.

Leinonen, Rasko, Hideaki Sugawara, and Martin Shumway (2011). "The Sequence Read Archive". *Nucleic Acids Research* 39. ISSN: 0305-1048. DOI: 10.1093/nar/gkq1019.

Li, Xiao-Long, Lin Liu, Dan-Dan Li, Ya-Ping He, Le-Hang Guo, Li-Ping Sun, Lin-Na Liu, Hui-Xiong Xu, and Xiao-Ping Zhang (2017). "Integrin β4 promotes cell invasion and epithelial-mesenchymal transition through the modulation of Slug expression in hepatocellular carcinoma". *Scientific Reports* 7.1. ISSN: 2045-2322. DOI: 10.1038/srep40464.

Liberzon, Arthur, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P. Mesirov, and Pablo Tamayo (2015). "The Molecular Signatures Database (MSigDB) hallmark gene set collection". *Cell systems* 1.6. ISSN: 2405-4712. DOI: 10.1016/j.cels.2015.12.004.

Liu, Jieya, Ting Shao, Jin Zhang, Qianyi Liu, Hui Hua, Hongying Zhang, Jiao Wang, Ting Luo, Yuenian Eric Shi, and Yangfu Jiang (2022). "Gamma synuclein promotes cancer metastasis through the MKK3/6-p38MAPK cascade". *International Journal of Biological Sciences* 18.8. ISSN: 1449-2288. DOI: 10.7150/ijbs.69155.

Lu, Mingyang, Mohit Kumar Jolly, Herbert Levine, José N. Onuchic, and Eshel Ben-Jacob (2013). "MicroRNA-based regulation of epithelial–hybrid–mesenchymal fate determination". *Proceedings of the National Academy of Sciences* 110.45. DOI: 10.1073/pnas.1318192110.

Lüönd, Fabiana, Nami Sugiyama, Ruben Bill, Laura Bornes, Carolina Hager, Fengyuan Tang, Natascha Santacroce, Christian Beisel, Robert Ivanek, Thomas Bürglin, Stefanie Tiede, Jacco van Rheenen, and Gerhard Christofori (2021). "Distinct contributions of partial and full EMT to breast cancer malignancy". *Developmental Cell* 56.23. ISSN: 1534-5807. DOI: 10.1016/j.devcel.2021.11.006.

MacLean, Adam L., Tian Hong, and Qing Nie (2018). "Exploring intermediate cell states through the lens of single cells". *Current Opinion in Systems Biology* 9. ISSN: 24523100. DOI: 10.1016/j.coisb.2018.02.009.

MacLean, Adam L., Maia A. Smith, Juliane Liepe, Aaron Sim, Reema Khorshed, Narges M. Rashidi, Nico Scherf, Axel Krinner, Ingo Roeder, Cristina Lo Celso, and Michael P. H. Stumpf (2017). "Single Cell Phenotyping Reveals Heterogeneity Among Hematopoietic Stem Cells Following Infection". *Stem Cells* 35.11. ISSN: 1066-5099. DOI: 10.1002/stem.2692.

Martin, Marcel (2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads". *EMBnet Journal* 17.1. ISSN: 2226-6089. DOI: 10.14806/ej.17.1.200.

Masugi, Y., K. Yamazaki, K. Emoto, K. Effendi, H. Tsujikawa, M. Kitago, O. Itano, Y. Kitagawa, and M. Sakamoto (2015). "Upregulation of integrin β4 promotes epithelial–mesenchymal transition and is a novel prognostic marker in pancreatic ductal adenocarcinoma". *Laboratory Investigation* 95.3. ISSN: 1530-0307. DOI: 10.1038/labinvest.2014.166.

McGinnis, Christopher S., David M. Patterson, Juliane Winkler, Daniel N. Conrad, Marco Y. Hein, Vasudha Srivastava, Jennifer L. Hu, Lyndsay M. Murrow, Jonathan S. Weissman, Zena Werb, Eric D. Chow, and Zev J. Gartner (2019). "MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices". *Nature Methods* 16.7. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0433-8.

McInnes, Leland, John Healy, and James Melville (2020). "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction". *arXiv*. DOI: 10.48550/arXiv.1802.03426.

Medici, Damian, Elizabeth D. Hay, and Bjorn R. Olsen (2008). "Snail and Slug Promote Epithelial-Mesenchymal Transition through β-Catenin–T-Cell Factor-4-dependent Expression of Transforming Growth Factor-β3". *Molecular Biology of the Cell* 19.11. ISSN: 1059-1524. DOI: 10.1091/mbc.e08-05-0506.

Miano, Carmen, Alessandra Morselli, Francesca Pontis, Chiara Bongiovanni, Francesca Sacchi, Silvia Da Pra, Donatella Romaniello, Riccardo Tassinari, Michela Sgarzi, Elvira Pantano, Carlo Ventura, Mattia Lauriola, and Gabriele D'Uva (2022). "NRG1/ERBB3/ERBB2 Axis Triggers Anchorage-Independent Growth of Basal-like/Triple-Negative Breast Cancer Cells". *Cancers* 14.7. ISSN: 2072-6694. DOI: 10.3390/cancers14071603.

Moon, Kevin R., David van Dijk, Zheng Wang, Scott Gigante, Daniel B. Burkhardt, William S. Chen, Kristina Yim, Antonia van den Elzen, Matthew J. Hirn, Ronald R. Coifman, Natalia B. Ivanova, Guy Wolf, and Smita Krishnaswamy (2019). "Visualizing structure and transitions in high-dimensional biological data". *Nature Biotechnology* 37.12. ISSN: 1546-1696. DOI: 10.1038/s41587-019-0336-3.

Moses, Lambda, Pétur Helgi Einarsson, Kayla Jackson, Laura Luebbert, A. Sina Booeshaghi, Sindri Antonsson, Nicolas Bray, Páll Melsted, and Lior Pachter (2023). "Voyager: exploratory single-cell genomics data analysis with geospatial statistics". *bioRxiv*. DOI: 10.1101/2023.07.20.549945.

Neufeld, Anna, Lucy L Gao, Joshua Popp, Alexis Battle, and Daniela Witten (2024). "Inference after latent variable estimation for single-cell RNA sequencing data". *Biostatistics* 25.1. ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxac047.

Ni, Man, Yue Zhao, and Xiaoguang Wang (2021). "Suppression of synuclein gamma inhibits the movability of endometrial carcinoma cells by PI3K/AKT/ERK signaling pathway". *Genes & Genomics* 43.6. ISSN: 2092-9293. DOI: 10.1007/s13258-021-01080-5.

Nieto, M. Angela, Ruby Yun-Ju Huang, Rebecca A. Jackson, and Jean Paul Thiery (2016). "EMT: 2016". *Cell* 166.1. ISSN: 00928674. DOI: 10.1016/j.cell.2016.06.028.

Panchy, Nicholas, Kazuhide Watanabe, Masataka Takahashi, Andrew Willems, and Tian Hong (2022). "Comparative single-cell transcriptomes of dose and time dependent epithelial–mesenchymal spectrums". *NAR Genomics and Bioinformatics* 4.3. ISSN: 2631-9268. DOI: 10.1093/nargab/lqac072.

Pastushenko, Ievgenia et al. (2018). "Identification of the tumour transition states occurring during EMT". *Nature* 556.7702. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-018-0040-3.

Puram, Sidharth V. et al. (2017). "Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer". *Cell* 171.7. ISSN: 00928674. DOI: 10.1016/j.cell.2017.10.044.

Qiu, Chengxiang, Beth K. Martin, et al. (2024). "A single-cell time-lapse of mouse prenatal development from gastrula to birth". *Nature* 626.8001. ISSN: 1476-4687. DOI: 10.1038/s41586-024-07069-w.

Qiu, Xiaojie, Yan Zhang, et al. (2022). "Mapping transcriptomic vector fields of single cells". *Cell* 185.4. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2021.12.045.

Rackauckas, Christopher and Qing Nie (2017). "DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia". *Journal of Open Research Software* 5.1. ISSN: 2049-9647. DOI: 10.5334/jors.151.

Regev, Aviv et al. (2017). "The Human Cell Atlas". *eLife* 6. ISSN: 2050-084X. DOI: 10.7554/eLife.27041.

Robin, Fabien, Gaëlle Angenard, Luis Cano, Laetitia Courtin-Tanguy, Elodie Gaignard, Zine-Eddine Khene, Damien Bergeat, Bruno Clément, Karim Boudjema, Cédric Coulouarn, and Laurent Sulpice (2020). "Molecular profiling of stroma highlights stratifin as a novel biomarker of poor prognosis in pancreatic ductal adenocarcinoma". *British Journal of Cancer* 123.1. ISSN: 1532-1827. DOI: 10.1038/s41416-020-0863-1.

Rommelfanger, Megan K. and Adam L. MacLean (2021). "A single-cell resolved cell-cell communication model explains lineage commitment in hematopoiesis". *Development* 148.24. ISSN: 0950-1991. DOI: 10.1242/dev.199779.

Rood, Jennifer E., Aidan Maartens, Anna Hupalowska, Sarah A. Teichmann, and Aviv Regev (2022). "Impact of the Human Cell Atlas on Medicine". *Nature Medicine* 28.12. ISSN: 1546-170X. DOI: 10.1038/s41591-022-02104-7.

Roskelley, Calvin D. and Mina J. Bissell (2002). "The dominance of the microenvironment in breast and ovarian cancer". *Seminars in Cancer Biology* 12.2. ISSN: 1044-579X. DOI: 10.1006/scbi.2001.0417.

Ruscetti, Marcus, Bill Quach, Eman L. Dadashian, David J. Mulholland, and Hong Wu (2015). "Tracking and Functional Characterization of Epithelial-Mesenchymal Transition and Mesenchymal Tumor Cells during Prostate Cancer Metastasis". *Cancer Research* 75.13. ISSN: 1538-7445. DOI: 10.1158/0008-5472.CAN-14-3476.

Saelens, Wouter, Robrecht Cannoodt, Helena Todorov, and Yvan Saeys (2019). "A comparison of single-cell trajectory inference methods". *Nature Biotechnology* 37.5. ISSN: 1546-1696. DOI: 10.1038/s41587-019-0071-9.

Sanford, Eric M, Benjamin L Emert, Allison Coté, and Arjun Raj (2020). "Gene regulation gravitates toward either addition or multiplication when combining the effects of two signals". *eLife* 9. ISSN: 2050-084X. DOI: 10.7554/eLife.59388.

Satija, Rahul, Jeffrey A. Farrell, David Gennert, Alexander F. Schier, and Aviv Regev (2015). "Spatial reconstruction of single-cell gene expression data". *Nature Biotechnology* 33.5. ISSN: 1546-1696. DOI: 10.1038/nbt.3192.

Sha, Yutong, Daniel Haensel, Guadalupe Gutierrez, Huijing Du, Xing Dai, and Qing Nie (2019). "Intermediate cell states in epithelial-to-mesenchymal transition". *Physical Biology* 16.2. ISSN: 1478-3967. DOI: 10.1088/1478-3975/aaf928.

Sha, Yutong, Shuxiong Wang, Federico Bocci, Peijie Zhou, and Qing Nie (2021). "Inference of Intercellular Communications and Multilayer Gene-Regulations of Epithelial–Mesenchymal Transition From Single-Cell Transcriptomic Data". *Frontiers in Genetics* 11. ISSN: 1664-8021. DOI: 10.3389/fgene.2020.604585.

Sha, Yutong, Shuxiong Wang, Peijie Zhou, and Qing Nie (2020). "Inference and multiscale model of epithelial-to-mesenchymal transition via single-cell transcriptomic data". *Nucleic Acids Research* 48.17. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkaa725.

Shan, Zhi-Guo, Zhen-Wei Sun, Li-Qun Zhao, Qiang Gou, Zhi-Fu Chen, Jin-Yu Zhang, Weisan Chen, Chong-Yu Su, Nan You, Yuan Zhuang, and Yong-Liang Zhao (2021). "Upregulation of Tubulointerstitial nephritis antigen like 1 promotes gastric cancer growth and metastasis by regulating multiple matrix metallopeptidase expression". *Journal of Gastroenterology and Hepatology* 36.1. ISSN: 1440-1746. DOI: 10.1111/jgh.15150.

Shaw, Tanya J and Paul Martin (2016). "Wound repair: a showcase for cell plasticity and migration". *Current Opinion in Cell Biology* 42. ISSN: 0955-0674. DOI: 10.1016/j.ceb.2016.04.001.

Shen, Minhong et al. (2019). "Tinagl1 Suppresses Triple-Negative Breast Cancer Progression and Metastasis by Simultaneously Inhibiting Integrin/FAK and EGFR Signaling". *Cancer Cell* 35.1. ISSN: 1535-6108. DOI: 10.1016/j.ccell.2018.11.016.

Sherman, Brad T., Ming Hao, Ju Qiu, Xiaoli Jiao, Michael W. Baseler, H. Clifford Lane, Tomozumi Imamichi, and Weizhong Chang (2022). "DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update)". *Nucleic Acids Research*. ISSN: 1362-4962. DOI: 10.1093/nar/gkac194.

Shi, Dong-Min, Li-Xin Li, Xin-Yu Bian, Xue-Jiang Shi, Li-Li Lu, Hong-Xin Zhou, Ting-Jia Pan, Jian Zhou, Jia Fan, and Wei-Zhong Wu (2018). "miR-296-5p suppresses EMT of hepatocellular carcinoma via attenuating NRG1/ERBB2/ERBB3 signaling". *Journal of Experimental & Clinical Cancer Research* 37.1. ISSN: 1756-9966. DOI: 10.1186/s13046-018-0957-2.

Shiba-Ishii, Aya, Yunjung Kim, Toshihiro Shiozawa, Shinji Iyama, Kaishi Satomi, Junko Kano, Shingo Sakashita, Yukio Morishita, and Masayuki Noguchi (2015). "Stratifin accelerates progression of lung adenocarcinoma at an early stage". *Molecular Cancer* 14.1. ISSN: 1476-4598. DOI: 10.1186/s12943-015-0414-1.

Simeonov, Kamen P., China N. Byrns, Megan L. Clark, Robert J. Norgard, Beth Martin, Ben Z. Stanger, Jay Shendure, Aaron McKenna, and Christopher J. Lengner (2021). "Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states". *Cancer Cell* 39.8. ISSN: 1535-6108. DOI: 10.1016/j.ccell.2021.05.005.

Srivastava, Avi, Laraib Malik, Tom Smith, Ian Sudbery, and Rob Patro (2019). "Alevin efficiently estimates accurate gene abundances from dscRNA-seq data". *Genome Biology* 20.1. ISSN: 1474-760X. DOI: 10.1186/s13059-019-1670-y.

Street, Kelly, Kimberly Siegmund, and Darryl Shibata (2023). "Epigenetic Conservation Infers That Colorectal Cancer Progenitors Retain The Phenotypic Plasticity Of Normal Colon". *Research Square*. DOI: 10.21203/rs.3.rs-2609517/v1.

Stumpf, Patrick S., Rosanna C. G. Smith, Michael Lenz, Andreas Schuppert, Franz-Josef Müller, Ann Babtie, Thalia E. Chan, Michael P. H. Stumpf, Colin P. Please, Sam D. Howison, Fumio Arai, and Ben D. MacArthur (2017). "Stem Cell Differentiation as a Non-Markov Stochastic Process". *Cell Systems* 5.3. ISSN: 2405-4712. DOI: 10.1016/j.cels.2017.08.009.

Subramanian, Aravind, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov (2005). "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles". *Proceedings of the National Academy of Sciences* 102.43. DOI: 10.1073/pnas.0506580102.

Sun, Qi, Yukui Shang, Fengkai Sun, Xiwen Dong, Jun Niu, and Fanni Li (2020). "Interleukin-6 Promotes Epithelial-Mesenchymal Transition and Cell Invasion through Integrin β6 Up-regulation in Colorectal Cancer". *Oxidative Medicine and Cellular Longevity* 2020. ISSN: 1942-0900. DOI: 10.1155/2020/8032187.

Tagliazucchi, Guidantonio Malagoli, Anna J. Wiecek, Eloise Withnell, and Maria Secrier (2023). "Genomic and microenvironmental heterogeneity shaping epithelial-to-mesenchymal trajectories in cancer". *Nature Communications* 14.1. ISSN: 2041-1723. DOI: 10.1038/s41467-023-36439-7.

Takemura, Yusuke, Hidenori Ojima, Go Oshima, Masahiro Shinoda, Yasushi Hasegawa, Minoru Kitago, Hiroshi Yagi, Yuta Abe, Shutaro Hori, Yoko Fujii-Nishimura, Naoto Kubota, Yuki Masuda, Taizo Hibi, Michiie Sakamoto, and Yuko Kitagawa (2021). "Gamma-synuclein is a novel prognostic marker that promotes tumor cell migration in biliary tract carcinoma". *Cancer Medicine* 10.16. ISSN: 2045-7634. DOI: 10.1002/cam4.4121.

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. ISSN: 1757-4684. DOI: 10.15252/emmm.201404208.

The UniProt Consortium (2023). "UniProt: the Universal Protein Knowledgebase in 2023". *Nucleic Acids Research* 51.D1. ISSN: 0305-1048. DOI: 10.1093/nar/gkac1052.

Theodoris, Christina V., Ling Xiao, Anant Chopra, Mark D. Chaffin, Zeina R. Al Sayed, Matthew C. Hill, Helene Mantineo, Elizabeth M. Brydon, Zexian Zeng, X. Shirley Liu, and Patrick T. Ellinor (2023). "Transfer learning enables predictions in network biology". *Nature* 618.7965. ISSN: 1476-4687. DOI: 10.1038/s41586-023-06139-9.

Thiery, Jean Paul (2003). "Epithelial–mesenchymal transitions in development and pathologies". *Current Opinion in Cell Biology* 15.6. ISSN: 0955-0674. DOI: 10.1016/j.ceb.2003.10.006.

Thiery, Jean Paul, Hervé Acloque, Ruby Y. J. Huang, and M. Angela Nieto (2009). "Epithelial-Mesenchymal Transitions in Development and Disease". *Cell* 139.5. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2009.11.007.

Thomas, Paul D., Dustin Ebert, Anushya Muruganujan, Tremayne Mushayahama, Laurent-Philippe Albou, and Huaiyu Mi (2022). "PANTHER: Making genome-scale phylogenetics accessible to all". *Protein Science: A Publication of the Protein Society* 31.1. ISSN: 1469-896X. DOI: 10.1002/pro.4218.

Thorpe, Hannah, Abdulaziz Asiri, Maham Akhlaq, and Mohammad Ilyas (2017). "Cten promotes epithelial-mesenchymal transition through the post-transcriptional stabilization of Snail". *Molecular Carcinogenesis* 56.12. ISSN: 1098-2744. DOI: 10.1002/mc.22704.

Tian, Xiao-Jun, Hang Zhang, and Jianhua Xing (2013). "Coupled Reversible and Irreversible Bistable Switches Underlying TGFβ-induced Epithelial to Mesenchymal Transition". *Biophysical Journal* 105.4. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2013.07.011.

Ting, David T. et al. (2014). "Single-Cell RNA Sequencing Identifies Extracellular Matrix Gene Expression by Pancreatic Circulating Tumor Cells". *Cell Reports* 8.6. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2014.08.029.

Tirosh, Itay et al. (2016). "Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq". *Science* 352.6282. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aad0501.

Traag, V. A., L. Waltman, and N. J. van Eck (2019). "From Louvain to Leiden: guaranteeing well-connected communities". *Scientific Reports* 9.1. ISSN: 2045-2322. DOI: 10.1038/s41598-019-41695-z.

Tripathi, Shubham, Jianhua Xing, Herbert Levine, and Mohit Kumar Jolly (2021). "Mathematical Modeling of Plasticity and Heterogeneity in EMT". *The Epithelial-to Mesenchymal Transition: Methods and Protocols*. Springer Protocols. ISBN: 978-1-07-160779-4. DOI: 10.1007/978-1-0716-0779-4_28.

van Dijk, David, Roshan Sharma, Juozas Nainys, Kristina Yim, Pooja Kathail, Ambrose J. Carr, Cassandra Burdziak, Kevin R. Moon, Christine L. Chaffer, Diwakar Pattabiraman, Brian Bierie, Linas Mazutis, Guy Wolf, Smita Krishnaswamy, and Dana Pe'er (2018). "Recovering Gene Interactions from Single-Cell Data Using Data Diffusion". *Cell* 174.3. ISSN: 00928674. DOI: 10.1016/j.cell.2018.05.061.

Weinreb, Caleb, Samuel Wolock, Betsabeh K. Tusi, Merav Socolovsky, and Allon M. Klein (2018). "Fundamental limits on dynamic inference from single-cell snapshots". *Proceedings of the National Academy of Sciences* 115.10. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1714723115.

Williams, Elizabeth D., Dingcheng Gao, Andrew Redfern, and Erik W. Thompson (2019). "Controversies around epithelial–mesenchymal plasticity in cancer metastasis". *Nature Reviews Cancer* 19.12. ISSN: 1474-1768. DOI: 10.1038/s41568-019-0213-x.

Wolf, F. Alexander, Philipp Angerer, and Fabian J. Theis (2018). "SCANPY: large-scale single-cell gene expression data analysis". *Genome Biology* 19.1. ISSN: 1474-760X. DOI: 10.1186/s13059-017-1382-0.

Wu, Xiaojun, Roy Wollman, and Adam L. MacLean (2023). "Single-cell Ca^{2+} parameter inference reveals how transcriptional states inform dynamic cell responses". *Journal of The Royal Society Interface* 20.203. ISSN: 1742-5662. DOI: 10.1098/rsif.2023.0172.

Xing, Jianhua and Xiao-Jun Tian (2019). "Investigating epithelial-to-mesenchymal transition with integrated computational and experimental approaches". *Physical Biology* 16.3. ISSN: 1478-3975. DOI: 10.1088/1478-3975/ab0032.

Yang, Jing et al. (2020). "Guidelines and definitions for research on epithelial–mesenchymal transition". *Nature Reviews Molecular Cell Biology* 21.6. ISSN: 1471-0080. DOI: 10.1038/s41580-020-0237-9.

Ye, Shan-Ping, Hong-Xin Yu, Wei-Jie Lu, Jun-Fu Wang, Tai-Yuan Li, Jun Shi, and Xiao-Ye Cheng (2023). "Stratifin Promotes Hepatocellular Carcinoma Progression by Modulating the Wnt/β-Catenin Pathway". *International Journal of Genomics* 2023. ISSN: 2314-436X. DOI: 10.1155/2023/9731675.

Yu, Min et al. (2013). "Circulating Breast Tumor Cells Exhibit Dynamic Changes in Epithelial and Mesenchymal Composition". *Science* 339.6119. DOI: 10.1126/science.1228522.

Yun, Sumi, Jiwon Koh, Soo Kyung Nam, Jung Ok Park, Sung Mi Lee, Kyoungyul Lee, Kyu Sang Lee, Sang-Hoon Ahn, Do Joong Park, Hyung-Ho Kim, Gheeyoung Choe, Woo Ho Kim, and Hye Seung Lee (2018). "Clinical significance of overexpression of NRG1 and its receptors, HER3 and HER4, in gastric cancer patients". *Gastric Cancer: Official Journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association* 21.2. ISSN: 1436-3305. DOI: 10.1007/s10120-017-0732-7.

Zhang, Jingyu, Xiao-Jun Tian, Hang Zhang, Yue Teng, Ruoyan Li, Fan Bai, Subbiah Elankumaran, and Jianhua Xing (2014). "TGF-β-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops". *Science Signaling* 7.345. ISSN: 1937-9145. DOI: 10.1126/scisignal.2005304.

Zhang, Yun, Joana Liu Donaher, et al. (2022). "Genome-wide CRISPR screen identifies PRC2 and KMT2D-COMPASS as regulators of distinct EMT trajectories that contribute differentially to metastasis". *Nature Cell Biology* 24.4. ISSN: 1476-4679. DOI: 10.1038/s41556-022-00877-0.

Zhao, Xinyu, Enqin Wang, Hongkun Xu, and Lihong Zhang (2023). "Stratifin promotes the growth and proliferation of hepatocellular carcinoma". *Tissue and Cell* 82. ISSN: 0040-8166. DOI: 10.1016/j.tice.2023.102080.

Zheng, Guixi, Hakim Bouamar, Matyas Cserhati, Carla R. Zeballos, Isha Mehta, Habil Zare, Larry Broome, Ruolei Hu, Zhao Lai, Yidong Chen, Francis E. Sharkey, Meenakshi Rani, Glenn A. Halff, Francisco G. Cigarroa, and Lu-Zhe Sun (2022). "Integrin alpha 6 is upregulated and drives hepatocellular carcinoma progression through integrin $\alpha 6\beta 4$ complex". *International Journal of Cancer* 151.6. ISSN: 1097-0215. DOI: 10.1002/ijc.34146.

Zhuang, Qing, Caiyun Liu, Like Qu, and Chengchao Shou (2015). "Synuclein-γ promotes migration of MCF7 breast cancer cells by activating extracellular-signal regulated kinase

pathway and breaking cell-cell junctions". *Molecular Medicine Reports* 12.3. ISSN: 1791-2997. DOI: 10.3892/mmr.2015.3799.