From cell types to cell states: adapting foundation models for cellular plasticity

Shi Pan[1], Eloise Withnell[1], Cenk Celik[1], Maria Secrier[1,*]

[1]UCL Genetics Institute, Department of Genetics, Evolution and Environment, University College London, London, WC1E 6BT, UK

*Contact: m.secrier@ucl.ac.uk

Abstract

Single-cell foundation models (scFMs) have been widely adopted for cell type annotation, yet their suitability for modelling cellular plasticity, where cells transition along continuous, context-dependent state trajectories, remains unclear. Here, we systematically benchmark state-of-the-art scFMs against conventional machine learning and bioinformatics approaches on epithelial-mesenchymal plasticity (EMP), a prototypical plastic cellular process. We show that naïve fine-tuning of scFMs often fails to resolve intermediate states and is strongly influenced by tissue- and stimulus-specific signals. We propose a parameter-efficient dual-task adaptation strategy for EMP foundation models (EMP-FM) that combines discrete classification with pseudotime-guided regression, which improves cell state resolution in controlled settings (up to 85% AUROC), but remains sensitive to domain shifts. Across diverse *in vitro* and *in vivo* datasets, scFMs do not consistently outperform conventional methods, which often achieve comparable performance with lower complexity. Together, our results delineate both the potential and current limitations of scFMs for modelling cellular plasticity and support their complementary use alongside established bioinformatics approaches.

INTRODUCTION

Single cell omics technologies have made it possible to characterise cellular identity and behaviour at unprecedented resolution. While much of the field has focused on cell type classification, many biologically important phenomena manifest not as discrete identities but as continuous, context-dependent cell states. Cell types represent relatively stable, discrete cellular identities (e.g., T cells, B cells, fibroblasts) that can be effectively distinguished through categorical classification approaches. In contrast, cell *states* (e.g. stemness, metabolic rewiring, activation) reflect dynamic, context-dependent behaviour that cells adopt along a

biological continuum. A classical example is the epithelial-to-mesenchymal transition (EMT), a developmental process in which epithelial cells gradually acquire mesenchymal characteristics, and which is frequently co-opted in cancer to promote invasion and metastasis[1,2]. EMT and other plastic processes, such as differentiation and immune activation, are characterised by a continuum of transformation rather than stable, well-separated transcriptional categories[3], are stimulus-dependent and may lack clear boundaries. As a result, they pose fundamentally different computational challenges.

Traditional bioinformatics approaches, typically relying on canonical markers (e.g. E-cadherin or vimentin[4-6] for EMT) or limited gene signatures[7,8], and conventional machine learning approaches designed for discrete classification offer valuable insights but are inherently ill-suited for modelling the fluid, continuous nature of cellular plasticity. As cancer cell trajectories vary depending on the tissue and external stimuli[9], these strategies struggle to resolve subtle intermediate states or to generalise across experimental conditions. Furthermore, continuous processes often involve long-range or nonlinear regulatory dependencies[10] that are not easily captured by linear models or shallow classifiers, limiting our ability to develop unified frameworks to characterise cellular plasticity in diverse systems and beyond tightly controlled *in vitro*/*in vivo* settings.

Single cell foundation models (scFMs) like scBERT[11], scGPT[12], Cell2Sentence[13], GenePT[14] or scFoundation[15], have introduced a powerful new paradigm that holds promise in addressing such challenges. These models are trained on extremely large and diverse scRNA-seq corpora, leveraging transformer architectures[16-18] that enable them to encode broad transcriptional programmes, learn long-range gene-gene dependencies and, in principle, provide transferable representations for downstream tasks[19]. However, current scFMs have been predominantly developed and evaluated for discrete annotation tasks, especially cell type classification. This creates a fundamental mismatch when applied to continuous biological processes like EMP: the learned representations tend to prioritise strong sources of variation, such as tissue of origin or stimulus, over the more subtle, dynamic signals that define cell state transitions. Furthermore, recent evaluations also reveal significant limitations of these models in a zero-shot cell-type classification scenario, with scFMs often underperforming compared to simpler baselines[20]. However, the potential of fine-tuning scFMs for cell state classification remains insufficiently explored in systematic settings.

These limitations raise two important questions. First, can existing scFMs be effectively adapted to model continuous cellular processes, or do they require fundamentally different training objectives and architectures? Second, what design principles are needed for foundation models to capture dynamic transcriptional changes that unfold along biological continua? To explore these questions, we focus on epithelial–mesenchymal plasticity (EMP) as a representative system. EMP provides a well-characterised yet challenging example of a continuous transition involving heterogeneous intermediate states, strong context dependency, and substantial variability across cancer types and stimuli. Importantly, it offers an experimentally grounded testbed for assessing the strengths and weaknesses of current scFM paradigms when repurposed for cell state modelling.

By systematically benchmarking conventional approaches and several state-of-the-art scFMs on modelling EMP across multiple cancer types and under different stimuli, we uncover fundamental limitations in how pre-trained scFMs represent cell states and demonstrate why naïve fine-tuning is insufficient to overcome these issues. To address this, we develop a parameter-efficient dual-path adaptation framework that reshapes the pretrained feature space to better capture context-dependent trajectories, while preserving the underlying structure learned during large-scale pretraining. Our results highlight both the promise and the current limitations of using foundation models for continuous cell-state modelling, and outline conceptual and methodological directions for future scFM development.

RESULTS

EMP-FM, a foundation model that predicts EMT states in single cell cancer data

To facilitate a better understanding of cellular plasticity, we explored whether a pre-trained foundation model can quantify EMP from scRNA-seq data. We introduce EMP-FM, a fine-tuned scFM designed to probe whether targeted architectural and training choices using both a classification and a regression strategy can improve the representation of cell states along the EMT continuum. To define the ground truth transcriptional states that cells occupy as they undergo EMT in different cancer tissues, we employed data from Cook and Vaderhyden[9]. This study encompasses single cell RNA sequencing from breast (MCF7), lung (A549), prostate (DU145) and ovarian (OVCA420) cancer cell lines undergoing EMT, profiled at 0 days, 8 hours, 1 day, 3 days and 7 days after stimulation with TGFβ, TNF or EGF. We labelled cells

sequenced at day 0 as epithelial (E), and ulterior states as hybrid E/M (EM1 at 8 hours, EM2 at 1 day, EM3 at 3 days) and mesenchymal (M, 7 days), respectively (**Supplementary Figure 1**). This experimental setup provides clean transcriptional profiles for distinct states along the EMT continuum.

The EMP transcriptional logic can be learned from the experimentally measured longitudinal profiles of this dataset. We followed the embedding strategy in other pre-trained scFMs[11,12] and designed a two-stage dual path pipeline consisting of extra residual connections to enhance the raw gene expression signal, a Parameter-Efficient Fine-Tuning (PEFT)[21] module and a Mixture of Expert (MoE)[22] architecture with: (1) a classification head to learn discrete cell state labels, and (2) a regression head using the pseudotemporal ordering calculated by psupertime[9] (**Figure 1a-b and Methods**). The dual modelling classification/regression strategy addresses the inherent challenge that EMP represents a continuous biological process that cannot be fully captured by discrete classification alone, thus transforming the pre-trained scFM from a general cell type classifier to a specialized cell state predictor capable of capturing fine-grained EMP cell state transitions.

The feature space from the pre-trained backbone of scBERT[11] was modified to learn a generalisable EMP process across different tissues, and each expert network (named scMultiNet) focused on an individual target cell state across multiple tissues. Finally, we add an interpretability module to allow the exploration of the gene regulatory programmes underlying these states (**Figure 1c**).

EMP-FM behind the scenes - gradual learning of EMP states

Understanding the patterns learned by EMP-FM is a crucial step in assessing whether pre-trained scFMs can prioritise subtle process-specific transitions like EMP over dominant signals such as tissue type. We explored the learned latent space evolution through five embedding stages using principal component analysis (PCA) (**Figure 2a)**: the original data distribution, the feature space of the data after being processed by the original scBERT, the feature space generated by scBERT after PEFT, the feature space of our proposed scMultiNet, and finally, the feature space resulting from the EMP-FM fusion. In the raw gene expression data, cells clustered by cancer cell line rather than experimental timepoints, demonstrating how tissue-specific signatures mask EMP patterns. Higher-order principal components (**Supplementary Figure 2a**) showed intermixing of EMP states rather than clear EMP trajectories, confirming

that EMP states are inseparable in the raw data. While the pre-trained scBERT reduced cell line batch effects, it failed to separate EMP timepoints, with cells remaining extensively intermixed (**Figure 2a**). This exemplifies how pre-trained scFMs, despite zero-shot cell type classification capabilities, struggle to capture subtle variations in biological processes across tissues and stimuli. Our proposed scMultiNet achieved more distinct boundaries and tighter clustering between EMP categories, while the final EMP-FM model showed distinct separation of all EMP states, demonstrating successful capture of EMP-related features (**Figure 2a**). Remarkably, although the model was trained using discrete labels without explicit temporal information, it also successfully captures the sequential nature of the EMT process, with cells naturally arranging themselves from top right to bottom left (blue to orange) on a chronological EMT continuum in the latent space (**Figure 2a,** also shown with PHATE[23] and Diffusion Pseudotime (DPT)[24] analysis in **Supplementary Fig 2b**). This demonstrates the model's ability to learn the underlying dynamics from the data.

Benchmarking of EMP-FM

Having established that EMP-FM can learn meaningful EMT representations across conditions, we benchmarked its classification performance against state-of-the-art scFMs, conventional machine learning and bioinformatics methods. All methods were trained on the same 80% randomly sampled single cells across distinct EMT-inducing stimuli (TGFβ, TNF, EGF) and tested on the remaining 20%. EMP-FM achieved 85% AUROC, outperforming scFMs with standard fine-tuning: scGPT[12] (72%), Geneformer[25] (73%), and Universal Cell Embedding (UCE)[26] (56.7%) (**Figure 2b**, **Supplementary Table 1**). EMP-FM's performance also exceeded that of conventional machine learning approaches including K-Nearest Neighbours (KNN, 57%), random forest (75%) and AdaBoost (83%) (**Figure 2b, Supplementary Table 1**). Among specialised bioinformatics methods, scPred[27], a supervised method to discretise cell types/states through feature selection on highly variable genes, showed similar performance (74%) to some of the other fine-tuned scFMs, while AUCell[28], which assigns a continuous score per cell based on the expression of a given gene set, achieved a lower performance of 62%. scType[29], a cluster annotation method based on known cell markers, could only distinguish epithelial and mesenchymal states, mis-annotating most hybrid states (**Figure 2b, Supplementary Figure 3**). Overall, EMP-FM achieved higher classification performance compared to both traditional methods and state-of-the-art scFMs in capturing targeted EMP cell states in this controlled setting.

154 We also built stimulus-specific EMP-FMs, which only adapted the classification head in the
155 fusion network and were trained exclusively on individual stimulation conditions across all
156 tissues (TGFβ, TNF, EGF). EMP-FM-TGFβ achieved 99% AUROC, while EMP-FM-TNF and
157 EMP-FM-EGF reached 89% and 76% AUROCs, respectively (**Figure 2c, Supplementary**
158 **Figure 4, Supplementary Tables 1-2**). Our results are in agreement with previous reports[9],
159 suggesting that the canonical EMT inducer TGFβ drives coordinated transcriptional changes
160 delineating a comparatively more deterministic process[30-32]. Capturing EGF-induced
161 transitions was more challenging, potentially due to increased transcriptional heterogeneity[33],
162 while in the case of TNF-induced EMT the models may struggle to capture complex pathway
163 synergies required for stable EMT phenotypes[34].

164 All EMP states showed good discrimination with AUROCs greater than 78% in the EMP-FM-
165 ALL model and greater than 98% in the TGFβ model (**Figure 2d**). Both TNF and EGF models
166 performed best at the 3-day timepoint (EM3: 98% and 85% AUROC, respectively), showing
167 superior discrimination of intermediate and late EMT compared to early 8-hour transitions
168 which are likely less stable (**Supplementary Table 2**). Notably, here each model captured
169 stimulus-specific programmes, as applying them to the other stimulus scenarios decreased
170 classification performance significantly (**Supplementary Table 3**).

171 These results show that generic scFMs, although minimally reliant on preprocessing or curated
172 gene lists, do not consistently outperform conventional machine learning or bioinformatics
173 approaches for cell state classification, with methods such as AdaBoost and scPred often
174 achieving comparable or superior performance. Carefully designed, parameter-efficient
175 adaptations have can substantially improve scFM performance, but remain sensitive to context
176 and experimental conditions.

177 **Intra-dataset and domain-shift validation**

178 To evaluate the generalisability and limitations of scFMs for cell state learning tasks, we
179 conducted validation in four EMP-centric experimental settings: mesenchymal-to-epithelial
180 reversal experiments from the same study by Cook and Vanderhyden, independent cell line
181 datasets under different experimental designs, genetically engineered mouse model (GEMM)
182 data, and human *in situ* datasets (spatial transcriptomics).

183 Cook and Vanderhyden[9] tested two EMT reversal scenarios in their study: (1) where the EMT-
184 inducing stimulus was removed from mesenchymal cells, and (2) where mesenchymal cells
185 were treated with EMT inhibitor LY364947 (targeting TGFBR1). Both scenarios should induce
186 partial epithelial reversion, which allows us to question whether the model has learned EMP
187 biology rather than memorising time-point associations. Neither scenarios would be expected
188 to show a complete reversion to the 0-day state, or to follow a similar trajectory from M to E
189 as from E to M. EMP-FM confirmed a significant reduction in EMT (Wilcoxon rank-sum test
190 $p < 0.001$) following both inhibitor treatment (**Figure 3a-b**) and stimulation removal (**Figure**
191 **3c, Supplementary Figure 5a**). These results demonstrate that scFMs can learn meaningful
192 biological transition patterns rather than noise or batch effects, with successful bidirectional
193 prediction validating its understanding of continuous EMT regulatory logic across distinct
194 experimental perturbations. However, not all EMP states remained consistently present during
195 the reversal process (**Figure 3d, Supplementary Figure 5b**), likely reflecting subtle
196 differences in dominant cell states and cancer trajectories under varying conditions, as posited
197 by Cook and Vanderhyden[9]. This is a common challenge in cell state tasks that most
198 classification-only scFM frameworks cannot adequately address. Our proposed dual-path
199 architecture incorporating a regression element can overcome this by assigning continuous
200 EMT scores, which revealed the expected downward trends in this experiment
201 (**Supplementary Figure 5a**).

202 We further explored the potential and limitations of scFMs in tracking EMT progression *in*
203 *vitro* using external cell line scRNA-seq datasets: MCF10A breast epithelial cells treated with
204 TGFβ from Paul et al[35], and MCF10A cells undergoing spontaneous or TGFβ-induced EMT
205 from McFaline-Figueroa et al[36]. Compared to our training data, the Paul et al[35] dataset used the
206 same TGFβ stimulus but with denser temporal sampling, while McFaline-Figueroa et al[36]
207 provided a different experimental design that captured both spontaneous and TGFβ-induced
208 EMP processes at two time points, before and after EMT. As directly applying the EMP-FM
209 classifier head to these datasets faced challenges due to mismatched class numbers between
210 training and testing datasets, we generated continuous scores instead to provide more
211 appropriate evaluation metrics and constructed a pseudo confusion matrix to demonstrate
212 limitations of scFMs. In the Paul et al[35] dataset, EMP-FM-TGFβ showed consistent EMT score
213 increases across T0-T4 timepoints, as expected, with T4-T7 reaching steady state (**Figure 4a-**
214 **b, Supplementary Tables 4-5**) and outperformed differentially expressed gene (DEG)
215 signatures that failed to capture T3 transitions (**Supplementary Figures 6, 7a**). In the

McFaline-Figueroa et al[36] dataset, EMP-FM-TGFβ specifically captured TGFβ-driven but not spontaneous EMT increases, while EMP-FM-TNF, EMP-FM-EGF and all-in-one models detected increases in both conditions, suggesting spontaneous EMT resembles TNF/EGF-induced pathways (**Supplementary Figures 8, 9a, Supplementary Table 5**). However, this does not indicate that scFMs simply outperform conventional approaches; conversely, in some cases we observed that traditional bioinformatics methods outperform scFMs, particularly for unseen experimental designs (i.e. spontaneous EMT scenarios, **Supplementary Table 5**).

We next extended our validation to experiments performed in GEMMs to evaluate the generalisation capabilities of scFMs with domain shifts *in vivo* in two scenarios: (1) breast cancer cells tracked as they become increasingly invasive (thus activating EMT) at 12 hours, 3 days, and 5 days from Grasset et al[37]; and (2) knockout of a key EMT regulator, Zeb1, in basal mammary epithelial cells from Han et al[38]. In the cell invasion dataset, EMP-FM captured the expected gradual EMT progression at days 3 and 5 (**Figure 4c**), while differentially expressed gene signatures showed bimodal distribution at day 3 and unexpected decreases at day 5 (**Supplementary Figure 9b**). In the Zeb1 knockout setting, we observed decreased EMT scores consistent with expected EMT attenuation following Zeb1 loss (**Figure 4d**), whereas gene signatures incorrectly predicted EMT increases (**Supplementary Figure 9c**). However, despite incorporating broader gene associations, EMP-FM showed comparable performance to conventional machine learning and bioinformatics methods in several instances (**Supplementary Table 4-5**). Furthermore, training data composition significantly influenced the scFMs' predictive performance across different GEMM datasets, while traditional methods provided relatively stable results with substantially lower computational overhead.

Finally, EMP-FM was further evaluated using spatial transcriptomics (ST) data from Cassier et al[39], profiling endometrial tumours from two patients before and after Netrin-1 EMT inhibitor treatment. It is worth noting this is a challenging test setting due to differences between scRNA-seq and ST data and distinct cancer tissue. Without additional processing steps, our model predicted decreased EMP scores post-treatment in both cases (**Figure 5e**), with reductions visible throughout tissue slides (**Figure 5f**). The fact that our model is able to capture this in endometrial cancer, which is not a cancer type it was originally trained on, suggests that aspects of the learned EMP-related signal may transfer beyond the cancer types represented in the training data.

247 Importantly, model-predicted scores exhibited different scales across datasets, reflecting
248 complexities regarding EMT entry timing, duration and progression dynamics that lack
249 established analytical frameworks.

250 **Gene programmes underlying TGFβ-induced EMT state transitions in cancer**

251 To understand the gene regulation patterns underlying scFMs' predictions, we developed the
252 Attention-Driven Expression Significance Index (ADESI) to prioritise biologically meaningful
253 changes that overcome the scFMs' sensitivity to minor expression variation (see Methods).
254 ADESI-identified genes showed distinct expression dynamics across EMT states in EMP-FM
255 (**Figure 5a**), delineating these states more effectively than differentially expressed genes that
256 exhibited extensive within-state heterogeneity (**Supplementary Figure 10**). Optimisation
257 analysis (see Methods) revealed that a list of 200 genes provided sufficient information for
258 reliable EMT state determination (**Figure 5b, Supplementary Table 6**). Validation using EMT
259 inhibitor datasets from Cook and Vanderhyden[9] and Cassier et al[39] confirmed the gene list's
260 biological relevance, with epithelial (E) signatures increasing upon inhibitor treatment as
261 expected (**Figure 5c-d**). The processes captured by the model included classical mesenchymal
262 regulation markers such as *VIM* upregulation in EM3, but also other genes involved in
263 processes linked with EMT such as oxidoreduction and ATP synthesis[40], metabolic
264 reprogramming, TGFβ signalling, motility and metastasis[41,42] (**Supplementary Table 7**).

265

266 **DISCUSSION**

267 In this work, we unveil the potential and limitations of applying scFMs as a tool for exploring
268 cellular plasticity. Defining EMP as a prime example of a dynamic and plastic cellular process
269 that involves multiple cell state switches, we systematically benchmarked scFMs against
270 conventional machine learning and bioinformatics approaches through a comprehensive
271 evaluation framework encompassing multiple datasets representing the same targeted
272 underlying biological process. Our results demonstrate that, while scFMs show promise in
273 capturing subtle gene regulatory patterns underlying EMP beyond up/downregulated genes
274 inferred through traditional bioinformatics, significant performance disparities persist.
275 Specifically, we identified that pre-trained models encode mixed signals from diverse
276 biological processes, creating fundamental challenges when targeting complex phenomena like

EMP that lack dominant transcriptional signatures - where subtle state transitions can be obscured by stronger tissue-type or cell-type signals.

Our results indicate that carefully designed adaptation strategies, such as the proposed dual-path classification/regression pipeline, can partially eliminate the gaps to achieve superior performance to simple fine-tuning pipelines and other SOTA methods. Our pipeline involved a naïve dual-path design with PEFT and MoE modules to systematically adapt pre-trained feature spaces while preserving foundational biological knowledge, enabling targeted optimisation for specific processes with reduced computational demands. Our proposed EMP-FMs can be considered as a proof-of-concept that scFMs, with careful design, could be adapted to model cell states and plasticity within limited settings. However, while our proposed approach shows some advantages, scFMs may underperform conventional methods in cross-dataset and domain shift scenarios. Substantial gaps remain for complex cell state classification tasks, and regression scores require recalibration. Thus, current scFMs require fundamental improvements in feature representation and architectural design to fully replace conventional approaches in future applications.

Compared to the cell type classification, predicting cell states is substantially more challenging due to their uncertainty. Single cell transcriptomic datasets offer time point snapshots during the process, with no guarantee that the states captured at those time points are stable. Conflicting evidence on EMP representing either continuous or discrete transitions further underscores this complexity[43-48]. This uncertainty demands models sensitive to subtle yet meaningful cellular perturbations - a requirement that distinguishes cell state classification from conventional cell type identification tasks. The proposed dual-path architecture addresses the challenge of capturing biological continuity from discrete, temporally sampled snapshots by jointly producing discrete state labels and continuous scores that reconstruct the underlying EMT continuum from potentially unstable cellular states. Continuous scores are preferable when experimental setups differ significantly from training conditions, but work to recalibrate these scores to new datasets will be required in the future.

Notably, the predictions from classical bioinformatics or machine learning methods aligned well with experimental measurements in several scenarios (e.g. spontaneous EMT) with much lower computational cost. Considering that, EMP is highly dependent on context[9,43], both tissue and stimulus[9,43], this emphasises the difficulty in developing a single unified framework for EMP and the value of traditional metrics for EMP alongside advanced foundation models,

309 which potentially could be used in a complementary manner in the future. The fact that our
310 EMP-FMs do not simply recapitulate classical EMT drivers but rather capture broader RNA
311 metabolism, cellular energetics and motility features that accompany EMT, suggest their
312 potential advantage over other methods in uncovering new regulators of this process, which
313 will need to be validated experimentally. Future work should focus on expanding the capability
314 of such models to capture a broader array of EMT programmes profiled under different stimuli
315 and in more physiological contexts integrating the tumour microenvironment. However, such
316 longitudinal datasets with ground truth EMT tracking are limited at the moment.

317 This study establishes a proof-of-concept that scFMs can be systematically adapted to
318 characterize complex biological processes exhibiting continuous rather than discrete
319 transitions, while also revealing persistent performance gaps relative to conventional
320 approaches. By introducing a dual-path adaptation framework compatible with PEFT and MoE
321 architectures, we outline a potential direction for extending pre-trained scFMs to other dynamic
322 cellular processes. Our results indicate that fundamental improvements are still needed in scFM
323 pre-training and architectural design, but scFMs hold potential in understanding biological
324 complexity beyond conventional classification tasks.

325

326 **METHODS**

327 **Training dataset**

328 We sourced gene expression data from Cook and Vanderhyden[8], made available at:
329 https://drive.google.com/drive/folders/1SIEIf7UswTv_0S6TypYsaRzMcfkfsgji
330 (GSE147405). The dataset encompasses 53,290 cells profiled through scRNA-seq from A549
331 (lung), DU145 (prostate), MCF7 (breast) and OVCA420 (ovarian) cancer cell lines, subjected
332 to TGFβ (9,084 cells), TNF (22,159 cells) and EGF (22,047 cells) induced EMT, paired with
333 EMT sampling times ranging from 0 day (epithelial state), 8 hours, 1 day, 3 days (hybrid states
334 labelled EM1, EM2, EM3, respectively) to 7 days (closer to a fully mesenchymal state). These
335 cell lines maintain an epithelial morphology and have been demonstrated to undergo EMT in
336 previous studies. Morphological changes consistent with EMT in each cell line have been
337 documented in Cook and Vanderhyden[9]. We employed raw counts as input to the model and
338 followed the same preprocessing steps described in the protocol of the backbone model

339 (scBERT) in all of our experiments. We also used the pre-computed pseudotime values
340 originally generated by Cook and Vanderhyden[9] as ground truth to train a regression model,
341 enabling it to predict a cell's position along this continuous spectrum.

342 In our training, we developed models using data from all tissue types, with specific models
343 tailored for different stimulation conditions. Based on our training strategy, each stimulation
344 type was associated with its own dedicated model, and we selected the best-performing models
345 for subsequent validation experiments.

346 **Learning discrete cell states and pseudotemporal ordering by using a single cell**
347 **foundation model with PEFT**

348 We create a two-stage dual path framework for EMP task. In training stage 1, we trained a
349 binary classifier for each EMP state, which served as an expert network in the following stages.
350 This proxy task was designed to force the model to focus on a generalisable EMP process
351 across different tissues. The expert network for each EMP state was adapted from scBERT[11]
352 with extra residual connections to enhance the raw gene expression signal and a Parameter-
353 Efficient Fine-Tuning (PEFT)[21] module. We named this expert network 'scMultiNet'. Its
354 purpose was to reshape the pre-trained feature manifold, redirecting the model's attention
355 towards more nuanced cellular state features relevant for EMP progression. EMP-FM involves
356 multiple scMultiNets for different cell states, combined through a fusion network with Mixture
357 of Expert (MoE)[22] architecture. We simultaneously train a classification head for cell states and
358 a regression head using the pseudotemporal ordering calculated by psupertime[9] in the training
359 set to capture the continuous nature of EMP (see Methods). This two-stage dual modelling
360 approach enables our model to predict continuous scores that reflect EMT progression while
361 preserving the integrity of the learned feature space.

362 Training stage II aims to train a fusion net to produce the final outcome. The gene expression
363 profile data was filtered as required by scBERT and encoded into 200-dimensional vector
364 spaces for each expert network (scMultiNet-E, scMultiNet-EM1, scMultiNet-EM2,
365 scMultiNet-EM3, or scMultiNet-M). The selected pre-trained foundational model (scBERT or
366 other potential single-cell foundation models in the future) is adept at analysing scRNA-seq
367 data, aiming to generalise cell state predictions across various tissues and patient cohorts. As a
368 pre-trained single-cell foundation model, the scBERT model underwent extensive pre-training
369 on a corpus of approximately 100,000 scRNA-seq samples with gene2vec and a vocabulary of

around 16,000 genes. The proposed method follows all the preprocessing steps in scBERT to make sure the proposed model can assign positional significance to genes.

We name our expert models 'scMultiNet' as we employ a simple multiplication layer consisting of fine-tuned foundation models. The proposed model considers the attention vector of the foundation model as a weight of raw gene expression profile data. We employed LoRA, a PEFT method, in our training stage I to adapt the embedding space foundation model's architecture for interpreting single-cell gene expression profiles (**Figure 1**). PEFT enhanced the model's adaptability to the targeted dataset while minimising the risk of overfitting. The design of scMultiNet ensured the model's focus remained on genes relevant to the EMT process, filtering out those with low expression or tangential relevance. Genes exhibiting minimal expression were removed through normalisation and selective filtering to mitigate their potential to skew the model's learning outcomes.

In all training stages, we utilised a one-dimensional convolution for dimensionality reduction. A three-layer neural network in training stage I is used to transform gene feature vectors into probabilistic cell type identities for binary classification proxy tasks. scMultiNet was trained with Adam optimizer (learning rate of 1e-4) and a cosine step scheduler.

For the target classification and regression tasks, we implemented a Mixture-of-Experts (MoE) architecture consisting of specialized expert mapper and a gating mechanism. Each expert mapper is structured as a two-layer perceptron with ReLU activation and dropout regularization. The input features are first processed through individual feature encoders, which project each feature group into a common representation space. A gating network, implemented as a multi-layer neural network, dynamically assigns input samples to the most relevant experts by computing importance weights. We employ a top-k gating mechanism with added noise during training to encourage diverse expert utilization. The expert outputs are then combined according to the gating weights, and the fused representations are processed through a final classifier/regression network comprising two fully connected layers with ReLU activation for the ultimate prediction tasks.

Model validation

Validation of the EMP-FM model's performance was carried out by randomly splitting the original data into training (80%) and testing (20%) sets, focusing on assessing the model's

accuracy. The model's feature space was visualised using Principal Component Analysis (PCA) with Scanpy and Potential of Heat-diffusion for Affinity-based Transition Embedding (PHATE) with its original implementation from the Krishnaswamy lab[23] to compare the changes before and after training, demonstrating that the training process effectively captured the key information related to EMT.

During training, we randomised the order of all sampling time points and discretised them into one-hot encodings. By not introducing temporal information during training, if the learned embedding space exhibited clear time-series associations, it would indicate that the model had learned the correct information (results shown in **Figure 2 and Supplementary Figure 2**).

To further validate the robustness and broad applicability of our approach, we tested our model on independent single-cell datasets where EMT is tracked longitudinally, with different experimental designs. These datasets are available for download from GSE194019, GSE114687, GSE155628 and GSE149299. The model's performance was analysed by comparing the categories provided by the datasets with the continuous scores inferred by the model.

In our analysis of the GSE155628 and GSE149299 datasets, we mapped mouse genes to human homologues using Ensembl BioMart. After data cleaning and integration with the gene annotation file, approximately 80% of mouse genes were successfully mapped to identifiable human homologues. For genes with multiple human orthologues, we used an averaging approach to consolidate expression values.

Comparison with traditional machine learning and bioinformatics methods

The evaluation process involved an initial dimensionality reduction via PCA across the first 10 PCs while retaining 95% of variance, followed by training with machine learning algorithms (KNN, random forest and AdaBoost from SciPy). We also conducted comprehensive comparisons against two categories of baseline methods. The first category includes state-of-the-art single cell foundation models: scGPT, scFoundation, and Universal Cell Embedding (UCE), which were fine-tuned on our training data. Conventional bioinformatics analysis methods (scType[29], AUCell[28] and scPred[27]) were also tested to demonstrate a full comparison with all potential state-of-the-art methods. For fair comparison, we implemented two distinct evaluation frameworks. On the Cook dataset, all methods were evaluated using identical train-

test splits, with performance measured by the average Area Under the Receiver Operating Characteristic curve (AUROC) for EMT state classification. For additional independent validation datasets where discrete states and temporal sampling designs differed from the training data, we adopted a continuous scoring approach, evaluating performance through Pearson Correlation Coefficient (PCC) between predicted EMT scores and experimental sampling timepoints. All methods share the same training set.

To identify differentially expressed genes (DEGs), we applied the FindAllMarkers function in Seurat. Marker genes were identified by comparing cell populations defined by the ground truth annotation on the Cook and Vanderhyden[9] full dataset. We focused on genes that were upregulated in each group by setting only.pos = TRUE. To ensure robust detection, we applied a minimum expression threshold (min.pct = 0.5), requiring that genes be expressed in at least 50% of cells in each group. Additionally, we set a low log-fold change threshold (logfc.threshold = 0.1) to capture subtle expression differences relevant to state transitions. Then, we applied AUCell scoring using DEGs for each category as one mega list, which performed better than using individual gene lists for every time point. Results were visualised as estimation plots[49] to demonstrate median differences from baseline category.

Attention-Driven Expression Significance Index (ADESI) scoring

We introduce the Attention-Driven Expression Significance Index (ADESI), a novel method designed to enhance the interpretability of complex biological datasets. Traditional single-cell foundation models use attention visualisation to demonstrate model interpretability and versatility. However, this may fail to capture changes between states that would be considered impactful enough to be biologically meaningful. ADESI addresses this issue by leveraging the EMP-FM architecture and Z-scores to provide biologically relevant gene weightings. Initially, mirroring traditional methods, we extracted attention scores from each expert model relevant to the designated category. These scores were then multiplied by the original expression levels of the respective genes to derive the EMF-FM output. The average score for each gene was calculated in the final classified category and the top K genes were prioritised. This design is used to aid researchers in understanding the genes that the model's attention mechanism focuses on, which could be potentially employed as markers of these states.

459 **Spatial transcriptomics analysis**

460 Endometrial carcinoma Visium slides were obtained from Cassier et al[39]. Filtered matrices
461 were loaded, merged per patient and spots with fewer than 1,000 detected genes were removed.
462 Pre-processing and normalisation were conducted using the Scanpy package. To more
463 precisely identify tumour cells we employed the copy number inference tool STARCH[50] and
464 only kept spots that showed evidence of copy number changes, which are likely to be tumour
465 specific. SpottedPy[51] was used to identify EMT niches and assess the spatial relationships.

466 **Gene signature analysis**

467 We used Gene Set Variation Analysis (GSVA) calculated using the GSEApy package[52] to score
468 the tumour samples with the EMP-FM gene signatures. We categorised the high attention genes
469 based on their upregulation or downregulation in the initial dataset for each time point. We
470 calculated a composite score by subtracting the GSVA score of downregulated genes from that
471 of the upregulated genes. The scaled score for comparison was calculated by subtracting the Z-
472 score of downregulated genes from the Z-score of the upregulated genes. Cohen's d score was
473 used for assessing how well the gene sets distinguish each cluster. Cohen's d was computed by
474 assessing the difference in means between the groups, standardised by their pooled variability.
475 G:profiler[53] was used for gene set enrichment analysis.

476 **Model building, statistical analysis and data visualisation**

477 The EMP-FM model was built with PyTorch for transformer blocks, and pytorch-lightning for
478 efficient training. The analyses of the results were conducted in Python and R. Groups were
479 compared using a two-sided Student's t-test, Wilcoxon rank-sum test or ANOVA as required.
480 Correlations were calculated using Spearman's rank order correlation. Graphs were generated
481 using the Seaborn and Matplotlib Python packages.

482

483 **DATA AVAILABILITY**

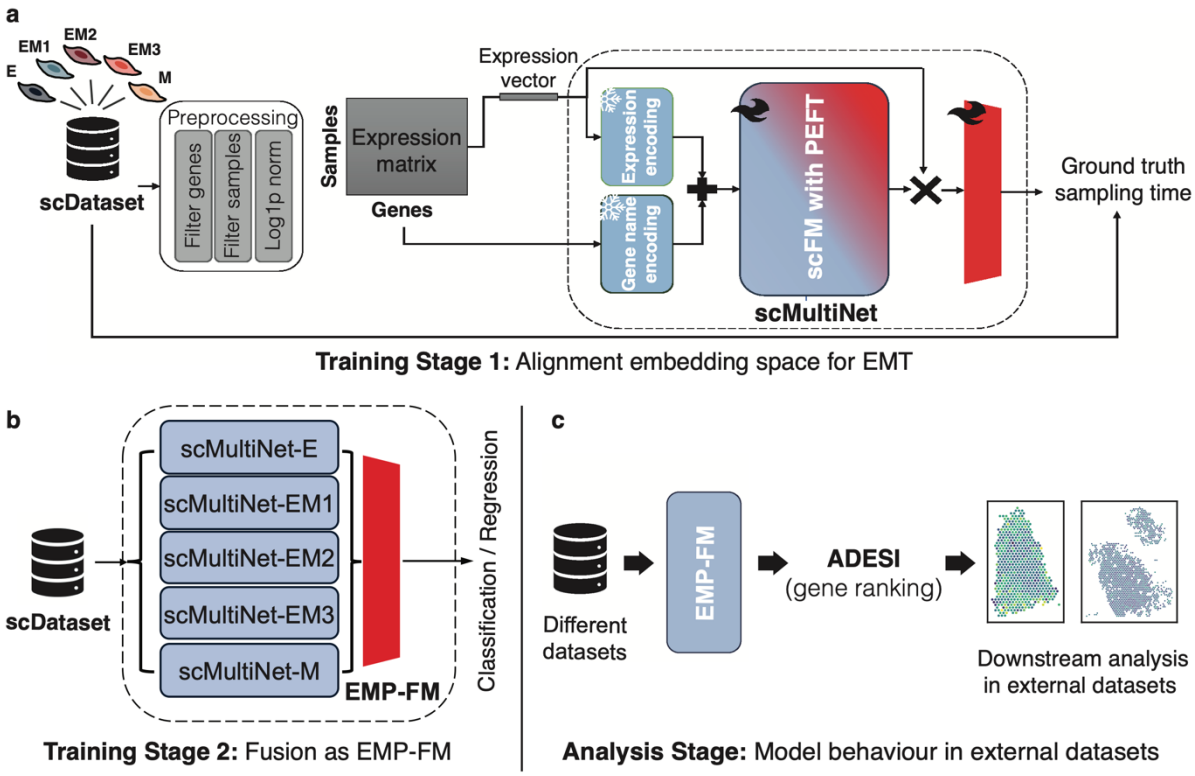
484 The results published here are based upon publicly available data made available at the Gene
485 Expression Omnibus (GEO) by Cook and Vanderhyden[9] (GSE147405), Paul et al[35]
486 (GSE194019) and McFaline-Figueroa et al[36] (GSE114687). The endometrial cancer spatial

transcriptomic slides were downloaded from Cassier et al[39] (GSE225691). Single cell RNA-seq data of the adult mouse mammary epithelium from control and Zeb1 MSKO mice were downloaded from Han et al[38] (GSE155628). Single-cell RNA sequencing data from C3(1)-Tag cancer cells tracked during invasion and colony formation were obtained from Grasset et al[37] (GSE149299).

CODE AVAILABILITY

The EMP-FM package can be found at the following GitHub repository: https://github.com/secrierlab/EMP-FM. The implementation of the basic module of EMP-FM, namely scMultiNet, can be accessed within this repository at: EMP-FM/scLLM/Models/scMultiNet. All the code developed for the purpose of this analysis is deposited at: EMP-FM/Experiment.

499 **FIGURES**



500

501 **Figure 1: EMP-FM pipeline in scRNA-seq data**. **(a)** The first training phase involves binary
502 classification training for each EMT time point using the PEFT method. This allows the
503 original pre-trained model's feature space to better represent EMT. **(b)** The final model is
504 obtained by fusing the five binary classification models for the individual states (E, EM1-3,
505 M). **(c)** Downstream analysis employs ADESI to rank genes that contribute significantly to the
506 model based on attention scores and gene expression.

507

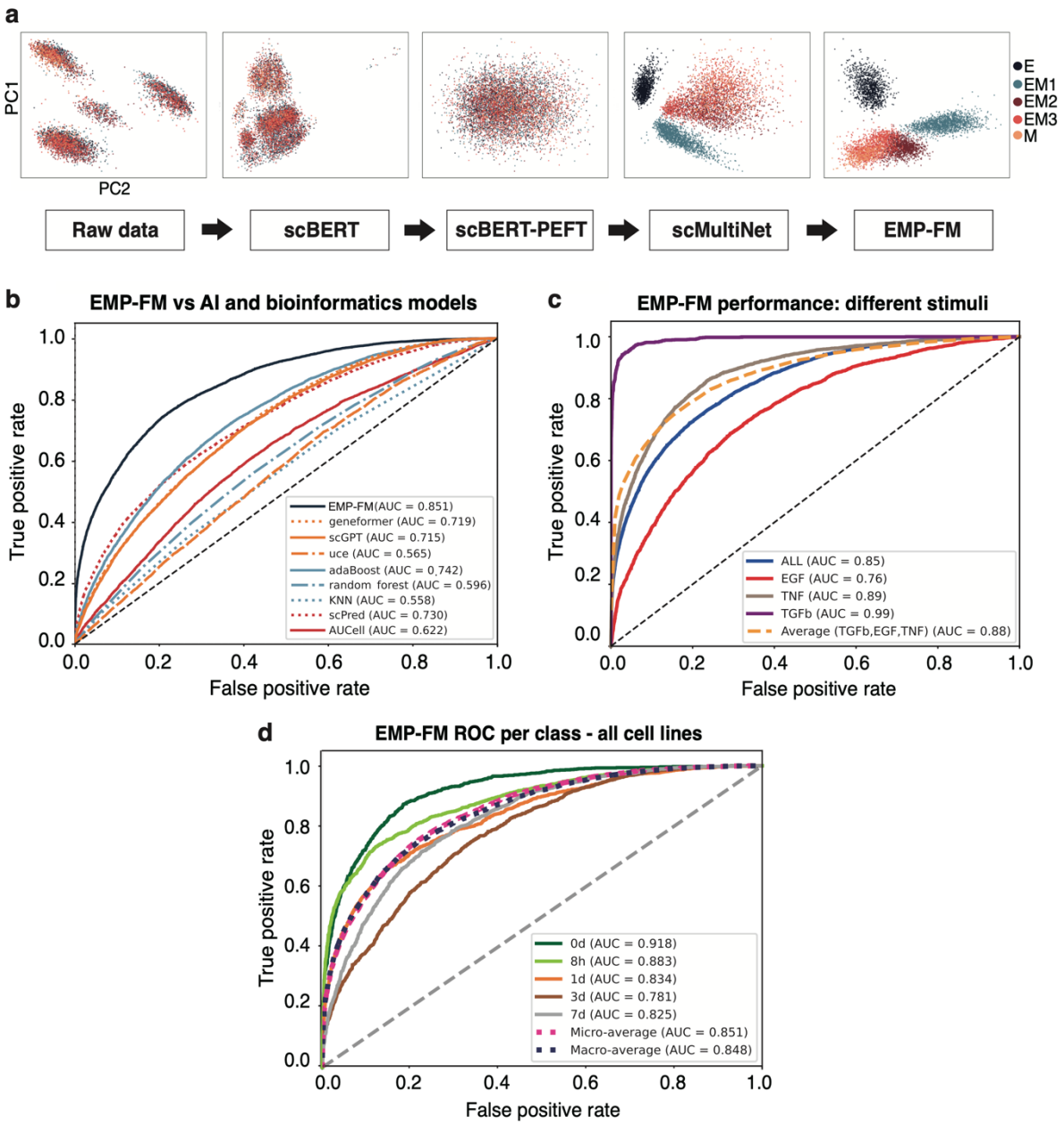


Figure 2: EMP-FM embedding space visualisation and benchmarking. (a) The changes in embedding space are shown through PCA projections from various models, highlighting the transformation of raw data into refined clusters following the application of scBERT, PEFT, scMultiNet and EMP-FM. Every point corresponds to a single cell that has been profiled. The five colours indicate the ground truth EMP states. **(b)** Comparison of EMP-FM with state-of-the-art foundation models including scGPT, Geneformer and UCE, conventional machine learning and bioinformatics methods. **(c)** EMP-FM performance under different EMT-inducing stimuli. ROC curves are shown for models of TGFβ-, TNF- and EGF-induced EMT, alongside an all-in-one model (ALL), with the average performance across the three stimulation types displayed as dotted curve. **(d)** EMP-FM-ALL performance across different timepoints and EMT states: epithelial (E), early EMT (EM1), intermediate EMT (EM2), late EMT (EM3), and mesenchymal (M).

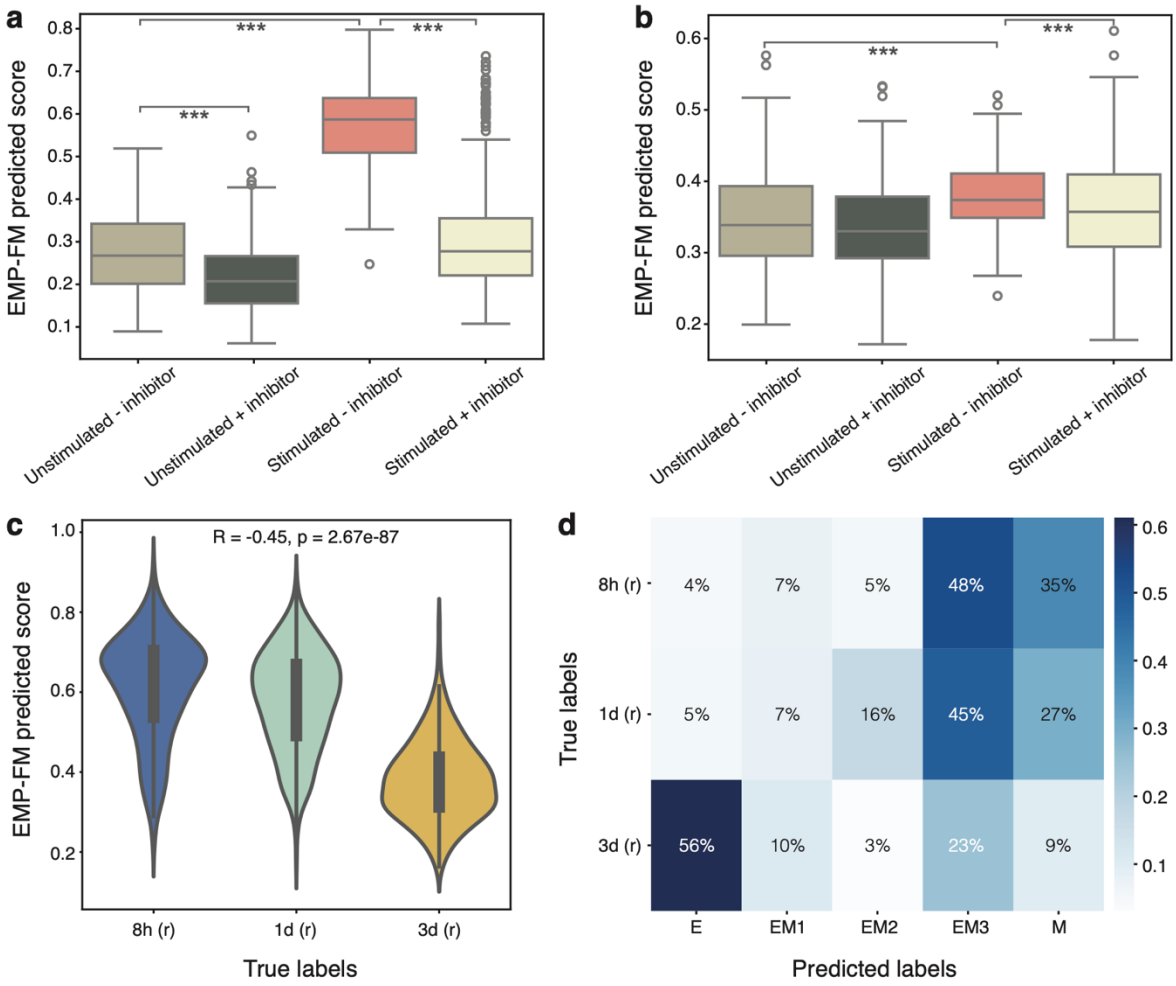


Figure 3: Model validation in EMT inhibitor and stimulus removal data in the original study. (a-b) Predicted EMP-FM-TGFβ scores in A548 (a) and MCF7 (b) cell lines that are either unstimulated or stimulated with TGFβ, and either treated (+inhibitor) and untreated (-inhibitor) with a TGFBR1 inhibitor. (*** p-value <0.001, ** <0.01, *<0.1). EMT scores are significantly lower after the addition of the inhibitor and in the unstimulated versus stimulated condition, as expected. **(c)** EMP-FM-TGFβ regression score distribution for the TGFβ stimulus removal data across all cancer cell lines, sequenced at 8h, 1 day and 3 days post-removal, respectively. The scores decrease with increasing time after removal, as expected. The Spearman correlation coefficient (R) and p-value are reported. **(d)** Confusion matrix of predicted cell states for the cells with TGFβ removal (averaged across all cancer cell lines) from the EMP-FM-TGFβ model.

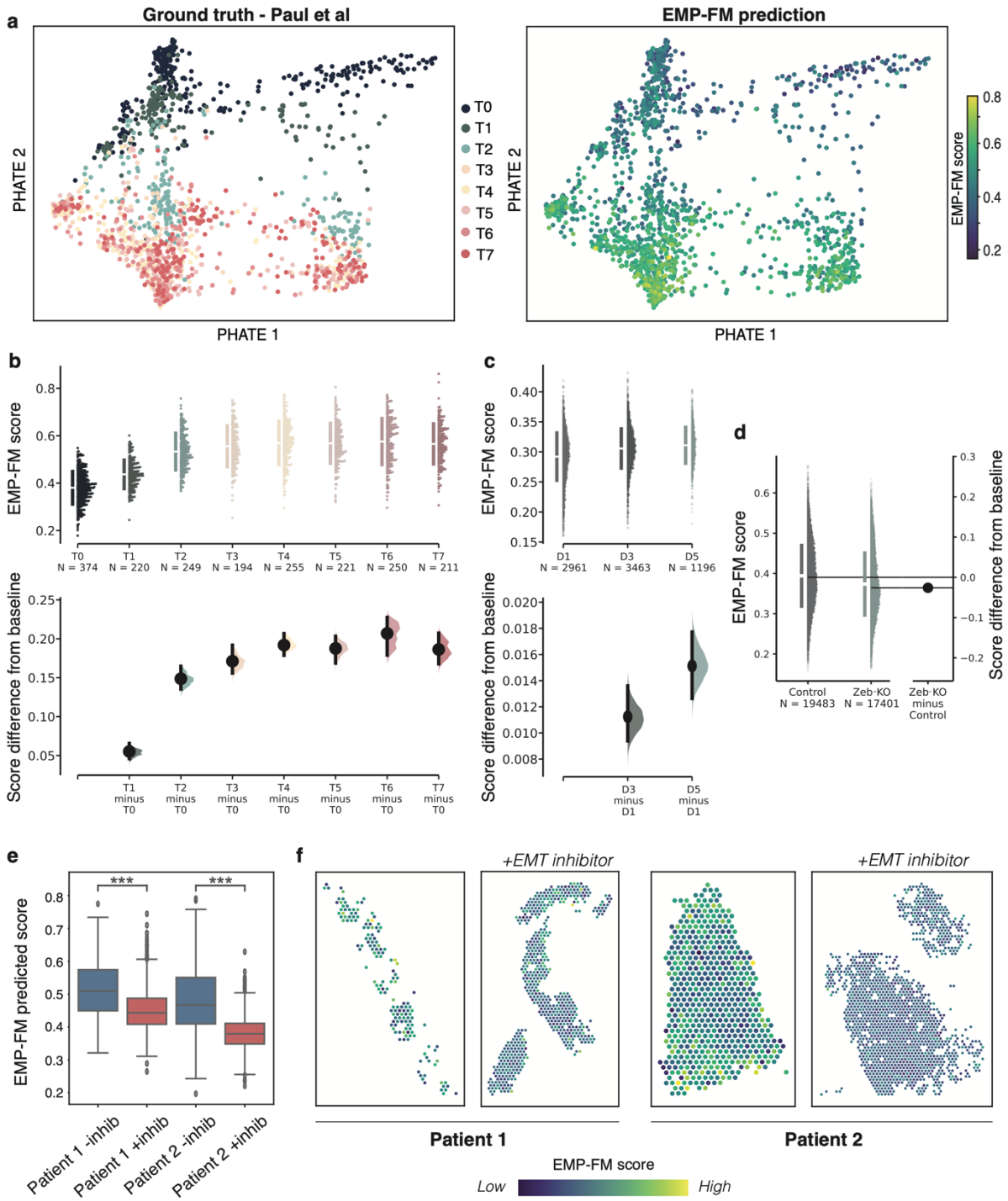
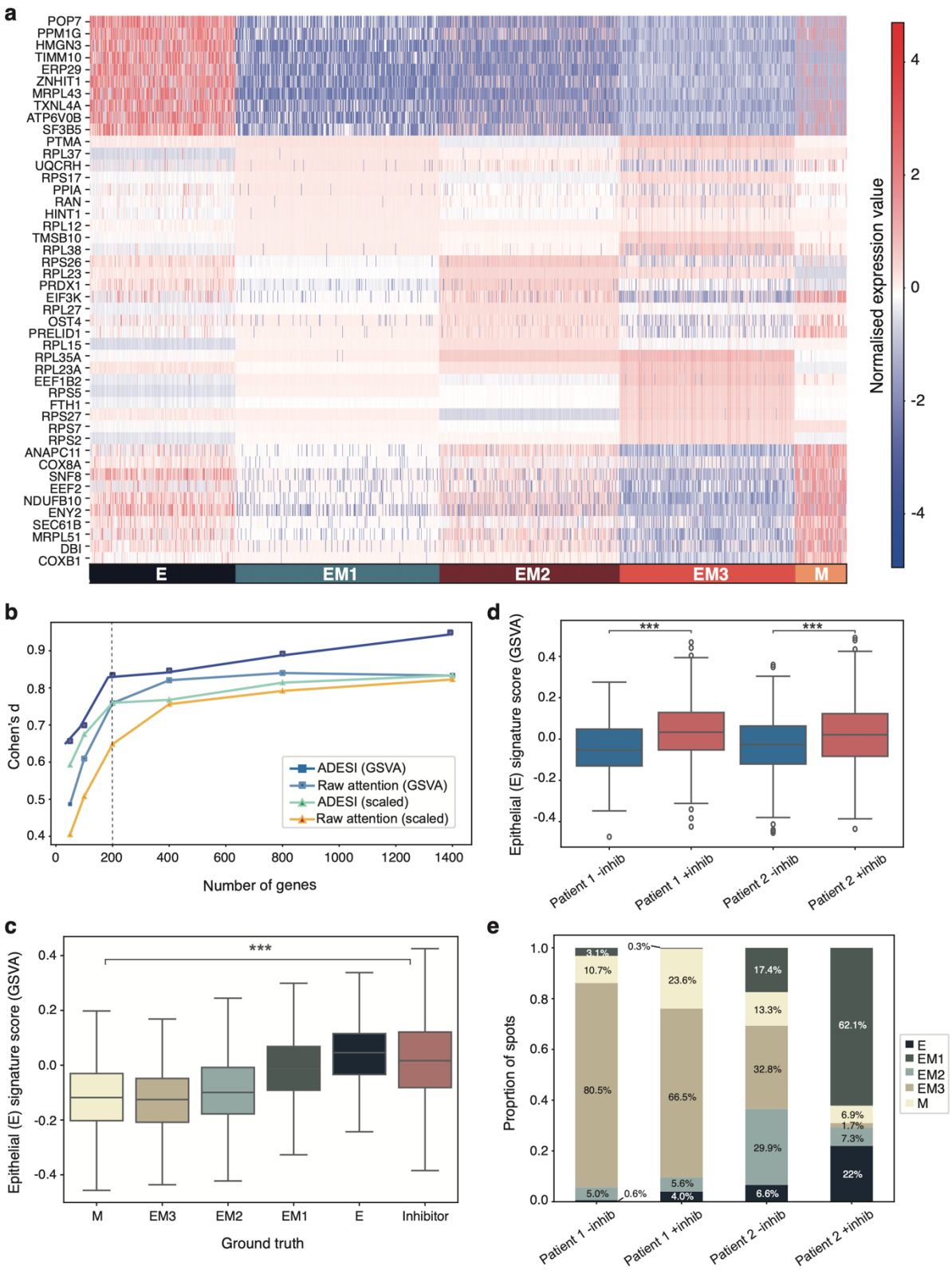


Figure 4: Validation of the EMP-FM model in external datasets. (a-b) Model validation in MCF10A cancer cell lines undergoing EMT (Paul et al[35]): **(a)** PHATE visualisation of individual cell projections coloured by ground truth sampling time (left panel, time points measured T1-T7) and scores predicted by EMP-FM (right panel, higher values indicate a more mesenchymal state); **(b)** Distribution of EMP-FM scores predicted for individual cells at the 7 time points (top) and differences in scores compared to the T0 baseline (bottom). Increases at all time points are significant (Kruskal-Wallis test $p<<0.001$, see statistics in Supplementary Table 4). **(c)** Model validation in C3(1)-Tag mouse breast cancer cells tracked during hybrid

544 E/M transition (Grasset et al[37]): Distribution of EMP-FM scores predicted for individual cells
545 at day 0 (12h), day 3 and day 5 (top) and differences in scores compared to the day 0 baseline
546 (bottom). Increases at both time points are significant (Kruskal-Wallis test, see statistics in
547 Supplementary Table 4). **(d)** Model validation in mouse basal mammary epithelial cells tracked
548 upon Zeb1 knockout (Han et al[38]): Distribution of EMP-FM scores predicted for individual
549 cells before and after Zeb1 knockout (left) and score difference compared to control (right).
550 Increases after knockout is significant (Kruskal-Wallis test, see statistics in Supplementary
551 Table 4). **(e-f)** Model validation in an external EMT inhibitor (Netrin-1) dataset from Cassier
552 et al[39]: **(e)** Score distribution across endometrial spatial transcriptomic slides before (-inhib)
553 and after (+inhib) treatment with the EMT inhibitor; **(f)** Visualisation of EMP-FM predicted
554 regression scores across the spatially profiled slides before and after EMT inhibition in Patient
555 1 (left) and Patient 2 (right).



556

557 **Figure 5: EMP gene signature inference and sensitivity analysis. (a)** Expression heat map
558 of key genes selected by the model using ADESI scoring across different categories, with a
559 gradient from blue to red indicating increasing expression. The colour bar at the bottom
560 distinguishes the EMP state for each profiled sample, offering an intuitive representation of
561 which genes are important in predicting various EMP states. **(b)** Sensitivity analysis plot

562 highlighting the effect of increasing the gene list size and alternating both methods to rank
563 genes (raw attention vs. ADESI) and methods to score the gene list (scaled versus GSVA) on
564 distinguishing the EMP states in the Cook and Vanderhyden[9] data. The dashed line highlights
565 the elbow point. **(d)** Validation of the epithelial gene signature in the EMT inhibitor (Netrin-1)
566 dataset from Cassier et al[39] (*** $p<0.001$). An increase in the epithelial signature is seen in
567 both patients after the treatment with the inhibitor (+inhib).

568

SUPPLEMENTARY TABLE CAPTIONS

Supplementary Table 1. Performance comparison of different state-of-the-art models for EMT prediction tasks on the datasets from Cook et al. The table presents AUROC scores for various models across three stimulation conditions (TGFβ, EGF, and TNF) and their averaged performance. Models are categorized into three groups: our proposed EMP-FM models with different parameter sizes (512 and 128), state-of-the-art foundation models (SOTA-FM including scGPT, Geneformer, and UCE), and conventional machine learning approaches (random forest, KNN, and AdaBoost). Higher scores indicate better predictive performance.

Supplementary Table 2. Performance of the three EMP-FM models (columns) in predicting EMT states across the different stimulation datasets (listed as rows).

Supplementary Table 3. Performance of the three EMP-FM models in predicting each EMT state (timepoint) in each cancer type.

Supplementary Table 4. Kruskal-Wallis test statistics showcasing group differences for individual figure panels where EMP-FM regression scores as well as AUCell-generated scores based on DEGs and classical EMT markers are compared in validation datasets.

Supplementary Table 5. Performance of EMP-FM, scGPT and random forest regression models in the external validation datasets. The top performing model in each dataset is highlighted in red. PCC - Pearson correlation coefficient**.**

Supplementary Table 6. The top 200 up-/down-regulated genes in each EMT state, as uncovered by the ADESI metric and robustness analysis.

Supplementary Table 7. Pathway enrichment analysis results for the EMT gene signatures. **(a)** The 0d (E) gene signature. **(b)** The 8h (EM1) gene signature. **(c)** The 1d (EM2) gene signature. **(d)** The 3d (EM3) gene signature. **(e)** The 7d (M) gene signature.

AUTHOR CONTRIBUTIONS

MS and SP designed the experiment. MS supervised the analyses. SP developed and tested the EMP-FM model and the ADESI score. EW prepared the training and validation datasets, performed the gene sensitivity analysis, pathway enrichment and spatial transcriptomics analyses. CC evaluated EMT in validation datasets using classic EMT markers and bioinformatics methods. All authors wrote and approved the manuscript.

ACKNOWLEDGEMENTS

MS, SP and CC were supported by a UKRI Future Leaders Fellowship (MR/T042184/1, MR/Y034031/1). EW was supported by a studentship award from the Health Data Research UK-The Alan Turing Institute Wellcome PhD Programme in Health Data Science (218529/Z/19/Z). Work in MS's lab was supported by a BBSRC equipment grant (BB/R01356X/1) and a Wellcome Institutional Strategic Support Fund (204841/Z/16/Z).

The illustrations in Supplementary Figure 1 were generated using the icon libraries available at https://bioicons.com/.

ETHICS DECLARATIONS

This study employs only publicly available data. All data comply with ethical regulations, with approval and informed consent for collection and sharing already obtained by the relevant consortia.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

1 Dongre, A. & Weinberg, R. A. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Biol* **20**, 69-84 (2019). https://doi.org/10.1038/s41580-018-0080-4

2 Lu, W. & Kang, Y. Epithelial-Mesenchymal Plasticity in Cancer Progression and Metastasis. *Dev Cell* **49**, 361-374 (2019). https://doi.org/10.1016/j.devcel.2019.04.010

3 Sinha, D., Saha, P., Samanta, A. & Bishayee, A. Emerging Concepts of Hybrid Epithelial-to-Mesenchymal Transition in Cancer Progression. *Biomolecules* **10** (2020). https://doi.org/10.3390/biom10111561

4 Malagoli Tagliazucchi, G., Wiecek, A. J., Withnell, E. & Secrier, M. Genomic and microenvironmental heterogeneity shaping epithelial-to-mesenchymal trajectories in cancer. *Nat Commun* **14**, 789 (2023). https://doi.org/10.1038/s41467-023-36439-7

5 Brown, M. S. *et al.* Phenotypic heterogeneity driven by plasticity of the intermediate EMT state governs disease progression and metastasis in breast cancer. *Sci Adv* **8**, eabj8002 (2022). https://doi.org/10.1126/sciadv.abj8002

6 Kroger, C. *et al.* 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). https://doi.org/10.1073/pnas.1812876116

7 Chakraborty, P., George, J. T., Tripathi, S., Levine, H. & Jolly, M. K. Comparative Study of Transcriptomics-Based Scoring Metrics for the Epithelial-Hybrid-Mesenchymal Spectrum. *Front Bioeng Biotechnol* **8**, 220 (2020). https://doi.org/10.3389/fbioe.2020.00220

8 Vasaikar, S. V. *et al.* EMTome: a resource for pan-cancer analysis of epithelial-mesenchymal transition genes and signatures. *Br J Cancer* **124**, 259-269 (2021). https://doi.org/10.1038/s41416-020-01178-9

9 Cook, D. P. & Vanderhyden, B. C. Context specificity of the EMT transcriptional response. *Nat Commun* **11**, 2142 (2020). https://doi.org/10.1038/s41467-020-16066-2

10 Celia-Terrassa, T. *et al.* Hysteresis control of epithelial-mesenchymal transition dynamics conveys a distinct program with enhanced metastatic ability. *Nat Commun* **9**, 5005 (2018). https://doi.org/10.1038/s41467-018-07538-7

11 Yang, F. *et al.* scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. *Nat Mach Intell* **4**, 852-+ (2022). https://doi.org/10.1038/s42256-022-00534-z

12 Cui, H. *et al.* scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat Methods* (2024). https://doi.org/10.1038/s41592-024-02201-0

13 Levine, D. *et al.* Cell2Sentence: Teaching Large Language Models the Language of Biology. *bioRxiv*, 2023.2009.2011.557287 (2024). https://doi.org/10.1101/2023.09.11.557287

14 Chen, Y. & Zou, J. GenePT: A Simple But Effective Foundation Model for Genes and Cells Built From ChatGPT. *bioRxiv* (2024). https://doi.org/10.1101/2023.10.16.562533

15 Hao, M. *et al.* Large-scale foundation model on single-cell transcriptomics. *Nat Methods* **21**, 1481-1491 (2024). https://doi.org/10.1038/s41592-024-02305-7

16 Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 (2018). <https://ui.adsabs.harvard.edu/abs/2018arXiv181004805D>.

17 Liu, X. *et al.* GPT understands, too. *AI Open* (2023). https://doi.org/https://doi.org/10.1016/j.aiopen.2023.08.012

18 Gu, A. & Dao, T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. arXiv:2312.00752 (2023). <https://ui.adsabs.harvard.edu/abs/2023arXiv231200752G>.

19 Szalata, A. *et al.* Transformers in single-cell omics: a review and new perspectives. *Nat Methods* **21**, 1430-1443 (2024). https://doi.org/10.1038/s41592-024-02353-z

20 Kedzierska, K. Z., Crawford, L., Amini, A. P. & Lu, A. X. Zero-shot evaluation reveals limitations of single-cell foundation models. *Genome Biol* **26**, 101 (2025). https://doi.org/10.1186/s13059-025-03574-x

21 Ding, N. *et al.* Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nat Mach Intell* **5**, 220-235 (2023). https://doi.org/10.1038/s42256-023-00626-4

22 Cai, W. *et al.* A survey on mixture of experts. *arXiv preprint arXiv:2407.06204* (2024).

23 Moon, K. R. *et al.* Visualizing structure and transitions in high-dimensional biological data. *Nat Biotechnol* **37**, 1482-1492 (2019). https://doi.org/10.1038/s41587-019-0336-3

24 Haghverdi, L., Buttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods* **13**, 845-848 (2016). https://doi.org/10.1038/nmeth.3971

25 Cui, Z., Liao, Y., Xu, T. & Wang, Y. Geneformer: Learned gene compression using transformer-based context modeling. *arXiv preprint arXiv:2212.08379* (2022).

26 Rosen, Y. *et al.* Universal Cell Embeddings: A Foundation Model for Cell Biology. *bioRxiv*, 2023.2011.2028.568918 (2023). https://doi.org/10.1101/2023.11.28.568918

27 Alquicira-Hernandez, J., Sathe, A., Ji, H. P., Nguyen, Q. & Powell, J. E. scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. *Genome Biol* **20**, 264 (2019). https://doi.org/10.1186/s13059-019-1862-5

28 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086 (2017). https://doi.org/10.1038/nmeth.4463

29 Ianevski, A., Giri, A. K. & Aittokallio, T. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nat Commun* **13**, 1246 (2022). https://doi.org/10.1038/s41467-022-28803-w

30 Hao, Y., Baker, D. & Ten Dijke, P. TGF-beta-Mediated Epithelial-Mesenchymal Transition and Cancer Metastasis. *Int J Mol Sci* **20** (2019). https://doi.org/10.3390/ijms20112767

31 Katsuno, Y., Lamouille, S. & Derynck, R. TGF-beta signaling and epithelial-mesenchymal transition in cancer progression. *Curr Opin Oncol* **25**, 76-84 (2013). https://doi.org/10.1097/CCO.0b013e32835b6371

32 Horiguchi, K. *et al.* TGF-beta drives epithelial-mesenchymal transition through deltaEF1-mediated downregulation of ESRP. *Oncogene* **31**, 3190-3201 (2012). https://doi.org/10.1038/onc.2011.493

33 Ramirez, D., Kohar, V. & Lu, M. Toward Modeling Context-Specific EMT Regulatory Networks Using Temporal Single Cell RNA-Seq Data. *Front Mol Biosci* **7**, 54 (2020). https://doi.org/10.3389/fmolb.2020.00054

34 Li, C. W. *et al.* Epithelial-mesenchymal transition induced by TNF-alpha requires NF-kappaB-mediated transcriptional upregulation of Twist1. *Cancer Res* **72**, 1290-1300 (2012). https://doi.org/10.1158/0008-5472.CAN-11-3123

35 Paul, I. *et al.* Parallelized multidimensional analytic framework applied to mammary epithelial cells uncovers regulatory principles in EMT. *Nat Commun* **14**, 688 (2023). https://doi.org/10.1038/s41467-023-36122-x

36 McFaline-Figueroa, J. L. *et al.* A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. *Nat Genet* **51**, 1389-1398 (2019). https://doi.org/10.1038/s41588-019-0489-5

37 Grasset, E. M. *et al.* Triple-negative breast cancer metastasis involves complex epithelial-mesenchymal transition dynamics and requires vimentin. *Sci Transl Med* **14**, eabn7571 (2022). https://doi.org/10.1126/scitranslmed.abn7571

38 Han, Y. *et al.* Coordinate control of basal epithelial cell fate and stem cell maintenance by core EMT transcription factor Zeb1. *Cell Rep* **38**, 110240 (2022). https://doi.org/10.1016/j.celrep.2021.110240

39 Cassier, P. A. *et al.* Netrin-1 blockade inhibits tumour growth and EMT features in endometrial cancer. *Nature* **620**, 409-416 (2023). https://doi.org/10.1038/s41586-023-06367-z

40 Guerra, F. *et al.* Mitochondrial Dysfunction: A Novel Potential Driver of Epithelial-to-Mesenchymal Transition in Cancer. *Front Oncol* **7**, 295 (2017). https://doi.org/10.3389/fonc.2017.00295

41 Zheng, S. *et al.* Role and mechanism of actin-related protein 2/3 complex signaling in cancer invasion and metastasis: A review. *Medicine (Baltimore)* **102**, e33158 (2023). https://doi.org/10.1097/MD.0000000000033158

42 Xu, Z. *et al.* Integrin beta1 is a critical effector in promoting metastasis and chemo-resistance of esophageal squamous cell carcinoma. *Am J Cancer Res* **7**, 531-542 (2017).

43 Yang, J. *et al.* Guidelines and definitions for research on epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* **21**, 341-352 (2020). https://doi.org/10.1038/s41580-020-0237-9

44 Li, D. *et al.* Heterogeneity and plasticity of epithelial-mesenchymal transition (EMT) in cancer metastasis: Focusing on partial EMT and regulatory mechanisms. *Cell Prolif* **56**, e13423 (2023). https://doi.org/10.1111/cpr.13423

45 Jolly, M. K. *et al.* Stability of the hybrid epithelial/mesenchymal phenotype. *Oncotarget* **7**, 27067-27084 (2016). https://doi.org/10.18632/oncotarget.8166

46 Bakir, B., Chiarella, A. M., Pitarresi, J. R. & Rustgi, A. K. EMT, MET, Plasticity, and Tumor Metastasis. *Trends Cell Biol* **30**, 764-776 (2020). https://doi.org/10.1016/j.tcb.2020.07.003

47 Aiello, N. M. *et al.* EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. *Dev Cell* **45**, 681-695 e684 (2018). https://doi.org/10.1016/j.devcel.2018.05.027

48 Sha, Y. *et al.* Intermediate cell states in epithelial-to-mesenchymal transition. *Phys Biol* **16**, 021001 (2019). https://doi.org/10.1088/1478-3975/aaf928

49 Ho, J., Tumkaya, T., Aryal, S., Choi, H. & Claridge-Chang, A. Moving beyond P values: data analysis with estimation graphics. *Nat Methods* **16**, 565-566 (2019). https://doi.org/10.1038/s41592-019-0470-3

50 Elyanow, R., Zeira, R., Land, M. & Raphael, B. J. STARCH: copy number and clone inference from spatial transcriptomics data. *Phys Biol* **18**, 035001 (2021). https://doi.org/10.1088/1478-3975/abbe99

51 Withnell, E. & Secrier, M. SpottedPy quantifies relationships between spatial transcriptomic hotspots and uncovers environmental cues of epithelial-mesenchymal plasticity in breast cancer. *Genome Biol* **25**, 289 (2024). https://doi.org/10.1186/s13059-024-03428-y

52 Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39** (2023). https://doi.org/10.1093/bioinformatics/btac757

760 53 Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler--a web-based
761 toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids*
762 *Res* **35**, W193-200 (2007). https://doi.org/10.1093/nar/gkm226

763