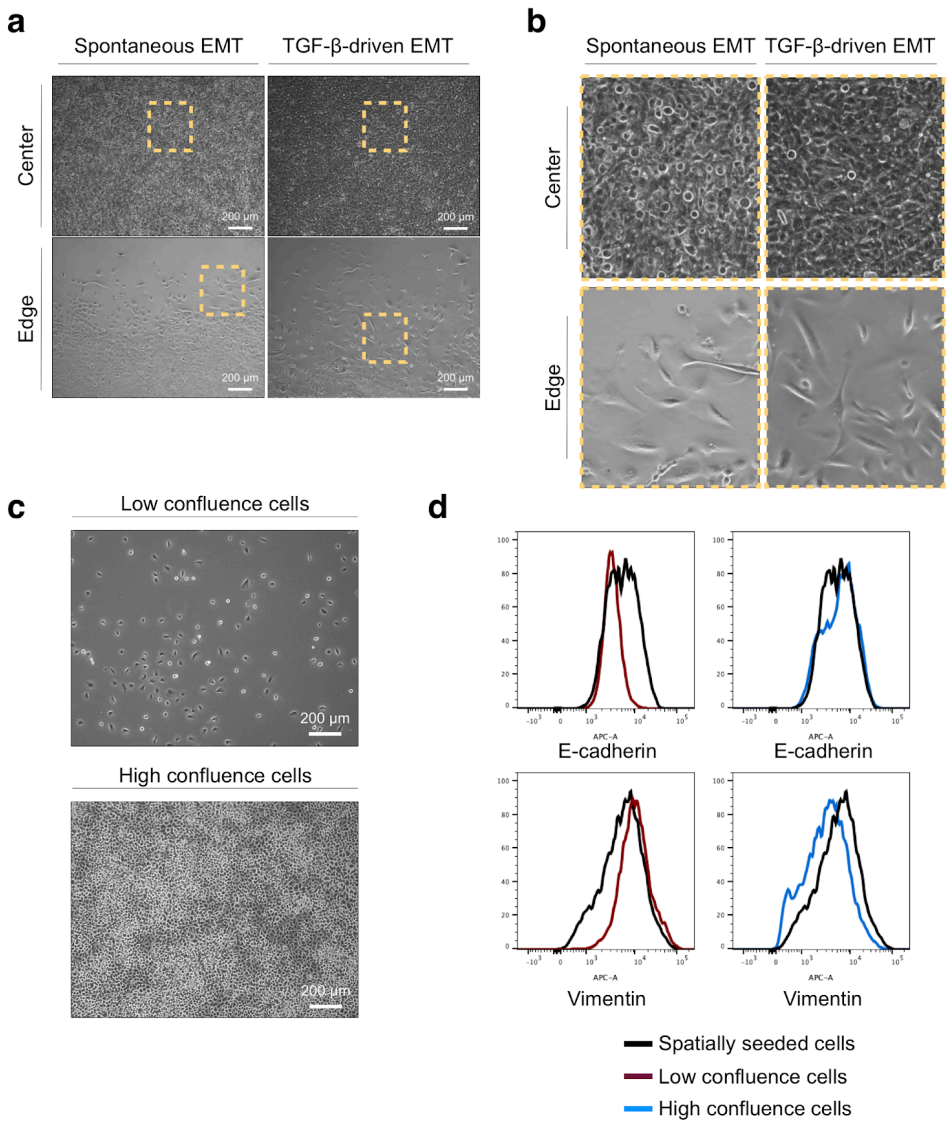
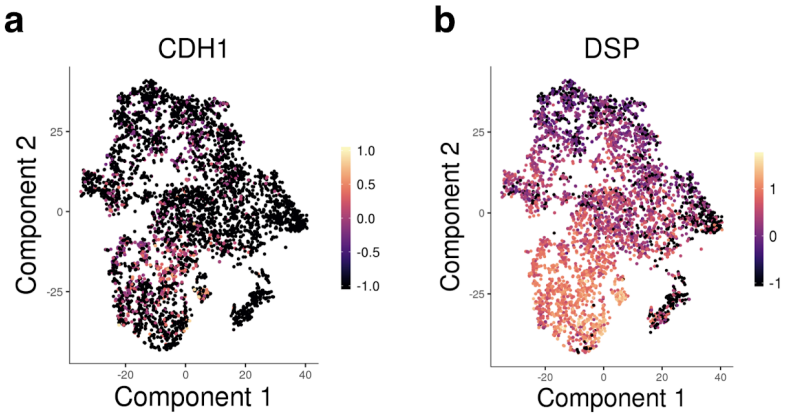


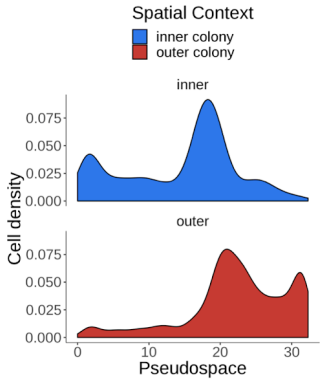
Supplementary Figure 1: Cells at the edge of seeded colonies acquire morphological characteristics of mesenchymal cells

a) Brightfield images of MCF10A cells at the center and edge of spontaneous or TGF-β exposed colonies (200 µm scale bar, representative fields from 2 independent experiments and 6 images collected at similar coordinates between conditions). **b)** Zoom-in of images as outlined in yellow boxes from **a**. **c)** Brightfield images on MCF10A cells cultured at low and high confluence confluence for 7 days (200 µm scale bar, representative fields from 2 independent experiments and 4 images collected at similar areas of the plate between conditions). **d)** Histograms of e-cadherin (top panel) and vimentin (bottom panels) protein levels in MCF10A cells seeded according to our spatial assay as well as MCF10A cells passaged every 96 hours and maintained at low (left panels) or high (right panels) confluence.



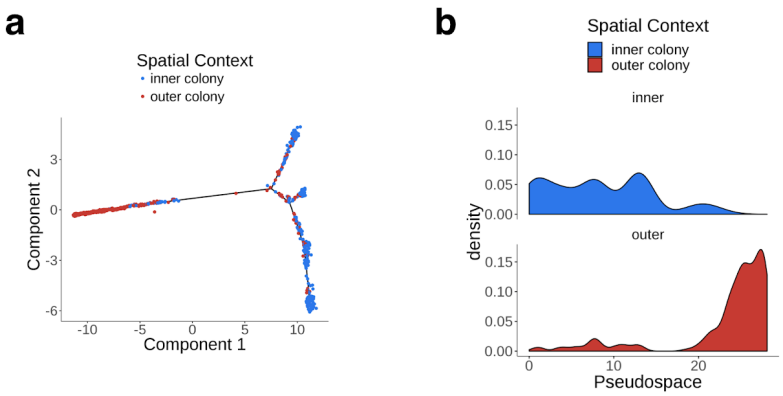
Supplementary Figure 2: Expression of epithelial markers suggest that EMT is a continuum

a-b) tSNE embedding of cells from our spontaneous EMT assay colored by CDH1 **(a)** or DSP **(b)** expression



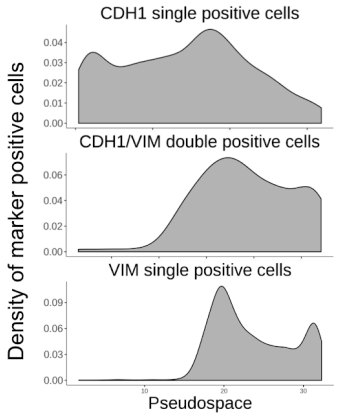
Supplementary Figure 3: Cells from inner and outer cell fractions are differentially distributed across pseudospace

Distribution of inner (top) and outer (bottom) cells across our spontaneous EMT pseudospatial trajectory.



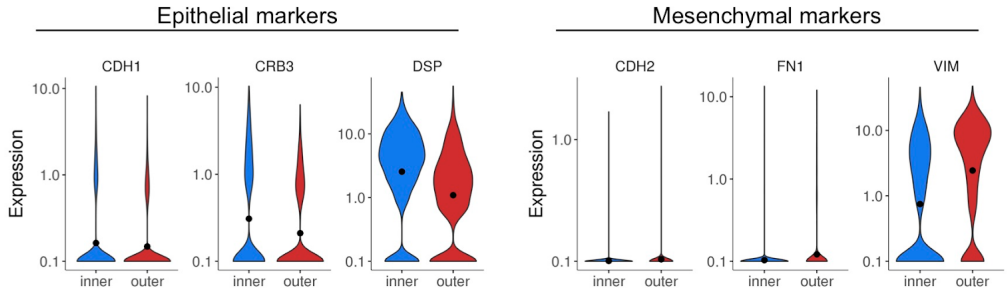
Supplementary Figure 4: Monocle does not identify a continuously populated trajectory for the subset of cells at the endpoints of spontaneous EMT

Although our spontaneous trajectory supports the view of a continuous EMT process we sought to ensure that it is not a result of Monocle arbitrarily connecting cells from a discrete EMT process. We selected cells in the first and last decile across pseudospace to simulate sampling from two distinct biological states and repeated our trajectory reconstruction with the expectation that a trajectory rom our subsetted cells should not be continuously populated. (**a**) Trajectory of subsetted inner and outer MCF10A cells upon spontaneous EMT progression. Note that the resulting trajectory from our subsetted cells contains a large gap in its center with inner and outer colony cells clustered at each end. (**b**) Distribution of inner (top) and outer (bottom) cells across our spontaneous EMT pseudospatial trajectory of subsetted cells from (**a**).



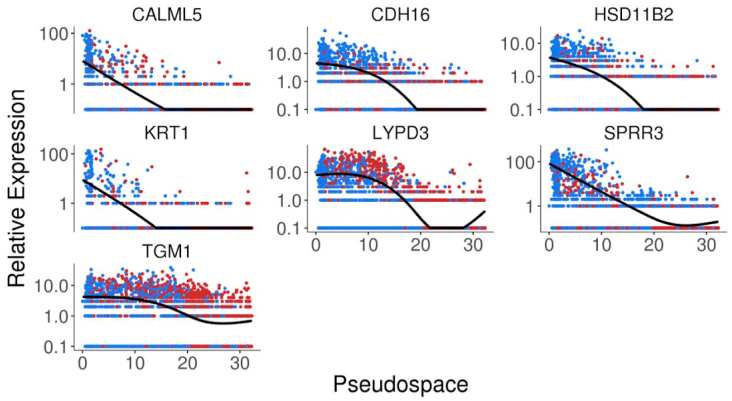
Supplementary Figure 5: Distribution of classically defined intermediate EMT states in our spontaneous EMT trajectory

Distribution of CDH1, VIM and CDH1/VIM double positive cells across pseudospace. Pseudospace was divided into 5 quantiles, cells expressing CDH1 and VIM over their respective mean expression levels were identified and the fraction of cells positive for CDH1, VIM or both were plotted.



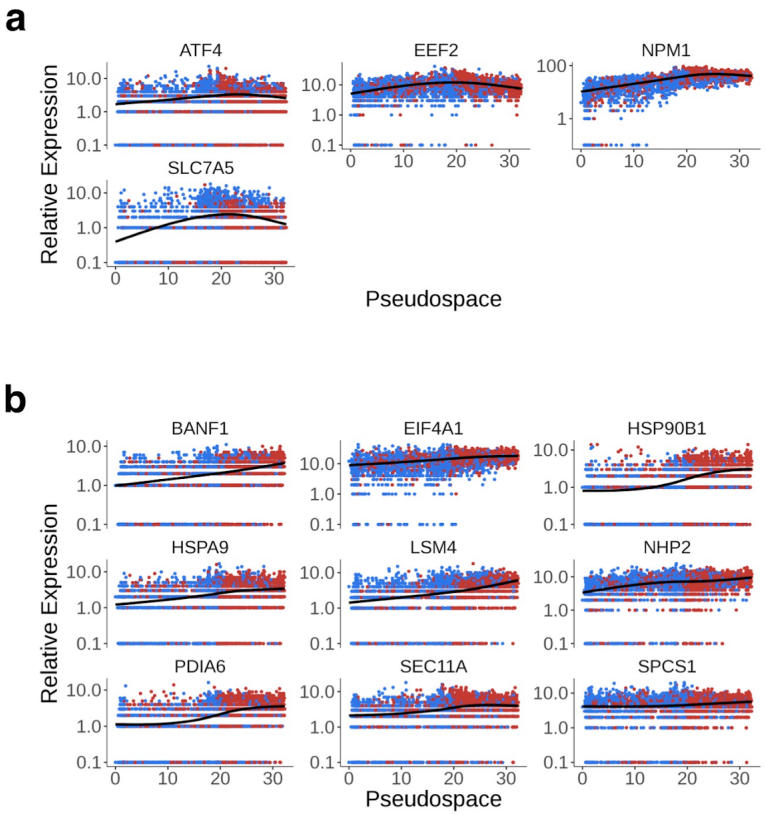
Supplementary Figure 6: Bulk averages of epithelial marker expression fail to capture the dynamics of spontaneous confluence-mediated EMT

Gene expression levels of epithelial (left) and mesenchymal (right) markers across averaged inner (blue, 2,440 cells) and outer (red, 2,564 cells) colony cells. Point within the violin depicts the mean expression level for each group with violin spanning the minimum and maximum expression value across cells.



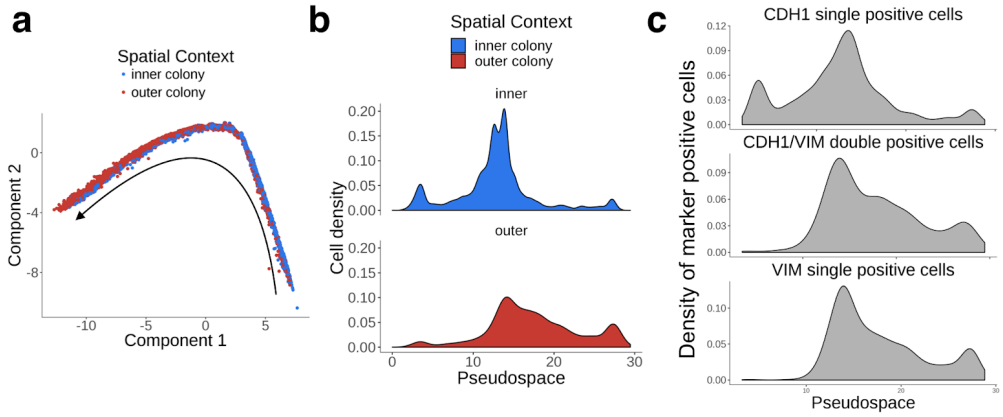
Supplementary Figure 7: Expression kinetics of genes repressed during spontaneous EMT and serve as a signature for KRAS activation associated with an epithelial exit

Color denotes whether cells were isolated from the interior (blue) or exterior (red) of the colony.



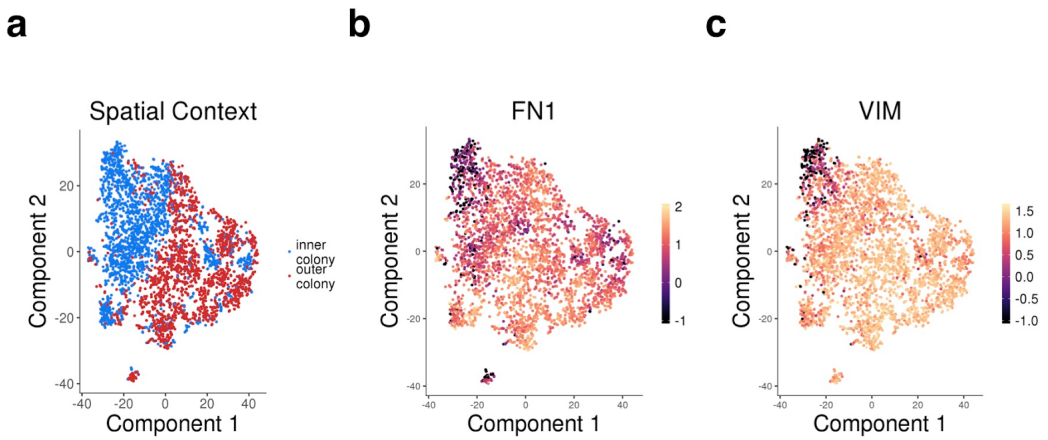
Supplementary Figure 8: Expression kinetics of unfolded protein response genes activated during spontaneous EMT

a) Unfolded protein response genes, including the UPR transcriptional regulator ATF4, from figure 1g cluster 6 that are activated towards the end of pseudospace with similarly to genes associated with an induction of an EMT. **b)** Unfolded protein response genes from figure 1g, cluster 5 that are activated towards the end of pseudospace. Color denotes whether cells were isolated from the interior (blue) or exterior (red) of the colony.



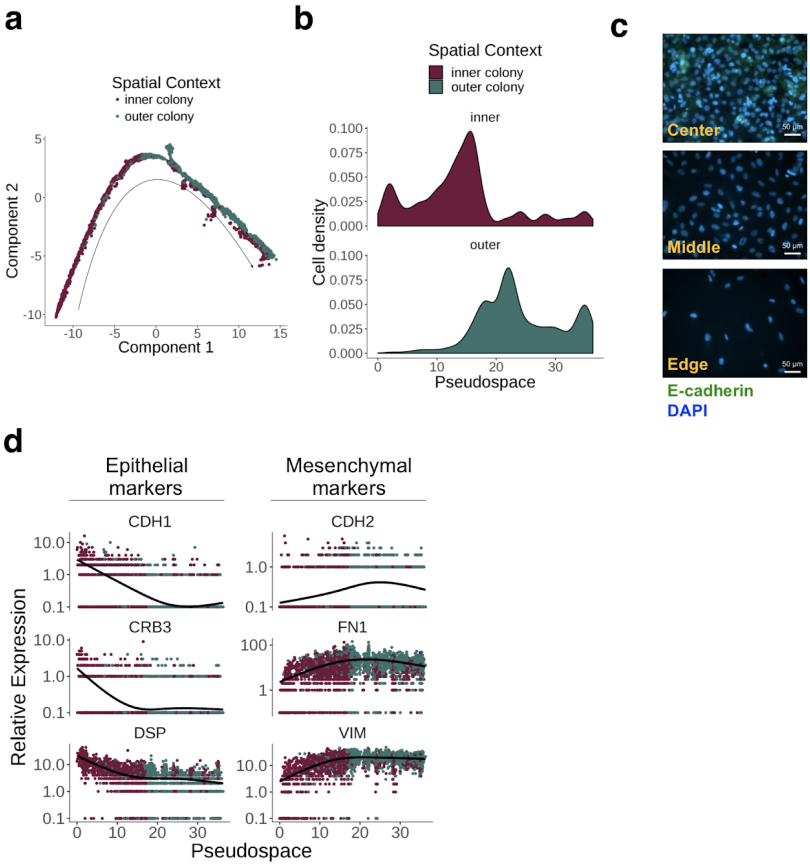
Supplementary Figure 9: Pseudospatial trajectory reconstruction of primary human mammary epithelial (HuMEC) cells undergoing spontaneous EMT

a) Trajectory of inner and outer HuMEC cells upon spontaneous EMT progression. Arrow denotes the progression of pseudoposition along the tree. **b)** Distribution of inner (top) and outer (bottom) cells across the pseudospatial trajectory. **c)** Distribution of CDH1, VIM and CDH1/VIM double positive cells across pseudospace. Pseudospace was divided into 5 quantiles, cells expressing CDH1 and VIM over their respective mean expression levels were identified and the fraction of cells positive for CDH1, VIM or both were plotted.



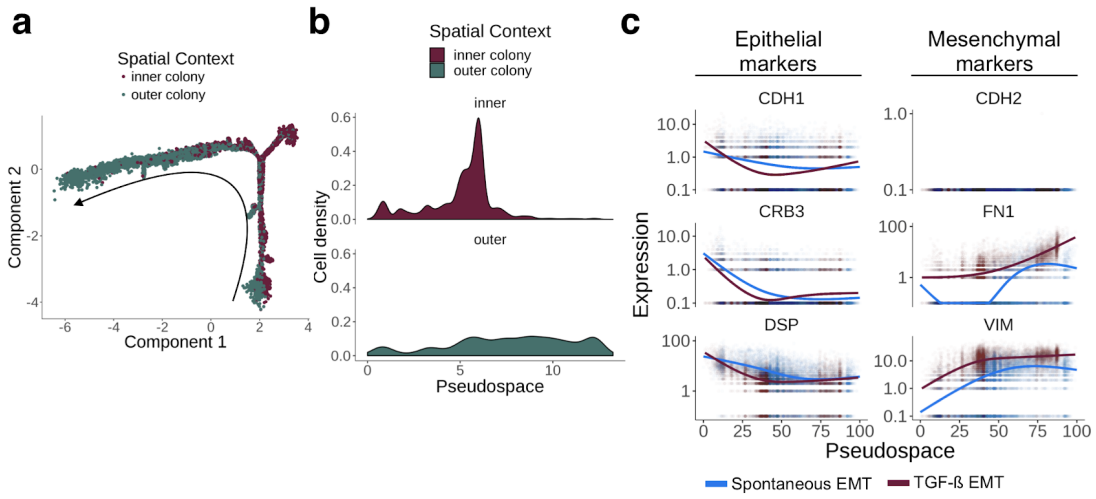
Supplementary Figure 10: Expression of mesenchymal markers across cells from our TGF-ß-driven EMT trajectory

a-c) tSNE embedding of cells from our TGF-ß-driven EMT assay. Cells are colored by the fraction from which they were isolated or by FN1 **(b)** or VIM **(c)** expression.



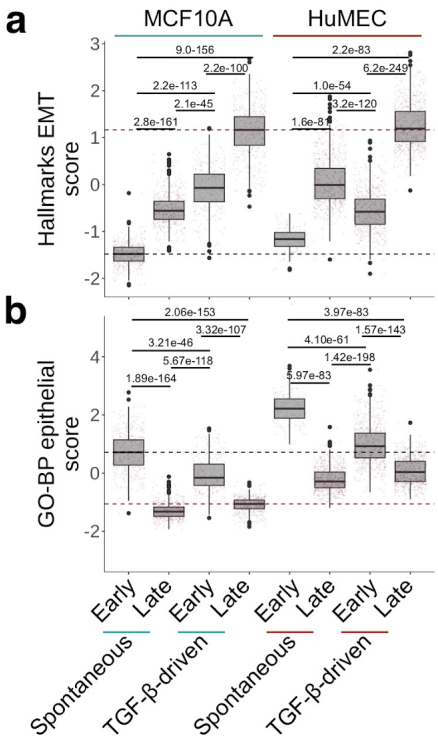
Supplementary Figure 11: Pseudospatial trajectory reconstruction of MCF10A cells undergoing TGF-ß-driven EMT

a) Trajectory of inner and outer MCF10A cells upon confluence and TGF-ß induced EMT progression. Arrow denotes the progression of pseudoposition along the tree. b) Distribution of inner (top) and outer (bottom) cells across the pseudopositional trajectory. c) E-cadherin and DAPI staining of cells from the center, middle and edge of TGF-ß exposed MCF10A colony (50 µM scale bar, representative fields from 6 images collected across 3 independent samples). d) Expression of select epithelial and mesenchymal markers across pseudospace. Cells are colored by the fraction from which they were isolated as in b.



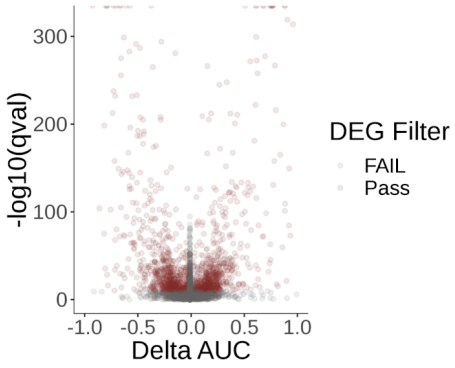
Supplementary Figure 12: Pseudospatial trajectory reconstruction of HuMEC cells undergoing TGF-ß-driven EMT

a) Trajectory of inner and outer HuMEC cells upon TGF-ß induced EMT progression. Arrow denotes the progression of pseudoposition along the tree. b) Distribution of inner (top) and outer (bottom) cells across the pseudopositional trajectory. c) Expression of select epithelial and mesenchymal markers across warped pseudospace with cells colored by treatment condition.

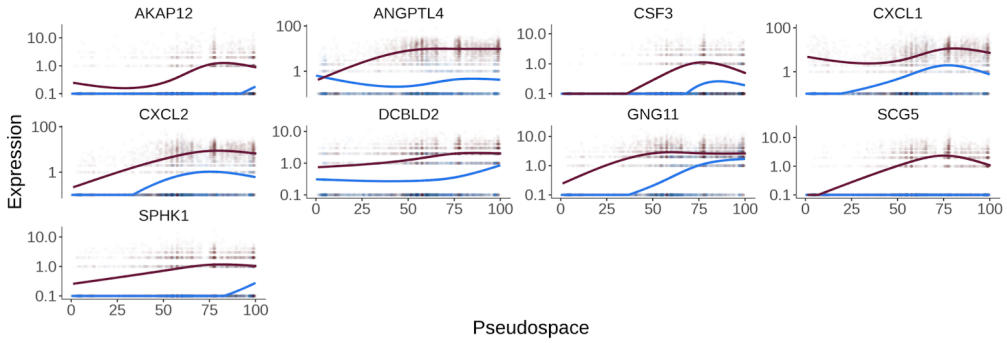


Supplementary Figure 13: EMT and epithelial scores of MCF10A and HuMEC cells at the endpoints of spontaneous and TGF-ß-driven EMT trajectories

a-b) Normalized aggregate expression scores of Hallmarks EMT (**a**) and GO-BP epithelial (**b**) genes of MCF10A and HuMEC cells at early and late positions in pseudospatial trajectories (MCF10A: spontaneous EMT early = 462 cells, spontaneous EMT late = 558 cells, TGF-ß-driven EMT early = 286 cells, TGF-ß-driven EMT late = 486 cells; HuMEC: spontaneous EMT early = 149 cells, spontaneous EMT late = 777 cells, TGF-ß-driven EMT early = 733 cells, TGF-ß-driven EMT late = 800 cells). Boxplots depict the median score (bold line within box) with lower and upper hinges depicting the 25th and 75th percentiles, respectively. Two-tailed wilcoxon rank test.

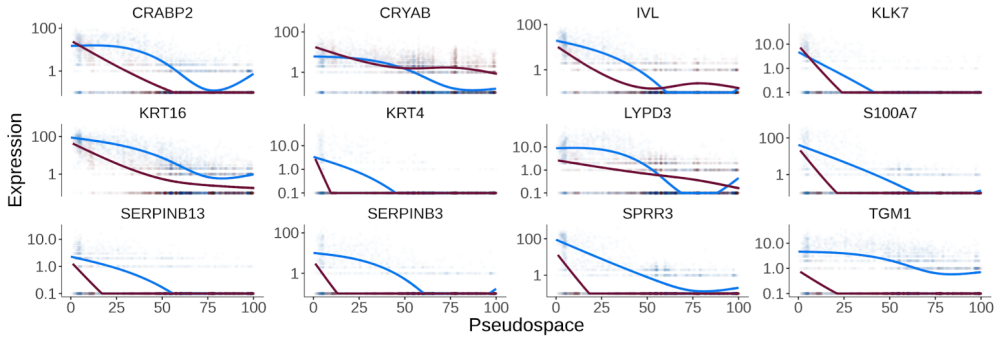


Supplementary Figure 14: The difference in AUC between our aligned spontaneous and TGF-ß-driven EMT trajectories follows the expected significance-magnitude relationship Volcano plot depicting the relationship between statistical significance and effect size as measured by our ΔAUC metric. Each dot represents an individual gene (12,606 genes). Genes are colored by whether they pass or fail our filtering criteria (likelihood ratio test, FDR q < $1x10^{-10}$ & |ΔAUC| > 0.02).



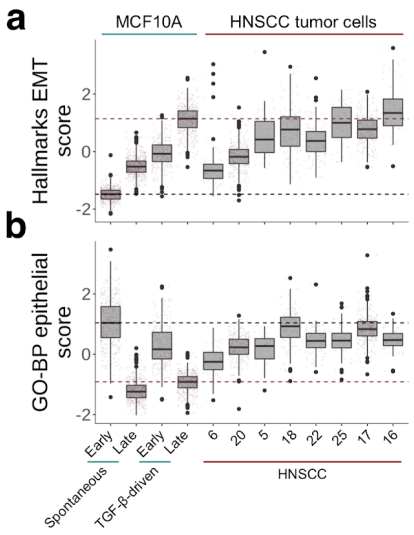
Supplementary Figure 15: Expression kinetics of genes activated during late stages of TGF-ß-driven EMT and serve as a signature for KRAS activation and acquisition of a mature mesenchymal state

Color denotes whether cells were come from spontaneous (blue) or TGF-ß-driven (maroon) EMT samples.



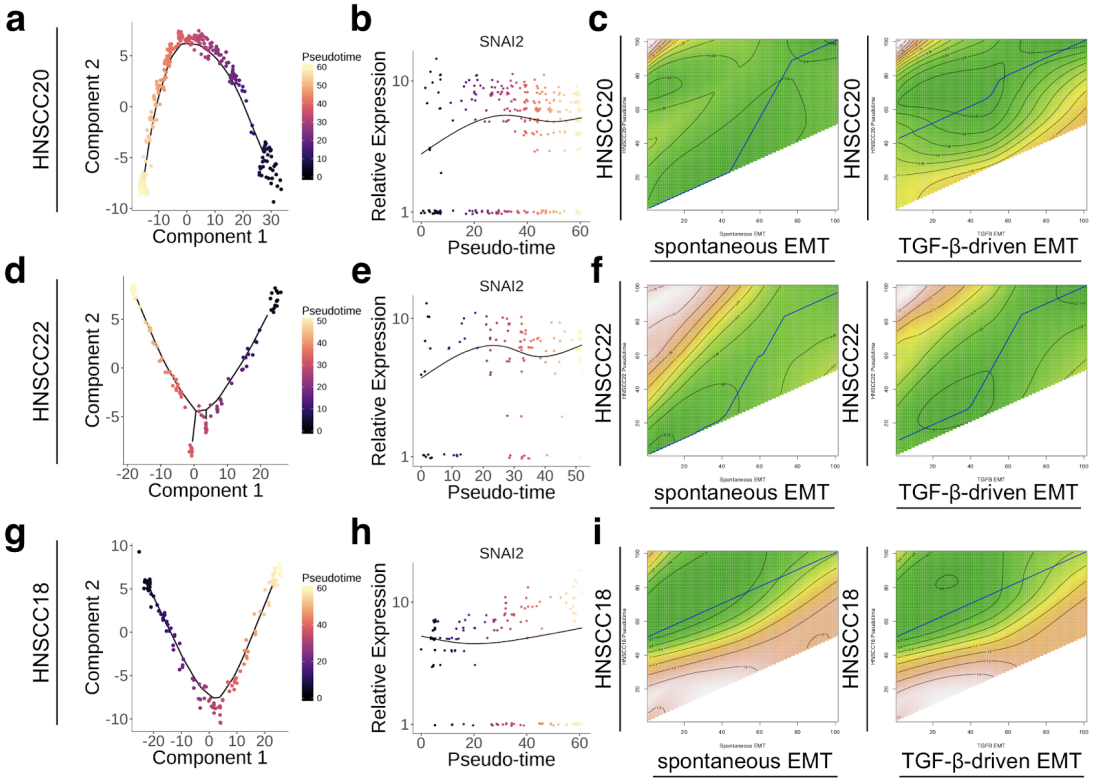
Supplementary Figure 16: Expression kinetics of genes repressed during early stages of spontaneous EMT and are low throughout TGF-ß-driven EMT and serve as a signature for KRAS mediated epithelial exit

Color denotes whether cells were come from spontaneous (blue) or TGF-ß (maroon) driven EMT condition.



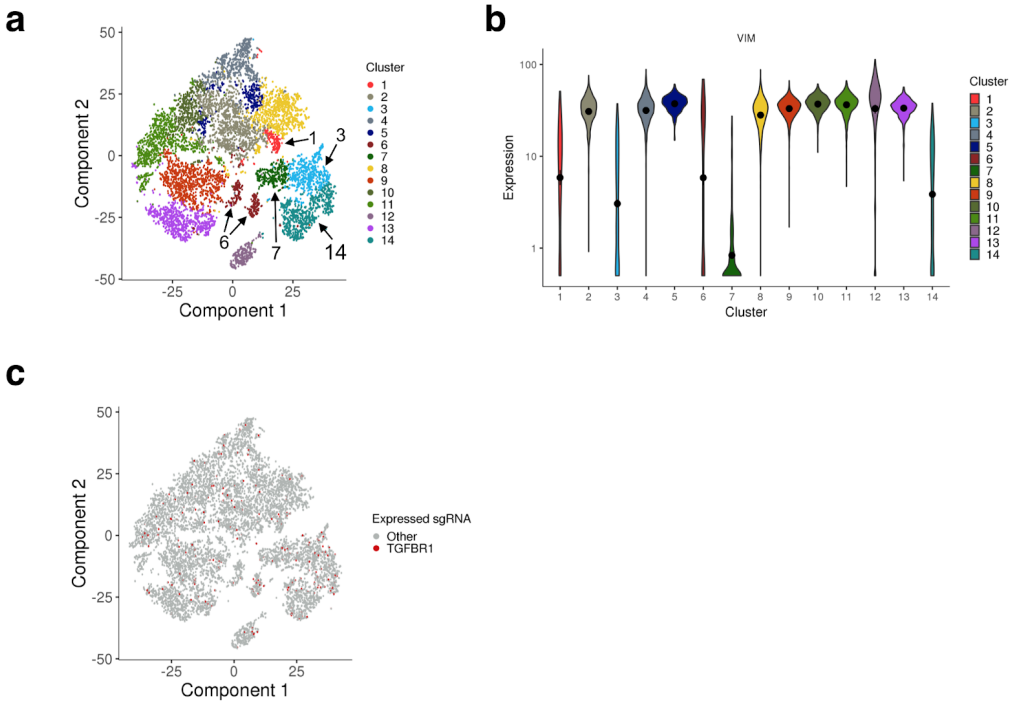
Supplementary Figure 17: EMT and epithelial scores of MCF10A cells at the endpoints of spontaneous and TGF-ß-driven EMT trajectories and cells from HNSCC tumors

a-b) Normalized aggregate expression scores of Hallmarks EMT (**a**) and GO-BP epithelial (**b**) genes of MCF10A at early and late positions in pseudospatial trajectories (Mock = 1,020 cells, TGF-ß = 772 cells) and HNSCC tumor cells tumors (6 = 80 cells, 20 = 321 cells, 5 = 41 cells, 18 = 140 cells, 22 = 119 cells, 25 = 54 cells, 17 = 330 cells, 16 = 56 cells). Boxplots depict the median score (bold line within box) with lower and upper hinges depicting the 25th and 75th percentiles, respectively.



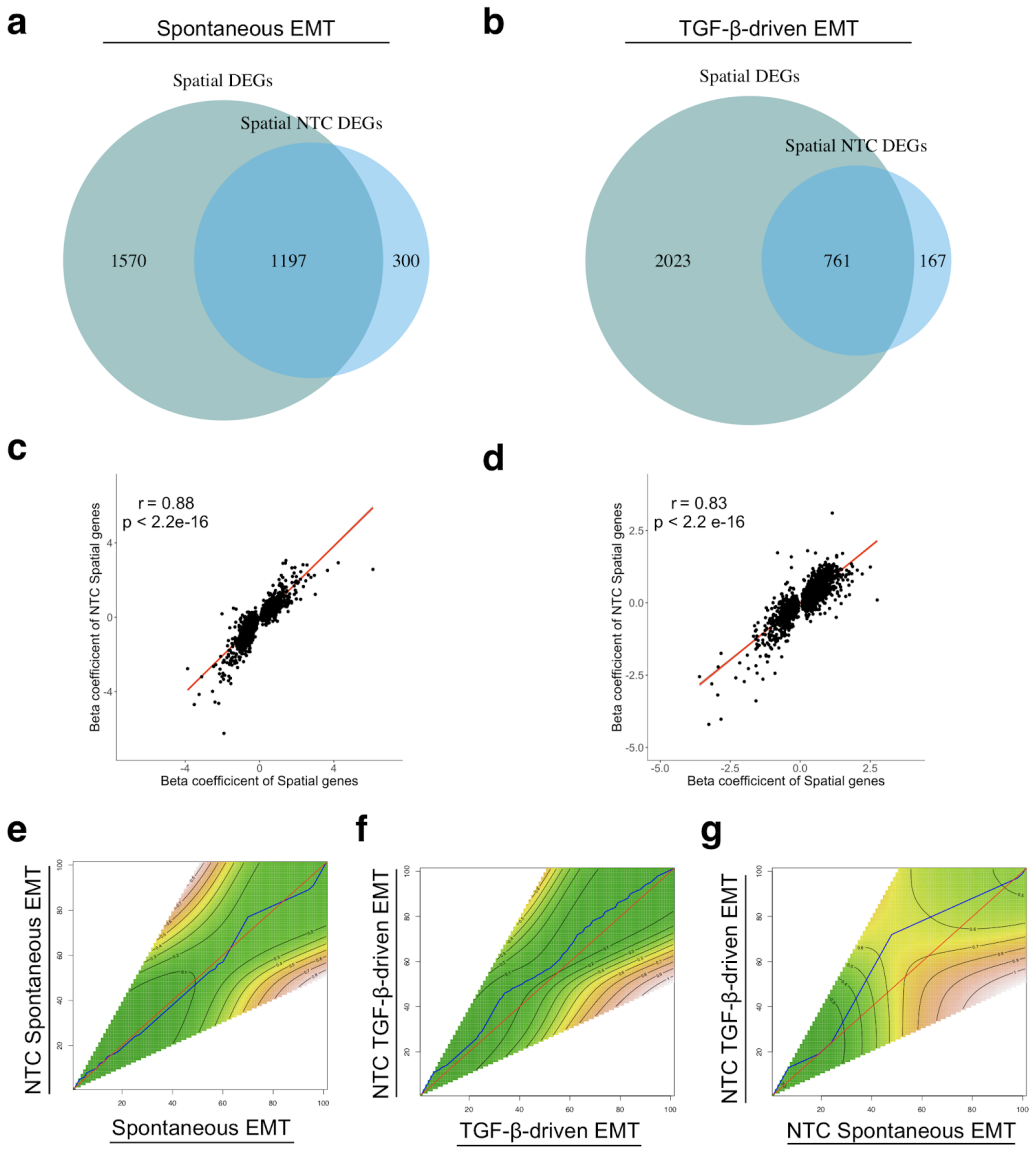
Supplementary Figure 18: Dynamic time warping of HNSCC tumor cell pseudotime and MCF10A spontaneous and TGF-ß-driven EMT trajectories

a/d/g) Pseudo-temporal trajectories of cells from HNSCC tumors 20 (**a**), 22 (**d**) and 18 (**g**). **b/e/h**) Expression of SNAI2 along pseudo-time for cells from from HNSCC tumors 20 (**b**), 22 (**e**) and 18 (**h**). The beginning of pseudotime was set to the end of the tree with the lowest levels of SNAI2. **c/f/i**) Alignment plots of dynamic time warped trajectories illustrates the warping of HNSCC tumor cell trajectories to spontaneous (left panel) or TGF-ß-driven EMT trajectories.



Supplementary Figure 19: tSNE embedding and clustering analysis of pooled loss of function screen identifies clusters that fail to EMT upon TGF-ß exposure

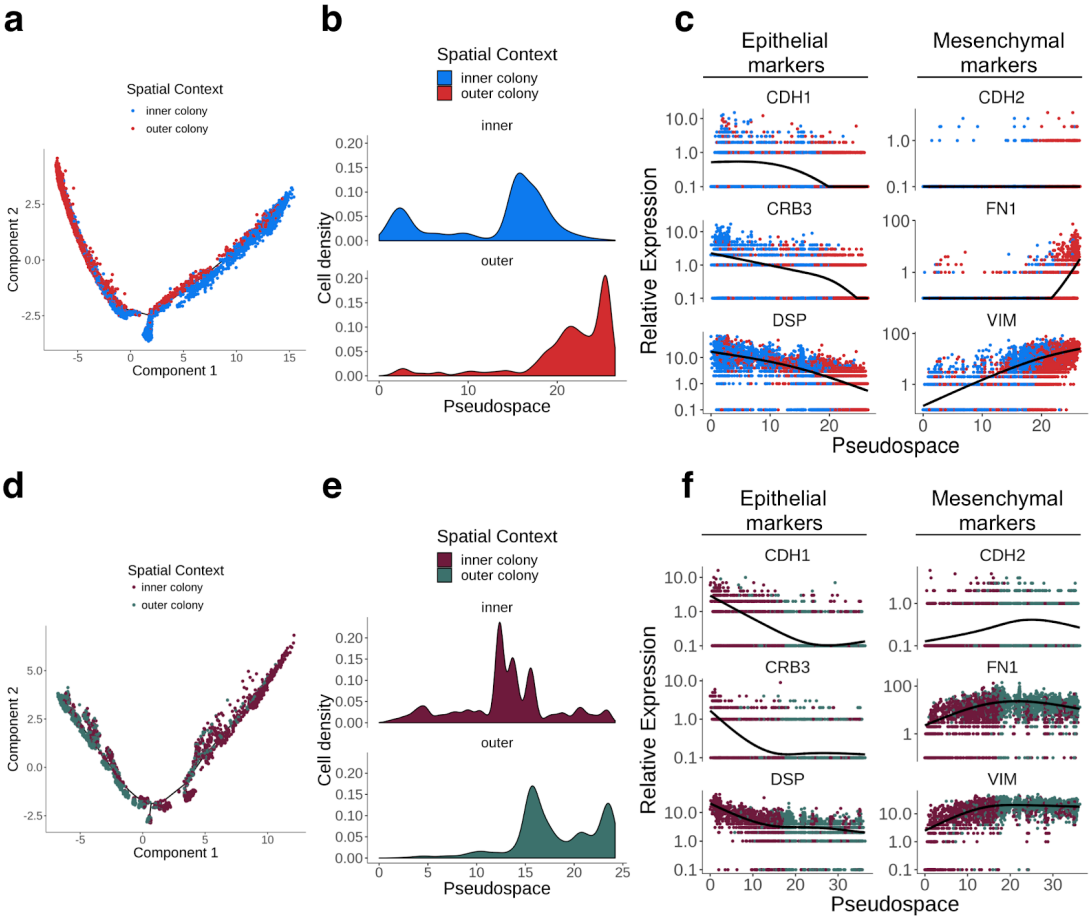
a) tSNE embedding of MCF10A cells upon TGF-ß-driven EMT progression. Cells are colored by their assignment to clusters identified using Louvain community detection in PCA space (9,951 cells). VIM low clusters are labeled as in **b**. **b)** VIM levels across tSNE clusters from **a**. Point within the violin depicts the mean expression level for each group with violin spanning the minimum and maximum expression values across cells. **c)** Distribution of TGFBR1 sgRNA containing cells across tSNE from **a**.



Supplementary Figure 20: Replicability of spatial differential gene expression between original and pooled loss of function spatial experiments

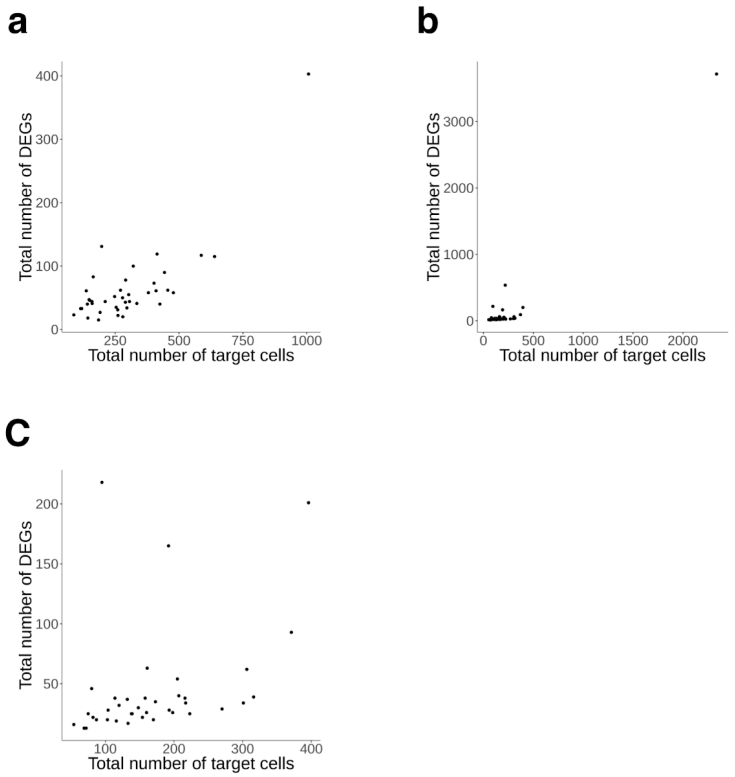
a-b) Venn diagram demonstrating the overlap between spatially differentially expressed genes (DEGs) for our original experiment pertaining to Figures 1 and 2 and cells expressing non-targeting controls in our pooled loss of function screen for spontaneous (**a**) and TGF-ß-driven (**b**) EMT. **c-d**) Correlation between beta coefficients between cells from our original spatial experiments and cells expressing non-targeting controls for our original spatial DEGs for spontaneous (**c**) and TGF-ß-driven (**d**) EMT (Pearson's r and Pearson chi-square test p-value). (**e-f**) Alignment plots of dynamic time warped trajectories illustrate the warping of non-targeting control cells from spontaneous (**e**) or TGF-ß-driven (**f**) EMT to their respective unedited counterparts from our original MCF10A experiments. (**g**) Alignment plot illustrating the warping of non-targeting control cells from our loss of function spontaneous EMT to our loss of function TGF-ß-driven EMT trajectory. Note that the cost of aligning non-targeting control cells from our

loss of function to our original experiments is minimal compared to aligning between EMT conditions.

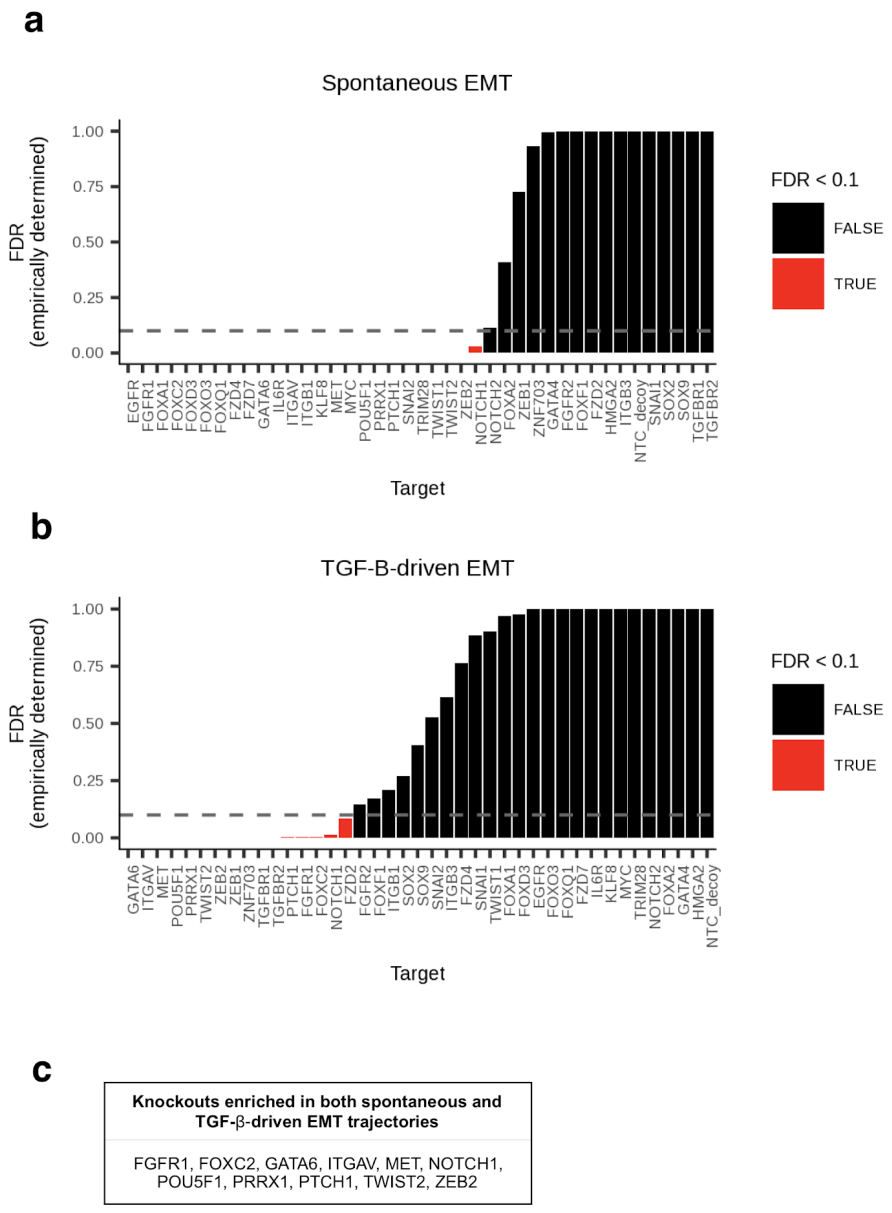


Supplementary Figure 21: Pseudospatial trajectory reconstruction of CROP-Seq edited cells undergoing spontaneous and TGF-ß-driven EMT

a/d) Trajectory of inner and outer CROP-Seq edited cells upon spontaneous and TGF-ß-driven EMT progression. Cells are colored according to pseudospatial region. **b/e)** Distribution of inner and outer cells amongst pseudospatial trajectories in the spontaneous **(b)** and TGF-ß-driven **(e)** conditions. Cells are colored by the fraction from which they were isolated. **c/f)** Expression of epithelial and mesenchymal markers across pseudospatial trajectories in the spontaneous **(c)** and TGF-ß-driven **(f)** conditions. Cells are colored by the fraction from which they were isolated.

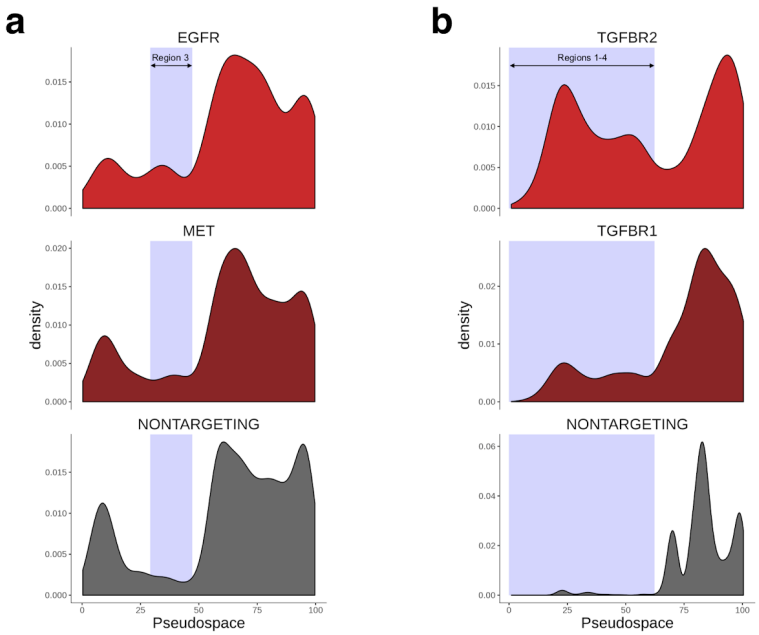


Supplementary Figure 22: Differential gene expression as a function of edited cell number **a-b**) Total number of DEGs vs total number of cells expressing guide RNAs against select cell surface receptors and transcription factors in our screen under spontaneous (**a**) and TGF-β (**b**) EMT conditions. **c**) Same as (**b**) with TGFBR1 and TGFBR2 knockouts removed.



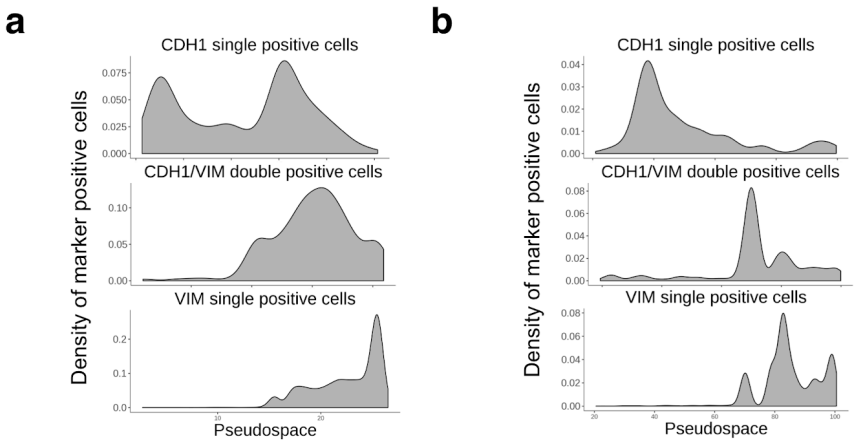
Supplementary Figure 23: Loss of function of a subset of EMT regulators results in differential distribution of knockout cells across pseudospace

Significance of the distribution of sgRNA containing cells across CROP-seq pseudospatial reconstruction for cells in the spontaneous (13,854 cells) **(a)** and TGF-ß-driven (9,951 cells) **(b)** conditions (FDR < 0.1, empirical FDR was determined by the rate at which each target was more significantly distributed according to Chi-square compared to a random subset of non-targeting controls (NTC-decoy) across 1000 iterations. **(c)** List of knockouts that are enriched in both spontaneous and TGF-ß-driven EMT conditions.



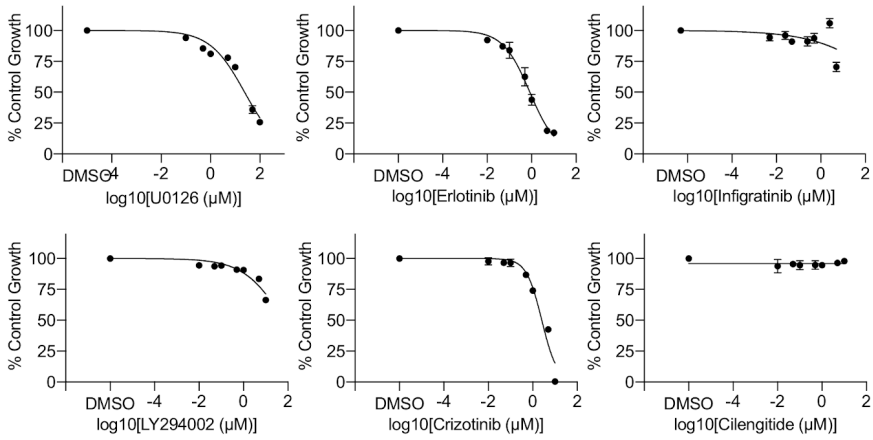
Supplementary Figure 24: Distribution of edited cells along pseudospatial trajectories reveals accumulation at various areas within the EMT continuum

Density plots of the distribution of cells expressing sgRNAs to the specified target or non-targeting controls under spontaneous and TGF-ß-driven EMT conditions. Shaded area corresponds to pseudospatial region(s) wherein EGFR and MET **(a)** or TGFBR1 and TGFBR2 **(b)** show accumulation according to our analysis.



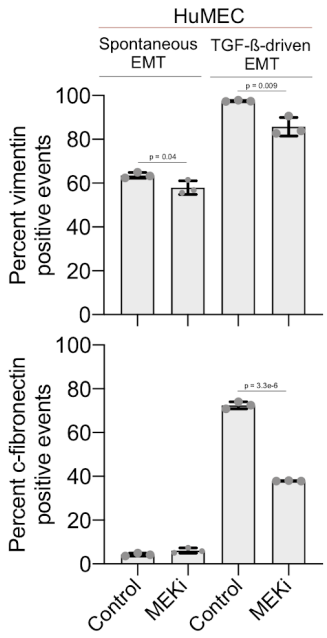
Supplementary Figure 25: Distribution of classically defined intermediate EMT states in confluence and TGF-ß-driven EMT CROP-Seq trajectories

Distribution of CDH1, VIM and CDH1/VIM double positive cells across pseudospatial regions for spontaneous **(a)** or TGF-ß-driven **(b)** EMT trajectories. Pseudospace was divided into regions depending on the density of cells across pseudospace, cells expressing CDH1 and VIM over their respective mean expression levels were identified and the fraction of cells positive for CDH1, VIM or both were plotted.



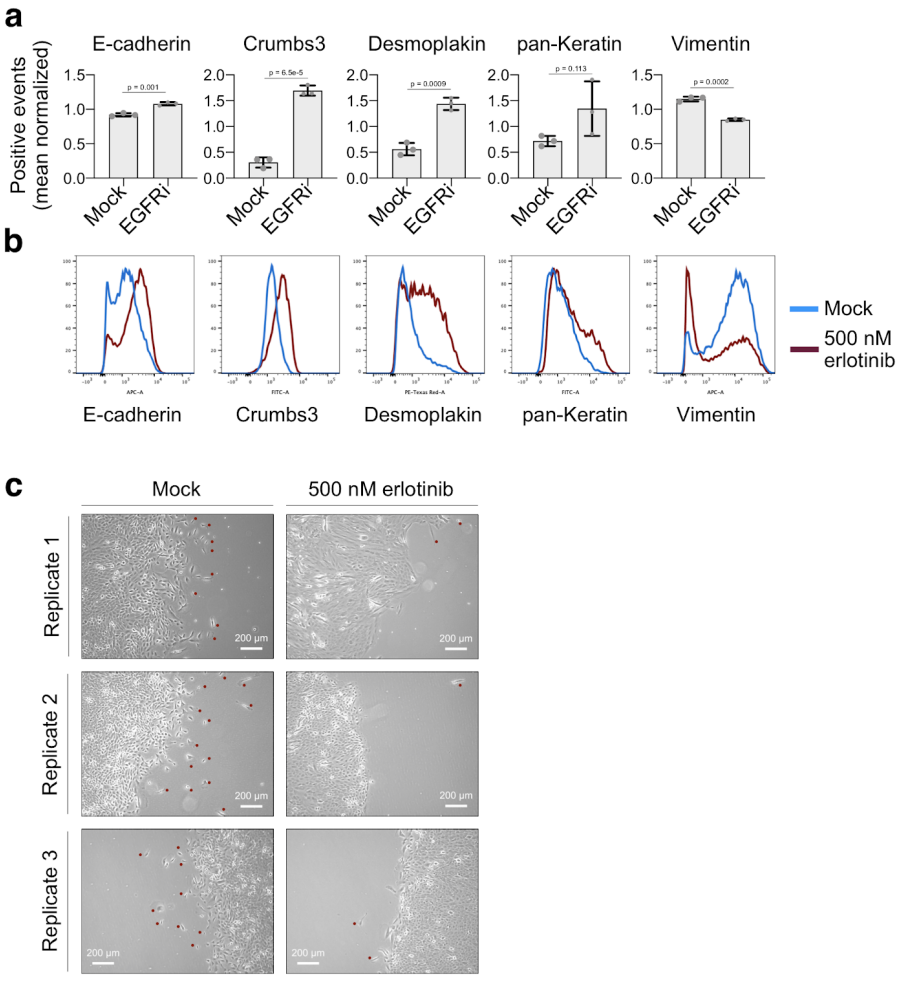
Supplementary Figure 26: Dose response curves of MCF10A cells to increasing concentrations of select inhibitors targeting proteins associated with KRAS/MEK/ERK signaling

Dose response curves obtained from treatment of MCF10A cells with the specified concentration of small molecule inhibitor for 96 hours. Control growth was assessed measuring ATP levels via the Cell Titer-Glo colorimetric assay (n = 3, culture replicates). Luminescence values for each drug dose combination were normalized to DMSO control. For every dose tested, point depicts the mean control growth, error bars depict standard deviation of the mean.



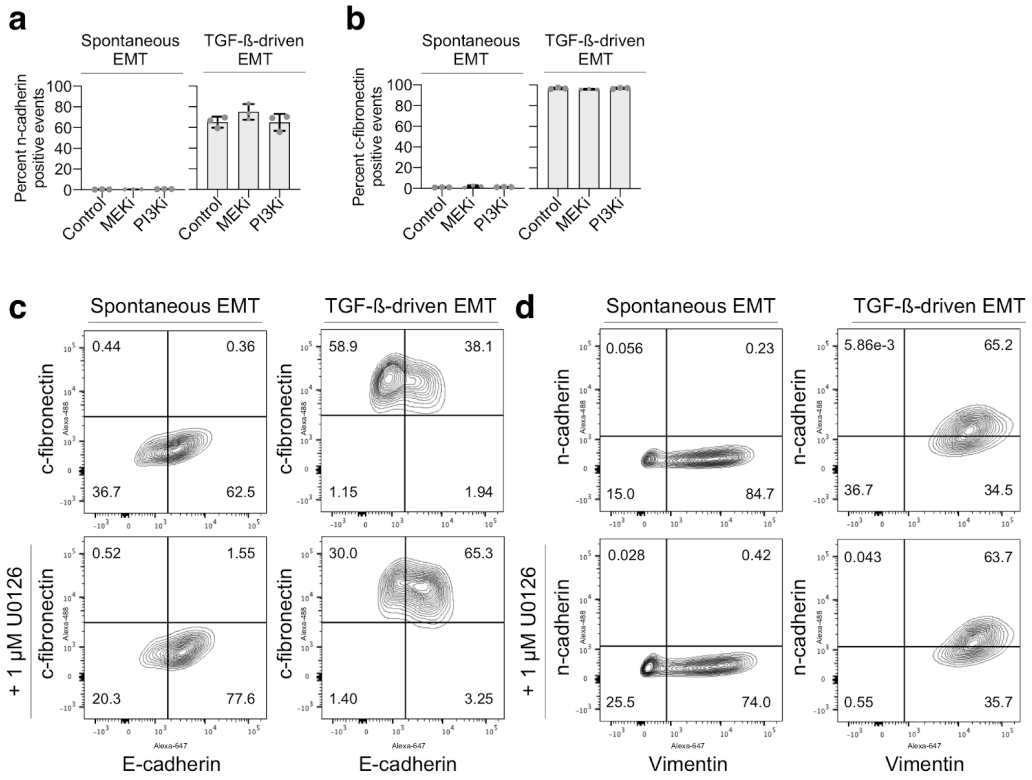
Supplementary Figure 27: MEK kinase inhibition decreases the acquisition of mesenchymal traits in HuMEC cells

Percent vimentin (top panel) or cytoplasmic fibronectin (bottom panel) positive cells in HuMEC colonies exposed to MEK (U0126) inhibition after spontaneous (Control and U0126) or TGF-ß-driven (TGFβ and U0126 + TGFβ) EMT. Error bars denote standard deviation from the mean (n = 3 independent samples, , two-tailed Student's t-test).



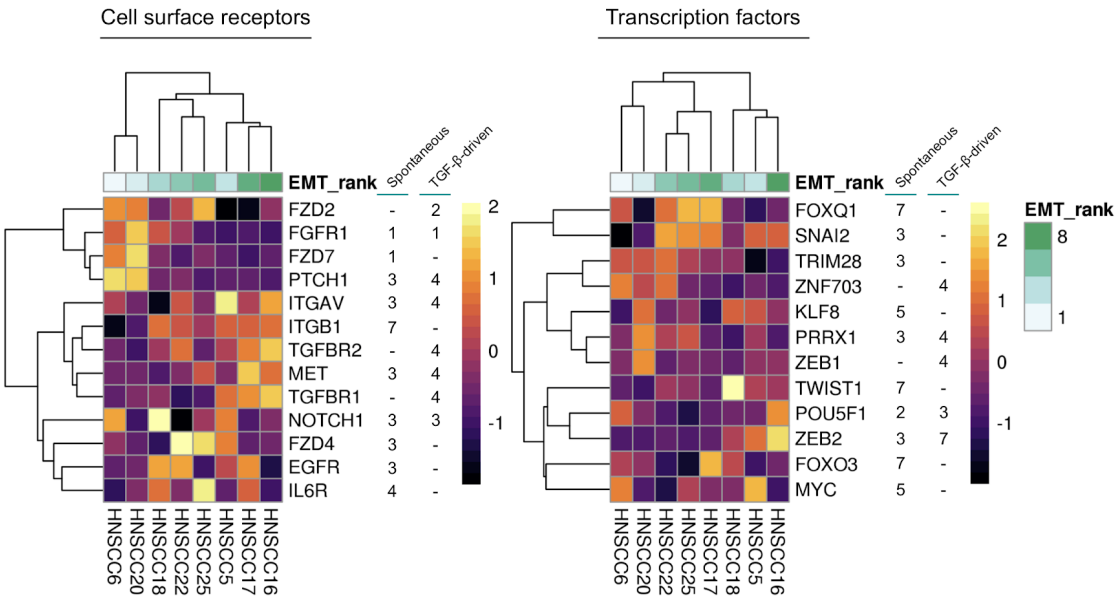
Supplementary Figure 28: EGFR signaling is necessary for exit from an epithelial state during spontaneous EMT in MCF10A cells

a) EMT marker protein levels in cells undergoing spontaneous EMT in the absence and presence of 500 nM of the EGFR inhibitor erlotinib. Error bars denote standard deviation from the mean (n = 3 independent samples, two-tailed Student's t-test). (**b**) Representative flow cytometry histograms of EMT marker protein levels from (**a**). (**c**) Representative brightfield images of cells at the boundary of our spontaneous EMT colonies in the absence and presence of erlotinib for the experiments described in (**a**) and (**b**) (200 µm scale bar, representative fields from 12 images collected, n = 3 independent experiments). Red dots have been placed immediately next to cells which do not appear to have cell-cell attachments to other cells in the colony and represent mesenchymal cells that have migrated away from the colony. Note the drastic reduction in these cells upon EGFR inhibition.



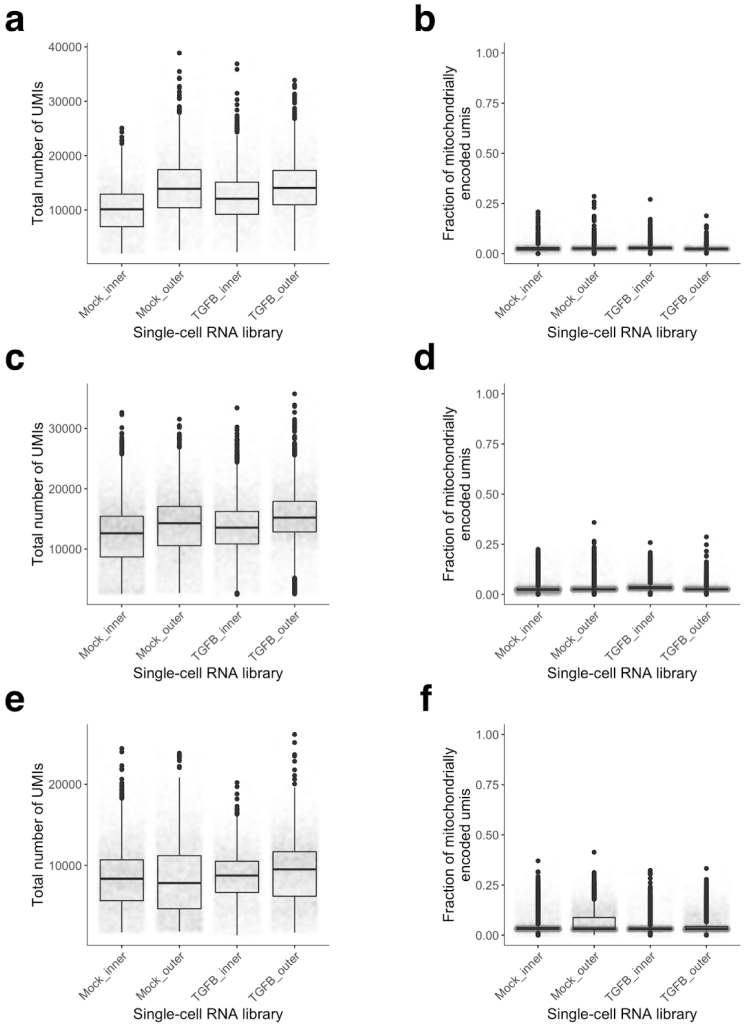
Supplementary Figure 29: Effect of MEK and PI3K inhibition on EMT marker express along spontaneous and TGF-ß-driven EMT

a) Percent n-cadherin positive cells in MCF10A colonies exposed to U0126 and LY294002 after spontaneous or TGF-ß-driven EMT. Error bars denote standard deviation from the mean (n = 3 independent samples). **b)** Percent cytoplasmic fibronectin positive cells in MCF10A colonies exposed to U0126 and LY294002 after spontaneous or TGF-ß-driven EMT. Error bars denote standard deviation from the mean (n = 3 independent samples). **c)** Contour plots depicting cytoplasmic fibronectin and e-cadherin protein levels of TGF-ß-driven EMT colonies as measured by flow cytometry. The numbers represent the proportion of the colony in each quadrant (n = 3 independent samples). **d)** Contour plots depicting N-cadherin and vimentin protein levels of TGF-ß-driven EMT colonies as measured by flow cytometry (n = 3 independent samples). The numbers represent the proportion of the colony in each quadrant. These are representative plots for the experiment presented in Figure 4c and panels a and b of this figure.



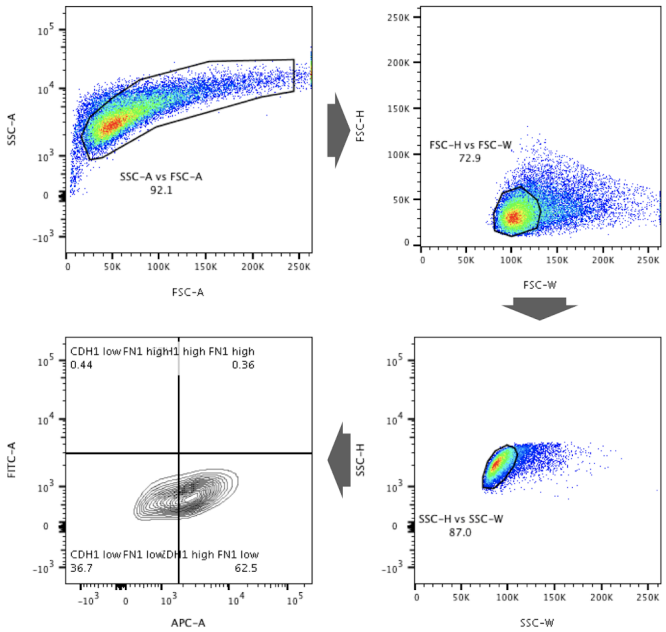
Supplementary Figure 30: Mean expression levels in HNSCC tumors cells of cell surface receptors and transcription factors identified in our screen as altering accumulation along MCF10A pseudospatial EMT trajectories

Each block represents the mean expression of a receptor (left heatmap) or transcription factor (right heatmap) in a HNSCC tumor sample. Annotation on top of the heatmap denotes the rank of the extent of partial EMT identified by Puram et al[15]. The number at right of the heatmap denote the region at which each receptor or transcription factor is maximally enriched within spontaneous or TGF-β-driven EMT trajectories.



Supplementary Figure 31: Median number of unique molecular identifiers (UMIs) and fraction of mitochondrial reads for single-cell RNA libraries from this study

(**a**/**c**/**e**) Total number of UMIs for MCF10A pseudospatial libraries (**a**) (Mock inner = 2,440 cells, Mock outer = 2,564 cells, TGF-β inner = 2,121 cells, TGF-β outer = 2,116 cells), MCF10A-Cas9 blast pseudospatial libraries (**c**) (Mock inner = 7,141 cells, Mock outer = 6,713 cells, TGF-β inner = 4,585 cells, TGF-β outer = 5,366 cells) and HuMEC pseudospatial libraries (Mock inner = 6,096 cells, Mock outer = 5,812 cells, TGF-β inner = 4,357 cells, TGF-β outer = 6,087 cells) (**e**). (**b**/**d**/**f**) Fraction of umis of mitochondrial encoded genes or MCF10A pseudospatial libraries (**b**), MCF10A-Cas9 blast pseudospatial libraries (**d**) and HuMEC pseudospatial libraries (**f**). Boxplots depict the median UMI (**a**/**c**/**e**) or mitochondrially encoded umis (**b**/**d**/**f**) (bold line within box) with lower and upper hinges depicting the 25th and 75th percentiles, respectively.



Supplementary Figure 32: Example of flow cytometry gating used to determine the protein levels of select EMT markers

Single cells were isolated via sequential gating on SSC-A vs. FSC-A, FSC-H vs FSC-W and SSC-H vs SSC-W according to standard flow cytometry practices. Gates for APC-A (describing Alexa-647 labeled e-cadherin or vimentin levels) and FITC-A (describing Alexa-488 labeled n-cadherin or fibronectin levels) were set using the spontaneous EMT sample as a negative control for n-cadherin and fibronectin low populations and the TGF-β driven EMT as a negative control for e-cadherin and vimentin low populations.

Sample	GEO sample accession	Cell line	Normalized mean reads per cell	Median genes per cell	Median number of UMIS	% mapped reads
MCF10A Spontaneous EMT inner colony	GSM3147293	MCF10A	47905	2473	10128	83.4%
MCF10A Spontaneous EMT outer colony	GSM3147294	MCF10A	47905	3070	13894	83.4%
MCF10A TGF-β-driven EMT inner colony	GSM3147295	MCF10A	47905	3084	12072	77.8%
MCF10A TGF-β-driven EMT inner colony	GSM3147296	MCF10A	47905	3523	14047	78.2%
HuMEC Spontaneous EMT inner colony	GSM3895092	HuMEC	30636	2282	8354	82.4%
HuMEC Spontaneous EMT outer colony	GSM3895093	HuMEC	30636	2402	7820	82.6%
HuMEC TGF-β-driven EMT inner colony	GSM3895094	HuMEC	30636	2527	8754	82.5%
HuMEC TGF-β-driven EMT inner colony	GSM3895095	HuMEC	30636	2676	9516	82.5%
MCF10A-Cas9 Spontaneous EMT inner colony	GSM3147297	MCF10A-Cas9-Blast	43557	2971	12612	84.9
MCF10A-Cas9 Spontaneous EMT outer colony	GSM3147298	MCF10A-Cas9-Blast	43557	3211	14277	85.2
MCF10A-Cas9 TGF-β-driven EMT inner colony	GSM3147299	MCF10A-Cas9-Blast	43557	3211	13537	83.9
MCF10A-Cas9 TGF-β-driven EMT inner colony	GSM3147300	MCF10A-Cas9-Blast	43557	3542	15224	84.1

Supplementary Table 1: Summary of quality metrics of single-cell RNA libraries reported in this study

GEO sample accession number and general quality control metrics for MCF10A, MCF10A-Cas9 and HuMEC cell experiments in this study.

Supplementary Table 2: Differential expression analysis between cell fractions of MCF10A cells undergoing spontaneous EMT
Results of differential gene expression analysis between inner (2,440 cells) and outer (2,564 cells) cell fractions of MCF10A cells undergoing spontaneous EMT (likelihood ratio test).

Supplementary Table 3: Pseudospatial differential expression analysis of spontaneous EMT
Results of pseudospatial differential gene expression analysis for 5,004 MCF10A cells undergoing spontaneous EMT (likelihood ratio test). Max AUC refers to the maximum area under the curve across 5 equally sized pseudospatial bins for every gene across the spontaneous EMT trajectory (see Methods).

Supplementary Table 4: Gene-set analysis of genes differentially expressed across pseudospace in cells undergoing spontaneous EMT
Results of gene-set analysis on differentially expressed gene clusters across 5,004 MCF10A cells undergoing spontaneous EMT using the GO Biological processes and MSigDB Hallmarks gene-set collections (hypergeometric test).

Supplementary Table 5: Differential expression analysis between cell fractions of HuMEC cells undergoing spontaneous EMT.
Results of differential gene expression analysis between inner (6,096 cells) and outer (5,812 cells) cell fractions of HuMEC cells undergoing spontaneous EMT (likelihood ratio test).

Supplementary Table 6: Differential expression analysis between aligned spontaneous and TGF-β-driven MCF10A EMT trajectories
Results of differential gene expression analysis between aligned MCF10A cells from spontaneous (5,004 cells) and TGF-β-driven (4,237 cells) EMT trajectories (likelihood ratio test). Max AUC difference refers to the maximum absolute relative difference of the area under the curve across 5 equally sized pseudospatial bins for every gene between spontaneous and TGF-β-driven EMT trajectories (see Methods).

Supplementary Table 7: Gene-set analysis of genes differentially expressed between aligned spontaneous and TGF-β-driven MCF10A EMT trajectories
Results of gene-set analysis on differentially expressed gene clusters between aligned MCF10A cells undergoing spontaneous (5,004 cells) and TGF-β-driven (4,237 cells) EMT trajectories using the GO Biological processes, MSigDB Hallmarks and MSigDB Oncogenic Signatures gene-set collections (hypergeometric test).

Supplementary Table 8: Pseudospatial differential expression analysis between knockout and non-targeting control MCF10A cells undergoing spontaneous EMT
Results of differential gene expression between knockouts and non-targeting control MCF10A cells undergoing spontaneous (13,854 cells) EMT (likelihood ratio test).

Supplementary Table 9: Pseudospatial differential expression analysis between knockout and non-targeting control MCF10A cells undergoing TGF-β-driven **EMT**
Results of differential gene expression between knockouts and non-targeting control MCF10A cells undergoing TGF-β-driven (9,951 cells) EMT (likelihood ratio test).

Supplementary Table 10: Sequences of oligonucleotides used in this study
Oligonucleotide sequences for oligonucleotides used to generate CROP-seq libraries and for enrichment of sgRNA containing transcripts from 10x cDNA samples.