Morphological Single Cell Profiling of the Epithelial-Mesenchymal Transition

Susan E. Leggett, Jea Yun Sim, Jonathan E. Rubins, Zachary J. Neronha, Evelyn Kendall Williams, and Ian Y. Wong

Supporting Information

Supporting Methods

Cell Culture

Human mammary epithelial cells stably transfected with an ER-Snail-1^{6SA} retroviral construct (MCF-10A Snail) as well as the parental cells (MCF-10A) were a gift from D.A. Haber (Massachusetts General Hospital).[35] The Snail-1^{6SA} variant is refractory to phosphorylation and is thus stably expressed and localized in the nucleus, where it initiates EMT induction.[36] Both MCF-10A variants were cultured following Brugge and coworkers.[37] Corning 96-well half area high content imaging microplates (Fisher Scientific, Cat. No. 15-100-170) were used for all imaging experiments. First, microplates were coated with fibronectin to aid cell adhesion; Briefly, fibronectin from human plasma (Sigma-Aldrich, Cat. No. F2006) was diluted with 1X phosphate buffered saline (PBS) to a final concentration sufficient to coat the well bottom with 5μg/cm^{2}of protein. Cells were seeded into the microplate wells at a concentration of 500 cells/well (unless otherwise specified), allowed to settle and adhere overnight, and were then subjected to exogenous application of stimuli and/or drugs as indicated. Special care was taken to ensure that cells remained subconfluent at the conclusion of the experiments. T-47D breast carcinoma cells were purchased from the DTP, DCTD Repository (Charles River Labs, Inc.), cultured in RPMI 1640 (ATCC Cat. No. 30-2001) supplemented with 10% Fetal Bovine Serum (Fisher Scientific, Cat. No. SH3007103), 0.2U/mL bovine insulin (Sigma-Aldrich, Cat. No. I1882), 100 units/mL penicillin, and 100 μg/mL streptomycin, and were passaged following standard ATCC protocol. MDA-MB-231 cells were purchased from ATCC and were cultured in DMEM containing L-Glutamine, L-glucose, and sodium pyruvate (Fisher Scientific, Cat. No. MT-10-013-CV) and supplemented with 10% FBS, 100 units/mL penicillin, and 100 μg/mL streptomycin. T-47D and MDA-MB-231 cells were cultured for 96 h in high content imaging plates as described and plated with 3,000 and 1,500 cells/well, respectively. These higher seeding densities with respect to MCF-10A experiments were experimentally determined, and compensate for differences in the characteristic doubling times of each cell line. All cell lines used were cultured in T25 flasks and passaged at 70-80% confluence; only low passage numbers were used ($P<15$). Cultures were maintained under standard conditions including humid incubation at 37° C and 5% CO_2.

EMT Induction with Snail-1 or TGF-β1

Snail-1 expression was induced in MCF-10A Snail cells through the addition of 4-hydroxytamoxifen (OHT; Sigma-Aldrich, Cat. No. H7904) resuspended in DMSO and used at a final concentration of 500 nM in media. To induce EMT to varying extents at the population level, a time course of Snail-1 induction was prepared with four conditions, all over a course of 72 h. The durations of OHT exposure included 0, 24, 48, and 72 h, where DMSO (0.05% final concentration, consistent with OHT dose) was applied during the remainder of the 72 h time course. Separately, EMT was induced in MCF-10A cells through treatment with 5 ng/mL recombinant human TGF-β1 (R&D Systems, Cat. No. 240-B) in growth media. TGF-β1 was prepared at 20 μg/mL in 4mM HCl containing 1mg/mL bovine serum albumin and the final media contained 0.025% of stock solution. In the absence of TGF-β1, cells received regular growth media. A comparable time course was conducted with four conditions over 72 h. All conditions were initially treated with growth media and TGF-β1 was added to different wells for durations of 0, 24, 48 and 72 h. For both EMT induction experiments, media was replenished at the 48 h time point, to prevent nutrient deprivation and growth factor depletion.

To examine whether or not the 72 h time course of OHT treatment was sufficient to induce EMT and yield a stable mesenchymal population, a 72 h washout of OHT was performed. Briefly, MCF-10A Snail cells were preinduced with OHT for 72 h, washed with media, and then media containing 0.05% DMSO was added for an additional 72 h (Fig. S8). Similarly, a TGF-β1 washout experiment was performed where TGF-β1 was applied to MCF-10A cells for 72 h, cells were washed with media, and supplemented with media only for an additional 72 h (Fig. S10). At the end of time courses, cells were fixed, immunostained, segmented using CellProfiler, and classified into epithelial and mesenchymal populations using GMM.

Density Dependent Induction of EMT

MCF-10A cells were plated in high content imaging plates, as described, where 375 cells/well was chosen as the lowest seeding density. This density was first validated during routine culture (equivalent to ~1:20 split) as it maintained the ability to produce characteristic epithelial monolayers several days after passaging cells. Thus, to generate a range of terminal densities, cells were plated at either 375 or 500 cells/well (the typical density for all other experiments) and subsequently cultured for either 72 or 96 h before fixation and immunostaining. The resulting density for each segmented (CellProfiler) and analyzed (GMM) condition was calculated by first counting the number of cells per 20x field of view of the acquired images, which was then divided by the area of the 20x field of view to yield: #Cells/Surface Area. For ease of comparison between conditions, the overall confluency of cells in the entire 20x field was estimated by eye. Thus, both qualitative and quantitative density measures were determined in order to effectively evaluate density dependent induction of EMT (Fig. S12). For the sparse density condition, multiple experimental replicates were pooled to obtain >100 cells for GMM and thus, the density was determined as an average estimate of 15% confluency across replicates.

Drug Treatment with Taxol for EMT Induction

MCF-10A Snail cells were cultured in media with OHT for 72 hours to induce a mesenchymal phenotype (preinduced) or with DMSO to maintain an epithelial phenotype (uninduced). Both preinduced and uninduced cells were seeded at a density of 750 cells/well. This elevated seeding density relative to OHT and TGF-β1 experiments was used due to the shortened duration of the time course with Taxol treatment (48 h Taxol vs. 72 h OHT and TGF-β1). Cells were then allowed to adhere to fibronectin-coated multiwell plates overnight as described, and the growth media containing OHT or DMSO was removed. Both cell types were then treated with either DMSO (0.05%, control) or 4 nM paclitaxel (Taxol) for 48 hours, and subsequently fixed and immunostained.

Cell Viability: The 4nM Taxol dose was determined to be sublethal by running two different viability assays. At the completion of the experiment, dead cells were labeled with DRAQ7 stain (Abcam, Cat. No. ab109202). Briefly, DRAQ7 (0.3mM) was added to each well at a final dilution of 1:200 and cells were incubated at 37˚C, 5% CO_2, for 10 minutes. Next, cells were imaged as described, using a 10X Plan Fluor objective (NA 0.3, long working distance); images were acquired using Phase contrast to capture images of live and dead cells, while Cy5 was used to capture images of DRAQ7 positive, dead cells. The indicated drug dose resulted in cell death comparable to untreated cells and cells treated with DMSO (Fig. S19A-C). To quantify the degree of cell death for each condition, cells were plated in 6-well tissue culture plates (Genesee scientific), allowed to settle and adhere overnight as in imaging experiments, then treated with media (control), 0.05% DMSO, or 4nM Taxol for 48 h. At the end of the 48 h treatment period, cells were washed with 1X PBS, and Accumax was added to obtain a single cell suspension. Once fully lifted, cells were resuspended in MCF-10A resuspension media, centrifuged, resupended in 200ul growth media, and then a small aliquot (20ul) of cell suspension was mixed at a 1:1 ratio with 0.2% Trypan Blue. Next, 20ul of the mixed suspension was added to a Cellometer Counting Chamber (Nexcelom Bioscience, Cat. No. SD100) and percent viability was determined using an automated imaging and analysis program: Trypan Blue Viability Assay with Cellometer Auto 1000 Cell Viability Counter (Nexcelom Bioscience) (Fig. S16D). Percent viability was determined for 3 replicates of each condition.

Immunostaining and Fluorescent Imaging

At the conclusion of time course experiments, cells were fixed for 20 minutes at 4˚C with 4% paraformaldehyde in 1X PBS (all solutions are in 1X PBS unless otherwise specified) containing 2mM calcium chloride and 2mM magnesium chloride. Cells were subsequently immunostained for the epithelial and mesenchymal biomarkers E-cadherin and vimentin, respectively. Cells were permeabilized with 0.1% Triton X-100 for <5 minutes, washed, and blocked with 10% goat serum. Cells were washed with sodium acetate buffer (7.5mM, pH 7.4, in 1X PBS) for 15 minutes and then rinsed several times with 1% nonfat dry milk. Cells were incubated overnight at 4˚C with primary antibodies: 250 μg/mL E-cadherin (Fisher Scientific, Cat. No. BDB610181) diluted at 1:500 and vimentin (Cell Signaling Technology, Cat. No. 5741S) diluted at 1:200 in 1% milk. Cells were then washed with 1% milk and incubated in the dark at room temperature with suitably matched secondary antibodies: 2mg/mL Alexa Fluor 488 and Alexa Fluor 555 (ThermoFisher Scientific, Cat. No. A-11001 and A-21428, respectively) diluted at 1:500 in 1% milk. Lastly, cells were washed with 1X PBS, incubated with 2 μg/mL Hoechst Pentahydrate and 2 μg/mL HCS CellMask Deep Red (ThermoFisher Scientific, Cat. No. 33258 & H32721, respectively) for 30 minutes at room temperature, and washed again.

Immunostained cells were imaged using an inverted epifluorescence microscope (Nikon TiE). Images were acquired with a 20X Super Plan Fluor objective (NA = 0.45, extra long working distance) and with a 14-bit resolution sCMOS camera (Andor Neo). Fluorescence illumination was provided by a light-guide coupled Lumencore Sola white light excitation system. Care was taken to ensure all images were acquired with identical acquisition parameters (exposure time, camera gain/gamma control, and microscope aperature settings). Post-acquisition, fluorescence thresholds for each experiment were set for qualitative comparison across conditions using built-in NIS Elements AR settings, specifically 10% Low (under-exposed) to eliminate background, and 0.5% High (over-exposed) to reduce pixel saturation. Thresholds were set on a channel-by-channel basis, with respect to the condition with the highest level of biomarker expression. Lastly, gamma was adjusted to 0.5 for all images of cytoplasmic features (Deep Red stain, Cy5 channel) to accentuate dim features of cells for visualization purposes only (Figures);

gamma was kept at 1 for all other images presented in the paper and for the segmentation of all images.

Image Analysis

Cell detection and shape feature analysis were performed using automated cell segmentation with CellProfiler 2.1 (Broad Institute) (Fig. S2).[38] First, the "Correct Illumination –Calculate and –Apply" modules were applied to correct for uneven illumination across channels and image sets. Second, fluorescently labeled nuclei (DAPI channel) were segmented as primary objects, which then served as seed objects for the detection of fluorescently labeled vimentin (TRITC channel), a biomarker of EMT, and fluorescently labeled cytoplasm (Cy5 channel), which became the secondary objects for the cell body. These fluorescent objects were manually verified and corrected as needed using the "EditObjectsManually" module. Detailed segmentation parameters are described in the Supporting Information (Fig. S2). Finally, shape measurements were extracted from the segmented objects, particularly nuclear max radius, vimentin area, cytoplasm form factor, and cytoplasm max feret diameter, using the "MeasureObjectSizeShape" module. Fluorescence intensity metrics were also extracted using the "MeasureObjectIntensity" module, however, subsequent analysis revealed that shape metrics were more valuable for consistent cell classification.

Phenotype Classification using a Gaussian Mixture Model (GMM)

An initial training set was developed by segmenting cells treated either with DMSO or 4-OHT for 72 h, followed by 72 h treatment with DMSO, corresponding to putative epithelial and mesenchymal phenotypes. For each condition, CellProfiler was used to segment 100-500 cells for 15 shape features (Figure S3). These single cell features from both conditions were then combined into a single dataset. For consistent comparison, shape features were rescaled between 0 and 1 by dividing by the maximum value of each shape feature for the combined dataset. Notable shape features used to distinguish between epithelial and mesenchymal phenotypes included the following (as defined by CellProfiler 2.1):

Area: The actual number of pixels in the region.

FormFactor: Calculated as $4*\pi*\text{Area}/\text{Perimeter}^2$. Equals 1 for a perfectly circular object.

MaxFeretDiameter: The Feret diameter is the distance between two parallel lines tangent on either side of the object (imagine taking a caliper and measuring the object at various angles). The maximum Feret diameter is the largest possible diameter, rotating the calipers along all possible angles.

MaxRadius: The maximum distance of any pixel in the object to the closest pixel outside of the object. For skinny objects, this is 1/2 of the maximum width of the object.

Cell intensity metrics were also extracted, where the most noteworthy differences between epithelial and mesenchymal phenotypes were observed for vimentin metrics including mean, median and integrated intensity:

IntegratedIntensity: The sum of pixel intensities within an object.

Based on the 15 shape parameters and vimentin intensity metrics, all possible combinations of features (up to 5) were screened for their classification accuracy. An expectation maximization (EM) algorithm was used to determine maximum likelihood estimates of the parameters for a Gaussian Mixture Model[39] (*gmdistribution.fit*, MATLAB R2013b). This analysis specifies two subpopulations and assumes that the total population can be reconstructed as a weighted mixture of two phenotypes defined by a Gaussian distribution around some distinct phenotype with some overall probability. Based on this reference subpopulation model, the heterogeneity of a particular cell population can be estimated. For each cell, the posterior probability that it is either epithelial or mesenchymal can be computed from the reference model using Bayes' rule (*posterior*, MATLAB R2013b).

This classifier was then tested against separate datasets with known phenotypes, which were partitioned into two subpopulations and then evaluated for accuracy (*cluster*, MATLAB R2013b). First, a test set was constructed using MCF-10A cells, again treated either with DMSO or 4-OHT for 72 h, followed by 72 h treatment with DMSO. Second, a renormalized test set was constructed using epithelial (T-47D) and mesenchymal (MDA-MB-231) cells. The accuracy of the partitioning was assessed by comparing the actual and predicted values of control and induced conditions (*classperf*, MATLAB 2013b).

Overall, an optimized four metric GMM distribution was selected based on vimentin area, cytoplasmic maximum feret diameter and form factor, as well as nuclear max radius. Vimentin area, cytoplasmic max feret diameter, and nuclear max radius all showed a consistent increase when comparing the control condition (DMSO) with induced EMT (OHT), while cytoplasmic form factor displayed a consistent decrease across these conditions. Overall, this classifier showed the highest accuracy for epithelial and mesenchymal phenotypes in both training sets. Based on this four metric GMM reference model, all subsequent datasets were partitioned into epithelial and mesenchymal subpopulations, respectively (*cluster*, MATLAB R2013b).

Statistical Analysis

To assess statistical significance between the distributions of data for cell morphologic features between Epithelial (DMSO) and Mesenchymal (OHT) train and test conditions, the two-sample Kolmogorov-Smirnov test in MATLAB (Mathworks) was used. The null hypothesis was rejected at $p \leq 0.05\%$ (5% significance level) and only highly statistically different morphologic features were used for cell classification of epithelial and mesenchymal subpopulations. Boxplots displaying these cell shape features were generated in MATLAB (Mathworks) using the BOXPLOT function, in which the dividing line of the box represents the median, box edges signify the 25th and 75th percentiles, and whiskers indicate 99.3% coverage of the data. Notched boxplots were plotted with extremes at $q_2 - 1.57(q_3 - q_1)/\mathrm{sqrt}(n)$ and $q_2 + 1.57(q_3 - q_1)/\mathrm{sqrt}(n)$, where q_2 is the median (50th percentile), q_1 and q_3 are the 25th and 75th percentiles, respectively, and n is the number of observations. Two medians will be significantly different at the 5% significance level if their notches do not overlap. The PLOTSPREAD function (Mathworks File Exchange) was additionally used to overlay data points corresponding to individual cell metrics, which were appropriately colored according to cell classification with Gaussian mixture modeling, where epithelial = green and mesenchymal = red. This function offsets data points horizontally to aid visualization of the statistical distribution; this offset should not be interpreted as meaningful. The percentage of cells classified as epithelial and mesenchymal for each experiment were plotted as histograms and bar graphs using the HIST and BAR functions, respectively, in MATLAB (Mathworks).

Supplemental Figures

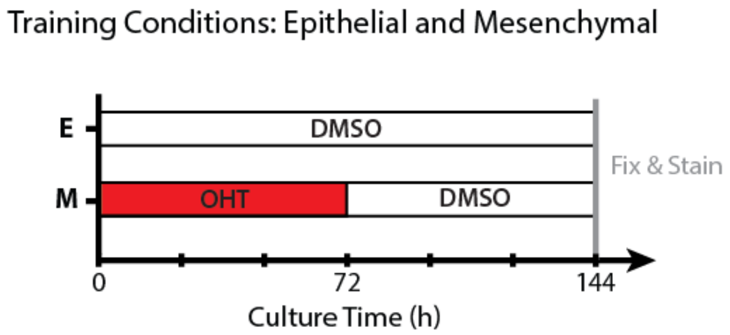
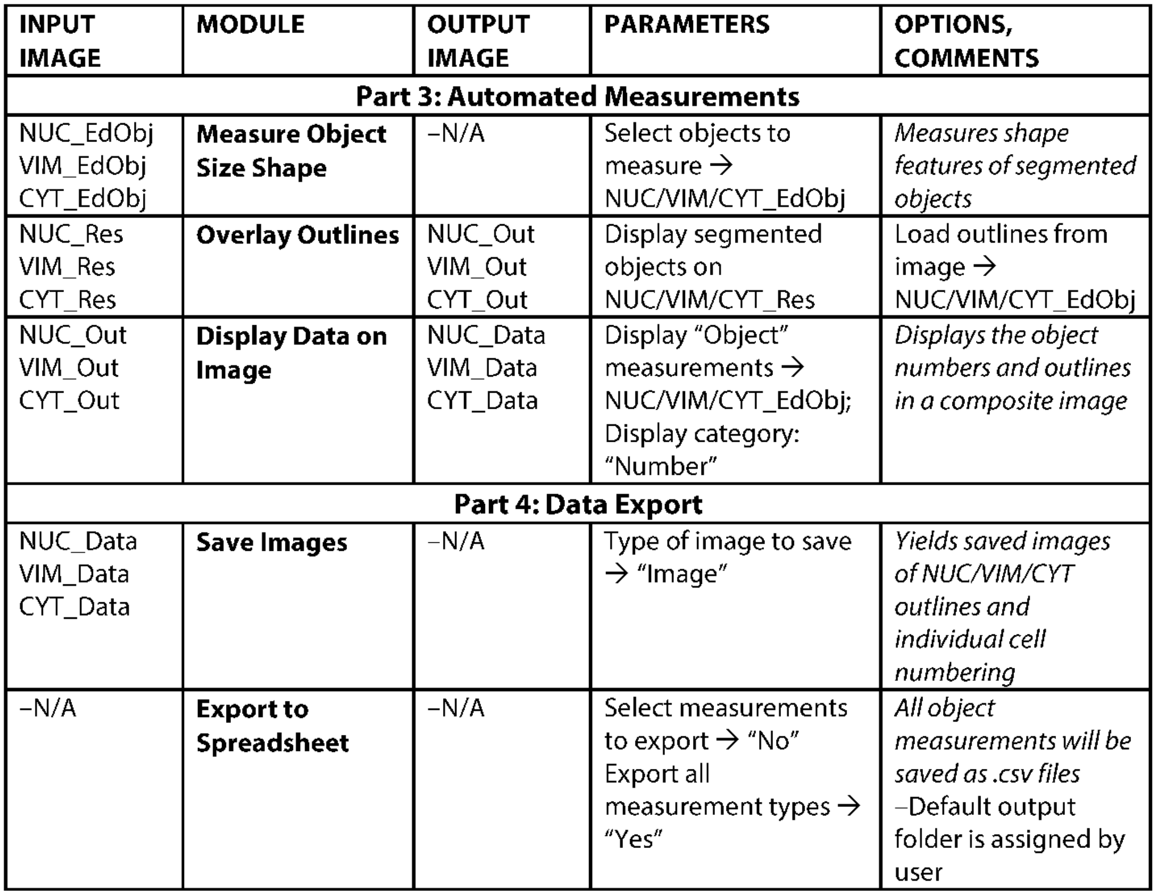


Figure S1. Schematic of Experimental Conditions for GMM Training Set. Overview of conditions that generate epithelial (E) and mesenchymal (M) populations used to train the Gaussian mixture model. MCF-10A Snail cells were seeded at t=0 (DMSO control, white bar), while OHT (500nM final concentration, red bar) was added to cells for a total duration of 72 h to induce a mesenchymal phenotype. From 72 h to 144 h, OHT was replaced with DMSO, to allow for the further progression of E and M phenotypes. At the completion of the time course, cells were fixed and subsequently immunostained for epithelial and mesenchymal biomarkers.

Figure S2. Flow chart of CellProfiler analysis pipeline

INPUT IMAGE	MODULE	OUTPUT IMAGE	PARAMETERS	OPTIONS, COMMENTS
Part 1: Image Enhancement				
NUC VIM CYT	**Correct Illumination Calculate**	NUC_Ilum VIM_Ilum CYT_Illum	Calculate by “Regular” method	Rescale → Yes Smoothing Method→ Fit Polynomial
NUC VIM CYT	**Correct Illumination Apply**	NUC_Corr VIM_Corr CYT_Corr	–Select Illumination Function: NUC/ VIM/CYT_Illum –Apply by “Divide”	*Corrects for Uneven Illumination*
VIM_Corr	**Enhance Or Suppress Features**	VIM_Enh	“Enhance” Operation	Feature type → Neurites; Method → Tubeness
VIM_Corr & VIM_Enh	**Image Math**	VIM_Math	Use “Add” Operation	*Adds enhanced features to objects*
NUC_Corr VIM_Math CYT_Corr	**Rescale Intensity**	NUC_Res VIM_ Res CYT_Res	Choose specific values to reset to full intensity range	Set intensity range for best visualization
Part 2: Object Segmentation				
NUC_Res	**Identify Primary Objects**	NUC_ Obj	–“Global” threshold strategy –“Otsu” threshold method –“Three-class thresholding”→assign middle intensity to foreground	Set diameter of objects (adjust per cell type) in pixel units; Distinguish clumped objects→ “Shape”; Draw dividing lines → “Intensity”
NUC_Obj	**Edit Objects Manually**	NUC_EdObj	Retain outlines of edited objects → Yes; “Retain” numbering of edited objects	Manually edit nuclei objects as needed; Use NUC_Res as guiding image
VIM_Res CYT_Res	**Identify Secondary Objects**	VIM_Obj CYT_Obj	Use “NUC_EdObj” as input objects and “propagation” method –“Global” threshold strategy –Threshold method: “Background” for VIM, “Otsu” for CYT	Use “Automatic” smoothing method for thresholding
VIM_Obj CYT_Obj	**Edit Objects Manually**	VIM_EdObj CYT_EdObj	Retain outlines of edited objects → Yes; “Retain” numbering of edited objects	Manually edit VIM/CYT objects as needed; Use VIM/CYT_Res as guiding images



INPUT IMAGE	MODULE	OUTPUT IMAGE	PARAMETERS	OPTIONS, COMMENTS
Part 3: Automated Measurements				
NUC_EdObj VIM_EdObj CYT_EdObj	**Measure Object Size Shape**	–N/A	Select objects to measure → NUC/VIM/CYT_EdObj	*Measures shape features of segmented objects*
NUC_Res VIM_Res CYT_Res	**Overlay Outlines**	NUC_Out VIM_Out CYT_Out	Display segmented objects on NUC/VIM/CYT_Res	Load outlines from image → NUC/VIM/CYT_EdObj
NUC_Out VIM_Out CYT_Out	**Display Data on Image**	NUC_Data VIM_Data CYT_Data	Display "Object" measurements → NUC/VIM/CYT_EdObj; Display category: "Number"	*Displays the object numbers and outlines in a composite image*
Part 4: Data Export				
NUC_Data VIM_Data CYT_Data	**Save Images**	–N/A	Type of image to save → "Image"	*Yields saved images of NUC/VIM/CYT outlines and individual cell numbering*
–N/A	**Export to Spreadsheet**	–N/A	Select measurements to export → "No" Export all measurement types → "Yes"	*All object measurements will be saved as .csv files* –Default output folder is assigned by user

Figure S2. Flow chart of CellProfiler analysis pipeline. General process overview for image analysis using CellProfiler 2.1 (Broad Institute). NUC, VIM, and CYT represent the original TIFs acquired for nuclear, vimentin, and cytoplasm channels, respectively, after fluorescence imaging of immunostained samples. CellProfiler modules were utilized in order as indicated, with the given input and output images. Parameters and comments provide key information for image enhancement, object segmentation, automated measurements, and data export used in the analysis.

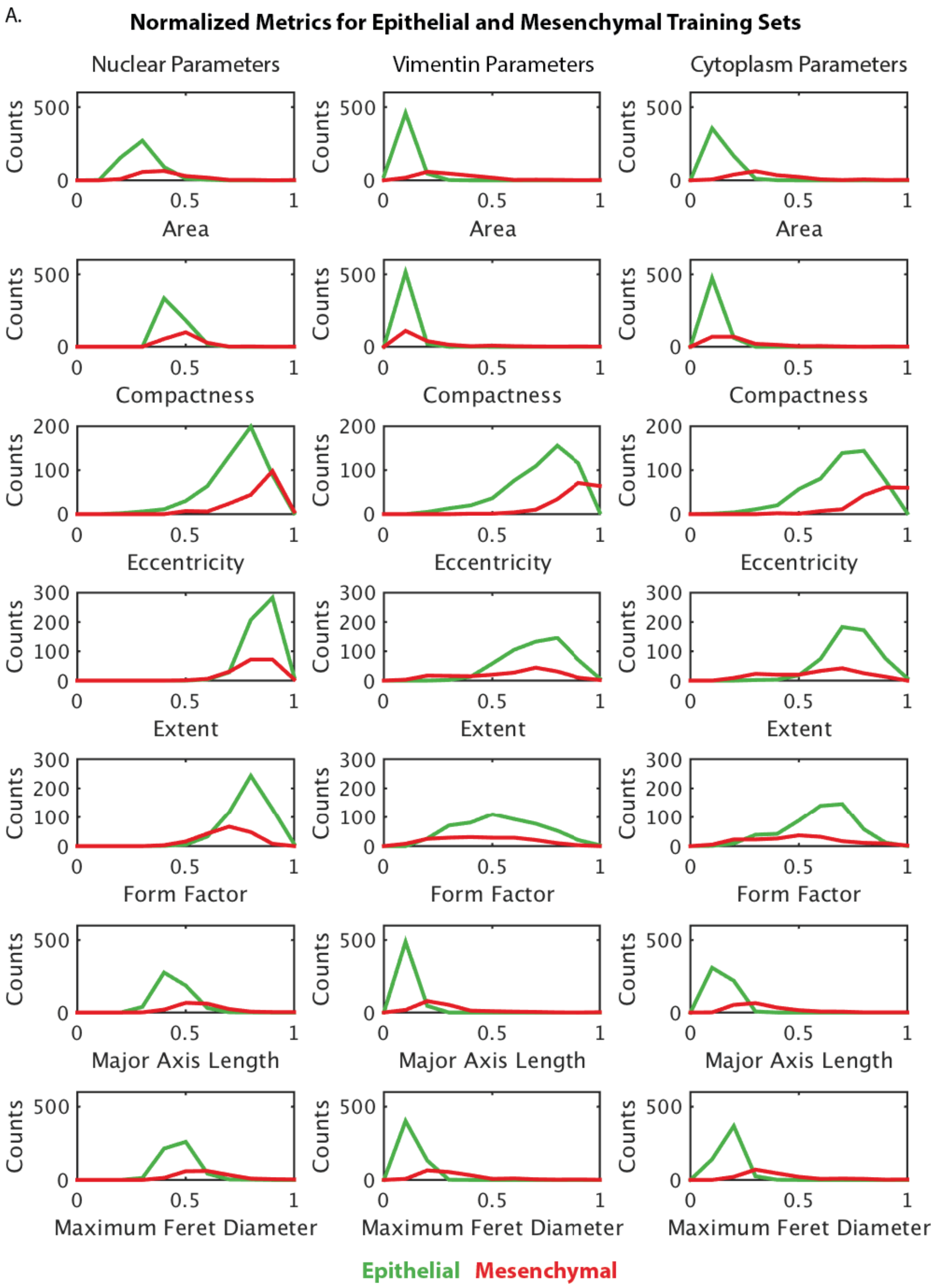


Figure S3A. Histograms of all Nuclear, Vimentin, and Cytoplasmic metrics for Epithelial and Mesenchymal Training Sets. Histograms of normalized distributions from 0 to 1 of all nuclear, vimentin, and cytoplasm shape metrics for the epithelial (DMSO) group (green), and mesenchymal (72 h OHT, 72 h DMSO) group (red). Counts (y-axis) represent the number of individual cells with the indicated normalized value (x-axis) for each metric.

B.

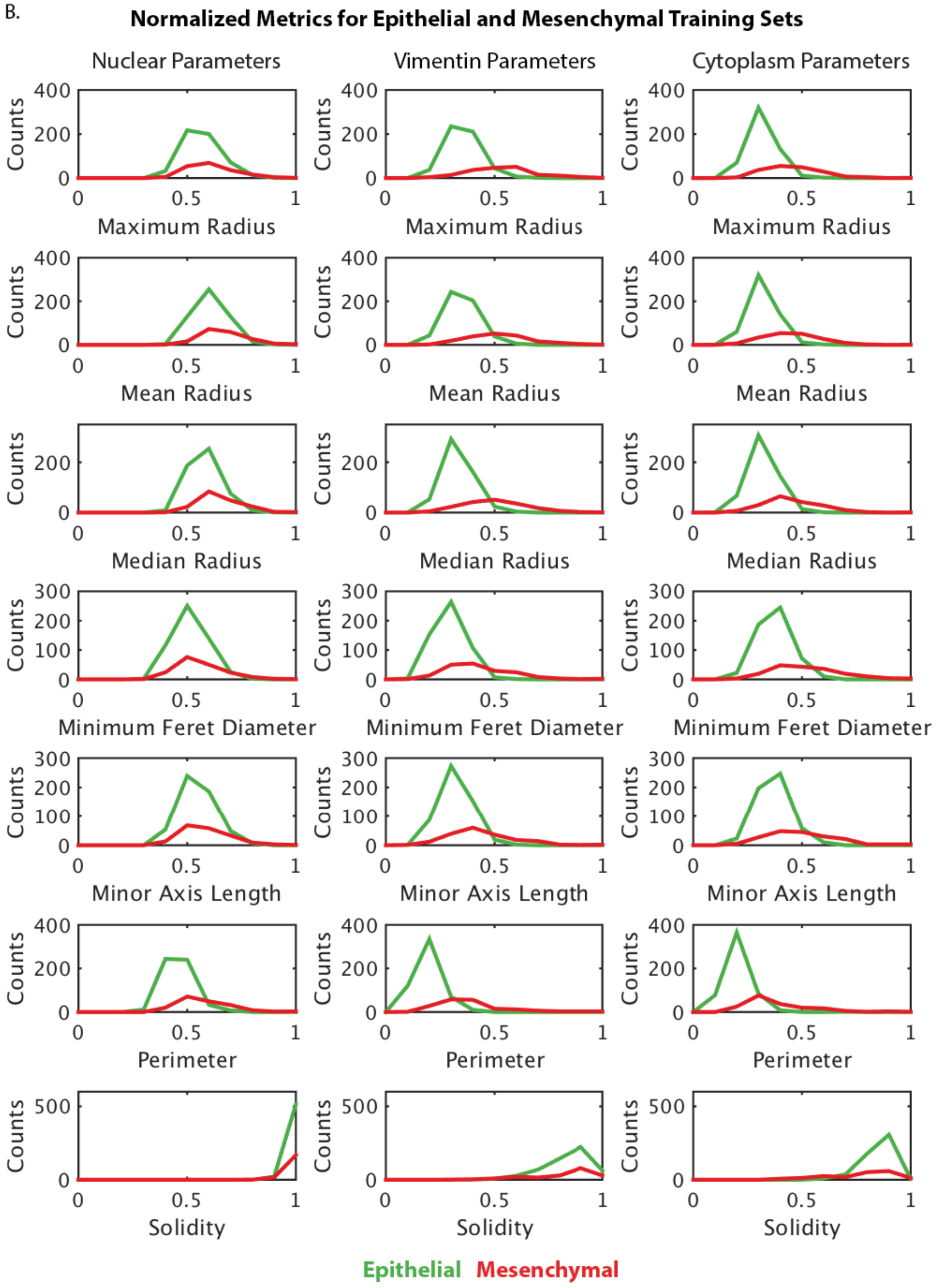


Figure S3B. Histograms of all Nuclear, Vimentin, and Cytoplasmic metrics for Epithelial and Mesenchymal Training Sets. Histograms of normalized distributions from 0 to 1 of all nuclear, vimentin, and cytoplasm shape metrics for the epithelial (DMSO) group (green), and mesenchymal (72 h OHT, 72 h DMSO) group (red). Counts (y-axis) represent the number of individual cells with the indicated normalized value (x-axis) for each metric.

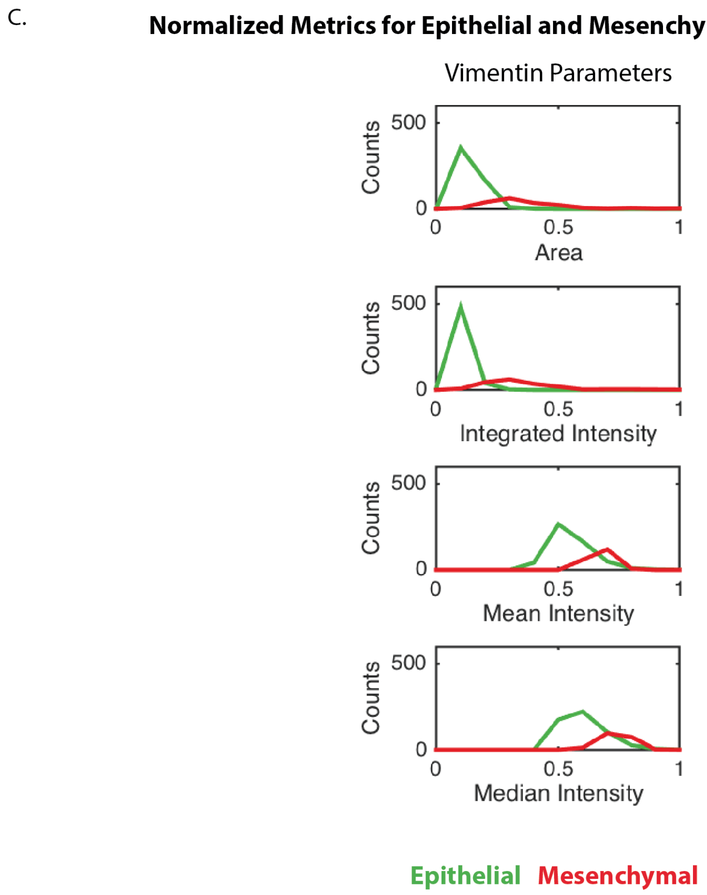


Figure S3C. Histograms of Vimentin Intensity metrics for Epithelial and Mesenchymal Training Sets. Histograms of normalized distributions from 0 to 1 of all vimentin intensity metrics including integrated intensity, mean intensity, and median intensity for the epithelial (DMSO) group (green), and mesenchymal (72 h OHT, 72 h DMSO) group (red). Vimentin area is also displayed for comparison, which demonstrates the overlap between area and integrated intensity distributions. Counts (y-axis) represent the number of individual cells with the indicated normalized value (x-axis) for each metric.

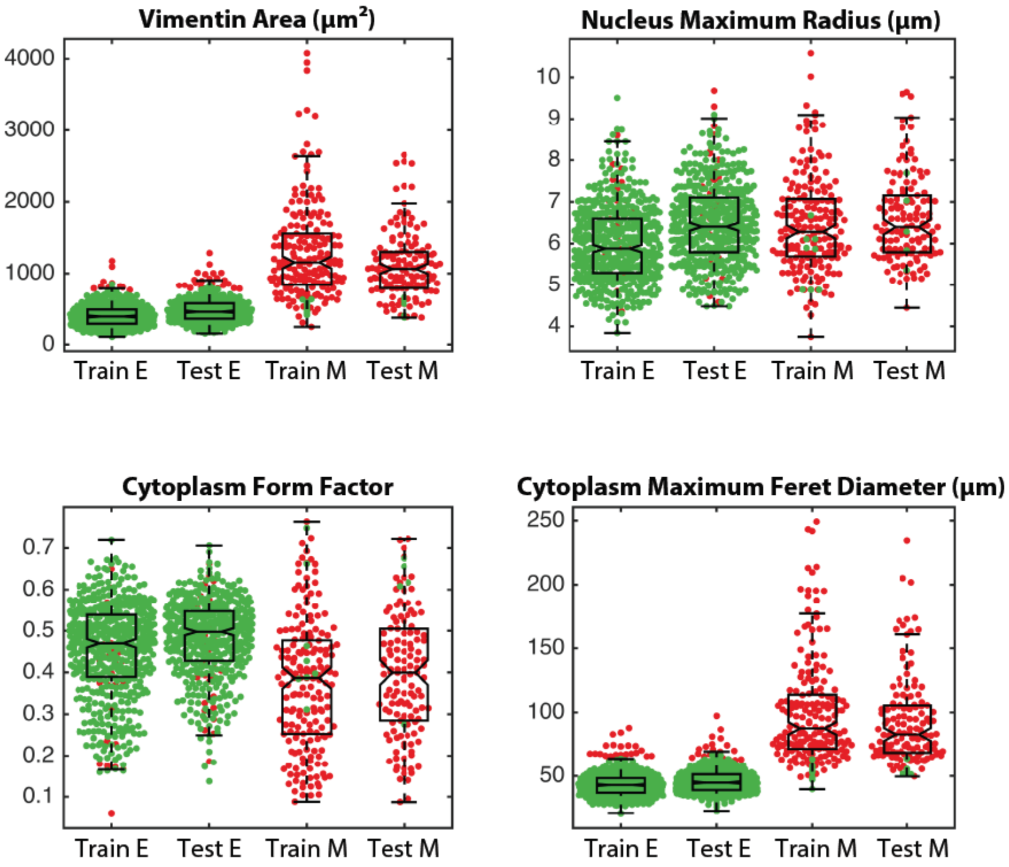
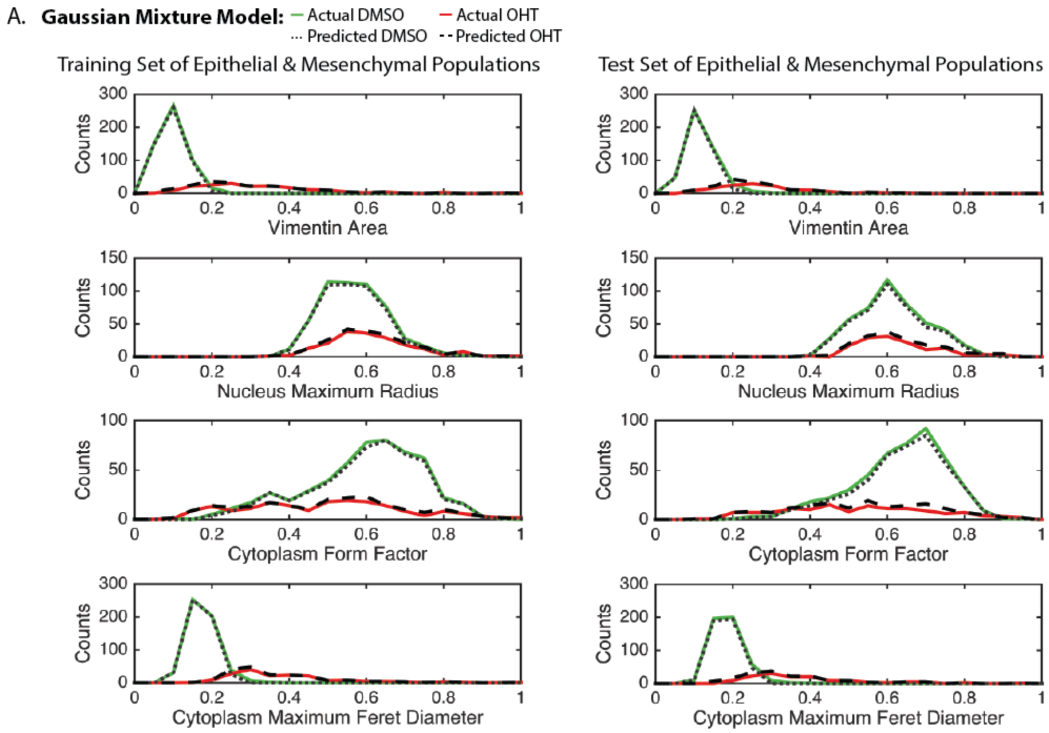


Figure S4. Boxplot of selected metrics used to distinguish Training and Test Epithelial and Mesenchymal conditions. Boxplots displaying individual cell raw values for each of the selected shape metrics of epithelial cells (E; DMSO condition, green) and mesenchymal cells (M; OHT condition, red) in the GMM training and test sets. For each of these selected metrics, the two-sample Kolmogorov-Smirnov test (*kstest2* in MATLAB) demonstrates that the data for E and M groups compromise different, continuous distributions.



B.

Training Set Accuracy

		Predicted DMSO	Predicted OHT
Actual	DMSO	503 (94%)	32 (6%)
	OHT	8 (4%)	177 (96%)

Test Set Accuracy

		Predicted DMSO	Predicted OHT
Actual	DMSO	436 (92%)	39 (8%)
	OHT	5 (4%)	129 (96%)

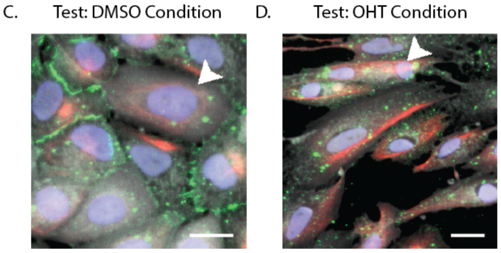


Figure S5. **Predicted vs. Actual Segmentation for GMM Training and Test sets.**

(A) Histograms of normalized distributions from 0 to 1 of selected metrics for the epithelial (DMSO condition, green) and mesenchymal (OHT condition, red) groups for training and test sets, displaying the GMM classification with predicted DMSO (small dashes), and predicted OHT (large dashes). Counts (y-axis) represent the number of individual cells with the indicated normalized value (x-axis). (B) Confusion matrix displaying the number and percentage of cells in each group. (C&D) Immunofluorescent staining of nuclei (blue), E-cadherin (green), vimentin (red), and cytoplasm (gray). Scale = 25μm. (C) Representative IF of the test set epithelial condition (72 h DMSO); Arrow indicates a misclassified cell predicted to be in the OHT group with GMM, likely due to phenotypic plasticity. (D) Representative IF of the test set mesenchymal condition (72 h OHT, 72 h DMSO); Arrow indicates a cell undergoing a division event that was misclassified by GMM (predicted DMSO).

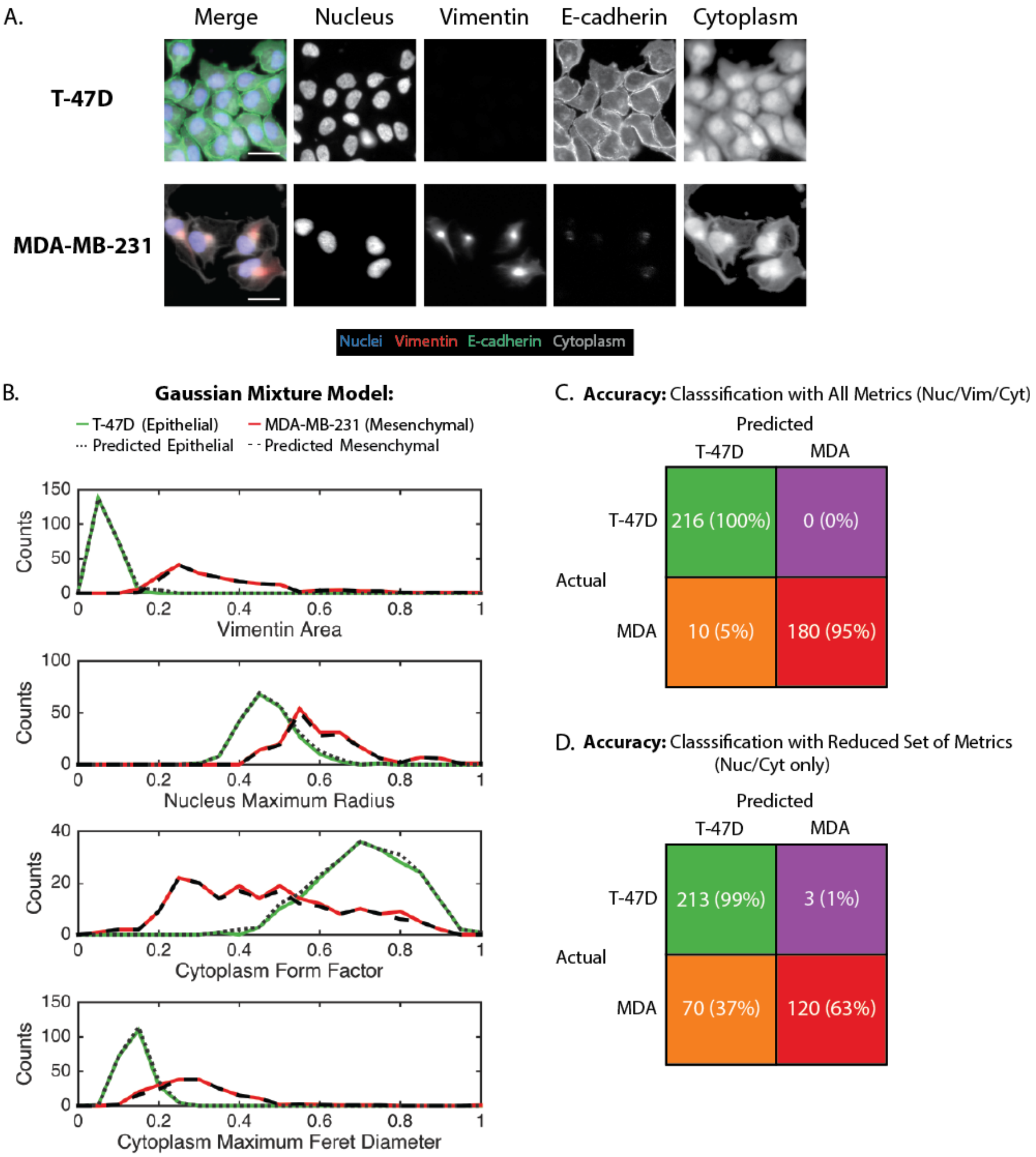


Figure S6. Predicted vs. Actual Segmentation for GMM of T-47D and MDA-MB-231 cells. (A) Immunofluorescent staining of nuclei (blue), E-cadherin (green), vimentin (red), and cytoplasm (gray) of epithelial and mesenchymal cell lines, T47-D and MDA-MB-231, respectively. Scale = 25µm. (B) Histograms of normalized distributions from 0 to 1 of selected metrics for T-47D (green) and MDA-MB-231 (red) groups, displaying the GMM classification with predicted T-47D (small dashes), and predicted MDA-MB-231 (large dashes). Counts (y-axis) represent the number of individual cells with the indicated normalized value (x-axis). (C&D) Confusion matrix displaying the percentage of cells in each group by classification using all 4 metrics (C) and a reduced set of 3 metrics with only nuclear and cytoplasm features (D).

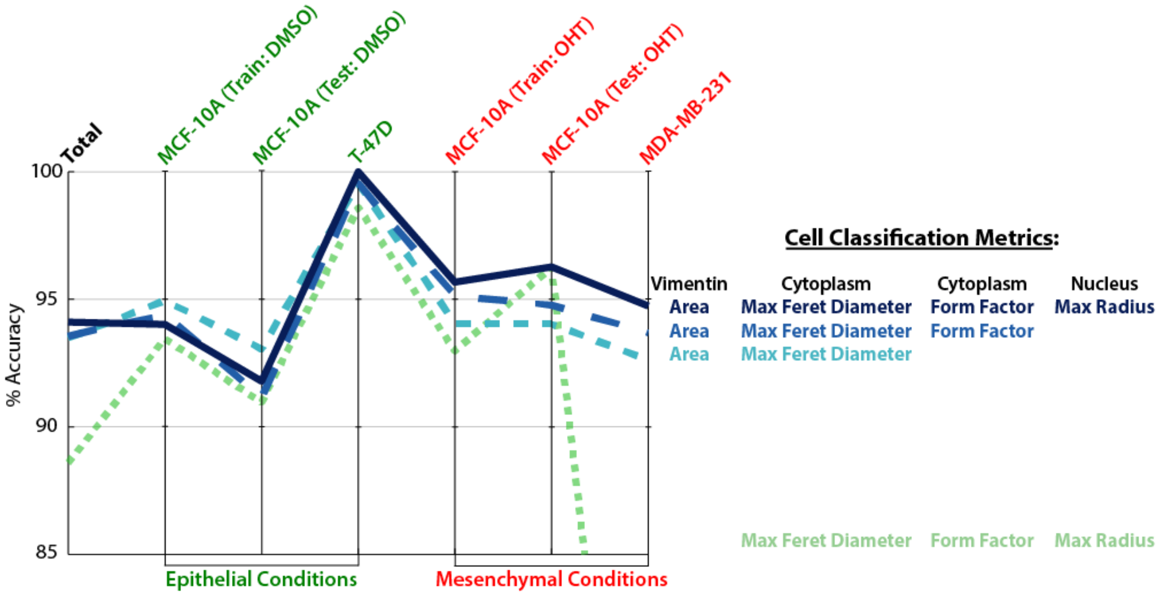


Figure S7. Optimal combinations of Nuclear, Vimentin, and Cytoplasmic metrics for Cell Classification. Parallel coordinates plot displaying the % accuracy (y-axis) for GMM classification of various epithelial and mesenchymal conditions (x-axis labels) with each of the following combinations of classification metrics (right). The 4-metric combination (solid, dark blue line) represents the best overall set of metrics for classifying both epithelial and mesenchymal conditions, and has the highest Total % accuracy (x-axis, far left).

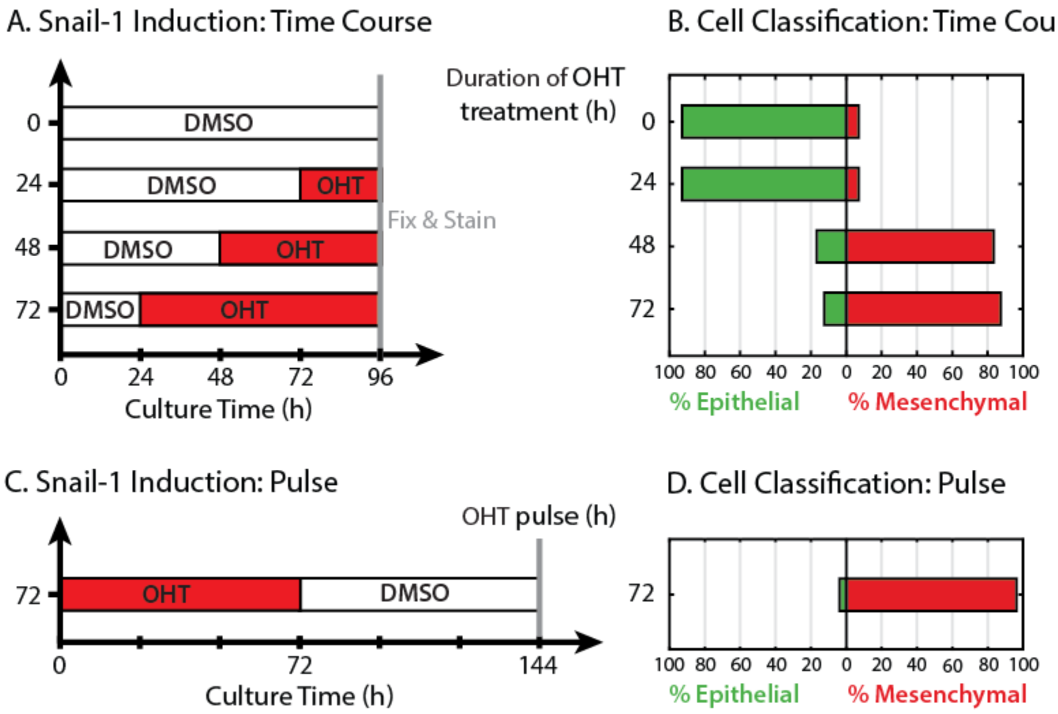


Figure S8. Schematic of experimental conditions and results for OHT induction.
(A) Overview for the timeline of Snail-1 induction with OHT. All cells were seeded at t=0 (DMSO control, white bar), while OHT (500nM final concentration, red bar) was added to cells at the indicated time points to yield 0, 24, 48, and 72 h OHT conditions. At the completion of the time course, cells were fixed and subsequently immunostained for epithelial and mesenchymal biomarkers. (B) Barplot of epithelial and mesenchymal fractions with duration of OHT induction. The percentage of cells classified as epithelial (green) largely decreases from 24 to 48 hours of OHT treatment, while the percentage of cells classified as mesenchymal (red) greatly increases between these conditions. (C&D) Overview of the timeline for the test condition (from Fig.1D, S4, S5) of OHT pulse (72 h OHT), withdrawal (72 h DMSO) (C), which yields a primarily mesenchymal population as represented by the barplot of epithelial and mesenchymal fractions (D).

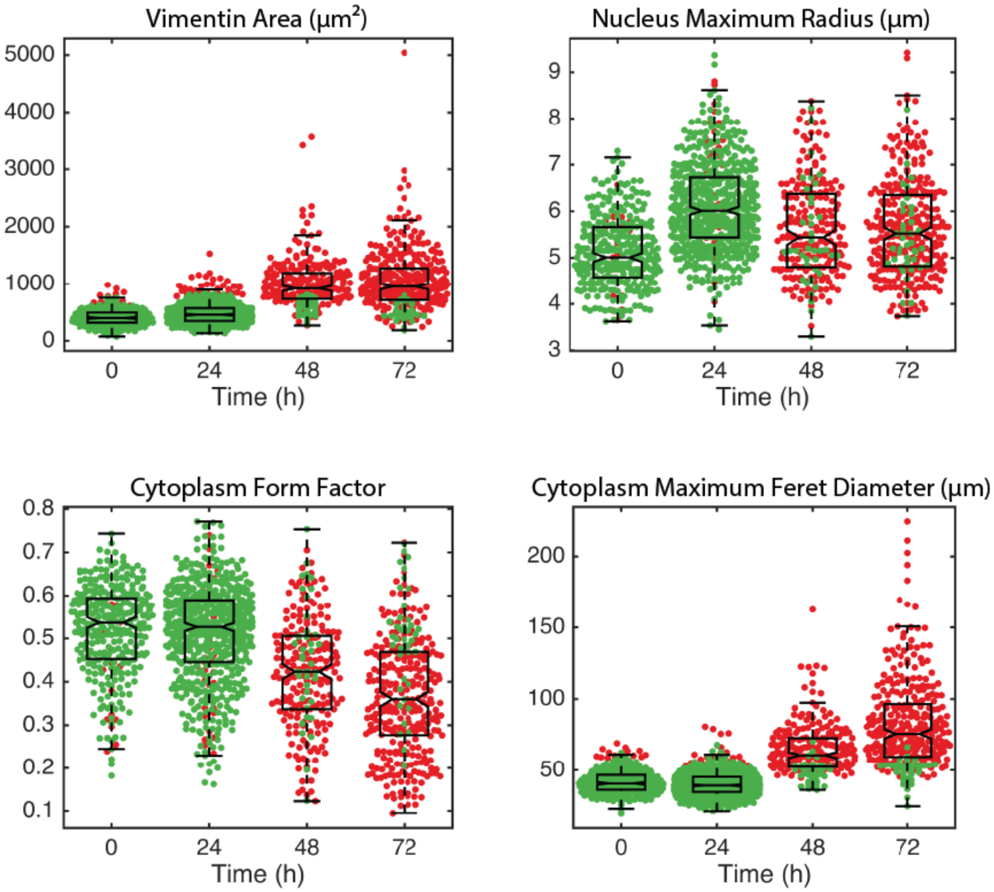


Figure S9. Boxplot of selected metrics with duration of OHT induction. Boxplots overlaid with data points representing individual cell values for each of the selected shape metrics of DMSO (control cells) and OHT (induced cells) for 0, 24, 48, and 72 h conditions. Individual cells were classified as epithelial (green) or mesenchymal (red) by GMM.

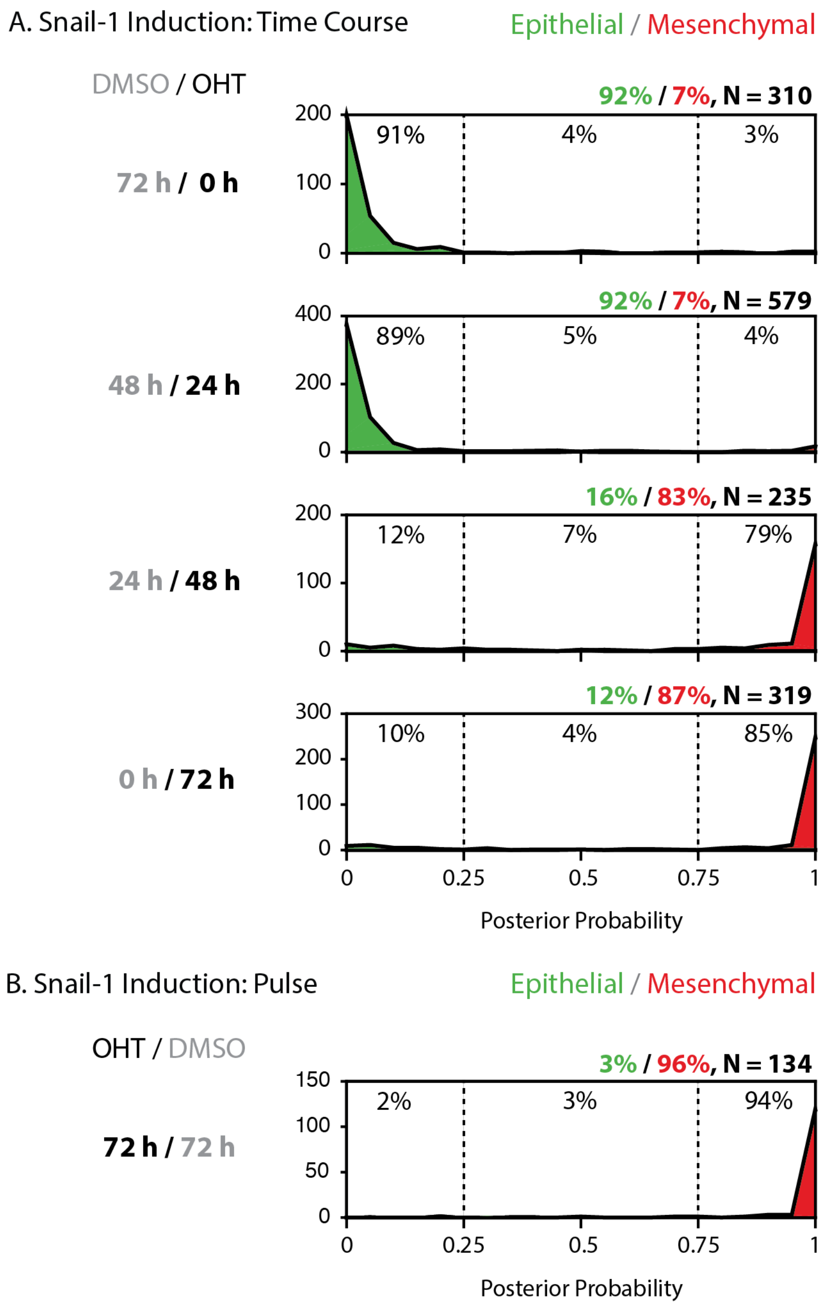


Figure S10. Posterior Probabilities of Gaussian Mixture Model for OHT induction. Posterior probability distributions depicting the conditional likelihood that a cell is mesenchymal (p=1). Percentage of cells classified as epithelial (green), mesenchymal (red) and total number of cells analyzed per conditioned (N) are displayed above graphs (right). For p = 0 to 0.25 and 0.75 to 1, there is a high probability that an individual cell is epithelial (green) and mesenchymal (red), respectively, with relative percentages displayed within graphs. The middle percentage where p = 0.25 to 0.75 represents the fraction of cells with less definitive classification. (A) For the time course of OHT induction and (B) for the OHT pulse, withdrawal condition.

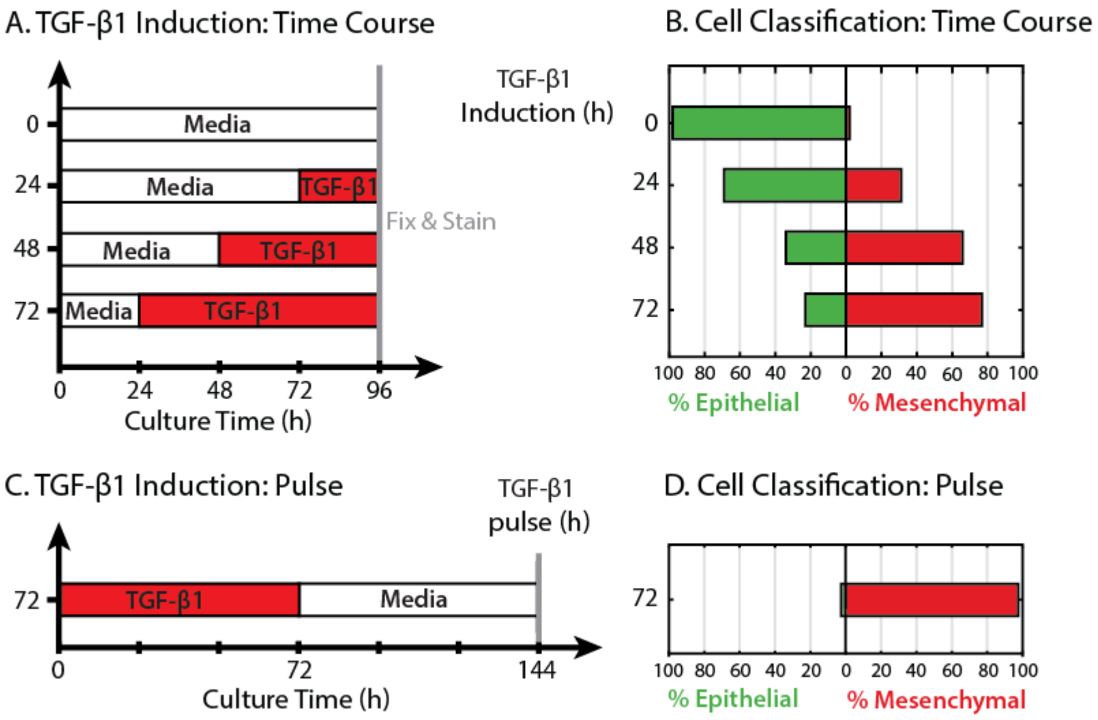


Figure S11. Schematic of experimental conditions and results for TGF-β1 induction. (A) Overview for the timeline of TGF-β1 induction. All cells were seeded at t=0 h (control, white bar), while TGF-β1 (5 ng/mL final concentration, red bar) was added to cells at the indicated time points to yield 0, 24, 48, and 72 h TGF-β1 conditions. At the completion of the time course, cells were fixed and subsequently immunostained for epithelial and mesenchymal biomarkers. (B) The percentage of cells classified as epithelial (green) gradually decreases from 0 to 72 hours of TGF-β1 treatment, while the percentage of cells classified as mesenchymal (red) increases for these conditions. (C&D) Overview of the timeline for the TGF-β1 pulse (72 h TGF-β1), withdrawal (72 h media) experiment (C), which yields a primarily mesenchymal population as represented by the barplot of epithelial and mesenchymal fractions (D).

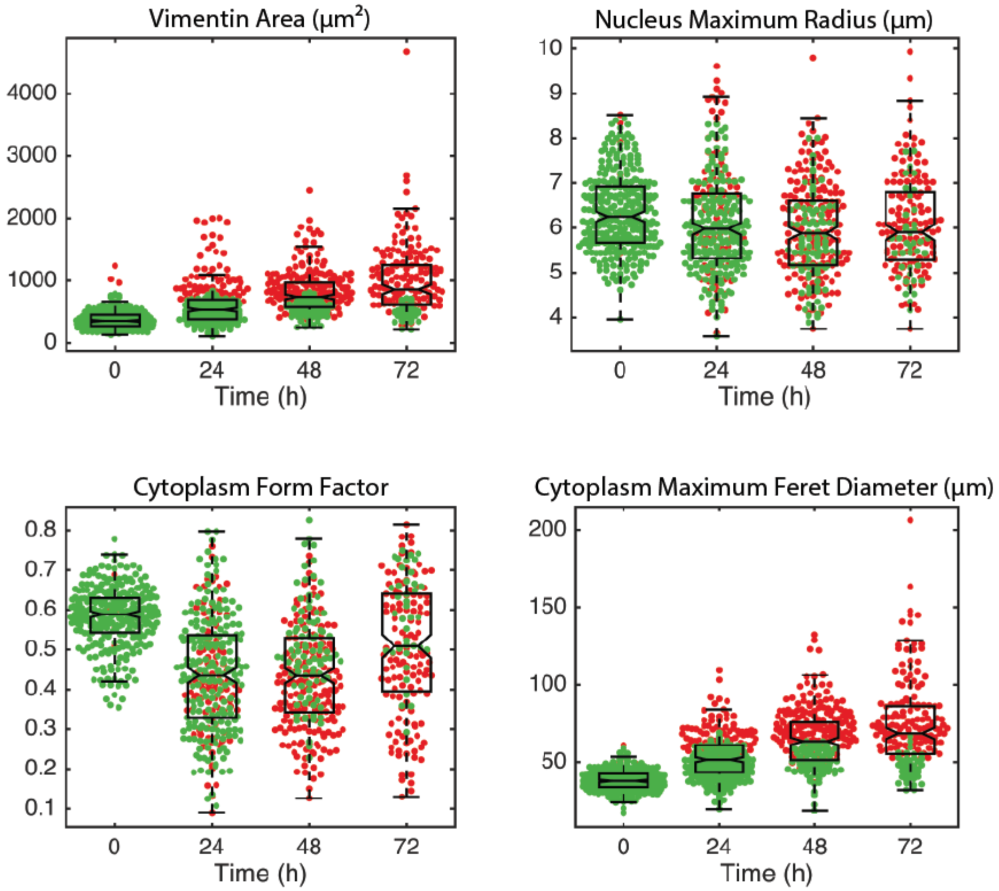


Figure S12. Boxplot of selected metrics with duration of TGF-β1 induction. Boxplots overlaid with data points representing individual cell values for each of the selected shape metrics for untreated cells (control) and TGF-β1 treated cells for the indicated durations. Individual cells were classified as epithelial (green) or mesenchymal (red) by GMM.

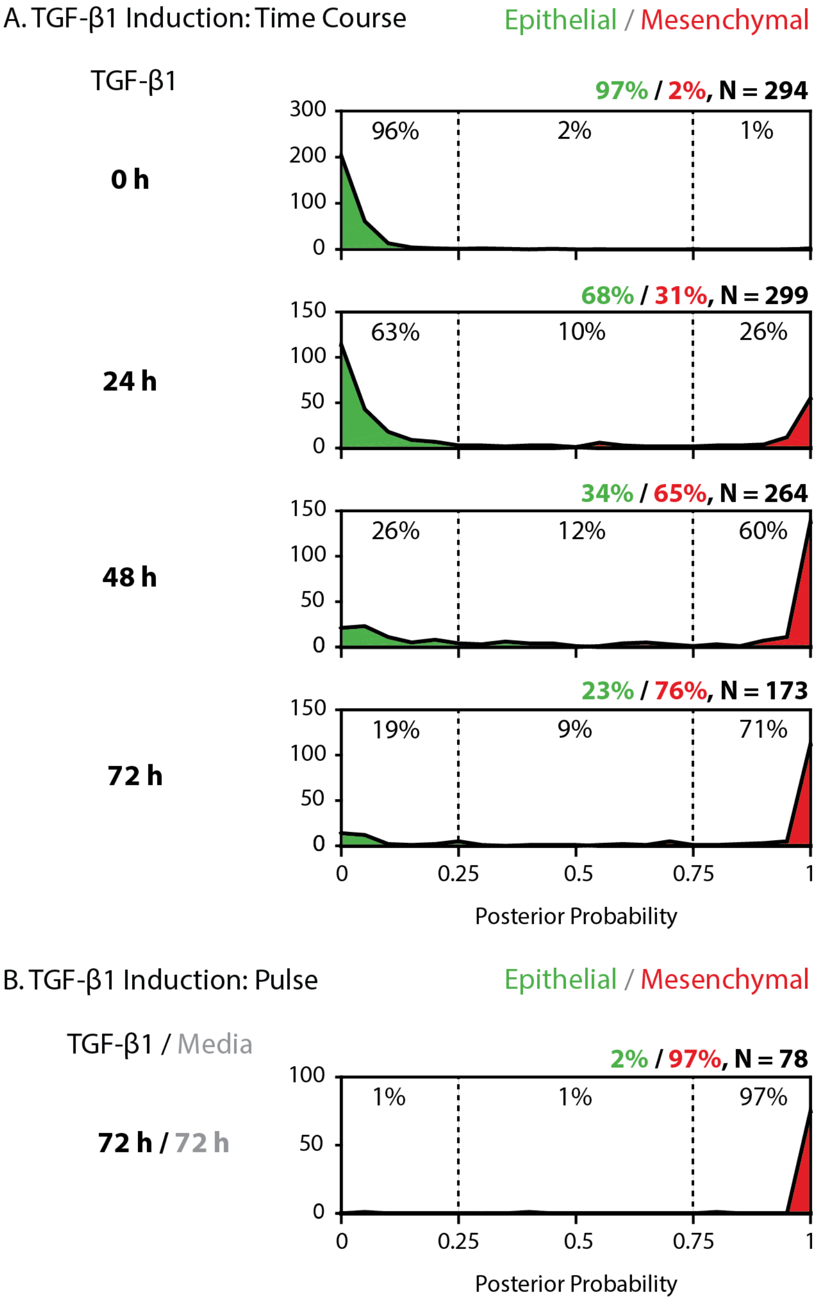
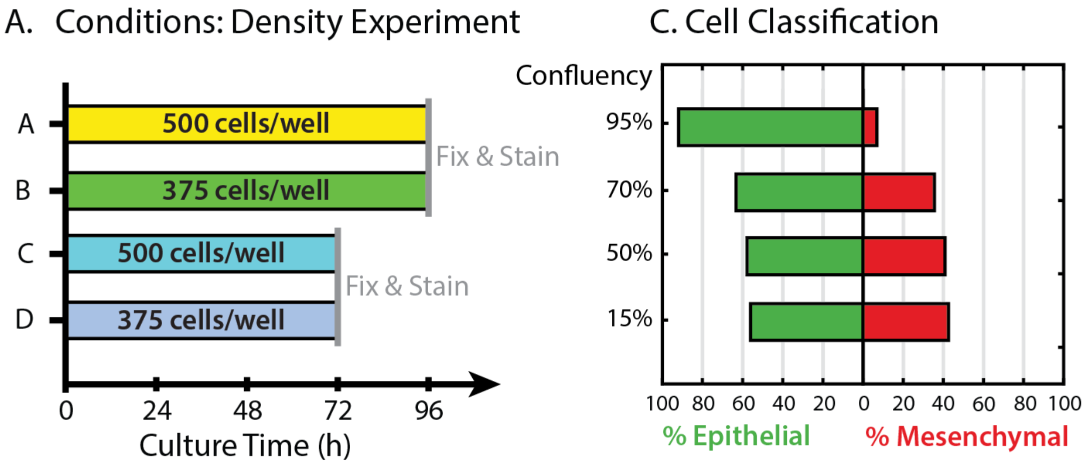


Figure S13. Posterior Probabilities of Gaussian Mixture Model for TGF-β1 induction. Posterior probability distributions depicting the conditional likelihood that a cell is mesenchymal (p=1). Percentage of cells classified as epithelial (green), mesenchymal (red) and total number of cells analyzed per conditioned (N) are displayed above graphs. For p = 0 to 0.25 and 0.75 to 1, there is a high probability that an individual cell is epithelial (green) and mesenchymal (red), respectively, with relative percentages displayed within graphs. The middle percentage where p = 0.25 to 0.75 represents the fraction of cells with less definitive classification. (A) For the time course of TGF-β1 induction and (B) for the TGF-β1 pulse, withdrawal condition.



B. Table of Cell Densities for Analyzed Conditions

Condition	# Cells/20x field	Cells/cm²	Description	Estimated Confluency
A	660	1.17E+05	monolayer	95%
B	353	6.23E+04	subconfluent	70%
C	287	5.07E+04	cell clusters	50%
D	45	7.95E+03	sparse	15%

Figure S14. Schematic of experimental conditions and results for Density experiments. (A) Overview for the timeline of cell density experiments. All cells were seeded at t=0 h, while cells were plated at two different densities, either 375 or 500 cells/well, and were subsequently cultured for both 72 h and 96 h durations to yield conditions with a wide range of endpoint densities (high to low confluency: yellow to blue bars). At the completion of the time course, cells were fixed and subsequently immunostained for epithelial and mesenchymal biomarkers. (B) Table of quantified values for the number of cells per image, cells per surface area (cm^2), description of density, and estimated confluency of cells at the experimental endpoint; resulting densities ranged from 95 to 15% confluence. (C) Barplot of epithelial and mesenchymal fractions that result from the indicated confluency condition. The percentage of cells classified as epithelial (green) gradually decreases from 95 to 15% confluence, while the percentage of cells classified as mesenchymal (red) gradually increases between these conditions, with the lowest density having the largest portion of mesenchymal cells at approximately 43% of the population.

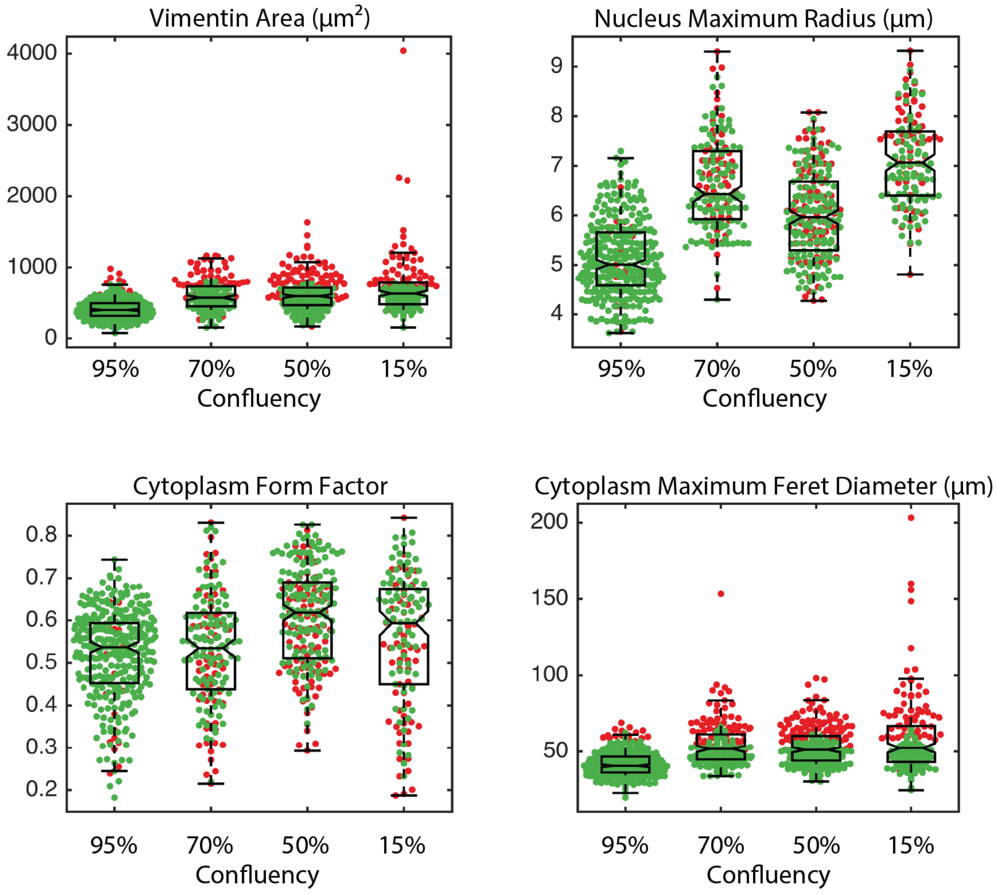


Figure S15. Boxplot of selected metrics with Density experiments. Boxplots overlaid with data points representing individual cell values for each of the selected shape metrics for cells at a range of densities (95 to 15% confluence). Individual cells were classified as epithelial (green) or mesenchymal (red) by GMM.

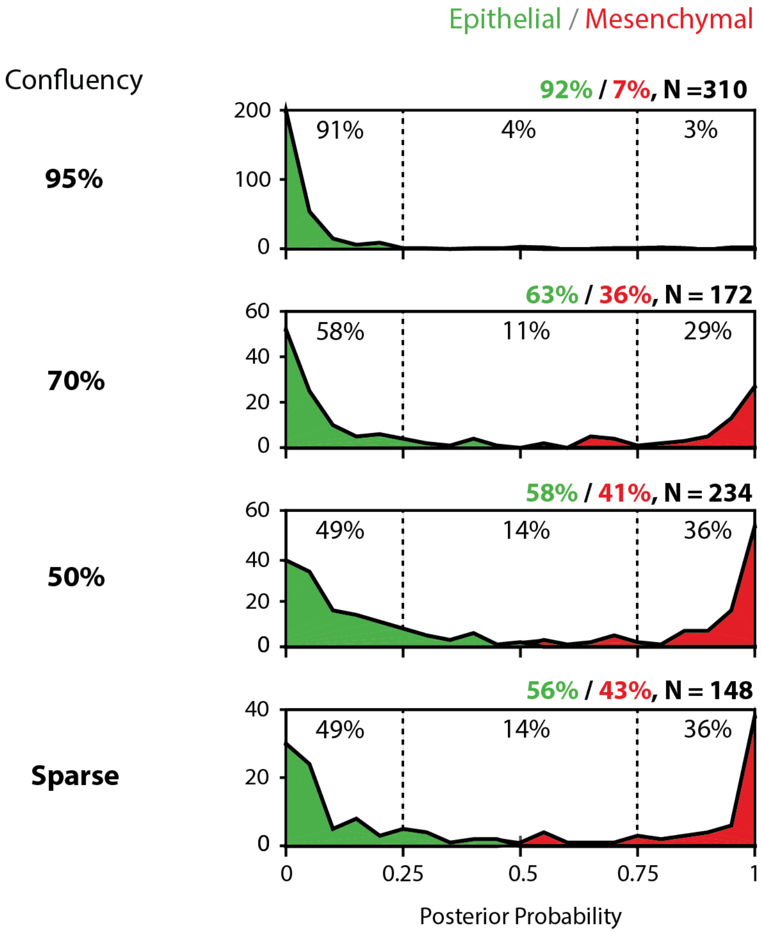


Figure S16. Posterior Probabilities of Gaussian Mixture Model with Density. Posterior probability distributions depicting the conditional likelihood that a cell is mesenchymal (p=1) for the indicated confluencies. Percentage of cells classified as epithelial (green), mesenchymal (red) and total number of cells analyzed per conditioned (N) are displayed above graphs. For p = 0 to 0.25 and 0.75 to 1, there is a high probability that an individual cell is epithelial (green) and mesenchymal (red), respectively, with relative percentages displayed within graphs. The middle percentage where p = 0.25 to 0.75 represents the fraction of cells with less definitive classification.

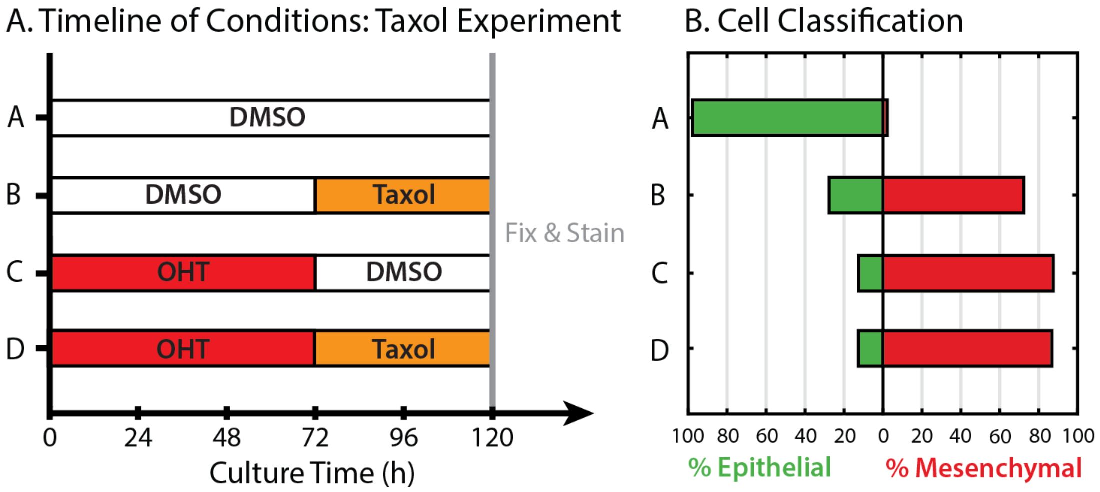


Figure S17. Schematic of experimental conditions and results for Taxol treatment. (A) Overview for the timeline of Taxol experiments. All cells were first treated with DMSO (uninduced; A, B) or 500nM OHT (preinduced; C, D) for 72 h, followed by treatment with DMSO (uninduced and preinduced controls; A & C, respectively) or 4nM Taxol (uninduced and preinduced Taxol; B & D, respectively) for 48 h. Cells were subsequently fixed and immunostained for epithelial and mesenchymal biomarkers. (B) Barplot of epithelial and mesenchymal fractions with the indicated conditions A-D. The percentage of cells classified as epithelial (green) largely decreases in uninduced cells treated with Taxol (A vs. B), while the percentage of cells classified as mesenchymal (red) greatly increases between these conditions, and is thus very similar to the preinduced OHT condition (C). At the population level, Taxol treatment in preinduced cells appears to maintain the mesenchymal phenotype (C vs. D).

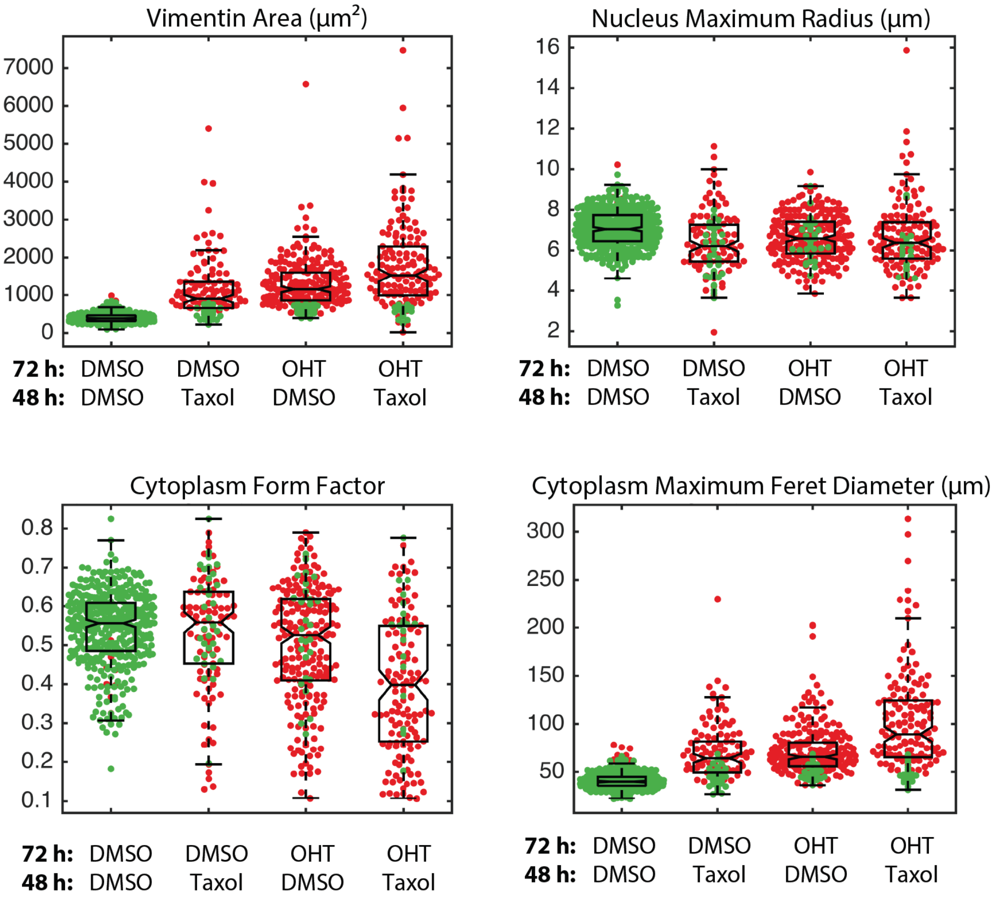


Figure S18. Boxplot of selected metrics with Taxol Treatment. Boxplots overlaid with data points representing individual cell values for each of the selected shape metrics for the indicated Taxol conditions (uninduced control, uninduced Taxol, preinduced control, and preinduced Taxol, from left to right). Individual cells were classified as epithelial (green) or mesenchymal (red) by GMM.

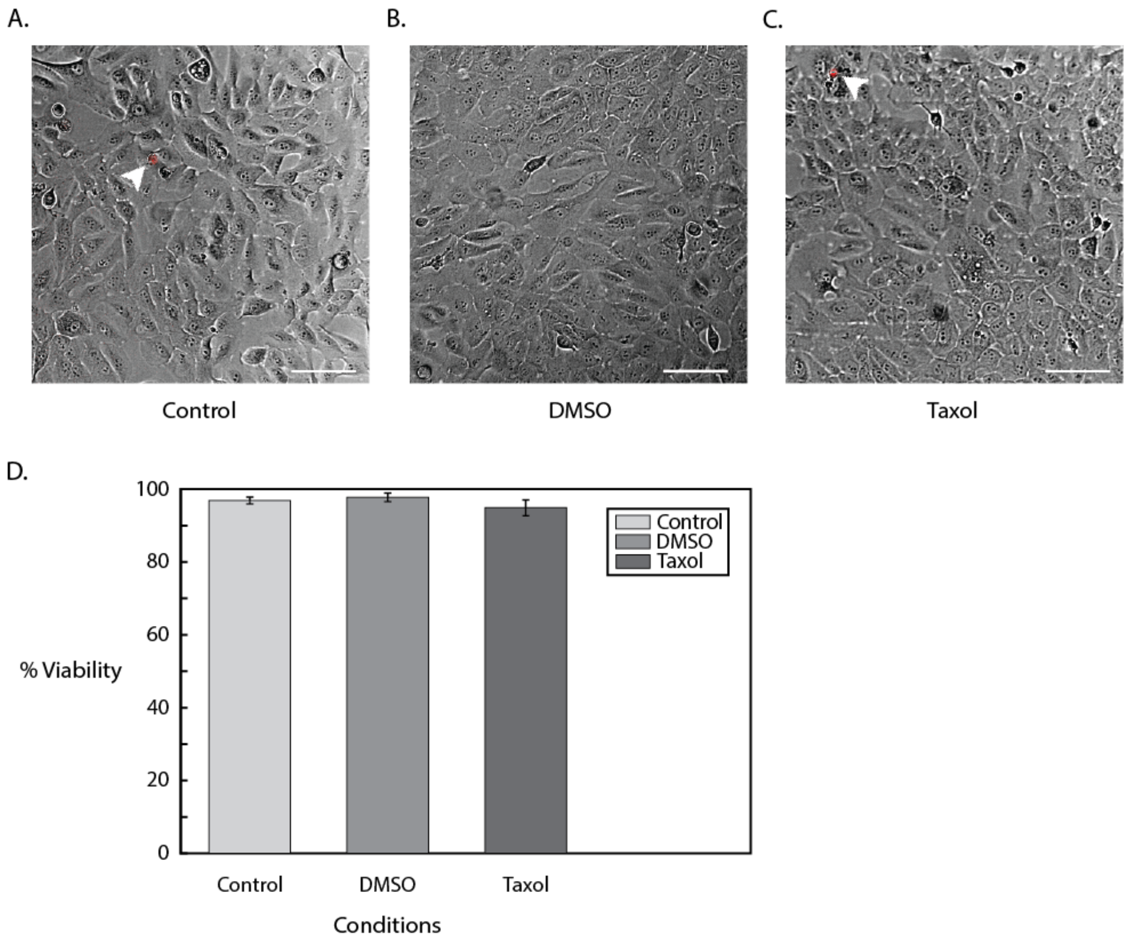


Figure S19. Live/Dead Staining and Percent Viability after Taxol treatment.
(A-C) Merged images for phase contrast and DRAQ7 (dead cell stain) of MCF-10A cells after 48 h of the following treatments: untreated (control, A), 0.05% DMSO (DMSO control, B), and Taxol treated (4nM Taxol, C). Scale = 100µm. Arrows indicate DRAQ7 positive staining cells (red). (D) Percent viability was also determined for triplicates of each condition using an automated program: Trypan Blue Viability Assay (Cellometer 1000 Auto, Nexcelom Bioscience). The percentage viability is comparable for each of the conditions.

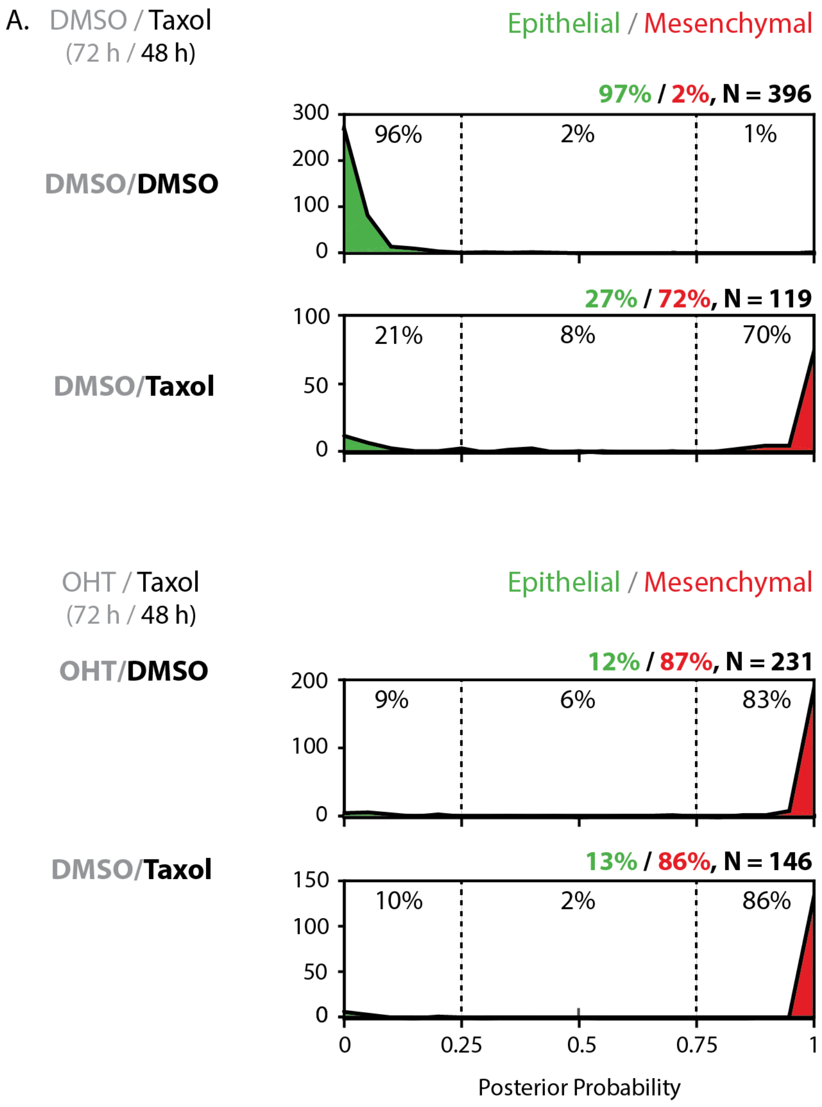


Figure S20. Posterior Probabilities of Gaussian Mixture Model with Taxol treatment. Posterior probability distributions depicting the conditional likelihood that a cell is mesenchymal (p=1). Percentage of cells classified as epithelial (green), mesenchymal (red) and total number of cells analyzed per conditioned (N) are displayed above graphs. For p = 0 to 0.25 and 0.75 to 1, there is a high probability that an individual cell is epithelial (green) and mesenchymal (red), respectively, with relative percentages displayed within graphs. The middle percentage where p = 0.25 to 0.75 represents the fraction of cells with less definitive classification. (A) For uninduced cells treated with DMSO (top) or Taxol (bottom) and (B) for preinduced (OHT) cells treated with DMSO (top) or Taxol (bottom).