SI Appendix for Cerchiari *et al.*

General materials and methods for cell sorting assays
A typical self-organization experiment begins by isolating luminal epithelial (LEP) or basal (myoepithelial, MEP) cells from uncultured primary human tissues or fourth passage cultured primary human mammary epithelial cells (HMEC) via fluorescently activated cell sorting (FACS). After sorting, these cells are reconstituted into aggregates in agarose or Matrigel using photolithographically defined microwells or by chemically-programmed assembly. Alternatively, cell-clusters can be preaggregated in agarose microwells and transferred to Matrigel (Fig. S1) before patterns of self-organization emerge (Fig. S2B). For all experiments, agarose was obtained from Allstar Scientific (490-050; Lot # 1004VIVCN) and used at a concentration of 3% (w/v) in Phosphate Buffer Saline (PBS). Different concentrations of agarose or different non-adhesive hydrogels such as polyethylene glycol (PEG) and polyacrylamide (PA) reproduce the same experimental outcome we describe without affecting the kinetics of cell-sorting. Agarose was chosen as the non-adhesive and non-fouling hydrogel material of choice because it is inexpensive and easy to prepare. Matrigel was obtained from BD Biosciences (354230; Lot # 07898) and was used without dilution. Different Matrigel Lot numbers or different formulations of adhesive gels (e.g. Collagen-Matrigel mixtures) reproduce the same experimental outcomes we describe but can significantly alter the kinetics of self-organization. In general, higher concentrations of Matrigel or added collagen increase the rate of self-organization. Differences in the kinetics of self-organization were also observed between different human specimens or between primary cells isolated from cultured HMEC or uncultured tissues. In general, cultured HMEC were found to self-organize into a stable conformation faster than uncultured primary HMEC. In adhesive microwells made out of PDMS (base:crosslinker ratio of 10:1) or functionalized agarose (3% w/v in PBS), MEP organize very rapidly to the tissue boundary but do not maintain their reconstituted architecture for more than 1 day. Unless otherwise noted, LEP and MEP are color-coded green and red respectively. The target diameter of the reconstituted cell-aggregates is always 100 μm.

Photolithography and micromolding
Using the Biomedical Micro- and Nano-Fabrication Center (BMNC) at UCSF, freestanding SU-8 features on silicon wafers were fabricated using standard photolithographic techniques. All recipes used for photopatterning were adapted from MicroCheM's technical specification sheets. For instance, in order to obtain circular microwells 120 μm in diameter and 80 μm deep, SU-8 2035 (MicroCheM) was spun on silicon wafers at a velocity of 500 rpm for 10 seconds followed by a 1250 rpm spin for 30 seconds. The wafer was then soft-baked for 5 minutes at 65 ˚C and for 10 minutes at 95 ˚C, UV-exposed in contact mode with an exposure energy of 215 mJ/cm^2 (through a photo-mask we designed in AutoCAD and purchased from Outputcity Co.), post-exposure baked for 5 minutes at 65 ˚C and for 10 minutes at 95 ˚C, and developed in SU-8 developer (MicroCheM) for at least 20 minutes. The patterned substrate was finally

washed with isopropanol/water and baked at 150 °C for 1 hour prior to measuring pillar heights using a stylus-profilometer (Ambios XP2) and visualizing the quality of the patterned wafer by SEM (NovelX mySEM). The silicon master was used to create PDMS micropillars by pouring a Sylgard 184 (Silicone Elastomer Kit, Dow Corning) onto the patterned wafer using a base:crosslinker ratio of 10:1. After degassing and curing at 60 °C overnight, the molded elastomer was finally peeled off the substrate and used to imprint agarose wells via micromolding a solution of 3 % agarose in 1X PBS (w/v).

HMEC material
Cultured normal finite lifespan primary HMEC were obtained from reduction mammoplasty tissue of a 19 year old individual (specimen 240L) and cultured in M87A medium with cholera toxin and oxytocin as previously described (1). All cultured primary HMEC were used for experiments at passage 4. Uncultured primary HMEC were directly isolated from frozen organoids of 23, 16, and 19-year old individuals (specimens 59L, 160L, and 168 respectively). All HMEC used in this study were provided as de-identified human samples by Dr. Martha Stampfer (Lawrence Berkeley National Laboratory).

Isolating LEP and MEP from primary HMEC and cultured primary HMEC
Primary human mammary epithelial cells at passage 4 were established and maintained in M87A medium according to previously reported methods (1). To lift the cells prior to sorting, a 10 cm dish was rinsed with PBS (Calcium and Magnesium free) with 0.04% EDTA and then incubated with PBS (Calcium and Magnesium free) with 0.04% EDTA at 37 °C until cells were rounded with a few detaching from the substrate. The PBS (Calcium and Magnesium free) with 0.04% EDTA was then transferred to a collection tube and replaced with 1.0 ml of 0.05% trypsin for 30-60 seconds at 37 °C. The flask was then tapped sharply to dislodge the remaining cells and all cells were collected in M87A medium. Cells were counted and pelleted at 160 x g for 4 minutes. Pelleted cells were washed once more with PBS and resuspended in growth medium at a concentration of 10^7/ml. Fluorescently tagged antibodies for CD10 and CD227 were then added to the cells at the indicated dilution (Table S1) and incubated for 30 minutes on ice. Finally, labeled cells were washed three times with PBS to remove unbound antibody and resuspended in FACS buffer (PBS containing 2 % BSA, 1 mM EDTA, and 1 μg/mL DAPI). Cells were then sorted on a BD FACSAria II controlled by FacsDiva software. LEP were gated as a population of CD227+/CD10- (Muc1+/Calla-) cells, MEP were gated as CD227-/CD10+ (Muc1-/Calla+) cells as shown in Fig. S1A. DAPI(+) cells were not collected. For uncultured primary HMEC, cryogenically preserved organoids were thawed and collected into a 50 ml tube. The organoids were then washed in PBS to remove freezing medium and resuspended in 5 ml of 0.25 % trypsin with saline and EDTA. Organoids were incubated with gentle agitation at room temperature for 10 minutes and then shaken vigorously for 30 seconds. 10 ml of cold M87A medium was added to inhibit trypsin activity and the cell suspension was filtered through a 40 μm mesh cell strainer to remove aggregates and debris. Cells were pelleted at 300 x g for 4 minutes and resuspended at a concentration of 10^7/ml in ice cold M87A medium, stained

with CD227 and CD10 antibodies as described above, and FACS purified as shown in Fig. S1B.

Isolating luminal and basal cells from human prostate tissue

All primary human prostate epithelial cells (HPEC) were obtained from the Urology Department of Stanford University following an Institutional Review Board approved protocol and in collaboration with the Laboratory of Dr. Donna Peehl. The procedure for isolating and reconstituting luminal and basal epithelia from primary prostate tissue was similar to the procedure used for isolating and reconsituting primary mammary tissue but required different reagents, antibodies, media-formulations, and an additional surgical step designed to isolate healthy human prostate tissue from cancerous tissue. All procedures were performed sequentially immediately after the prostectomy surgery. For example, for the benign left side of prostate sample MB (right side bearing a Gleason Score 3+3 tumor), 8 mm cores were removed and submerged in ice-cold HEPES-buffered saline (HBS) using an automated coring device (Alabama Research and Development, Mundford, AL) under aseptic conditions. A Krumdieck tissue slicer (Alabama Research and Development) was used to prepare precision-cut 300 μm thick tissue slices as previously described (2). 80 tissue slices were first washed with HBS in a sterile 5 mm dish, and then Collagenase (200 U/ml) was added in addition to MCDB 105 medium with 50 nM R1881 (Sigma), 10 μM Y27632 (Sigma) and 2 U/ml Dnase I (Amp. Grade, Invitrogen) (2). The tissue was minced with sterile scissors into small pieces (<1 mm^3) and pipetted vigorously. While in collagenase digestion medium, the sample was transported on ice from Stanford to UCSF. Immediately upon arrival to UCSF, the chunks of tissue were inserted in a rotator shaker and incubated at 37 ˚C (Note: For this particular sample, there were eight 15 mL conical tubes with 10 tissue slices/tube). After overnight digestion, the tissue suspension in each conical tube was pelleted and incubated with 5 mL of 0.2 % Trypsin/0.2 % EDTA (per tube) for 10 min at 37 ˚C before quenching with 10 mL of soybean trypsin Inhibitor (Cascade Biologics). The digested tissue was then strained through a 40 μm filter, pelleted, and resuspended in HEPES buffer. All tissue suspensions were combined in a single conical tube and placed on ice. At this point the cell count was approximately 2.3 x 10^7. These cells were pelleted and resuspended in 2.3 mL of DMEM (for an approximate concentration of 1 x 10^7/mL). 50 μl of the cell suspension were dispensed in 3 FACS tubes and incubated without antibody (unstained "control"), 2.5 μL of APC-conjugated CD44 antibody, or 10 μL of FITC-conjugated CD57 antibody (see Table S1 for details). The remaining cell solution was incubated with both antibodies (460 μL of anti-CD57 and 115 μL of anti-CD44) following the manufacturers' suggestions for the desired concentration per test. After 30 min incubation on ice, the cells were diluted in DMEM (0.5 mL for the controls and 20 mL for the double stained cells), pelleted, and resuspended in 0.5 mL (for controls) or 1 mL (for double stained cells) of FACS buffer (1:1 DMEM/F12 with 1 ug/mL DAPI). The cells were finally strained one last time through a 40 μm filter and brought to UCSF's Laboratory for Cell Analysis (LCA) for sorting on a FACS AriaII instrument as shown in Fig. S1C.

Cell staining using cytosolic or membrane dyes
For cell-substrate contact angle measurements, sorted LEP or MEP in 1 mL of PBS were separately stained with 5 µL of either DiI (Invitrogen Vybrant Multicolor cell-Catalog Kit V22889) for MEP or 5 µL of DiO (Invitrogen Invitrogen Vybrant Multicolor cell-Catalog Kit V22889) for LEP for 10 minutes at 37 ˚C, pelleted, and resuspended in M87A medium. For all other assays, sorted LEP or MEP were suspended in 10 mL of PBS and incubated with 1 µL of either 10 µM Cell Tracker Green (CTG, Invitrogen) for LEP or 1 µL of 10 µM Cell Tracker Red (CTR, Invitrogen) for MEP for 5 minutes at 37 ˚C, pelleted, and resuspended at a concentration of 10^6/mL in M87A. Although the membrane dyes gave crisper images than the cytosolic trackers, these were observed to be more toxic. Therefore, DiI and DiO were only used for our cell-substrate contact angle assays.

Chemically Programmed Assembly (CPA) of cells in Matrigel
Chemical aggregation of LEP and MEP, and full encapsulation into Matrigel gels was performed using an adaptation of previously reported methods (3-5). Briefly, one of two complementary single stranded DNA molecules bearing a 5'-membrane anchor were introduced to the surfaces of LEP and MEP cells. Mixing of cell populations bearing complementary sequences on their cell surfaces triggered their rapid aggregation allowing them to be embedded in Matrigel. DNA glue was removed by treatment with DNAse. Upon culture at 37 ˚C the MEP and LEP self-organize as shown in Fig. S1D.

Loading and visualizing cells in agarose
Homotypic or heterotypic populations of stained or unstained cells were centrifuged into microwells at 160 x g for 4 minutes and at a concentration of 10^6/mL. Excess cells were then washed away with culture medium and the remaining physically confined cell-aggregates were monitored for self-organization or circularity by time-lapse microscopy or by immunofluorescence as schematically illustrated in Fig. S1E and F.

Agarose to Matrigel transfers
To transfer cell-aggregates from agarose to Matrigel, aggregates were allowed to compact into scrambled cell-clusters (3 hours for HMEC and 1.5 days for primary cells) before carefully resuspending them in 15 mL of PBS. Then, after allowing the clusters to sediment to the bottom of the collection tube, the excess volume was discarded and the clusters were transferred to a thin layer of solid Matrigel, Clusters were allowed to settle onto the gel for 10 minutes. Finally, an additional layer of liquid Matrigel was overlaid on top and allowed to set for at least 1 hour at 37 ˚C before adding warm M87A medium (Fig. S1E).

Immunofluorescence
All microtissues were fixed with 4% PFA for 20 minutes and then incubated in blocking buffer (10 % heat inactivated goat serum in PBS + 0.5 % Triton X-100) at 4 ˚C for at least 1 day. Primary antibodies were diluted in blocking buffer and added to the sample as summarized in Table S1. After at least one day incubating at 4 ˚C with the primary

antibody, microtissues were washed several times with PBS + Triton-X for at least one day and incubated with alexa-conjugated secondary antibodies diluted 1:200 in blocking buffer for approximately 1 day. All sample were extensively washed with PBS + Triton-X + 1 ug/mL DAPI before imaging.

Image acquisition

All confocal microscopy images were acquired using a spinning disk confocal microscope (Zeiss Cell Observer Z1 equipped with a Yokagawa spinning disk and running Zeiss Zen Software). All other images were acquired using an inverted epifluoresent microscope (Zeiss Axiovert 200M running SlideBook software). For all assays, the use of long working-distance objectives facilitated data collection.

Quantifying self-organization

To quantify the self-organization of HMEC in agarose or Matrigel we adopted three approaches: visual scoring, average CTG and CTR intensity maps, and normalized keratin profiles.

1) Visual scoring: We inspected each reconstituted microtissue using phase contrast and fluorescence microscopy and counted the frequency of each tissue configuration (I-V) as shown in Fig. S2, B and E. The advantage of this method is throughput and the ability to appreciate subtle differences between similar architectures. The disadvantage is that this method is subject to human bias. To mitigate this limitation we assigned criteria for visual scoring as follows: (I) – correct: 75-100% MEP coverage of a LEP core; (II) – almost correct: 50-75% MEP coverage of a LEP core; (III) – almost inverted: 50-75% LEP coverage of a MEP core; (IV) – inverted: 75-100% LEP coverage of a MEP core; (V) – scrambled: 2 or more distinct MEP and LEP phases (Fig. S2B).

2) Average Cell Tracker Green (CTG) and Cell Tracker Red (CTR) intensity maps: To generate average intensity maps we imaged Cell-Tracker fluorescent intensity signals in multiple tissues using an epiflourescence microscope and averaged the fluorescence signal. This generated a 2D image that describes the average localization of green and red cytosolic dyes within a reconstituted tissue (Fig. S1D). These projections can be thought as heat maps that display in green and red the probability of finding LEP and MEP at that particular location in the reconstituted tissue. This quantification method does not introduce human bias but can be compromised by image-processing artifacts. More importantly, the final average fluorescent projections do not capture architectural differences that are revealed by inspection of the raw data. For instance, the average intensity maps for HMEC self-organizing in PDMS and Matrigel are similar, consistent with the basal positioning of MEP. However, the average intensity maps for these conditions do not reflect differences revealed by close inspection of individual tissues.

3) Normalized keratin profiles: To visualize the average localization of lineage-specific (i.e. luminal or basal) keratin intermediate filaments within a reconstituted tissue required collecting several high-magnification confocal images of fixed and stained structures,

cropping the images to the exact size of the spherical tissue, binarizing the fluorescent signal from the green (for K19) and red (for K14) channels, stacking (and resizing to smallest sample) all images collected, and plotting the normalized average (K14 and K19) signals as a function of the normalized radial distance from the tissue center (S2, F-I). Although this keratin profiling approach is the most quantitative method, it is time-consuming and may reflect biological processes (e.g. cell-differentiation) that relate to, but do not exclusively represent, the contribution of self-organization to structure formation. When possible, we use all three approaches to quantify self-organization. All image analysis was performed in FIJI.

Chemical functionalization of agarose microwells
Agarose microwells were washed 2 times with 0.1 M sodium carbonate at pH 8.5 (coupling buffer) and then 2 M sodium carbonate (activation buffer). Activation of hydroxyl groups was initiated by adding 5 M cyanogen bromide in acetonitrile for ~2 min at room temperature. Activated molds were subsequently washed with 1 mL of ice-cold deionized water and 1 mL of coupling buffer. Finally, wells were treated with 2 mg/mL BSA, 0.5 mg/mL ECM protein (e.g. Fibronectin1), or no protein in coupling buffer overnight at 4 ˚C. The wells were then given a fresh solution of protein in PBS for another incubation period of 24 hours at room temperature. On the day of the experiment, any residual reactive groups were blocked by treating the wells with 50 mM ethanolamine (pH 9.0) for 1 hour prior to extensive washing with M87A medium for several hours at room temperature. The ECM proteins used for the results presented in S3 were Fibronectin 1 (Fn1, Sigma F1141) and Collagen 1 (Col1, Advanced Biomatrix 5005-B).

Cell-Cell contact angle measurements
FACS-sorted HMECs stained with either Cell Tracker Green (CTG) or Cell Tracker Red (CTR) were centrifuged into small rectangular agarose microwells (20 μm x 40 μm x 15 μm) fabricated using the photolithography and micromolding methods described above. Wells were designed to accommodate two cells. After 4 hours in the wells, the cell doublets were imaged at 40X magnification and analyzed in FIJI in order to estimate the contact angle at the cell-cell interface. For these experiments we visually screened for microwells that contained only 2 healthy living cells of comparable size.

Cell-ECM contact angle measurements
Glass coverslips were submerged in a solution of Matrigel diluted in M87A (1:100 v/v) on an orbital shaker at room temperature. After overnight adsorption of Matrigel protein components, the coverslips were rinsed with PBS and incubated for a few hours with a solution of 4 nM QD605 (Invitrogen) in M87A to assist in resolving the interface. These coverslips were then washed with M87A and used as substrates for the spreading of FACS-sorted LEP and MEP stained with either DiI or DiO as previously described. The contact angle at the cell-substrate interface was estimated by analyzing 63X confocal Z-stacks (Fig. S4A) in FIJI or Zen software. For these experiments we visually screened for healthy living cells that were spreading onto the surfaces in a radially symmetric fashion

and that were not interacting with nearby cells. We observed that individual LEP cells interacted very weakly with the surface and, therefore, were frequently removed by washing or agitation. It is also important to note that for these experiments, the contact angle measured at the cell-substrate interface doesn't necessarily correlate with the spreading area seen in the XZ plane. As shown in Figure S4B, regions of cell-membrane can spread on the surface of the substrate without causing a pronounced change in overall cell shape in the XZ plane (Figure S4E). We therefore focused on the gross shape of the cell in the XZ plane rather than on the outline of the cell membrane. In other words, for all cell-substrate contact angle assays, we measure the contact angle as the angle formed between the substrate and the main body of the cell. This tended to concide with the inner position of intense actin staining seen in phalloidin stained cells (Figure S4E). This is a particularly important note for the contact angle measurements we did following Talin 1 knockdown in MEP because a fraction of these cells were observed to spread their membrane on the surface without affecting the overall shape of the cell in the XZ plane.

Cell-aggregate circularity and Cell-cluster spreading
For measuring the circularity of homogenous LEP or MEP clusters we loaded agarose microwells as described above and allowed the cells to interact overnight (or 1.5 days for uncultured primary cells) before collecting multiple images for each condition at 40X magnification (S4F). The contours of the microtissues were then extracted from the image and analyzed for circularity scores using FIJI's built-in circularity measurement tool. This function defined circularity (*C*) as follows:

$$c = 4\pi * \frac{(Area)}{(Perimeter)^2}$$

For the spreading assays of LEP and MEP-clusters over Matrigel-coated substrates we transferred the cell-aggregates after 4 hours (or 1.5 days for primary cells) in agarose microwells onto glass coverslips treated with Matrigel as described for cell-ECM contact angle measurements. These clusters were then allowed to spread on the substrate for 12-16 hours, before being fixed and stained for lineage-specific keratin markers and then analyzed for the extent of tissue-matrix interactions via phase contrast or confocal microscopy (S4G). While all MEP-clusters were observed to interact with the substrate, most LEP-clusters were so non-adherent that they were removed during the immunofluorescence washing steps. We therefore interpret our measurements as overestimates of LEP-ECM adhesion. Finally, it should also be mentioned that the spreading behavior of MEP-clusters onto the substrate is affected by how much time these cells are allowed to pre-aggregate in non-adhesive microwells. For fourth passage primary MEP, the cell clusters lose their ability to exchange cell-cell interactions in favor of cell-substrate interactions after 1 day of pre-aggregation.

Transcriptional profiling of HMEC MEP and LEP
On three separate days, batches of fourth passage HMEC were FACS-sorted into LEP and MEP populations. Immediately after sorting, RNA was isolated from each cell

population using the RNeasy Mini Kit (Qiagen 74104), converted to cDNA using a RT2 First Strand Kit (Qiagen 330401), mixed with QIAGEN's RT2 PCR master mix, and aliquoted across 2 SA-Biosciences PCR arrays (PAHS-146Z for cell-cell adhesion molecules and PAHS-013 for cell-ECM adhesion molecules); the target concentration of RNA loaded into each array was 2 µg. The plates were then analyzed on an ABI Vii7A RT-PCR instrument. Identical threshold values for all MEP (n=3) and LEP (n=3) plates were used in order to extract raw threshold cycles (Cts) from the data. These raw Ct values were used to calculate fold-changes using the $\Delta\Delta Ct$ method (http://www.sabiosciences.com/pcrarraydataanalysis.php). To normalize the data, we used the average of the arithmetic means of 3 genes: GAPDH, ACTB, and HPRT1. To interpret the data, the LEP population was used as the "control" (i.e. over or underexpression of MEP genes with respect to LEP genes) as shown in S5. Genes not amplified by cycle 35 were considered undetermined. P-values were calculated using a Student's t-test. Data analysis was performed using the online software provided by SA BioSciences(http://pcrdataanalysis.sabiosciences.com/pcr/arrayanalysis.php). Although several of the mRNAs identified as over or underexpressed by these qPCR screens were reflected at the protein level, some were not. For instance, at the RNA level, CDH1 (E-cadherin) was found to be expressed at very similar levels in MEP and LEP. However, for all of the cultured or uncultured primary cells we analyzed by flow cytometry, E-cadherin was always found to be preferentially expressed in the luminal lineage.

siRNA transfection and protein knockdown
Conditions for transfection of HMEC with siRNA were optimized using cells grown in 12 well plates and with siRNA targeting cell surface adhesion molecules α_6 or β_1 integrin (S6, A and B). Cells grown in 10 cm dishes were transfected with Smartpool siRNA (Dharmacon/Thermo) targeting Talin1, p120 catenin, or a non-silencing control (NSC) 72 hours prior to sorting (i.e. maximum knock-down). Transfections were performed with 75 nM siRNA using Dharmafect 2 transfection reagent following the manufacturer's protocol. Successful transfection was assessed using a siGLO Red Transfection indicator. Successful knock down of the target protein was confirmed by western blot and/or flow cytometery as shown in Fig. S6, A-D. The physical consequences of knockdown were quantified using contact angle measurements, circularity, cell-cluster spreading assays, or self-organization assays (Fig. S6). For contact angle measurements, we visually screened for cells that did not exhibit extensive membrane damage following the delivery of siRNA reagents. We noticed no appreciable differences in contact angle, circularity measurements, or self-organizing capacity between the non-silencing controls (NSC) and the untransfected control cells (control). Therefore, we use NSC and control cells interchangeably.

Western blots
Protein lysates were prepared from unsorted HMEC or from FACS purified MEP and LEP populations by lysis in RIPA buffer. 50 µg of protein was resolved on a 4–12 % polyacrylamide gradient gel and transferred to Immobilon-P membrane (Biorad). The membrane was blocked in TBST (25 mM Tris, pH 7.4, 137 mM NaCl, 2 mM KCl, 0.05%

Tween-20) containing 1 % nonfat powdered milk and incubated with antibodies against Talin1, or p120 catenin (Table S1). Binding of primary antibodies was detected using HRP-conjugated anti-mouse or anti-rabbit IgG and visualized by chemiluminescence on a Versadoc imaging station (Biorad) or by film.

Characterization of integrin expression after Talin 1 knockdown
HMEC were transfected with siRNA targeting Talin1 or non-silencing control. After 72 hours the cells were trypsinized and prepared for FACS analysis. Cells were incubated with PE-labeled anti- α_6 integrin, PE-labeled anti- β_1 integrin or PE labeled isotype control for 30 minutes on ice. Stained cells were washed with PBS and the fluorescence signal measured on a FacsCaliber instrument. Collected data was analyzed using FloJo software (S6, L and M).

Cell-ECM spreading and self-organization with Y27632 and Latrunculin A
For the Cell-ECM spreading assays and the self-organization assays reported in Fig. S6 (N and O), MEP and LEP were prepared as described above but incubated with either 50 μM Y27632 (Calbiochem) or 2 μM Latrunculin A (Sigma L5163).

Estimating the change in surface energy upon cell-cell and cell-ECM spreading

Summary of the approach: To examine the robustness of tissue self-organization to variable and plastic interfacial properties of cells, we use an energy-based (rather than mechanical) model of tissue self-organization. The goal of the below analysis, therefore, is to estimate and then place the relative energies of nearest neighbor interactions (e.g. cell-cell cohesion and cell-ECM cohesion) on a common scale for use in the computational model. For this analysis, we make a number of assumptions. We assume that a cell has characteristic interfacial tensions (units of energy/area) that are unique to each type of interface the cell is in contact with. By dimensional analysis, we also assume that the surface energy of the cell can be expressed as the sum of the products of each interfacial area with the interfacial tension characteristic of that area. We ignore the microscopic and molecular mechanisms giving rise to each interfacial energy, as well as second order effects that may arise due to the fact that multiple cells are interacting at once, and instead focus on the overall change in the energy of the system upon the formation of a single cell-cell or cell-ECM contact from dissociated cells. We treat the system as arriving at a steady state within a few hours.

At the steady-state, the balance of interfacial tensions of cells with other cells, surfaces, and media can be estimated using Young's equation. Ratios of interfacial tensions can be converted to relative energies of pairwise cell-cell and cell-ECM interactions by measuring contact angles and contact areas at each interface. To simplify this analysis we use previous AFM measurements of cortical elasticity (1) in order to assume a constant and similar surface-tension at the cell-medium interface of MEP and LEP. To estimate the change in surface energy upon MEP and LEP spreading on surfaces, we use Matrigel coated glass as substrates. Although the mechanics of glass are quite different than a 3D

Matrigel slab, the rigidity of the substrate allows key simplifications to our analysis because glass is a non-deformable substrate. By also assuming a constant cell surface area during the duration of the assays (4 hours), we estimate that, for primary human mammary epithelial cells at passage 4, the change in surface energy (work of cell-cell cohesion) between a MEP-MEP and a LEP-MEP is approximately 6.4 and 2.8 times stronger than the work of cell-cohesion of forming a LEP-LEP doublet. We also estimate that the change in surface energy of MEP-ECM spreading (on matrigel coated glass) is at least 2 times stronger than the work of MEP-MEP cohesion, and that the change in surface energy upon LEP interaction with the ECM is approximately 0 since LEP do not stick to nor spread on Matrigel coated surfaces. Importantly, we assume that this change in surface energy upon MEP-ECM spreading is an overestimate due to the high rigidity of glass relative to matrigel, which would be predicted to increase cell spreading. The details of our analysis follow.

Assumptions and details
We reason that the change in energy associated with forming a cell-cell interface (W_{cc}) is the difference in energy before (E_1) and after (E_2) contact formation.


$$W_{cc} = \Delta E = E_2 - E_1 \tag{1}$$

The energy of the system before (E_1) and after (E_2) cell-cell contact can be expressed as the product of the interfacial energies and the areas of each interface:

$$E_1 = \Upsilon_{cm(1)} * A_{cm(1)}$$
$$E_2 = \Upsilon_{cm(2)} * A_{cm(2)} + \Upsilon_{cc(2)} * A_{cc(2)}$$

Where Υ is the surface tension at the cell-medium (Υ_{cm}) or cell-cell (Υ_{cc}) interface in dimensions of energy/area, and A is the area of contact (normalized by the total surface area of a cell before and after contact) at the cell-medium (A_{cm}) or cell-cell (A_{cc}) interfaces.
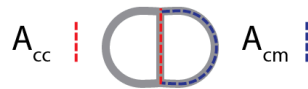
To simplify analysis, we assume that the cells have a constant surface tension at the cell-cell or cell-medium interface throughout the assay (4 hours). We also assume that the interfacial tension at the cell-cell interface is equal at either side of a cell-cell junction, an assumption supported by the straight interfaces formed in both homo and heterotypic cell doublets (Fig 1G). We also assume that the surface tension at the cell-cell interface can be expressed as a function of the surface tension at the cell-medium interface (2) based on geometric considerations and Young's equation. In the context of the energy-based

model, we assume that the formation of additional cell-cell interfaces will have similar interfacial tension as the first cell-cell interface.
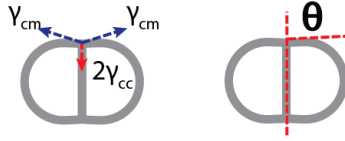
(Young) $$\Upsilon_{cc} = \Upsilon_{cm} * \cos\theta_{cc} \tag{2}$$

Combining (1) and (2) generates a new equation (3) that expresses the energy change (or work of cohesion) upon forming the interface between two cells.

$$W_{cc} = \Upsilon_{cm} * [A_{cm(2)} + \cos\theta_{cc} * A_{cc(2)} - A_{cm(1)}] \tag{3}$$

Based on previous measurements of the cortical elasticity of LEP and MEP by AFM (6), we assume that these cells have a similar surface tension at the cell-medium interface (i.e. $\Upsilon_{LEP-m} = \Upsilon_{MEP-m} = \Upsilon_{cm}$). Therefore, we can relate MEP cohesion to LEP cohesion as a function of measurable contact angles and contact areas:

$$\frac{W_{MEP-MEP}}{W_{LEP-LEP}} = \tag{4}$$
$$\frac{A_{MEP-medium(2)} + \cos\theta_{MEP-MEP} * A_{MEP-MEP(2)} - A_{MEP-medium(1)}}{A_{LEP-medium(2)} + \cos\theta_{LEP-LEP} * A_{LEP-LEP(2)} - A_{LEP-medium(1)}}$$

To simplify, we assume that cell surface area does not change significantly over the course of the assay and that the initial shape of the cell in the non-spread state has the geometry of a sphere.

$\theta_{MEP-MEP}$ and $\theta_{LEP-LEP}$ are then determined by image analysis of cell-doublets. The average of these values is used to determine $\cos\theta_{MEP-MEP}$ and $\cos\theta_{LEP-LEP}$ and used in combination with (4) to approximate the relative energetics of MEP-MEP and LEP-LEP cohesion.

We note that MEP and LEP form a straight interface at their cell-cell junctions indicating balanced mechanical coupling across the heterotypic interface. Therefore, a similar analysis was used to estimate the work of heterotypic MEP-LEP cohesion and are summarized as follows.

	$A_{cm(1)}$	$A_{cm(2)}$	$A_{cc(2)}$	$\cos(\theta_{cc})$	$W_{cc}/W_{LEP\text{-}LEP}$
MEP-MEP	1.000	0.795	0.205	0.358	6.417
MEP-LEP	1.000	0.857	0.143	0.602	2.775
LEP-LEP	1.000	0.930	0.070	0.707	1.000

The energy of forming a cell-ECM interface was estimated using similar reasoning: the energy released upon formation of a cell-matrix interface (W_{cx}) is the difference in energy states before (E_3) and after (E_4) interface formation.
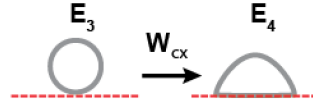
$$W_{cx} = \Delta E = E_4 - E_3 \tag{5}$$

The energy of the system before and after cell-ECM interface formation can be expressed as follows:

$$E_3 = \Upsilon_{cm(3)} * A_{cm(3)} + \Upsilon_{xm(3)} * A_{xm(3)}$$
$$E_4 = \Upsilon_{cm(4)} * A_{cm(4)} + \Upsilon_{cx(4)} * A_{cx(4)} + \Upsilon_{xm(4)} * A_{xm(4)}$$

Where Υ is the surface tension at the cell-medium (Υ_{cm}), matrix-medium (Υ_{xm}), or cell-matrix (Υ_{cx}) interface and A is the area of contact (normalized by the total surface area of a cell before and after contact) at the cell-medium (A_{cm}), cell-matrix (A_{cx}), or matrix-medium (A_{xm}) interface.

We similarly express the surface tension at the matrix-medium interface as a function of the surface tension at the cell-medium interface (6) based on geometric considerations and Young's equation, which are much simplified on non-deformable substrates such as glass because Υ_{cx} and Υ_{xm} act in opposing directions.
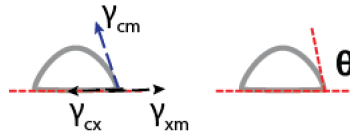
$$\Upsilon_{xm} = \Upsilon_{cx} + \Upsilon_{cm} * \cos(180° - \theta_{cx}) \tag{6}$$

To further simplify the analysis, we assume that the change in area at the matrix-medium interface is equal to the area formed at the cell-matrix interface and that cell-matrix spreading reaches the steady state by 4 hours. Combining (5) and (6) generates a new equation (7) that expresses the change in surface energy upon cell ECM spreading.

$$W_{cx} = \Upsilon_{cm} * [A_{cm(4)} - A_{cm(3)} - A_{cx(4)} * \cos(180° - \theta_{cx})] \tag{7}$$

Using similar reasoning as with (4) allows us to express the relative change in surface energy of a MEP-matrix spreading with respect to a LEP-LEP cohesion as a function of measureable contact angles and contact areas.

$$\frac{W_{MEP-matrix}}{W_{LEP-LEP}} \tag{8}$$

$$= \frac{A_{MEP-medium(4)} - A_{MEP-medium(3)} - A_{MEP-matrix(4)} * \cos(180° - \theta_{MEP-matrix})}{A_{LEP-medium(2)} + \cos\theta_{LEP-LEP} * A_{LEP-LEP(2)} - A_{LEP-medium(1)}}$$

The surface area in contact with medium was calculated by assuming the geometry of the spread cell to be that of a cone with a radius and slant height equal to the average radius and slant height of a MEP spread on the substrate as measured by confocal microscopy. The surface area in contact with ECM was calculated by measuring the perimeter of cells in contact with the substrate by analyzing Z-stack confocal images of spread MEP. This surface area was normalized to the average surface area of a cell modeled as a cone.

$\theta_{MEP-matrix}$ was measured by image analysis of Z-stack confocal reconstructions of cells spread on ECM-coated surfaces. The angle was measured at the point of cortex engagement (bright actin staining in confocal image S4E. The average of these values was substituted into (8) to estimate the relative energy of interaction between MEP-ECM and LEP-LEP based on the average cell-cell contact angle, cell-cell area, and cell-medium area estimated as discussed above and summarized below. We note that the vast majority of LEP do not adhere nor spread on the Matrigel-coated substrate. Therefore, we estimate the work of cell-ECM cohesion for LEP and the matrix to be zero ($W_{LEP-matrix} = 0$).

	$A_{cm(3)}$	$A_{cm(4)}$	$A_{cx(4)}$	$\cos(180-\theta_{cx})$	$W_{cx}/W_{LEP-LEP}$
MEP-X	1.000	0.532	0.468	0.193	27.222
LEP-X	N/A	N/A	N/A	N/A	0.000

To test the effect of our assumptions on the relative values of cell-cell cohesion and cell-ECM cohesion, we re-evaluated the values reported above based on different assumptions. For instance, if we estimate the total area of a cell before spreading as the area of a sphere but the total area of the cell after spreading as the area of a cone the cell surface area is increased by approximately 27 %. Although this would imply that MEP-ECM cohesion is 21.3 (and not 27.2) times stronger than LEP-LEP cohesion, the MEP-matrix interaction is still 3.4 times stronger than MEP-MEP cohesion. Since cell spreading on the tissue boundary can dominate self-organization even when it is considerably weaker than cell-cell cohesion, these values are still strikingly higher than what would be necessary to drive robust self-organization. In the computational model, we use a value for $W_{MEP-ECM}$ = 16 as a conservative estimate for MEP-Matrigel adhesion since the glass substrates used in these measurements are considerably stiffer than Matrigel (Guo WH *et al* Biophys J. 2006). To test the consequences of our assumptions regarding cell-medium interfacial tension, we note that the energy of MEP-MEP cohesion and MEP-ECM cohesion relative to LEP-LEP cohesion scale directly with the ratio of MEP-medium and LEP-medium interfacial tension. Thus, if MEP have higher interfacial

tension than LEP at the cell-medium interface, we would expect the difference in the energies of cell-cell cohesion to increase accordingly.

We point out that MEP-ECM spreading (cell-ECM cohesion) is likely mediated through a modulation of cortical tension at the cell-ECM interface. Several lines of evidence support this notion. First, a bright ring of actin is observed at the cell cortex just inside of the protruding edge of spreading cells. Second, KD of Talin1, a key adapter for linking integrins to the contractile cell cortex, blocks cells spreading and self-organization without affecting cell-surface expression of integrins (Fig. S6, L and M). Third, small molecules that interfere with actomyosin contractility (e.g. Latrunculin A and Y-27632) block self-organization (Fig. S6O). Given that perturbation to cell-cell cohesion in LEP and MEP do not affect self-organization (Figures 2-4), these data suggest that the effects of Latrunculin and Y-27632 on cell-sorting are primarily actining on MEP-ECM cohesion. Despite these preliminary data, future studies will be necessary to more precisely dissect the precise relationship between ECM elasticity, cell-cell, cell-ECM, and cell-medium interfacial tensions on the energy of cell-cell and cell-ECM contact formation.

Cell Sorting Model Background

In the main text of the paper, we analyze a mathematical model of tissue self-organization through cell sorting and differential adhesion to study the impact of a strong interaction between one cell type, the MEP cells, and the ECM on the robustness of self-organization to variation in cell-cell interactions. We show that a strong interaction between the MEP cells and the ECM tissue boundary can make self-organization of tissue structure robust to a) changes in adhesive interactions between cells (Fig. 2) and b) underlying variability in cellular interactions due to heterogeneity (Fig. 4).

We formulated a simple exact model of cell sorting inspired by models of Steinberg and Glazier (7,10). These classic models analyze self-organization of a tissue through cell sorting as an energy minimization problem analogous to well known problems in statistical mechanisms where an equilibrium macroscopic ordering of a system is achieved through local interactions between particles (Ising model, XY-model). The fundamental assumption in statistical mechanics inspired models of cell sorting is that the dynamics of cells within a tissue bring the system to a kind of equilibrium state that maximizes the total cellular cohesion by minimizing an effective internal energy.

Several different flavors of lattice cell sorting models have been proposed. Here, we develop a computationally efficient model that enables complete reconstruction of tissue phase diagrams (Fig. 2); analysis of how variability in interaction energies between cells impacts self-organization (Fig. 4); and analytic calculations (below).

Model Summary

First, we give a brief intuitive description of the model, and then in later sections we develop the model formally and state computational details. In our model, cells occupy sites on a square lattice and are of two types, LEPs and MEPs. The lattice is bounded by an ECM boundary. Each cell interacts with the cells or ECM at neighboring lattice sites, and these local interactions carry a local interaction energy E_{ij} (interaction energy between cells at sites i and j) that depends on the type of cell (or ECM) sitting at site i and j.

Local interaction energies are summed together across the lattice to generate a global interaction energy, E_{total}. Following Steinberg and Glazier, we find the equilibrium configuration of the tissue (minimizing E_{total}) by allowing cells to make stochastic transitions between neighboring lattice sites. The probability of a cell transitioning from one lattice site to the next increases if that transition will decrease E_{total}. We iterate this energy minimization process through thousands of local cell moves and accept the final configuration of the tissue.

To introduce variability into the model, we draw the local interaction energies between two cells (E_{ij}) from a probability distribution, so that MEP and LEP cells across the tissue will interact through differing E_{ij} values. In this case, the mean and variance of E_{ij} is fixed for

pair-wise cell-cell or cell–ECM interactions. For example, the mean and standard deviation of E_{ij} for MEP-ECM interaction is fixed, but individual values vary.

The key simplification of our model compared with Glazier is that cells occupy a single lattice site and do not have internal structure. In Glazier, cells are composed of a series of lattice sites, and this allows the Glazier model to allow cellular deformations to occur. However, this introduces considerable complexity and computational cost into the model.

The strength of our lattice model is two-fold. First, it allows comprehensive exploration of the tissue parameter space. We can adjust the interaction rules for all cell-cell and cell-ECM pairs (MEP-MEP,MEP-LEP,LEP-LEP,LEP-ECM, MEP-ECM) and ask how the tissue will self-organize for each rule set. Such comprehensive analysis is possible because we can computationally equilibrate each model tissue in a matter of seconds , and so we can explore thousands of different local interaction rules to generate the tissue phase diagrams shown in the text. Second, the lattice model is analytically tractable. Specifically, we can calculate E_{total} analytically for different tissue configurations and analytically explore the influence of a matrix interaction on the low energy tissue configuration. The analytic tractability allows us to make strong statements about the sufficiency of MEP-matrix interaction rules for driving formation of the correct configuration of the mammary tissue.

Detailed Description of Lattice Model
Formally, we consider a set of cells, c_i, that occupy sites on a square lattice. The lattice is enclosed by an irregular boundary that approximates the circular boundary provided by PDMS wells or the ECM gel in experiments. Cells come in two different types, LEP and MEP. Lattice sites occupied by ECM are modeled as being occupied by a cell of a third type but that are fixed on the lattice. The function $\rho(i)$ specifies the type of cell (or ECM) at site i, so that $\rho(i) \in (1, 2, 3)$. Here, we identify $\text{MEP} = 1, \text{LEP} = 2$, and $\text{ECM} = 3$.

Cells interact locally with neighboring cells through pairwise interaction energies that depend upon the type of each cell. Cells interact with all adjacent and diagonal nearest neighboring lattice sites (Fig. 2). We define a symmetric matrix W that contains the magnitude (by convention this is a positive number) of the cohesive interaction between cell types. For example, $W_{12} = W_{21}$ is the strength of the cohesion between the LEP and MEP cells, and W_{nm} is the cohesion of cell types n and m.

To make the analogy with statistical mechanics, we transform the cohesive interaction into an energy and define a local interaction energy, E_{ij}, between cells at nearest neighbor lattice sites, i, j, as:

$$E_{ij} = -W_{\rho(i)\rho(j)}, \tag{2}$$

where we have made $E_{ij} < 0$ so that we can preserve the analogy with physical systems that minimize energy.

The global cohesive energy of the cell aggregate, E_{total} is the sum of all local pairwise E_{ij}:

$$E_{\text{total}} = \sum_{<ij>} E_{ij}, \tag{3}$$

where this sum, $\sum_{<ij>}$ is taken over all nearest and next nearest pairs on the lattice (where $d(i,j) < \sqrt{2}$). Following Steinberg, we model the cell aggregate as dynamically finding a configuration that minimizes E_{total} and hence maximizes the total strength of cohesive cell interactions in the system.

Computational equilibration of tissues via Metropolis Algorithim

Computationally, we minimize E_{total} using the Metropolis-Hastings algorithm (8). At each iteration of the algorithm we select a random test cell and allow the cell to make a local test move uniformly at random in any lattice direction. For the test configuration, we calculate:

$$\Delta E_{\text{total}} = E_{\text{total}}(\text{test}) - E_{\text{total}}(0), \tag{4}$$

where $E_{\text{total}}(0)$ is the energy of the tissue prior to the test move, and $E_{\text{total}}(\text{test})$ is the energy following the test move.

If $\Delta E_{\text{total}} < 0$, we accept the test move, and if $\Delta E_{\text{total}} \geq 0$, we accept the move with probability:

$$p_{\text{move}} = \exp(-\frac{\Delta E_{\text{total}}}{\tau}), \tag{5}$$

where τ is an effective "temperature" that modulates the probability of accepting a test move that increases the global energy of the aggregate.

To achieve a minimum energy configuration without jamming, we use an annealing schedule where we first linearly and then exponentially decrease the temperature,τ, of the cellular ensemble as a function of time. In all runs of the model, we explore 2×10^5 single cell moves. The annealing strategy allows the cells to first explore a large area of configuration

space, and the structure is more finely refined at progressively lower temperature where move acceptance criteria are more stringent. Empirically, we found the results to be insensitive to the details of the annealing schedule. As an important note, our computational strategy is an equilibration strategy that seeks, but will not necessarily find, the lowest energy configuration of the tissue. Importantly, we do not allow the ECM to move. The ECM remains fixed during the entire process of energy minimization providing a rigid boundary condition for the cell aggregate.

A key simplification of our model is that we consider cells to be objects without internal structure. In other models in the literature (7), cells are composed of a set of lattice sites that are allowed to move. We selected a simpler modeling framework approximating each cell as an atomic unit, permitting a computationally efficient exploration of tissue phase diagrams as well as analytic calculations.

In figure S8A, we show a computationally generated tissue phase diagram for a two component LEP-MEP tissue in the regime, $W_{11} > W_{22}$. On the same plot, we show that phase boundaries predicted by Steinberg.

Introduction of variable interaction energies
In the modeling framework described above, cell-cell and cell-matrix interactions are stereotyped and fixed, so that cells of the same type are identical objects. In this framework, all MEP cells are identical, and all LEP cells are identical. In real tissue, cellular properties vary due to fluctuations in gene expression, cellular history, and local microenvironmental conditions. Therefore, we analyzed how variability in cell-cell interaction energies affects the tissue morphologies predicted by cell-sorting models. In the spirit of studies of quenched disorder in statistical mechanics (9), we introduce variability by drawing pairwise cell-cell interaction energies from a distribution for each run of the Metropolis Hasting model, and then, we hold these pairwise interaction terms fixed for simulated annealing and minimization of E_{total}.

Specifically, we label the cells with a number, n, that is independent of lattice site. We now define $\rho(c_n)$ as the cell type of cell n. For two cells, n and m, the probability of drawing an interaction energy of magnitude $|E_{nm}|$, $P(|E_{nm}|)$ is distributed according to:

$$P(|E_{nm}|) = \text{LogNorm}(W_{\rho(n)\rho(m)}, \sigma_{\rho(n)\rho(m)}), \tag{6}$$

where $\text{LogNorm}(\mu, \sigma)$ signifies a log normal distribution of mean, μ and standard deviation, σ. All energies are negative numbers, so the magnitude of each E_{nm} is distributed according to the distribution above, but the negative sign is preserved. Cell types have a mean strength of cohesive interaction specified by the symmetric matrix W where, again,

W_{12} is the mean interaction strength between cell type one and two. $\sigma_{\rho(n)\rho(m)}$ is the standard deviation of cohesive interactions between cell types $\rho(n)$ and $\rho(m)$. We selected a log normal distribution to model the interaction energy distribution because distributions of protein numbers in the cell have been shown to follow log normal distributions empirically. Again, we assume the convention that W is a matrix of positive values, but energies are negative numbers so that the total energy is minimized in our Metropolis equilibration.

Once pairwise cell-cell interaction energies are set, we proceed in equilibrating the tissue (minimizing global energy) with the identical Metropolis routine described above. The total energy of the cell aggregate is sum over all E_{nm} representing local interaction energies, so that:

$$E_{\text{total}} = \sum_{<nm>} E_{nm}, \tag{7}$$

again where this sum, $\sum_{<nm>}$ is taken over all neighboring lattice sites. In Fig. 4 (Main Text), we fix the mean interaction strength,W_{ij}, and vary the standard deviation, σ_{ij}, to study the influence of variability on the process of self-assembly.

Other strategies for introducing variability are possible, for example, one could evolve the specific values of the cell-cell interaction energies over the simulation or have the interaction energies determined by relative levels of simulated adhesion receptors in single cells, but we selected a strategy which required as few possible new assumptions or complications in the model.

Computational classification of tissue phases
Following tissue equilibration, we computationally classify the "phase" of the tissue as correct, inverted, disordered or phase separated using a scheme that incorporates three measures of global ordering a) the number of connected components of each cell type, b) the radial distribution function of cell types in the aggregate, and c) boundary uniformity. We combine these three measures to classify the tissue.

Connected tissue components: First, for each of the two cell types, MEPs and LEPs, we construct a cell graph where each cell is a vertex and edges exist between nearest and next nearest neighboring cell pairs of the same type (including diagonal neighbors as before). The cell graph will have a number of connected components that correspond to contiguous tissue units of a single cell type (fig. S8C). We calculate the number of connected components for each cell type C_{lep} and C_{mep}, and then calculate the total number of connected components in the structure $C_{\text{lep}} + C_{\text{mep}} = C_{\text{total}}$. Ordered tissues (fig. S8C, top), either correct or inverted, have large cell type contiguous regions, and therefore a

small number of connected tissue components. Disorganized structures have a large number of isolated cells (fig. S8C,bottom) increasing the number of disconnected components in the tissue and C_{total}. An absolutely ordered tissue of either the correct or inverted configuration has $C_{\text{total}} = 2$, but we used a slightly larger number to account for variation in equilibration. Empirically, we found that a threshold $C_{\text{total}} < 4$ demarcates the disorganized and organized phases in the Steinberg model and so used this cut off to separate the disorganized and organized phases in our computation.

Tissue polarity: Within the organized structures ($C_{\text{total}} < 4$), we sought to distinguish between structures with the same inside outside topology but distinct polarity. The "correct" configuration has a LEP core and MEP exterior layer; the inverted configuration has a MEP core and exterior LEP layer. We measure the polarity of the structure by defining a radial distribution function, $\langle \theta(r) \rangle$ that quantifies the local composition of the tissue. First, we define a function, $\theta(c(i))$, where $\theta(c(i)) = 1$ if $c(i)$ is a MEP and $\theta(c(i)) = -1$ if $c(i)$ is a LEP. We can then analyze the average value of θ as a function of radius in the tissue aggregate by computing:

$$\langle \theta(r) \rangle = \sum_{c(i) \in r} \frac{\theta(c(i))}{N_r}, \tag{8}$$

where the sum is taken over all cells within a radial band distance r from the center of mass of the tissue, and N_r is the number of cells within the radial band. $\langle \theta(r') \rangle = 1$ for a radial band, r' that is uniformly MEP, and $\langle \theta(r') \rangle = -1$ for a uniformly LEP radial band, at r', and 0 for a band of equal composition. Therefore, the function provides a measure of radial composition and uniformity. An example of radial distribution functions for two tissue phases are shown in fig. S8B.

The slope of $\langle \theta(r) \rangle$ measures changes in cellular composition as a function of distance and hence the polarity of the structure. Specifically, $\langle \theta(r) \rangle$ is flat for disordered aggregates. The function is monotonically increasing for the correct configuration (going from -1 to 1) or decreasing (fig. S8 B) for the inverted configuration.

We obtained a slope, s_θ, of this radial distribution function by fitting the function to a line and used the slope of the line to measure the polarity of the structure. Empirically, we found that a slope of $s_\theta < .2$ (correct) and $s_\theta > .2$ (inverted), separate the polarity of the correct and inverted structures.

Boundary uniformity: Finally, to separate ordered structures from "phase separated" structures, we used a calculation similar to that used in the experimental calling of phases. We devised a simple measure of boundary uniformity:

$$u = \frac{1}{N_b} | \sum_{i \in b} \theta(c(i)) |, \tag{9}$$

where here the sum is taken over cells in the tissue boundary, b. N_b is the number of cells in the boundary layer. Since $\theta(c(i)) = 1, -1$, the absolute value of the sum counts the excess of one cell type vs the other on the boundary. We, then, divide by N_b the number of cells in the boundary to achieve $u \in [0, 1]$ a measure of boundary uniformity. u is, then, the average value of θ, our order parameter, on the boundary. We selected $u = .8$ as the cutoff for determining the correct or inverted phase.

Scoring function and empirical evaluation using Steinberg rules: Using these three measures, we derived and tested our coloring and scoring system.

$$\text{Red (Correct): } CC_T < 4 \text{ AND } s_\theta > .2 \text{ AND } u > .8 \tag{10}$$

$$\text{Green (Inverted): } CC_T < 4 \text{ AND } s_\theta < -.2 \text{ AND } u > .8 \tag{11}$$

$$\text{Yelow (Phase Separated): } CC_T < 4 \text{ AND } (|s_\theta| < .2 \text{ OR } u < .8) \tag{12}$$

$$\text{Blue (Disorganized): } CC_T > 4. \tag{13}$$

To benchmark our phase calling framework, we classified tissue phases in the two component cell model where phase boundaries have been studied previously by Steinberg. In the fig. S8A, we show a computationally generated phase diagram for a two component tissue in the regime $W_{22} > W_{11}$ (LEP-LEP self interaction is stronger than MEP-MEP self interaction). A complementary phase diagram with similar qualitative structure exists for the $W_{11} > W_{22}$ region of parameter space where the "correct" configuration replaces the region occupied by the "inverted" configuration in S8A.

To compare with Steinberg, we plot the ratio W_{22}/W_{11} ($W_{\text{lep lep}}/W_{\text{mep mep}}$) against W_{12} ($W_{\text{mep lep}}$). We demarcate key regions predicted by Steinberg in the diagram. Steinberg demarcates the triangular region between the lines as generating the phase we call "inverted". The region above this line is "disorganized". Qualitatively our diagram strongly resembles Steinberg's diagram in the shape of the organized and disorganized region. Steinberg defines a region of phase separation below the lines $W_{12} = 1$ for the $W_{11} = 1$ value. We find that the majority of yellow, phase separated points, fall below this line.

Quantitatively, Steinberg defines the critical phase boundary between the disorganized configuration and the configuration where a LEP core is wrapped by a MEP boundary layer as occurring at $(W_{11} + W_{22})/2$ and $W_{12} > W_{11}$. To quantify the performance of our computational strategy, we sampled computationally generated points according to the Steinberg

Rules, and compared the computationally called phase with the phase from the Steinberg rules. Fig. S8D. that our classification scheme performed at greater than 90% success rate for each of the three critical phases examined in the main text.

Analytical results and inversion inequality
In this section, we derive an analytical expression for the magnitude of MEP-ECM interaction required to achieve the inversion phenomena observed in the experimental portion of the work. Experimentally we observed that in the absence of an ECM interaction a two component tissue of MEP and LEP cells assembles into a configuration with a MEP core surrounded by a LEP boundary; we call this the inverted configuration . In the presence of an ECM-MEP interaction, we observe that the cell aggregate achieves a flipped configuration with the LEP cells forming an interior core that is surrounded by an outer MEP cell layer, the "correct" configuration. Here, we calculate the strength of the MEP-ECM interaction, $W_{\text{MEP ECM}} = W_{13}$ required to accomplish tissue inversion given the strength of the cell-cell interactions. We derive a simple expression for the strength of the MEP-ECM interaction required to flip a tissue from the inverted to the correct configuration.

In his classic paper, Steinberg found that a two component tissue aggregate will assemble into a configuration with an inner core and outer boundary cell layer for well defined ranges of cell-cell interaction strengths (10). The Steinberg rules define a region of parameter space in which the two component tissue will form the inverted configuration:

$$W_{11} > W_{22} \tag{14}$$

$$W_{22} < W_{12} < \frac{W_{11} + W_{22}}{2}, \tag{15}$$

where MEP cells have the index 1 and LEP cells the index 2. The experimentally measured cell-cell interaction parameters fall within this Steinberg range.

Now, to study the influence of a new interaction between the MEP cells and the matrix on a two component tissue, we imagine a system where the tissue begins in the inverted configuration (i.e. with a W matrix consistent with the above rules) , and then we "turn-on" an interaction between the MEP cells and the ECM. We can then ask when the MEP-ECM, W_{13}, interaction will cause the equilibrium configuration of the tissue to switch from the inverted configuration to the correct configuration by comparing the energy of the correct and inverted configurations in the presence of the MEP-ECM interaction.

For a tissue aggregate composed of an identical number of LEP and MEP cells, the inverted tissue and the correct tissue are identical geometrically except that the cell labels have been

permuted, so that for all cells MEP → LEP and LEP → MEP. The case where $N_{\text{mep}} = N_{\text{lep}}$ is a special case where the correct and inverted configurations have an important symmetry. The number of interactions in each region of the tissue will be preserved under a permutation of cell labels, an exchange of all MEP and LEP cells. Importantly, the number of interactions in the core and boundary of the structure will not be the same. In general, a difference in the number of MEP and LEP cells can be accounted for by a correction term that is specific to the particular geometry.

We can write down an expression for the energy of the inverted configuration, E_{inverted}, by simply counting the number of each type of cell interaction (MEP-MEP, LEP-LEP, MEP-LEP) and multiplying by the appropriate interaction strength parameter, W_{ij} . In the inverted configuration, MEP cells define the core and LEP cells define the outer cell layer (fig. S8E). Therefore:

$$-E_{\text{inverted}} = W_{11}\ \phi_{\text{core}} + W_{12}\ \phi_{\text{interface}} + W_{22}\ \phi_{\text{outer}}. \qquad (16)$$

Here, we use ϕ to represent the number of cell-cell interactions in each part of the structure, so that ϕ_{core} is the number of cell-cell interactions in the interior core of the structure (fig. S8E). ϕ_{outer} represents the number of cell-cell interactions in the outer LEP cell layer. $\phi_{\text{interface}}$ is the number of MEP-LEP interactions at the interface of the LEP and MEP cell layers. We are using the thermodynamic convention where low energy states are favorable which leads to the negative sign in $-E_{\text{inverted}}$. This energy expression is a sum of terms that count the number of each type of interaction scaled by the interaction strength. For example, the term $W_{11}\ \phi_{\text{core}}$ counts the number of MEP-MEP cell-cell interactions in the core and multiplies these by their interaction strength W_{11}.

The energy of the correctly formed structure is identical in form to E_{inverted} except that (i) we permute the cell type labels $1 \rightarrow 2$ and $2 \rightarrow 1$, and (ii) we add an additional term proportional to W_{13} to account for the boundary interaction between MEP cells and ECM:

$$-E_{\text{correct}} = W_{22}\ \phi_{\text{core}} + W_{12}\ \phi_{\text{interface}} + W_{11}\ \phi_{\text{outer}} + W_{13}\ \phi_{\text{boundary}}, \qquad (17)$$

where the term ϕ_{boundary} represents the number of tissue boundary interactions between MEP cells and the ECM.

Now, we can ask when $E_{\text{inverted}} > E_{\text{correct}}$ so that the tissue adopts the correct configuration as a function of W_{13}, and find this to occur when:

$$W_{13} > (W_{11} - W_{22})(\frac{\phi_{\text{core}} - \phi_{\text{outer}}}{\phi_{\text{boundary}}}), \tag{18}$$

so that the threshold inversion energy is simply the difference in the homotypic interaction strengths for MEP cells and LEP cells scaled by a geometric factor. The cross interaction term W_{12} does not appear in this expression because it has canceled due to the symmetry of the boundary in the problem. The ϕ factors on the RHS of the equation are geometric factors that can be determined precisely for any specific aggregate configuration. Again, MEP $= 1$, LEP $= 2$, and ECM $= 3$. The analysis can be modified to account for cases where surface to core volume ratio differ from one ($N_{\text{mep}} \neq N_{\text{lep}}$) through introduction of correction factors.

In the main text, we consider a lattice geometry that is approximately square. The internal core lattice is approximately length 7, and the outer cell layer has approximately two cell layers. In this configuration, $\phi_{\text{core}} = 8(7^2)/2$ because there are 49 cells in the inner core each cell coordinates with 8 other cells, and the factor of two occurs because each interaction requires two cells; $\phi_{\text{outer}} = 150$, and $\phi_{\text{boundary}} = 96$. These numbers can be simply calculated from the square lattice geometry and the coordination numbers 8 in the core, 5 in the outer cell layer layer, and 3 on the ECM tissue boundary. Therefore, the inversion inequality for the lattice geometry studied in our computational model is:

$$W_{13} > (W_{11} - W_{22}) \times .48. \tag{19}$$

This inequality predicts the value of W_{13} required to invert a tissue given the values of W_{11} and W_{22}. We validate this result computationally in two ways. fig. S8F shows a plot of the polarity of a computationally determined structure for the interaction energies ($W_{11} = 6$ and $W_{22} = 1$ as a function of W_{13}). The equation predicts inversion at a value of W_{13} near 2.5, and in fact, in the fig. S8F, the polarity of the structure transitions from -1 to +1 at almost exactly this value.

To test the expression, we computationally selected 800 structures where the parameters in W are sampled so that the structure will form the inverted configuration in the absence of ECM. We set $W_{22} = 1$ for simplicity; sampled W_{11}, W_{13} uniformly at random; and selected W_{12} so the the tissue would form the inverted configuration without the ECM. In fig. S8G, we plot the polarity of the structure as a function of W_{11} and W_{13}. On the diagram, we have plotted the line $W_{13} = (W_{11} - 1) \times .48$, the analytically predicted phase boundary. We see that this line demarcates the correct (Red) from the inverted (Green) configuration.

In general, we can use the inversion inequality to study the scaling relationship between the critical value of W_{13} and the size of the structure. For a circular aggregate with $r > 2$,

$\phi_{\text{core}} \sim 8\pi r^2/2$, $\phi_{\text{outer}} \sim (5(2\pi r) + (l-2)*2\pi r6)/2$ where $l, l > 1$ is the number of single cell layers in the outer cell layer, and $\phi_{\text{boundary}} \sim 3(2\pi r)$, so that:

$$W_{13} > (W_{11} - W_{22})\frac{4r - (5 + 6(l-2))}{6}. \tag{20}$$

In this way, for fixed interaction energies, the critical value of W_{13} is proportional to the radius of the aggregate. For equal LEP, MEP cell numbers, a structure becomes three layered at $r = 6$. At this radius, the scaling factor is $13/6 \sim 2$. The experimentally observed values of W_{13} should, therefore, be consistent with aggregates of a larger size size than analyzed experimentally in the main text.

In the above derivation, we considered interactions between the LEP and ECM, W_{23} to be negligible. This interaction can be introduced into the inverted configuration, and the resulting inequality is simply:

$$W_{13} > (W_{11} - W_{22})(\frac{\phi_{\text{core}} - \phi_{\text{outer}}}{\phi_{\text{boundary}}}) + W_{23}. \tag{21}$$

Since W_{13}, W_{23} both act at the boundary of the tissue, their contribution to the energy of the aggregate carries an identical geometric factor. Increases in W_{23} linearly increase the required interaction between MEP and ECM to drive the tissue into the correct configuration.

Figure S1

Figure S2

A

Functionalized agarose

C

PDMS

B

	No Protein	BSA	Fibronectin1	Collagen1

10X

40X

LEP MEP

D

	Agarose		Functionalized agarose	
	1%	4%	1%	4%

5 hrs

LEP MEP

10 hrs

20 hrs

E

LEP

Phase	Fn1 Actin DAPI

4 hrs

16 hrs

MEP

Phase	Fn1 Actin DAPI

4 hrs

16 hrs

Figure S3

Figure S4

A

B

C

GENE	Fold Reg	P-Value
DLL1	7.1376	0.009465
CTNNA3	7.085	0.017444
P2RX6	5.4168	0.029744
PVRL1	4.3714	0.004773
PVRL3	4.3302	0.013716
DOCK4	3.7041	0.005895
DSC1	2.9617	0.000481
DSC3	2.9095	0.000445
DSG3	2.8431	0.000684
CDH3	2.537	0.020976
DSG4	2.4219	0.001867
PKP1	2.255	0.005657
FLNA	2.0187	0.00554
CDH2	-17.1788	0.000043
MAPRE2	-3.8368	0.004983
PVRL4	-3.3766	0.000044
PVRL2	-2.5301	0.006147
MLLT4	-2.4098	0.002287
BAIAP2	-2.3678	0.000015
DNM1	-2.1688	0.017836
DSG2	-2.1261	0.001182
DSC2	-2.0584	0.001646
TJP1	-2.0133	0.016625
ARVCF	-2.0049	0.016034
CDH1	-1.094	0.108166
HPRT1	-1.1428	0.718387
ACTB	-1.1855	0.080876
GAPDH	1.3017	0.023373
RPLP0	-1.098	0.132623

D

GENE	Fold Reg	P-Value
MMP3	20.58	0.045848
LAMA1	19.9	0.001156
MMP1	18.9	0.020323
COL7A1	10.07	0.001856
ITGA4	8.36	0.002981
THBS2	6.8	0.020468
ITGA6	6.5	0.012699
FN1	6.28	0.014437
LAMA3	4.85	0.000031
SPARC	4.63	0.000653
ITGA1	4.6	0.009826
KAL1	4.5	0.010793
COL6A1	3.96	0.031306
ITGA3	3.62	0.041596
COL5A1	3.34	0.006622
ITGA2	3.27	0.002634
ITGB4	3.06	0.02833
MMP7	-17.26	0.011632
ITGB2	-8.72	0.023136
MMP15	-4.73	0.045259
ICAM1	-3.43	0.019627
COL11A1	23.18	0.138137
MMP13	11.52	0.138647
COL16A1	6.22	0.133479
TIMP1	6.36	0.099112
ITGB3	3.02	0.201825
CD44	2.14	0.017713
VCAN	1.96	0.006662
ECM1	2.06	0.001285
ITGB1	1.6	0.000247
LAMA2	1.77	0.060047
LAMB1	2.38	0.001648
LAMB3	2.68	0.003246
TNC	2.61	0.010367
VTN	2.16	0.125593
CDH1	1.1	0.180546
NCAM1	-9.56	0.095578
TGFBI	-2.28	0.008675
TIMP2	-2.3	0.016754
VCAM1	-1.11	0.595206
RPL13A	1.26	0.945848
HPRT1	1.17	0.393554
B2M	-1.91	0.126736
GAPDH	1.47	0.09394
ACTB	1.11	0.634511

Figure S5

A

Alpha 6 Integrin KD at 48 hrs

Isotype
100nM siRNA
75nM siRNA
50nM siRNA
25nM siRNA
NO siRNA

% MAX

PE Signal

B

Beta1 Integrin KD with 75nM

Isotype
72hrs siRNA
48hrs siRNA
NO siRNA

% MAX

PE Signal

C

NSC   Talin1 siRNA

Talin1

beta-Actin

D

LEP   MEP
Control p120 siRNA   NSC p120 siRNA

p120 Catenin

beta-Actin

E

Control   Beta1 KD   Talin1 KD

XY plane   MEP   MEP   MEP

XZ plane

Matrigel   LEP MEP

agarose   LEP MEP

F

Cell-Cell Contact Angle

Talin-1   p120   Control

G

Cell-ECM Contact Angle

Control   Talin-1   p120

H

Cell-Cell   Circularity   Cell-ECM

NSC

p120 KD

Talin1 KD

I

self-organization in agarose
Control   p120 KD
LEP MEP

self-organization in Matrigel
Control   Talin1 KD
LEP MEP

J

10X   40X
Control

10X   40X
Talin1 KD

K

Circularity

NSC   p120   Talin1

L

Beta1 Integrin after Talin1 KD

Isotype   NSC   Talin1 KD

% MAX

PE Signal

M

Alpha6 Integrin after Talin1 KD

Isotype   NSC   Talin1 KD

% MAX

PE Signal

N

DMSO   Y-27632   Latrunculin A

Actin
K14

20X - Confocal

63X - Confocal
XY   XY   XY
XZ   YZ   XZ   YZ   XZ   YZ

O

self-organization in Matrigel

Phase   Fluorescence   Merge   Avg. Int.

DMSO

Y-27632

Lat A

Figure S6

A

agarose

Avg. Intensity

LEP MEP

Matrigel

Avg. Intensity

LEP MEP

B

MEP | LEP | Merge
K14 | K19

40X Confocal Microscopy

Orthogonal Views

MEP | LEP | Merge
K14 | K19

40X Confocal Microscopy

Orthogonal Views

C

Avg. Intensity

MEP | LEP

Avg. Intensity

MEP | LEP

D

Norm. Intensity — K14, K19 vs Radial Distance

Frequency — I II III IV V

Norm. Intensity — K14, K19 vs Radial Distance

Frequency — I II III IV V

E

agarose

K8 K5

Norm. Intensity vs Radial Distance

Matrigel

K8 K5

Norm. Intensity vs Radial Distance

Basal
Luminal

K5 K8

F

prostate — basal
mammary — LEP

agarose | Matrigel

Figure S7

A

$W_{mep\ lep}$

$\frac{1}{2}(W_{lep\ lep}+W_{mep\ mep}) = W_{mep\ lep}$

B)

$W_{mep\ lep} = 1$

$\frac{W_{lep\ lep}}{W_{mep\ mep}}$

C Connected Tissue Components

CC = 1   CC = 1

CC = 8   CC = 1

B

Inverted

$\langle\theta(r)\rangle$

disorganized

r

D

Fraction Correct

Correct   Inverted   Random

Tissue Configuration from Steinberg Rules

E

ECM

Interface

Surface

Core

Boundary Layer

F

Tissue   Polarity

$W_{mep\ mat}$

Predicted Threshold

G

$.48\ (W_{mep\ mep} - 1)$

$W_{mep\ mat}$

$W_{mep\ mep}$

Figure S8

Supplementary Figures Legends

Figure S1. *Fluorescence Activated Cell Sorting (FACS) and methods for full encapsulation of cell-clusters in Matrigel.* (A) Representative FACS scatter plot used for isolating luminal and basal cells from primary human mammary epithelial cells (HMEC) at passage 4. Insets show the FACS scatter plots for unstained controls as well as the gates used for sorting luminal (in green) and basal (in red) cells. Green events are sorted as luminal, red events are myoepithelial (or basal), and blue events are unsorted cells. (B) Sample FACS scatter plot used for purifying luminal and myoepithelial cells from uncultured primary HMEC isolated from human reduction mammary tissue. (C) Sample FACS scatter plot used for purifying luminal and basal cells from uncultured primary human prostate epithelial cells (HPEC) directly isolated from human prostate tissue. (D) Schematic (top) and representative 10X, and 40X fluorescent microscopy images (bottom) of luminal cells (stained with Cell Tracker Green - CTG) and myoepithelial cells (stained with Cell Tracker Red - CTR) aggregated by chemical programmed assembly (CPA) prior to encapsulation in Matrigel. (E) Schematic diagram and representative images of luminal and basal cells mechanically aggregated for 3 hours in non-adhesive microwells prior to transfer and full encapsulation of cell-aggregates in Matrigel. (F) Schematic diagram and representative images of cell-aggregates left in agarose microwells for 16 hours. All Scale bars are 100 μm.

Figure S2. *Self-organization of HMECs in agarose and Matrigel.* (A) Sample FACS scatter plot used for isolating pure lineages from primary human mammary epithelial cells (HMEC) at passage 4 (LEP in green and MEP in red). (B) Representative 40X phase and fluorescent microscopy images of heterotypic aggregates of LEPs (stained with Cell Tracker Green - CTG) and MEPs (stained with Cell Tracker Red - CTR) in microwells immediately after mechanical aggregation (left). After 24 hours, these mixtures of cells can self-organize into structures represented by the icons and microscopy images to the right. (C) Representative 40X images of HMEC clusters after 1 day in agarose (top) or Matrigel (bottom). LEPs are stained with CTG and MEPs were stained with CTR. Scale bars are 50 μm. (D) CTG and CTR average intensity maps for 30 HMEC clusters after 1 day in agarose (top) or Matrigel (bottom). (E) Visual Scoring for HMEC clusters after 1 day in agarose (N= 268) or Matrigel (N= 235). (F) Representative 40X confocal microscopy image of a heterotypic aggregate of LEPs and MEPs fixed and stained for lineage-specific keratin markers (K14 for MEPs and K19 for LEPs) immediately after transfer to Matrigel. (G) Representative 40X confocal microscopy images of a microtissue (100 μm in diameter) stained for K14 (red), K19 (green), and nuclei (blue) after 24 hours in agarose (top) or Matrigel (bottom). (H) Average K14, K19 intensity signals for 15 different HMEC clusters after 24 hours in agarose (top) or Matrigel (bottom). (I) Normalized average keratin intensity profiles as a function of normalized radial distance for the data presented in H. (J) Representative 40X confocal microscopy images of a lumenized HMEC cluster stained for K14 (red), K19 (green), actin (yellow), and nuclei (blue) after 96 hours in Matrigel. Scale bar is 50 μm. (K) Average K14, K19, and merged intensity profiles for 9 different lumenized HMEC clusters after 96 hours in Matrigel. (L) Normalized average radial keratin profiles for the data presented in K. (M) Sample 40X confocal microscopy images of a lumenized HMEC cluster stained for α6

integrin (red), Mucin 1 (green), Desmoglein 2 (yellow), and nuclei (blue) after 96 hours in Matrigel. Scale bar is 50 μm. (N) Representative 20X confocal microscopy images of lumenized acini fixed and stained for Laminin 5 (red) and DAPI (blue). Scale bar is 20 μm. (O) Sample confocal orthogonal view (in the XZ plane) of a lumenized bilayered acinus fixed and stained for K19 (green) and K14 (red) after 96 hours in Matrigel. Scale bar is 50 μm.

Figure S3. *Self-Organization of HMECs in functionalized agarose and PDMS microwells*. (A) Schematic of cyanogen bromide mediated activation and functionalization of agarose surfaces. (B) 10X (top) and 40X (bottom) representative fluorescent microscopy images of LEPs and MEPs stained with Cell Tracker Green (CTG) and Cell Tracker Red (CTR), respectively, and allowed to self-organize for 24 hours in 3% agarose microwells functionalized with no protein (i.e. ethanolamine), Bovine Serum Albumin (BSA), Fibronectin1 (Fn1), or Collagen1 (Col1). The inset is an average intensity projection of 30 Col1-functionalized agarose microwells. Scale bars in 10X images are 100 μm. Scale bar in 40X image is 50 μm. (C) Representative images of LEPs and MEPs stained with CTG and CTR after 24 hours in PDMS microwells. Inset is an average intensity projection of 30 microwells. (D) Representative 40X fluorescent microscopy images of LEPs and MEPs stained with Cell Tracker Green and Cell Tracker Red, respectively, and allowed to self-organize for 5, 10, and 20 hours in 1% or 4% agarose or Collagen 1 functionalized agarose microwells. (E) 40X phase contrast and confocal images of LEP (top) and MEP (bottom) fixed and stained for Fibronectin 1 (red), Actin (green), and nuclei (blue) after culturing dissociated cells on PDMS for 4 hours and 16 hours. While MEPs spread, LEPs formed 3D aggregates that were easily washed away. Those LEP-clusters that remained attached to the substrate were anchored by a Fibronectin 1 expressing-cell (white arrow). Scale bar is 10 μm.

Figure S4. *Contact angle, circularity, and spreading of LEP and MEP*.
(A) Orthogonal views of a representative 63X confocal microscopy image showing a MEP cell spreading onto a Matrigel-coated substrate. At 0 hours, the membrane dye (DiO – red) is localized to the surface of the cell (left). After 4 hours, the dye is also found in the interior of the cell (right). The interface with the substrate is visualized with a layer of physisorbed quantum dots (QD-605 - magenta). Scale bar is 10 μm. (B) Sample 63X confocal microscopy images of a MEP spread onto the surface of the Matrigel-coated substrate and exhibiting membrane protrusions that extend onto the surface but do not deform the overall shape of the cell in the XZ plane. (C) Representative 63X confocal microscopy images reconstructed in the XZ plane showing the spreading of single LEPs (top) and MEPs (bottom) stained with membrane dyes (DiO - green and DiI - red) on Matrigel coated coverslips after 4 (left) and 12 hours (right). (D) Quantification of cell-ECM contact angles for MEPs after 4 and 12 hours and for LEPs after 4 hours. (E) Representative MEP (top) and LEP (bottom) cells fixed and stained for Actin and K14 or K19 respectively after 12 hours of spreading onto the Matrigel-coated substrate. (F) Representative 40X phase contrast microscopy images of homotypic LEP-aggregates (top) and MEP-aggregates (bottom) after 16 hours in large circular agarose microwells (left). 40X confocal microscopy images of representative LEP and MEP clusters fixed and stained for K14 and K19 (MEP and LEP, respectively) confirm the identity of the

homotypic aggregates (right). A few contaminating cells of the opposing lineage are visible in each aggregate. Scale bars are 100 μm. (G) Representative 10X phase contrast microscopy images (top) and 40X confocal microscopy images (bottom) of homotypic (MEP or LEP) or heterotypic (MEP and LEP) clusters after spreading on Matrigel coated coverslips for 12 hours. K14 and K19 staining confirms the identity of each lineage. Scale bars are 10 μm.

Figure S5. *qPCR screen for cell-cell and cell-ECM adhesion molecules in MEP and LEP*. (A) The expression of a panel of cell-cell adhesion genes in MEP (on Y-axis) relative to LEP (on X-axis). The indicated dots are arbitrarily selected genes that showed a change in expression found in the MEP lineage with respect to the LEP lineage of at least 2-fold and with a P-Value less than 0.05. The red color signifies overexpression in the MEP lineage with respect to the LEP lineage while the green color signifies genes with a reduced level of expression in the MEP lineage with respect to the LEP lineage. (B) The expression of a panel of cell-ECM adhesion genes in MEP (on Y-axis) relative to LEP (on X-axis). Labeled genes were arbitrarily selected and illustrate several integrins and ECM molecules of interest. (C) Table summarizing fold-change in expression for the most significant cell-cell adhesion molecules overexpressed (in red) or underexpressed (in green) in MEP relative to LEP. (D) Table summarizing fold-change in expression for the most significant cell-ECM adhesion molecules overexpressed (in red) or underexpressed (in green) in MEP relative to LEP.

Figure S6. *Perturbations to cell-cell and cell-ECM cohesion*. (A) Flow cytometry analysis of protein knockdown as a function of siRNA concentration using α_6 integrin as a model target. (B) Flow cytometry analysis of protein knockdown as a function of time (after transfection) using β_1 integrin as a model target. (C) Western blot of HMEC transfected with siRNA targeting Talin1. (D) Western blot of HMEC transfected with siRNA targeting p120-catenin. After transfection, HMEC were sorted into LEP and MEP to confirm knock down in both lineages. (E) XY and XZ orthogonal views of a control MEP, β_1 knock-down (KD) MEP, and Talin1 KD MEP stained with DiO (red) and spreading on the surface of a Matrigel coated substrate (top). In β_1 KD MEP, the spreading area of the cell in the XY plane appeared to be slightly reduced but this was not reflected in a substantial decrease in contact angle at the MEP-substrate interface and did not significantly perturb the self-organization of β_1 KD MEPs with control LEPs in agarose or Matrigel (bottom). For most MEP, Talin1 KD resulted in a reduction in both the spreading area of the cell and the Cell-ECM contact angle (right). In addition, the self-organization of Talin1 KD MEPs with control LEPs was altered in Matrigel but not agarose (bottom). Scale bars are 10 μm. (F) Quantification of cell-cell contact angle upon perturbations to Talin1 and p120 expression in MEP (control = untransfected). (G) Quantification of cell-ECM contact angle upon perturbations to Talin1 and p120 catenin expression in MEP (control = untransfected). (H) Representative images obtained from cell-cell contact angle (left), circularity (middle), and cell-ECM contact angle (right) assessing the effect of siRNA-mediated knockdown of p120 catenin and Talin1 in MEP. The non-silencing control is labeled NSC. (I) Sample 10X fluorescent microscopy images of control and Talin1 KD MEP (stained with Cell Tracker Red) and reconstituted with control LEP (stained with Cell Tracker Green) after 1 day in Matrigel (top). Sample 10X

fluorescent microscopy images of Control and p120-catenin KD MEP (stained with Cell Tracker Red) reconstituted with control LEP (stained with Cell Tracker Green) after 1 day in agarose (bottom). Scale bars are 100 μm. (J) Sample 10X (left) and 40X (right) microscopy images of control and talin1 KD MEPs after 8 hours of pre-aggregation in agarose wells and 12 hours of spreading onto a Matrigel coated substrate. Scale bars are 100 μm. (K) Quantification of circularity 4 hours after mechanical aggregation of MEP transfected with siRNA targeting Talin1 or p120 catenin. The non-silencing control is labeled NSC. (L) Flow cytometry analysis of β_1 integrin expression after siRNA-mediated Talin1 knockdown. (M) Flow cytometry analysis of α_6 integrin expression upon siRNA-mediated Talin1 knockdown. (N) 20X and 63X confocal microscopy images of MEP with DMSO (control), 50 μM Y-27632, or 2 μM Latrunculin A in M87A medium after 12 hours of spreading onto a Matrigel-coated coverslip. The 63X confocal images offer orthogonal views of representative clusters in the ZX and YX planes (Green = Actin; Red = K14). Note the lack of an organize ring of actin circling the cell periphery in the presence of latrunculinA or Y-27632. Also note the appearance of cell protrusions rich in cytokeratins upon treatment with Y-27632. (M) Self-organization of MEP and LEP stained with Cell Tracker Red and Cell Tracker Green, respectively, after 16 hours of self-organization in Matrigel in the presence of DMSO (control), 50 μM Y-27632, or 2 μM Latrunculin A. Normalized average intensity profiles are shown to the right (n=15). Scale bars are 50 μm.

Figure S7. *Self-organization of uncultured primary human mammary and prostate epithelial cells in agarose or Matrigel.* (A) Representative 40X epifluorescence microscopy image and average fluorescent intensity (N=18) of uncultured primary LEPs and MEPs in agarose (left) or Matrigel (right) (green = LEP, Cell Tracker Green; red = MEP, Cell Tracker Red). (B) Representative 40X confocal microscopy images and orthogonal reconstructions of primary LEPs and MEPs in agarose (left) or Matrigel (right) (green = K19; red = K14). (C) Average K14 (red) and K19 (green) intensity for 13 primary HMEC clusters in agarose (left) or Matrigel (right). (D) Normalized average keratin radial intensity plot for primary HMEC (N=13) derived from sample 59L in agarose (far left) or Matrigel (mid right) and visual scoring of primary HMEC clusters derived from two additional donors (samples 160L and 168) in agarose (mid left; N= 39) or Matrigel (N = 28; far right). (E) Normalized average radial K5 (for basal cells) and K8 (for luminal cells) intensity plots for primary human prostate epithelial cells (N=12) derived from sample MB in agarose (left) or Matrigel (right). Also shown is a representative 40X confocal microscopy image of a primary prostate aggregate in Matrigel. (F) 40X fluorescent microscopy images of the self-organization of first passage primary basal cells (stained with CTR) derived from human prostatectomies with fourth passage primary human LEP (stained with CTG) after 24 hours in agarose (left) or Matrigel (right). The insets are CTG and CTR average fluorescent intensity maps (N = 30). All scale bars are 10 μm.

Figure S8. *Analytical calculations and validation of computational model* (A) Computational equilibration of the two-component (MEP/LEP) tissue (without ECM) as compared with the tissue phase diagram derived by Steinberg. We selected N>5000

interaction matrices (W) uniformly at random in the regime $W\_lep\_lep > W\_mep\_mep$ and determined the equilibrium tissue configuration using the computational equilibration and classification procedure. Each point on the plot indicates a single value of W. The axes of the diagram are set to be consistent with Steinberg, so that the y-axis is W_mep_lep and the x-axis is W_lep_lep/W_mep_mep. The diagram depicts phase boundaries as demarcated by Steinberg (dotted lines). A phase boundary separates the disorganized (blue) and organized (green) phase at $W\_mep\_lep = (W\_lep\_lep + W\_mep\_mep)/2$. Another phase boundary at $W\_mep\_lep = 1$ demarcates a phase separated regime. Our simulation conforms to the Steinberg predictions and blue points (disorganized) are above the line and green points (MEP inside, LEP outside or "inverted") are below the line. The phase separation region was less well demarcated from the organized regime in our model due to the closed and rigid boundary configuration that we selected in order to simulate the experimental conditions in our study. B) Examples of radial distribution functions for the inverted (green line) and disorganized configurations (blue line). The x-axis indicates radial distance and the y-axis is $<\Theta(r)>$. C) Examples of connected tissue components are shown as black lines for two different tissue configurations. In each, the MEP and LEP connected components are shown separately. The key feature is that the disorganized tissue has a large number of MEP connected components. D) The success rate of the computational model and scoring function are shown. We selected N>500 computational tissues with $W\_mep\_ECM = 0$ using the Steinberg two component tissue rules. The bar plot indicates the fraction of correct computational calls for each phase. The graph shows that each of the three phases discussed in the paper is determined at greater than 90% accuracy (.95, .92, .96 fraction correct respectively). E) A diagram depicting the regions of the tissue aggregate (core, outer cell layer, interface, boundary) referred to in the analytical derivation. F) Polarity, s_theta, of a simulated tissue as a function of W_mep_ECM for fixed cell interaction parameters ($W\_mep\_mep = 6$ and $W\_lep\_lep = 1$ and $W\_mep\_lep = 3$). The polarity of the tissue is shown on the y-axis and the value of W_mep_ECM on the x-axis. The polarity (s_Θ) of the tissue switches from -1 (inverted) to +1 (correct) as a function of W_mep_ECM. The critical value of $W\_mep\_ECM \sim 2.5$ as predicted by the analytical calculation is indicated by the dashed line. G) Polarity, s_Θ, for a section of the tissue phase space is shown in detail and compared with the analytical inversion inequality. N>800 tissues were selected uniformly at random from the inverted region of the two component phase diagram where W_mep_mep in [4,8], W_mep_ECM in [0,8], $W\_lep\_lep = 1$, and $W\_mep\_lep = .75 + W\_mep\_mep/4$. Points are colored according to the polarity of the structure (color bar), so that red indicates structures with a MEP outside and LEP inside polarity indicative of the "correct" configuration, and green indicates structures with a LEP outside and MEP inside configuration indicative of the inverted configuration. The black line is set by the inversion inequality (Equation 26) with .48 ($W\_mep\_mep - 1$) since $W\_lep\_lep = 1$.

Table S1: List of antibodies used, product numbers, and applications (IF – Immunofluorescence; FACS – Fluorescently activated cell sorting; FC – Flow Cytometry; WB – Western Blot). The dilution used for each application is reported in parenthesis.

Antibody	Product	Application
Anti-human Keratin 5	NeoMarkers XM26 (clone XM26)	IF (1:30)
Anti-human Keratin 8	Epitomics 1925-1 (poly)	IF (1:100)
Anti-human Keratin 19	Sigma C6930 (clone A53-B/A2)	IF (1:50)
Anti-human Keratin 14	Thermo RB-9020-P (poly)	IF (1:50)
Anti-human CD324	BioLegend 324112 (clone 67A4)	IF (1:100)
Anti-human Laminin 5	Millipore MAB19563 (clone D4B5)	IF (1:50)
Anti-human CD49f	Millipore MAB1378 (clone GoH3)	IF (1:50)
Anti-human CD227	BD 559774 (clone HMPV)	IF (1:50)
Anti-human Desmoglein 2	Novus NBP1-857096 (poly)	IF (1:50)
Anti-human Keratin 14	NeoMarkers XM26 (clone XM26)	IF (1:30)
Anti-human CD227-FITC	BD 559774 (clone HMPV)	FACS (1:50)
Anti-human CD10-APC	Biolegen 312210 (clone HI10a)	FACS (1:200)
Anti-human CD44-APC	BD 560532 (clone G44-26)	FACS (1:20)
Anti-human CD57-FITC	BD 555619 (clone NK-1)	FACS (1:5)
Anti-human CD49f-PE	Biolegend 313612 (clone GoH3)	FC (1:200)
Anti-human CD29-PE	Biolegend 555443 (clone T52-16)	FC (1:200)
Anti-human p120-catenin	BD 610133 (clone98/pp120)	WB (1:1000)
Anti-human Talin1	Cell Signaling 4021 (poly)	WB (1:500)
Anti-human β-actin HRP	Abcam ab49900 (clone AC15)	WB (1:2000)

Supplementary References

1. Stampfer MR, LaBarge MA, Garbe JC (2013) An Integrated Human Mammary Epithelial Cell Culture System for Studying Carcinogenesis and Aging. *Cell and Molecular Biology of Breast Cancer*:323–361.
2. Peehl DM (2002) (Culture of epithelial cells, New York).Second Edition.
3. Selden NS et al. (2012) Chemically programmed cell adhesion with membrane-anchored oligonucleotides. *J Am Chem Soc* 134:765–768.
4. Liu JS, Farlow JT, Paulson AK, LaBarge MA, Gartner ZJ (2012) Programmed Cell-to-Cell Variability in Ras Activity Triggers Emergent Behaviors during Mammary Epithelial Morphogenesis. *Cell Reports* 2:1461–1470.
5. Gartner ZJ, Bertozzi CR (2009) Programmed assembly of 3-dimensional microtissues with defined cellular connectivity. *Proc Natl Acad Sci USA* 106:4606–4610.
6. Chanson L et al. (2011) Self-organization is a dynamic and lineage-intrinsic property of mammary epithelial cells. *Proc Natl Acad Sci USA* 108:3264–3269.
7. Graner and Glazier. (1992) Simulation of biological cell sorting using a two dimensional extended Potts model. *Physical review letters.*
8. Metropolis et al. (1953). Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys. 21, 1087.*
9. Sherrington et al. (1975). *Solvable Model of a Spin-Glass. Physical Review Letters. 35.1792.*
10. M Steinberg. (1963). Reconstruction of Tissue by Disassociated cells. *Science.* 141:401–408.