Supplemental Figure 1: Cell positioning is intrinsically heterogeneous in vivo and in vitro (related to Figure 1).

A. Pixel classification and post-processing workflow for image segmentation. Following initial segmentation, images were processed to remove small features, holes, and debris.
B. Structural distribution for human mammary gland sections with similar size composition reveals intrinsic heterogeneity in cell position. Data from **Fig. 1A** was divided by tissue size and LEP proportions. Distributions are shown for tissues smaller than 150 μm in diameter, with LEP proportion between 0.35 and 0.65 (red) or between 0.65 and 0.9 (blue). The number of tissues analyzed in each category are noted on the graph.
C. Schematics illustrating the experimental workflow for mammary organoid reconstitution. Fourth passage HMEC were infected with lentivirus expressing cytoplasmic GFP or mCherry and purified by FACS between after 5-7 days. Equal numbers of GFP LEP and mCh MEP were aggregated in non-adherent microwells for 4-6 h prior to transfer to Matrigel. Organoids were imaged two days after reconstitution.
D. Cell isolation by FACS. Single cells were isolated based on forward and side scatter intensity. From singlets, GFP+ cells were isolated as GFP+ and mCherry-, while mCherry+ cells were isolated as GFP- and mCherry+. GFP+ LEP were isolated as GFP+, EpCAM-hi and CD271- or GFP+, EpCAM-hi and CD49f-low. mCherry+ MEP were isolated as mCherry+, EpCAM-low and CD271+ or mCherry+, EpCAM-low and CD49f- hi.
E. Immunostaining of mammary organoids two and six days after culture in Matrigel. GFP (yellow) and mCh (purple) expression was used as markers for sorted LEP and MEP respectively. The expression of keratin-19 (red) and keratin-14 (cyan) marks luminal and myoepithelial lineages.
F. Principal component analysis plot on different structural metrics for organoids containing roughly equal number of LEP and MEP. Organoids were manually annotated as LEP shell, MEP shell, mixed or split (schematics shown in panel g). 500-1000 images were sampled from each category for the analysis.
G. Pearson's correlation matrix for the structural metrics used in the analysis. Boundary LEP fraction and intermixing score were uncorrelated measures of tissue structure.
H. The probability density for intermixing score for mammary organoids and MEP spheroids two days post-reconstitution.
I. The probability density for the normalized inter-centroid distance for mammary organoids and MEP spheroids two days post-reconstitution.
J. Boundary LEP occupancy and intermixing score are effective in separating the 4 annotated structural categories.
K. The distribution of LEP boundary occupancy of mammary organoids in Matrigel at different times

post-reconstitution.

L. LEP boundary occupancy for mammary organoids from different patient specimens in Matrigel, 2 days post-assembly.

M. LEP/GFP boundary occupancy for mammary organoids and MEP spheroids in Matrigel across experimental replicates two days post-assembly.

Data was collected across multiple experiments. The number of observations is noted at the bottom or top of the graphs. The lines and hinges for boxplots show the median and the 1st and 3rd quartiles. Asterisks represent significance of difference from the reference group (day 2 for panel K, 240L for panel L, and base-mean for panel M), as follows ns: $p > 0.05$; *: $p < 0.05$, **: $p < 0.005$; ***: $p < 0.0005$ based on Wilcoxon test.

Supplemental Figure 2: Tissues dynamically sample from the ensemble steady state distribution (related to Figure 2).

A. The trends for average ϕ_b and Φ over time for mammary organoids (gold) and MEP spheroids (blue) analyzed by time lapse microscopy. The whisker plots show the mean and standard deviation for the data.

B. The probability of transitioning between any two structural states over a 3-hour window is represented by the size of the circles, similar to **Fig. 2C**. Any transitions not observed are marked by '+'.

C. The autocorrelation function (ACF) of ϕ_b for mammary organoids and MEP spheroids. The ACF approaches zero after approximately 2 hours (dashed line). The points and error bars are the mean and 95% confidence intervals from binning across all organoids. The lines represent the average for each organoid.

D. Small spheroids of GFP- or mCh- were reaggregated to make spheroids with pre-sorted initial structures instead of random structures. There reaggregated spheroids relaxed to random configurations within a few hours, with their average ϕ_b and intermixing scores resembling the steady state ensemble average. The whisker plots show the mean and 95% confidence interval for the data.

Supplemental Figure 3: A statistical mechanical framework provides a quantitative description of organoid structural distributions (related to Figure 3).

A. A two-dimensional hexagonal lattice model for mammary organoids. Each cell is modeled as a hexagon with 6 neighbors. The cell centers are arranged in a hexagonal grid.

B. A three-dimensional body-centered cubic (BCC) lattice model for mammary organoids. Each cell is modeled as a truncated octagon with 6 nearest neighbors (square interfaces) and 8 next nearest neighbors (hexagonal interfaces). The cell centers are arranged in a hexagonal grid. Tissue surface was modeled as a sphere to avoid overestimating the ECM surface area.

C. The proportion of cell-cell interfaces (either LEP-LEP or LEP-MEP) do not predict the mechanical energies of the simulated tissues using the BCC lattice. The energies within a macrostate defined by either of these variables are asymmetrically distributed about the average.

D. The distribution of microstate energies for the macrostates with $\phi_b = 0.1, 0.2 \text{ and } 0.5$. For each macrostate, the energies are symmetrically distributed about the average macrostate energy (vertical line). The difference in energy across macrostates is much larger than that variance within a macrostate. All energies were estimated using the BCC model.

E. Bootstrapping was used to obtain confidence intervals for the mechanical potential (ΔE). An example of 50 bootstrapping iterations is shown. The inset shows the distribution of ΔE from 1000 iterations.

F. A schematic for the two-compartment lattice model for entropy estimation. Each boundary (dark gray) or core (light gray) lattice site can be occupied by either a LEP (gold) or MEP (purple). Macrostates are defined by the fraction of LEP in boundary (ϕ_b).

G. A schematic representation of structural degeneracy showing many microstates within the same macrostate (e.g., 3 LEP in the boundary). Cells with the same label are identical, and swapping two cells of the same identity yields identical microstates.

H. Estimated number of cells and boundary fraction for tissues of different sizes, as a function of the diameter of a single cell and the thickness of the boundary compartment. Based on the average organoid size, tissues with diameter = 80 μm and a boundary thickness of 8 μm were used unless specified otherwise (dashed line). For these parameters, lattice sites were equally distributed between the boundary and the core.

I. Schematic showing the effect of sampling the middle tissue section on the measured LEP fraction and LEP boundary occupancy.

J. The relationship between the measured LEP boundary occupancy and the measured LEP fraction for different imaging depths in organoids with equal proportion of LEP and MEP. For sorted organoids, the measured LEP fraction is higher than the expected value.

K. Estimated standard deviation for ϕ_b due to volume sampling for different imaging depths.

L. Predicted structural distributions for MEP spheroids resulting from combining structural degeneracy

and volume sampling for different imaging depths.

M. Experimentally measured structural distributions for spheroids containing equal number of GFP+ and mCh+ LEP, MCF7 or MCF10A cells. The blue line is the fit for the distribution for MEP spheroids.

N. Comparison of experimental and predicted structural distributions for mammary organoids and MEP spheroids. Histograms show the experimental distribution. Dashed and dotted lines are predictions based on degeneracy only or volume sampling only. The solid line is the final analytical model prediction.

Supplemental Figure 4: Tissue activity sets the balance between the mechanical potential and macrostate degeneracy (related to Figure 4).

A. Example MSRD vs Δt curves for MEP (purple) or LEP (gold) spheroids in Matrigel and agarose. Time intervals less than 3 h we used to fit the data to the diffusion models to calculate D_{eff} and α, as the curves were mostly linear at these time scales.
B. The diffusion exponents for MEP (purple) or LEP (gold) spheroids in Matrigel and agarose. The cells show diffusive-like behaviors (α = 1).
C. Example segmentations of mammary organoids and MEP spheroids in agarose at steady state (24 h). Scale bar = 50 μm.
D. The structural distribution for MEP spheroids in agarose. The blue line is the distribution of MEP spheroids in Matrigel. The number of spheroids is noted on the graph.
E. Representative segmented mammary organoids and MEP spheroids cultured in agarose and followed by time lapse microscopy. Scale bar = 50 μm. Graph shows the quantification for the organoids shown.

Supplemental Figure 5: Engineering the structural ensemble by programming the mechanical potential and activity (related to Figure 5).

A. Depletion of Talin-1 and p120 catenin protein expression in MEP upon treatment with shRNA against TLN1 and CTNND1 respectively. Top panels show representative western blots for Talin 1 and p120 catenin expression along with α-tubulin expression (loading control). All intensity measurements were normalized to the loading control. The graphs show percent reduction in normalized protein expression in shRNA-treated cells compared to non-targeting shRNA-treated cells.
B. Interfacial tension measurements for KD-MEP. Cortical tension, cell-cell contact angles and cell-ECM contact angles are shown using boxplots. Estimated cell-cell and cell-ECM interfacial tensions are listed in the table.
C. Apparent diffusion coefficients for KD-MEP in Matrigel (top) and agarose (bottom). The number of spheroids analyzed is noted at the bottom of the graph.

Asterisks represent significance of difference from the reference group (ctrl shRNA), as follows ns: $p > 0.05$; *: $p < 0.05$, **: $p < 0.005$; ***: $p < 0.0005$ based on Wilcoxon test.

Supplemental Figure 6: Engineering the structural ensemble by programming macrostate degeneracy (related to Figure 6).

A. The macrostate energy calculations for mammary organoids with increasing Φ using the BCC lattice model.

B. Apparent diffusion coefficients for MEP in Matrigel for organoids with varying Φ. The number of spheroids analyzed is noted at the bottom of the graph.

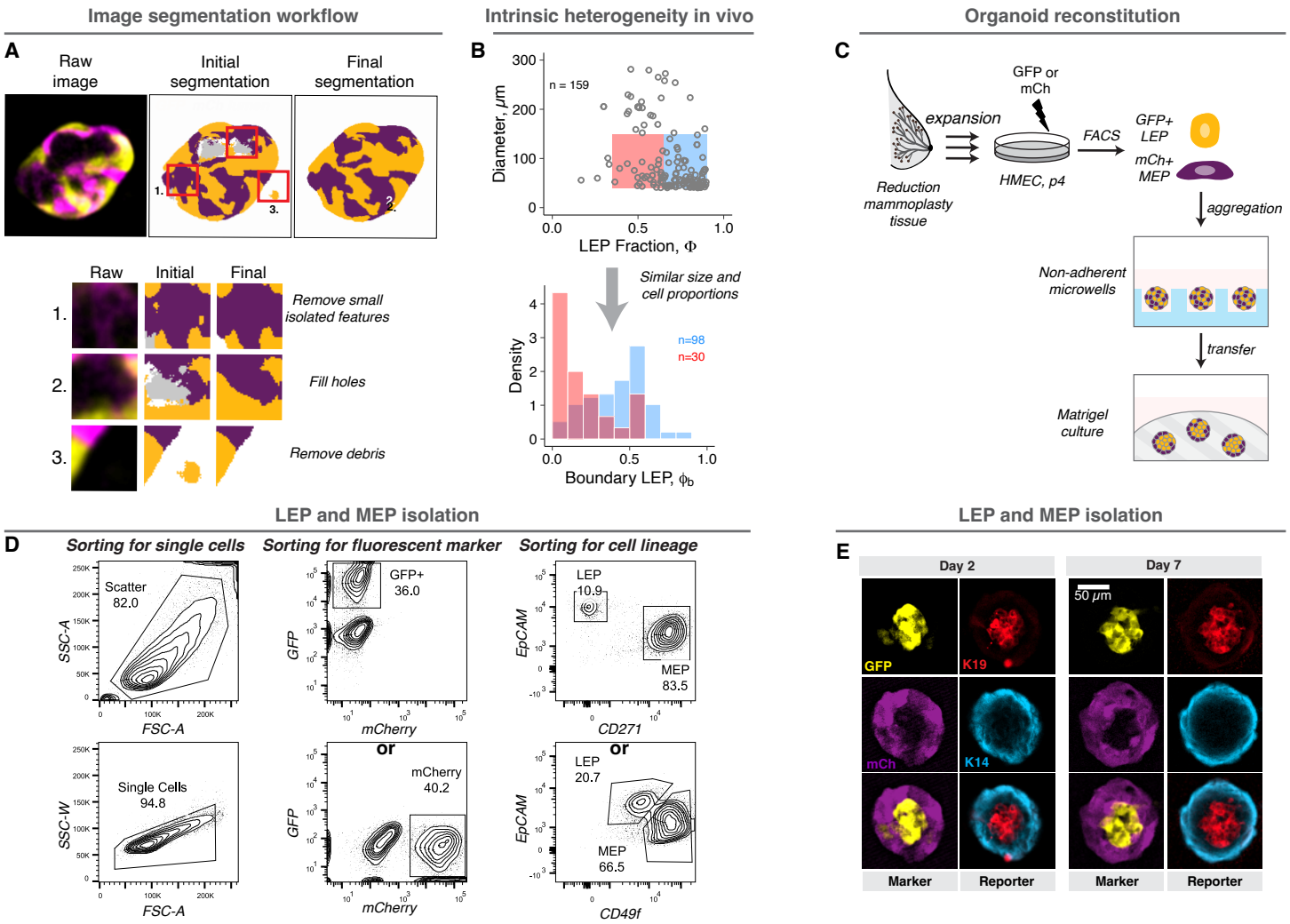
C. The macrostate energy calculations for mammary organoids with increasing diameter using the BCC lattice model.

D. Cell speeds near the tissue boundary for MEP spheroids of varying sizes in Matrigel. The number of spheroids analyzed is noted at the bottom of the graph.

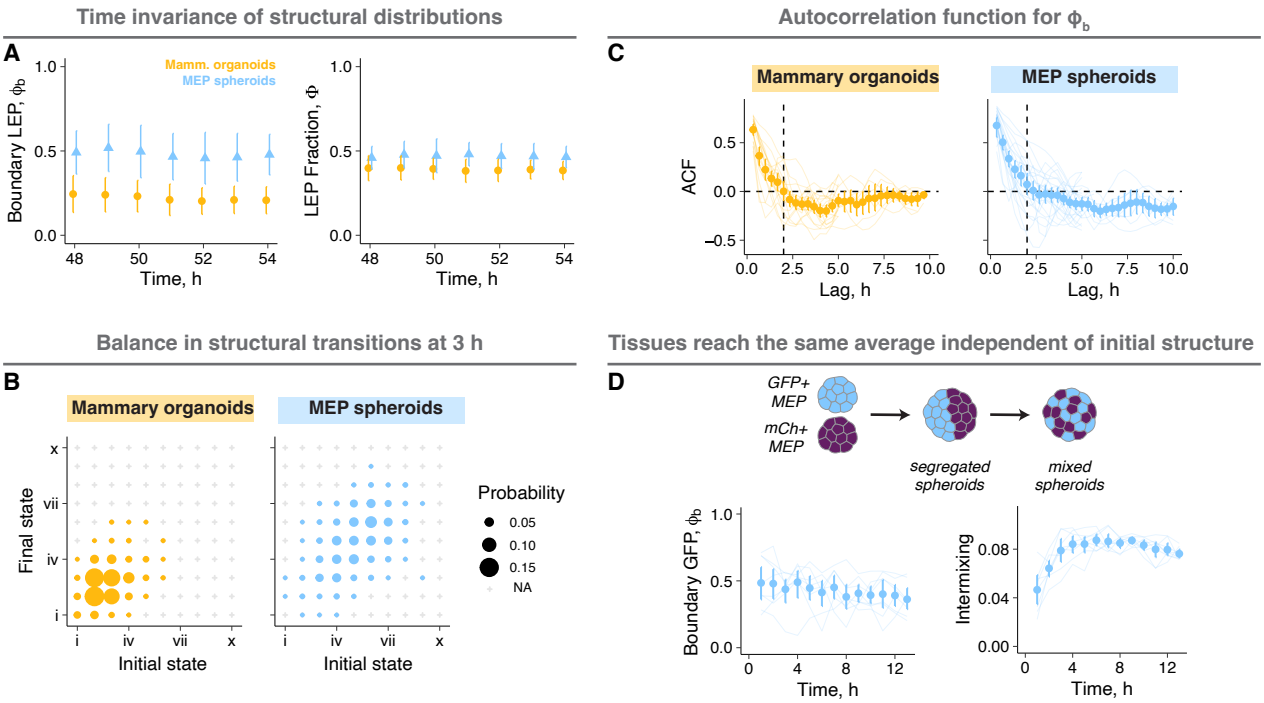
E. Calculations of ΔG, average ϕ_b, and standard deviation of ϕ_b for different mechanical potentials and activities in tissues with a diameter = 80 μm containing equal number of LEP and MEP. The lines are predictions from the model and are colored by the value of ΔE. Estimated values for different experimental conditions are also shown, where points and error bars are the average and standard deviations. The symbols represent different conditions (○: mammary organoids, △: MEP spheroids, ◇: TLN1-KD spheroids, □: CTNND1-KD spheroids), and the points are colored by their calculated ΔE.

F. Calculations of ΔG, average ϕ_b, and standard deviation of ϕ_b for different mechanical potentials and LEP fractions in tissues with a diameter = 80 μm and activity corresponding to Matrigel. Estimated values for different experimental conditions are also shown, where points and error bars are the average and standard deviations. The symbols represent different conditions (○: mammary organoids, △: MEP spheroids), and the points are colored by their calculated ΔE.

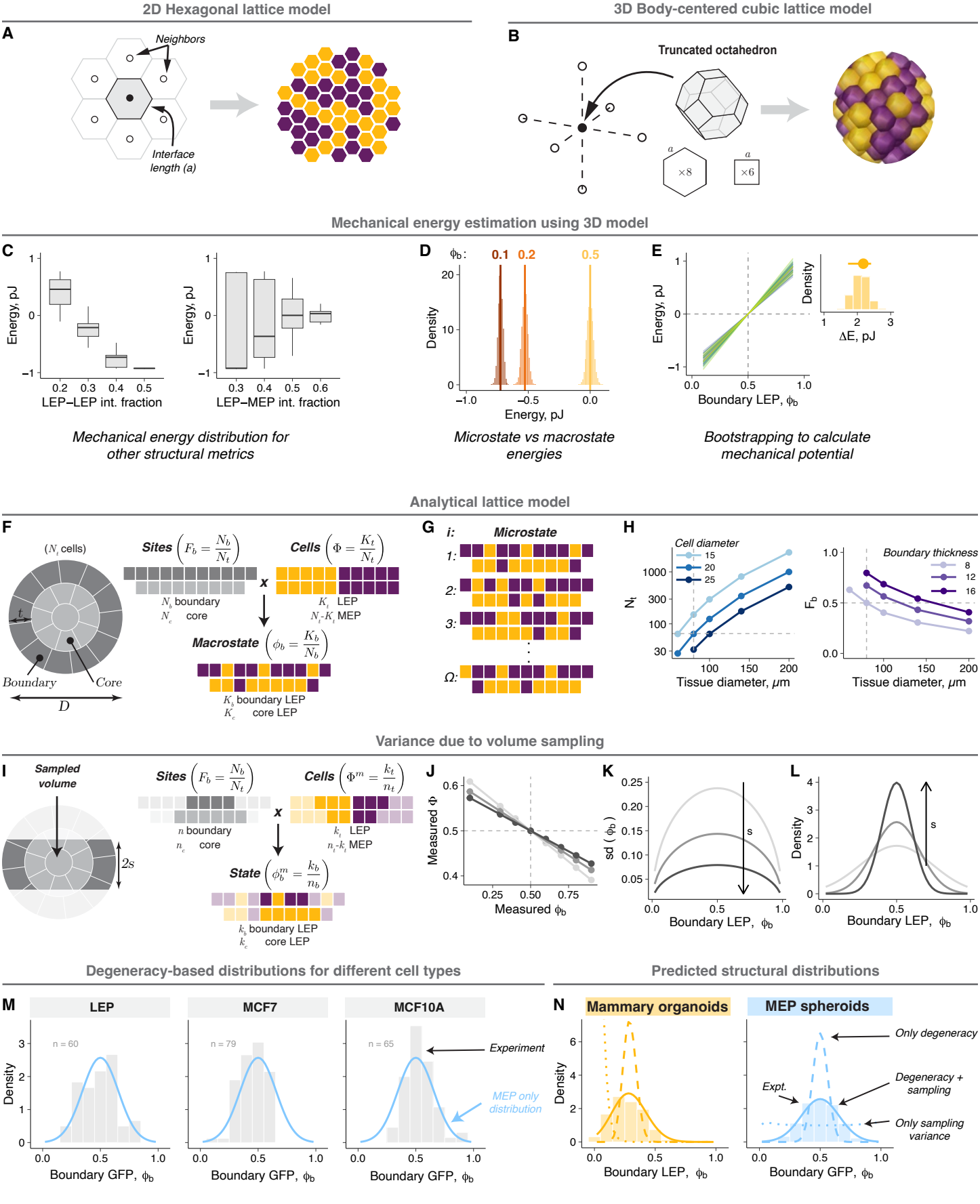
G. Calculations of ΔG, average ϕ_b, and standard deviation of ϕ_b for different mechanical potentials and tissue sizes LEP fraction = 0.5 and activity corresponding to Matrigel. Estimated values for different experimental conditions are also shown, where points and error bars are the average and standard deviations. The symbols represent different conditions (○: mammary organoids, △: MEP spheroids), and the points are colored by their calculated ΔE.



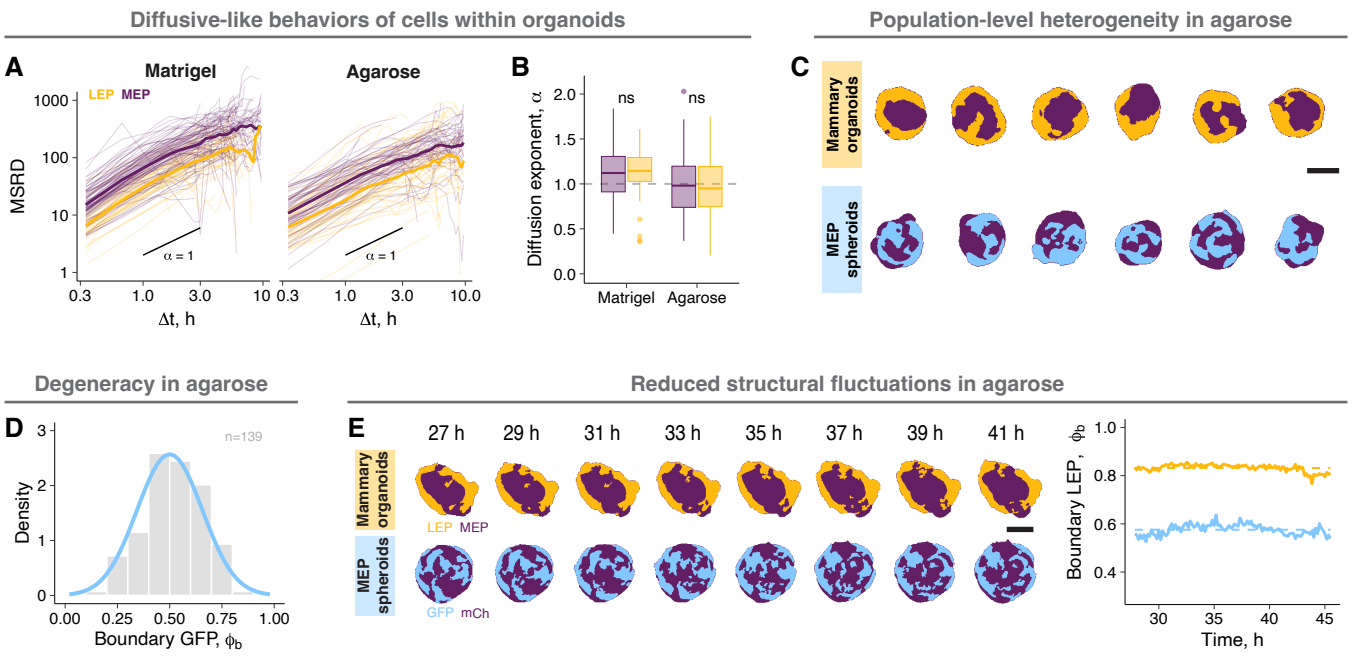
SUPPLEMENTAL FIGURE 1



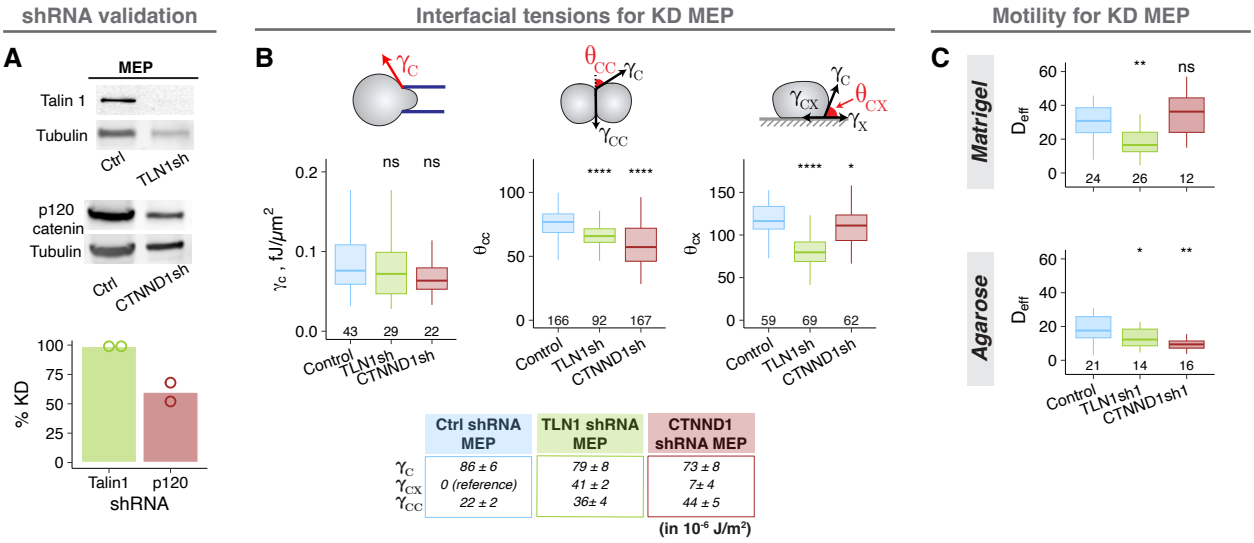
SUPPLEMENTAL FIGURE 2



SUPPLEMENTAL FIGURE 3



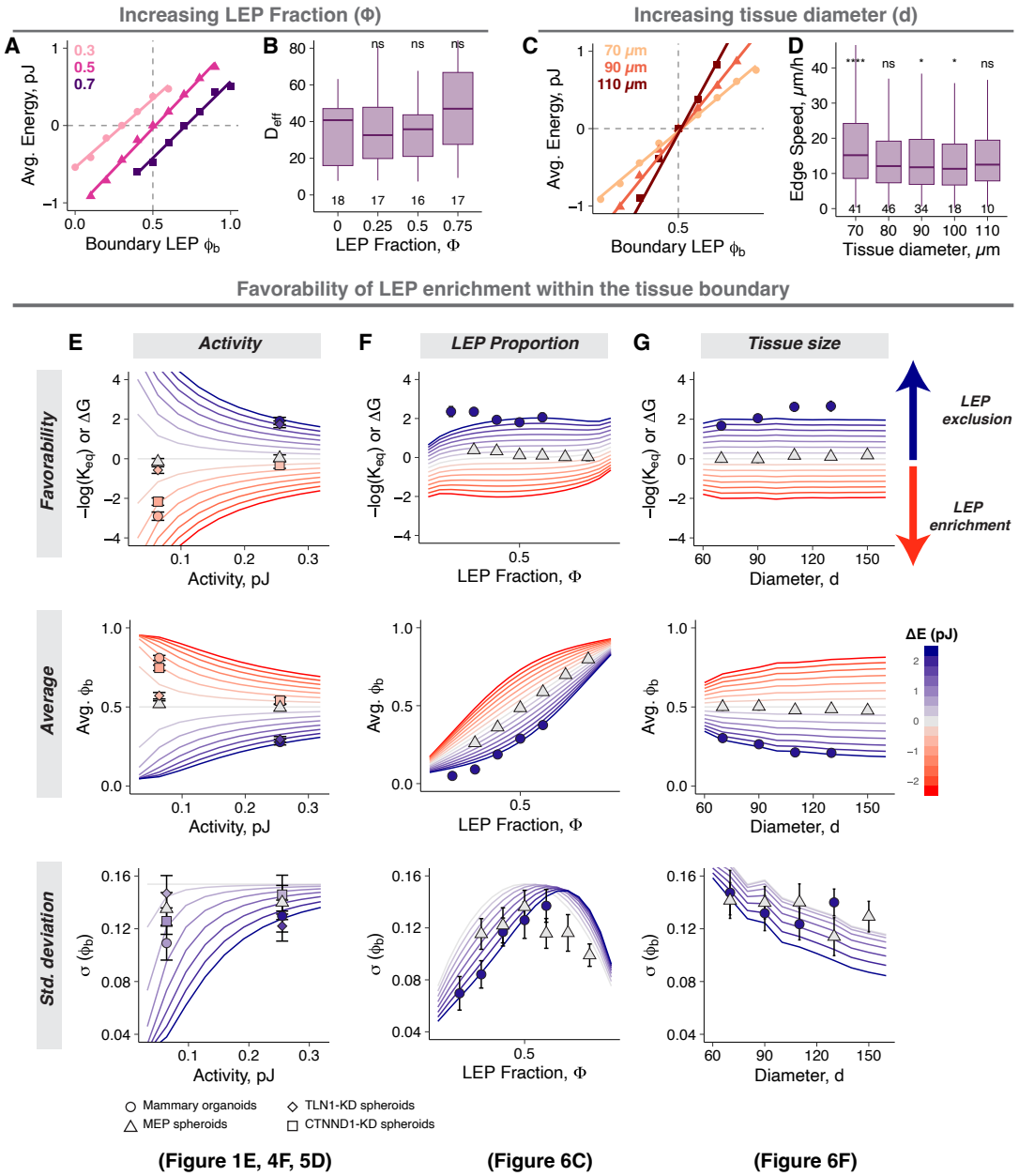
SUPPLEMENTAL FIGURE 4



	Ctrl shRNA MEP	TLN1 shRNA MEP	CTNND1 shRNA MEP
γ_C	86 ± 6	79 ± 8	73 ± 8
γ_{CX}	0 (reference)	41 ± 2	7± 4
γ_{CC}	22 ± 2	36± 4	44 ± 5

(in 10^{-6} J/m^2)

SUPPLEMENTAL FIGURE 5



SUPPLEMENTAL FIGURE 6