

**UCSF**

**UC San Francisco Electronic Theses and Dissertations**

**Title**

A Statistical Mechanical Approach to Steady-State Tissue Structures in Human Mammary Epithelial Organoids

**Permalink**

<https://escholarship.org/uc/item/7pw2p38z>

**Author**

Hu, Jennifer Liu

**Publication Date**

2021

Peer reviewed|Thesis/dissertation

A Statistical Mechanical Approach to Steady-State Tissue Structures in Human Mammary Epithelial Organoids

by  
Jennifer Hu

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of  
DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

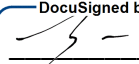
of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

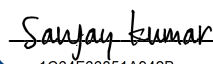
AND

UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

DocuSigned by:  
  
6E99B613A5754DD... Zev Gartner  
Chair

DocuSigned by:  
  
Jeffrey Bush

DocuSigned by:  
  
1C64F66851A942B... Sanjay Kumar

---

Committee Members



## Dedication and Acknowledgments

This dissertation was made possible with the support of many individuals, first and foremost my co-author Vasudha Srivastava: a brilliant scientist and engineer, a perceptive mentor, and a compassionate human being. I am blessed to call her my colleague and friend. She contributed much of this work, in particular the micropipette aspiration, cloning, and diffusion analysis. Chasing the science with all its twists and turns has been an incredible experience with my adviser, Zev Gartner, and his infectious enthusiasm for pushing forward both the science and my development as a scientist. I also thank Mark LaBarge for providing guidance with warmth, humor, and seriousness. My dissertation committee, Sanjay Kumar, Jeff Bush, and Valerie Weaver, has always engaged deeply and generously offered feedback and support. Their thoughtful questions led to many fruitful directions. Insightful discussions with Greg Huber, Matt Thompson, Rob Phillips, and L. Mahadevan, have helped inspire me to think outside of the box.

My friendships have always been life-giving, but even more so this past year, with weekly bursts of happiness and restoration: board games with Dominic and Katie, movies and games with the Michaels and Julie, D&D with David and Kara, dance with Tina, Lo-Hua, and LuShuang, yoga with Caroline, phone calls from Aisha. It has been so lovely to be known and to know each other over these years as we become ourselves. And finally, I am grateful for the unconditional support of my family, especially Jelle, who has been my partner and my home through it all.

# A Statistical Mechanical Approach to Steady-State Tissue Structures in Human Mammary Epithelial Organoids

Jennifer Liu Hu

## Abstract

Tissues in the body adopt regular structures that are actively maintained far from thermodynamic equilibrium and yet achieve a steady state. This control over cellular arrangements both ensures tissue function and protects against the spread of cancer cells in the mammary gland, where structural disruption marks the transition from in situ to invasive breast cancer and a major increase in patient risk. With a reconstituted human mammary organoid system derived from cultured primary cells, I model the steady-state structural distributions of tissues as a function of cellular mechanical properties, configurational entropies, and active fluctuations. These experimental results demonstrate a surprising concordance with equilibrium statistical mechanics. Activating one of the most commonly dysregulated breast cancer genes, *PIK3CA*, causes decreased self-organization in mammary organoids via a change in cells' mechanical properties. Making targeted changes to parameters of the tissue energy and active fluctuations with RNA interference and small molecule inhibition corrects organoid structure. This approach directly connects cell mechanics, cell dynamics, and tissue geometry to quantitative structural outcomes using a statistical mechanical framework. It also ties specific physical parameters to phenotypes of risk in breast cancer progression and lays the groundwork for future studies of the dynamics of tissue structure and state.

# Table of Contents

1	Introduction . . . . .	1
1.1	Tissue structure in the mammary gland . . . . .	2
1.2	Cell sorting in tissues . . . . .	4
1.3	Statistical mechanics in biology . . . . .	10
2	Methods . . . . .	15
2.1	HMEC material and cultivation . . . . .	15
2.2	Viral transduction of HMEC . . . . .	16
2.3	Photolithography and PDMS stamps . . . . .	18
2.4	Reconstituted mammary organoids . . . . .	18
2.5	Microscopy . . . . .	19
2.6	Immunofluorescence . . . . .	20
2.7	Interfacial tension measurements . . . . .	21
2.8	Image processing and quantification . . . . .	22
2.9	Three-dimensional nuclear diffusion assay . . . . .	26
2.10	Predicted tissue energies . . . . .	26
2.11	Inferred tissue energies . . . . .	29
3	Theory . . . . .	31
3.1	Statistical mechanics for multicellular tissues . . . . .	31
3.2	Defining structural coordinates . . . . .	32
3.3	Quantifying tissue energy . . . . .	34
3.4	Enumerating structural microstates . . . . .	38
3.5	Steady state and activity . . . . .	41

4	Results . . . . .	46
4.1	Organoid structures follow a Boltzmann distribution . . . . .	46
4.2	Cell ratio and tissue entropy . . . . .	47
4.3	Estimated tissue energies predict experimental outcomes . . . . .	49
4.4	Agarose microenvironment decreases activity . . . . .	51
4.5	Oncogenic PI3K shifts the steady-state distribution . . . . .	55
4.6	Reversion to the steady state by altering energy and activity . . . . .	57
5	Discussion . . . . .	60
5.1	Cell sorting in the context of cancer . . . . .	60
5.2	Structural steady states in development and differentiation . . . . .	62
5.3	Organoids as a model of active matter . . . . .	64
5.4	Caveats and alternative approaches . . . . .	66
	Conclusion . . . . .	70
	References . . . . .	72
	Appendix A Selected interfacial tensions . . . . .	94
	Appendix B Predicted and inferred tissue energy gradients . . . . .	96
	Appendix C Estimation of interfacial energies in tissues . . . . .	98
	Appendix D Combinatoric microstate enumeration in polyhedral lattice . . . . .	103
	Appendix E Cross-sectional sampling of organoid features . . . . .	105

# List of Figures

1.1	Structure of the mammary gland . . . . .	3
1.2	Statistical mechanics in biology . . . . .	12
2.1	FACS . . . . .	16
2.2	Lentiviral backbone . . . . .	17
2.3	Image processing . . . . .	24
2.4	Truncated octahedral honeycomb . . . . .	27
2.5	Bootstrapping energy derivative estimation . . . . .	29
3.1	Categorical outcomes . . . . .	33
3.2	Morphological parameters and structural outcomes . . . . .	34
3.3	Interfacial tensions of LEP and MEP . . . . .	36
3.4	Simulated tissue energies . . . . .	38
3.5	Combinatorics of cell arrangements . . . . .	39
3.6	Structural distributions of MEP-only organoids . . . . .	41
3.7	Timescale of mixing . . . . .	42
3.8	Timescale of self-organization . . . . .	43
3.9	Fluctuations in the steady state . . . . .	44
4.1	Calculation of the Boltzmann distribution for GFP organoids. . . . .	47
4.2	Shifted cell ratios for GFP and MEP-only organoids. . . . .	49
4.3	Predicted and inferred energies of <i>TLN1</i> sh MEP organoids. . . . .	51
4.4	Activity in Matrigel and agarose. . . . .	53
4.5	<i>CTNND1</i> sh MEP + MEP tissue structures and energies. . . . .	54
4.6	Breast cancer driver genes in LEP. . . . .	56
4.7	<i>PIK3CA</i> LEP shift the steady state with altered interfacial energy. . . . .	57
4.8	<i>TLN1</i> sh rescues LEP-ECM energy and steady state organization. . . . .	58



4.9	AKT inhibition by MK2206 decreases tissue activity. . . . .	59
A.1	Comparison of interfacial tensions for GFP, agarose, and PIK3CA. . . . .	95
C.1	Cell interfacial tensions. . . . .	98
C.2	Truncated octahedral honeycomb . . . . .	100
C.3	Formation of cell contacts . . . . .	100
D.1	Combinatorics of cell arrangements in polyhedral lattice . . . . .	104
E.1	Simulated organoid cross-sections . . . . .	106
E.2	Cross-section and whole tissue metrics . . . . .	107
E.3	Tissue size and structural distribution . . . . .	108

# List of Tables

2.1	Morphological parameters . . . . .	25
A.1	Combinations of interfacial tensions . . . . .	94
B.1	Predicted and inferred energy gradients . . . . .	96

## List of Abbreviations

### Acronyms

BSA	Bovine serum albumin
CI	Confidence interval
DAPI	4',6-diamidino-2-phenylindole
DCIS	Ductal carcinoma in situ
DITH	Differential interfacial tension hypothesis
ECM	Extracellular matrix
EMT	Epithelial to mesenchymal transition
EO	Edge occupancy
FACS	Fluorescence-activated cell sorting
FDT	Fluctuation dissipation theorem
FIJI	FIJI is just ImageJ
GFP	Green fluorescent protein
HMEC	Human mammary epithelial cell
IDC	Invasive ductal carcinoma
IMS	Intermixing score
LEP	Luminal epithelial cell
LSM	Laser scanning microscope

MEP	Myoepithelial cell
MOI	Multiplicity of infection
MSD	Mean squared displacement
NESS	Non-equilibrium steady state
PBS	Phosphate-buffered saline
PDMS	Polydimethylsiloxane
PFA	Paraformaldehyde
PI3K	Phosphatidylinositol 3-kinase
shRNA	Short hairpin RNA
TDLU	Terminal ductal-lobular unit
WT	Wild type

### **Genes**

<i>CCND1</i>	Gene name of cyclin D1, a cyclin that inhibits pRb to begin cell division.
<i>CDH1</i>	Gene name of E-cadherin, an epithelial cell-cell adhesion molecule.
<i>CTNND1</i>	Gene name of p120 catenin, an adapter protein linking cadherins to the actin cytoskeleton.
<i>ERBB2</i>	Gene name of ERBB2 or HER2, a receptor tyrosine-protein kinase and proto-oncogene.
<i>GATA3</i>	Gene name of GATA3, a transcription factor that regulates the luminal epithelial phenotype.

*MAP2K4* Gene name of MAPKK4 or MEK4, part of the RAF-MEK-ERK pathway, inducing proliferation and anti-apoptotic signals.

*MAP3K1* Gene name of MAPKKK1 or MEKK1, part of the RAF-MEK-ERK pathway.

*MYC* Gene name of MYC, a proto-oncogene and transcription factor regulating cell cycle, metabolism, proliferation, and apoptosis.

*PIK3CA* Gene name of p110 $\alpha$ , one of three variants of the catalytic subunit of PI3K responsible for phosphorylating phosphatidylinositol (4,5)-bisphosphate.

*PTEN* Gene name of phosphatase and tensin homolog, which dephosphorylates phosphatidylinositol (3,4,5)-trisphosphate and reverses the effect of PI3K.

*RB1* Gene name of pRb, tumor suppressor that blocks cell division in response to stress.

*TLN1* Gene name of Talin-1, an adapter protein linking integrins to the actin cytoskeleton.

*TP53* Gene name of p53, a tumor suppressor that blocks cell division in response to DNA damage.

# 1. Introduction

Self-organization is a fundamental process in living systems, which require well-ordered structures to function. The initially indistinguishable, unordered cells of an embryo develop identity and structure through a variety of physical and biochemical self-organizing processes that specify cell fates and arrange them in specific spatial positions. However, in most animals, these processes are not so stereotypical as to produce the same structures every single time. Substructures within tissues frequently differ from one another in size, shape, and composition. Heterogeneity is a feature of normal tissues, both in vivo and in organotypic in vitro cultures like organoids. This balance between structural order and disorder that determines tissue structure is disrupted during disease progression. The transition from in situ, enclosed lesions to fully invasive cancer is defined by the breakdown of tissue structure and of the boundaries between tissue compartments. Improving our understanding of the biological and physical forces driving self-organization and disorganization will allow us to predict and manage tissue structures in the context of development, disease, and tissue engineering.

This dissertation demonstrates that the observed distribution of self-organized tissue structures can be modeled and predicted by combining the principles of statistical mechanics with energetic approaches to cell sorting. In a stable yet fluctuating dynamic tissue, the addition of a quantitative measure of configurational Boltzmann entropy more accurately describes the observed distribution of tissues than the traditional approach to cell sorting, which considers interfacial energies alone. Instead, the free energy of a tissue is a function both of the interfacial energies and the structural entropy of the cellular arrangement. Theoretical predictions are tested by directly perturbing the enthalpic and entropic parameters, causing the resulting structural distribution to shift in predictable ways. These results pro-

vide support for the statistical mechanical model and suggest that these parameters are key to understanding structural transitions in tissues.

## 1.1: Tissue structure in the mammary gland

The mammary gland is a remarkable tissue that develops postnatally in response to circulating hormones, the levels of which drive the expansion, differentiation, and involution cycles that occur throughout pregnancy, the menstrual cycle, and aging (Parmar and Cunha, 2004) (**Fig. 1.1A**). Luminal epithelial cells (LEP) are enclosed by a layer of myoepithelial cells (MEP) at the basal surface in contact with the basement membrane. The basement membrane is a specialized extracellular matrix (ECM) that is generated by secretion and remodeling activity of MEP and stromal fibroblasts (Nelson and Larsen, 2015). This thin ( $\approx 50$  nm to 100 nm) but dense network of proteoglycans and glycoproteins is organized around a scaffold of laminins and collagen IV, and provides physical and biochemical signals that modulate the epithelial phenotype (Yurchenco and Patton, 2009; Barcellos-Hoff et al., 1989; Novaro, 2003). The basement membrane separates the epithelium from the connective tissue, a molecularly and physically distinct ECM primarily composed of fibrillar collagens I and III and fibronectin in the loose intralobular stroma and the denser interstitial or interlobular stroma (**Fig. 1.1B**). This structural organization ensures that apically secreted milk produced by LEP in the terminal ductal-lobular units (TDLU) can be expelled by the enclosing contractile MEP and funneled along the ducts towards the nipple. The mammary gland can create and maintain this cellular arrangement despite the dynamic remodeling throughout the normal life of the organ, as well as the acquisition of cancer driver genes in the LEP compartment.

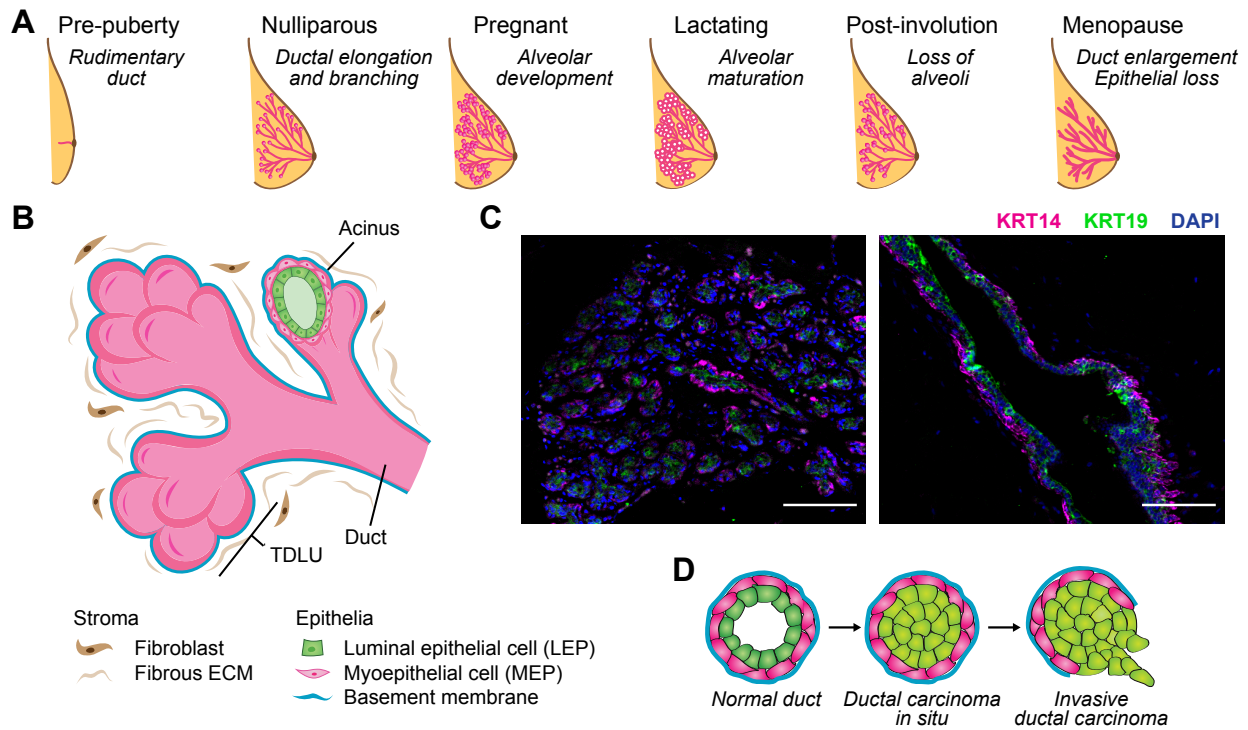


Figure 1.1: Development and dynamics of mammary gland structure throughout life in (A) adapted from Srivastava et al. (2020). Branched epithelial trees are composed of luminal and myoepithelial cells separated by a laminin-rich basement membrane from the fibrous collagen-I stromal ECM, adapted from Hu et al. (2018) (B). Human tissue sections of TDLU and duct stained for cytokeratins provided by Sundus Shalabi (City of Hope) and Alexander Borowsky (UC Davis). MEP, marked by cytokeratin 14 (KRT14), cover LEP (stained with cytokeratin 19, KRT19) more uniformly in ducts than in acini. A variety of cellular arrangements are visible even within the same tissue of a single patient. Scale bar: 100  $\mu\text{m}$  (C). Most breast carcinomas originate in the luminal compartment. A significant minority of patients are diagnosed with non-invasive DCIS, in which transformed cells are still enclosed within the duct. Progression to invasive disease occurs in a minority of pure DCIS as transformed LEP bypass the MEP and basement membrane (D).

The transition from the epithelial to the stromal microenvironment is conducive to cancer invasion. The epithelial compartment is associated with tumor suppressive signals secreted by MEP (Hu et al., 2008; Cowell et al., 2013), while stromal secreted factors and ECM remodeling have been implicated in cancer aggressiveness (Bussard et al., 2016). Furthermore, the physical composition of the stroma itself promotes cancer progression. Fibrous collagen I is known to entrain directed migration along fibers in both normal and invasive mammary cells and propagate cellular traction forces over long distances, aiding collective



migration and ECM remodeling (Nguyen-Ngoc and Ewald, 2013; Riching et al., 2014; Shi et al., 2014). The basement membrane acts as a physical barrier to invasion in the absence of protease or force-mediated disruption (Chang et al., 2017; Chang and Chaudhuri, 2019). In addition, live imaging studies in mouse breast cancer organoids show that normal MEP dynamically self-organize around LEP, greatly decreasing the dissemination of invasive tumor cells in a dose-dependent manner, indicating their direct role in tumor suppression by dynamic control of tissue architecture (Sirka et al., 2018). Unsurprisingly, it is the transition to invasive disease – the structural breach of the MEP layer and surrounding basement membrane, and the entry into the stroma – that marks the critical increase in risk for patients (Onega et al., 2017). In an intact tissue architecture, transformed LEP can hyperproliferate in the lumen in the commonly diagnosed ductal carcinoma in situ (DCIS) stage, but the majority of pure DCIS lesions never progress to invasive disease (Sanders et al., 2005) (**Fig. 1.1D**). Predicting which DCIS patients are likely or unlikely to progress would reduce the morbidity associated with prophylactic surgery, radiation, and chemotherapy, and is currently an active field of research with a paucity of clear molecular markers (Kuerer, 2015; Groen et al., 2017). Thus, understanding normal tissue architecture in the mammary gland will provide insight into predicting and preventing its disruption in the transition from in situ to invasive disease.

## **1.2: Cell sorting in tissues**

Biological systems generate complex, functional structures from initially disorganized or uniform cells. Cell identities and positions are determined through multiple self-organizing processes. One such process is cell sorting, where cells of different identities separate into distinct compartments of the tissue. Cell sorting has been the focus of a long line of investigation, from Wilson’s dissociated sponges in 1907 to Holtfreter’s amphibian embryos in 1939-1955, which showed that cells from different germ layers would spontaneously re-segregate from initially randomly ordered cell aggregates (Townes and Holtfreter, 1955). Holtfreter

concluded that sorting owed to something he termed tissue affinity, “which may serve as a reminder for the existence of analogous phenomena in chemistry,” laying the groundwork for a physical understanding of tissue self-organization (Steinberg and Gilbert, 2004). In a series of experiments investigating tissue sorting in the 1960s, Steinberg proposed that this tissue affinity corresponded to the quantitative surface tension of a tissue aggregate generated by adhesion between its component cells. In sorting, contacts between the most cohesive cells would be maximized, at the expense of weakly cohesive ones, and in 1970, Steinberg showed that the relative sorting of six different embryonic tissues could be predicted by their relative strengths of tissue cohesion.

This differential adhesion hypothesis would be further built upon and developed in the decades to follow, which included the discovery and characterization of the proteins responsible for cell adhesion, such as cadherins, as well as a growing understanding of the biomechanics of the cellular actin cytoskeleton. Harris (1976) proposed that rather than cell adhesion strength, cell sorting owed more to the contractility of the cortical actin network, coupled to cell adhesion molecules. With the refinement of increasingly sensitive biophysical techniques, from micropipette aspiration to atomic force microscopy, and specific inhibitors of actin and myosin, it became easier to quantify and disentangle the relative contributions of adhesion strength and cell contractility. These quantitative studies found that the weaker forces associated with adhesion molecules alone must be accompanied by differences in active contractile forces to drive cell sorting and tissue shape changes (Krieg et al., 2008; Youssef et al., 2011; Maître et al., 2012). The modern conception of cell sorting considers cell adhesion proteins to effect sorting by coupling the underlying actomyosin networks to the cell surface and by signaling changes in cytoskeletal architecture that ultimately provide the mechanical forces that rearrange cells (Yamada and Nelson, 2007; Winklbauer, 2015). The resulting differential interfacial tension hypothesis (DITH), formally proposed by Brodland (2002), harmonizes the combined biophysical contributions of cell cytoskeletal components, adhesion molecules, and active contractility.

The DITH remains a widespread and useful concept in understanding how cells reach and maintain their positions in tissues. A considerable array of different model organisms and systems have implicated DITH in a variety of morphogenetic processes. For example, the *Drosophila* wing antero-posterior compartment boundary is enforced by increased interfacial tension between cells of different compartments, measured by laser ablation to be approximately 2.5-fold higher (Landsberg et al., 2009). Similarly, Eph-ephrin signaling plays a key role in sharpening the boundary between tissues by decreasing the favorability of heterotypic contacts. In *Xenopus*, Eph/Ephrin signaling inhibits cadherin clustering and increases local actomyosin contractility between the notochord and presomitic mesoderm tissues, forcing the two tissues to stay distinct and sharply separated (Fagotto et al., 2013). Kindberg et al. (2021) found that not only was interfacial tension increased at heterotypic EPHB2:EPHRIN-B1 contacts, but tension at the cell-medium interface was increased in EPHB2 cells, driving both increased cohesion of EPHB2 cells and decreased contacts between the two cell types. Increasing the favorability of cohesive interactions between cells of the same type is also associated with developmental boundary formation. The zebrafish spinal cord initially uses a sonic hedgehog morphogen gradient to trigger differentiation of three progenitor populations, roughly located in zones of increasing distance from the morphogen source. The final, sharply delineated regions of progenitors are ensured by homotypic cell sorting of these progenitors, each of which has higher homotypic than heterotypic adhesion forces caused by their type-specific expression profiles of three cell-cell adhesion molecules, *cdh2*, *cdh11*, and *pcdh19* (Tsai et al., 2020). Favoring contacts between different cell types has the opposite morphogenetic outcome. During convergent extension in *Drosophila*, a Toll receptor code stabilizes heterotypic adhesion, thereby promoting the intercalation of cells with different receptor profiles to elongate the body axis (Paré et al., 2014). Dynamic alterations of cell adhesion in response to cellular neighborhoods can generate cell sorting phenotypes that change over time, as demonstrated by Toda et al. (2018). This synthetic Notch-driven adhesion system contained cell contact-driven feedback loops such that cells adopted mechanical

properties, sorted according to their energies, received signals from their new neighboring cells, then adopted new phenotypes that either reinforced or changed their sorting behavior. Of additional interest are non-cell autonomous phenomena, where cell collectives develop different mechanical phenotypes than isolated cells. In *Hydra vulgaris*, groups of ectodermal cells develop an epithelialized phenotype with a polarized non-adhesive apical surface and preferentially engulf other cells via collective cell spreading, including isolated ectodermal cells that have not undergone epithelialization (Skokan et al., 2020).

Our lab has previously explored the parameters and predictions of the DITH, in particular its extension to cell-matrix interactions, using microtissues of cultured primary human mammary epithelial cells (HMEC), which provide an excellent model system for studying cell sorting. Unlike immortalized lines like MCF10A, these finite-lifespan, early-passage cells derived from normal reduction mammoplasties retain the two major cell types of the breast epithelium, luminal and myoepithelial cells (LEP, MEP) (Stampfer et al., 2013). The separation of outer MEP from inner LEP is driven by a strong difference in the interfacial energies of each cell type with the surrounding ECM. When reaggregated and embedded in an adhesion-supportive ECM, LEP and MEP self-organize in hours to days into a core-shell configuration, with MEP contacting the ECM as they would in the in vivo tissue. LEP and MEP self-organize by exchanging neighbors and their layered position in the tissue is due to the relative interfacial energies of each cell-cell and cell-ECM interface. In HMEC, the interfacial energy between cells and ECM is central to determining tissue structure: the LEP-ECM interfacial energy is highly unfavorable compared to that of MEP-ECM. Unlike commonly described cell-cell adhesion-based systems, HMEC organoids organize very differently depending on the presence or absence of ECM. In the absence of ECM – in a non-adhesive agarose microwell – the cohesive MEP sort to the center. With an adhesion-permissive environment, the MEP will sort to the outer layer, as they do in vivo (Cerchiarri et al., 2015).

Cell sorting is also driven by other processes which have been observed and proposed in various tissue systems. In particular, the DITH is primarily based on static cell properties, rather than dynamics, such as directed cell migration or differences in cell motility or dynamics. Contact-mediated repulsion driven by Delta/Notch has been implicated in cell sorting in zebrafish stripe formation, forming the long-range inhibition component of a cellular Turing pattern (Hamada et al., 2014). In HEK293 cells, heterotypic Eph/ephrin signaling triggers localized collapse of cell processes, or contact inhibition of locomotion, and repulsive migration between the signaling cells, a mechanism that is predicted to be more efficient at sharpening tissue borders than DITH alone (Taylor et al., 2017). During gastrulation, mesoderm involution into the ectoderm is dependent primarily upon directed cell motility in a Rac-dependent manner, while tension-dependent cell sorting of mesoderm progenitor and ectoderm progenitor cells in vitro is unaffected by dominant-negative Rac but strongly affected by the osmolarity of the culture medium (Krens et al., 2017). Undirected differences in cellular motility have also been modeled to drive sorting in cells and active particle systems (Stenhammar et al., 2015). Cells with high directional persistence spatially sort to the ends of a microtissue’s long axis (Mori et al., 2009). In the mouse blastocyst, the sorting of cells in the inner cell mass may be driven by both a DITH mechanism as well as differential surface blebbing associated with localized, subcellular heterogeneity in cortical actin (Yanagida et al., 2020).

Many computational tools have been developed to model cell sorting, including subcellular lattice models, e.g., Cellular Potts (Graner and Glazier, 1992), polyhedral lattice models (Flenner et al., 2012), finite-element models (Chen and Brodland, 2000; Viens, 2007), vertex-based meshes (Fletcher et al., 2014; Neumann et al., 2018), and self-propelled particles and Voronoi tessellations (Belmonte et al., 2008). These models typically incorporate cellular forces (interfacial tensions, cell motility) and mechanical constraints (area/volume conservation, tissue boundaries, cell and substrate elasticity and viscosity) into a global Hamiltonian or energy functional, or into a total force acting on the individual cell for cell-centric meth-

ods. The resulting energy differences between cell configurations or force imbalances acting on cells lead to positional changes and sorting towards the minimum energy structure.

However, traversing an energy landscape is not without its hazards, as is appreciated in molecular structure prediction, where metastable states and kinetic bottlenecks are common (Caves et al., 1998; Jack et al., 2007; Shaffer et al., 2016; Coveney and Wan, 2016). In deterministic fluctuation-free models of cell sorting, cells only move due to the action of explicit forces, e.g., imbalances in interfacial tensions, meaning that once a tissue reaches an energy minimum, it will not randomly exit it (Hutson et al., 2008). These systems are sensitive to local energy minima: for example, a partially sorted tissue with multiple cores might have a global energy minimum with all cores merged into one, but in order to connect the separate cores, the tissue would need to temporarily increase its heterotypic area to create a bridge. Deterministic energy minimization alone may not be sufficient to lead to robust sorting, except in small and highly connected systems. Recent work on glass-like behavior in tissues has shown that highly confluent tissues can be “jammed” by cell shape constraints and kinetically trapped away from an energy minimum (Bi et al., 2015; Park et al., 2015; Ilina et al., 2020), even in the presence of collective cell migration (Angelini et al., 2011). Such a system may not reach a hypothetical equilibrium distribution because it may be jammed in its initial configuration, and not all microstates are equally accessible to it.

Tissue activity or energy fluctuation is frequently included in models of cell sorting as a parameter that enables sampling of the configuration space, even against an energy gradient. Effectively, it captures the occurrence of random cellular positional changes. In the Cellular Potts model, for example, it exists as a temperature  $T$  that modifies the probability of lattice site-switching (Graner and Glazier, 1992). Agent-based systems like self-propelled particles incorporate more complex parameters of cell motility, such as a cell-intrinsic semi-random driving force with directional persistence (Belmonte et al., 2008). Tissue activity has two main consequences: first, creating dynamic fluctuations about the steady state, and second,

allowing the system to adequately sample the configuration space and approach the steady state from an initially distant state.

### **1.3: Statistical mechanics in biology**

Dynamic, nonequilibrium systems are often of great interest in biology, in particular their ability to reach and maintain not-so-dynamic homeostatic states. Despite the fact that active biological systems are inherently nonequilibrium at the molecular level, it has been observed that nonequilibrium systems can regain equilibrium behavior at the macroscopic level, including detailed balance, fluctuation dissipation, and Gibbs distributions, given the appropriate choice of spatiotemporal scale (Egolf, 2000). Dynamic nonequilibrium systems can develop stable cycles or steady states with significant robustness to fluctuations, and indeed this behavior is essential for the active maintenance of biological homeostasis (Wang et al., 2008).

Nonequilibrium steady states (NESS) are stationary (time-invariant) states of nonequilibrium systems and can be described in similar terms as equilibrium theory (Oono and Paniconi, 1998). They feature a steady state that is maintained by constant mass or energy inputs from external driving forces and dissipation of heat energy out of the system. These forces and fluxes are seen in active systems containing internal motors or periodic driving, like molecular motors in cells (Battle et al., 2016), or systems with consistent transfer of mass or energy in and out of the system, like the flow of blood passing through a certain stretch of blood vessel or populations of cells transitioning between different states while maintaining the overall population structure (Hayashi et al., 2008; Chang and Marshall, 2019).

Statistical mechanics links the macroscopic observed quantities of a system to the statistical behavior of its microscopic components (Frigg, 2008; Uffink, 2007). Concepts from statistical mechanics have been used in describing the behavior of biological systems, from dynamics to steady states, including protein folding, gene transcription, oncogenic mutations in protein-protein interaction networks, and cell pluripotency (Frauenfelder et al., 1991; Bintu

et al., 2005; Vilar and Leibler, 2003; AlQuraishi et al., 2014; Teschendorff and Feinberg, 2021). It provides a generalizable approach to describing the probability distribution of a system as a function of the energy of its possible states and the number of microstates consistent with the macrostate.

A *microstate* represents the low-level arrangement of the system. Examples of microstates in biological systems include the locations of ligands in solution, the locations of transcription factors bound to the genome, the nucleotides of a gene sequence, and the levels of gene expression in a cell. The relative probability of each possible microstate is the Boltzmann weight,  $e^{-\beta\epsilon}$ , where  $\epsilon$  is the energy of the microstate and  $\beta$  is a measure of the random fluctuations of the system. Not every individual microstate is of biological interest. Typically, we are interested in sets of significant microstates, for example: the set of synonymous mutations, the set of all ligand configurations where one ligand is bound to a receptor, or the set of all transcriptomic profiles consistent with a particular cell state (**Fig. 1.2**). These sets of microstates are *macrostates*, which capture some “macroscopic” property of interest, such as receptor occupancy or cell state, by combining many functionally equivalent microstates. Naturally, the relative probability of such a macrostate  $i$  is the sum of the Boltzmann weights of the component microstates,  $P_i = \sum e^{-\beta\epsilon}$ . This can be stated as

$$P_i = \frac{W_i}{Z} e^{-\beta E_i} \tag{1.1}$$

where  $W_i$  is the number of microstates in the macrostate  $i$ ,  $E_i$  is the relative energy of the macrostate  $i$ , and  $Z$  is the partition function defined such that  $\sum_i P_i = 1$ .



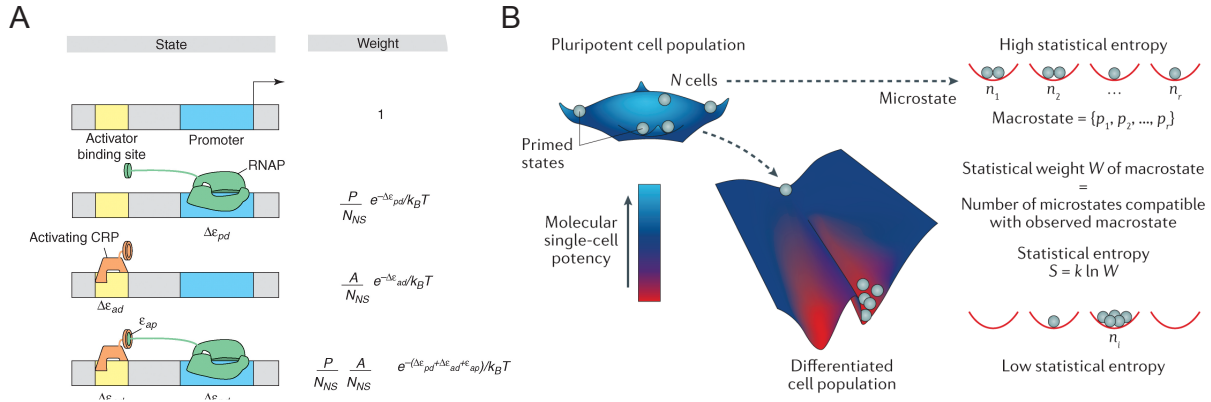


Figure 1.2: Some examples of statistical mechanics applied to problems in biology. Biological phenomena can be considered to take place in different states, such as bound/unbound, or naive/differentiated. Statistical mechanical analysis can be performed to determine the relative probabilities of different states of interest by comparing their relative energy costs and multiplicity. Energy can be directly related to physical energies, like protein-DNA binding energies (A), or more abstract, like the activating-inhibiting interactions of a gene regulatory network (B). Figures are reproduced from Bintu et al. (2005) and Teschendorff and Feinberg (2021).

The “energy” is often estimated using binding rates or molecular interaction energies, but need not be directly related to kinetic or potential energies at all. For example, the iconic image of cell differentiation as balls rolling down phenotypic valleys in a three-dimensional epigenetic landscape was introduced by Waddington (1957). It remains a compelling image today, as modern techniques and statistical mechanics are applied to characterizing these “energy” valleys as transcriptional and epigenetic attractors formed by gene regulatory networks. For relatively small gene regulatory networks, energy landscapes can be predicted from known or estimated activating and inhibitory interactions (Bhattacharya et al., 2011; Li and Wang, 2013). Molecular binding energies can be more literally translated into an energy landscape. AlQuraishi et al. (2014) applied a statistical mechanical approach to predicting the functional effects of SH2 domain mutations on protein-protein interaction networks. Beginning with first-principles biophysical data on binding assays and crystal structures, they developed a Hamiltonian to predict changes in binding affinity for mutant SH2 proteins for all possible protein configurations. Once combined with bioinformatic network information

on protein-protein interaction networks and pathways, their model represented a description of systems-level phenotypic changes based upon the energies and multiplicities of the possible mutants affecting protein binding, thereby establishing a link between microstates (mutant proteins and their paired interactions) and macrostates (downstream effects on pathway activity).

Given the tremendous complexity of the biological processes underlying many energy landscapes of interest, it is not always feasible to predict their structures from first principles, so the problem may be inverted in biological practice: given the observed state occupancies, what is the underlying energy landscape (Chau Nguyen et al., 2017)? Complex cell states spanning many genes, for which small gene regulatory networks are insufficient, are now being probed using high-dimensional and single-cell biology. High-dimensional datasets probing chromatin accessibility or transcript expression of large numbers of single cells can be used to infer the differentiation potential landscapes underlying observed cell distributions or inferred cellular trajectories as cells move towards energy minima (Lang et al., 2014; Jenkinson et al., 2017; Jin et al., 2018; Teschendorff and Feinberg, 2021). Similarly, energy landscape inference has been used for high-dimensional viral mutation fitness analysis, which lacks a robust bottom-up approach to estimating functional outcomes. In paired studies, Shekhar et al. (2013) and Ferguson et al. (2013) used a Boltzmann analysis of the observed prevalence of HIV strains in patients to infer the fitness landscape associated with specific mutational states, using the relation  $E_i - E_{WT} = -\ln \frac{P_i}{P_{WT}}$ . Their model predictions were comparable to experimental in vitro HIV replication data generated in-house and from the literature, and corresponded to clinically relevant features of HIV strains in vivo, including escape mutants, the pathway of viral evolution in individual patients, and high fitness costs to viral evolution in elite controller patients. These inferred energy landscapes can supplement or validate predictions from first principles, and using both approaches together can provide a compelling description of the putative energetics of the system.

What about active systems, where the system’s position in state space is influenced not only by the underlying energy landscape but also by the periodic or stochastic driving forces it experiences? In gene expression space, such fluctuations can arise from transcriptional noise that transiently drives the cells away from an attractor steady state. Such systems are of interest in modeling pluripotency, as stem cells stochastically leaving the steady state may be differentially susceptible to differentiation, even while the ensemble as a whole maintains a fluctuation-resistant, stable distribution of cells with dynamically varying states (Chang et al., 2008; Kalmar et al., 2009). In physical state spaces, cells metabolize chemical energy and use it to carry out driven motion, which ultimately leads to heat dissipation and entropy generation in the surroundings. At small length scales, where active motor fluctuations and thermal fluctuations are of similar orders of magnitude ( $\sim 1$ - to  $100$ -fold  $k_B T$ ), the nonequilibrium nature of the observed states can only be detected by violations of the equilibrium requirements of detailed balance or the fluctuation-dissipation theorem, by observing net probability fluxes in state space or nonlinear relationships between spontaneous and induced fluctuations respectively (Seifert, 2012). The activity can generate an apparent or effective temperature very similar to the thermodynamic concept, as though the NESS were a closed system at equilibrium at  $T_{\text{EFF}}$ , where  $T_{\text{EFF}} \neq T$  indicates a nonequilibrium state maintained by active processes rather than thermal noise (Cugliandolo, 2011; Gnesotto et al., 2018). For example, hair cells’ mechanosensitive hair bundles oscillate to amplify sound and adjust sensitivity. Martin et al. (2001) observed that frequency-stimulated hair bundle fluctuations displayed active amplification of oscillations at high frequencies,  $T_{\text{EFF}} > T$ , and active dampening at low frequencies,  $T_{\text{EFF}} < T$ , while control hair cells’ motion was consistent with the external temperature. Similarly, Turlier et al. (2016) used active and passive microrheology to measure the active contribution to red blood cell membrane flickering, a phenomenon previously thought to reflect thermal fluctuations. Such studies demonstrate that active inputs can generate NESS strongly resembling equilibrium behavior.

## 2. Methods

### 2.1: HMEC material and cultivation

De-identified cultured normal finite-lifespan primary HMEC were provided by Drs. Martha Stampfer and James Garbe (Lawrence Berkeley National Laboratory). The majority of experiments were performed using cells derived from benign surgical discard material from a reduction mammoplasty of a 19 year old individual with bilateral breast hypertrophy (specimen 240L). HMEC were cultured in M87A medium with cholera toxin and oxytocin exchanged every other day as previously described (Stampfer et al., 2013), with penicillin/streptomycin beginning at passage 4. All cultured primary HMEC were used for experiments at passage 4.

LEP and MEP were isolated by fluorescence-activated cell sorting (FACS), modified from Cerchiari et al. (2015). Briefly, at 80 – 90% confluence in 10-cm dishes or 6-well plates, HMEC were gently rinsed once with 4 ml of 0.05% trypsin and incubated with 2 ml trypsin at 37°C until cells were detaching from the substrate, or in half the volume for 6-well plates. Cells were detached by rinsing with 10 ml M87A medium and passed through a 40- $\mu$ m nylon filter into a 50-ml centrifuge tube on ice. Cells were centrifuged at 160 $\times$ g for 4 min then incubated in FACS antibody staining solution in M87A for 30 min on ice. FACS antibody cocktail comprised 10  $\mu$ L/ml APC/Fire 750 CD271 and Pacific Blue EpCAM mouse monoclonal antibodies (BioLegend 345116 and 324218). Cells were washed with 10 ml of PBS and resuspended in 500  $\mu$ L FACS buffer per 10-cm dish (PBS + 2% w/v BSA). Cells were sorted on a BD FACSAria III fitted with a 100- $\mu$ m nozzle running FacsDiva software using the 4-way purity setting. GFP-positive cells were gated as GFP<sup>+</sup>/mCherry<sup>-</sup> and mCherry-positive cells were gated as GFP<sup>-</sup>/mCherry<sup>+</sup>. LEP were gated as EpCAM<sup>high</sup>/CD271<sup>low</sup> and MEP were gated as EpCAM<sup>low</sup>/CD271<sup>high</sup>, as shown in **Fig. 2.1**.

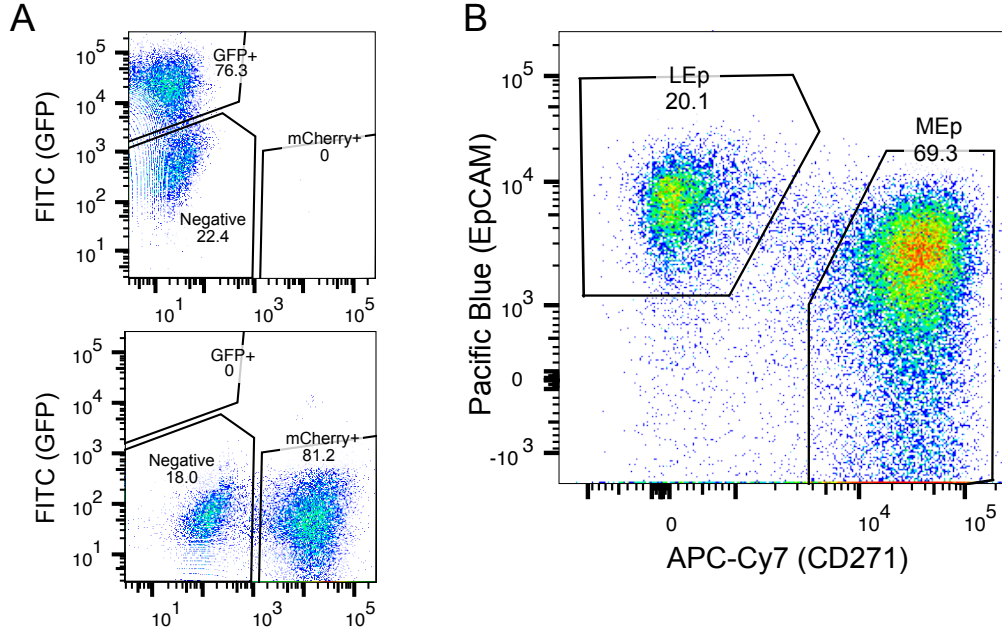


Figure 2.1: Representative flow cytometry plot of GFP or mCherry-transduced fourth passage HMEC (A) stained with APC/Fire 750 CD271 (MEP marker) and Pacific Blue EpCAM (LEP marker) (B).

## 2.2: Viral transduction of HMEC

Transgenes were cloned into the pSicoR lentiviral backbone using BsiWI/XbaI restriction sites under the EF1 $\alpha$  promoter and short hairpin RNA (shRNA) sequences under the U6 promoter, as mapped in **Fig. 2.2**. Enhanced GFP or mCherry reporter fluorescent proteins were separated from the insert by a P2A ribosomal skipping sequence. Transgene expression and shRNA knockdown was confirmed by Western blot. Lentiviral vectors with a puromycin resistance marker were prepared for a subset of shRNA sequences to be co-transduced with a different transgene marked by GFP.

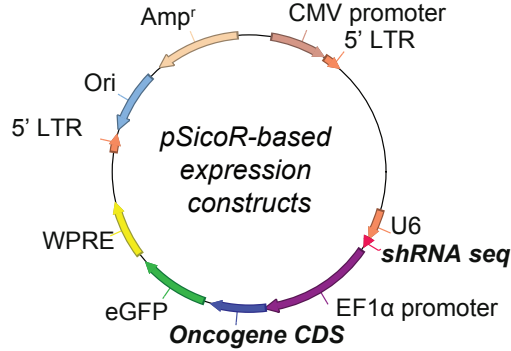


Figure 2.2: Vector map of the pSicoR lentiviral backbone. Oncogenes were inserted downstream of the EF1 $\alpha$  and shRNA sequences were inserted downstream of the U6 promoter.

Concentrated third-generation lentiviral vectors were purchased from UCSF Viracore or generated in-house by triple transfection of HEK293T/17 cells (UCSF Cell Culture Facility) grown in DMEM + GlutaMAX (ThermoFisher Scientific 10566016) with 10% FBS and 1mM sodium pyruvate, and switched to lentivirus packaging medium (Opti-MEM + GlutaMAX (ThermoFisher Scientific 51985034) with 5% FBS and 1mM sodium pyruvate) 12 – 16 hr before transfections. Cells at 80 – 90% confluence were transfected with plasmids for the gene of interest, pCMV-VSV-G (Bob Weinberg, Addgene plasmid 8454) and pCMV delta R8.2 (Didier Trono, Addgene plasmid 12263) using Lipofectamine 3000 (ThermoFisher Scientific, L3000015). Supernatant was harvested and filtered after one and two days, then centrifuged with Lenti-X Concentrator (Takara Bio 631231). Viruses were titered by serial dilution and infection of 10,000 passage 5 MEP per well of a 96-well plate. Small volumes were aliquoted and stored at  $-80^{\circ}\text{C}$ . Thawed aliquots were stored at  $4^{\circ}\text{C}$  for up to one week. For transgenic HMEC, frozen ampules of HMEC were thawed into plates at 2500 – 3500 cells/cm<sup>2</sup> in a half-volume of M87A medium containing 2  $\mu\text{g}/\text{ml}$  polybrene and lentivirus, initially at 1 MOI based on titer, and afterwards adjusted for real performance to target 50-80% transduction efficiency. After 3 hr, M87A medium was added to full volume. Virus medium was allowed to remain on cells until 2 days after initial seeding. Puromycin selection was initiated at 2 days at 1  $\mu\text{g}/\text{ml}$ .

### **2.3: Photolithography and PDMS stamps**

Silicon masters were prepared matching the desired agarose features, either rectangular cell-cell microwells of dimensions  $10\ \mu\text{m} \times 20\ \mu\text{m} \times 20\ \mu\text{m}$ , or cylindrical aggregation microwells of approximately  $200\ \mu\text{m}$  depth and  $120$  or  $180\ \mu\text{m}$  diameter. SU-8 UV-curing resin (MicroChem) was distributed on pre-cleaned silicon wafers using a spin coater. SU-8 2050 was spun 30 seconds at 1000 rpm for  $200\ \mu\text{m}$ -high features, while SU-8 2025 was coated at 4000 rpm for approximately  $20\ \mu\text{m}$ -high features. The wafer was baked at  $95^\circ\text{C}$  to remove excess resist solvent, then exposed under a 365-nm UV lamp with a printed photomask (CAD Art Services) for approximately 6 min for thick features and 2 min for thin features. After baking at  $95^\circ\text{C}$  for 1 min, cooled wafers were submerged in PGMEA developer at a depth of 2 cm on a rotary shaker until uncured resist was dissolved away. Excess was rinsed off with PGMEA developer, then the wafer was rinsed with isopropanol, dried, and baked for a few minutes at  $95^\circ\text{C}$  to remove excess solvent. PDMS polymer and crosslinker at a 10:1 ratio were poured on the silicon master to a depth of approximately 1 cm, degassed in a vacuum desiccator, and baked at  $60^\circ\text{C}$  overnight. The PDMS was removed, cut into stamps approximately  $1\ \text{cm} \times 0.5\ \text{cm}$ , and stored in 75% ethanol. Agarose microwells were cast from 3% w/v high-melting point agarose (AllStar Scientific 490-050) in PBS by dropping PDMS stamps into molten agarose and allowing to set at  $4^\circ\text{C}$ . PBS was added and the PDMS stamps were removed with curved-tip tweezers. Agarose microwells were stored under PBS at  $4^\circ\text{C}$  and used within one day of preparation.

### **2.4: Reconstituted mammary organoids**

Fluorescent protein-expressing HMEC were reconstituted into organoids, modified from Cerchiari et al. (2015). Sorted GFP and mCherry cells were combined in M87A media at a 1:1 ratio at a total cell concentration of 300,000 cells/ml. Approximately  $100\ \mu\text{L}$  of the cell suspension was pipetted onto agarose microwells in a 24-well plate and centrifuged twice, facing

opposite directions to distribute cells evenly across the microwells, at  $160\times g$  at  $4^{\circ}\text{C}$  at half acceleration in a swinging-bucket rotor centrifuge. Excess cells were removed by aspiration of the remaining medium before culturing the cells in microwells in M87A medium until they formed cohesive cell aggregates, approximately 4 hr for passage 4 HMEC. After aggregation, cell clusters could be transferred to 3D culture. First, 10  $\mu\text{L}$  of Matrigel was spread in a thin layer in the center of a glass-bottom culture chamber (either a 24-well plate or an 8-chamber coverslip), and allowed to gel at  $37^{\circ}\text{C}$  for at least 30 minutes. Microwells were flushed by pipetting 1 ml of PBS over the microwells and transferring to a 15-ml conical. Trituration was performed twice with a wide-bore p1000 tip then once with a standard p1000 tip, until most microwells were empty by visual inspection. The collected aggregates were centrifuged briefly at  $15\times g$  at moderate acceleration and deceleration in a swinging-bucket rotor centrifuge. The supernatant was discarded and the aggregates resuspended in ECM using a wide-bore p200 tip, typically Matrigel at 9 mg/ml, or Matrigel with 2 mg/ml collagen-I. The aggregates were distributed over the pre-gelled Matrigel underlay in a 40  $\mu\text{L}$  droplet and incubated at  $37^{\circ}\text{C}$  for 30 – 45 minutes before being covered in M87A medium. For drug and growth factor treatments, a  $100\times$  concentrated solution was mixed into the ECM, i.e., 0.4  $\mu\text{L}$  in 40  $\mu\text{L}$  of Matrigel, and mixed thoroughly before plating, in order to reduce potential for uneven diffusion of compounds through ECM. Aggregates were cultured with M87A media changes every other day. To generate pre-segregated aggregates, GFP cells and mCherry cells were spun in separate 120  $\mu\text{m}$  microwells and incubated at  $37^{\circ}\text{C}$  for 3 hr before the single-color aggregates were flushed out and recombined into 240  $\mu\text{m}$  microwells and incubated again for 3 hr.

## **2.5: Microscopy**

Organoids were imaged for cytoplasmic fluorescent proteins GFP and mCherry at 2 days and 7 days on either a spinning disk confocal microscope or a laser scanning confocal microscope running ZEN software. Z-stacks contained approximately 7 to 11 slices spaced 5  $\mu\text{m}$



apart. Live imaging was performed under environmental control with 5% CO<sub>2</sub> at 37°C with timepoints every 5 to 20 minutes apart.

Technical limitations generally prevent long time-lapse acquisitions of living organoids, which experience phototoxicity and declining health after several hours of imaging. Organoids that have undergone live imaging are typically less sorted than organoids left in the incubator for the same amount of time. To reduce experimental variability due to cell health, we restrict our analysis of the approach to steady state to a time window of eight hours, and for steady state begin analysis after 48 hr.

## **2.6: Immunofluorescence**

For 2D immunofluorescence, cells were fixed in 4% PFA for 15 min at room temperature, permeabilized with 0.5% Triton X-100 for 5 min at 4°C, blocked with blocking buffer (wash buffer + 10% heat-inactivated goat serum) for 1 hr at room temperature, incubated with primary antibody in blocking buffer overnight at 4°C, rinsed in wash buffer (PBS + 0.1% w/v BSA + 0.2% v/v Triton X-100 + 0.041% v/v Tween-20) 3 times for 10 min at room temperature, incubated with secondary antibody in blocking buffer for 30 min at room temperature, rinsed in wash buffer 2 times for 10 min at room temperature, and incubated with DAPI or phalloidin in PBS for 30 min in room temperature. For 3D immunofluorescence, cells were fixed in 2% PFA warmed to 37°C and allowed to cool to room temperature over 45 min, permeabilized with 0.5% Triton X-100 for 15 min at room temperature, blocked with blocking buffer for 2 hr at room temperature or overnight at 4°C, incubated with primary antibody in blocking buffer overnight at 4°C, rinsed in wash buffer 3 times for 1 hr at room temperature, incubated with secondary antibody in blocking buffer overnight at 4°C, and rinsed in wash buffer 3 times for 1 hr at room temperature. Samples were stored in PBS at 4°C prior to imaging.

## 2.7: Interfacial tension measurements

Cortical tensions were measured by micropipette aspiration of single cells in suspension (Srivastava and Robinson, 2015). Glass micropipettes of 8  $\mu\text{m}$  to 11  $\mu\text{m}$  inner diameter were pulled and microforged from glass capillary tubes, then mounted into a Narishige MM-92 micromanipulator and stage mounted to a Zeiss Axiovert 200M running SlideBook software. Cells were either used immediately after lifting or FACS sorted and held on ice for a total of no more than 5 hr after initial passaging. Cells were resuspended at 5000 cells/ $\mu\text{m}^2$  and warmed to room temperature before aspiration. A cell was aspirated into a glass micropipette at initial pressure of 0.03 kPa, and pressure was increased in 0.03 kPa steps until the cell was sufficiently deformed. At each pressure, after waiting for 30 seconds, three images were taken using a 40 $\times$  air objective. The average cellular deformation inside the pipette ( $L_p$ ) was measured in FIJI, and critical pressure ( $\Delta P_{\text{crit}}$ ) was identified as the aspiration pressure where  $L_p = R_p$ , pipette inner radius. Any cells that blebbed were discarded. Cortical tension was calculated as  $\gamma_c = \Delta P_{\text{crit}} \left( \frac{1}{R_p} - \frac{1}{R_c} \right)^{-1}$ .

Contact angle measurements were modified from Cerchiari et al. (2015). For cell-ECM measurements, 8-chambered coverslips (Nunc Lab-Tek II, ThermoFisher Scientific) were covered in reduced growth factor Matrigel diluted in M87A medium at 2% v/v (0.18–0.2 mg/ml) at room temperature for at least 2 hr during FACS sorting. The coverslips were washed once with M87A medium prior to seeding FACS-sorted LEP and MEP at 5,000 cells per 0.7  $\text{cm}^2$  chamber. The coverslips were washed again and cultured until 4 hr, when they were fixed in 4% PFA and immunostained for cytokeratin 14 and cytokeratin 19. Healthy, radially symmetric cells not touching nearby cells were imaged with a 63 $\times$  oil objective (NA 1.4) on a spinning disk confocal Zeiss Cell Observer Z1, Yokagawa spinning disk microscope. Slices in  $z$ -stacks were taken 0.26  $\mu\text{m}$  apart and spanned the cell surface and most of the cell height. Cytokeratin 14 and cytokeratin 19 were used to verify cell identity of MEP and LEP, respectively. The contact angle at the cell-ECM interface was measured with the FIJI angle

tool at four points around the circumference of each cell, taken from coronal and sagittal projections, using the gross shape of the cell and disregarding thin, flat membrane extrusions. The four measurements were averaged on a per-cell basis for analysis. Asymmetrical cells with standard deviation greater than  $30^\circ$  were excluded.

For cell-cell measurements, FACS-sorted GFP and mCherry cells were centrifuged into small rectangular agarose microwells designed to hold two cells ( $20\ \mu\text{m} \times 40\ \mu\text{m} \times 20\ \mu\text{m}$ ) and cultured for 4 hr. With environmentally controlled live imaging, or after fixing in 4% PFA for 10 min at room temperature, pairs of similar-size cells were imaged with either a  $20\times$  air objective or a  $40\times$  water objective on the spinning disk confocal at the center plane of contact. The FIJI angle tool was used to measure four contact angles at each cell-cell interface, which were averaged for each cell doublet for analysis.

## 2.8: Image processing and quantification

Initially, organoid  $z$ -stacks were qualitatively categorized as correct, mixed, split, inverted, and in between, using a random channel-swapping approach to blind the user to the true identity of LEP and MEP pixels. Center  $z$ -slices were visually identified by size. The center three slices (the center and one above and below, or the center two slices if the center was at the edge of the  $z$ -stack) were isolated from each  $z$ -stack and prepared for segmentation by background subtraction and automatic contrast adjustment in FIJI. The estimated error associated with using a three-slice average of the organoid center cross-sections ( $-5, 0, 5\ \mu\text{m}$ , spanning  $10\ \mu\text{m}$  in the  $z$ -axis) is discussed in Appendix E. LSM images were additionally processed with a Gaussian blur to reduce the pixel noise associated with LSM. Channels were standardized by swapping where reasonable, so that an organoid with mCherry LEP and GFP MEP would be segmented the same as an organoid with GFP LEP and mCherry MEP, yielding pixels classifications by cell type, not fluorescent marker.

Separate Ilastik pipelines were generated for each of two microscopes (spinning disk confocal and laser scanning confocal) and two levels of fluorescent protein expression (bright

and dim) to account for image variation between microscopes and constructs (Berg et al., 2019). Ilastik’s pixel classification module uses a Random Forest classifier on pixels with pre-defined features, allowing for fast interactive annotation. The training datasets were generated from random samples of images from different outcome categories to ensure an even representation of different organoid configurations. The training dataset was sparsely labeled into four pixel categories: MEP, LEP, ECM, and hole. The hole category denotes regions within the organoid with very low signal, such as lumens, vacuoles, or the interiors of large organoids where light penetration is weak. Image features were set to two-dimensional filters of the background-subtracted two-channel fluorescence data. Annotation of the training data proceeded until regions of uncertainty were localized to boundaries between pixel classes, and high certainty was established in all other areas.

Segmentations were exported and processed in MATLAB. Due to natural variations in fluorescence and cell morphology, some regions are very difficult to segment either manually or automatically. Processing was used to smooth the border of the organoid and reasonably infer unassigned or ambiguous pixel identities. It removed labeled areas outside of the main organoid, removed small speckles, reassigned internal holes as either cells or ECM by interpolation, and used a disk structuring element to smooth out crevices in the organoid that can be replaced with cell pixels. The majority of images required minimal processing, unlike the more challenging partially lumenized all-LEP mixed organoid shown in **Fig. 2.3**, as an example of these image processing steps. Pixel classifications were in general accurate enough to describe the overall structure of the organoid and the proportion of the tissue edge occupied by each cell type.

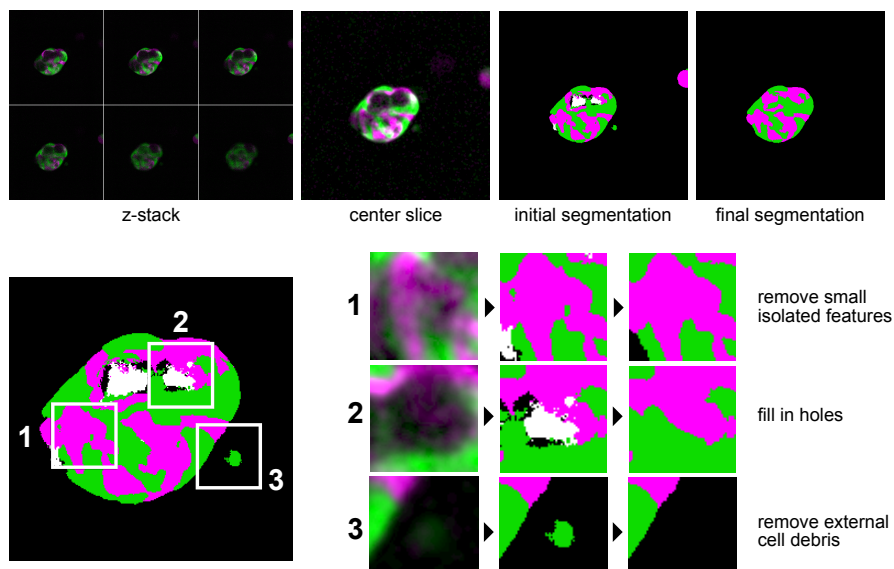


Figure 2.3: Image processing involves manually choosing the center slice, background subtraction, pixel assignment by a random forest classifier, and morphological operations designed to remove small isolated features, fill in holes, and remove external cell debris.

For the corresponding processing of organoid timelapses, the same automated pixel classification and processing procedure was applied to each individual  $z$ -slice at each time-point. The center slice was assumed to be the slice with the largest area; in the case of a contiguous near-tie (within 15% of the largest area), the center slice among those was selected.

The morphological parameters of each segmented slice were calculated in MATLAB. A complete list of variables, their purpose, and their derivations are provided in **Table 2.1**. Where applicable, absolute lengths and areas were converted from pixels to microns or microns squared. Quality control cutoffs for cell fraction, area and circularity were defined to exclude poorly segmented organoids and organoids merged with cell debris. Organoids were subsetted to have no more than 60% of one cell type, cross-sectional area between  $3000\ \mu\text{m}^2$  to  $10\ 000\ \mu\text{m}^2$  inclusive, and circularity greater than 0.7. For live imaging, area cutoffs were defined to exclude datapoints with area more than 2-fold different from the mean track area to remove aberrant frames from  $z$  drifting and collisions with nearby cells.

Table 2.1: Morphological parameters recorded for organoid images. Pixel connectivity is such that each pixel is adjacent to its 8 neighbors (Chebyshev distance 1). [Type] refers to LEP in heterotypic organoids and the green cells in homotypic organoids.

Name	Measures	Description
[Type] fraction	composition	Proportion of [Type] pixels.
Area	shape	Total area in $\mu\text{m}^2$ .
Circularity		Roundness, defined as $\frac{4\pi \cdot \text{Area}}{\text{Perimeter}^2}$ .
Diameter		Effective diameter of the organoid, $2\sqrt{\text{Area}/\pi}$ .
Intermixing	cohesion	Number of pixels adjacent to a pixel of the opposite cell type, divided by $2 \cdot \text{Area}$ . Higher values mean less cohesion/segregation.
Blob Size		The mean feature size: the reciprocal of the mean fast Fourier transform frequency of the mask with two cell types -1, 1, and 0 elsewhere.
Solidity		Proportion of [Type] pixels in the smallest convex enclosing polygon of [Type]. Higher values mean more cohesion of [Type].
Intercentroid	linear sorting	Distance between the centroids of the cell types, divided by Diameter. Larger values mean more separation.
Edge occupancy Outer occupancy Inner occupancy	radial sorting	Proportion of [Type] in the region. Edge: directly adjacent to ECM. Outer/Inner: the lower/upper third of pixels ranked by distance to ECM.
Correctness		Proportion of pixels that match the segmentation generated by moving all MEP pixels as close to the ECM as possible.

## 2.9: Three-dimensional nuclear diffusion assay

To quantify the undirected random motion within cell aggregates, we used timelapse imaging of organoids labeled with cytoplasmic GFP into which were introduced a small number (comprising  $1/10$  of the total cells spun into microwells) of doubly transfected GFP and H2B-mScarlet-expressing cells of the same type. Organoids containing 1 – 10 red nuclei with no neighboring organoids were chosen for live imaging of 7 5- $\mu\text{m}$  slices at 10 min resolution.

For nuclear motility, 7 5- $\mu\text{m}$  slices were analyzed using the TrackMate pipeline in FIJI to assign spot identities to tracks with manual curation of track assignments. Timepoints were assigned based on imaging frequency (5, 10, or 20 min intervals). Image scaling information and centroid Ilastik segmentations were used to define the  $x$  and  $y$  coordinates relative to the centroid in  $\mu\text{m}$ . Quality control on centroid assignments was performed using deviation from mean area and mean centroid. Frameshift calculations of  $x, y$ , and  $t$  were calculated and used to calculate instantaneous speeds. Extremely high instantaneous speeds were considered mislabeled spots and removed prior to analysis. Tracks of at least 5 hr duration and beginning 10 hr after aggregate formation were considered. Lag times were limited to 2 hr. The limit of detection of cell position fluctuations is close to the fluctuations of the tissue centroid. Comparison of mean squared displacement (MSD) to lag times across experimental conditions indicate a diffusion coefficient  $\alpha$  near 1.2, indicating slightly super-diffusive behavior, so diffusion coefficients were estimated using the relationship  $\text{MSD} = \langle (\vec{x}_\tau - \vec{x}_0)^2 \rangle \propto D_{\text{EFF}} \cdot \tau^{1.2}$ . Different dates were analyzed as separate batches.

## 2.10: Predicted tissue energies

To link the measured mechanical energies of isolated cellular interfaces to the global energy of a tissue composed of many cellular interfaces, I built a three-dimensional simulated tissue, representing a cluster of cells as a uniform polyhedral foam of near-minimal surface area.

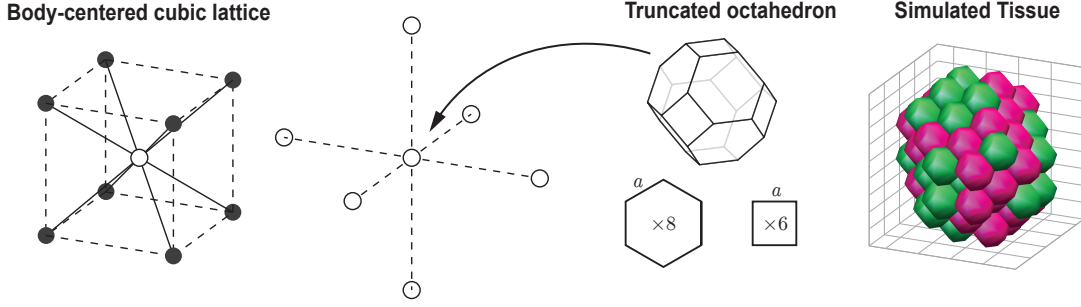


Figure 2.4: The Voronoi tessellation of the base-centered cubic lattice is a close packing of cells. Each cell is a truncated octahedron with six square faces along the main axes and eight hexagonal faces along the diagonals. All polygonal faces have side length  $a$ . Lattice points were chosen by excluding points beyond a specified radial distance from the centroid, and the outermost layer of points, not pictured, are assigned to be substrate.

The tissue enthalpy can then be quantified in terms of the cell contact angles  $\theta$ , cell cortical tension  $\gamma_c$ , and the areas of the cellular interfaces  $A$ , as described in Appendix C. These calculations allow us to estimate the tissue energies in terms of units of energy (pJ), based upon the empirical measurement of  $\gamma_c$  (nN  $\mu\text{m}$  or  $10^{-3}$  pJ  $\mu\text{m}^{-2}$ ) and the estimated cellular areas ( $\mu\text{m}^2$ ). Some interfacial tensions not measured were substituted with the most similar available interfacial tension. For example, the micropipette aspiration assay is extremely low-throughput and was performed for a minimal set of control and transduced LEP and MEP with and without drug, the results of which suggested that cell cortical tensions are very similar across cell types. Additionally, cell-cell contact angles for *PIK3CA*<sup>H1047R</sup> with MK2206 were substituted with those for the same transgene without drug.

$$\begin{aligned} \Delta E &= \sum_{i,j} \Delta E_{c_i c_j} + \sum_k \Delta E_{c_k x} \\ &= \sum_{i,j} 2\gamma_c A_{c_i c_j} (\cos \theta_{c_i c_j} - 1) + \sum_k \gamma_c A_{c_k x} (2 \cos \theta_{c_k x} - 1) \\ A_{\text{hexagon}} &= \frac{3\sqrt{3}}{2} \left( \frac{\pi}{6\sqrt{2}} \right)^{\frac{2}{3}} R^2 \approx 134 \mu\text{m}^2 \\ A_{\text{square}} &= \left( \frac{\pi}{6\sqrt{2}} \right)^{\frac{2}{3}} R^2 \approx 52 \mu\text{m}^2 \end{aligned}$$



As the total estimated tissue energy is a complicated function of the five interfacial energies (between the two cell types and the substrate), each of which is associated with its own degree of measurement uncertainty, the point estimate and confidence interval of the resulting energy landscape was quantified with a bootstrap approach. First, the standard error for each interfacial energy ( $\Delta E_{cc}$  or  $\Delta E_{cx}$ ) was estimated using the Gaussian error propagation approach,

$$\sigma_{\Delta E}^2 = \left| \frac{\partial \Delta E}{\partial \gamma_c} \right|^2 \sigma_{\gamma_c}^2 + \left| \frac{\partial \Delta E}{\partial \theta} \right|^2 \sigma_{\theta}^2 = (2 \cos \theta - 1)^2 \sigma_{\gamma_c}^2 + (-2\gamma_c \sin \theta)^2 \sigma_{\theta}^2$$

Next, randomized parameter sets of interfacial energies were generated by drawing each of the five interfacial energies from the corresponding normal distributions with mean and standard deviation determined by the error propagation. For each of 1000 sets, the energies were used to calculate the energy landscapes of 1,000,000 randomly generated simulated tissues randomly subsampled for computational efficiency. Phase space was binned into  $25 \times 25$  two-dimensional bins and the configurations were randomly resampled without replacement so that no bin held more than 50 tissues. The estimated energies generated from ten such parameter sets based on the wild-type and PIK3CA measurements are shown with respect to EO in **Fig. 2.5**. PIK3CA parameter sets have higher variance, resulting in a wider distribution of estimated slopes. An unweighted least squares linear regression was performed to calculate  $dE_i$  as a function of Edge Occupancy and Intermixing Score.

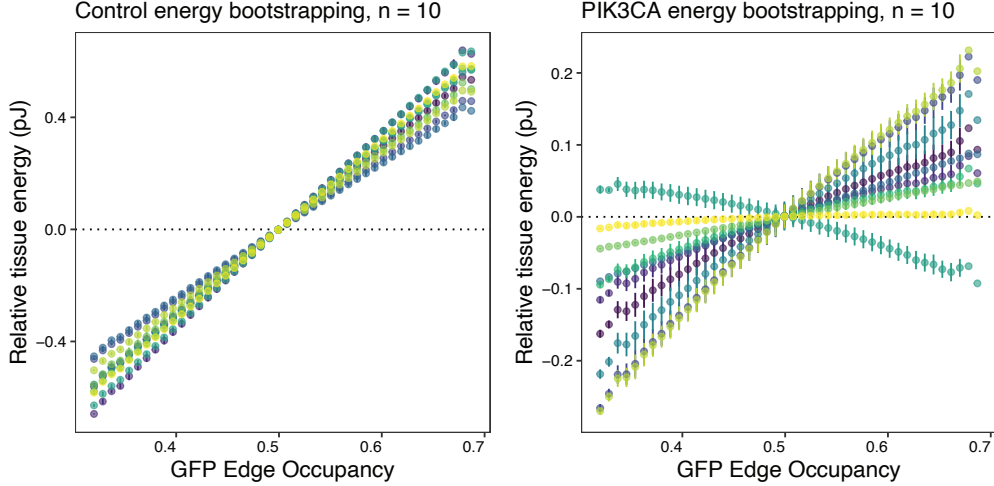


Figure 2.5: The change in total tissue energy with respect to macrostate  $i$  is estimated by bootstrapped linear regression on randomly drawn parameter sets based on the measured values and uncertainties of  $\gamma$ . For clarity, only 10 parameter sets on the Edge Occupancy axis are displayed. Points are mean  $\pm$  standard deviation.

## 2.11: Inferred tissue energies

We may solve the inverse Ising problem by comparing the actual data observed from sorting and non-sorting organoids,  $P_i$  and  $P_{\text{MEP},i}$ . I approximate  $W_i$  from  $P_{\text{MEP},i}$  with a normal distribution fitted to the observed EO frequencies,  $\frac{1}{\sigma\sqrt{2\pi}}e^{-(i-\mu)^2/2\sigma^2}$ . The MEP-only organoids are subsetting to match the range of compositions and sizes. The inferred  $W_i$  truncated normal is used to calculate the dimensionless quantity  $-\ln\left(\frac{P_i}{W_i}\right)$ , which is equal to  $\beta E_i$  within a constant<sup>1</sup>. Invalid data points due to 0 observed events are omitted.

The propagated uncertainty associated with  $\beta E_i = -\ln\frac{P_i}{W_i}$  is expected to contain the term  $\frac{1}{P_i W_i}$ , which is extremely large for many regions of phase space where configurational entropy is low or where very few organoids have been observed, not to mention the regions where  $\beta E_i$  is undefined due to 0 observed events. Where the inferred  $\beta E_i$  is displayed for visualization purposes it is accompanied by a weighted least squares linear regression with

---

<sup>1</sup> $\beta E_i = -\ln\left(\frac{P_i}{W_i}\right) + c$ . The constant  $c$  captures terms which are independent of  $i$ , such as the partition function  $Z$ , the tissue reference energy  $E_0$ , and the scaling factor converting the very large numbers of microstates  $W_i$  into values of  $P_{\text{MEP},i} < 1$ .

weights proportional to  $P_i W_i$  to reduce the influence of high variance regions. However, this results in significant data loss. For inferring the slope term  $\beta dE_i$  and estimating its CI, I take the nonlinear least squares fit of a resampling of the observed events, assuming the form  $P_i = \frac{W_i}{Z} e^{-\beta dE_i \cdot i}$  and solving for  $\beta dE_i$  over 1000 bootstraps resampling 200 points.

### 3. Theory

We are interested in the generation and maintenance of tissue structure, where tissue structure is defined as the spatial arrangement of the component cells with respect to each other and their surroundings. In the mammary gland, these structural features ensure that the cells that produce milk and the cells that contract to expel milk are arranged in a core-shell structure; they also create a cellular barrier to cancer progression (Sirka et al., 2018). However, not every part of the tissue is found in the ideal configuration, both across different parts of the same organ and across individuals. Instead, we see a distribution of structures that is biased towards the sorted, ideal structure, but nonetheless shows a great deal of variability. In vitro, self-organizing forces create organoids with a LEP core and MEP shell (Cerchiari et al., 2015), but structures are variable even though so many of the inputs and constraints of the system can be controlled.

This control over system constraints makes HMEC organoids an attractive model system for characterizing and perturbing the driving forces of tissue self-organization. The ability to analyze large numbers of organoids of defined compositions makes it possible to isolate and quantitatively explain the observed structural distributions as a function of statistical mechanical parameters. Furthermore, by using the concepts of the DITH, we can directly estimate the relative mechanical energies of different structures.

#### 3.1: Statistical mechanics for multicellular tissues

We can begin to define a statistical mechanical framework for tissue self-organization by establishing tissue-level analogs to microstates, macrostates, energy, and entropy.

- *Microstate*: The tissue arrangement, or the positions of all cells in an organoid.
- *Macrostate*: The set of all arrangements of cells consistent with a macroscopic observable. For tissue self-organization, this will be a structural coordinate value.

- *Energy*: The global tissue energy is the sum of the interfacial energies, as in DITH.
- *Entropy*: Boltzmann entropy  $S_i = k \ln W_i$ , where  $W_i$  is the number of microstates consistent with macrostate  $i$ .
- *Activity*: The baseline level of energy fluctuations in the system, which manifests in tissues as random undirected cell motility.

Towards the goal of establishing a tissue Boltzmann distribution,  $P_i = \frac{W_i}{Z} e^{-\beta E_i}$ , this chapter will define the quantities used to distinguish between the structural macrostates, describe the derivation of energy and entropy associated with each macrostate, describe the concept of tissue activity, and discuss the applicability of steady-state and ergodic approximations to tissue self-organization.

### 3.2: Defining structural coordinates

The choice of structural coordinates determines the interpretation of all other terms. We seek a way to identify macrostates that can be quantitatively measured from microscopy data, and can be associated with a global tissue energy and a set of equivalent microstates.

We begin by considering the expected and observed variations in organoid structure. HMEC organoids are roughly spherical, and as they are composed of non-invasive cells, do not extend or invade into the ECM. They are generally solid rather than lumenized, allowing cells to take on positions throughout the sphere. Given these constraints, two different cell types can be arranged in some qualitatively distinct patterns. For HMEC organoids, the most relevant patterns are 1) “GFP core”, with an mCherry shell, 2) “split”, with GFP and mCherry cores segregated, 3) “mixed”, with GFP and mCherry intermingled, and 4) “GFP shell”, with mCherry core (**Fig. 3.1**). Other studies of cell sorting observe similar structures, using terms like *checkerboard*, *segregated* or *phase-separated*, and *engulfment*, and larger groups of cells may display multiple cores or multiple poles (Brodland, 2004; Toda et al., 2018).

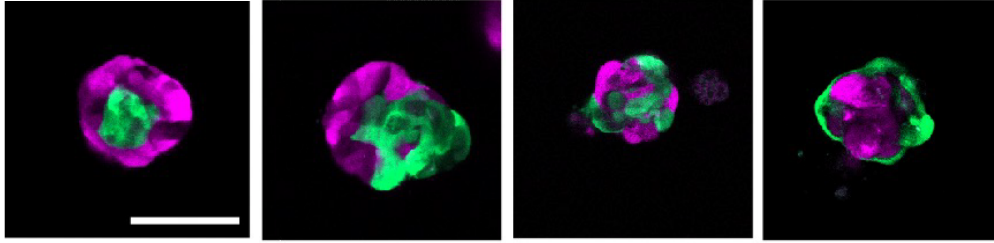


Figure 3.1: GFP and mCherry cells in a tissue can be organized in core, split, mixed, and shell arrangements. Scale bar: 100  $\mu\text{m}$ .

I scored a large dataset of organoid images into the four categories alongside image analysis to extract the morphological information described in **Table 2.1**. Borderline structures showing partial engulfment or multiple cores were assigned to the appropriate core-shell category for clarity of visualization. An initial set of 10,000 organoids was subsampled evenly across the four categorical outcomes, yielding almost 5,000 images. After initial analysis, highly correlated redundant variables were removed for the PCA shown in **Fig. 3.2**. The qualitative structural categories, despite their significant degree of overlap, can be readily separated using two parameters, the Edge Occupancy (EO) and the Intermixing Score (IMS), which represent PC1 and PC2 respectively. EO is high for shell, low for core, and intermediate for split and mixed, while Intermixing Score is high for mixed, low for split, and intermediate for the core/shell arrangements.

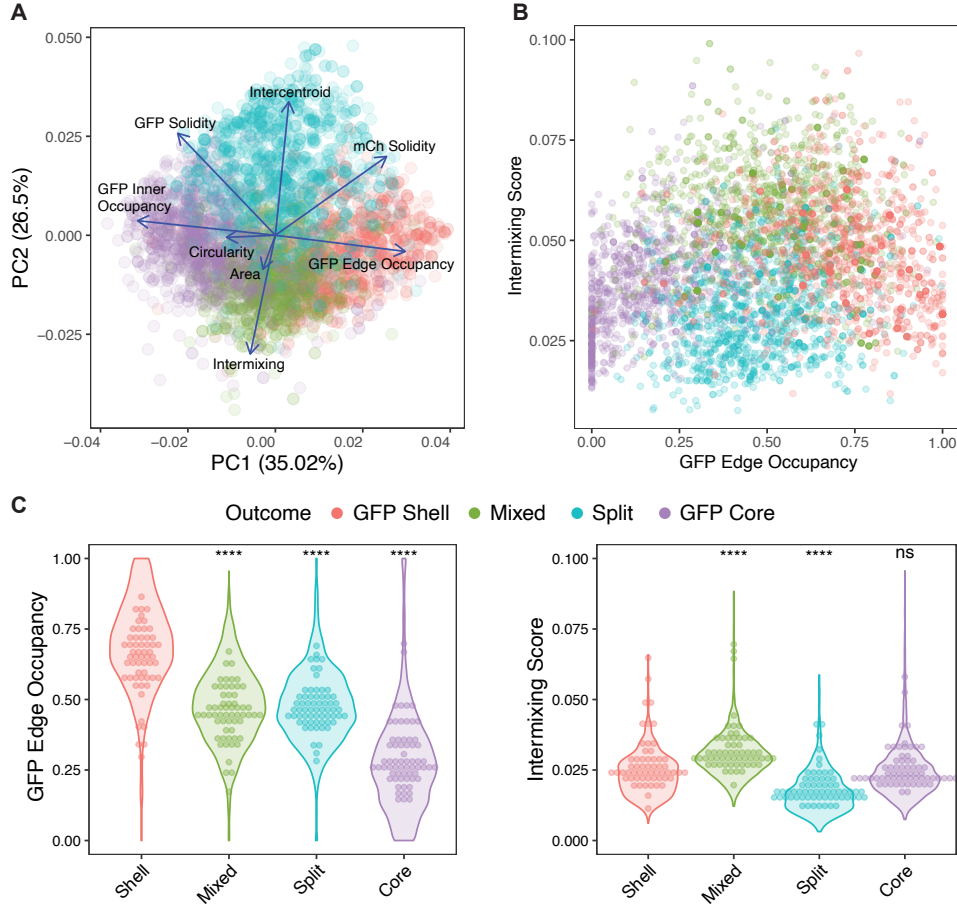


Figure 3.2: Outcomes can be described with relatively few morphological parameters. Members of four categorical structural outcomes were equally subsampled from a large organoid dataset. Principal components were derived from the most informative morphological variables (A). The raw values of two parameters, the Edge Occupancy and Intermixing Score, are shown in (B) and (C) for the full organoid dataset. In (C), points represent technical replicates (means of experiment dates) and statistics were performed using the Wilcoxon signed-rank test against the GFP Shell reference.

Edge Occupancy and Intermixing Score are natural choices for defining cell sorting outcomes, but how well do they perform at defining tissue energy and entropy?

### 3.3: Quantifying tissue energy

Having identified some potential structural coordinates, we next determine how well these quantities can predict tissue energy. How does the interfacial energy of a tissue change between different macrostates? Can each macrostate be assigned an energy level?

Cerchiari et al. (2015) reported a DITH framework for mammary self-organization that identified minimum energy configurations. In the DITH, the global tissue energy Hamiltonian is the sum of the products of all interfacial areas with their corresponding tensions. We expand on these ideas to estimate the energy landscape by interpolation. The number of possible arrangements of all the cells and their membranes is unquantifiably large, but it is not necessary to identify every possible cell arrangement and calculate its tissue energy. Rather, we can generate a much smaller, discrete subset of arrangements that samples from a range of possible structures, generated by permutation of polyhedral lattice points in a simulated tissue, and use this finite set of points to interpolate the tissue energy landscape as a function of structure and cellular parameters (section 2.10).

These parameters are the cell-cell and cell-ECM energies, which are reflected in the free energy changes between cells in suspension and adherent cells. These energy changes can be quantified using various approaches; for example, Shishvan et al. (2018) drew upon stress fiber formation, traction force microscopy, and cell morphologies to estimate energy of adhesion for spreading smooth muscle cells with heterogeneous shapes on substrates of different stiffnesses. Adherent HMEC are approximated in shape by simple spherical caps, allowing us to measure interfacial tensions by analyzing cell contact angle geometries as a statics problem, that is, the Young equation for droplet spreading (Cerchiari et al., 2015). Relative tensions can be scaled to the cell's actomyosin cortex tension in  $\text{nN } \mu\text{m}$  or, equivalently,  $10^{-3} \text{ pJ } \mu\text{m}^{-2}$ , measured by micropipette aspiration. The derivation of these equations is detailed in Appendix C. The values of these quantities for the wild-type LEP and MEP are shown in **Fig. 3.3**.



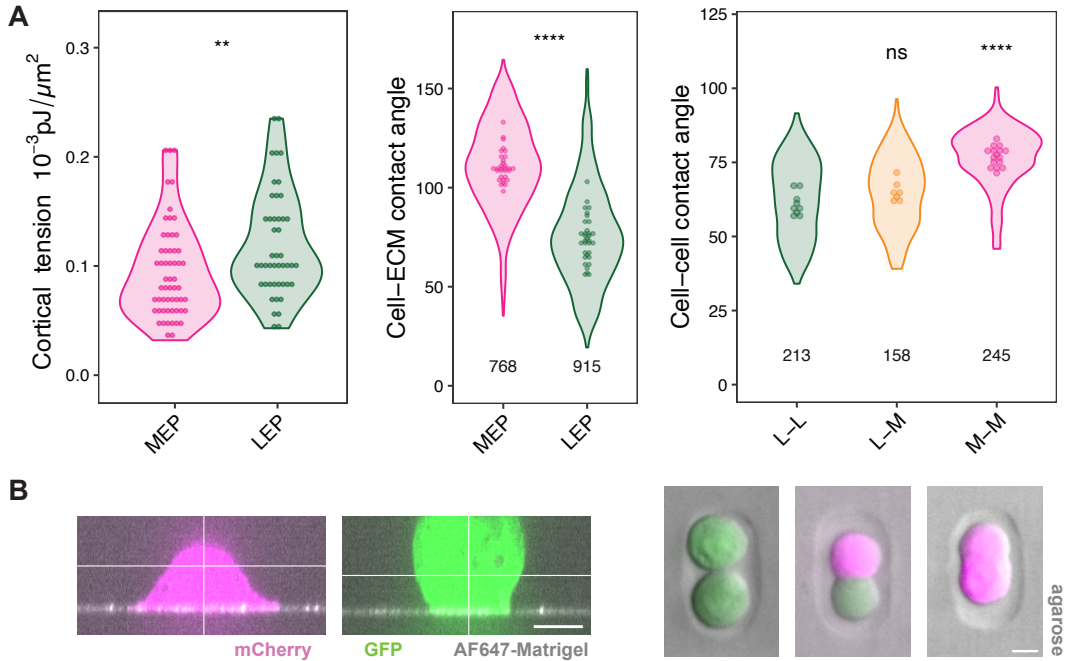


Figure 3.3: Cortical tensions and contact angles of LEP and MEP are used to estimate tissue interfacial energy (A). For cortical tensions, points indicate individual cells; for contact angles, points indicate technical replicates (experiments performed on different dates), while biological replicates (individual cells or cell doublets) are listed at the bottom. P-values were calculated with the Wilcoxon signed-rank test. Representative images of cell-ECM and cell-cell contact angles (B). Scale bar: 10  $\mu\text{m}$ .

Combining the polyhedral lattice point model with empirically measured quantities for cell-cell and cell-matrix interfacial tensions shows that for wild-type tissues containing LEP and MEP, global tissue energy is best determined by the proportion of the tissue boundary that is occupied by LEP (**Fig. 3.4**). This structural coordinate is not only useful in describing the key disruption during cancer progression but is also nearly linearly related to total tissue energy. Each EO macrostate is associated with multiple possible energies, representing differences in cell-cell connectivity, but these differences are small relative to differences in energy between EO macrostates.

Macrostates with energy distributions can be considered to have a single effective energy, depending on the nature of the energy distribution across microstates. For a macrostate  $i$  with normally distributed microstate energies  $\epsilon_i$ , the effective energy is the average energy

minus a term related to variance,  $E_i = \bar{\epsilon}_i - \frac{\beta\sigma^2}{2}$  (Phillips et al., 2019). We observed that  $\bar{\epsilon}_i$  is close to a linear function of  $i$  in simulated tissues (**Fig. 3.4A**). It will be the case that  $dE_i = d\bar{\epsilon}_i$  if  $\sigma^2$  and  $\beta$  are not also functions of  $i$ . In our simulated tissues, the energy distributions for each LEO macrostate can be approximated as normal distributions of equal variance (**Fig. 3.4B**). The true variance of the energy distributions for real tissues will likely be broader, as deformable cells can take on larger and smaller interfacial areas, but the geometrical behavior of the cells should be similar and it is unlikely to acquire a dependence on  $i$ . Activity,  $1/\beta$ , is a property of the system as a whole rather than of a macrostate  $i$ . Then, modeling tissue energy as a linear function of  $i$  with derivative  $dE_i = d\bar{\epsilon}_i$ ,  $E_i \approx i \cdot dE_i + E_0$ . As we are interested in relative energy differences between states rather than absolute energy values, the added  $E_0$  reference energy level, which is constant with respect to  $i$ , is not important<sup>1</sup> and I use  $E_i \approx i \cdot dE_i$ . The calculation of the estimated slope  $dE_i$  and its associated 95% confidence interval, accounting for uncertainty in our measurements of interfacial energies, is described in section 2.10.

---

<sup>1</sup>Any constant  $E_0$  added to  $E_i$  in the exponent of the Boltzmann weight is equivalent to a uniform scalar multiple  $e^{E_0}$  on the probability density, which will be normalized out in the partition function  $Z$  by the constraint that  $\sum_i P_i = 1$ .

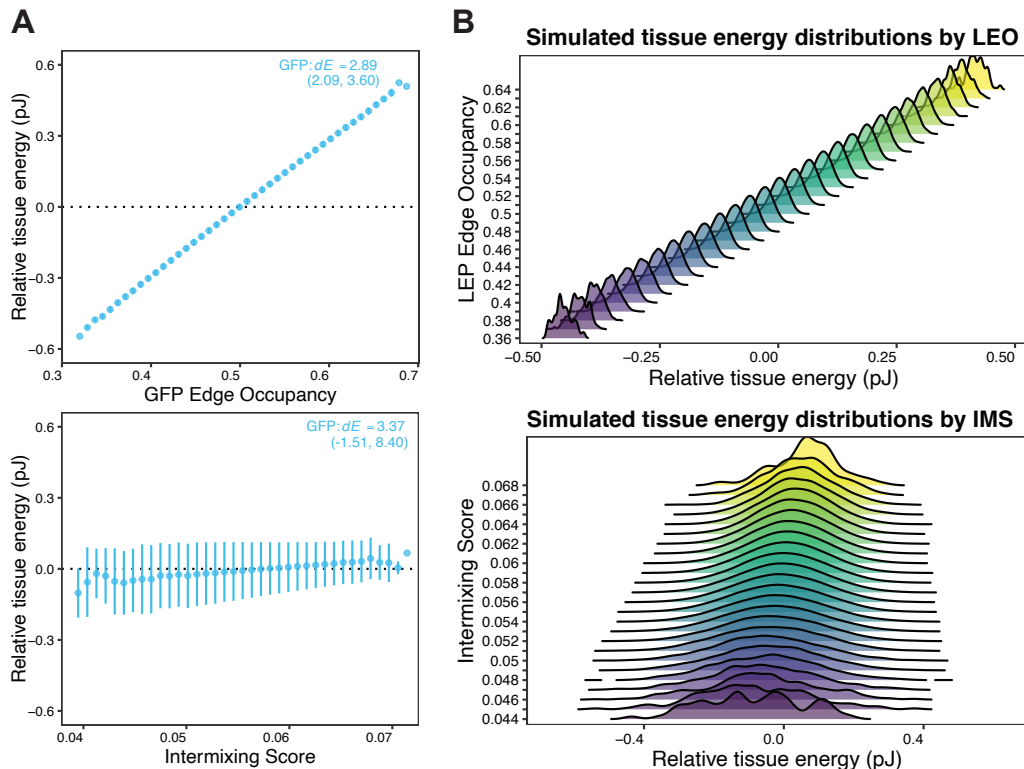


Figure 3.4: One million simulated tissues were randomly populated with 50% LEP and 50% MEP cells using the mean interfacial tensions measured for  $\gamma_{cx}$ ,  $\gamma_{cc}$ , and  $\gamma_c$ . These tissues show a strong relationship between Edge Occupancy and relative tissue energy, while the Intermixing Score is less cleanly associated with energy. Points denote binned means with bars  $\pm 1$  standard deviation (A). At each EO or IMS value, there are multiple microstates, resulting in energy distributions for each macrostate which are approximately normally distributed (B). The predictive power of IMS is considerably lower than EO, indicating that it is a less informative single state variable.

### 3.4: Enumerating structural microstates

The Boltzmann entropy,  $S = k \ln W$ , is a function of the number of ways ( $W$ ) the microscopic components of a system can be arranged to be compatible with the macroscopic observed parameters used to define each macrostate<sup>2</sup>. Of course, the cellular components of a tissue can take any real value, making the number of possible arrangements uncountably infinite. Instead, the standard statistical mechanical approach is to use a “coarse-graining” procedure

<sup>2</sup>Note that this identity leads naturally to a free-energy formulation of the Boltzmann distribution,  $P_i \propto W_i e^{-\beta E_i} = e^{-\beta E_i + \ln W_i} = e^{-\frac{1}{kT} E_i + \frac{S_i}{k}} = e^{-\frac{1}{kT} (E_i - TS_i)} = e^{-\beta G_i}$ , which is simply the Boltzmann weight, where the energy of a microstate is replaced with the free energy of a macrostate.

to partition the fine-grained true microstates into a finite number of disjoint microstates, or arrangements (Frigg, 2008; Uffink, 2007). We then define macrostates by grouping together arrangements with the same macroscopic features.

Consider a lattice that has been partitioned into  $n$  core sites and  $n$  shell sites. How can  $k_{\text{tot}}$  green cells be arranged across these sites so that  $k$  green cells are in the shell? For  $k = 0$ , there must be 0 in the shell and  $k_{\text{tot}}$  in the core; in the case shown in **Fig. 3.5**, there is only one possible configuration. For  $k = 1$ , we put green cells in 1 shell site and  $k_{\text{tot}} - 1$  core sites, choosing from  $n$  both times. For  $k = 2$ , we choose 2 and  $k_{\text{tot}} - 2$ ; for three, 3 and  $k_{\text{tot}} - 3$ ; and so on, increasing to a maximum at  $k_{\text{tot}}/2$  and decreasing symmetrically until reaching the structure with all  $k_{\text{tot}}$  green cells in the shell and none in the core. In this example, for  $k$  green shell cells out of  $k_{\text{tot}}$ , the number of possible configurations is  $W_k = \binom{n_{\text{shell}}}{k} \cdot \binom{n_{\text{core}}}{k_{\text{tot}} - k}$ , where  $\binom{n}{k}$  is the binomial coefficient,  $\frac{n!}{k!(n-k)!}$ . A three-dimensional variant of this combinatoric example using the surface areas of the truncated octahedral lattice can be calculated with the appropriate generating function and is described in Appendix D.

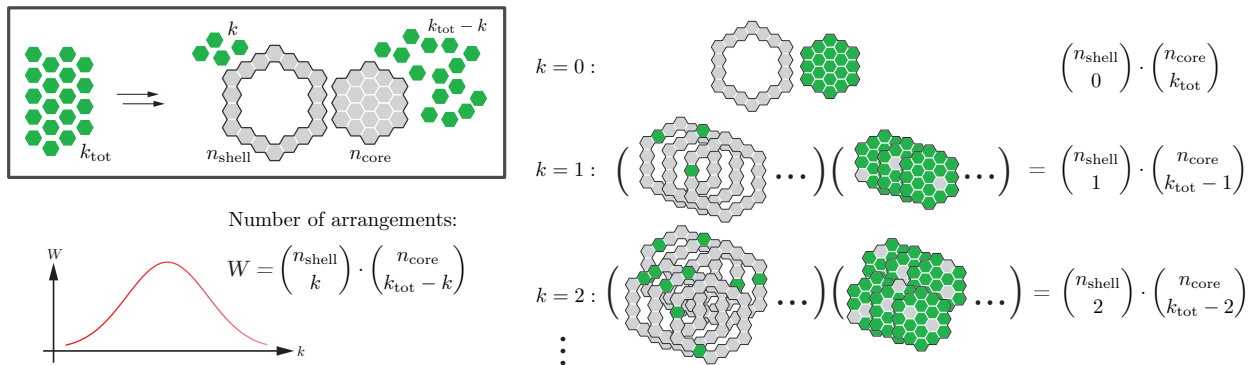


Figure 3.5: Combinatorics of cell arrangements. This simple discrete model demonstrates that the number of possible arrangements for a given number of green cells in the shell,  $k$ , is a function of the geometry of the system ( $n_{\text{shell}}, n_{\text{core}}, k_{\text{tot}}$ ) and can be precisely calculated with the binomial.

Here, we defined each macrostate by the number of green cells in the shell. Each macrostate is composed of a multiplicity of microstates, or number of possible cell arrange-

ments that satisfy the definition of the macrostate. The illustrative example treats each cell as occupying exactly one site, and the tissue size as equivalent to the number of cells or sites.

However, cells are not actually rigid, discrete objects, nor do they arrange themselves into idealized polyhedral foams in real tissues. The lattice model, whether two- or three-dimensional, lacks degrees of freedom associated with the possible configurations of the cells and their membranes. Hypothetically, cell deformation could be represented as random variation in the sizes of connected interfaces, convolving the discrete case with a Gaussian and widening the multiplicity distribution, or by breaking each cell into multiple connected lattice sites as in the Cellular Potts model. A more straightforward solution, however, is to directly measure the empirical distribution of real organoids, which implicitly includes hidden degrees of freedom, in order to infer tissue state multiplicity. In a tissue that has no self-organizing forces at all, the distribution of tissue structures is purely determined by the multiplicity of different states. And as we intuit from the example of **Fig. 3.5**, there should be statistically fewer well-ordered than disorganized structures, with a distribution similar to the binomial, which is Gaussian for large  $n$  by the Central Limit Theorem.

Organoids composed of only MEP labeled with different fluorophores have no differences between cell types that could cause self-organization. The EO distribution of the resulting all-MEP tissues closely resembles the distribution of tissue structures predicted by the combinatoric approach, a bell curve centered at 50/50 (**Fig. 3.6**). Supposing that the observed all-MEP tissues are drawn from the underlying entropy-driven distribution, we will approximate  $W_i$  as the truncated normal distribution with support  $[0, 1]$  with mean and standard deviation fitted to the MEP-only data.

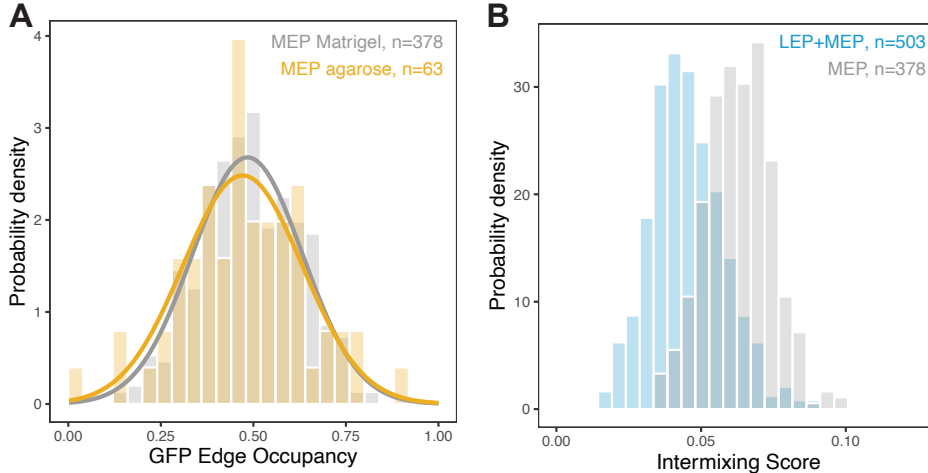


Figure 3.6: Structural distributions of MEP-only organoids in both Matrigel and agarose environments show that the distribution is approximately Gaussian with respect to Edge Occupancy. Curves are truncated normal fits with median 0.487,  $\sigma = 0.156$  for Matrigel and median 0.509,  $\sigma = 0.164$  for agarose (A). Unlike sorted LEP+MEP organoids (blue), MEP-only organoids (gray) have high Intermixing Scores, indicating that the two cell populations are interspersed rather than segregated (B).

### 3.5: Steady state and activity

With statistical mechanics, we wish to predict the steady-state distribution of an ensemble of organoids on a phase space of structural states, where the probability of observing a system in a given energy state is given by **Eq.** (1.1). In the case of tissue self-organization, it is reasonable to treat self-organized tissues as a quasi-steady system if the constraints of the system (such as cell proportion, state, and physical properties) are stable over the timeframe of interest, even if they are expected to change over the long term. Maintaining these constraints is analogous to treating the tissue as a closed system conserving “particle number” (cells) while exchanging metabolic energy with its environment, over a timescale of interest during which the energy and entropy landscapes are preserved.

We estimated the time it takes for organoids to reach the steady-state mean structure after initial seeding by quantifying organoid structures from initial configurations that are far from the predicted steady-state distribution. This experiment can be designed in two ways: first, by measuring the cell mixing (Intermixing Score) of MEP-only pre-segregated tissues,

and second, by measuring the cell sorting (EO) of wild-type MEP and LEP aggregates from initially mixed aggregates.

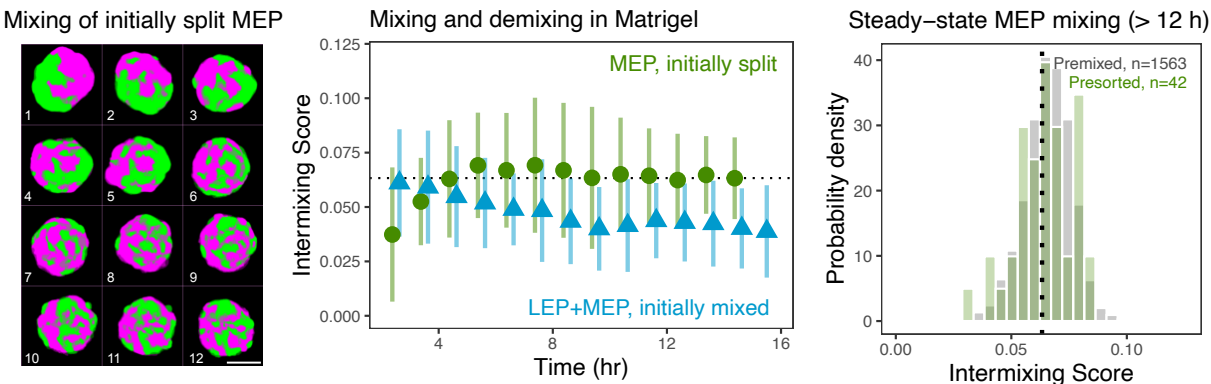


Figure 3.7: Timelapse montage of a pre-sorted MEP organoid from 1-12 hr after embedding in Matrigel, segmentation mean over 3 center cross-sections. Scale bar: 50  $\mu\text{m}$ . Live imaging shows initially split MEP+MEP organoids ( $n = 15$ ) quickly mix over several hours, while initially mixed LEP+MEP organoids ( $n = 35$ ) demix over a similar time period. After reaching the steady state after half a day of culture, initially split MEP-only organoids have high Intermixing Scores indistinguishable from initially mixed MEP-only organoids. Dotted lines indicate mean IMS of organoids at day 2.

Pre-segregated MEP-only tissues were generated by allowing GFP MEP and mCherry MEP to aggregate in separate 180- $\mu\text{m}$  wells, transferring them to shared 240- $\mu\text{m}$  wells, and incubating them together for approximately 3 hr to allow initial adhesions to form before transferring to Matrigel. These pre-sorted tissues mixed extremely quickly, reaching an Intermixing Score like that of standard MEP-only tissues after only a few hours, while initially mixed LEP+MEP organoids segregated more slowly (**Fig. 3.7**). Sorting of LEP+MEP organoids along the Edge Occupancy coordinate tends to take longer than feasible for live imaging, but static timepoint measurements of organoids over time show that LEP+MEP aggregates in Matrigel approach the steady-state distribution over two days of culture, and the distribution is maintained after one week of culture. In agarose, the sorted distribution is also stable from day 1 through 3, though for cell viability reasons they cannot be cultured in suspension much longer without undergoing anoikis (**Fig. 3.8**).

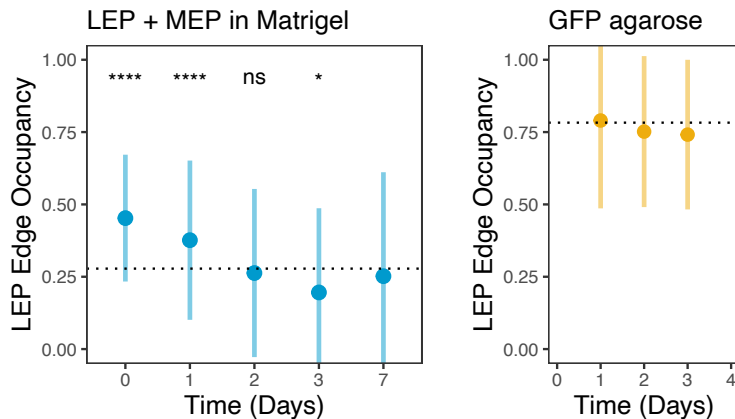


Figure 3.8: Static timepoint measurements show that LEP+MEP tissues reach the steady-state configuration over one to two days of culture in Matrigel. Dotted line indicates mean EO of LEP+MEP organoids on day 2. For days 0, 1, 2, 3, and 7,  $n = 167, 1174, 1475, 72$  and  $98$ . Comparison of the means was performed against timepoint 7 with the Wilcox test and Holm post-hoc correction. Agarose LEP+MEP organoids appear to maintain a stable structural distribution after 1 day ( $n = 575, 1138$  and  $409$ ). Points are mean  $\pm$  standard deviation.

To what extent can each individual system evenly explore the configuration space? In an ergodic system, the trajectory of a single organoid, in the limit of infinite time, would lie densely (pass arbitrarily close to every point) on the entire accessible configuration space, spending an amount of time in each region proportional to the probability of observing that configuration in an ensemble of many organoids. The ergodic hypothesis is useful as it implies that a organoid initially anywhere in the phase space will eventually be able to reach the steady state, rather than becoming trapped in a local energy minimum. It also implies that the stationary probability distribution of a single isolated system must be the microcanonical ensemble, and has historically been used to justify various arguments in both the Boltzmann and Gibbs frameworks of statistical mechanics (Uffink, 2007).

In practice, it is extremely difficult to prove that even the simplest realistic physical systems meet the requirements of ergodicity (Moore, 2015; Uffink, 2017), but we can make a number of less-stringent observations consistent with ergodicity in our system. For example, we have shown that a MEP-only tissue’s final steady-state distribution is not dependent on



its initial conditions, but rather that trajectories initiated in either the initially mixed or the initially split regions of configuration space both converge to a mixed distribution (**Fig. 3.7**). Furthermore, we performed analyses of LEP+MEP timelapses taken after reaching steady state (40 hr to 60 hr) (**Fig. 3.9**). The LEP EO of this small ensemble does not change over 10 hr, indicating that the steady-state structure is stable over time. The steady state is dynamic, with random fluctuations as cells exchange positions. Individual organoids' EO can vary despite spending the majority of the time close to the mean. These fluctuations tend to return to the mean: an organoid far below the mean EO will typically increase, whereas an organoid with unusually high EO will decrease (**Fig. 3.9A**).

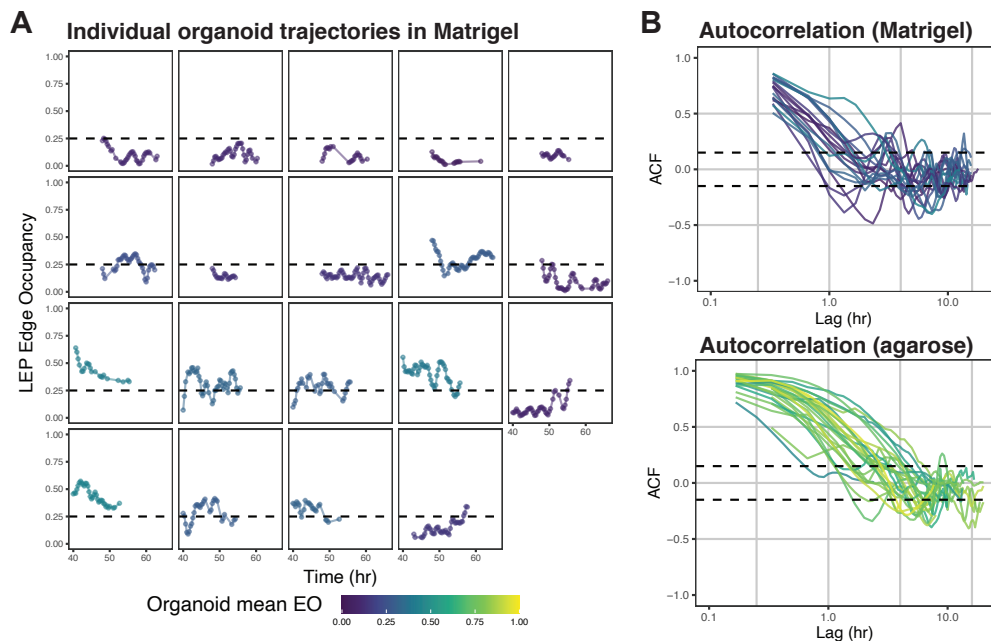


Figure 3.9: After two days of culture in Matrigel, LEP+MEP organoids have reached their steady state structural distribution, but this steady state is dynamic and varies across individual organoids. LEP Edge Occupancy of individual organoids fluctuates near the mean (dashed line) (A). Organoid trajectories rapidly lose autocorrelation over a couple hours in Matrigel and in agarose, though the decay is slower in agarose (B). Traces are colored by their mean Edge Occupancy value.

We also observe that HMEC organoids are readily able to exchange cell positions and explore the configurational space from their initial state and at their steady state. Strikingly, cells appear to swirl and diffuse through the bulk of the tissue, and even after an initially

mixed organoid reaches a sorted state, cells frequently move to different locations or shift in and out of sorted positions (**Fig. 3.9A**). Dynamic changes in organoid structure, cellular rearrangements, and full-organoid collective rotation, termed coherent angular motion, have also been observed in mammary organoids by other groups (Sirka et al., 2018; Neumann et al., 2018; Tanner et al., 2012). Despite the important role of DITH and cell mechanics in determining the most energetically favorable structure, cells in HMEC organoids make cell position changes that are not determined solely by favorable energy changes. Indeed, even organoids composed of MEP alone will display active cell migration despite the absence of any forces driving sorting (**Fig. 3.7**). A dynamic system like the actively moving cells in an HMEC organoid can explore the structural energy landscape even in its steady state as a result of active fluctuations, which we term *activity*. Activity represents the random cellular energy fluctuations that are large enough to drive cell position changes in the absence of or even against mechanical energy gradients. This parameter is analogous to thermal energy,  $k_B T$ , in a microscopic system and functions similarly in tuning the influence of energy relative to entropy in determining the steady-state distribution. The term  $\beta$  in **Eq. (1.1)** corresponds to the inverse of activity, or  $\frac{1}{k_b T}$  in a thermodynamic system.

We can easily show that the activity of an organoid structure cannot be the thermodynamic temperature. The thermodynamic temperature  $k_B T \approx 4 \times 10^{-21}$  J while the energy associated with cortical tension alone in a single cell is approximately  $5 \times 10^{-13}$  J (**Fig. 3.3A**), a separation of 8 orders of magnitude. This distinction between thermal and active fluctuations is more difficult to make at smaller scales, where, for example, a molecular motor may use energy on the order of  $10k_B T$  per step (Ritort, 2007; Milo and Phillips, 2015).

## 4. Results

Given the definitions of microstates and macrostates and their relationships to energy and entropy we established in chapter 3, we proceed to experiments manipulating these parameters and demonstrate that tissue structural distributions are well described using this framework. I first describe experiments demonstrating how tissue structural distributions are influenced by cell ratio (configurational entropy), adhesion knockdowns (tissue energy), and activity, then describe experimental work using breast cancer oncogene *PIK3CA* to generate a new disordered tissue steady state, and targeted perturbations that revert this phenotype.

### 4.1: Organoid structures follow a Boltzmann distribution

The Boltzmann distribution for our organoid structures is

$$P_i = \frac{W_i}{Z} e^{-\beta dE \cdot i}$$

where:

- $i$  : The Edge Occupancy (EO) of the organoid, the proportion of its outer boundary that is LEP (or GFP). Estimated as a weighted average of the EO of the boundaries of the three center cross-sections.
- $P_i$  : The probability of observing an organoid with  $\text{EO} = i$ .
- $E_i \approx i \cdot dE$  : The relative energy of a tissue with  $\text{EO} = i$ , approximated as a linear function of  $i$ . Estimated by linear regression on the total tissue energies of simulated tissues parameterized with measured interfacial tensions.
- $W_i \propto P_{\text{MEP},i}$  : The relative frequency of tissue arrangements with  $\text{EO} = i$ , equal to the probability of observing an equivalent MEP-only organoid with  $\text{EO} = i$ . Esti-

ated by fitting a truncated normal distribution to the observed MEP-only organoid distribution.

- $1/\beta$  (activity): The tissue activity, or random energy fluctuations that drive cell position changes independent of the average energy landscape defined by  $E_i$ .
- $Z$ : The partition function, a scaling factor that ensures  $\sum_i P_i = 1$ .

The points are very close to linear, as predicted by energy simulations, and slope of the regression is approximately  $\beta dE$ . We find that the Boltzmann distribution generated from the inferred value of the regression constant,  $\frac{W_i}{Z} e^{-\beta dE \cdot i}$ , demonstrates a good fit to our data,  $P_i$ . The numerical values of  $\beta dE$  will be inferred by nonlinear regression directly on  $P_i$  (section 2.11).

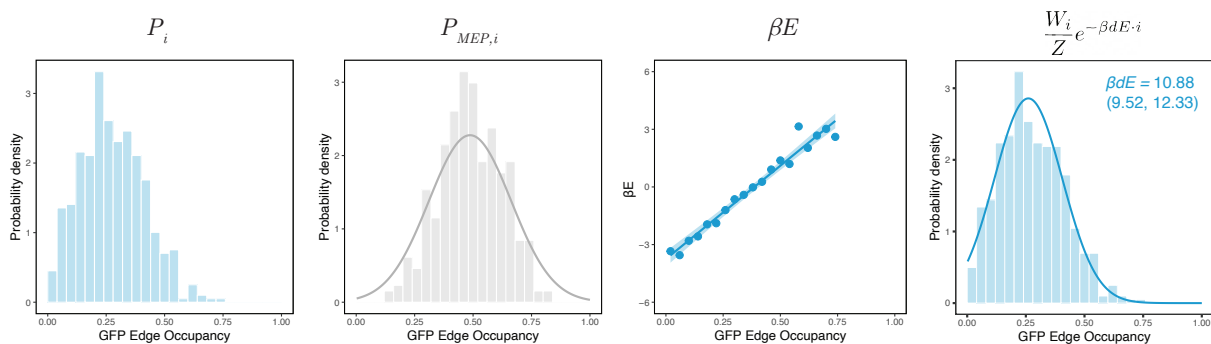


Figure 4.1: Organoids of GFP LEP and mCherry MEP generate a structural distribution, representing  $P_i$ . GFP and mCherry MEP-only tissues generate the probability distribution  $P_{MEP,i}$ , which is proportional to  $W_i$ . The MEP-only distribution is matched to a normal distribution, and  $\beta E_i$  is calculated as  $-\ln\left(\frac{P_i}{W_i}\right)$ . These parameters can be used to calculate a Boltzmann distribution,  $\frac{W_i}{Z} e^{-\beta dE \cdot i}$ , that is similar to the original organoid structure distribution  $P_i$ .

## 4.2: Cell ratio and tissue entropy

The underlying structural entropy is determined by the geometrical constraints of the system, particularly the cellular composition. Changing the ratio of green to red cells causes a corresponding shift in the structural distribution, both for MEP-only tissues and for LEP/MEP

tissues, and the magnitude of that shift is consistent with the Boltzmann model  $P_i = \frac{W_i}{Z} e^{-\beta E_i}$ , changing  $W_i$  while maintaining  $\beta E_i$ .

Sets of organoids with defined GFP fractions were generated by seeding aggregates with different ratios of GFP and mCherry cells, and by subsetting the images by the total proportion of GFP pixels (**Fig. 4.2**). For 25%, 50%, and 75% tissues, subset bounds were defined as [0.15, 0.35], [0.4, 0.6], and [0.65, 0.85]. For MEP-only tissues, GFP and mCherry fluorophores were considered interchangeable, so the same tissues are used for 25% and 75% with colors reversed as needed. The observed  $P_i$  LEP+MEP distributions produce inferred energies that show that these distributions are indeed consistent with changing  $W_i$  while maintaining  $\beta E_i$  (**Fig. 4.2A-B**). Both simulated and observed  $P_{\text{MEP},i}$  distributions are centered on 25%, 50%, and 75% Edge Occupancy (**Fig. 4.2C**).

One feature I anticipated, but that did not appear to have a prominent role in either the inferred or predicted energies, was an increasing expected slope with increasing LEP fraction. A relationship between LEP fraction and  $dE_i$  would arise from the slightly larger predicted energy cost of moving a MEP from the shell to the core if the MEP-ECM interaction is replaced with MEP-LEP rather than MEP-MEP interactions due to a higher baseline probability of attaching to a LEP. Our energy parameters, which have minimal differences between cell-cell energies, do not display this effect, but it appears in simulated tissues with larger variations in cell-cell energies, as shown in a simulated example with all other interfacial tensions set to 0 and only the LEP-LEP set to -0.1 (**Fig. 4.2D**). In HMEC, we are likely unable to detect any difference between  $dE$  between tissues of differing LEP fractions, but it is possible that other self-organizing cell combinations will have cell-cell interfacial energies that make cell ratio a critical factor in determining the energies,  $E_i$ , rather than primarily the entropy,  $W_i$ , as in HMEC organoids.

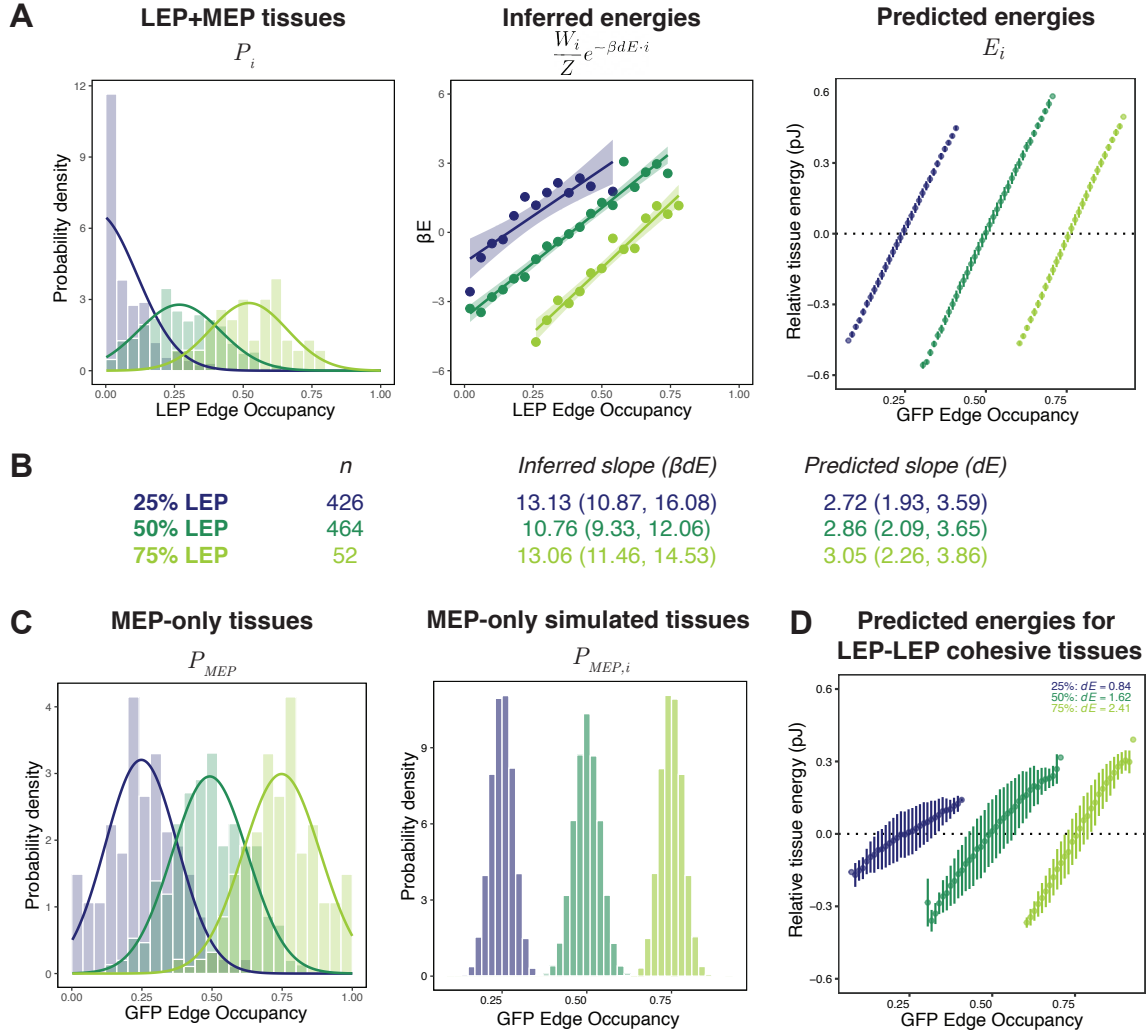


Figure 4.2: The probability distributions of the LEP/MEP tissues of varying cell ratios have entropies, but not energies, that are dependent on cell ratio. Inferred  $\beta dE$  slopes are shown with 95% CI envelope. Simulated tissues predict very little change in  $dE$  (A). Bootstrapped predicted and inferred slopes with 95% CI. Point estimates of  $dE$  increase with increasing LEP proportion (B). Randomly generated simulated tissues have Gaussian distributions centered at the total proportions of green cells, much like the observed MEP-only tissue distributions (C). LEP fraction has a stronger influence on  $dE$  if the relative importance of cell-cell cohesion is increased, as in simulated tissues where LEP-LEP is the only favorable interface (D).

### 4.3: Estimated tissue energies predict experimental outcomes

In our system, differences in energy between tissue configurations derive from different interfacial energies of cell-cell and cell-ECM interfaces. In an adhesive microenvironment,

the large difference in cell-ECM interfacial energies between MEP and LEP drives tissue self-organization, an energy gradient that can be estimated by tissue simulations and inferred from the observed structural distribution of LEP/MEP organoids relative to MEP-only organoids. A set of estimated and inferred energy slopes from across experimental conditions is reproduced in Appendix B.

MEP-only organoids have no differences between cells other than fluorophore, but we can synthetically generate an energy gradient by knocking down proteins involved in cell adhesion. As individual integrins are often dispensable in the mammary gland (Klinowska et al., 2001; Naylor et al., 2005), we used talin-1 knockdown to disrupt cell-ECM adhesion in MEP. Talin-1 is a protein that plays a critical role in integrin function by coupling the actin cytoskeleton to integrins' cytoplasmic tails and by triggering activating conformational change upon binding (Tadokoro, 2003; Calderwood, 2004). As previously demonstrated, targeting talin-1 in MEP with a silencing short hairpin RNA (shRNA) construct dramatically decreases MEP-ECM contact angle without affecting cell-cell adhesions (Cerchiari et al., 2015). GFP *TLN1*sh MEP and normal MEP sort similarly to GFP LEP and normal MEP in Matrigel because both are characterized by a large difference in cell-ECM interfacial tensions. By contrast, the same cell combinations have moderate to high EO steady states in non-adherent agarose microenvironments, where cell-cell interactions dominate (**Fig. 4.3**).

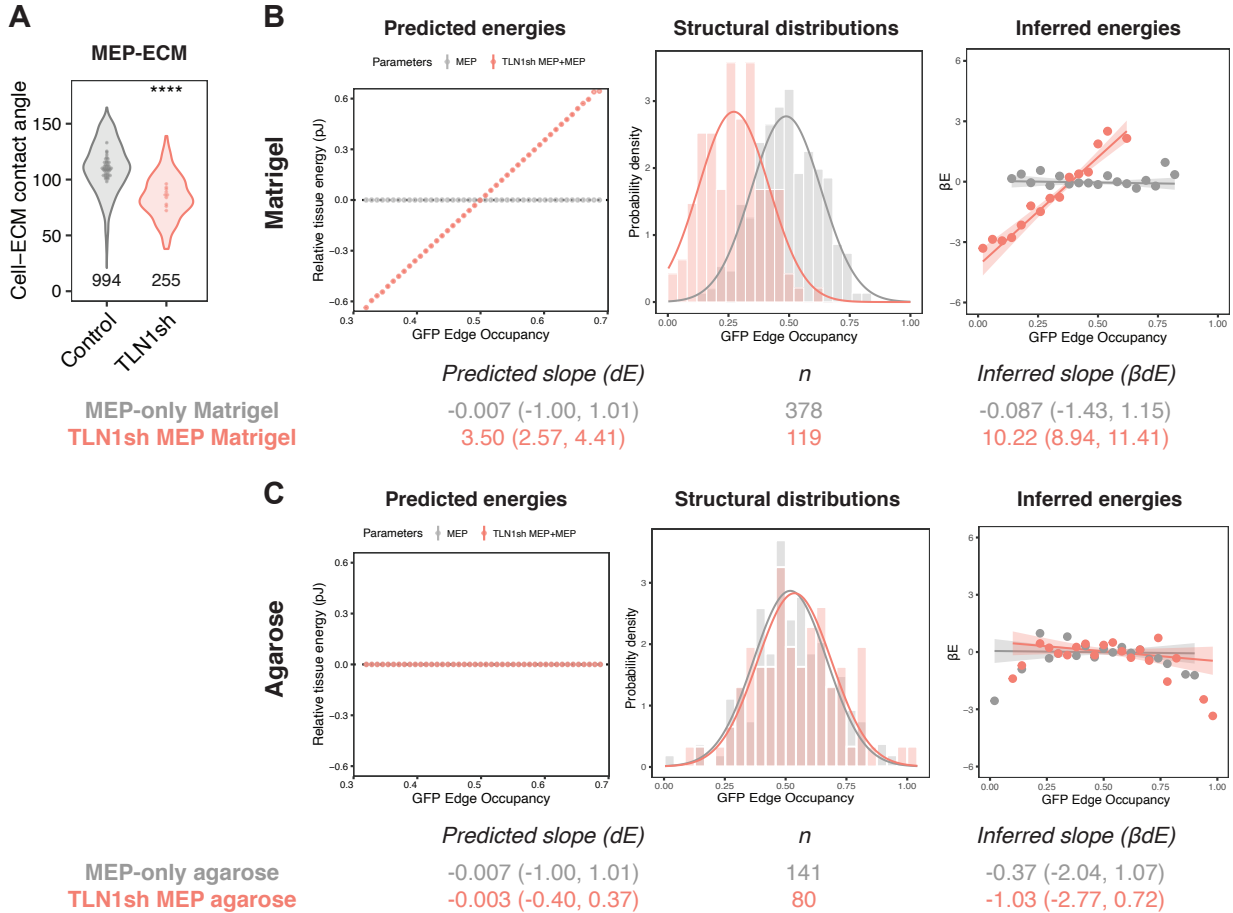


Figure 4.3: MEP-ECM contact angles for control and *TLN1sh* MEP (A). In the presence of ECM, the difference in cell-substrate interfacial tension ( $\gamma_{cx}$ ) between control and *TLN1sh* MEP causes the *TLN1sh* MEP to sort to the interior of the tissue due to a predicted positive slope of the tissue energy,  $dE \approx 3.5$ ;  $\beta dE \approx 10.22$  (B). However, in nonadherent agarose conditions, with no predicted energy difference, tissues are indistinguishable from MEP-only aggregates (C).

#### 4.4: Agarose microenvironment decreases activity

Using both a forward and inverse approach to estimate the tissue energy landscape for control organoids in Matrigel and agarose yields different estimates of the strength of the energy gradient (Fig. 4.4A). These differences can be reconciled because of the presence of the activity term,  $\beta$ , which scales the strength of the energy gradient. Intuitively, the positional spread of marbles in a dish can be controlled both by the steepness of its walls and by how vigorously the dish is shaken. Random positional fluctuations driven by cell motility play



an important role in allowing the tissue to sample the state space in order to reach the steady state, but also cause transient movements in and out of the steady state. To capture this random, undirected cell motion, we used live imaging of sparsely labeled nuclei within homotypic aggregates of single cell types, ensuring that the measured cell motility is only a function of random fluctuations and not driven by sorting along an energy gradient. In both LEP and MEP, the diffusion coefficient is decreased in agarose relative to Matrigel, suggesting that  $\beta_{\text{Matrigel}} > \beta_{\text{agarose}}$  (**Fig. 4.4B**). This decreased cell neighbor exchange can also be observed in pre-sorted MEP-only organoids, which relax towards a high Intermixing Score steady state over a few hours. The mixing process is slower in agarose than in Matrigel, indicating reduced cell position fluctuations (**Fig. 4.4C**).

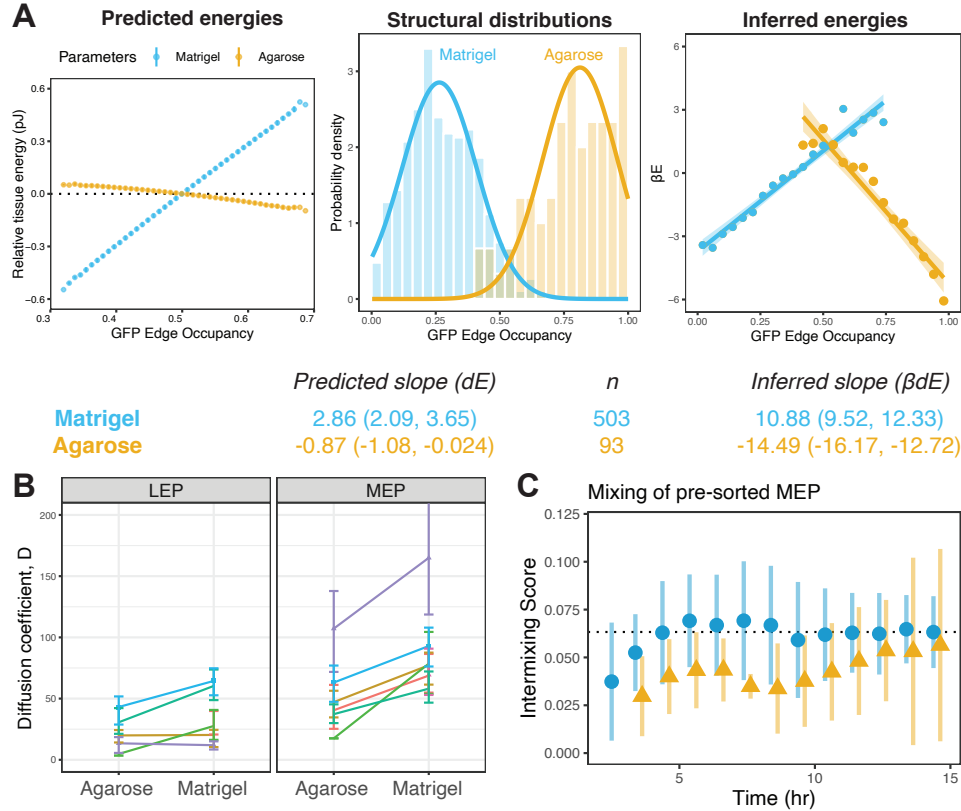


Figure 4.4: Based upon predicted tissue energies for the parameters measured from cell contact angles and cortical tensions, cells in agarose should barely sort, or not sort at all. However, organoids in Matrigel and agarose sort to a similar degree and have inferred  $\beta dE$  values that are very close in magnitude (A). Nuclear tracking of cell diffusion in homotypic aggregates indicate that LEP and MEP differ substantially in their capacity to move through a tissue, but for both cell types, cell position exchange is lower in agarose. Points are paired by experiment date and bars show bootstrapped estimates of the standard error (B). The Intermixing Score of pre-sorted MEP organoids in Matrigel (blue) increases more rapidly than in agarose (yellow) (C). Points are mean  $\pm$  standard deviation.

Similarly, we used a p120-catenin (*CTNND1sh*) knockdown to target cell-cell adhesion in MEP to make MEP-only tissues predicted to sort with the less-cohesive *CTNND1sh* MEP in the shell. This adapter protein is a master regulator of cadherin stability and membrane localization. While bound to the cytoplasmic tail of cadherins, it is thought to regulate cell-cell junctions by inhibiting cadherin degradation; it also regulates cytoskeletal dynamics via interactions with small GTPases (Xiao et al., 2007). In addition to the expected reduction of cell-cell adhesion, *CTNND1sh* MEP are also less adhesive to ECM. Together

these perturbations lead to a modest energy gradient predicting sorting in both Matrigel and agarose. However, only tissues in agarose demonstrate sorting behavior, indicating a difference between the inferred  $\beta dE$  and the predicted  $dE$  (Fig. 4.5).

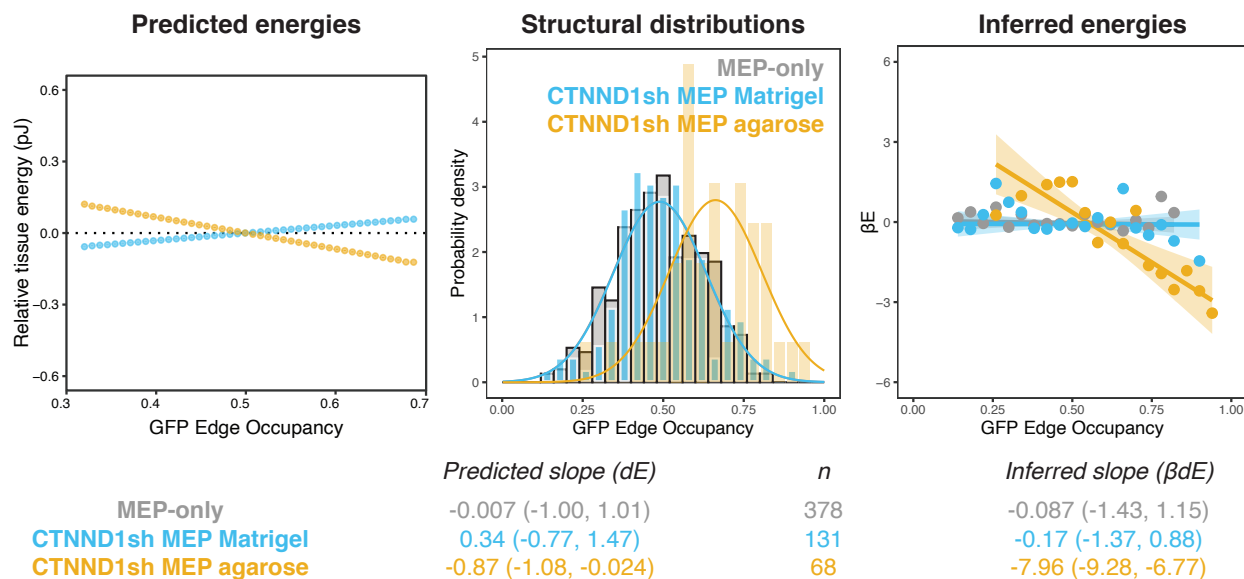


Figure 4.5: Despite the weak energy gradient predicted for *CTNND1sh* MEP, these tissues demonstrate noticeable sorting in agarose. On the other hand, sorting is not observed in Matrigel, where high activity and/or low  $dE$  together cause the structural distribution to be indistinguishable from the MEP-only distribution. MEP-only histogram is outlined in black to improve its visibility under the overlying *CTNND1sh* MEP agarose distribution.

The predicted  $dE$  in *CTNND1sh* MEP Matrigel is extremely modest, however, and these data are also consistent with a model where downregulation of p120-catenin, through decreased cadherin stability and decreased actin cytoskeletal dynamics, alters both cell-cell and cell-ECM interfacial interactions to a similar magnitude, with the net effect of no predicted sorting in Matrigel. Accordingly, while the predicted and inferred  $dE$  and  $\beta dE$  slopes are statistically significantly different from 0 in *CTNND1sh* MEP agarose, the *CTNND1sh* MEP Matrigel slope 95% CIs contain 0.

## 4.5: Oncogenic PI3K shifts the steady-state distribution

In the context of breast cancer, this framework implies that among dysregulations associated with breast cancer, some perturbations or combinations of perturbations should potentiate LEP invasion by stabilizing the structure with LEP at the tissue edge. We tested a panel of the most commonly dysregulated genes in breast cancer by lentiviral transduction in LEP mixed with wild-type MEP, and discovered that expression of PI3K bearing the H1047R or E545K point mutations had a marked effect on self-organization capacity. This decrease in tissue organization correlated with an increase in LEP-ECM contact angle, indicating an alteration of LEP-ECM interfacial energy (**Fig. 4.6**).

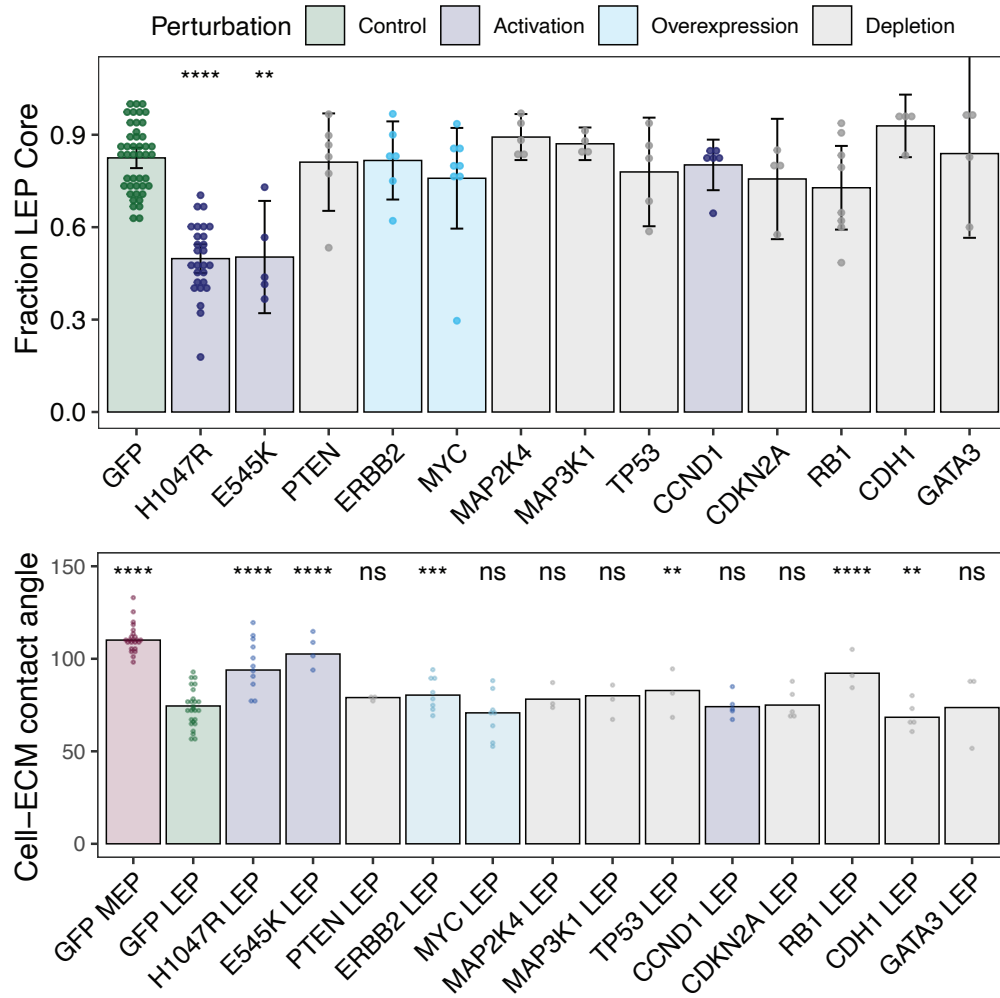


Figure 4.6: A panel of the most commonly dysregulated driver genes in breast cancer was transduced individually into passage 4 LEP. Organization frequency was measured as a proportion of tissues observed in the LEP core class. Constitutively active PI3K (either of two p110 $\alpha$  point mutants, H1047R and E545K) had the strongest effect on organoid structure and the strongest corresponding increase in LEP-ECM contact angle. Points mark means of different experiment dates and error bars indicate standard deviation of experiment means. Multiple Wilcox comparisons were performed against wild-type organoids and GFP LEP references with the Holm post-hoc correction.

*PIK3CA*<sup>H1047R</sup>-expressing LEP have a cell-ECM contact angle and interfacial tension intermediate between wild-type LEP and MEP, which predicts a decrease in the magnitude of the predicted energy slope  $dE$ , but not a change in its sign. Indeed, measuring the GFP Edge Occupancy of *PIK3CA* organoids confirmed that although the structures were disrupted, the distribution as a whole was still slightly more organized than expected from random chance,

e.g.  $P_i \neq \frac{W_i}{Z}$ . Inferring the exponent term  $\beta dE$  from the observed distributions confirmed the prediction that for *PIK3CA*,  $dE$  is reduced but still positive (**Fig. 4.7**).

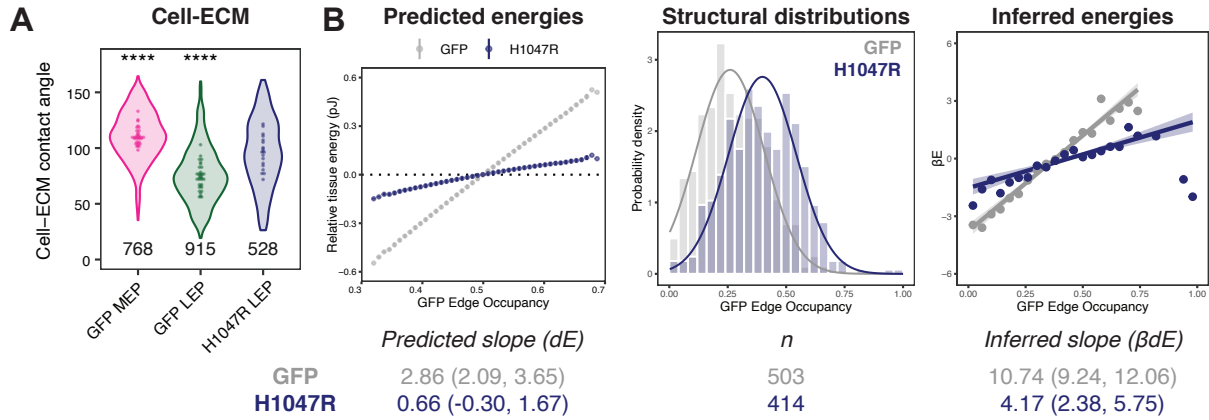


Figure 4.7: Characterization of the effect of *PIK3CA*<sup>H1047R</sup> expression on cell mechanical phenotype and tissue structural outcomes. Transformed LEP adopt high cell-ECM contact angles, reflecting their decreased interfacial tension at that interface (A). The new energies predict a statistically significantly moderated energy landscape where tissues are no longer driven to sort to the same extent, which corresponds to separations between wild-type and *PIK3CA* observed structural distributions and their inferred energies (B).

#### 4.6: Reversion to the steady state by altering energy and activity

PI3K sits at the head of a major signaling pathway with many downstream effects, including cell growth and proliferation, metabolism, migration, and secretion (Hemmings and Restuccia, 2012; Lien et al., 2017). It is reasonable to wonder whether the increased cell-ECM contact angle is in fact necessary for the structural disruption, or if it is merely a side effect and some other PI3K-related phenotype is responsible. To determine the importance of cell-ECM interaction in maintaining tissue structure, we used *TLN1*sh, which we previously described as inducing a cell-ECM energy change sufficient to drive cell sorting in MEP-only tissues. As it interacts directly with integrins and the actin cytoskeleton, talin-1 knockdown is a good candidate for an orthogonal perturbation that specifically targets cell-ECM adhesion rather than more general aspects of PI3K/AKT signaling. Talin-1 knockdown in *PIK3CA* LEP was performed by co-transfection of two lentiviruses containing GFP and puromycin

markers. Control cells were transduced with the empty puromycin vector and treated with the same puromycin selection (**Fig. 4.8**). The knockdown successfully reduced cell-ECM energy, restoring the predicted energies, structural distributions, and inferred energies to near wild-type levels.

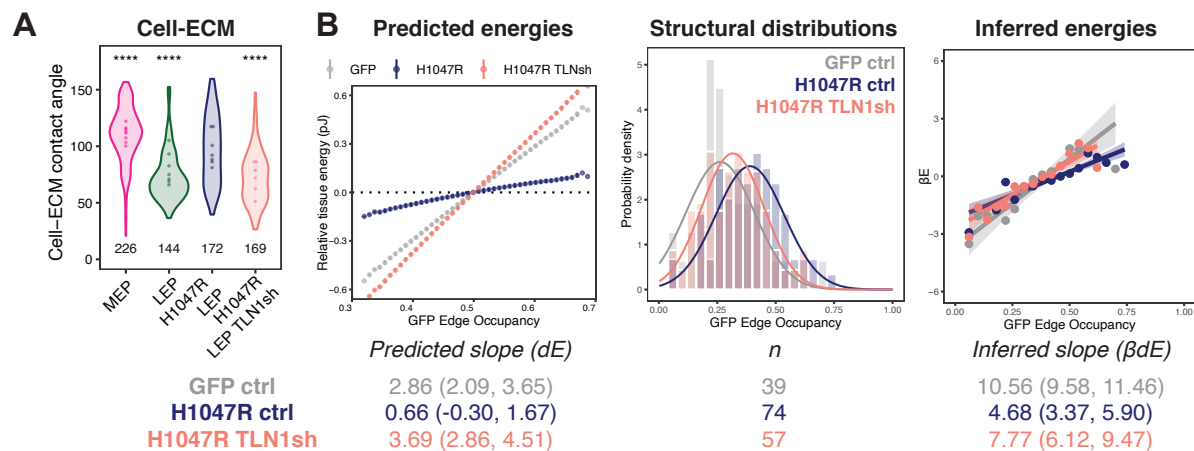


Figure 4.8: Co-transduction of shRNA targeting talin-1 in  $PIK3CA^{H1047R}$  LEP causes a drop in cell-ECM contact angle. Points represent means of different experiment dates. Significance testing by Wilcox test with Holm correction for multiple comparisons against the H1047R LEP reference group (A). The predicted energy gradient of  $PIK3CA$  TLN1sh organoids increases to be similar to the wild type. The observed tissue distribution returns close to the control, with an inferred energy slope  $\beta dE$  that is not statistically different from that of wild type (B).

We tested several small-molecule inhibitors to determine if they could revert the  $PIK3CA$  phenotype, including inhibitors of PI3K, AKT, RAC, PDK, WNT/GSK3 $\beta$ , MEK, ROCK, MTOR, and TGF $\beta$ , which were validated by suppression of phosphorylation of downstream targets in Western blots. We determined that MK-2206, an allosteric inhibitor of AKT, induced a restoration of tissue structural distribution (Hirai et al., 2010). However, no change in the predicted energies was observed. This implicates activity in restoring the probability distribution back towards the lower energy tissue configurations, and indeed, cell diffusion assays demonstrate that MK-2206 decreases activity in homotypic aggregates (**Fig. 4.9**). These results show that the probability distribution of an ensemble of organoids

at the steady state can be shifted by altering either energy or activity even when the initial perturbation was due to altering only the system's energetics.

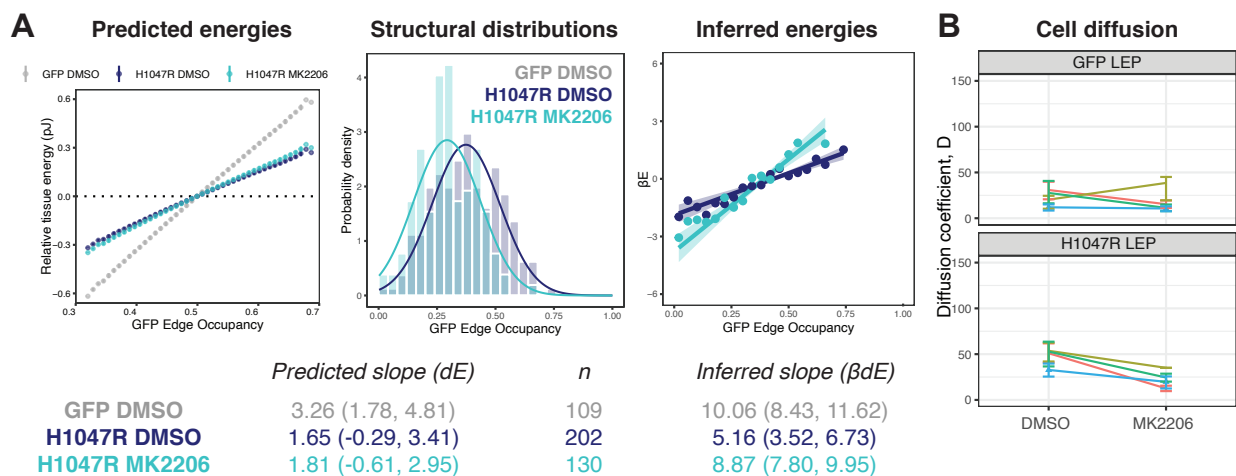


Figure 4.9: MK-2206, an inhibitor of AKT, restores self-organization of *PIK3CA* LEP without altering the cell-ECM energies. Predicted energy gradients remain the same for MK-2206 and DMSO conditions, despite there being a difference in the structural distributions and the inferred energies for MK-2206 (A). Instead, MK-2206 appears to act by suppressing tissue activity, as indicated by a decrease in cell diffusion in LEP-only and *PIK3CA* LEP-only aggregates treated with DMSO vehicle control or MK-2206. Points are paired by experiment date and bars show standard deviation across different individual aggregates (B).



## 5. Discussion

### 5.1: Cell sorting in the context of cancer

These results provide a unique perspective on a number of clinically relevant observations, viewed through the lens of tissue self-organization and the active processes that maintain normal tissue structure. Tissue structure itself is maintained by the activity of normal MEP, both as a physical barrier and a source of tumor-suppressive biochemical signals. It is reported that alterations to and loss of normal MEP may increase risk of DCIS progression (Allinen et al., 2004; Hilson et al., 2009; Russell et al., 2015; Sirka et al., 2018), and these changes to cell ratio and MEP-ECM phenotype have direct consequences for increasing configurational entropy and reducing energetic barriers to structural disruption. Similarly, invasive LEP are associated with phenotypes such as hyperproliferation, epithelial-mesenchymal transition, altered cell-cell and cell-ECM interactions, and increased cellular tension, which reflect changes in cell proportion and cell interfacial tensions (Choi et al., 2013). Environmental factors, such as post-pregnancy involution and hormone replacement therapy, are known to increase breast cancer risk through involution, pro-inflammatory signals, ECM remodeling, and cell proliferation, suggestive of links to increased tissue activity or dynamics, altered cell-ECM interfacial energies, cell state changes, and cell ratio changes (Schedin, 2006; Narod, 2011). These cellular and tissue changes associated with cancer progression could be considered to cause changes in  $E_i$ ,  $W_i$ , and  $\beta$ , altering the underlying parameters determining the steady-state tissue configuration, and ultimately leading to increased occupancy of LEP at the tissue edge. These parameters describe specific phenotypes that render the breast more susceptible to progression from DCIS to invasive disease, a significant clinical and conceptual contribution to understanding this critical transition.

In our panel of driver perturbations, while many other oncogenes are associated with hallmarks of cancer, such as cell proliferation and survival (i.e., *MYC*, *TP53*, *CCND1*,

*CDKN2A*, *RB1*), direct structural changes were only observed in the oncogene that also changed the cell-ECM interaction (**Fig. 4.6**). Interestingly, *ERBB2* did not have a similar effect, although HER2-overexpressing breast cancers are known to have increased risk of metastasis and are associated with ECM adhesion and remodeling phenotypes (Eccles, 2001). Rather than the constitutively active mutant NEU-NT traditionally used in mouse tumorigenesis (Ursini-Siegel et al., 2007; Fry et al., 2017), we chose to overexpress wild-type HER2, as breast cancer patients rarely ( $\approx 4\%$ ) show activating mutations in HER2 (Cocco et al., 2019). However, transduction of only the coding sequence excludes splice isoforms known to increase aggressiveness in patients, such as  $\Delta 16$ HER-2 (Jackson et al., 2013; Volpi et al., 2019), and we do not co-transduce with HER3, which is considered to be critical in heterodimerizing with HER2 to activate PI3K/Akt pathway signaling (Junttila et al., 2009; Ruiz-Saenz et al., 2018). A different choice of construct or a co-expression strategy may show a stronger transforming phenotype. We also observed that shRNA targeting of E-cadherin (*CDH1*) was insufficient to induce a structural change in mammary organoids, though one might expect decreased cell-cell adhesion to play a role in cell positioning. However, as previously described, LEP-LEP interactions are relatively weak and play very little role in DITH-driven cell sorting of HMEC organoids in three-dimensional aggregates. Others have observed a role for E-cadherin-dependent cohesion of LEPs in PDMS microwells (Chanson et al., 2011); my closer examination of HMEC in these wells suggests these results are consistent with delamination of MEP and LEP in PDMS, with MEP spreading on the PDMS walls as a 2D substrate and LEP forming a detached central aggregate (data not shown). Loss of E-cadherin is a hallmark of epithelial-to-mesenchymal transition (EMT) and is particularly prevalent in lobular carcinomas, which often exhibit diffusive growth and metastasis patterns rather than collective migration (Cowin et al., 2005). E-cadherin downregulation alone is not sufficient to disrupt HMEC cell positioning, and its loss may be less relevant to cell sorting in the context of a noninvasive tissue than other EMT-associated changes, such as acquisition of mesenchymal phenotypes like migration and ECM remodeling (Felipe Lima

et al., 2016). Overall, genetic perturbations must be interpreted in light of their phenotypic outcomes. In our finite-lifespan HMEC, single perturbations rarely affected cell mechanical or organization phenotypes; in the in vivo context of multiple hits, it is likely the case that many disparate genetic and epigenetic alterations and their combinations can converge to similar mechanical phenotypes.

The particular *PIK3CA*<sup>H1047R</sup> used in our study was also reported in a mouse model of breast cancer, with a very similar cell-sorting phenotype (Van Keymeulen et al., 2015; Koren et al., 2015). Tamoxifen-inducible mouse *Pik3ca*<sup>H1047R</sup> was expressed in luminal or basal (myoepithelial) cells. Transformed MEP produced a non-representative spectrum of tumor types, while transformed LEP gave rise to a variety of heterogeneous tumor types. Interestingly, prior to tumorigenesis, low levels (5-30%) of transformed LEP were found within the basal MEP tissue compartment, suggesting the acquisition of MEP-like cell sorting properties. These cells also displayed phenotypic plasticity, adopting MEP molecular markers as they translocated to the basal compartment, suggesting that cell fate decisions could be influenced by their position within the mammary gland and prompting another interesting avenue of future work.

## **5.2: Structural steady states in development and differentiation**

Cell identity and plasticity are tied to their microenvironment. In the mammary gland, it has been shown that adult basal stem cells give rise to only one lineage under physiological conditions, where cells experience a full complement of neighboring cell types. However, these progenitor cells can reactivate multipotency when transplanted into epithelium-free fat pads or when luminal cells are ablated in situ. This context-specific plasticity is also observed in other bilayered epithelia like prostate, salivary gland, and sweat glands (Centonze et al., 2020). The arrangement of cells within a tissue relative to one another and their extracellular matrix is an essential feature of the stem cell niche, as cell potency and fate are controlled by paracrine cues, cell-cell contact, and ECM signaling (Walker et al., 2009; Chen et al.,

2013; Ottone et al., 2014; Pinho and Frenette, 2019). The self-organizing organoid model system has the potential to integrate these physical and biological features of development and stem cell biology.

Because of cell motility, proliferation, and death, stem cell niches can be dynamic, involving changes in the relative strengths of these cues over time as cells stochastically sample different local microenvironments. Using live intravital imaging of intestinal crypts, Ritsma et al. (2014) showed that the cells of the crypt base dynamically “compete” for limited niche access, and that stem cells positioned near the boundary are more likely to be lost or replaced, whereas stem cells spatially distant from the edge are more likely to be maintained. Similarly, live lineage tracing in the hair follicle niche revealed position-dependent stem cell fate decisions, while also demonstrating that more differentiated epithelial cells repopulate the stem cell compartment after laser ablation (Rompolas et al., 2013). These results suggest that cell potency and plasticity are regulated by cell position. Cell position, in turn, is regulated by cell mechanical phenotypes through sorting phenomena. In the generation of intestinal organoids from single LGR5-positive stem cells, prior to the generation of distinct crypt zones, the initially identical cells break symmetry through transient fluctuations in YAP1 expression (Serra et al., 2019). These first symmetry-breaking events give rise to the first Paneth cell and the morphogenesis of the crypt, a spatially defined stem cell niche that then concentrates and maintains secreted niche cues, demonstrating how cell state can trigger tissue morphogenesis that then reinforces cell state.

It would be fascinating to use organoids as individual miniature stem cell niches, where cells actively and transiently sample different positions with a frequency based on their relative mechanical properties. If a stem cell is retained at the tissue edge for a longer period, does its increased exposure to ECM lead it to differentiate toward a more MEP-like state? Conversely, if a stem cell spends more time buried within the tissue, does it differentiate toward a more LEP-like state? The time evolution of such a tissue model would follow the cell sorting process, but also incorporate a time-varying Hamiltonian due to changes

in phenotype that could then stabilize or alter the original steady state, reminiscent of the cell position and adhesion feedback loops developed by Toda et al. (2018). We could quantitatively measure the probability distributions of these fate decisions, generating a potency landscape composed not only of gene regulatory networks, but also of cell structural coordinates.

### **5.3: Organoids as a model of active matter**

This work has generated a number of exciting future directions in developing an active matter model of multicellular tissue. We have proposed a macroscopic model system which appears to behave according to statistical mechanical principles, with experimental accessibility to previously inferred values such as the state of an individual system, the mechanical energies forming the energy landscape, and the configurational entropy of a non-sorting system. Despite the experimental and technical challenges and variability inherent to dealing with living cells, it has the appealing property that we are able to directly measure probability distributions (on discrete bins and with sufficiently large sample sizes) of organoid structures in addition to physically or genetically perturbing the system parameters.

We defined a short timescale for steady state to allow for the analysis of systems that change over longer timescales. A number of processes could occur over several days to a week that could disrupt the parameters of self-organization or the geometrical constraints of the system. For example, the energetic constraints of the system could be altered by changes in cell mechanical properties or phenotype. LEP in organoid culture for an extended period can polarize and begin to develop lumens, representing a change in cell state and an altered interfacial tension at the apical surface (Camelo and Luschnig, 2020). In addition, MEP can secrete their own basement membrane proteins and appear to remodel the surrounding ECM, potentially creating a new substrate surface whose properties depend on its exposure to MEP and which can serve to physically constrain a growing tissue and guide morphogenesis (Wang et al., 2013; Morrissey and Sherwood, 2015; Haigo and Bilder, 2011). The overall

composition and size of the organoid can also change over time, if there are unequal cell proliferation rates or enough total cell proliferation to change the total size of the tissue. If these disruptions or “reactions” occur slowly compared to the rate of self-organization and the neighbor exchanges between cells in tissues, they can be approximated as negligible over the timecourses of interest. A typical timescale for reaching a steady-state organization for LEP and MEP is in the range of a few hours to a few days (**Fig. 3.7**), and has been the focus of the investigations in this dissertation. However, the dynamics of this process could be explored in further detail by embracing its nonequilibrium properties and applying principles from stochastic thermodynamics.

The role of tissue activity is of particular interest, and our analysis would be greatly improved with experimental approaches for perturbing cell dynamics independently of cell mechanics, and for measuring activity or  $T_{\text{EFF}}$  more directly. Large datasets of ensemble measurements have been very useful for developing our understanding of steady-state distributions, but there is much yet to discover with dynamic information. Live organoid imaging data with technique refinements to improve cell viability and imaging resolution will be important to future investigations. For example, with additional timelapse data, we may be able to demonstrate detailed balance at steady state, or identify timescales over which detailed balance is satisfied or broken (Gnesotto et al., 2018). The fluctuation-dissipation theorem (FDT) of linear responses to perturbations is a requirement of equilibrium. With sufficiently high-resolution timelapse imaging, we may be able to measure deviations from FDT to infer an effective temperature, or to quantify the degree to which the tissue activity resembles an effective temperature (Kubo, 1966; Loi et al., 2008; Turlier et al., 2016). Using a photopatterning approach, we can pattern cells with single-cell spatial resolution in a high-throughput manner, essentially generating “perturbed” tissues of arbitrary distance from the steady state (Cabral et al., 2021). Following their trajectories through state space should allow us to better understand the dynamics of tissue self-organization towards a steady state attractor. Further characterization of the time-dependent portion of self-organization, the

approach to steady state, and the response to being perturbed far from steady state may allow us to connect tissue structure transitions to a quantitative measure of entropy production as a function of the statistical time-irreversibility of forward and reverse transitions (Parrondo et al., 2009; Ciliberto, 2017; Li et al., 2019; Martínez et al., 2019).

## 5.4: Caveats and alternative approaches

The mammary organoid cell sorting system sits at the edge of the “small” systems. It is small enough to approximate a Hamiltonian for tissue energy from physical principles, in this case from the cell interfacial tensions, but large enough that simplifications are required to distill the complex mechanobiology into quantifiable components. Rather than modeling each molecule involved in the cell adhesion machinery, the contributions of many microscopic biological processes are averaged into single terms with the expectation that across interfaces spanning many molecular complexes and over a tissue composed of many cells, average values are appropriate approximations.

However, it is possible that individual cells will have intrinsic, non-transient differences in mechanics from other cells of ostensibly the same type. Cell to cell variation, in particular, anti-correlated variations across multiple parameters, could have significant impacts on the structures of individual organoids. For example, if a LEP were both unusually less cohesive to other LEP and more adhesive to ECM, it would have a strong energetic push to the tissue edge, shifting the minimum energy structure to higher EO. We could detect persistent cell to cell heterogeneity with more long-term live imaging of individual organoid trajectories, as unusual cells would cause their organoids to fluctuate around a different steady state structure from the population. On the other hand, because cortical tension, cell-cell contacts, and cell-matrix contacts share molecular machinery and pathways, it is possible that these parameters are positively correlated with each other on the level of individual cells. Similar to the p120 catenin knockdown experiments (**Fig. 4.5**), correlated variations tend to cancel themselves out in our system; the combined increase in  $\gamma_{cc}$  and  $\gamma_{cx}$  resulted in a very shallow

$dE$  for *CTNND1*sh MEP + MEP aggregates. In such an adhesion hierarchy, heterogeneity at the cellular level would play a limited role. Alternatively, cellular heterogeneity could contribute to structural disorder in conjunction with configurational entropy, analogous to transcriptional noise in a gene expression state space (Chang et al., 2008). Indeed, spatiotemporally heterogeneous deviations in the energy Hamiltonian would cause fluctuations in cell position, essentially representing an activity term.

A linear relationship between EO and tissue energy is predicted by simulations and supported by our inferred energies,  $-\ln \frac{P_i}{W_i}$ . Yet, there are some reasons we might expect to see nonlinearity. First, in systems where differences in cell-cell interfacial energies are significant relative to differences in cell-ECM, the energy gradient is accordingly steeper along the IMS coordinate, which will affect the intermediate values of EO, which are dominated by high IMS mixed structures. Such curves are visible in a number of our parameter sets, such as those in agarose, with a slight concavity near  $EO = 0.5$ . Second, the use of an effective energy term  $E_i = \bar{\epsilon}_i - \frac{\beta\sigma^2}{2}$  introduces two potential sources of nonlinearity,  $\beta$  and  $\sigma^2$ . Variance may be more pronounced at low values of IMS because we sample only a range of the center of each organoid slice, which may by chance exclude or include the segregated region. Activity may also be dependent on structure, as we have observed that activity differs dramatically between LEP and MEP (**Fig. 4.4**). Although we typically observe intermediate levels of diffusion in mixed organoids for nuclei of either cell type, it is also possible that activity will be inhomogeneously distributed in highly segregated, low IMS tissues, preventing cell exchange between the low and high activity tissue compartments in a manner reminiscent of motility-driven active matter sorting or motility transitions in jammed tissues (Stenhammar et al., 2015; Angelini et al., 2011). Third, the predicted energy landscape developed from the use of a simulated tissue does not span the parameter space due to cell shape limitations. The high and low extremes of EO are geometrically inaccessible to non-deformable cells. With the inclusion of deformable or off-lattice cells, the energy landscape may be similar to the linear interpolation, or it may differ, for example if cell deformation incurs an energy cost.



While we found the linear approximation to perform well for our system, more significant observed deviations could easily be incorporated into a new nonlinear formulation for  $E_i$ .

Cutoffs for size, cell ratio, and circularity ensure that the inputs to the structural distribution analysis are controlled and uniform in their overall constraints. However, they also can lead to low numbers of observations, which hampers an analysis based on sometimes subtle shifts in probability distributions. This analysis would benefit from methods development to reduce organoid variability and increase imaging throughput, such that more organoids can be imaged per experiment. Future approaches could incorporate membrane-anchored oligonucleotide-assisted cell micromolding, where cells are immediately and temporarily condensed within microwells upon centrifugation and can be moved to Matrigel directly rather than after a few hours (Weber et al., 2016). Another appealing approach involves microwells that are integrated directly into the microscopy workflow (Zhao et al., 2019; Brandenburg et al., 2020), though some optimization will be required to determine whether dilute ECM added over the top will provide a sufficiently homogenous and stable substrate for cell-ECM interface-driven cell sorting (Kakni et al., 2020), or if the fully enclosed organoids can generate their own ECM in a short timeframe (Todhunter et al., 2021). Alternatively, micro-tissues may be spun directly into micromolded ECM wells (Mori et al., 2009). Maintaining the culture in microwells would remove the aggregate transfer step and thus the chance of destroying weakly condensed organoids, which is particularly marked for cells with weak cell-cell interactions, like LEP and *CTNND1*sh cells.

Organoid culture is typically manual, introducing the potential for handling variability. Brain and kidney organoid researchers have also observed batch variability of cell differentiation and state between experiment dates (Quadrato et al., 2017; Phipson et al., 2019). However, in an automated, high-throughput assay platform for brain organoids, the donor and clone source were more significant sources of variability than batch (Shah et al., 2020), suggesting that even in stem cell-derived organoids, which are highly sensitive to small deviations in cell number and identity, automation is a powerful tool for reducing organoid

variability. The process for generating HMEC organoids could be automated by bioprinting cell-dense suspensions directly into or on top of ECM (Brassard et al., 2021; Lawlor et al., 2021), or by encapsulating cells in Matrigel droplets with microfluidics (Dolega et al., 2015; Jiang et al., 2020), which can generate organoids with uniform size and cell number.

That said, it is critical to ensure that added complexity does not lead to longer handling times or extended periods on ice. In our experience, these cells are sensitive and day-to-day variability in culture stress or handling stress can negatively affect their ability to be FACS sorted, condense, and self-organize. Furthermore, care must be taken to avoid introducing local heterogeneities into the ECM, which could manifest as a local energy gradient attracting cells into cracks or along aligned ECM fibers. For example, some optimization may be necessary before using a multi-step gelation process where organoids sit at the exposed face of one hydrogel surface before being fully embedded.

## Conclusion

Reaching and maintaining a steady state is a critical element of life. Variations on the theme of stability, fluctuations, and attractor states have been applied to biological systems across length scales, though they take their most concrete and rigorous form for small systems for which the interaction energies can be defined in terms of molecular binding energies. For larger biological systems, energy landscape approaches have generally been abstract, inferring “energy” as a measure of cell potency or organismal fitness from the distribution of observed states, usually of the form  $E = -\ln P$ , derived from a simple Boltzmann weight of  $e^{-E}$  without considering entropy and temperature, equivalent to using a free energy (Shekhar et al., 2013; Jenkinson et al., 2017).

By contrast, our system has distinctive features with measureable influences from entropy and activity (“temperature”). Our choice of structural coordinate, which compresses the full dimensionality of the spatial position of each individual cell into an informative structural macrostate, allowed us to make an explicit combinatorics argument from traditional statistical mechanics, which we confirmed by generating an energy-neutral non-sorting system (section 3.4). Building upon the extensive pre-existing theory and experimental evidence of the DITH, we defined a tissue-scale Hamiltonian from “first principles”. Empirical measurements of isolated cell interactions and a simplified three-dimensional lattice model were used to estimate the energy landscape (section 2.10), while observations of ensembles of organoids with varying parameters were quantified and transformed into inferred energy landscapes with an included activity term (section 2.11). It does not merely infer a deterministic energy landscape, but also defines the entropic role of random chance in steady-state tissue structure. Here, random chance is neither unpredictable nor unquantifiable. Configurational entropy represents a baseline level of heterogeneity in a system imposed by geometrical constraints, which can be quantified and estimated under controlled conditions. Live imag-

ing verified that  $\beta$ , the difference between predicted and inferred energy in Matrigel and agarose, could be attributed to differences in cell neighbor exchange throughout the tissue, which we quantified by nuclear diffusion, an approach similar to the use of tracer particles in active matter systems.

Together these experiments verified the applicability of statistical mechanical concepts to cell sorting and highlighted parameters important in determining and maintaining the steady-state structure of tissues. These parameters were borne out in practice, as the second-most dysregulated gene in breast cancer, *PIK3CA*, demonstrated its ability to shift the steady state away from the bilayered core-shell and towards a more mixed configuration. This change arose from altered cell-ECM interactions and could be reverted by tuning the Boltzmann parameters. Specifically targeting the cell-ECM interface with *TLN1sh* could directly reverse the energetic phenotype imposed by *PIK3CA* activation, but consistent with the Boltzmann parameters, an alternative, indirect approach was also effective. Reducing tissue activity, which we achieved using an inhibitor of AKT, restored the structural steady state to wild-type levels without changing the energetics of the system. Together these results verify the importance of all three components of the Boltzmann weight in determining the probability of a state: energy, entropy, and activity.

Tissue self-organization is a dynamic and integrated process in which cell identity, mechanical properties, position, and signaling regulate each other to generate tissue structure and function. We used the organoid system to define the relationships between cell mechanical properties and position at the steady state in this work, but it is exciting to consider applying similar ideas to the wide variety of processes that determine tissue structure. Our system could be amenable to studying the kinetics of self-organization or the reciprocity between cell position, niche, and state in an evolving tissue architecture. The Boltzmann equation has been remarkably useful in describing, predicting, and perturbing the steady-state structure of a macroscopic multicellular system, demonstrating the potential for physical principles to guide our understanding of biological development.

## References

- Allinen, M., Beroukhi, R., Cai, L., Brennan, C., Lahti-Domenici, J., Huang, H., Porter, D., Hu, M., Chin, L., Richardson, A., Schnitt, S., Sellers, W.R., and Polyak, K. Molecular characterization of the tumor microenvironment in breast cancer. *Cancer Cell*, 6(1):17–32, 2004. DOI: 10.1016/j.ccr.2004.06.010.
- AlQuraishi, M., Koytiger, G., Jenney, A., MacBeath, G., and Sorger, P.K. A multiscale statistical mechanical framework integrates biophysical and genomic data to assemble cancer networks. *Nature Genetics*, 46(12):1363–1371, 2014. DOI: 10.1038/ng.3138.
- Angelini, T.E., Hannezo, E., Trepat, X., Marquez, M., Fredberg, J.J., and Weitz, D.A. Glass-like dynamics of collective cell migration. *Proceedings of the National Academy of Sciences*, 108(12):4714–4719, 2011. DOI: 10.1073/pnas.1010059108.
- Barcellos-Hoff, M.H., Aggeler, J., Ram, T.G., and Bissell, M.J. Functional differentiation and alveolar morphogenesis of primary mammary cultures on reconstituted basement membrane. *Development*, 105:223–235, 1989.
- Battle, C., Broedersz, C.P., Fakhri, N., Geyer, W.F., Howard, J., Schmidt, C.F., and Mackintosh, F.C. Broken detailed balance at mesoscopic scales in active biological systems. *Science*, 352(6285):604–607, 2016. DOI: 10.1126/science.aac8167.
- Belmonte, J.M., Thomas, G.L., Brunnet, L.G., de Almeida, R.M.C., and Chaté, H. Self-Propelled Particle Model for Cell-Sorting Phenomena. *Physical Review Letters*, 100(24):248702, 2008. DOI: 10.1103/PhysRevLett.100.248702.
- Berg, S., Kutra, D., Kroeger, T., Straehle, C.N., Kausler, B.X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., Eren, K., Cervantes, J.I., Xu, B., Beuttenmueller, F., Wolny, A., Zhang, C., Koethe, U., Hamprecht, F.A., and Kreshuk, A. ilastik: interactive machine

- learning for (bio)image analysis. *Nature Methods*, 16(12):1226–1232, 2019. DOI: 10.1038/s41592-019-0582-9.
- Bhattacharya, S., Zhang, Q., and Andersen, M.E. A deterministic map of Waddington’s epigenetic landscape for cell fate specification. *BMC Systems Biology*, 5(1):85, 2011. DOI: 10.1186/1752-0509-5-85.
- Bi, D., Lopez, J.H., Schwarz, J.M., and Manning, M.L. A density-independent rigidity transition in biological tissues. *Nature Physics*, 11(12):1074–1079, 2015. DOI: 10.1038/nphys3471.
- Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. Transcriptional regulation by the numbers: models. *Current Opinion in Genetics & Development*, 15(2):116–124, 2005. DOI: 10.1016/j.gde.2005.02.007.
- Brandenberg, N., Hoehnel, S., Kuttler, F., Homicsko, K., Ceroni, C., Ringel, T., Gjorevski, N., Schwank, G., Coukos, G., Turcatti, G., and Lutolf, M.P. High-throughput automated organoid culture via stem-cell aggregation in microcavity arrays. *Nature Biomedical Engineering*, 4(9):863–874, 2020. DOI: 10.1038/s41551-020-0565-2.
- Brassard, J.A., Nikolaev, M., Hübscher, T., Hofer, M., and Lutolf, M.P. Recapitulating macro-scale tissue self-organization through organoid bioprinting. *Nature Materials*, 20(1):22–29, 2021. DOI: 10.1038/s41563-020-00803-5.
- Brodland, G.W. The Differential Interfacial Tension Hypothesis (DITH): A Comprehensive Theory for the Self-Rearrangement of Embryonic Cells and Tissues. *Journal of Biomechanical Engineering*, 124(2):188–197, 2002. DOI: 10.1115/1.1449491.
- Brodland, G.W. Computational modeling of cell sorting, tissue engulfment, and related phenomena: A review. *Applied Mechanics Reviews*, 57(1):47–76, 2004. DOI: 10.1115/1.1583758.

- Bussard, K.M., Mutkus, L., Stumpf, K., Gomez-Manzano, C., and Marini, F.C. Tumor-associated stromal cells as key contributors to the tumor microenvironment. *Breast Cancer Research*, 18(1):84, 2016. DOI: 10.1186/s13058-016-0740-2.
- Cabral, K.A., Patterson, D.M., Scheideler, O.J., Cole, R., Abate, A.R., Schaffer, D.V., Sohn, L.L., and Gartner, Z.J. Simple, Affordable, and Modular Patterning of Cells using DNA. *JoVE (Journal of Visualized Experiments)*, (168):e61937, 2021. DOI: 10.3791/61937.
- Calderwood, D. Talin controls integrin activation. *Biochemical Society Transactions*, 32(3):434–437, 2004. DOI: 10.1042/bst0320434.
- Camelo, C. and Luschnig, S. Cells into tubes: Molecular and physical principles underlying lumen formation in tubular organs. In: *Current Topics in Developmental Biology*. Academic Press, 2020. DOI: 10.1016/bs.ctdb.2020.09.002.
- Caves, L.S.D., Evanseck, J.D., and Karplus, M. Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Science*, 7(3):649–666, 1998. DOI: <https://doi.org/10.1002/pro.5560070314>.
- Centonze, A., Lin, S., Tika, E., Sifrim, A., Fioramonti, M., Malfait, M., Song, Y., Wuidart, A., Van Herck, J., Dannau, A., Bouvencourt, G., Dubois, C., Dedoncker, N., Sahay, A., de Maertelaer, V., Siebel, C.W., Van Keymeulen, A., Voet, T., and Blanpain, C. Heterotypic cell–cell communication regulates glandular stem cell multipotency. *Nature*, 584(7822):608–613, 2020. DOI: 10.1038/s41586-020-2632-y.
- Cerchiari, A.E., Garbe, J.C., Jee, N.Y., Todhunter, M.E., Broaders, K.E., Peehl, D.M., Desai, T.A., LaBarge, M.A., Thomson, M., and Gartner, Z.J. A strategy for tissue self-organization that is robust to cellular heterogeneity and plasticity. *Proceedings of the National Academy of Sciences*, 112(7):2287–2292, 2015. DOI: 10.1073/pnas.1410776112.

- Chang, A.Y. and Marshall, W.F. Dynamics of living cells in a cytomorphological state space. *Proceedings of the National Academy of Sciences*, 116(43):21556–21562, 2019. DOI: 10.1073/pnas.1902849116.
- Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*, 453(7194):544–547, 2008. DOI: 10.1038/nature06965.
- Chang, J. and Chaudhuri, O. Beyond proteases: Basement membrane mechanics and cancer invasion. *Journal of Cell Biology*, 218(8):2456–2469, 2019. DOI: 10.1083/jcb.201903066.
- Chang, T.T., Thakar, D., and Weaver, V.M. Force-dependent breaching of the basement membrane. *Matrix Biology*, 57-58:178–189, 2017. DOI: 10.1016/j.matbio.2016.12.005.
- Chanson, L., Brownfield, D., Garbe, J.C., Kuhn, I., Stampfer, M.R., Bissell, M.J., and LaBarge, M.A. Self-organization is a dynamic and lineage-intrinsic property of mammary epithelial cells. *Proceedings of the National Academy of Sciences*, 108(8):3264–3269, 2011. DOI: 10.1073/pnas.1019556108.
- Chau Nguyen, H., Zecchina, R., and Berg, J. Inverse statistical problems: from the inverse Ising problem to data science. *Advances in Physics*, 66(3):197–261, 2017. DOI: 10.1080/00018732.2017.1341604.
- Chen, H.H. and Brodland, G.W. Cell-Level Finite Element Studies of Viscous Cells in Planar Aggregates. *Journal of Biomechanical Engineering*, 122(4):394–401, 2000. DOI: 10.1115/1.1286563.
- Chen, S., Lewallen, M., and Xie, T. Adhesion in the stem cell niche: biological roles and regulation. *Development*, 140(2):255–265, 2013. DOI: 10.1242/dev.083139.
- Choi, Y., Lee, H.J., Jang, M.H., Gwak, J.M., Lee, K.S., Kim, E.J., Kim, H.J., Lee, H.E., and Park, S.Y. Epithelial-mesenchymal transition increases during the progression of



- in situ to invasive basal-like breast cancer. *Human Pathology*, 44(11):2581–2589, 2013. DOI: 10.1016/j.humpath.2013.07.003.
- Ciliberto, S. Experiments in Stochastic Thermodynamics: Short History and Perspectives. *Physical Review X*, 7(2):021051, 2017. DOI: 10.1103/PhysRevX.7.021051.
- Cocco, E., Lopez, S., Santin, A.D., and Scaltriti, M. Prevalence and role of HER2 mutations in cancer. *Pharmacology & Therapeutics*, 199:188–196, 2019. DOI: 10.1016/j.pharmthera.2019.03.010.
- Coveney, P.V. and Wan, S. On the calculation of equilibrium thermodynamic properties from molecular dynamics. *Physical Chemistry Chemical Physics*, 18(44):30236–30240, 2016. DOI: 10.1039/C6CP02349E.
- Cowell, C.F., Weigelt, B., Sakr, R.A., Ng, C.K., Hicks, J., King, T.A., and Reis-Filho, J.S. Progression from ductal carcinoma *in situ* to invasive breast cancer: Revisited. *Molecular Oncology*, 7(5):859–869, 2013. DOI: 10.1016/j.molonc.2013.07.005.
- Cowin, P., Rowlands, T.M., and Hatsell, S.J. Cadherins and catenins in breast cancer. *Current Opinion in Cell Biology*, 17(5):499–508, 2005. DOI: 10.1016/j.ceb.2005.08.014.
- Cugliandolo, L.F. The effective temperature. *Journal of Physics A: Mathematical and Theoretical*, 44(48):483001, 2011. DOI: 10.1088/1751-8113/44/48/483001.
- Dolega, M.E., Abeille, F., Picollet-D’hahan, N., and Gidrol, X. Controlled 3D culture in Matrigel microbeads to analyze clonal acinar development. *Biomaterials*, 52:347–357, 2015. DOI: 10.1016/j.biomaterials.2015.02.042.
- Eccles, S.A. The Role of c-erbB-2/HER2/neu in Breast Cancer Progression and Metastasis. *Journal of Mammary Gland Biology and Neoplasia*, 6(4):393–406, 2001. DOI: 10.1023/A:1014730829872.

- Egolf, D.A. Equilibrium Regained: From Nonequilibrium Chaos to Statistical Mechanics. *Science*, 287(5450):101–104, 2000. DOI: 10.1126/science.287.5450.101.
- Fagotto, F., Rohani, N., Touret, A.S., and Li, R. A Molecular Base for Cell Sorting at Embryonic Boundaries: Contact Inhibition of Cadherin Adhesion by Ephrin/Eph-Dependent Contractility. *Developmental Cell*, 27(1):72–87, 2013. DOI: 10.1016/j.devcel.2013.09.004.
- Felipe Lima, J., Nofech-Mozes, S., Bayani, J., and Bartlett, J.M.S. EMT in Breast Carcinoma—A Review. *Journal of Clinical Medicine*, 5(7):65, 2016. DOI: 10.3390/jcm5070065.
- Ferguson, A.L., Mann, J.K., Omarjee, S., Ndung’u, T., Walker, B.D., and Chakraborty, A.K. Translating HIV Sequences into Quantitative Fitness Landscapes Predicts Viral Vulnerabilities for Rational Immunogen Design. *Immunity*, 38(3):606–617, 2013. DOI: 10.1016/j.immuni.2012.11.022.
- Fleener, E., Janosi, L., Barz, B., Neagu, A., Forgacs, G., and Kosztin, I. Kinetic Monte Carlo and cellular particle dynamics simulations of multicellular systems. *Physical Review E*, 85(3):031907, 2012. DOI: 10.1103/PhysRevE.85.031907.
- Fletcher, A., Osterfield, M., Baker, R., and Shvartsman, S. Vertex Models of Epithelial Morphogenesis. *Biophysical Journal*, 106(11):2291–2304, 2014. DOI: 10.1016/j.bpj.2013.11.4498.
- Frauenfelder, H., Sligar, S.G., and Wolynes, P.G. The energy landscapes and motions of proteins. *Science (New York, N.Y.)*, 254(5038):1598–1603, 1991. DOI: 10.1126/science.1749933.
- Frigg, R. A field guide to recent work on the foundations of statistical mechanics. In: D. Rickles (ed.), *The Ashgate Companion to Contemporary Philosophy of Physics*, 99–196. Ashgate Publishing, London, UK, 2008.

- Fry, E.A., Taneja, P., and Inoue, K. Oncogenic and tumor-suppressive mouse models for breast cancer engaging HER2/neu: Mouse models for BC engaging HER2/neu. *International Journal of Cancer*, 140(3):495–503, 2017. DOI: 10.1002/ijc.30399.
- Gnesotto, F.S., Mura, F., Gladrow, J., and Broedersz, C.P. Broken detailed balance and non-equilibrium dynamics in living systems: a review. *Reports on Progress in Physics*, 81(6):066601, 2018. DOI: 10.1088/1361-6633/aab3ed.
- Graner, F. and Glazier, J.A. Simulation of biological cell sorting using a two-dimensional extended Potts model. *Physical Review Letters*, 69(13):2013–2016, 1992. DOI: 10.1103/PhysRevLett.69.2013.
- Groen, E.J., Elshof, L.E., Visser, L.L., Rutgers, E.J.T., Winter-Warnars, H.A., Lips, E.H., and Wesseling, J. Finding the balance between over- and under-treatment of ductal carcinoma in situ (DCIS). *The Breast*, 31:274–283, 2017. DOI: 10.1016/j.breast.2016.09.001.
- Haigo, S.L. and Bilder, D. Global Tissue Revolutions in a Morphogenetic Movement Controlling Elongation. *Science*, 331(6020):1071–1074, 2011. DOI: 10.1126/science.1199424.
- Hamada, H., Watanabe, M., Lau, H.E., Nishida, T., Hasegawa, T., Parichy, D.M., and Kondo, S. Involvement of Delta/Notch signaling in zebrafish adult pigment stripe patterning. *Development*, 141:8, 2014. DOI: 10.1242/dev.099804.
- Harris, A.K. Is Cell Sorting Caused by Differences in the Work of Intercellular Adhesion? A Critique of the Steinberg Hypothesis. *Journal of Theoretical Biology*, 61:267–285, 1976.
- Hayashi, K., Lopes, S.M.C.d.S., Tang, F., and Surani, M.A. Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States. *Cell Stem Cell*, 3(4):391–401, 2008. DOI: 10.1016/j.stem.2008.07.027.

- Hemmings, B.A. and Restuccia, D.F. PI3K-PKB/Akt Pathway. *Cold Spring Harbor Perspectives in Biology*, 4(9):a011189, 2012. DOI: 10.1101/cshperspect.a011189.
- Hilson, J.B., Schnitt, S.J., and Collins, L.C. Phenotypic Alterations in Ductal Carcinoma In Situ-associated Myoepithelial Cells. *Am J Surg Pathol*, 33(2):6, 2009.
- Hirai, H., Sootome, H., Nakatsuru, Y., Miyama, K., Taguchi, S., Tsujioka, K., Ueno, Y., Hatch, H., Majumder, P.K., Pan, B.S., and Kotani, H. MK-2206, an Allosteric Akt Inhibitor, Enhances Antitumor Efficacy by Standard Chemotherapeutic Agents or Molecular Targeted Drugs In vitro and In vivo. *Molecular Cancer Therapeutics*, 9(7):1956–1967, 2010. DOI: 10.1158/1535-7163.MCT-09-1012.
- Hu, J.L., Todhunter, M.E., LaBarge, M.A., and Gartner, Z.J. Opportunities for organoids as new models of aging. *The Journal of Cell Biology*, 217(1):39–50, 2018. DOI: 10.1083/jcb.201709054.
- Hu, M., Yao, J., Carroll, D.K., Weremowicz, S., Chen, H., Carrasco, D., Richardson, A., Violette, S., Nikolskaya, T., Nikolsky, Y., Bauerlein, E.L., Hahn, W.C., Gelman, R.S., Allred, C., Bissell, M.J., Schnitt, S., and Polyak, K. Regulation of In Situ to Invasive Breast Carcinoma Transition. *Cancer Cell*, 13(5):394–406, 2008. DOI: 10.1016/j.ccr.2008.03.007.
- Hutson, M.S., Brodland, G.W., Yang, J., and Viens, D. Cell Sorting in Three Dimensions: Topology, Fluctuations, and Fluidlike Instabilities. *Physical Review Letters*, 101(14):148105, 2008. DOI: 10.1103/PhysRevLett.101.148105.
- Iliina, O., Gritsenko, P.G., Syga, S., Lippoldt, J., La Porta, C.A.M., Chepizhko, O., Grosser, S., Vullings, M., Bakker, G.J., Starruß, J., Bult, P., Zapperi, S., Käs, J.A., Deutsch, A., and Friedl, P. Cell–cell adhesion and 3D matrix confinement determine jamming transitions in breast cancer invasion. *Nature Cell Biology*, 22(9):1103–1115, 2020. DOI: 10.1038/s41556-020-0552-6.

- Jack, R.L., Hagan, M.F., and Chandler, D. Fluctuation-dissipation ratios in the dynamics of self-assembly. *Physical Review E*, 76(2):021119, 2007. DOI: 10.1103/PhysRevE.76.021119.
- Jackson, C., Browell, D., Gautrey, H., and Tyson-Capper, A. Clinical Significance of HER-2 Splice Variants in Breast Cancer Progression and Drug Resistance. *International Journal of Cell Biology*, 2013:e973584, 2013. DOI: 10.1155/2013/973584.
- Janiš, V. Introduction to Mean-Field Theory of Spin Glass Models. *arXiv:1506.07128 [cond-mat]*, 2015.
- Jenkinson, G., Pujadas, E., Goutsias, J., and Feinberg, A.P. Potential energy landscapes identify the information-theoretic nature of the epigenome. *Nature Genetics*, 49(5):719–729, 2017. DOI: 10.1038/ng.3811.
- Jiang, S., Zhao, H., Zhang, W., Wang, J., Liu, Y., Cao, Y., Zheng, H., Hu, Z., Wang, S., Zhu, Y., Wang, W., Cui, S., Lobie, P.E., Huang, L., and Ma, S. An Automated Organoid Platform with Inter-organoid Homogeneity and Inter-patient Heterogeneity. *Cell Reports Medicine*, 1(9):100161, 2020. DOI: 10.1016/j.xcrm.2020.100161.
- Jin, S., MacLean, A.L., Peng, T., and Nie, Q. scEpath: energy landscape-based inference of transition probabilities and cellular trajectories from single-cell transcriptomic data. *Bioinformatics*, 34(12):2077–2086, 2018. DOI: 10.1093/bioinformatics/bty058.
- Junttila, T.T., Akita, R.W., Parsons, K., Fields, C., Lewis Phillips, G.D., Friedman, L.S., Sampath, D., and Sliwkowski, M.X. Ligand-Independent HER2/HER3/PI3K Complex Is Disrupted by Trastuzumab and Is Effectively Inhibited by the PI3K Inhibitor GDC-0941. *Cancer Cell*, 15(5):429–440, 2009. DOI: 10.1016/j.ccr.2009.03.020.
- Kakni, P., Hueber, R., Knoops, K., López-Iglesias, C., Truckenmüller, R., Habibovic, P., and Giselbrecht, S. Intestinal Organoid Culture in Polymer Film-Based Microwell Ar-

- rays. *Advanced Biosystems*, 4(10):2000126, 2020. DOI: <https://doi.org/10.1002/adbi.202000126>.
- Kalmar, T., Lim, C., Hayward, P., Muñoz-Descalzo, S., Nichols, J., Garcia-Ojalvo, J., and Arias, A.M. Regulated Fluctuations in Nanog Expression Mediate Cell Fate Decisions in Embryonic Stem Cells. *PLOS Biology*, 7(7):e1000149, 2009. DOI: [10.1371/journal.pbio.1000149](https://doi.org/10.1371/journal.pbio.1000149).
- Kindberg, A.A., Srivastava, V., Muncie, J.M., Weaver, V.M., Gartner, Z.J., and Bush, J.O. EPH/EPHRIN regulates cellular organization by actomyosin contractility effects on cell contacts. *Journal of Cell Biology*, 220(6):e202005216, 2021. DOI: [10.1083/jcb.202005216](https://doi.org/10.1083/jcb.202005216).
- Klinowska, T.C., Alexander, C.M., Georges-Labouesse, E., Van der Neut, R., Kreidberg, J.A., Jones, C.J., Sonnenberg, A., and Streuli, C.H. Epithelial Development and Differentiation in the Mammary Gland Is Not Dependent on  $\alpha 3$  or  $\alpha 6$  Integrin Subunits. *Developmental Biology*, 233(2):449–467, 2001. DOI: [10.1006/dbio.2001.0204](https://doi.org/10.1006/dbio.2001.0204).
- Koren, S., Reavie, L., Couto, J.P., De Silva, D., Stadler, M.B., Roloff, T., Britschgi, A., Eichlisberger, T., Kohler, H., Aina, O., Cardiff, R.D., and Bentires-Alj, M. PIK3CAH1047R induces multipotency and multi-lineage mammary tumours. *Nature*, 525(7567):114–118, 2015. DOI: [10.1038/nature14669](https://doi.org/10.1038/nature14669).
- Krens, S.F.G., Veldhuis, J.H., Barone, V., Čapek, D., Maître, J.L., Brodland, G.W., and Heisenberg, C.P. Interstitial fluid osmolarity modulates the action of differential tissue surface tension in progenitor cell segregation during gastrulation. *Development*, 144(10):1798–1806, 2017. DOI: [10.1242/dev.144964](https://doi.org/10.1242/dev.144964).
- Krieg, M., Arboleda-Estudillo, Y., Puech, P.H., Käfer, J., Graner, F., Müller, D.J., and Heisenberg, C.P. Tensile forces govern germ-layer organization in zebrafish. *Nature Cell Biology*, 10(4):429–436, 2008. DOI: [10.1038/ncb1705](https://doi.org/10.1038/ncb1705).

- Kubo, R. The fluctuation-dissipation theorem. *Reports on Progress in Physics*, 29:255–283, 1966. DOI: 10.1088/0034-4885/29/1/306.
- Kuerer, H.M. Ductal carcinoma *in situ* : treatment or active surveillance? *Expert Review of Anticancer Therapy*, 15(7):777–785, 2015. DOI: 10.1586/14737140.2015.1043897.
- Landsberg, K.P., Farhadifar, R., Ranft, J., Umetsu, D., Widmann, T.J., Bittig, T., Said, A., Jülicher, F., and Dahmann, C. Increased Cell Bond Tension Governs Cell Sorting at the *Drosophila* Anteroposterior Compartment Boundary. *Current Biology*, 19(22):1950–1955, 2009. DOI: 10.1016/j.cub.2009.10.021.
- Lang, A.H., Li, H., Collins, J.J., and Mehta, P. Epigenetic Landscapes Explain Partially Reprogrammed Cells and Identify Key Reprogramming Genes. *PLOS Computational Biology*, 10(8):e1003734, 2014. DOI: 10.1371/journal.pcbi.1003734.
- Lawlor, K.T., Vanslambrouck, J.M., Higgins, J.W., Chambon, A., Bishard, K., Arndt, D., Er, P.X., Wilson, S.B., Howden, S.E., Tan, K.S., Li, F., Hale, L.J., Shepherd, B., Pentoney, S., Presnell, S.C., Chen, A.E., and Little, M.H. Cellular extrusion bioprinting improves kidney organoid reproducibility and conformation. *Nature Materials*, 20(2):260–271, 2021. DOI: 10.1038/s41563-020-00853-9.
- Li, C. and Wang, J. Quantifying Cell Fate Decisions for Differentiation and Reprogramming of a Human Stem Cell Network: Landscape and Biological Paths. *PLOS Computational Biology*, 9(8):e1003165, 2013. DOI: 10.1371/journal.pcbi.1003165.
- Li, J., Horowitz, J.M., Gingrich, T.R., and Fakhri, N. Quantifying dissipation using fluctuating currents. *Nature Communications*, 10(1):1666, 2019. DOI: 10.1038/s41467-019-09631-x.
- Lien, E.C., Dibble, C.C., and Toker, A. PI3K signaling in cancer: beyond AKT. *Current Opinion in Cell Biology*, 45:62–71, 2017. DOI: 10.1016/j.ceb.2017.02.007.

- Loi, D., Mossa, S., and Cugliandolo, L.F. Effective temperature of active matter. *Physical Review E*, 77(5):051111, 2008. DOI: 10.1103/PhysRevE.77.051111.
- Martin, P., Hudspeth, A.J., and Jülicher, F. Comparison of a hair bundle's spontaneous oscillations with its response to mechanical stimulation reveals the underlying active process. *Proceedings of the National Academy of Sciences*, 98(25):14380–14385, 2001. DOI: 10.1073/pnas.251530598.
- Martínez, I.A., Bisker, G., Horowitz, J.M., and Parrondo, J.M.R. Inferring broken detailed balance in the absence of observable currents. *Nature Communications*, 10(1):3542, 2019. DOI: 10.1038/s41467-019-11051-w.
- Maître, J.L., Berthoumieux, H., Krens, S.F.G., Salbreux, G., Jülicher, F., Paluch, E., and Heisenberg, C.P. Adhesion Functions in Cell Sorting by Mechanically Coupling the Cortices of Adhering Cells. *Science*, 338(6104):253–256, 2012. DOI: 10.1126/science.1225399.
- Milo, R. and Phillips, R. *Cell Biology by the numbers*. Garland Science, 1 edition, 2015. DOI: 10.1201/9780429258770.
- Moore, C.C. Ergodic theorem, ergodic theory, and statistical mechanics. *Proceedings of the National Academy of Sciences*, 112(7):1907–1911, 2015. DOI: 10.1073/pnas.1421798112.
- Mori, H., Gjorevski, N., Inman, J.L., Bissell, M.J., and Nelson, C.M. Self-organization of engineered epithelial tubules by differential cellular motility. *Proceedings of the National Academy of Sciences*, 106(35):14890–14895, 2009. DOI: 10.1073/pnas.0901269106.
- Morrissey, M.A. and Sherwood, D.R. An active role for basement membrane assembly and modification in tissue sculpting. *Journal of Cell Science*, 128(9):1661–1668, 2015. DOI: 10.1242/jcs.168021.



- Narod, S.A. Hormone replacement therapy and the risk of breast cancer. *Nature Reviews Clinical Oncology*, 8(11):669–676, 2011. DOI: 10.1038/nrclinonc.2011.110.
- Naylor, M.J., Li, N., Cheung, J., Lowe, E.T., Lambert, E., Marlow, R., Wang, P., Schatzmann, F., Wintermantel, T., Schüetz, G., Clarke, A.R., Mueller, U., Hynes, N.E., and Streuli, C.H. Ablation of  $\beta$ 1 integrin in mammary epithelium reveals a key role for integrin in glandular morphogenesis and differentiation. *The Journal of Cell Biology*, 171(4):717–728, 2005. DOI: 10.1083/jcb.200503144.
- Nelson, D.A. and Larsen, M. Heterotypic control of basement membrane dynamics during branching morphogenesis. *Developmental Biology*, 401(1):103–109, 2015. DOI: 10.1016/j.ydbio.2014.12.011.
- Neumann, N.M., Perrone, M.C., Veldhuis, J.H., Huebner, R.J., Zhan, H., Devreotes, P.N., Brodland, G.W., and Ewald, A.J. Coordination of Receptor Tyrosine Kinase Signaling and Interfacial Tension Dynamics Drives Radial Intercalation and Tube Elongation. *Developmental Cell*, 45(1):67–82.e6, 2018. DOI: 10.1016/j.devcel.2018.03.011.
- Nguyen-Ngoc, K.V. and Ewald, A. Mammary ductal elongation and myoepithelial migration are regulated by the composition of the extracellular matrix. *Journal of Microscopy*, 251(3):212–223, 2013. DOI: 10.1111/jmi.12017.
- Novaro, V. Collagen-IV and laminin-1 regulate estrogen receptor expression and function in mouse mammary epithelial cells. *Journal of Cell Science*, 116(14):2975–2986, 2003. DOI: 10.1242/jcs.00523.
- Onega, T., Weaver, D.L., Frederick, P.D., Allison, K.H., Tosteson, A.N., Carney, P.A., Geller, B.M., Longton, G.M., Nelson, H.D., Oster, N.V., Pepe, M.S., and Elmore, J.G. The diagnostic challenge of low-grade ductal carcinoma in situ. *European Journal of Cancer*, 80:39–47, 2017. DOI: 10.1016/j.ejca.2017.04.013.

- Oono, Y. and Paniconi, M. Steady State Thermodynamics. *Progress of Theoretical Physics Supplement*, 130:29–44, 1998. DOI: 10.1143/PTPS.130.29.
- Ottone, C., Krusche, B., Whitby, A., Clements, M., Quadrato, G., Pitulescu, M.E., Adams, R.H., and Parrinello, S. Direct cell–cell contact with the vascular niche maintains quiescent neural stem cells. *Nature Cell Biology*, 16(11):1045–1056, 2014. DOI: 10.1038/ncb3045.
- Park, J.A., Kim, J.H., Bi, D., Mitchel, J.A., Qazvini, N.T., Tantisira, K., Park, C.Y., McGill, M., Kim, S.H., Gweon, B., Notbohm, J., Steward Jr, R., Burger, S., Randell, S.H., Kho, A.T., Tambe, D.T., Hardin, C., Shore, S.A., Israel, E., Weitz, D.A., Tschumperlin, D.J., Henske, E.P., Weiss, S.T., Manning, M.L., Butler, J.P., Drazen, J.M., and Fredberg, J.J. Unjamming and cell shape in the asthmatic airway epithelium. *Nature Materials*, 14(10):1040–1048, 2015. DOI: 10.1038/nmat4357.
- Parmar, H. and Cunha, G.R. Epithelial-stromal interactions in the mouse and human mammary gland in vivo. *Endocrine-Related Cancer*, 11(3):437–458, 2004. DOI: 10.1677/erc.1.00659.
- Parrondo, J.M.R., Broeck, C.V.d., and Kawai, R. Entropy production and the arrow of time. *New Journal of Physics*, 11(7):073008, 2009. DOI: 10.1088/1367-2630/11/7/073008.
- Paré, A.C., Vichas, A., Fincher, C.T., Mirman, Z., Farrell, D.L., Mainieri, A., and Zallen, J.A. A positional Toll receptor code directs convergent extension in *Drosophila*. *Nature*, 515(7528):523–527, 2014. DOI: 10.1038/nature13953.
- Phillips, R., Belliveau, N.M., Chure, G., Garcia, H.G., Razo-Mejia, M., and Scholes, C. Figure 1 Theory Meets Figure 2 Experiments in the Study of Gene Expression. *Annual Review of Biophysics*, 48(1):121–163, 2019. DOI: 10.1146/annurev-biophys-052118-115525.
- Phipson, B., Er, P.X., Combes, A.N., Forbes, T.A., Howden, S.E., Zappia, L., Yen, H.J., Lawlor, K.T., Hale, L.J., Sun, J., Wolvetang, E., Takasato, M., Oshlack, A., and Little,

- M.H. Evaluation of variability in human kidney organoids. *Nature Methods*, 16(1):79–87, 2019. DOI: 10.1038/s41592-018-0253-2.
- Pinho, S. and Frenette, P.S. Haematopoietic stem cell activity and interactions with the niche. *Nature Reviews Molecular Cell Biology*, 20(5):303–320, 2019. DOI: 10.1038/s41580-019-0103-9.
- Quadrato, G., Nguyen, T., Macosko, E.Z., Sherwood, J.L., Min Yang, S., Berger, D.R., Maria, N., Scholvin, J., Goldman, M., Kinney, J.P., Boyden, E.S., Lichtman, J.W., Williams, Z.M., McCarroll, S.A., and Arlotta, P. Cell diversity and network dynamics in photosensitive human brain organoids. *Nature*, 545(7652):48–53, 2017. DOI: 10.1038/nature22047.
- Riching, K., Cox, B.L., Salick, M., Pehlke, C., Riching, A., Ponik, S.M., Bass, B., Crone, W., Jiang, Y., Weaver, A.M., Eliceiri, K., and Keely, P. 3D Collagen Alignment Limits Protrusions to Enhance Breast Cancer Cell Persistence. *Biophysical Journal*, 107(11):2546–2558, 2014. DOI: 10.1016/j.bpj.2014.10.035.
- Ritort, F. Nonequilibrium Fluctuations in Small Systems: From Physics to Biology. In: *Advances in Chemical Physics*, 31–123. John Wiley & Sons, Ltd, 2007. DOI: 10.1002/9780470238080.ch2.
- Ritsma, L., Ellenbroek, S.I.J., Zomer, A., Snippert, H.J., de Sauvage, F.J., Simons, B.D., Clevers, H., and van Rheenen, J. Intestinal crypt homeostasis revealed at single-stem-cell level by in vivo live imaging. *Nature*, 507(7492):362–365, 2014. DOI: 10.1038/nature12972.
- Rompolas, P., Mesa, K.R., and Greco, V. Spatial organization within a niche as a determinant of stem-cell fate. *Nature*, 502(7472):513–518, 2013. DOI: 10.1038/nature12602.

- Ruiz-Saenz, A., Dreyer, C., Campbell, M.R., Steri, V., Gulizia, N.P., and Moasser, M.M. HER2 amplification in tumors activates PI3K/Akt signaling independent of HER3. *Cancer Research*, canres.0430.2018, 2018. DOI: 10.1158/0008-5472.CAN-18-0430.
- Russell, T.D., Jindal, S., Agunbiade, S., Gao, D., Troxell, M., Borges, V.F., and Schedin, P. Myoepithelial Cell Differentiation Markers in Ductal Carcinoma in Situ Progression. *The American Journal of Pathology*, 185(11):3076–3089, 2015. DOI: 10.1016/j.ajpath.2015.07.004.
- Sanders, M.E., Schuyler, P.A., Dupont, W.D., and Page, D.L. The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. *Cancer*, 103(12):2481–2484, 2005. DOI: <https://doi.org/10.1002/cncr.21069>.
- Schedin, P. Pregnancy-associated breast cancer and metastasis. *Nature Reviews Cancer*, 6(4):281–291, 2006. DOI: 10.1038/nrc1839.
- Seifert, U. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics*, 75(12):126001, 2012. DOI: 10.1088/0034-4885/75/12/126001.
- Serra, D., Mayr, U., Boni, A., Lukonin, I., Rempfler, M., Challet Meylan, L., Stadler, M.B., Strnad, P., Papasaikas, P., Vischi, D., Waldt, A., Roma, G., and Liberali, P. Self-organization and symmetry breaking in intestinal organoid development. *Nature*, 569(7754):66–72, 2019. DOI: 10.1038/s41586-019-1146-y.
- Shaffer, P., Valsson, O., and Parrinello, M. Enhanced, targeted sampling of high-dimensional free-energy landscapes using variationally enhanced sampling, with an application to chignolin. *Proceedings of the National Academy of Sciences*, 113(5):1150–1155, 2016. DOI: 10.1073/pnas.1519712113.

Shah, K., Bedi, R., Rogozhnikov, A., Ramkumar, P., Tong, Z., Rash, B., Stanton, M., Sorokin, J., Apaydin, C., Batarse, A., Bergamaschi, J., Blattner, R., Brown, S., Bosshardt, A., Castrillo, C., Dang, B., Drusinsky, S., Enriquez, L., Grayson, D., Hilliard, J., Hsu, P.K., Johnson, C., Jones, R., Lash, A., Lee, C.Y., Li, K., McKay, A., Mount, E., Nicola, J., Oumzil, I., Paek, J., Pascoe, D., Piepho, A., Poust, S., Quang, D., Schultz, M., Sims, J., Taylor, P., Treiman, G., Wueseke, O., Young, N., Pollen, A., Flanzer, D., Chao, D., Skibinski, G., Kato, S., and Escola, G.S. Optimization and scaling of patient-derived brain organoids uncovers deep phenotypes of disease. *bioRxiv*, 2020.08.26.251611, 2020. DOI: 10.1101/2020.08.26.251611.

Shekhar, K., Ruberman, C.F., Ferguson, A.L., Barton, J.P., Kardar, M., and Chakraborty, A.K. Spin models inferred from patient-derived viral sequence data faithfully describe HIV fitness landscapes. *Physical Review E*, 88(6):062705, 2013. DOI: 10.1103/PhysRevE.88.062705.

Shi, Q., Ghosh, R.P., Engelke, H., Rycroft, C.H., Cassereau, L., Sethian, J.A., Weaver, V.M., and Liphardt, J.T. Rapid disorganization of mechanically interacting systems of mammary acini. *Proceedings of the National Academy of Sciences*, 111(2):658–663, 2014. DOI: 10.1073/pnas.1311312110.

Shishvan, S.S., Vigliotti, A., and Deshpande, V.S. The homeostatic ensemble for cells. *Biomechanics and Modeling in Mechanobiology*, 17(6):1631–1662, 2018. DOI: 10.1007/s10237-018-1048-1.

Sirka, O.K., Shamir, E.R., and Ewald, A.J. Myoepithelial cells are a dynamic barrier to epithelial dissemination. *The Journal of Cell Biology*, 217(10):3368–3381, 2018. DOI: 10.1083/jcb.201802144.

- Skokan, T.D., Vale, R.D., and McKinley, K.L. Cell Sorting in *Hydra vulgaris* Arises from Differing Capacities for Epithelialization between Cell Types. *Current Biology*, 2020. DOI: 10.1016/j.cub.2020.07.035.
- Srivastava, V., Huycke, T.R., Phong, K.T., and Gartner, Z.J. Organoid models for mammary gland dynamics and breast cancer. *Current Opinion in Cell Biology*, 66:51–58, 2020. DOI: 10.1016/j.ceb.2020.05.003.
- Srivastava, V. and Robinson, D. Mechanical Stress and Network Structure Drive Protein Dynamics during Cytokinesis. *Current Biology*, 25(5):663–670, 2015. DOI: 10.1016/j.cub.2015.01.025.
- Stampfer, M.R., LaBarge, M.A., and Garbe, J.C. An Integrated Human Mammary Epithelial Cell Culture System for Studying Carcinogenesis and Aging. In: H. Schatten (ed.), *Cell and Molecular Biology of Breast Cancer*, 323–361. Humana Press, Totowa, NJ, 2013. DOI: 10.1007/978-1-62703-634-4\_15.
- Steinberg, M.S. and Gilbert, S.F. Townes and Holtfreter (1955): Directed movements and selective adhesion of embryonic amphibian cells. *Journal of Experimental Zoology*, 301A(9):701–706, 2004. DOI: 10.1002/jez.a.114.
- Stenhammar, J., Wittkowski, R., Marenduzzo, D., and Cates, M.E. Activity-Induced Phase Separation and Self-Assembly in Mixtures of Active and Passive Particles. *Physical Review Letters*, 114(1):018301, 2015. DOI: 10.1103/PhysRevLett.114.018301.
- Tadokoro, S. Talin Binding to Integrin Tails: A Final Common Step in Integrin Activation. *Science*, 302(5642):103–106, 2003. DOI: 10.1126/science.1086652.
- Tanner, K., Mori, H., Mroue, R., Bruni-Cardoso, A., and Bissell, M.J. Coherent angular motion in the establishment of multicellular architecture of glandular tissues. *Proceedings of the National Academy of Sciences*, 109(6):1973–1978, 2012. DOI: 10.1073/pnas.1119578109.

- Taylor, H.B., Khuong, A., Wu, Z., Xu, Q., Morley, R., Gregory, L., Poliakov, A., Taylor, W.R., and Wilkinson, D.G. Cell segregation and border sharpening by Eph receptor–ephrin-mediated heterotypic repulsion. *Journal of The Royal Society Interface*, 14(132):20170338, 2017. DOI: 10.1098/rsif.2017.0338.
- Teschendorff, A.E. and Feinberg, A.P. Statistical mechanics meets single-cell biology. *Nature Reviews Genetics*, 1–18, 2021. DOI: 10.1038/s41576-021-00341-z.
- Toda, S., Blauch, L.R., Tang, S.K.Y., Morsut, L., and Lim, W.A. Programming self-organizing multicellular structures with synthetic cell-cell signaling. *Science*, eaat0271, 2018. DOI: 10.1126/science.aat0271.
- Todhunter, M.E., Miyano, M., Moolamalla, D.S., Filippov, A., Sayaman, R.W., and LaBarge, M.A. Volume-constrained microcontainers enable myoepithelial functional differentiation in highly parallel mammary organoid culture. *iScience*, 24(4):102253, 2021. DOI: 10.1016/j.isci.2021.102253.
- Townes, P.L. and Holtfreter, J. Directed movements and selective adhesion of embryonic amphibian cells. *Journal of Experimental Zoology*, 128(1):53–120, 1955. DOI: 10.1002/jez.1401280105.
- Tsai, T.Y.C., Sikora, M., Xia, P., Colak-Champollion, T., Knaut, H., Heisenberg, C.P., and Megason, S.G. An adhesion code ensures robust pattern formation during tissue morphogenesis. *Science*, 370(6512):113–116, 2020. DOI: 10.1126/science.aba6637.
- Turlier, H., Fedosov, D.A., Audoly, B., Auth, T., Gov, N.S., Sykes, C., Joanny, J.F., Gompper, G., and Betz, T. Equilibrium physics breakdown reveals the active nature of red blood cell flickering. *Nature Physics*, 12(5):513–519, 2016. DOI: 10.1038/nphys3621.
- Uffink, J. Compendium of the Foundations of Statistical Mechanics. In: J. Butterfield and J. Earman (eds.), *Philosophy of Physics*, Handbook of the Philosophy of Science, 923–1074. Elsevier, Amsterdam, 2007. DOI: 10.1016/B978-044451560-5/50012-9.

- Uffink, J. Boltzmann's Work in Statistical Physics. In: E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2017 edition, 2017.
- Ursini-Siegel, J., Schade, B., Cardiff, R.D., and Muller, W.J. Insights from transgenic mouse models of ERBB2-induced breast cancer. *Nature Reviews Cancer*, 7(5):389–397, 2007. DOI: 10.1038/nrc2127.
- Van Keymeulen, A., Lee, M.Y., Ousset, M., Brohée, S., Rorive, S., Giraddi, R.R., Wuidart, A., Bouvencourt, G., Dubois, C., Salmon, I., Sotiriou, C., Phillips, W.A., and Blanpain, C. Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature*, 525(7567):119–123, 2015. DOI: 10.1038/nature14665.
- Viens, D. A Three-dimensional Finite Element Model for the Mechanics of Cell-Cell Interactions. *Journal of Biomechanical Engineering*, 129(5):651, 2007. DOI: 10.1115/1.2768375.
- Vilar, J.M.G. and Leibler, S. DNA Looping and Physical Constraints on Transcription Regulation. *Journal of Molecular Biology*, 331(5):981–989, 2003. DOI: 10.1016/S0022-2836(03)00764-2.
- Volpi, C.C., Pietrantonio, F., Gloghini, A., Fucà, G., Giordano, S., Corso, S., Pruneri, G., Antista, M., Cremolini, C., Fasano, E., Saggio, S., Faraci, S., Di Bartolomeo, M., de Braud, F., Di Nicola, M., Tagliabue, E., Pupa, S.M., and Castagnoli, L. The landscape of d16HER2 splice variant expression across HER2-positive cancers. *Scientific Reports*, 9(1):3545, 2019. DOI: 10.1038/s41598-019-40310-5.
- Völkel, A.R. and Noolandi, J. Meanfield Approach to the Thermodynamics of Protein–Solvent Systems with Application to P53. *Biophysical Journal*, 80(3):1524–1537, 2001. DOI: 10.1016/S0006-3495(01)76125-5.



- Waddington, C.H. *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*. Routledge, London, 1 edition, 1957. DOI: 10.4324/9781315765471.
- Walker, M.R., Patel, K.K., and Stappenbeck, T.S. The stem cell niche. *The Journal of Pathology*, 217(2):169–180, 2009. DOI: <https://doi.org/10.1002/path.2474>.
- Wang, H., Lacoche, S., Huang, L., Xue, B., and Muthuswamy, S.K. Rotational motion during three-dimensional morphogenesis of mammary epithelial acini relates to laminin matrix assembly. *Proceedings of the National Academy of Sciences*, 110(1):163–168, 2013. DOI: 10.1073/pnas.1201141110.
- Wang, J., Xu, L., and Wang, E. Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations. *Proceedings of the National Academy of Sciences*, 105(34):12271–12276, 2008. DOI: 10.1073/pnas.0800579105.
- Weber, R.J., Cerchiari, A.E., Delannoy, L.S., Garbe, J.C., LaBarge, M.A., Desai, T.A., and Gartner, Z.J. Rapid Organoid Reconstitution by Chemical Micromolding. *ACS Biomaterials Science & Engineering*, 2(11):1851–1855, 2016. DOI: 10.1021/acsbomaterials.6b00421.
- Winklbauer, R. Cell adhesion strength from cortical tension – an integration of concepts. *Journal of Cell Science*, 128(20):3687–3693, 2015. DOI: 10.1242/jcs.174623.
- Xiao, K., Oas, R.G., Chiasson, C.M., and Kowalczyk, A.P. Role of p120-catenin in cadherin trafficking. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1773(1):8–16, 2007. DOI: 10.1016/j.bbamcr.2006.07.005.
- Yamada, S. and Nelson, W.J. Localized zones of Rho and Rac activities drive initiation and expansion of epithelial cell–cell adhesion. *The Journal of Cell Biology*, 178(3):517–527, 2007. DOI: 10.1083/jcb.200701058.

- Yanagida, A., Revell, C., Stirparo, G.G., Corujo-Simon, E., Aspalter, I.M., Peters, R., Belly, H.D., Cassani, D.A., Achouri, S., Blumenfeld, R., Franze, K., Paluch, E.K., Nichols, J., and Chalut, K. Cell surface fluctuations regulate early embryonic lineage sorting. preprint, *Developmental Biology*, 2020. DOI: 10.1101/2020.08.16.250084.
- Youssef, J., Nurse, A.K., Freund, L.B., and Morgan, J.R. Quantification of the forces driving self-assembly of three-dimensional microtissues. *Proceedings of the National Academy of Sciences*, 108(17):6993–6998, 2011. DOI: 10.1073/pnas.1102559108.
- Yurchenco, P. and Patton, B. Developmental and Pathogenic Mechanisms of Basement Membrane Assembly. *Current Pharmaceutical Design*, 15(12):1277–1294, 2009. DOI: 10.2174/138161209787846766.
- Zhao, L., Mok, S., and Moraes, C. Micropocket hydrogel devices for all-in-one formation, assembly, and analysis of aggregate-based tissues. *Biofabrication*, 11(4):045013, 2019. DOI: 10.1088/1758-5090/ab30b4.

## A. Selected interfacial tensions

Table A.1: Parameters for different cell combinations are composed of the five interfacial energies.  $L$  and  $M$  represent the two cell types, typically LEP and MEP. For MEP-only aggregates, the first named MEP type is considered  $L$ .  $X$  represents extracellular substrate, either Matrigel or agarose.

Parameters	$\gamma_{LL}$	$\gamma_{LM}$	$\gamma_{LX}$	$\gamma_{MM}$	$\gamma_{MX}$
GFP	-0.105	-0.107	-0.052	-0.131	-0.154
<i>PIK3CA</i> <sup>H1047R</sup>	-0.123	-0.114	-0.129	-0.131	-0.154
GFP agarose	-0.105	-0.107	0.000	-0.131	0.000
<i>PIK3CA</i> <sup>H1047R</sup> agarose	-0.123	-0.114	0.000	-0.131	0.000
MEP+MEP	-0.131	-0.131	-0.154	-0.131	-0.154
<i>TLN1</i> sh MEP+MEP	-0.131	-0.131	-0.045	-0.131	-0.154
<i>CTNND1</i> sh MEP+MEP	-0.090	-0.110	-0.123	-0.131	-0.154
<i>TLN1</i> sh MEP+MEP agarose	-0.131	-0.131	0.000	-0.131	0.000
<i>CTNND1</i> sh MEP+MEP agarose	-0.090	-0.110	0.000	-0.131	0.000
LEP+ <i>TLN1</i> sh MEP	-0.105	-0.107	-0.052	-0.131	-0.045
LEP+ <i>TLN1</i> sh MEP agarose	-0.105	-0.107	0.000	-0.131	0.000
GFP DMSO	-0.105	-0.107	-0.060	-0.131	-0.174

Parameters	$\gamma_{LL}$	$\gamma_{LM}$	$\gamma_{LX}$	$\gamma_{MM}$	$\gamma_{MX}$
GFP copanlisib	-0.105	-0.107	-0.045	-0.131	-0.147
GFP MK2206	-0.105	-0.107	-0.061	-0.131	-0.142
<i>PIK3CA</i> <sup>H1047R</sup> DMSO	-0.123	-0.114	-0.121	-0.131	-0.174
<i>PIK3CA</i> <sup>H1047R</sup> <i>TLN1</i> sh	-0.116	-0.114	-0.033	-0.131	-0.154
<i>PIK3CA</i> <sup>H1047R</sup> LEP+ <i>CTNND1</i> sh MEP	-0.123	-0.082	-0.129	-0.090	-0.123
<i>PIK3CA</i> <sup>H1047R</sup> copanlisib	-0.123	-0.114	-0.090	-0.131	-0.147
<i>PIK3CA</i> <sup>H1047R</sup> MK2206	-0.123	-0.114	-0.084	-0.131	-0.142
<i>PIK3CA</i> <sup>H1047R</sup> LEP+LEP	-0.123	-0.105	-0.129	-0.105	-0.052
<i>PIK3CA</i> <sup>H1047R</sup> MEP+MEP	-0.171	-0.141	-0.241	-0.131	-0.154

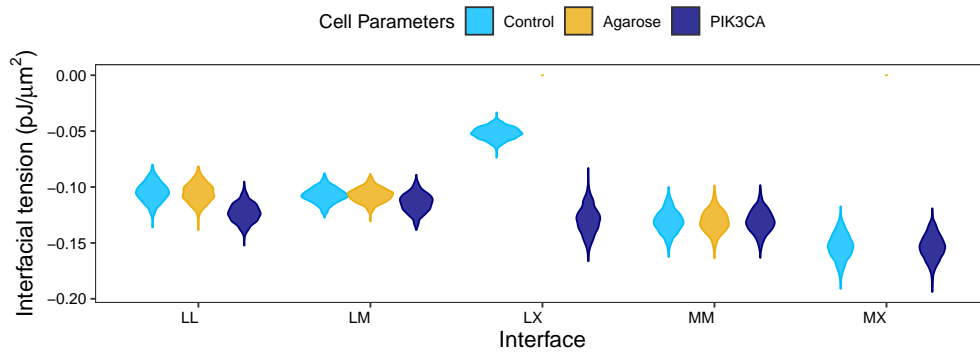


Figure A.1: Bootstrapping of 1000 random draws from each of 5 interfacial tensions, used for predicting tissue energies from simulated cell arrangements (section 2.10). Lower values are more energetically favorable. Cell-substrate (X) interactions with agarose are set to 0, assuming nonadherence.

## B. Predicted and inferred tissue energy gradients

Table B.1: Interfacial tension measurements were used to parameterize simulated tissues to predict  $dE$  (section 2.10). Observed organoid structure frequencies were used to infer  $\beta dE$  (section 2.11). A subset of predicted and inferred slopes with 95% CI are shown for cell combinations with at least 30 observed events. CI which span 0 may be considered to be predicted or inferred to be non-sorting. Agarose conditions are noted with (a).

Parameters	$dE$ (95% CI)	$\beta dE$ (95% CI)	n
GFP	2.864 (2.092, 3.649)	10.736 (9.244, 12.061)	503
<i>PIK3CA</i> <sup>H1047R</sup>	0.660 (-0.301, 1.666)	4.171 (2.376, 5.751)	414
GFP (a)	-0.413 (-0.765, -0.083)	-15.567 (-17.161, -13.776)	75
MEP+MEP	-0.007 (-1.004, 1.012)	-0.087 (-1.43, 1.149)	378
<i>TLN1</i> sh MEP+MEP	3.501 (2.574, 4.407)	10.219 (8.945, 11.413)	119
<i>CTNND1</i> sh MEP+MEP	0.341 (-0.771, 1.471)	-0.146 (-1.211, 0.934)	131
<i>TLN1</i> sh MEP+MEP (a)	-0.003 (-0.4, 0.372)	1.460 (-0.449, 3.04)	36
<i>CTNND1</i> sh MEP+MEP (a)	-0.659 (-1.08, -0.244)	-8.418 (-9.882, -7.06)	41
LEP+ <i>CTNND1</i> sh MEP	2.571 (1.566, 3.534)	2.826 (2.228, 3.392)	55
LEP+ <i>TLN1</i> sh MEP	-0.627 (-1.279, 0.047)	-3.997 (-4.701, -3.353)	40
GFP DMSO	3.257 (1.78, 4.806)	10.055 (8.434, 11.625)	109
<i>PIK3CA</i> <sup>H1047R</sup> DMSO	1.650 (-0.294, 3.412)	5.163 (3.516, 6.731)	202

Parameters	$dE$ (95% CI)	$\beta dE$ (95% CI)	n
$PIK3CA^{H1047R}$ TLNsh	3.691 (2.858, 4.513)	7.772 (6.122, 9.473)	57
$PIK3CA^{H1047R}$ copanlisib	1.708 (0.208, 3.285)	3.406 (1.648, 5.035)	66
$PIK3CA^{H1047R}$ MK2206	1.813 (0.606, 2.951)	8.873 (7.802, 9.946)	130

## C. Estimation of interfacial energies in tissues

Using cell doublets and individual cells to calculate the relative energies associated with cellular interfaces requires that we calculate the change in energy associated with forming each type of interface. In general terms,  $E = \sum_{\ell} \gamma_{\ell} A_{\ell}$ , so we need to determine  $\gamma_{\ell}$  and  $A_{\ell}$  for cells before and after adhesion. First, we use micropipette aspiration to measure the cell-intrinsic cortical tension  $\gamma_c$ , using established methods (Srivastava and Robinson, 2015). Then, we use the Young equation for droplet wetting to express the other interfacial tensions in terms of the steady-state contact angle of each interface,  $\theta_{cc}$  and  $\theta_{cx}$ , and the cortical tension  $\gamma_c$ , as previously shown (Cerchiari et al., 2015). These are diagrammed in **Fig. C.1**.

$$\gamma_c = \frac{\Delta P_{\text{crit}}}{2 \left( \frac{1}{R_p} + \frac{1}{R_c} \right)}$$

$$\gamma_{cc} = \gamma_c \cos \theta_{cc}$$

$$\gamma_{cx} = \gamma_c \cos \theta_{cx} + \gamma_x$$

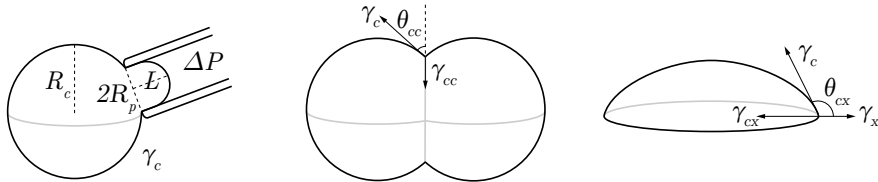


Figure C.1: Cell interfacial tensions. Cortical tension  $\gamma_c$  is calculated by micropipette aspiration. Cell-cell and cell-matrix interfacial tensions  $\gamma_{cc}$  and  $\gamma_{cx}$  are calculated by force balance of steady-state cell doublets or single adherent cells, respectively.  $\gamma_x$  is the substrate-medium tension.

Note that for a sufficiently large difference in the cortical tensions of two cells in contact with one another, the contact interface will be asymmetrical and the stiffer cell may

be partially engulfed by the other. In our assays, we do not see engulfment and cell cortical tensions are very close to each other, so we use the contact geometry shown.

To estimate the global tissue energy, or the total change in surface energies during the adhesion of the component cells to each other and the surrounding substrate, we will make the following assumptions<sup>1</sup>:

- Each type of cell interface can be assigned a characteristic interfacial tension or interfacial energy density. This characteristic tension emerges from the combined molecular mechanisms of cell contractility, cell adhesion molecule binding, cell membrane tension, and so on.
- Cell adhesion reaches a steady state within a few hours and the underlying interfacial tension does not change over the timescale of interest (hours to days).
- Cell interfacial tensions are unaffected by multiple cells interacting simultaneously.
- The total 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,  $E = \sum_i \gamma_i A_i$ .
- Cell volume is conserved during adhesion over the timescale of hours.

For simplicity, cell shapes are approximated by a bitruncated cubic honeycomb, a structure with one of the most efficient packings of equal-volume cells in a three-dimensional volume (**Fig. C.2**). While a less symmetrical, 0.3% more efficient solution exists, this lattice has the convenient property that all cells have the same shape (bitruncated cube or truncated octahedron) and the cells are evenly spaced in a body-centered cubic lattice.

---

<sup>1</sup>These assumptions resemble coarse graining or mean field approximations commonly used to reduce the degrees of freedom in spin glass Hamiltonians and protein structure prediction, namely the treatment of the microscopically complex cell interface as behaving with a single macroscopic mean tension, and the assumption that multiple contacts are independent (Völkel and Noolandi, 2001; Janiš, 2015).



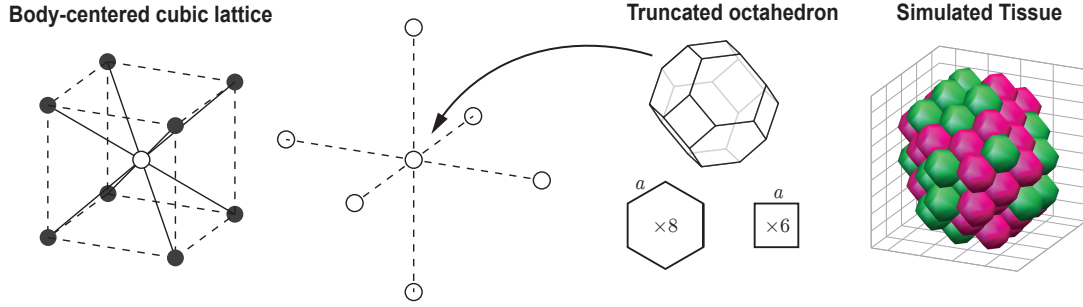


Figure C.2: The Voronoi tessellation of the base-centered cubic lattice is a highly efficient packing of cells into a volume. Each cell is a truncated octahedron with six square faces along the main axes and eight hexagonal faces along the diagonals, all with side length  $a$ . Cells are spaced  $a\sqrt{3}$  apart.

In order to assign an enthalpy value to a tissue configuration, we can consider the reference state to be the equivalent number and type of cells floating in suspension, making no contacts at all. Then the enthalpy of any given tissue is the change in surface energy associated with forming every cell-cell and cell-substrate interface. As we assume that contacts are independent, we assign every contact in the tissue an energy value  $\Delta E$  equal to the change in energy associated with forming a single such contact (**Fig. C.3**).

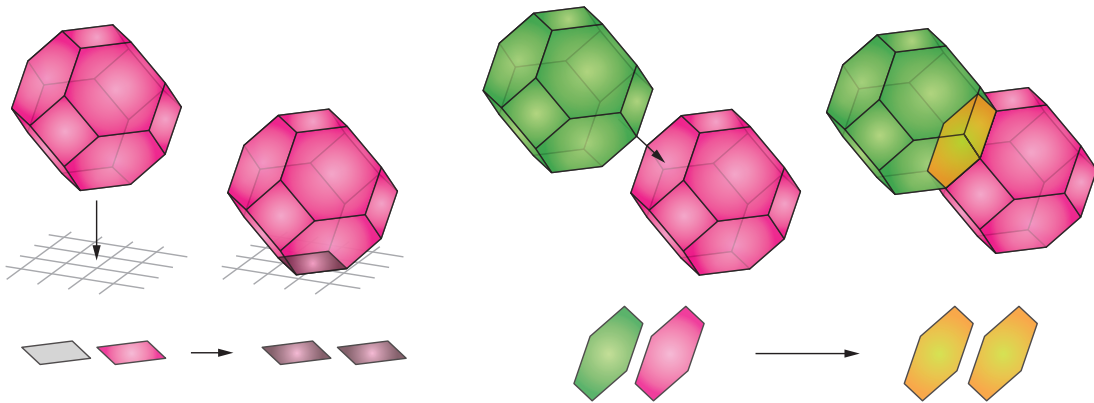


Figure C.3: The change in surface energy associated with forming a cell-matrix or a cell-cell interface is dependent on the area of the interface and the interfacial tension associated with it. The total surface energy of a tissue can be calculated as the sum of the  $\Delta E$  for all cell-cell and cell-matrix contacts.

Recall that  $\gamma_{cc} = \gamma_c \cos \theta_{cc}$  and  $\gamma_{cx} = \gamma_c \cos \theta_{cx} + \gamma_x$ . Then the energy change associated with the creation of a new interface,  $\Delta E = E_f - E_i$ , for either a cell-cell or cell-substrate interface, is:

$$\begin{aligned}
\Delta E_{cx} &= 2\gamma_{cx}A_{\text{interface}} - (\gamma_x A_{\text{interface}} + \gamma_c A_{\text{interface}}) \\
&= A_{\text{interface}} (2\gamma_{cx} - \gamma_x - \gamma_c) \\
&= A_{\text{interface}} (2\gamma_c \cos \theta_{cx} + 2\gamma_x - \gamma_x - \gamma_c) \\
&= \gamma_c A_{\text{interface}} (2 \cos \theta_{cx} - 1) + \gamma_x A_{\text{interface}} \\
\Delta E_{c_1 c_2} &= 2\gamma_{c_1 c_2} A_{\text{interface}} - (\gamma_{c_1} A_{\text{interface}} + \gamma_{c_2} A_{\text{interface}}) \\
&= A_{\text{interface}} (2\gamma_{c_1 c_2} - \gamma_{c_1} - \gamma_{c_2}) \\
&= A_{\text{interface}} (2\gamma_c \cos \theta_{c_1 c_2} - 2\gamma_c) \text{ for } \gamma_{c_1} = \gamma_{c_2} = \gamma_c \\
&= 2\gamma_c A_{\text{interface}} (\cos \theta_{c_1 c_2} - 1)
\end{aligned}$$

Therefore, the total energy of the tissue is

$$E = \sum_{\ell} \Delta E_{\ell} = \sum_{i,j} \Delta E_{c_i c_j} + \sum_k \Delta E_{c_k x}$$

where  $c_k x$  = each interface with the substrate

and  $c_i c_j$  = each interface between a unique pair of adhered cells

$$\begin{aligned}
&= \sum_{i,j} 2\gamma_c A_{c_i c_j} (\cos \theta_{c_i c_j} - 1) + \sum_k (\gamma_c A_{c_k x} (2 \cos \theta_{c_k x} - 1) + \gamma_x A_{c_k x}) \\
&= \sum_{i,j} 2\gamma_c A_{c_i c_j} (\cos \theta_{c_i c_j} - 1) + \sum_k \gamma_c A_{c_k x} (2 \cos \theta_{c_k x} - 1) + \gamma_x A_{\text{outer}}
\end{aligned}$$

We find a constant term,  $\gamma_x A_{\text{outer}}$ , which is the surface energy of the substrate alone, with no cells attached to it. Because this value is independent of tissue configuration, we can

consider it to be a reference energy  $E_0$  and redefine tissue energy as

$$\Delta E = E - E_0 = \sum_{i,j} 2\gamma_c A_{c_i c_j} (\cos \theta_{c_i c_j} - 1) + \sum_k \gamma_c A_{c_k x} (2 \cos \theta_{c_k x} - 1)$$

For the special case where the substrate is non-adherent (e.g., agarose), the final and initial states of cell-substrate surfaces are both cell-media interfaces. Since there is no change in interface, the energy change associated with cell-substrate interfaces  $\sum_k \Delta E_{c_k x} = 0$ . Tissue energy in suspension is simply calculated as

$$\Delta E = \sum_{i,j} 2\gamma_c A_{c_i c_j} (\cos \theta_{c_i c_j} - 1)$$

This approach leaves only terms that we can measure empirically using contact angle and micropipette aspiration assays ( $\theta_{c_k x}$ ,  $\theta_{c_i c_j}$ ,  $\gamma_c$ ), and the areas of each interface. These areas can be estimated for our cells, which typically have a radius of 10  $\mu\text{m}$ . Assuming that a cell does not change in volume when it takes the shape of a truncated octahedron,

$$\begin{aligned} V_{\text{sphere}} &= \frac{4\pi}{3} R^3 = V_{\text{trunc. octa.}} = 8\sqrt{2} a^3 \\ a^3 &= \frac{4\pi}{24\sqrt{2}} R^3 = \frac{\pi}{6\sqrt{2}} R^3 \\ A_{\text{hexagon}} &= \frac{3\sqrt{3}}{2} a^2 = \frac{3\sqrt{3}}{2} \left( \frac{\pi}{6\sqrt{2}} \right)^{\frac{2}{3}} R^2 \approx 134 \mu\text{m}^2 \\ A_{\text{square}} &= a^2 = \left( \frac{\pi}{6\sqrt{2}} \right)^{\frac{2}{3}} R^2 \approx 52 \mu\text{m}^2 \end{aligned}$$

## D. Combinatoric microstate enumeration in polyhedral lattice

The truncated octahedral honeycomb can also more closely approximate the expected multiplicity of microstates in a three-dimensional tissue. Rather than merely quantifying site occupancy in two categories, core and shell, cells can now occupy sites with varying amounts of their surface in contact with the substrate, the combination of the 14 hexagonal and square faces determined by the geometry of the tissue. In general, there will be  $m$  different possible areas that any given cell has exposed:  $A_1, A_2, \dots, A_m$ . Distinct lattice sites that have these areas can be grouped into sets with sizes  $S_1, S_2, \dots, S_m$ . If the number of LEP in each such group is  $L_i$ , the total LEP area in contact with the outside will be  $A_L = \sum_i^m A_i \cdot L_i$ . What are the possible values of  $A_L$ , and how many arrangements exist for each possible value of  $A_L$ ? We seek the generating function

$$G(x, y) = \sum_{A_L, L_{tot}} g_{A_L, L_{tot}} x^{A_L} y^{L_{tot}}$$

which takes the form

$$G(x, y) = \prod_{i=1}^m \left( \sum_{j=0}^{S_i} \binom{S_i}{j} x^{j \cdot A_i} y^j \right)$$

where the coefficient of  $x^{A_L} y^{L_{tot}}$  is the number of ways to arrange  $L_{tot}$  LEP into  $m$  groups such that the sum of their exposed surface areas is  $A_L$ .

The polyhedral lattices were generated and quantified in MATLAB, and the expansion of the polynomial was performed using SAGE after converting areas to integer values for efficient calculation in the integer polynomial ring. The resulting multiplicity data were plotted as histograms in **Fig. D.1**.

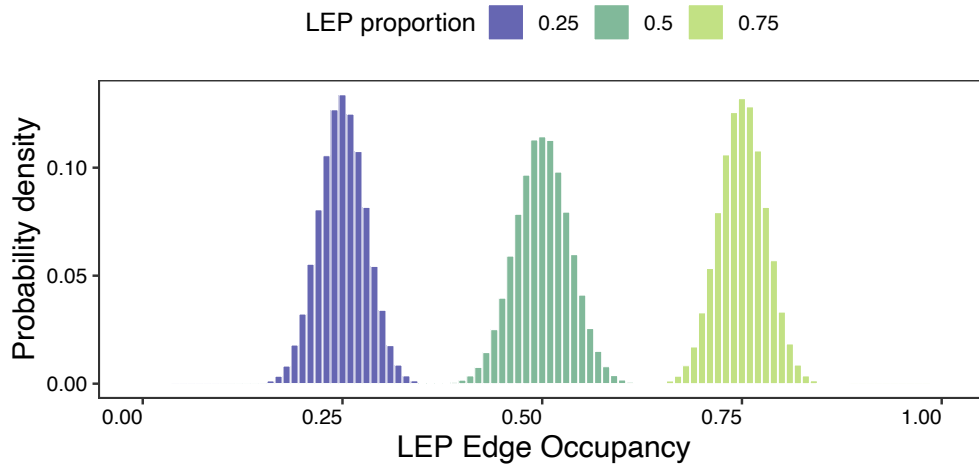


Figure D.1: Combinatorics of cell arrangements in a polyhedral lattice. The number of possible arrangements for a given edge occupancy can be calculated precisely given the geometry of the lattice sites, the size of the tissue, and the composition. Shown here are the probabilities of edge occupancies for 156-cell tissues with 25, 50, and 75% LEP total, demonstrating the predicted Gaussian-like distributions centered at LEP proportion.

## E. Cross-sectional sampling of organoid features

Due to limited imaging resolution and acquisition in the  $z$ -axis, it is impractical to quantify the full three-dimensional organoid. While our choice of the organoid's center is random in the sense that the organoid's orientation in the Matrigel or agarose microwell is arbitrary, it is naturally only a partial sampling of the cells. In this section, I use simulated tissues to characterize the error associated with using a weighted average of the three center slices in a  $z$ -stack to quantify the full organoid, specifically comparing the values calculated for Edge Occupancy, Intermixing Score, and predicted energy gradient.

Using a polyhedral lattice tissue with 156 total points (76 cells and 80 ECM), I made cross-sectional layers through the center planes by selecting all points with  $x = \lfloor R/2 \rfloor$ , and similarly for  $y$  and  $z$ . For any cross-section, only the fully enclosed interfaces (i.e., existing between two points both included in the cross-section) were analyzed. For a close-packed lattice, this is highly restrictive due to the loss of secondary neighbor connectivity, and each cell only makes 4 contacts rather than 14. A two-layer cross-section additionally includes the points with coordinate  $\lfloor R/2 \rfloor + 1$ . A three-layer cross-section also adds  $\lfloor R/2 \rfloor - 1$ , and, similarly, the four-layer cross-section contains all the points  $-1, 0, 1$  and  $2$  away from the central plane. These cross-sections contain 13, 26, 39 and 59 cells, respectively, out of a total of 76 cells in the entire tissue. The cells of first three cross-sections in the  $(x, y)$  plane are shown in **Fig. E.1**.

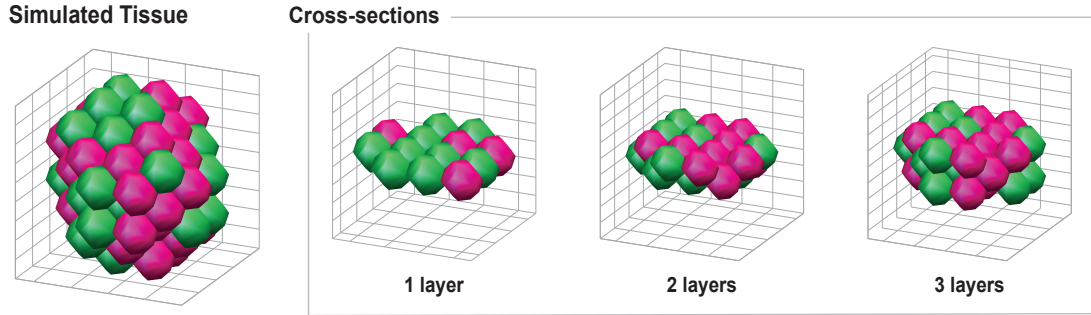


Figure E.1: Simulated organoid cross-sections for the tissue with radius 6. ECM points are not shown for clarity.

To populate the phase space, green and red cells fixed at a 1:1 ratio were randomly assigned to cell points in the tissue to generate a total of 1 000 000 random tissues. Edge Occupancy, Intermixing Score, and  $E$  were calculated for each cross-section and total tissue as usual. Similar to the quality control step for image processing, I excluded cross-sections with more than 60% of one cell type. For one-layered cross-sections, which contain only 14 cells total, this had a significant effect on the allowed structures, as only 6, 7, and 8 are legal numbers of any one cell type. Metrics calculated from all cross-section sizes were correlated with the corresponding tissue metrics but noticeably deviate from EO near high and low values, underestimating and overestimating, respectively, primarily due to missing the end caps of the tissue.

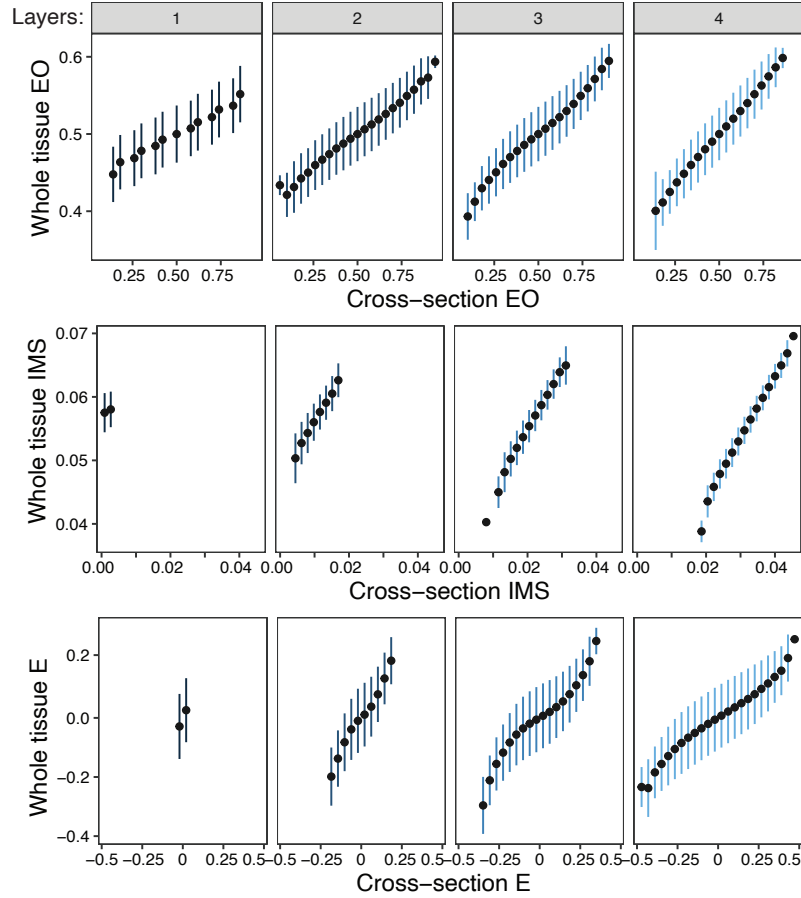
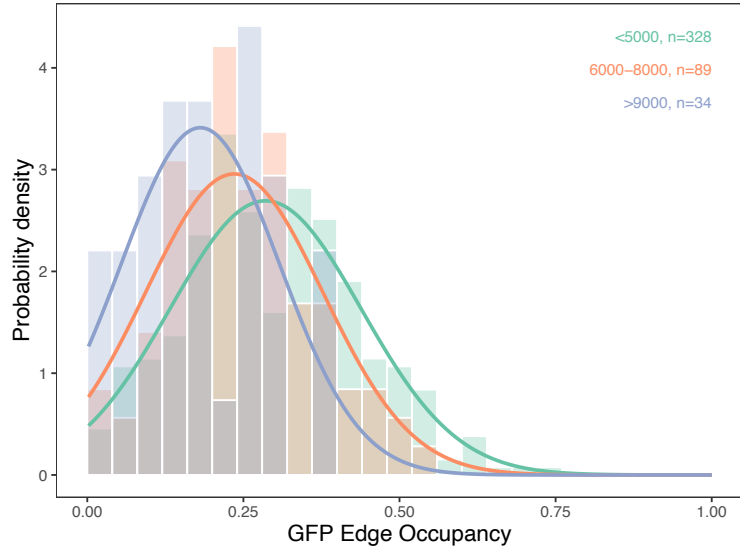


Figure E.2: Comparison between cross-section metrics and the corresponding full tissue metric for Edge Occupancy, Intermixing Score, and  $E$  calculated using control LEP and MEP parameters. Binned means  $\pm$  standard deviation are shown.

The effect of undersampling the total organoid should be related to the total size of the organoid, which suggests that this will be more pronounced in larger organoids. Using different size cutoffs for analysis of wild-type tissues demonstrates a noticeable trend of larger organoids having lower LEP Edge Occupancy (**Fig. E.3**).





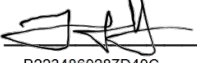
Area	$\beta dE_i$	lower	upper	n
<5000	8.053	6.684	9.265	328
6000–8000	13.371	12.244	14.426	89
>9000	18.536	16.883	20.202	34

Figure E.3: Comparison between GFP tissues subsetted to have small, medium, and large areas. The apparent energy gradient of large tissues is higher, potentially because the center slices of a very large organoid take up a smaller fraction of its total volume.

## Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:  
  
B2234860287D40C... Author Signature

6/1/2021  
Date