Winter 12-15-2016

Structure, Dynamics, and Regulation of Collective Cell Migration

Andrew J. Loza
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology & Biomedical Sciences
Computational and Molecular Biophysics

Dissertation Examination Committee:
Gregory D. Longmore, Chair
Elliot Elson
Guy Genin
Robert Mecham
Amit Pathak
Denis Wirtz

Structure, Dynamics, and Regulation of Collective Cell Migration

by
Andrew J. Loza

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2016
St. Louis, Missouri

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would first like to express my gratitude to my thesis mentor, Greg Longmore. His passion for scientific understanding, focus on quantitative biological approaches, and emphasis on the importance of a mechanistic understanding of systems have helped me to grow as a scientist. The Longmore Lab has been a great environment for research, and fun, over the last few years, and I would like to thank my fellow lab members for their input, humor, and support.

I would also like to thank my thesis committee for their guidance during my thesis proposal and subsequent updates. I would especially like to thank Elliot Elson for teaching the Mesoscopic Systems class one final time, which I was fortunate enough to take. The concepts covered in that class, along with his infectious scientific curiosity, provided the inspiration for the final chapter of this dissertation.

Thank you to all of our collaborators at Washington University in St. Louis and elsewhere who have given me the opportunity to explore new systems and fields. This work has broadened my scientific knowledge, and I am truly thankful for the chance to participate in these projects. I would also like to thank our collaborators at Johns Hopkins for their work and many discussions, without which the second chapter of this dissertation would not have been possible.

The Medical Scientist Training Program directors and administration have created a wonderful environment for physician-scientist training at Washington University, and I am thankful for the opportunity to be a part of this program. Thank you to Wayne Yokoyama, Brian Sullivan, Christy Durbin, Liz Bayer, and Linda Perniciaro for all of your hard work making this program outstanding.

I would like to thank my additional sources of funding through the Physical Sciences in Oncology Center grant and the Mechanobiology T32 grant. These programs have provided much inspiration and support for cross-disciplinary research that bridges the physical and biological sciences.

I would also like to express my gratitude to my family for their love and support. I would like to thank my parents for reinforcing my curiosity in how the world works from a young age, for being inspiring scientists themselves, and supporting me in all of my pursuits both in and out of science. I would like to thank my sister for encouraging me in science and being an important role model in hard work and dedication. I would also be remiss not to acknowledge Emily's and my dog, Sophia, because in the words of George Eliot, "Animals are such agreeable friends - they ask no questions, they pass no criticisms."

Most importantly, I would like to thank my wife, Emily Olfson, for her love, encouragement, and support. Her passion for medical research is inspiring, and her focus on helping others in research and in her friendships shows her true compassion. I am looking forward to the next steps in our lives.

Andrew Loza

Washington University in St. Louis

December 2016

ABSTRACT OF THE DISSERTATION

Structure, Dynamics, and Regulation of Collective Cell Migration

by

Andrew J Loza

Doctor of Philosophy in Biological and Biomedical Sciences

Computational and Molecular Biophysics

Washington University in St. Louis, 2016

Professor Greg Longmore, Chair

Collective migration is the process by which cells organize individual motions to productively migrate as a group and plays a fundamental role in organism development, tissue regeneration, and cancer invasion. In development, coordinated migration facilitates the formation of complex organ structures and is required for proper dissemination of neural crest cells throughout an organism. After injury, this process allows breaches in epithelial layers to be repaired while maintaining tissue integrity, and in cancer, collective behavior enhances invasion of tumor cells into the surrounding tissue. Chapter 1 provides an introduction for the role of collective migration across an organism's lifespan, the mechanisms used by cells to generate motile force, and the emergence of collective behavior. Chapter 2 dissects the intertwined roles of three fundamental parameters often altered in collective migration processes: cell density, cell adhesion, and cell-cell contractility through the Rho-ROCK-Myosin II signaling axis. Through quantitative analysis of large-scale time-lapse imaging and mathematical modeling, I identify force-sensitive contractility and cell packing as mediators of two distinct classes of collective migration. From these results, I formulate a phase-diagram of collective cell migration and test

predictions in an in-vivo epithelium using genetic manipulations to drive collective motion between predicted migratory phases. In Chapter 3, the effect of phenotypic heterogeneity on the organization of cells is examined, providing insight into the effects of early cancer progression on epithelial dynamics. I find that mutant cells within an otherwise wild-type tissue impact organization through local and field-effects, disrupting normal dynamics and leading to cell-type segregation. Chapter 4 provides a theoretical framework for quantitatively understanding and predicting the dynamics of protein interactions underlying biological processes including collective migration. Traditional chemical kinetics approaches break down in situations where components are slow diffusing or in countable numbers, requiring the formulation of new models that take into account this level of complexity. Here I develop an event-driven algorithm that bridges well-mixed and unmixed systems and use it to predict the effect of apparent changes in enzymatic efficiency due to alterations in mobility that may be caused by protein complex formation. Overall the work in this dissertation advances our understanding of the structure and dynamics of collective migration and the parameters governing this process by combining quantitative statistical analysis, mathematical modeling, and in-vivo live imaging.

"The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them."

– Sir William Bragg

CHAPTER ONE:

Introduction: Collective migration in development, regeneration, and disease

1.1 FROM INDIVIDUAL CELLS TO COLLECTIVES

The ability of individual cells to coordinate behaviors and function as groups defines multicellular organisms. Across all stages of life, individual cells organize motions to migrate as ensembles in a processes termed collective migration (Friedl and Gilmour, 2009). In development, it underlies the patterning of embryos and organs, allowing neural crest cells to travel throughout an organism and complex ductal structures to form (Ewald et al., 2008; Shih and Yamada, 2012). In regeneration, it allows for wound healing with maintained tissue integrity, and in disease, it plays a role in enhancing directed invasion of tumor cells into the surrounding tissue (Friedl and Wolf, 2003; Krawczyk, 1971).

Much research has focused on the mechanisms of single cell migration and elucidated a five-step model for migration: leading edge extension, formation and stabilization of cell-substrate contacts, generation of traction forces, and trailing edge detachment (Lauffenburger and Horwitz, 1996). Technological advancements in visualization and analysis, accompanied by a surge of interest in emergent collective phenomena, have led to a significant increase in the study of collective migration to determine how these fundamental pathways of motile force production extend to groups of cells. Despite these efforts, the mechanisms by which cells coordinate motion among neighbors remain incompletely understood.

1.2 CLASSIFYING COLLECTIVE MIGRATION

Migrating collectives adopt a broad spectrum of appearances that can be classified on the basis of cohesiveness and translocation. In mesenchymal-type collective migration, loosely coherent collections of cells migrate persistently but frequently exchange neighbors, while in epithelial-type collective migration, neighbor contacts remain relatively fixed as the group migrates (Friedl and Gilmour, 2009; Vedel et al., 2013; Weijer, 2009). Not all observed motions of cells require migration, per-se. Large rearrangements of cells can occur through coordinated neighbor exchange as seen in convergent-extension or preferential constriction of either the apical or basal surface of a tissue leading to folding (Friedl and Gilmour, 2009). Motion is present in both of these cases, little translocation relative to an external substrate may occur. While these differences are important considerations for comparison of systems, it is likely that the shared mechanisms and components are involved across the collective motion spectrum.

1.3 COLLECTIVE CELL MIGRATION IN DEVELOPMENT, REGENERATION, AND CANCER

In development, coordinated motions are required for the basic patterning of embryonic germ layers and the morphogenetic events that lead to intricate structures in organ formation (Scarpa and Mayor, 2016; Weijer, 2009). These include, but are not limited to, gastrulation, neural crest cell migration, lateral line migration, branching morphogenesis, border cell migration, and egg chamber rotation (Caussinus et al., 2008; Haigo and Bilder, 2011; Montell, 2003; Shook and Keller, 2003; Theveneau and Mayor, 2012). The diversity of these processes provide model systems for the study of both mesenchymal and epithelial-type collective motion and have demonstrated the importance of chemoattractants, leader-follower structure, and regulation of cell-cell contacts.

Of particular interest here is the collective migration of follicular epithelial cells in insect ovarioles. First identified in *Heteropeza pygmaea* in 1976 and re-characterized in *Drosophila melanogaster* in 2011, the epithelial layer surrounding the developing eggs undergoes a transition to rapid migration on an underlying layer of extracellular matrix (Fux et al., 1978; Haigo and Bilder, 2011). This rotation is important for the elongation and efficient extrusion of eggs (Horne-Badovinac, 2014). The genetic tools available in Drosophila and the ease of isolation of this tissue make it an ideal candidate for a model system of epithelial collective migration.

Collective migration plays a role in both epithelial and endothelial regeneration. Following injury, the remaining epithelium must seal the wound and restore normal barrier function. In epithelial injury, a cellular monolayer brides the wound gap while

maintaining tissue integrity through E-cadherin based adherens junctions and desmosomes (Moll et al., 1999; Poujade et al., 2007). The free space created by the wound may serve as a signal for directed migration (Mayor and Carmona-Fontaine, 2010), but inflammatory cytokines released by stromal fibroblasts play an important role in guiding migration (Werner et al., 2007). Part of this wound healing process is an additional collective migration event, revascularization. Endothelial cells in existing vessels generate protrusions and form a sprout that elongates through migration of the leading tip cells and proliferation of the trailing stalk cells (Sainson et al., 2005).

The collective nature of tumor cell invasion has long been recognized in pathological samples, which show tumor cells clustered in sheets, strands, and nests (Sahai, 2005). In *in-vivo* and *in-vitro* experiments, tumor cells have been observed to retain cell-cell contacts and migrate in a cooperative manner, enhancing the efficiency of invasion (Alexander et al., 2008; Nabeshima et al., 1999). Collective groups of cells have also been observed beyond the primary tumor site in both the lymphatic system and blood stream (Hendrix et al., 2003; Liotta et al., 1974). The factors influencing this process are incompletely understood, but growing evidence points to a complex interplay between cellular and environmental features, including alterations in cell-cell adhesion, extracellular matrix remodeling, release of cytokines, and interactions with stromal cells (Gaggioli et al., 2007; Soria et al., 2011; Wolf et al., 2007). Lastly, understanding the mechanisms of collective migration is of particular importance in regards to cancer heterogeneity. By bringing along multiple cells, collective invasion results in a greater diversity in genetic and phenotypic characteristics for each metastatic event, which may in turn lead to greater resistance or chance of recurrence (Heppner and Miller, 1983).

1.4 COLLECTIVE BEHAVIOR ACROSS PHYSICAL AND BIOLOGICAL SYSTEMS

Emergence of collective behavior in general has garnered significant interest in recent years. Nature is full of patterns and collectives: sand piles into dunes, ameboid *Dictyostelium* aggregate into fruiting bodies, ants assemble bridges, and birds form synchronized flocks with thousand of members (Bagnold, 2012; Bonner and Savage, 1947; Cavagna et al., 2010; Reid et al., 2015). While aesthetically pleasing, the fundamental question remains: to what extent is apparent organization simply pattern and to what extent is it functional? Sand dunes are inert, yet their formation demonstrates a mechanism by which simple rules can lead to organization, and these rules may be stumbled upon and reinforced through evolution. In the case of sand dunes, the mechanism responsible is local activation with long-range inhibition and is thought to be the same general rule that allows cells to polarize (Goehring and Grill, 2013). The diverse set of fields studying collective behavior provide a multitude of statistical techniques, analytical approaches, and underlying theory, which has helped to inform these and other studies of cellular collective migration (Cavagna et al., 2010; Katz et al., 2011). In particular, extensive parallels to collective cell migration have been made to other complex and crowded systems including solids, liquids, and amorphous glasses, which have provided insight into the physical mechanisms that control migration (Angelini et al., 2011; Bi et al., 2015; Bi et al., 2016; Deforet et al., 2014; Garcia et al., 2015; Nnetu et al., 2013; Puliafito et al., 2012; Sadati et al., 2013).

1.5 IMAGE PROCESSING IN THE STUDY OF COLLECTIVE MIGRATION

The study of collectives has been greatly enhanced by application of computer vision algorithms to collected data. Tracking individual features of thousands of cells is difficult and error prone, and to combat this a set of flow visualization methods has been developed The work presented here relies extensively on these techniques to extract usable data from time-lapse videos of collective migration. As these methods are commonly used throughout the following work, a brief introduction is provided here. Motion is identified through two techniques here, Particle Image Velocimetry (PIV) and Optical Flow. PIV is a technique that originated in fluid mechanics to study patterns that arose in fluid flow by adding particulate tracers to the flow. The motion of the fluid is estimated by finding a shift in space that provides the best alignment between corresponding sub-regions of consecutive frames (Westerweel 2013). This technique can be applied to a variety of image modalities, but the majority of implementations rely on the assumption that the particles within a sub-region undergo rigid translation. An alternative approach, known as Optical Flow, stems directly from the equations of mass transport that define fluid flow (Lucas 1981). Objects are encoded in images as specific values of pixel intensity and as they move in space, their pixel intensities shift as well. Optical flow works by solving for the spatial shift in the pixel intensities of frame N that best predicts the change in intensity between image N and N+1 (Vig 2016).

1.6 SCOPE OF THE DISSERTATION

The primary goal of this dissertation is to improve our understanding of the parameters that govern collective migration and how common alterations to cellular phenotypes seen in health and disease may restructure collective behavior.

Chapter 2 dissects the roles of three fundamental cell parameters in collective migration: cell-density, cell-cell adhesion, and cell-cell contractility. The effect of these variables on the size, shape, and leader-follower dynamics of collectives is determined and summarized in a phase diagram. The predictive nature of this model is then tested *in-vivo*. This chapter is in revision at Molecular Biology of the Cell:

> **Loza AJ**, Koride S, Schimizzi GV, Li B, Sun SX, Longmore GD. Cell Density Organizes Collective Migration through Changes in Actomyosin Contractility. (*in revision at MBoC*)

Chapter 3 examines the role of heterogeneous apical-basal polarity loss on the patterning and dynamics of epithelia, modeling a state of early cancer where abnormal cells arise within an otherwise normal epithelium. At the time of dissertation defense, this chapter was in preparation for submission.

> **Loza AJ**, Longmore GD. Populations of apical-basal polarity deficient cells disrupt an otherwise normal epithelium through direct contact and field effects.

Chapter 4 provides a theoretical framework and computational implementation of an algorithm for simulating biochemical networks where components are slow diffusing and in countable numbers. This develops the computational tools required for studying the signaling systems that underlie collective migration. At the time of dissertation defense, this chapter was in submission to PLoS One.

Loza AJ, Sherman MS, Elson EL. General Simulation of a Reaction Transport System. (in preparation for submission to PLoS One)

In summary, this work illustrates how quantitative imaging, statistical analysis, and mathematical modeling can advance our knowledge of the mechanisms organizing the motion of groups of cells. Furthermore, these fundamental principles can be used in a predictive manner to drive changes in migration *in-vivo*.

1.7 REFERENCES

Alexander, S., Koehl, G.E., Hirschberg, M., Geissler, E.K., and Friedl, P. (2008). Dynamic imaging of cancer growth and invasion: a modified skin-fold chamber model. Histochemistry and cell biology *130*, 1147-1154.

Angelini, T.E., Hannezo, E., Trepat, X., Marquez, M., Fredberg, J.J., and Weitz, D.A. (2011). Glass-like dynamics of collective cell migration. Proc Natl Acad Sci U S A *108*, 4714-4719.

Bagnold, R.A. (2012). The physics of blown sand and desert dunes (Courier Corporation).

Bi, D., Lopez, J.H., Schwarz, J.M., and Manning, M.L. (2015). A density-independent rigidity transition in biological tissues. Nature Physics *11*, 1074-1079.

Bi, D., Yang, X., Marchetti, M.C., and Manning, M.L. (2016). Motility-driven glass and jamming transitions in biological tissues. Physical Review X *6*, 021011.

Bonner, J.T., and Savage, L. (1947). Evidence for the formation of cell aggregates by chemotaxis in the development of the slime mold Dictyostelium discoideum. Journal of Experimental Zoology *106*, 1-26.

Caussinus, E., Colombelli, J., and Affolter, M. (2008). Tip-cell migration controls stalk-cell intercalation during Drosophila tracheal tube elongation. Current biology *18*, 1727-1734.

Cavagna, A., Cimarelli, A., Giardina, I., Parisi, G., Santagati, R., Stefanini, F., and Viale, M. (2010). Scale-free correlations in starling flocks. Proceedings of the National Academy of Sciences *107*, 11865-11870.

Deforet, M., Hakim, V., Yevick, H., Duclos, G., and Silberzan, P. (2014). Emergence of collective modes and tri-dimensional structures from epithelial confinement. Nature communications *5*.

Ewald, A.J., Brenot, A., Duong, M., Chan, B.S., and Werb, Z. (2008). Collective epithelial migration and cell rearrangements drive mammary branching morphogenesis. Dev Cell *14*, 570-581.

Friedl, P., and Gilmour, D. (2009). Collective cell migration in morphogenesis, regeneration and cancer. Nat Rev Mol Cell Biol *10*, 445-457.

Friedl, P., and Wolf, K. (2003). Tumour-cell invasion and migration: diversity and escape mechanisms. Nat Rev Cancer *3*, 362-374.

Fux, T., Went, D., and Camenzind, R. (1978). Movement pattern and ultrastructure of rotating follicles of the paedogenetic gall midge, Heteropeza pygmaea Winnertz (Diptera: Cecidomyiidae). International Journal of Insect Morphology and Embryology *7*, 415-426.

Gaggioli, C., Hooper, S., Hidalgo-Carcedo, C., Grosse, R., Marshall, J.F., Harrington, K., and Sahai, E. (2007). Fibroblast-led collective invasion of carcinoma cells with differing roles for RhoGTPases in leading and following cells. Nat Cell Biol *9*, 1392-1400.

Garcia, S., Hannezo, E., Elgeti, J., Joanny, J.-F., Silberzan, P., and Gov, N.S. (2015). Physics of active jamming during collective cellular motion in a monolayer. Proceedings of the National Academy of Sciences *112*, 15314-15319.

Goehring, N.W., and Grill, S.W. (2013). Cell polarity: mechanochemical patterning. Trends in cell biology *23*, 72-80.

Haigo, S.L., and Bilder, D. (2011). Global tissue revolutions in a morphogenetic movement controlling elongation. Science *331*, 1071-1074.

Hendrix, M.J., Seftor, E.A., Hess, A.R., and Seftor, R.E. (2003). Molecular plasticity of human melanoma cells. Oncogene *22*, 3070-3075.

Heppner, G.H., and Miller, B.E. (1983). Tumor heterogeneity: biological implications and therapeutic consequences. Cancer and Metastasis Reviews *2*, 5-23.

Horne-Badovinac, S. (2014). The Drosophila egg chamber—a new spin on how tissues elongate. Integrative and comparative biology *54*, 667-676.

Katz, Y., Tunstrøm, K., Ioannou, C.C., Huepe, C., and Couzin, I.D. (2011). Inferring the structure and dynamics of interactions in schooling fish. Proceedings of the National Academy of Sciences *108*, 18720-18725.

Krawczyk, W.S. (1971). A pattern of epidermal cell migration during wound healing. J Cell Biol *49*, 247-263.

Lauffenburger, D.A., and Horwitz, A.F. (1996). Cell migration: a physically integrated molecular process. Cell *84*, 359-369.

Liotta, L.A., Kleinerman, J., and Saidel, G.M. (1974). Quantitative relationships of intravascular tumor cells, tumor vessels, and pulmonary metastases following tumor implantation. Cancer research *34*, 997-1004.

Mayor, R., and Carmona-Fontaine, C. (2010). Keeping in touch with contact inhibition of locomotion. Trends in cell biology *20*, 319-328.

Moll, I., Houdek, P., Schäfer, S., Nuber, U., and Moll, R. (1999). Diversity of desmosomal proteins in regenerating epidermis: immunohistochemical study

using a human skin organ culture model. Archives of dermatological research *291*, 437-446.

Montell, D.J. (2003). Border-cell migration: the race is on. Nature Reviews Molecular Cell Biology *4*, 13-24.

Nabeshima, K., Inoue, T., Shimao, Y., Kataoka, H., and Koono, M. (1999). Cohort migration of carcinoma cells: differentiated colorectal carcinoma cells move as coherent cell clusters or sheets. Histology and histopathology *14*, 1183-1187.

Nnetu, K.D., Knorr, M., Pawlizak, S., Fuhs, T., and Käs, J.A. (2013). Slow and anomalous dynamics of an MCF-10A epithelial cell monolayer. Soft Matter *9*, 9335.

Poujade, M., Grasland-Mongrain, E., Hertzog, A., Jouanneau, J., Chavrier, P., Ladoux, B., Buguin, A., and Silberzan, P. (2007). Collective migration of an epithelial monolayer in response to a model wound. Proc Natl Acad Sci U S A *104*, 15988-15993.

Puliafito, A., Hufnagel, L., Neveu, P., Streichan, S., Sigal, A., Fygenson, D.K., and Shraiman, B.I. (2012). Collective and single cell behavior in epithelial contact inhibition. Proc Natl Acad Sci U S A *109*, 739-744.

Reid, C.R., Lutz, M.J., Powell, S., Kao, A.B., Couzin, I.D., and Garnier, S. (2015). Army ants dynamically adjust living bridges in response to a cost–benefit trade-off. Proceedings of the National Academy of Sciences *112*, 15113-15118.

Sadati, M., Taheri Qazvini, N., Krishnan, R., Park, C.Y., and Fredberg, J.J. (2013). Collective migration and cell jamming. Differentiation *86*, 121-125.

Sahai, E. (2005). Mechanisms of cancer cell invasion. Current opinion in genetics & development *15*, 87-96.

Sainson, R.C., Aoto, J., Nakatsu, M.N., Holderfield, M., Conn, E., Koller, E., and Hughes, C.C. (2005). Cell-autonomous notch signaling regulates endothelial cell branching and proliferation during vascular tubulogenesis. The FASEB journal *19*, 1027-1029.

Scarpa, E., and Mayor, R. (2016). Collective cell migration in development. The Journal of cell biology *212*, 143-155.

Shih, W., and Yamada, S. (2012). N-cadherin as a key regulator of collective cell migration in a 3D environment. Cell Adh Migr *6*, 513-517.

Shook, D., and Keller, R. (2003). Mechanisms, mechanics and function of epithelial–mesenchymal transitions in early development. Mechanisms of development *120*, 1351-1383.

Soria, G., Ofri-Shahak, M., Haas, I., Yaal-Hahoshen, N., Leider-Trejo, L., Leibovich-Rivkin, T., Weitzenfeld, P., Meshel, T., Shabtai, E., and Gutman, M. (2011). Inflammatory mediators in breast cancer: coordinated expression of TNFα & IL-1β with CCL2 & CCL5 and effects on epithelial-to-mesenchymal transition. BMC cancer *11*, 1.

Theveneau, E., and Mayor, R. (2012). Neural crest delamination and migration: from epithelium-to-mesenchyme transition to collective cell migration. Developmental biology *366*, 34-54.

Vedel, S., Tay, S., Johnston, D.M., Bruus, H., and Quake, S.R. (2013). Migration of cells in a social context. Proc Natl Acad Sci U S A *110*, 129-134.

Weijer, C.J. (2009). Collective cell migration in development. Journal of cell science *122*, 3215-3223.

Werner, S., Krieg, T., and Smola, H. (2007). Keratinocyte–fibroblast interactions in wound healing. Journal of Investigative Dermatology *127*, 998-1008.

Wolf, K., Wu, Y.I., Liu, Y., Geiger, J., Tam, E., Overall, C., Stack, M.S., and Friedl, P. (2007). Multi-step pericellular proteolysis controls the transition from individual to collective cancer cell invasion. Nature cell biology *9*, 893-904.

CHAPTER TWO:

Cell density and actomyosin contractility control the organization of migrating collectives within an epithelium

2.1 SUMMARY

The mechanisms underlying collective migration are important for understanding development, wound healing, and tumor invasion. Here we focus on cell density to determine its role in collective migration. Our findings show that increasing cell density, as might be seen in cancer, transforms groups from broad collectives to small narrow streams. Conversely, diminishing cell density, as might occur at a wound front, leads to large broad collectives with a distinct leader-follower structure. Simulations identify force-sensitive contractility as a mediator of how density impacts collectives, and guided by this prediction, we find the baseline state of contractility can enhance or reduce organization. Lastly, we test predictions from these data in an in-vivo epithelium by using genetic manipulations to drive collective motion between predicted migratory phases. This work demonstrates how commonly altered cellular properties can prime groups of cells to adopt migration patterns that may be harnessed in health or exploited in disease.

2.2 INTRODUCTION

Collective migration of cells underlies embryonic development, tissue regeneration, and tumor invasion (Ewald et al., 2008; Friedl and Gilmour, 2009; Krawczyk, 1971). In these processes, cells coordinate their individual motions to productively migrate in groups, giving rise to intricate branching structures (e.g. breast development), covering injured areas while retaining tissue integrity (e.g. skin wound healing), or invading into surrounding tissue (e.g. cancer metastasis).

To understand the complex behaviors of groups of cells, parallels have been made to other crowded physical systems, including solids, liquids, and amorphous glasses (Angelini et al., 2011; Bi et al., 2015; Bi et al., 2016; Deforet et al., 2014; Garcia et al., 2015; Nnetu et al., 2013; Puliafito et al., 2012; Sadati et al., 2013). These studies find similarities, but also highlight that unlike standard physical systems, cellular tissues fall into a class of materials known as active matter (Angelini et al., 2011), meaning cells internally generate their own migratory forces. Furthermore, cells are mechanosensitive, mechanoresponsive, and can be physically coupled to one another, allowing cells to directly impact their neighbors (Chen et al., 2004; Das et al., 2015). The types of collective motion that can form in these force-generating systems and the properties that control organization and motion are only partially understood.

Cell density has been implicated as an important property influencing the behavior of groups of cells, but how collective migration is specifically impacted and through which mechanisms are not completely known (Angelini et al., 2011; Nnetu et al., 2013; Vedel et al., 2013). Understanding this effect is of broad importance because shifts in cell density underlie many biological processes where collective migration occurs. In development and early cancer, for example, the level of cell proliferation is high (Cooper, 2000; Hanahan and Weinberg, 2011).

In the former, it is closely regulated to maintain normal tissue growth and homeostasis, whereas in the latter this proliferation is dysregulated leading to abnormal expansion of cells within tissues. In both instances, cell proliferation may increase cell-density within a tissue and thereby influence the migratory dynamics of constituent cells. A wound to an epithelial surface creates free space and the stimulus for directed migration to repair the denuded surface (Poujade et al., 2007). In contrast to development and cancer, this wound additionally produces a local drop in cell density extending multiple rows back from the site of injury (Krawczyk, 1971). This decrease in cell density therefore may impact the organization and collective migration of cells at the leading edge. A more complete understanding of the role of cell density in collective migration may allow for predictions as to how changes in tissue states generate the migratory behaviors that emerge.

In this study, we asked how cell density affects the collective organization and movement of cells within a model epithelium. The express purpose was to develop a quantitative model of collective cell migration, and through experiment and simulation, we identify force sensitive contractility as a mediator of the response of collective migration to increasing density. This iterative approach led to the development of a phase diagram of cell migration summarizing changes in collective behavior in response to changes in cell density and cell-cell contractility. We then tested the predictive value of this phase diagram in an in-vivo epithelium. Through genetic manipulations altering cell density and contractility, we were able to drive the *Drosophila* ovarian follicular epithelium from one predicted migratory phase to another.

2.3 RESULTS

Collective cell migration within an epithelium exhibits two distinct responses to increasing cell density

To develop a quantitative model describing how cells organize into collective groups within an epithelium, we asked how changing cell density within a confluent epithelial monolayer affected collective migration. Human breast epithelial cells (MCF10A) were plated in confluent monolayers and time-lapse videos spanning 2.3 mm^2 were acquired across a range of cell densities. Cell density was measured for each field of view using a nuclear label (Sup Fig 2.1A). Density was varied by plating a constant number of cells and taking measurements at progressively longer times, however measurements made at the same time on more densely plated cells yielded similar results (Sup Fig 2.1B-E).

Qualitatively, motion within the monolayer was highly organized at low density, with large groups of cells moving together in similar directions (Fig 2.1A and Video S1). At high density, cell movement was more heterogeneous and groups of similarly moving cells appeared smaller (Fig 2.1B and Video S2). Although mobility decreased with increasing density, migration remained superdiffusive across the range in cell density (Sup Fig 2.1H-K). To quantify organization while adjusting for changes in cell size, we computed a radial spatial autocorrelation function for the measured velocity field (see Supplemental Experimental Procedures) with distance normalized by mean cell radius. This function has a value of one when motion is aligned and zero when motion is random and is referred to here as radial correlation because correlations are computed only as a function of radial distance between observations. At low density, radial correlation decayed to approximately zero at a distance of 34 cell radii (red arrow, Fig 2.1C). This value, ρ, served as a reference distance over which organization turned to

random motion in low-density epithelia. Increasing density led to a faster early decay in radial correlation, however the curve then plateaued, remaining greater than zero at the reference distance of ρ=34 cell radii (Fig 2.1C and S1G).

These results suggested the presence of two length-scales of organization. To determine this, we measured the initial decay rate and the correlation plateau value by fitting each correlation curve to the following function:

$$C(r) = \left(1 - C_\rho\right)e^{-r/\lambda} + C_\rho, \quad (1)$$

Where C is the correlation at distance r, C_ρ is the correlation plateau value evaluated at a distance of ρ=34 cell radii, and λ is the correlation decay length (Fig 2.1D). The two variables λ and C_ρ identify two scales of organization. The parameter λ is a measure of local organization and C_ρ is a measure of broader regional organization describing how local groups are organized relative to each other. Higher values of both λ and C_ρ correspond to increased order. Correlation curves were fit using data out to 40 cell radii where the form of the fitting function remained valid, however fitting to a mixed exponential-linear decay model, which matches data out to 100 cell radii, produced results in close agreement (Sup Fig 2.2E-I).

We found that epithelia exhibited two distinct responses to increasing cell density that differentially affected these two length-scales. First, local organization (λ) decreased with increasing density, indicating a paradoxical decrease in local collective cell organization with increasing density. Second, regional organization (C_ρ) increased with increasing density (Fig 2.1F and 1G).

The observed effects at high density could involve specific cell-cell adhesion contacts (e.g. E-cadherin mediated) or a more general phenomenon involving cell packing (Huttenlocher et al., 1998; Trappe et al., 2001). To distinguish between these possibilities, we depleted α-

catenin using a specific shRNAi (Sup Fig 2.1F) (Bajpai et al., 2008). Alpha-catenin is a key component linking cadherin bonds to the actin cytoskeleton and as such is required for strong epithelial cell-cell adhesion (Nagafuchi et al., 1994). In α-catenin depleted cells, local organization (λ) remained low across all densities, but regional organization (C_ρ) dramatically increased with increasing density (Fig 2.1E). Therefore, cell-cell adhesion appeared necessary to produce the high local organization seen in WT epithelia at low density, and the increase in regional organization seen in WT epithelia at high density was enhanced by a reduction of functional cell-cell adhesion.

To facilitate determining the underlying physical cause(s) of this density response, we formulated a 2D vertex model of monolayer migration (Fig 2.1H). Vertex models are advantageous for simulating epithelial cell migration because they incorporate the connected nature of epithelia, cell shape changes, neighbor rearrangements, and a range of forces driving cell motion (Fletcher et al., 2014). For a detailed description of the model and the values of parameters used see Supplemental Experimental Procedures and Tables S2 and S3. Briefly, each cell is represented as a polygon with shared vertices and edges. Migration is governed by motion of their vertices and arises from passive (elastic and adhesive), frictional, and active (protrusive and contractile) forces.

Based upon previous modeling of cell sheets confined to circular regions (Li and Sun, 2014), we simulated the migration of 1000 cells under conditions of constant protrusive and contractile forces, referred to as the constant contractility model. These simulations captured low-density WT behavior (Fig 2.1I), however, organization was not responsive to changes in density (Fig 2.1K). We therefore modified the vertex model to incorporate dynamic signaling of Rho-ROCK-Myosin II activation sensitive to the physical environment of each cell, referred to

as the biomechanical contractility model (see Supplemental Experimental Procedures and Sup Fig 2.2.2K). This model is based on the observation that cells actively respond to externally applied mechanical forces by increasing myosin activation (Koride et al., 2014). The biomechanical contractility model predicted local organization across all densities including the paradoxical decrease in local organization with increasing density (Fig 2.1J, K). The experimental trend in regional organization was not observed, however. This reflects a limitation of the periodic boundary conditions used in the simulation, which influence long-distance correlations. Performing simulations of larger size reduced this effect, but still showed influence of the periodic boundary conditions (Sup Fig 2.1L-M). For this reason, we focused the simulation on examining local organization.

Given the requirement of α-catenin for local organization at low density, we asked whether the simulation could inform how it does so. Reducing α-catenin may impact a number of model parameters, including cell-cell adhesion and cell-cell friction. Reducing cell-cell adhesion or cell-cell friction alone did not reduce local organization (λ) at low density (Fig 2.1L, M), however it is unlikely these properties can be modified in isolation in cells. Beyond reducing adhesion, depletion of α-catenin likely reduced the coupling of contractile forces between cells. Reducing the parameter governing contractile force coupling diminished local organization at low density (Fig 2.1L and 1M), consistent with experimental observations (Fig 2.1E).

In summary, simulations identified a critical role for active contractility in organizing collective motion within an epithelium in response to changing cell density.

Regulated actomyosin contractility drives collective migration

Migration in the biomechanical model is driven by the state of the active contractility signaling system, providing experimentally testable predictions regarding activation of the Rho-ROCK-Myosin II pathway. Mean simulated cell-cell contractile activity was found to decrease with increasing density (Fig 2.2A). To experimentally test this prediction, cells were plated in a density gradient spanning 8 to 20μm in mean cell radius across a distance of 1cm (Fig 2.2B). As a measure for actomyosin contractility, the pMLC staining intensity per cell was measured by quantitative immunofluorescence and was found to decrease with increasing cell-density (Fig 2.2C). Supporting these data, measurement of pMLC intensity per cell with the same methods used to control density in migration experiments produced equivalent results (Sup Fig 2.2D).

The biomechanical contractility model also predicted the presence of contractile domains, defined as regions of high contractile activity surrounded by regions with lower contractile activity (Fig 2.2D). In simulation, contractile domains were more evident at low density than high density (Fig 2.2D, F). Experimentally, clusters of cells with high pMLC levels surrounded by regions with low pMLC levels were also more evident at low density (Fig 2.2E, G). Therefore, two density dependent features predicted by the biomechanical contractility model, namely mean contractility and the presence of contractile domains, were found experimentally.

Given the identified important role of cell-cell contractility in regulating collective migration in response to cell density, we hypothesized that changing baseline levels of actomyosin contractility would alter how cells responded to changes in density. Actomyosin contractility was diminished by treating cells with ROCK inhibitors Y27632 or Fasudil (Sup Fig 2.2B) or increased by stably expressing constitutively active Rho (RhoCA) (Sup Fig 2.2C). At low density, both decreasing and increasing actomyosin contractility diminished local

organization (λ) without affecting regional organization (C_ρ) (Fig 2.2H-K). At high density, decreasing or increasing contractility led to small differences in local organization relative to WT (Fig 2.2J). In contrast, regional organization showed larger differences relative to WT, increasing in cells with reduced contractility and decreasing in cells with increased contractility (Fig 2.2K).

In summary, these experiments indicated that the baseline state of actomyosin contractility altered the effect of cell-density on collective migration.

Directional correlations reveal transitions in collective shape

Inspection of time-lapse videos suggested that changes in the shape or pattern of collective groups accompanied changes in collective size (Videos S3 and S4) and that measuring radial correlation alone masked these patterns. For example, in contracting regions (i.e., cells converging on a point), correlations in the direction of travel may be low or negative while correlations in the reverse direction of travel may be high (Fig 2.3A, Sup Fig 2.3). These values are combined in the measurement of radial correlation, and thus this difference would not be apparent. In the case of streaming motions (i.e., follow-the-leader), the average correlation in the direction of travel remains higher than correlations perpendicular to the direction of travel (Fig 2.3B, Sup Fig 2.3). Again, these two disparate values would be averaged in radial correlation.

To measure the shape of collectively migrating groups, we defined a polar spatial correlation function using a mathematical approach related to previous methods (Kabla, 2012; Kuriyama et al., 2014; Szabo et al., 2010) (see Supplemental Experimental Procedures), referred to here as directional-correlation. In this analysis, correlations were assigned both a distance and direction relative to the index vector (Fig. 3C). In WT cells at low density, directional-correlations from a single field of view often exhibited dramatic asymmetries relative to the

direction of travel, suggesting the presence of large contracting, stretching, or rotating domains (Fig. 3D). When combining multiple tiled fields of view, directional correlation lost front-to-back asymmetry due to averaging of many contracting and expanding regions, but additional features remained. At low density, WT cells displayed a high degree of organization in the direction of travel and perpendicular to it, indicating that collectives were broad (Fig. 3E). At medium density, directional correlation decayed slower perpendicular to the direction of travel, also consistent with broad collective groups (Fig. 3E). At high density, directional correlation was highest in the direction of travel, indicating a transition to streaming motion where cells moved in a follow-the-leader manner and slid past lateral neighbors (Fig. 3E).

In contrast, αcatKD cells exhibited streaming motions at all densities, with greater correlation parallel to the direction of travel (Fig. 3G). Decreasing actomyosin contractility with Y27632 mimicked α-catenin depletion, producing streaming behavior even at low densities (Fig. 3F). Increasing contractility with RhoCA, resulted in directional correlations consistent with broad collectives even at high density (Fig. 3H). These data indicated that actomyosin contractility is important for organizing broad collectives. Reducing cell-cell contractility (either pharmacologically, by increasing density, or by uncoupling via α-catenin depletion) led to streaming type motions.

Flow of information within an epithelial sheet

Low-density WT epithelia displayed the greatest degree of local organization. To determine how these large and broad collective groups might form, we generalized the directional correlation function to include time correlations, referred to here as spatiotemporal correlation (see Supplemental Experimental Procedures). In this measurement, the velocity of an index position

at an initial time is compared to the velocities of neighboring positions as they evolve through time, where the difference between the two time-points is referred to as delay time (Fig 2.4A).

Spatiotemporal correlations provided information on how the motion of a cell at a given time may affect the motion of neighboring cells at later times. Consider two groups of cells where the neighboring group is behind an index group relative to the index group direction of travel (Fig 2.4B). The neighboring group has a certain degree of correlation with the index group at an initial reference time. As time passes, correlation between the neighboring group and initial state of the reference group may change. If the correlation rises, this means the migration of the neighboring group aligned to the index group over the delay time (Fig 2.4B). Using this analysis, we found that neighboring cells behind and beside aligned their motion to the direction of migration of an index cell.

For low-density WT cells, spatiotemporal correlations normalized to values at zero delay time were used to assess changes in organization. At a delay of 4 minutes, neighboring positions behind the index position in an arc of approximately 145 degrees were more correlated with the index position than at 0 minutes, while positions in front were not (Fig 2.4C). At later times, the region of enhanced correlation grew to include lateral neighboring positions in an arc of approximately 240 degrees. This asymmetry in spatiotemporal correlation shows that information flows from front to back even within a confluent epithelium, whereby the current migration of a given cell provides information on the future migration of neighbors in back or beside, but not those in front (where front and back are defined by the index cell migration direction).

As delay time increased, the peak in spatiotemporal correlation appeared to shift to distances further behind the index point. This peak represents the response time of the system

(Katz et al., 2011), and a shift of the peak to increasing distances at longer delay-times provides the rate at which this response propagates. To measure this, two 45-degree sectors in front of and behind the index point were used to define index-to-back neighbors and index-to-front neighbors (Fig 2.4D3). For index-to-back neighbors, the peak in normalized correlation occurred at later times for neighboring positions farther from the index position, indicating that the response indeed propagated rearward (Fig 2.4D1). In contrast, no peaks were observed for index-to-front neighbors (Fig 2.4D2), consistent with the previously determined direction of information flow. Plotting neighbor distance versus time-delay for each peak produced a linear trend (slope of 5.0 ± 0.1 μm/min, intercept of 47±2 μm) (Fig 2.4E). The slope of this line is the propagation rate of cellular response, and notably, it is significantly greater than the mean cell speed (0.29 μm/min). The difference in propagation rate and migration speed means the relative alignment was not due to neighboring cells migrating in a similar direction once reaching the position of an index cell.

This transfer of information depended on the Rho-ROCK-Myosin II signaling axis. Inhibiting contractility using Y27632 largely prevented trailing neighbors from aligning to the heading of an index position (Fig 2.4F). In epithelia with increased contractility, alignment occurred but did not include as many lateral neighbors (Fig 2.4G). The correlation peak for RhoCA cells propagated rearward at a rate similar to WT cells of 4.4 ± 0.3 μm/min. Furthermore, the alignment of neighbor cells to an index cell was dependent on cell density, with no detectable alignment response at medium or high density (Fig. S4).

These results demonstrated that the formation of broad collectives depends on actomyosin contractility that serves to direct the migration of neighboring cells behind and beside an index cell.

Cell density and actomyosin contractility define phases of collective migration

Taken together, these data indicated that epithelial collective migration can be described in terms of size (radial correlation) and shape (directional correlation) in a phase diagram based on two parameters: cell-density and functional cell-cell contractility (i.e., contractility contingent on cell-cell adhesion) (Fig 2.5A). Cell-density and cell-cell contractility differentially impact collective size at two length scales: local and regional organization. Local organization was greatest for low-density epithelia with WT levels of cell-cell contractility (Fig 2.5A and 5B, purple shading) and decreased with increasing density or manipulations of actomyosin contractility. Regional organization was greatest for high-density cells with low levels of functional cell-cell contractility (Fig 2.5A and 5B, red shading).

Three general shapes of collective migration were observed through the directional correlation analysis (Fig 2.5A, dashed black lines). In the first region (I), correlation was high in all directions relative to the direction of travel. In the second region (II), directional correlation was greatest perpendicular to the migration heading of a given cell, consistent with a mixture of contracting and stretching regions: cells are likely to be moving with lateral neighbors whether converging to or diverging from a point, but may be migrating in the opposite direction of neighbors ahead (contracting) or behind (stretching). In the third region, directional correlation was low perpendicular to the migration heading of an index cell, consistent with streaming motions.

Contractility and cell-density drive in-vivo changes in collective migration

To test these predictions in an in-vivo epithelium with three-dimensional geometry, we turned to the *Drosophila* ovarian follicular epithelium. The developing *Drosophila* egg chamber has a basement membrane and apical-basal polarized epithelium surrounding an oocyte and supporting nurse cells. During the course of egg development, the epithelial layer undergoes a migratory transition (Cetera and Horne-Badovinac, 2015; Cetera et al., 2014; Haigo and Bilder, 2011). In early stages (1-5), cells exhibit random jostling motions with slow drift in the future direction of organized migration. At stage 6, the follicular epithelium transitions to fast and highly organized collective motion. This organized migration is thought to be essential for egg chamber elongation allowing for efficient egg extrusion (Horne-Badovinac, 2014), yet the mechanisms which initiate and accelerate rotation are not fully understood. The proposed phase diagram predicts that both cell density and contractility must be tightly regulated for the onset and maintenance of rapid collective migration.

We first tested predictions regarding the initiation of organized collective motion, focusing on stages 4-5 when egg chambers were large enough to measure velocity fields but had not yet begun rapid rotation. Stage 4-5 ovarian follicular epithelia were low density, had low levels of basolateral actomyosin contractility, and displayed a low degree of organized motion, consistent with a phase in the bottom left of the diagram (Fig 2.6A-D). A confluent monolayer and the spheroidal layer of ovarian follicle cells may both be considered to be edgeless epithelia, however an important distinction between the two affects the ability to separate local versus regional organization. Moving radially outward from a given cell in a monolayer leads only to neighboring cells of greater absolute distance from a given cell. Moving radially outward from a given cell within a circularly connected epithelium, like the follicular epithelium, eventually

leads back to the same cell. Because of this connected nature, regional organization has the potential to directly influence local organization (i.e. the "region" may include the local group itself). For this reason we hypothesized that manipulations that affected either type of organization in 2D monolayers might lead to overall increases in collectivity.

Our findings predict that either increasing contractility or cell density would lead to the early onset of organized collective migration while decreasing functional contractility by disrupting cell-cell adhesion would increase disorganization (Fig 2.6A, blue arrows). To increase contractility, a constitutively active *Drosophila* ROCK allele (Rok-CAT) was expressed specifically in the follicular epithelium using the *traffic jam-Gal4* promoter (*tj-Gal4*) (Sup Fig 2.5A). To increase cell density, *Drosophila* Myc (dMyc) was overexpressed using *tj-Gal4*. Myc is a proto-oncogene that increases cell proliferation through induction of cyclin D2 (Bouchard et al., 1999). Expression of dMyc increased cell density compared to stage matched WT egg chambers, leading to smaller cells (equivalently, more per unit area) for any given stage (Fig 2.6B). Consistent with predictions, highly organized motion of stage 4-5 egg chambers emerged by increasing either contractility or density (Fig 2.6D, 6E, and 6H; Video S5). Reducing functional cell-cell contractility through α-catenin RNAi expression (αcat-IR) increased the degree of disorganization seen in WT stage 5 follicular epithelia leading to extensive neighbor exchange (Fig 2.6E and Video S5). In summary, increasing density or contractility drove stage 4-5 ovarian follicular epithelium to a state of highly organized collective rotation, while depleting α-catenin to reduce cell-cell adhesion and functional cell-cell contractility drove the epithelium into a state of greater disorder.

We next tested predictions regarding the maintenance of collective migration for stage 7-8 egg chambers, which exhibit highly uniform collective migration. Contractility of stage 7-8

egg chambers was greater than earlier stages as measured by pMLC staining, especially on the basolateral surfaces, consistent with published findings (Fig 2.6C) (Wang and Riechmann, 2007). Cell density peaked during stage 6 and lowered again for stage 7-8 egg chambers (Fig 2.6B). Therefore, we hypothesized that this epithelium was in a phase corresponding to mid-level contractility and slightly higher density than stage 4-5 egg chambers, offering multiple predictions. Increasing contractility, increasing cell density, or decreasing contractility should drive this epithelium from a state in which maximal order is produced to one with disorder (orange arrows Fig 2.6A). Increasing cell density by *tj-Gal4* Myc overexpression decreased organization and speed of stage 7-8 follicular epithelia (Fig 2.6F and 6H; Video S6). Increasing contractility by expressing Rok-CAT showed the predicted reduction in organization with regions of cells undergoing swirling motions that led to decreased order and regions with slower speed (Fig 2.6G and 6H; Video S7). Reducing cell-cell contractility through expression of α-catenin RNAi in stage 7-8 egg chambers dramatically disrupted organization (Fig 2.6G and 6H; Video S7). We controlled for disparities between morphological staging and developmental progression by measuring the size of egg chambers undergoing a border cell delamination and saw no difference (Fig. S5B, S5C).

In summary, these data demonstrated that the state of collective migration in an in-vivo epithelium could be predictably transformed by controlling cell-density and cell-cell contractility.

2.4 DISCUSSION

In this study we took a multi-faceted approach to determine the fundamental physical and biological mechanisms underlying how cells within an epithelium organize to collectively migrate. Focusing on the response of collective migration to changes in cell-density revealed that increasing cell-density drove changes in actomyosin contractility that diminished local organization, increased regional organization, and changed the shape of collective motion from large broad groups to small narrow streams. By defining these phases of collective migration and the parameters governing transitions among them, we were able to predictably alter collective migration in an *in-vivo* epithelium.

The observation that local organization decreased while regional order increased with increasing cell-density suggests that multiple length scales of coordination exist within an epithelium. Absolute local organization has been observed to decrease with increasing cell density (Angelini et al., 2010), and here we confirm that this reflects a true reduction in local group size by normalizing to cell radius. Interestingly, as local group organization decreased, the relative organization between these local groups increased. A potential explanation is that at high density, cells generate motile forces but are no longer able to align their motions to neighboring cells via a system of active contractility, leading cells to slide past each other and reduce local organization. This is not to say that these motions are uncoordinated: at high density, cells are less deformable and individual rearrangements cannot occur in isolation and may be coupled (Sadati et al., 2013), leading to an increase in regional organization. This two-level structure of organization bridges reports of decreases in correlation length with those that describe increases in the size of clusters of fast motion, termed dynamic heterogeneities, in the migration of cells at high-density (Trepat and Fredberg, 2011). Notably, increasing contractility prevented increases

in regional organization while diminishing contractility increased regional organization, supporting the idea that cell-cell contractility may prevent the growth of regional organization by organizing local motions at high density.

Contact inhibition of locomotion (CIL) is a process whereby cells alter their motion in response to interactions with neighbors (Abercrombie, 1970), and has been suggested to explain, in part, how collectives prioritize migration towards a free edge (Mayor and Carmona-Fontaine, 2010). CIL could, in principle, contribute to the observed alterations in local organization and motility with changing density. Cell-cell contacts and signal transduction of this event are critical for CIL, but the type of cell-cell contacts that form between two cells or groups of cells are likely context dependent. In a confluent epithelium contacts are ubiquitous and dynamic, but rather than generally inhibiting or redirecting motions, collective behavior emerges. Indeed, it has been proposed that cell-cell adhesions within a chain of cells signal differently from newly formed adhesions that are thought to induce CIL (Desai et al., 2013). Similarly, contacts within epithelial sheets may be different than nascent contacts that mediate CIL upon collision and act to mediate cohesion rather than repulsion. The leader-follower structure we observed through spatiotemporal analysis suggests there is prioritization among these contacts and that cell-cell tension may be a mediator. We also find that increasing cell density within an already confluent sheet reduces local collective organization, suggesting that signaling from cell-cell contacts may change as cell density increases.

Beyond demonstrating the impact of cell-density and contractility on collective size, we find that these variables control transitions between specific shapes of collective migration. We observed a key transition between broad collectives and streaming collectives that occurs with increasing density or loss of functional cell-cell contractility. This change in collective shape

may have important implications for collective invasion. Kuriyama et al. demonstrated that a reduction in N-cadherin was essential for collectives of neural crest cells to productively migrate during *Xenopus* development (Kuriyama et al., 2014). Correlation analyses performed by Kuriyama et al. show the effect of N-cadherin reduction is consistent with a collective shape change from broad collectives to streaming motions, enabling collectives to reshape and migrate through geometrically constrained regions. Alpha-catenin and other components linking cell adhesion to actomyosin contractility are commonly altered in cancers (Hiscox and Jiang, 1996; Shiozaki et al., 1994), and our results suggest that these losses could enhance invasion by changing collective group shape to streaming patterns.

Spatiotemporal analyses of cell migration revealed that broad collectives arise through information transfer whereby cells align to neighbors that they are behind or beside. Monolayer stress microscopy demonstrates that cells within epithelial sheets are under tension, and migration direction tends to align with the principal direction of normal stress. Our data give a dynamic view of cellular response to this picture of forces, demonstrating that the migration of a given cell can initiate changes in the migration of specific neighboring cells at low density. In the context of wound healing, the presence of a free edge may initially bias the migration direction of the first row of cells, reducing local cell density and putting the rest of the sheet in the "front-to-back" neighbor category. This directional signal can then propagate rearward, leading to a cycle of density reduction, neighbor-response, and directed migration.

The propagation speed of the alignment response is significantly faster than the speed of cell migration, suggesting cellular communication rather than external cues may be involved. This may be related to previously identified waves of contractility (Serra-Picamal et al., 2012). While these waves propagated at 1um/min, they were computed from the time required to reach

maximum response and significant changes in cellular strain and stress appear earlier. Additionally, the observed propagation rate predicts the average cellular response time to be approximately 6 minutes (5um/min, 32um diameter), similar to the estimated time of 5 minutes required for cells to change direction in galvanotactic studies (Cohen et al., 2014). Recent data on Rho GTPase polarization provide a potential molecular mechanism for this response. Mechanical tension, as may be generated between an index cell and trailing neighbor, can polarize Cdc42 activity in a P-cadherin and β-Pix dependent manner and Rac1 activity due to re-localization of Merlin (Das et al., 2015; Plutoni et al., 2016). Supporting this mechanism, cells with constitutively active Rho or inhibited ROCK show a poorer alignment response, potentially reflecting a decreased ability to generate polarized RhoGTPase activity. Furthermore, we find that the alignment response extends to encompass lateral positions. This indicates that while front-to-back polarization is likely a dominant component in organizing collective migration, lateral cell-cell associations may be important in maximizing the organization of large collectives.

Previous hypothesized cell migration phase diagrams have included three axes: cell density, cell force production, and cell-cell adhesion (Sadati et al., 2013). These data indicate that two of these axes, cell force production and cell-cell adhesion, might be highly coupled in epithelial systems as α-catenin depletion and ROCK inhibition had similar effects on both collective size and shape. We furthermore find that phases are not simply defined by a single parameter. The studied cell types have similar local organization high density, yet this motion differs in both regional and directional organization. This phase diagram still represents a simplified model of cell behavior. Many more cell-intrinsic and extrinsic properties play a role in governing collective migration including chemokines, extracellular matrix (ECM) proteins and

physical properties, as well as cell-ECM adhesion (Haas and Gilmour, 2006; Nguyen-Ngoc et al., 2012). It is likely that these features may impact how cells respond to increasing density or changes in contractility. Additionally, it is almost certain than alterations to an important component of one of these general cellular properties (i.e. an alteration to ROCK to affect cell-cell contractility) will influence another (for example, cell-ECM force generation) and that these additional effects may be of importance for an observed phenotype. Indeed, here we see that the WT response of collective migration to changes in density is itself driven by changes in contractility signaling.

The *Drosophila* egg-chamber follicular epithelium provided an opportunity to test these predictions in a system with 3D geometry and native ECM structure. Using this system, we demonstrated that basic cell migration studies can be used to provide insight into in-vivo migratory processes and suggests that although many factors govern cell migration, cell density and contractility are two major influences. Increasing contractility or cell-density led to high-speed collective rotational migration of the early follicular epithelium. Likely this represents an enhancement of underlying emerging collective rotation rather than initiation of collective motion. The follicular epithelium in stage 4-5 egg chambers has been shown to drift slowly in the direction of future rapid rotation (Cetera et al., 2014), but migration is not yet fast and uniform. Increasing contractility or cell-cell density may serve to better couple local and regional motions, respectively, until the entire egg chamber begins rotating. Although the connected nature of both the egg chambers and simulation provide coupling of local and regional motions, it is worth noting an important difference exists between migration on a spheroidal surface and the surface of a torus to which the periodic boundary conditions apply. While it is possible to have uniform migration on the surface of a torus, no such velocity field can exist on a sphere,

which must contain at least one stationary point or “cowlick”(Milnor, 1978). The formation and maintenance of these stationary points, especially when the egg chamber is approximately spherical presents an interested avenue for further exploration.

In a developmental context, these predictions suggest that both cell-density and cell-cell contractility may be endogenously regulated to control the onset of collective motion. The observed levels of actomyosin contractility and cell-density in the development of WT egg chambers in fact follow a trajectory consistent with the onset of rotation at stage 6: at this stage, WT egg chambers reach a peak in cell-density and have increased levels of actomyosin contractility (Fig 2.6B,C). Once rotating, however, an excess of contractile force generation or cell density may disrupt communication leading to disorganized motion.

Overall these data demonstrate how fundamental properties of individual cells impact the migratory properties of entire tissues, controlling the size and structure of collective groups. In development and wound healing, transitions in density and contractility can be harnessed for productive migration. However in cancer, these same systems may be exploited to enhance invasion.

2.5 EXPERIMENTAL PROCEDURES

Cell Culture and Density Control

Density variation was achieved by plating $6x10^5$ cells per 30mm dish and making measurements at 24, 48, and 72 hours (low, medium, and high-density respectively). The values for each density condition are listed in Table S1. Media was replaced each day and cell density was measured using Hoechst 33342 dye. Separate plates were used for each density to avoid phototoxicity. Distance values were normalized by cell size in each field of view before combining into a single density bin. Gradients in cell density were generated by swirling the culture dish for 30 seconds immediately after plating a uniform suspension of cells, increasing the relative concentration of cells in the center of the dish. To deplete a-catenin, the lentiviral pFLRu vector containing anti-alpha-catenin shRNA and puromycin resistance was used as described previously (Bajpai et al., 2008). MCF10A cells expressing constitutively active Rho were generated in the same manner using a QL activating mutant Rho (Coso et al., 1995) subcloned into a pFLRu vector. Puromycin (Sigma-Aldrich) was used in cell selection and maintenance at a concentration of 1.5 µg/mL. Fasudil (Sigma-Aldrich) and Y27632 (Sigma-Aldrich) were both used at a concentration of 30uM for 5 hours for ROCK inhibition.

Drosophila Lines, Staging, and Dissections

All crosses and staging were performed at 25°C. Stocks are described in FlyBase (http://flybase.org/). UAS-Myc (dm) (B#9675), and UAS-Rok-CAT-3.1 (B#6669) were provided by the Bloomington *Drosophila* Stock Center (BDSC, Indiana University, Bloomington, IN). En-Gal4, UAS-GFP was provided by J. Skeath (Washington University in St. Louis, St. Louis, MO). Nrg-GFP; TJ-Gal4; Indy-GFP was provided by S. Horne-Badovinac

(University of Chicago, Chicago, IL). ECs were staged according to morphological features described by King (Cummings and King, 1969) by matching z-stack projected area of each EC to the expected areas across stages. For ovariole experiments, 1-3 day-old females were aged on yeast for 1-3 days before dissection. Ovaries were dissected as described Prasad et al. in Live Imaging Media (Schneider's S2 Media, 0.6x Pen/Strep, 15% vol/vol fetal bovine serum (FBS), 0.2mg/mL insulin) and stained for 10min with CellMask Deep Red Plasma Membrane Stain (Thermo-Fisher) at 1:1000 dilution in Live Imaging Media. Cell size measurements were made at the basal surface. Border cell delamination was defined as the time-point where a distinct cluster of cells emerged from the surrounding single-layered follicular epithelium.

Immunofluorescence

Cells were fixed in 4% paraformaldehyde (VWR) in PBS, permeabilized with 0.5% triton-X100 (Sigma) in PBS, blocked in 1% BSA (Sigma) PBS overnight at 4C, incubated with primary antibodies overnight at 4C, and incubated with secondary bodies for 2 hours at room temperature with 5 minute washes in PBS in between. Ovarioles were fixed for 40min in 4% Paraformaldehyde in PBS, followed by a 5min wash in PBX (PBS with 0.1% Triton X-100), 2x 20min washes in PAXD (PBS with 1% bovine serum albumin, 0.3% Triton X-100, and 0.3% deoxycholate), and one 20 min wash in PAXDG (PAXD with 5% normal goat serum). The mean level of pMLC staining intensity per cell within the density gradient was calculated by normalizing the total intensity of pMLC signal within a region by the number of cells within that region at 13 points along the density gradient. All regions contained at least 200 cells. Ovarioles were incubated overnight at 4°C in primary antibody diluted in PAXDG and washed three times in PBX at room temperature. Ovarioles were then incubated overnight at 4°C in secondary

antibody diluted in PAXDG, washed twice in PBX, and once in PBS, all at room temperature. Prepared tissues were mounted in Vectashield mounting media with or without DAPI (Vector Laboratories, Burlingame, CA). See Supplemental Experimental Procedures for antibodies and concentrations.

Imaging

Cells: Live cell and fixed epifluorescence imaging was performed on a Nikon Ti-E microscope with a humidity and temperature controlled incubation chamber (LiveCell). Data were collected from 2.3 mm^2 regions by tiling 4 10X regions of view. Time-lapse imaging was performed with 60-minute duration and 2-minute frame intervals. Reference images of the dish surface were taken at each frame and field of view to correct for stage drift. *Drosophila*: Timelapse imaging was performed on either an inverted Zeiss LSM510 or upright LSM700 confocal microscope over 3.5 hours with 5-minute frame intervals. See also: Supplemental Experimental Procedures.

Imaging Analysis and Statistics

Image analysis was performed using Matlab (The Mathworks, Natick MA) and ImageJ. Cellular motion was measured with Particle Image Velocimetry for monolayers (Thielicke and Stamhuis, 2014), or Optical Flow for *Drosophila* (Vig et al., 2016). Mean cell radius was computed for each field of view using custom software to segment Hoechst 33342 stained nuclei and is defined as the radius of a circle with area equivalent to the mean cell area. Correlation computations were performed using custom Matlab software. Curve fitting was performed using Matlab software and data beyond a distance of 40 cell radii was not used to fit Eq. 1 due to departure

from the form of the fitting function beyond this distance. See also: Supplemental Experimental Procedures.

2.6 AUTHOR CONTRIBUTIONS

A.J.L. and G.V.S. conceived experimental design, performed experiments, and interpreted results. A.J.L., S.K., and B.L. designed, wrote, and performed computational simulations. A.J.L. wrote the manuscript. G.D.L and S.X.S conceived experimental design, interpreted results, supervised work, and edited the manuscript.

2.7 ACKNOWLEDGEMENTS

We would like to thank the Horne-Badovinac lab for readily providing fly lines and reagents. This work was supported by NIH Grants GM080673 to G.D.L. and 1U54CA143868 to G.D.L and S.X.S. F30 CA180244 and T32 GM007200 to G.V.S., and T32 EB018266 and T32 GM007200 to A.J.L.

2.8 REFERENCES

Abercrombie, M. (1970). Contact inhibition in tissue culture. In vitro *6*, 128-142.

Angelini, T.E., Hannezo, E., Trepat, X., Fredberg, J.J., and Weitz, D.A. (2010). Cell migration driven by cooperative substrate deformation patterns. Phys Rev Lett *104*, 168104.

Angelini, T.E., Hannezo, E., Trepat, X., Marquez, M., Fredberg, J.J., and Weitz, D.A. (2011). Glass-like dynamics of collective cell migration. Proc Natl Acad Sci U S A *108*, 4714-4719.

Bajpai, S., Correia, J., Feng, Y., Figueiredo, J., Sun, S.X., Longmore, G.D., Suriano, G., and Wirtz, D. (2008). {alpha}-Catenin mediates initial E-cadherin-dependent cell-cell recognition and subsequent bond strengthening. Proc Natl Acad Sci U S A *105*, 18331-18336.

Bi, D., Lopez, J.H., Schwarz, J.M., and Manning, M.L. (2015). A density-independent rigidity transition in biological tissues. Nature Physics *11*, 1074-1079.

Bi, D., Yang, X., Marchetti, M.C., and Manning, M.L. (2016). Motility-driven glass and jamming transitions in biological tissues. Physical Review X *6*, 021011.

Bouchard, C., Thieke, K., Maier, A., Saffrich, R., Hanley‐Hyde, J., Ansorge, W., Reed, S., Sicinski, P., Bartek, J., and Eilers, M. (1999). Direct induction of cyclin D2 by Myc contributes to cell cycle progression and sequestration of p27. The EMBO journal *18*, 5321-5333.

Cetera, M., and Horne-Badovinac, S. (2015). Round and round gets you somewhere: collective cell migration and planar polarity in elongating Drosophila egg chambers. Current opinion in genetics & development *32*, 10-15.

Cetera, M., Ramirez-San Juan, G.R., Oakes, P.W., Lewellyn, L., Fairchild, M.J., Tanentzapf, G., Gardel, M.L., and Horne-Badovinac, S. (2014). Epithelial rotation promotes the global alignment of contractile actin bundles during Drosophila egg chamber elongation. Nature communications *5*.

Chen, C.S., Tan, J., and Tien, J. (2004). Mechanotransduction at cell-matrix and cell-cell contacts. Annu Rev Biomed Eng *6*, 275-302.

Cohen, D.J., Nelson, W.J., and Maharbiz, M.M. (2014). Galvanotactic control of collective cell migration in epithelial monolayers. Nature materials *13*, 409-417.

Cooper, G. (2000). The Cell: A Molecular Approach, 2 edn (Sunderland, MA: Sinauer Associates).

Coso, O.A., Chiariello, M., Yu, J.C., Teramoto, H., Crespo, P., Xu, N., Miki, T., and Gutkind, J.S. (1995). The small GTP-binding proteins Rac1 and Cdc42 regulate the activity of the JNK/SAPK signaling pathway. Cell *81*, 1137-1146.

Cummings, M., and King, R. (1969). The cytology of the vitellogenic stages of oogenesis in Drosophila melanogaster. I. General staging characteristics. Journal of morphology *128*, 427-441.

Das, T., Safferling, K., Rausch, S., Grabe, N., Boehm, H., and Spatz, J.P. (2015). A molecular mechanotransduction pathway regulates collective migration of epithelial cells. Nat Cell Biol *17*, 276-287.

Deforet, M., Hakim, V., Yevick, H., Duclos, G., and Silberzan, P. (2014). Emergence of collective modes and tri-dimensional structures from epithelial confinement. Nature communications *5*.

Desai, R.A., Gopal, S.B., Chen, S., and Chen, C.S. (2013). Contact inhibition of locomotion probabilities drive solitary versus collective cell migration. Journal of The Royal Society Interface *10*, 20130717.

Ewald, A.J., Brenot, A., Duong, M., Chan, B.S., and Werb, Z. (2008). Collective epithelial migration and cell rearrangements drive mammary branching morphogenesis. Dev Cell *14*, 570-581.

Fletcher, A.G., Osterfield, M., Baker, R.E., and Shvartsman, S.Y. (2014). Vertex models of epithelial morphogenesis. Biophysical journal *106*, 2291-2304.

Friedl, P., and Gilmour, D. (2009). Collective cell migration in morphogenesis, regeneration and cancer. Nat Rev Mol Cell Biol *10*, 445-457.

Garcia, S., Hannezo, E., Elgeti, J., Joanny, J.-F., Silberzan, P., and Gov, N.S. (2015). Physics of active jamming during collective cellular motion in a monolayer. Proceedings of the National Academy of Sciences *112*, 15314-15319.

Haas, P., and Gilmour, D. (2006). Chemokine signaling mediates self-organizing tissue migration in the zebrafish lateral line. Developmental cell *10*, 673-680.

Haigo, S.L., and Bilder, D. (2011). Global tissue revolutions in a morphogenetic movement controlling elongation. Science *331*, 1071-1074.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Hiscox, S., and Jiang, W. (1996). Expression of E-cadherin, alpha, beta and gamma-catenin in human colorectal cancer. Anticancer research *17*, 1349-1354.

Horne-Badovinac, S. (2014). The Drosophila egg chamber—a new spin on how tissues elongate. Integrative and comparative biology *54*, 667-676.

Huttenlocher, A., Lakonishok, M., Kinder, M., Wu, S., Truong, T., Knudsen, K.A., and Horwitz, A.F. (1998). Integrin and cadherin synergy regulates contact inhibition of migration and motile activity. The Journal of cell biology *141*, 515-526.

Kabla, A.J. (2012). Collective cell migration: leadership, invasion and segregation. J R Soc Interface *9*, 3268-3278.

Katz, Y., Tunstrøm, K., Ioannou, C.C., Huepe, C., and Couzin, I.D. (2011). Inferring the structure and dynamics of interactions in schooling fish. Proceedings of the National Academy of Sciences *108*, 18720-18725.

Koride, S., He, L., Xiong, L.P., Lan, G., Montell, D.J., and Sun, S.X. (2014). Mechanochemical regulation of oscillatory follicle cell dynamics in the developing Drosophila egg chamber. Mol Biol Cell *25*, 3709-3716.

Krawczyk, W.S. (1971). A pattern of epidermal cell migration during wound healing. J Cell Biol *49*, 247-263.

Kuriyama, S., Theveneau, E., Benedetto, A., Parsons, M., Tanaka, M., Charras, G., Kabla, A., and Mayor, R. (2014). In vivo collective cell migration requires an LPAR2-dependent increase in tissue fluidity. J Cell Biol *206*, 113-127.

Li, B., and Sun, S.X. (2014). Coherent motions in confluent cell monolayer sheets. Biophysical journal *107*, 1532-1541.

Mayor, R., and Carmona-Fontaine, C. (2010). Keeping in touch with contact inhibition of locomotion. Trends in cell biology *20*, 319-328.

Milnor, J. (1978). Analytic proofs of the" hairy ball theorem" and the Brouwer fixed point theorem. The American Mathematical Monthly *85*, 521-524.

Nagafuchi, A., Ishihara, S., and Tsukita, S. (1994). The roles of catenins in the cadherin-mediated cell adhesion: functional analysis of E-cadherin-alpha catenin fusion molecules. The Journal of Cell Biology *127*, 235-245.

Nguyen-Ngoc, K.-V., Cheung, K.J., Brenot, A., Shamir, E.R., Gray, R.S., Hines, W.C., Yaswen, P., Werb, Z., and Ewald, A.J. (2012). ECM microenvironment regulates collective migration and local dissemination in normal and malignant mammary epithelium. Proceedings of the National Academy of Sciences *109*, E2595-E2604.

Nnetu, K.D., Knorr, M., Pawlizak, S., Fuhs, T., and Käs, J.A. (2013). Slow and anomalous dynamics of an MCF-10A epithelial cell monolayer. Soft Matter *9*, 9335.

Plutoni, C., Bazellieres, E., Le Borgne-Rochet, M., Comunale, F., Brugues, A., Seveno, M., Planchon, D., Thuault, S., Morin, N., Bodin, S.*, et al.* (2016). P-cadherin promotes collective cell migration via a Cdc42-mediated increase in mechanical forces. J Cell Biol *212*, 199-217.

Poujade, M., Grasland-Mongrain, E., Hertzog, A., Jouanneau, J., Chavrier, P., Ladoux, B., Buguin, A., and Silberzan, P. (2007). Collective migration of an epithelial monolayer in response to a model wound. Proc Natl Acad Sci U S A *104*, 15988-15993.

Puliafito, A., Hufnagel, L., Neveu, P., Streichan, S., Sigal, A., Fygenson, D.K., and Shraiman, B.I. (2012). Collective and single cell behavior in epithelial contact inhibition. Proc Natl Acad Sci U S A *109*, 739-744.

Sadati, M., Taheri Qazvini, N., Krishnan, R., Park, C.Y., and Fredberg, J.J. (2013). Collective migration and cell jamming. Differentiation *86*, 121-125.

Serra-Picamal, X., Conte, V., Vincent, R., Anon, E., Tambe, D.T., Bazellieres, E., Butler, J.P., Fredberg, J.J., and Trepat, X. (2012). Mechanical waves during tissue expansion. Nature Physics *8*, 628-634.

Shiozaki, H., Iihara, K., Oka, H., Kadowaki, T., Matsui, S., Gofuku, J., Inoue, M., Nagafuchi, A., Tsukita, S., and Mori, T. (1994). Immunohistochemical detection of alpha-catenin expression in human cancers. The American journal of pathology *144*, 667.

Szabo, A., Unnep, R., Mehes, E., Twal, W.O., Argraves, W.S., Cao, Y., and Czirok, A. (2010). Collective cell motion in endothelial monolayers. Phys Biol *7*, 046007.

Thielicke, W., and Stamhuis, E. (2014). PIVlab–Towards user-friendly, affordable and accurate digital particle image velocimetry in MATLAB. Journal of Open Research Software *2*.

Trappe, V., Prasad, V., Cipelletti, L., Segre, P., and Weitz, D.A. (2001). Jamming phase diagram for attractive particles. Nature *411*, 772-775.

Trepat, X., and Fredberg, J.J. (2011). Plithotaxis and emergent dynamics in collective cellular migration. Trends Cell Biol *21*, 638-646.

Vedel, S., Tay, S., Johnston, D.M., Bruus, H., and Quake, S.R. (2013). Migration of cells in a social context. Proc Natl Acad Sci U S A *110*, 129-134.

Vig, Dhruv K., Hamby, Alex E., and Wolgemuth, Charles W. (2016). On the Quantification of Cellular Velocity Fields. Biophysical Journal *110*, 1469-1475.

Wang, Y., and Riechmann, V. (2007). The role of the actomyosin cytoskeleton in coordination of tissue growth during Drosophila oogenesis. Current biology *17*, 1349-1355.

2.9 FIGURES

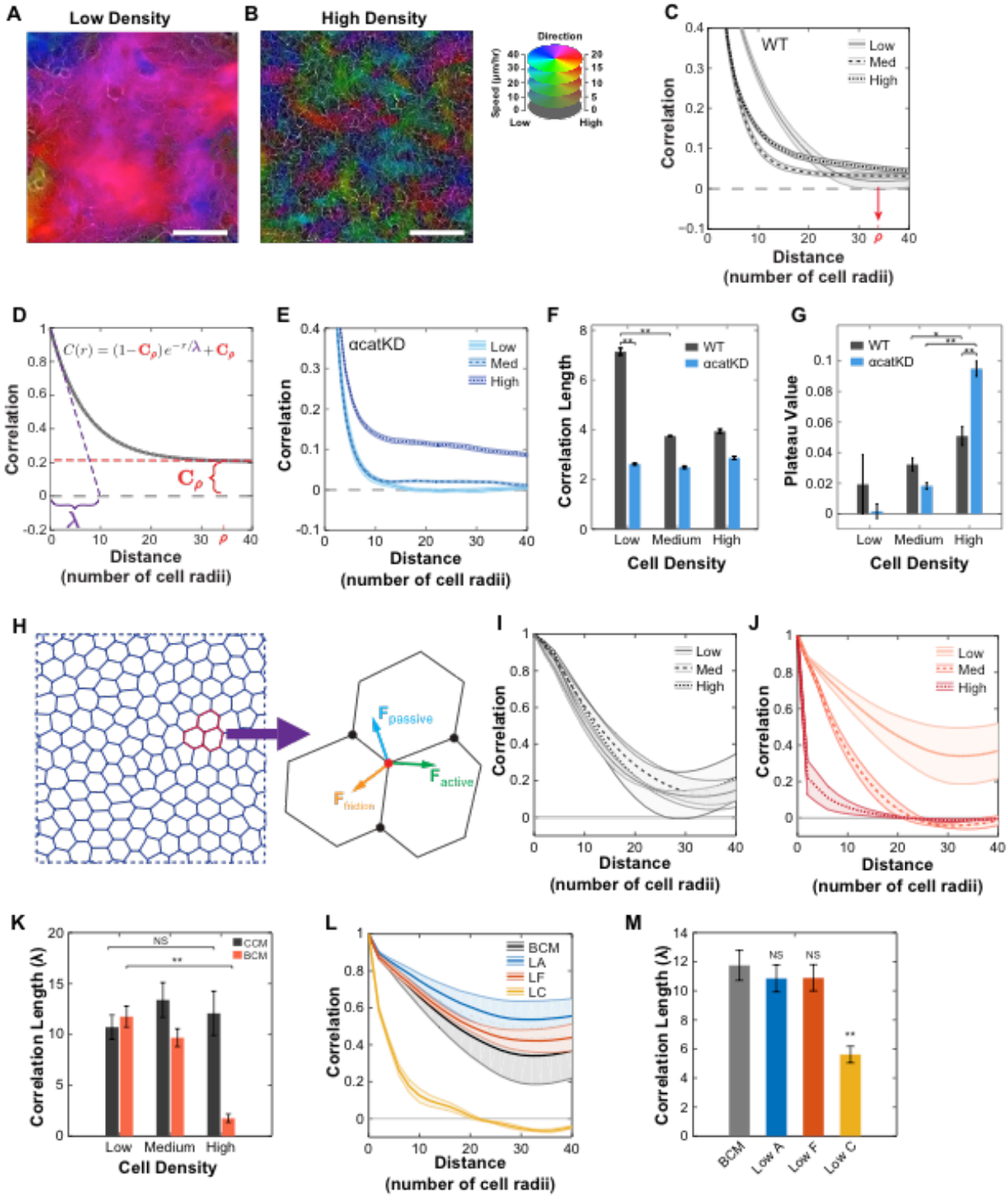


Figure 2.1: Local and regional organization of collective migration within an epithelium.

A. Heatmap overlay of cell velocity for low density WT cells with migration direction and speed encoded with color hue and intensity, respectively. B. Same as A for high-density WT cells. Scale bars for A and B 100 μm. C. Radial correlation for WT cells across a range of densities. Red arrow denotes ρ, the distance over which radial correlation decayed to zero for low-density WT cells. D. Diagram of fitting function used to measure correlation length (purple, λ) and plateau value (red, C_ρ). E. Radial correlation for acatKD cells across a range of densities. F. Value of correlation length (λ) obtained by fitting curves to Eq. 1. G. Value of plateau value (C_ρ) obtained by fitting curves to Eq. 1. H. Diagram of the vertex model and vertex force-balance used to simulate epithelial migration. I. Radial correlation curves for simulated epithelial migration using the constant contractility model. J. Radial correlation curves for simulated epithelial migration using the biomechanical contractility model. K. Value of correlation length (λ) obtained by fitting constant contractility (CCM) and biomechanical contractility (BCM) correlation curves. L. Radial correlations for simulated epithelial migration at low density for the biomechanical contractility model (BCM) and variations of the model: low adhesion (LA), low friction (LF), and low contractility coupling (LC). M. Value of correlation length (λ) for the various models in L. Error bars and envelopes are SEM. C,E-G: Each density is the mean of 14 (WT) or 6 (αcatKD) 2.3 mm^2 regions (four 10x fields per region). I-M: data are mean values from 3 simulations. See Tables S2 and S3 for simulated cell type parameters. Student's t-test used to assess significance. NS = not significant; * = $p<0.05$; ** = $p<0.01$. See also Videos S1 and S2.

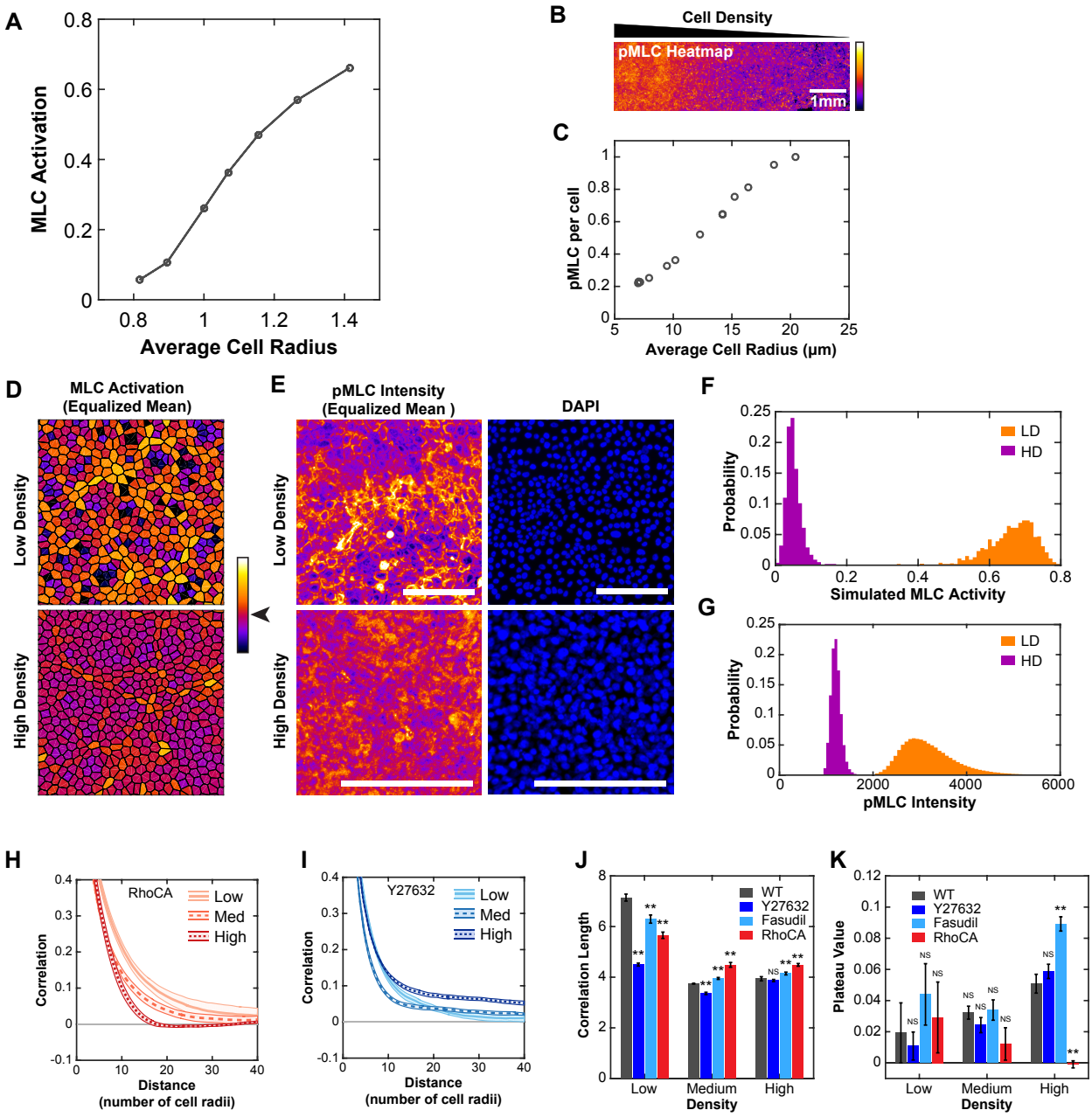


Figure 2.2: Actomyosin contractility drives collective behavior.

A. Mean level of simulated MLC activity across a range of densities. B. Intensity of pMLC immunofluorescence staining across a gradual density gradient of WT cells (high on left, low on right). C. Quantification of mean pMLC immunofluorescence signal per cell, normalized to the lowest density measurement. D. Simulated MLC activity for 500 individual cells at low and high density scaled such that mean activity levels are equal (mean value color denoted by arrowhead

on colorbar). E. Intensity of pMLC immunofluorescence staining of regions of approximately 500 cells at low and high density (magnified from B) and scaled such that mean pMLC intensity levels are equal. Scale bars are all 100 um. F. Histogram of absolute (without equalizing means) MLC activity for cells from D. G. Histogram of absolute pMLC immunofluorescence intensity for the regions in E. H. Radial correlation curves for RhoCA cells. I. Radial correlation curves for Y27632 treated WT cells. J. Correlation length (λ) for WT, RhoCA, and ROCK inhibition. K. Plateau value (C_ρ) for WT, RhoCA, and ROCK inhibition. Each density is the mean value from 14 (WT), 6 (Y27632), 6 (Fasudil), or 8 (RhoCA) 2.3 mm^2 regions (4 10x fields per region). Error bars and envelopes are SEM. Student's t-test used to assess significance. NS = not significant; * = $p<0.05$; ** = $p<0.01$.

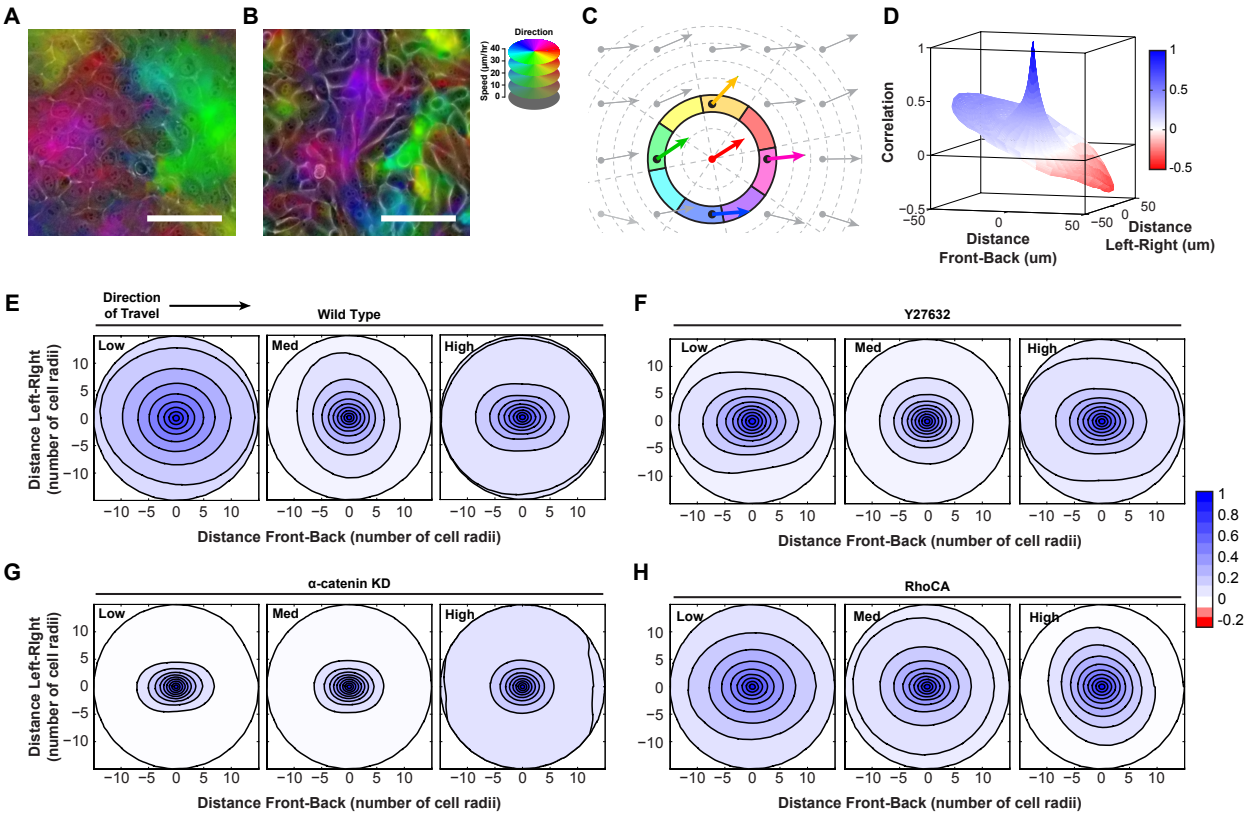


Figure 2.3: Cell density and contractility control transitions in collective shape.

A. Contracting region in WT cells at low-density (migration direction and speed correspond to color hue and intensity, respectively). Magenta and green regions are converging. B. Streaming motion in α-catenin depleted (αcatKD) cells at low density. Purple region elongated in direction of cell migration. C. Diagram of directional correlation. D. Representative directional correlation from a single field of view for low-density WT epithelial cells. *E-H.* Experimental polar correlation function at low, medium, and high density for WT, αcatKD, Y27632, and RhoCA epithelial cells respectively. Each density is the mean value from 14 (WT), 6 (Y27632), or 8 (RhoCA) regions (4 fields of view per region). See also Videos S3 and S4.

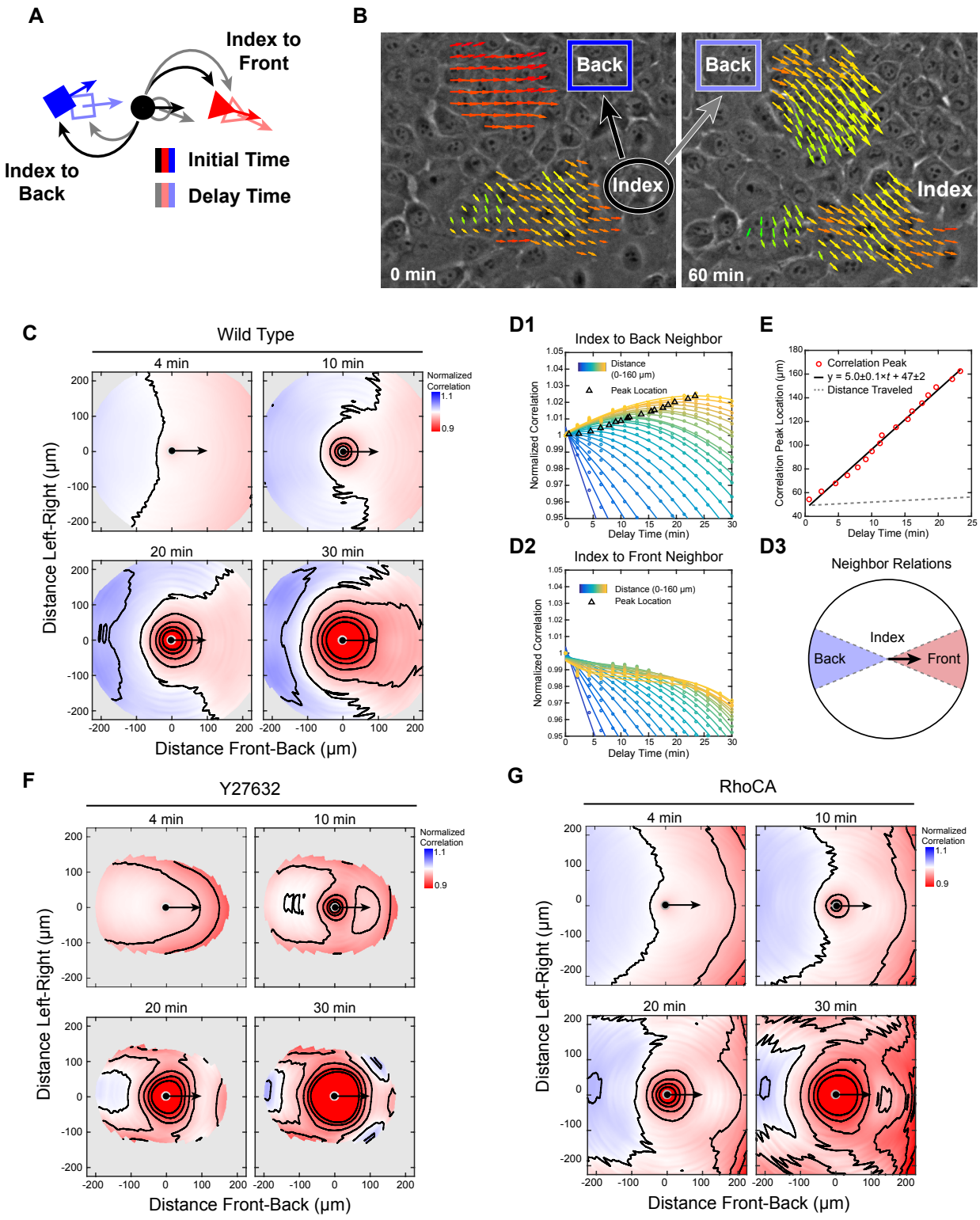


Figure 2.4: Dynamics of collective migration: directional transfer of information from leaders to followers.

A. Diagram of spatiotemporal correlation. Neighbors are defined by distance and direction from an index position (solid shapes). Correlations are computed at the initial time (black arrows) and as a function of delay time (gray arrows) as the neighbor positions migrate (outlined shapes). B. Representative leader-follower relationship. Neighbors (dark blue box) positioned behind an index group are initially not aligned (black arrow) to the index group, but increase alignment (gray arrow) over time. C. Spatiotemporal correlation normalized to zero delay time values for WT low-density cells at increasing delay times. D1. Normalized spatiotemporal correlation as a function of delay time for index-to-back neighbors of increasing distance from the index point (points) with cubic fit (lines). D2. Same as D1 for index-to-front neighbors. D3. Diagram of index-to-back and index-to front neighbors. E. Delay time for peak normalized correlation versus neighbor distance, linear fit (black line), and mean distance traveled (dashed gray line). F. Same as C for Y27632 treated low-density WT cells. G. Same as C for low-density RhoCA cells. Correlations for each cell type are the mean value from 7 (WT), 6 (Y27632), or 8 (RhoCA) 2.3 mm^2 regions (4 10x fields per region).

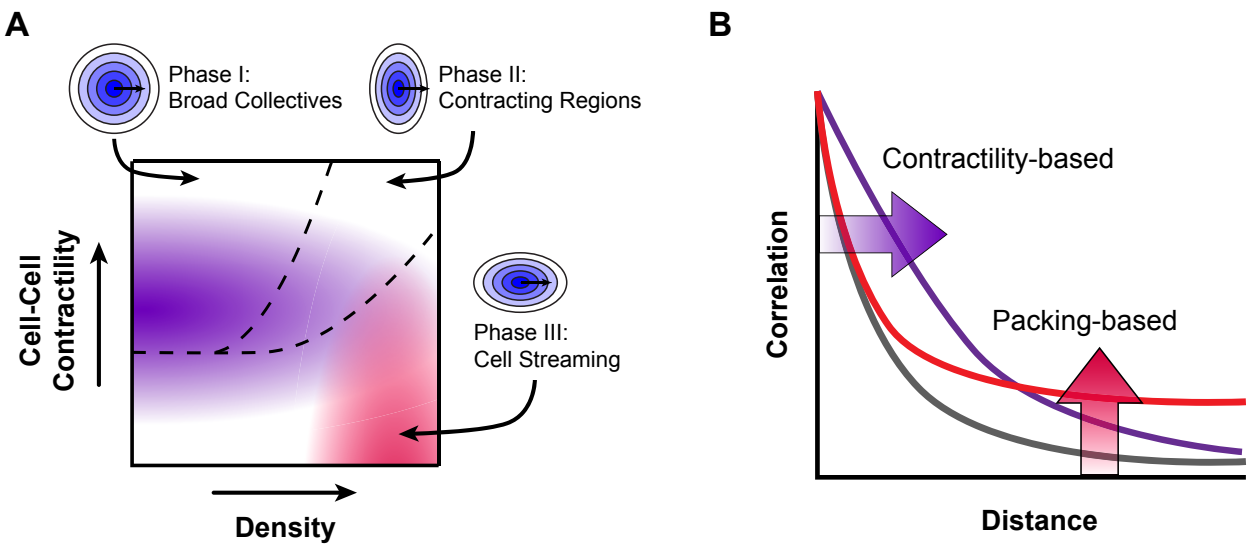


Figure 2.5: Phases of collective migration as a function of cell-density and contractility. A. Phase diagram depicting collective size in terms of local organization (purple) and regional organization (red) as well as collective shape (dashed black lines). B. Diagram of radial correlation for local organization (purple) and regional organization (red).

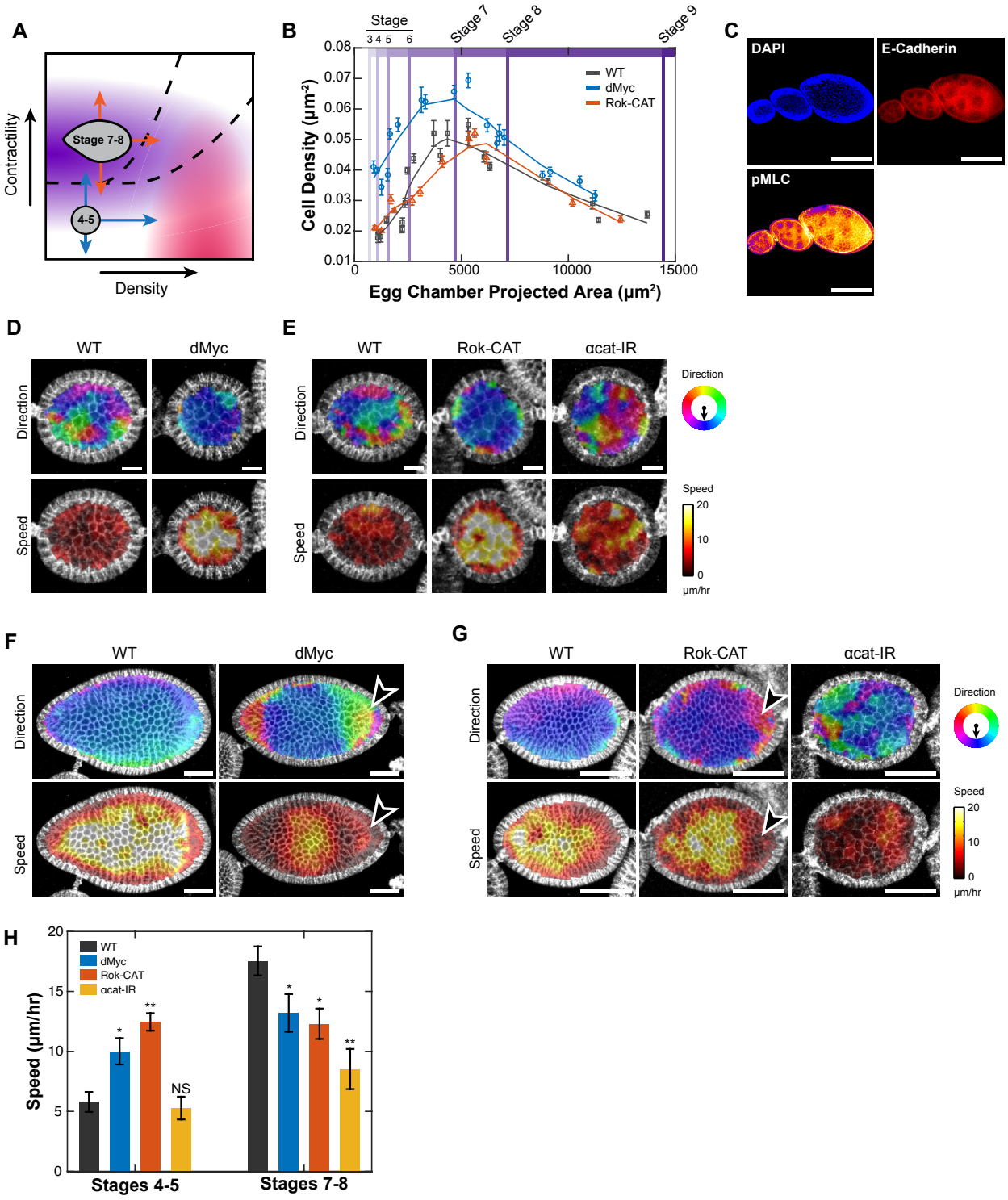


Figure 2.6: Prediction of in-vivo collective migration.

A. Phase diagram showing hypothesized states of early (stage 4-5) and late (stage 7-8) follicular epithelium. Predicted phase transitions denoted by blue and orange arrows. B. Cell density as a

function of projected egg chamber area and stage for developing WT, dMyc, and Rok-CAT *Drosophila* eggs. C. Fixed WT egg chambers at three stages of development (5, 7, and 8 from left to right) stained for DAPI, anti-E-cadherin and anti-pMLC. D. Migration direction (top) and speed (bottom) of WT versus dMyc stage 5 egg chambers. Measurements show speed and direction of migration over a 5-minute interval. E. Migration direction (top) and speed (bottom) of WT versus Rok-CAT and αcat-IR expressing stage 5 egg chambers. F. Migration direction (top) and speed (bottom) of WT and dMyc expressing stage 8 egg chambers. Arrowhead denotes region of decreased speed and organization for Myc follicular epithelium. G. Migration direction (top) and speed (bottom) for WT versus Rok-CAT and acat-IR expressing stage 7 egg chambers. Arrowhead denotes region of decreased speed and organization in Rok-CAT follicular epithelium. H. Mean migration speed of follicular epithelial cells. Student's t-test used to assess significance. NS = not significant; * = $p<0.05$; ** = $p<0.01$.

2.10 SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cell Culture

MCF10A cells were purchased from the ATCC and were cultured in 1:1 DMEM:F12 (Gibco) supplemented with 5% Horse Serum (Gibco), 10ug/mL Insulin (Sigma-Aldrich), 500 ng/mL Hydrocortisone (Sigma-Aldrich), 100 ng/mL Cholera Toxin (Sigma-Aldrich), 20 ng/mL EGF (Gibco), and Penicillin/Streptomycin (Gibco). 293T cells (ATCC) used for Lentiviral production were cultured in DMEM (Gibco) supplemented with 10% heat-inactivated FBS. In regular culture MCF10A cells were grown to 75% confluence, dissociated using Trypsin-EDTA 0.05% (Gibco), and split 1:5.

Immunofluorescence

Cells were incubated overnight at 4C with rabbit anti-pMLC (Cell Signaling, 3671S, 1:100). Secondary antibodies used were mouse anti-rabbit Alexa Fluor 488 and 568 (Invitrogen, 1:500) with appropriate phalloidin 488 or 568 (Molecular Probes, 1:100). Slides were mounted using Vectasheild mounting medium containing DAPI (Vector Labs) for imaging. The following primary antibodies were used for ovariole staining: rat anti-DE-cadherin (1:20) from Developmental Studies Hybridoma Bank (University of Iowa, Iowa City, IA), and rabbit anti-pMLC Ser19 (1:100; Cell Signaling, Beverly, MA). Secondary antibodies were Alexa 488 and 568 (Invitrogen) and Cy5 (Jackson ImmunoResearch).

Western Blotting

Cells were lysed in RIPA buffer (50mM Tris pH 8.0, 150mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) with 2μg/mL Aprotinin/Leupeptin, 2μM Pepstatin A, 200nM PMSF, 1mM

Na_3VO_4, and 2mM NaF, followed by centrifugation at 12,000 rpm for 10 minutes. Lysate concentration was determined by Bradford assay (Bio-Rad). Equal protein amounts were boiled in SDS containing sample buffer, resolved by 10% SDS-PAGE, and transferred to PVDF membranes (Millipore) in transfer buffer (25mM Tris, 192mM glycine, 5% methanol). Membranes were blocked with TBST (25mM Tris pH 7.4, 150mM NaCl, 2mM KCl, 0.5% Tween-20) containing 5% BSA and incubated overnight with primary antibodies. Detection was performed by incubating with HRP-conjugated secondary antibodies and developing with SuperSignal West Pico ECL (Pierce). Images were collected on a Bio-Rad ChemiDoc XRS+. The following antibodies were used: rabbit anti-alpha-Catenin (Sigma, C2081) and mouse anti-beta-Tubulin (Sigma, Tub 2.1).

Imaging

The Nikon Ti-E microscope was controlled by MetaMorph software (version 7.7.0.0, Molecular Devices). Images were acquired using metal halide lamp illumination (Prior), Semrock Brightline filter sets, a Nikon Pan Fluor 10x objective (NA 0.3, WD 16mm), and a cooled CCD camera (CoolSNAP HQ2, Photometrics). The LSM510 was controlled by Zeiss software. Images were acquired with either a C-Apochromat 40X (1.2 NA, water immersion) or a Plan-Apochromat 20X (0.75 NA) objective, with illumination from an Argon ion laser (458, 477, 488, 514 nm lines) or HeNe laser (633 nm line). The LSM700 was controlled by Zeiss software and images were acquired using a 10X Neofluor air (NA 0.3) lens. After dissection, ovarioles were washed for 5min in Live Imaging Media, and transferred to an imaging slide. For the upright LSM700, an agar pad (Live imaging media plus 0.4% NuSieve GTG Low-melt agarose) was made on top of a gas permeable membrane (Sarstedt, Lumox, 1-well, PCA#94.6150.101, well removed from membrane). A coverslip was placed on top of the agarose pad plus sample and vacuum grease was used to secure the coverslip

at its four corners. Halocarbon oil 27 (Sigma) was around to fill in the remaining free space under the coverslip to minimize evaporation of the Live Imaging Media. For the inverted LSM510, a custom aluminum imaging slide was made. This consisted of a 75x26x2mm aluminum sheet with two centered 20mm diameter holes. A 22x50mm coverglass was fixed to the bottom of this aluminum slide by sandwiching a sheet of parafilm (22x50mm with holes cut in the center corresponding to the holes in the aluminum slide) between the coverglass and aluminum slide and heating on a 100C hotplate for 30 seconds. Ovarioles in a small volume of live imaging media were placed in one of the wells formed by attaching the coverslip and the remaining volume was filled in with agar prepared as described above. The second well was filled with water and a gas permeable membrane was placed over both wells for humidity control.

Image Analysis

For PIV measurements, velocity fields were computed using three passes of 64, 32, and 16 pixels windows and 50% overlap to generate a displacement field with 5.2 um spacing. The resulting velocity field, produced by dividing by time between frames, was first corrected for stage drift by subtracting out any motion in the reference image and then smoothed using a boxcar moving average filter of width equal to 3 frames. Reference image motion was computed by using whole-image cross-correlation and subpixel peak finding. Notably, the accurate measurement of C_ρ (regional organization) depends on large imaging fields and directly compensating for microscope drift because analysis of the velocity fluctuation field $(u = v - \langle v \rangle)$ would directly alter this long-scale order. Optical Flow measurements were performed on maximum intensity projection images of z-stacks corresponding to half the thickness of an EC. Velocity fields were computed at a spatial resolution of 2.3 um using a Gaussian blur filter of 1.5 pixel standard deviation. Velocity fields

were smoothed using a boxcar moving average filter of width equal to 3 frames. Prior to velocity measurements, each video was-de-drifted using the Matlab function imregister, which employs intensity-based automatic image registration. Egg chamber projected area was determined by automated segmentation of egg chambers followed by manual editing to correct errors in separating touching ECs. Cell area was measured manually in ImageJ. Correction for non-uniform illumination was performed when tiling immunofluorescence images by using data from an empty field of view. pMLC staining intensity across the density gradient was quantified by measuring the total amount of pMLC signal at 12 uniformly spaced locations and normalizing by the total number of nuclei within that region (number of nuclei >100 at each location).

2.11 VERTEX MODEL OF COLLECTIVE CELL MIGRATION

Cells are represented by two dimensional polygons with shared edges and vertices as shown in. Migration of cells is governed by the motion of their vertices according to the following equation of motion:

$$\eta \frac{\mathrm{d}\boldsymbol{r}_i}{\mathrm{d}t} + \boldsymbol{F}_{fi} = \boldsymbol{F}_{ei} + \boldsymbol{F}_{ai}, \tag{1}$$

Where η is the friction coefficient between cells and their substrate, $\boldsymbol{r}_i$ is the position of the i^{th} vertex, t is time, and $\boldsymbol{F}_{fi}$, $\boldsymbol{F}_{ei}$, and $\boldsymbol{F}_{ai}$ denote friction, passive, and active forces, respectively, acting on the i^{th} vertex.

Passive Force

The passive force, $\boldsymbol{F}_{ei}$, is related to mechanics of the cell and cell-cell adhesion. It can be modeled using an energy-like function U_e:

$$\boldsymbol{F}_{ei} = -\frac{\partial U_e}{\partial \mathbf{r}_i}, \tag{2}$$

Where U_e is defined as:

$$U_e = \sum_{J=1}^{N} \frac{K}{2} \left(A_J - A_0\right)^2 + \sum_{i,j} \Lambda l_{ij}. \tag{3}$$

Here, J labels the cell, K is the area elastic modulus, Λ is the adhesion energy per unit length, A_J is the current area of cell J, A_0 is the preferred cell area, and $l_{ij} = |\boldsymbol{r}_i - \boldsymbol{r}_j|$ denotes the edge length between vertex i and its neighbor j.

Active Force

The active force, $\boldsymbol{F}_{ai}$, is the sum of three components: contractile force ($\boldsymbol{F}_{ci}$), persistent force ($\boldsymbol{F}_{pi}$), and random force ($\boldsymbol{F}_{Ri}$):

$$\boldsymbol{F}_{ai} = \boldsymbol{F}_{ci} + \boldsymbol{F}_{pi} + \boldsymbol{F}_{Ri}. \tag{4}$$

The contractile force, $\boldsymbol{F}_{ci}$, is computed as $\boldsymbol{F}_{ci} = -\partial U_c / \partial \mathbf{r}_i$ where the contractile energy function U_c is defined as:

$$U_c = \sum_{J=1}^{N} \frac{\Gamma_J}{2} L_J^2. \tag{5}$$

Here, L_J is the perimeter of cell J, and Γ_J denotes the level of contractility. Depending on the contractility model employed, the value of Γ_J is either set to a constant (*constant contractility*) or is determined by the output of a signaling model of stretch-activated force generation analogous to Rho-ROCK-Myosin signaling (*biomechanical contractility*). In this simple model of active contractility, two components are considered: an upstream activator, ρ, that responds to the degree to which a cell is stretched, and a downstream effector, M, that is activated by ρ and leads to contractile force generation. Here, ρ is analogous to Rho or ROCK and M is analogous to phosphorylated Myosin Light Chain. This system is modeled mathematically as a system of differential equations. The level of the activator ρ in the J-th cell obeys:

$$\frac{\mathrm{d}\rho_J}{\mathrm{d}t} = A_\rho H(s_J) \frac{s_J^m}{K_s + s_J^m} (1 - \rho_J) - D_\rho \rho_J, \tag{6}$$

Where H is the Heaviside step function, A_ρ is the maximum activation rate of ρ_J, $s_J = L_J - L_0$ is the current state of stretch of the cell, and K_s is the half-maximal response constant. The variable L_0 is the preferred cell perimeter and is equal to the perimeter of a regular hexagon with area A_0.

The activation of the effector M_J follows:

$$\frac{\mathrm{d}M_J}{\mathrm{d}t} = A_m \rho_J (1 - M_J) - D_M M_J \tag{7}$$

where M_J is the relative level of the activated force generating component in the J-th cell. M_J is related to the level of contractility via:

$$\Gamma_J = k M_J, \tag{8}$$

Where k is a contractility coupling constant. In the *constant contractility* model, the value of kM_J is fixed at a constant value.

The persistent force on the i-th vertex, $\boldsymbol{F}_{pi}$, is defined as:

$$\boldsymbol{F}_{pi} = \alpha \frac{\int_{-\infty}^{t} e^{-\beta(t-t)} \boldsymbol{v}_i(t) \mathrm{d}t}{\left| \int_{-\infty}^{t} e^{-\beta(t-t)} \boldsymbol{v}_i(t) \mathrm{d}t \right|} \tag{9}$$

where $\boldsymbol{v}_i = \mathrm{d}\boldsymbol{r}_i / \mathrm{d}t$, α is the strength of the persistent force, and β is a constant that determines the decay rate of persistence. The persistent force arises from protrusions developed by the cell and the decay in time reflects the time of actin-assembly and disassembly.

The random force reflects the fact that an individual cell will not move indefinitely in one direction and satisfies the following two relations:

$$\langle \boldsymbol{F}_{Ri} \rangle = \mathbf{0} \tag{10}$$

$$\langle F_{Rq}(t) F_{Rs}(t) \rangle = \sigma^2 \delta(t) \delta_{qs}, \tag{11}$$

where σ is a parameter that specifies the magnitude of variance, and $\delta(t)$ and δ_{qs} are Dirac's and Kroneker's δ-functions, respectively.

Friction Force

The frictional force on each vertex, $\boldsymbol{F}_{fi}$, models the friction between a cell, its neighbors, and the substrate. The cell-substrate friction coefficient is η_s, see Eq. (1), and was taken as a constant for all simulations. The friction between a cell and its neighbors is computed using a finite volume method. The cell-cell friction force on a given vertex is computed as the total shear force on the area element defined by the centers of neighboring cells and midpoints of neighboring edges. The shear force is assumed to arise from fluid-like behavior of cells and can be calculated from the deviatoric stress within the volume element:

$$\sigma_s = \eta_c \left(\nabla \boldsymbol{v} + (\nabla \boldsymbol{v})^T - \frac{2}{3} \nabla \cdot \boldsymbol{v} \boldsymbol{I} \right) \tag{12}$$

$$\boldsymbol{F}_{fc} = \oint \sigma_s \cdot \boldsymbol{n} \mathrm{d}S \tag{13}$$

where η_c is the viscosity of cells, $\nabla \boldsymbol{v}$ is the velocity gradient within the volume element, S is the surface enclosing the volume element, and $\boldsymbol{n}$ is the outward directed normal to the volume. The fluid is approximately incompressible, so that $\nabla \cdot \boldsymbol{v} = 0$, reducing Eq. 12 to:

$$\sigma_s = \eta_c (\nabla \boldsymbol{v} + (\nabla \boldsymbol{v})^T) \tag{14}$$

Substituting Eq. 14 into Eq. 13 yields:

$$\boldsymbol{F}_{fc} = \eta_c \oint \nabla \boldsymbol{v} \cdot \boldsymbol{n} \mathrm{d}S \tag{15}$$

The velocity field $\boldsymbol{v}$ along the area element path S can be computationally determined by decomposing $\boldsymbol{v}$ into x and y components and expanding these using a Taylor series about the midway point of each edge of the area element. Using the edge $\boldsymbol{c}_1\boldsymbol{x}_1$ with midpoint $\boldsymbol{x}_0$ as and example, the

expansion is:

$$
\begin{aligned}
v_x(\boldsymbol{x}) &\approx v_{x0} + (x - x_0)\left.\frac{\partial v_x}{\partial x}\right|_{x_0} + (y - y_0)\left.\frac{\partial v_x}{\partial y}\right|_{x_0} \\
v_y(\boldsymbol{x}) &\approx v_{y0} + (x - x_0)\left.\frac{\partial v_y}{\partial x}\right|_{x_0} + (y - y_0)\left.\frac{\partial v_y}{\partial y}\right|_{x_0}
\end{aligned}
\tag{16}
$$

Where the nodes $\boldsymbol{c}_1$, $\boldsymbol{r}_1$, and $\boldsymbol{r}_3$ are used to obtain:

$$
\begin{aligned}
v_{x0} &= \frac{2v_x(\boldsymbol{c}_1) + v_x(\boldsymbol{r}_1) + v_x(\boldsymbol{r}_3)}{4} \\
\left.\frac{\partial v_x}{\partial x}\right|_{x_0} &= \frac{r_{1y}[v_x(\boldsymbol{r}_3) - v_x(\boldsymbol{c}_1)] + r_{3y}[v_x(\boldsymbol{c}_1) - v_x(\boldsymbol{r}_1)] + c_{1y}[v_x(\boldsymbol{r}_1) - v_x(\boldsymbol{r}_3)]}{r_{1y}(r_{3x} - c_{1x}) + r_{3y}(c_{1x} - r_{1x}) + c_{1y}(r_{1x} - r_{3x})} \\
\left.\frac{\partial v_x}{\partial y}\right|_{x_0} &= -\frac{r_{1x}[v_x(\boldsymbol{r}_3) - v_x(\boldsymbol{c}_1)] + r_{3x}[v_x(\boldsymbol{c}_1) - v_x(\boldsymbol{r}_1)] + c_{1x}[v_x(\boldsymbol{r}_1) - v_x(\boldsymbol{r}_3)]}{r_{1y}(r_{3x} - c_{1x}) + r_{3y}(c_{1x} - r_{1x}) + c_{1y}(r_{1x} - r_{3x})}
\end{aligned}
\tag{17}
$$

and

$$
\begin{aligned}
v_{y0} &= \frac{2v_y(\boldsymbol{c}_1) + v_y(\boldsymbol{r}_1) + v_y(\boldsymbol{r}_3)}{4} \\
\left.\frac{\partial v_y}{\partial x}\right|_{x_0} &= \frac{r_{1y}[v_y(\boldsymbol{r}_3) - v_y(\boldsymbol{c}_1)] + r_{3y}[v_y(\boldsymbol{c}_1) - v_y(\boldsymbol{r}_1)] + c_{1y}[v_y(\boldsymbol{r}_1) - v_y(\boldsymbol{r}_3)]}{r_{1y}(r_{3x} - c_{1x}) + r_{3y}(c_{1x} - r_{1x}) + c_{1y}(r_{1x} - r_{3x})} \\
\left.\frac{\partial v_y}{\partial y}\right|_{x_0} &= -\frac{r_{1x}[v_y(r_3) - v_y(c_1)] + r_{3x}[v_y(c_1) - v_y(r_1)] + c_{1x}[v_y(r_1) - v_y(r_3)]}{r_{1y}(r_{3x} - c_{1x}) + r_{3y}(c_{1x} - r_{1x}) + c_{1y}(r_{1x} - r_{3x})}
\end{aligned}
\tag{18}
$$

Analogous equations for the faces $\boldsymbol{x}_1\boldsymbol{c}_1$, $\boldsymbol{c}_2\boldsymbol{x}_2$, $\boldsymbol{x}_2\boldsymbol{c}_3$, $\boldsymbol{c}_3\boldsymbol{x}_3$, and $\boldsymbol{x}_3\boldsymbol{c}_1$ can be derived to compute the entire the path integral in Eq. (15).

Topology Changes and Restrictions

Topological changes (i.e. changes in the connectivity of vertices) occur though T1 transformations in this model. In this type of transformation, when two vertices connected by an edge are within a critical distance from each other, connectivity changes such that the edge connecting the vertices is rotated by 90° while maintaining its length. Transformations are accepted only when the net tension force on the resulting vertices after the transition would cause them to separate. Additionally, to maintain the integrity of the two dimension system, motions of single vertices and pairs undergoing a T1 transformation are restricted to prevent the formation of complex polygonal systems (edges

of polygons may not cross).

Implementation of Vertex Model

The vertex model of migration was implemented in both the Matlab and Fortran programming languages (code available upon request). The parameters used to define normal cells are listed in Supplemental Table 1. In Supplemental Table 2, the additional cell types and the values of parameters changed to model those cell types are listed.

Parameter Definitions

All parameters were rendered dimensionless through normalization. Units of length were normalized by $\sqrt{A_0}$, units of time were normalized by $\frac{\eta_s}{KA_0}$, and units of energy were normalized by $KA_0^{3/2}$.

The density factor parameter, df, was used to modulate the cell density and reflects the mean cell size in the simulation relative to A_0. Using this convention, all internal cell parameters remain constant across density, and only the size of the simulation region is modified.

2.12 CORRELATION ANALYSES

Radial Spatial Correlation

For n observations of velocities $\boldsymbol{v}_i$ at locations $\boldsymbol{x}_i$, the radial spatial autocorrelation function (referred to as radial correlation) is defined as:

$$C(\rho) = \frac{\sum_{ij} \frac{\boldsymbol{v}_i \cdot \boldsymbol{v}_j}{\|\boldsymbol{v}_i\| \|\boldsymbol{v}_j\|} \delta(\rho - d_{ij})}{\sum_{ij} \delta(\rho - d_{ij})} \tag{19}$$

$$\delta(x) = \begin{cases} 1, & \text{if } 0 \leq x < w \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

where ρ is the spatial lag (distance from an index position), d_{ij} is the distance between the i^{th} and j^{th} observations, and the double vertical bars denote a Euclidean norm. The function δ ensures that only pairs of observations whose separation distance is within an annulus of width w contribute to the sum at the separation distance ρ.

Matrix Representation of Correlation Functions

The generalization of the radial correlation function is facilitated by a system of notation that uses matrices to define relationships between observations in a pairwise or n-way manner, including spatial relationships, temporal relationships, or velocity correlations. These matrices can be combined in a systematic manner to generate correlation functions.

The radial correlation function depends on two matrices, a correlation matrix to measure similarity in velocity and a distance matrix to measure spatial relationships. The correlation matrix

describes the directional similarity between all pairs of observations:

$$\boldsymbol{C}_{ij} = \frac{\boldsymbol{v}_i \cdot \boldsymbol{v}_j}{\|\boldsymbol{v}_i\|\|\boldsymbol{v}_j\|} \tag{21}$$

The distance matrix describes the distance between all pairs of observations and is defined as:

$$\boldsymbol{D}_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\| \tag{22}$$

Using these matrix definitions, the radial correlation function can be defined as:

$$C(\rho) = \frac{\sum_{ij} \boldsymbol{C}_{ij}\boldsymbol{I}_{ij}}{\sum_{ij} \boldsymbol{I}_{ij}} \tag{23}$$

$$\boldsymbol{I}_{ij} = \begin{cases} 1, & \text{if } \boldsymbol{D}_{ij} \in \rho \\ 0, & \text{otherwise} \end{cases} \tag{24}$$

where $\boldsymbol{I}_{ij}$ is an indicator matrix that takes on a value of 1 when the pair belongs to a specific radial annulus of diameter ρ. The set member symbol $\in$ is used to denote that ρ is a set of distance values within an annulus of finite width.

Polar Spatial Correlation (Directional-Correlation)

The radial correlation function measures the relative magnitude of organization within a velocity field but relies on an assumption that organization is symmetric, and cellular motion often breaks this assumption (Fig. S3). The polar spatial correlation function measures organization as a function of both distance and direction from an index point using a coordinate system defined by the heading of the index vector. This is achieved by defining a matrix which describes the direction from the i^{th} observation to the j^{th} observation relative to the heading of the i^{th} observation:

$$\boldsymbol{P}_{ij} = \arccos\left(\frac{\boldsymbol{v}_i \cdot (\boldsymbol{x}_i - \boldsymbol{x}_j)}{\|\boldsymbol{v}_i\|\|\boldsymbol{x}_i - \boldsymbol{x}_j\|}\right). \tag{25}$$

Unlike $\boldsymbol{C}$ and $\boldsymbol{D}$, the direction matrix is asymmetric. Using these matrix definitions, the polar correlation function can be defined as:

$$C(\rho, \theta) = \frac{\sum_{ij} \boldsymbol{C}_{ij} \boldsymbol{I}_{ij}}{\sum_{ij} \boldsymbol{I}_{ij}} \tag{26}$$

$$\boldsymbol{I}_{ij} = \begin{cases} 1, & \text{if } \boldsymbol{D}_{ij} \in \rho \text{ and } \boldsymbol{P}_{ij} \in \theta \\ 0, & \text{otherwise} \end{cases} . \tag{27}$$

$\boldsymbol{I}_{ij}$ is an indicator matrix that takes on a value of 1 when the pair belongs to a specific (ρ, θ) sector, where ρ is a radial annulus and θ is an angular segment.

Spatiotemporal Correlations

Spatiotemporal correlations provide information on the time-evolution of organization within a velocity field. The type of spatiotemporal analysis that can be performed depends on the form of the velocity field. Consider a velocity field $\boldsymbol{v}(\boldsymbol{X}, t)$ with positions $\boldsymbol{X}$ and time t. Repeated observations across time-points can be made either at the same absolute positions in space (Eulerian perspective) or on the same material elements which occupied the positions $\boldsymbol{X}$ at $t = 0$ (Lagrangian perspective). Cell migration naturally fits a Lagrangian perspective, but similar functions could be developed in a Eulerian framework.

Formalizing the Lagrangian perspective, let $\boldsymbol{X}$ define a set of reference positions at time $t = 0$. As time advances, the material elements that initially occupied the positions $\boldsymbol{X}$ move to the current positions $\boldsymbol{x} = f(\boldsymbol{X}, t)$. Therefore, the Lagrangian definition of a velocity field is $\boldsymbol{v}(\boldsymbol{x}(\boldsymbol{X}, t), t)$, or simply $\boldsymbol{v}(\boldsymbol{x}, t)$.

Due to the asymmetric direction matrix (Eq. 25) and the two perspectives from which velocity fields can be specified, multiple spatiotemporal correlation functions may be defined. An exami-

nation of the full set of functions that may be generated is beyond the scope of this manuscript, and here we restrict our approach to a correlation function that addresses whether neighboring positions align to the direction of an index position over time. The general form of this function is:

$$C(\boldsymbol{\rho}, \tau) = f(\boldsymbol{v}(\boldsymbol{x}, t), \boldsymbol{v}(\boldsymbol{x}(\boldsymbol{X} + \boldsymbol{\rho}), t + \tau)), \tag{28}$$

where $\boldsymbol{\rho} = (\rho, \theta)$ is a vector spatial lag comprised of a radial and angular component.

For discrete observations, this function can be defined in matrix notation. Let $\boldsymbol{x}_{ij}$ be the position of n labeled material elements over m synchronous time-points such that $\boldsymbol{x}_{i2}$ is the position at the second time-point of the i^{th} material element that occupied the $\boldsymbol{x}_{i1}$ at the first time-point. Similarly, let $\boldsymbol{v}_{ij}$ be the velocity of n labeled material elements over m synchronous time-points. The correlation matrix for this function is defined as:

$$\boldsymbol{C}_{ijkl} = \frac{\boldsymbol{v}_{ij} \cdot \boldsymbol{v}_{kl}}{\|\boldsymbol{v}_{ij}\| \|\boldsymbol{v}_{kl}\|}, \tag{29}$$

Where $\boldsymbol{C}_{ijkl}$ is the directional correlation between the i^{th} element at the j^{th} time-point and the k^{th} element at the l^{th} time-point. The distance matrix for this function is defined as:

$$\boldsymbol{D}_{ijkl} = \|\boldsymbol{x}_{ijkl} - \boldsymbol{x}_{ijkl}\| \tag{30}$$

The direction matrix for this function is defined as:

$$\boldsymbol{P}_{ijkl} = \arccos\left(\frac{\boldsymbol{v}_{ij} \cdot (\boldsymbol{x}_{ij} - \boldsymbol{x}_{kl})}{\|\boldsymbol{v}_{ij}\| \|\boldsymbol{x}_{ij} - \boldsymbol{x}_{kl}\|}\right) \tag{31}$$

The time-differential matrix for this function is defined as:

$$\boldsymbol{T}_{ijkl} = j - l \tag{32}$$

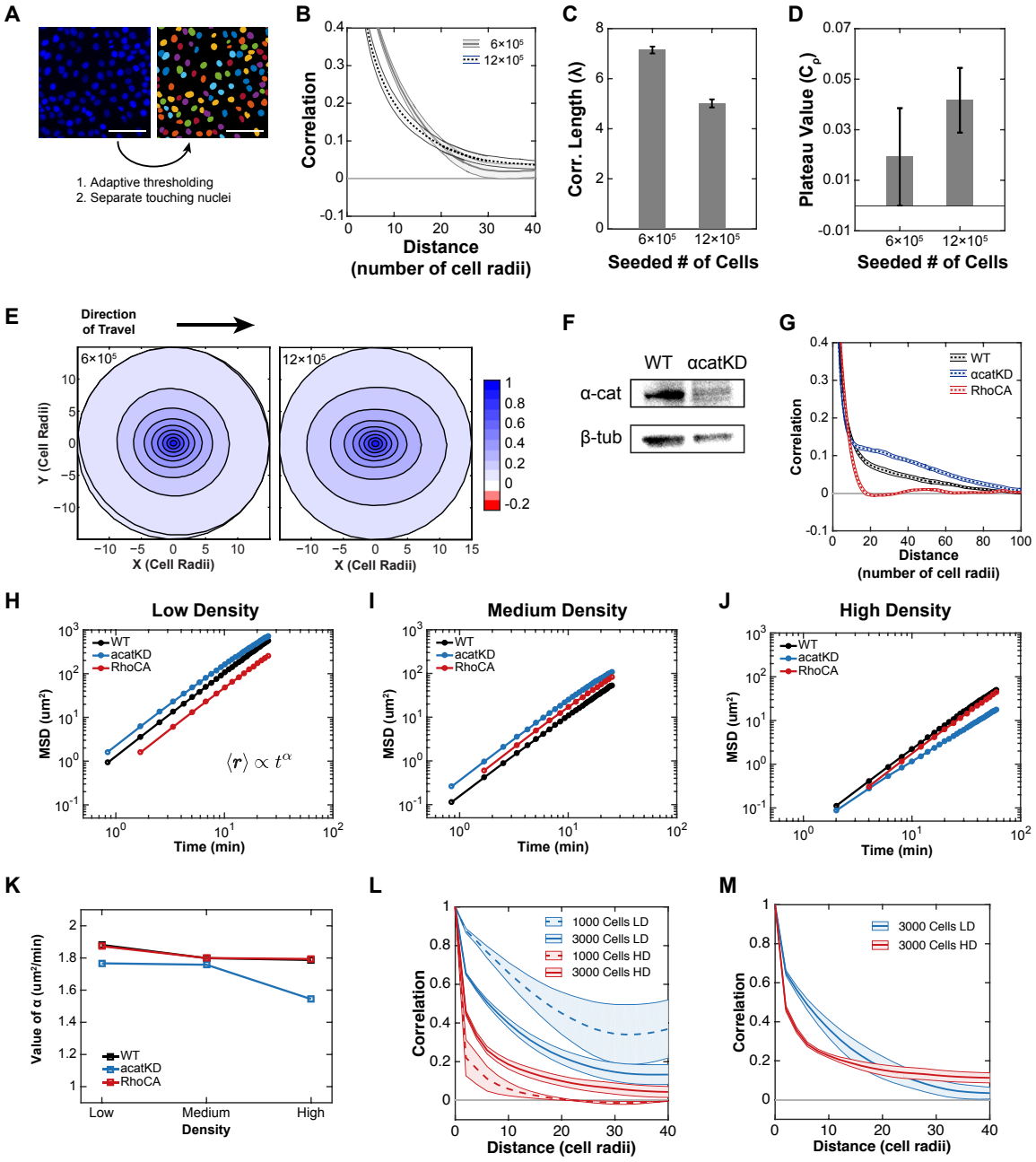
The correlation function can then be defined as:

$$C(\rho, \theta, \tau) = \frac{\sum_{ijkl} \boldsymbol{C}_{ijkl} \boldsymbol{I}_{ijkl}}{\sum_{ijkl} \boldsymbol{I}_{ijkl}} \tag{33}$$

$$\boldsymbol{I}_{ijkl} = \begin{cases} 1 & \text{if } \boldsymbol{D}_{ijkj} \in \rho, \boldsymbol{P}_{ijkj} \in \theta, \text{ and } \boldsymbol{T}_{ijkl} \in \tau \\ 0 & \text{otherwise} \end{cases} . \quad (34)$$

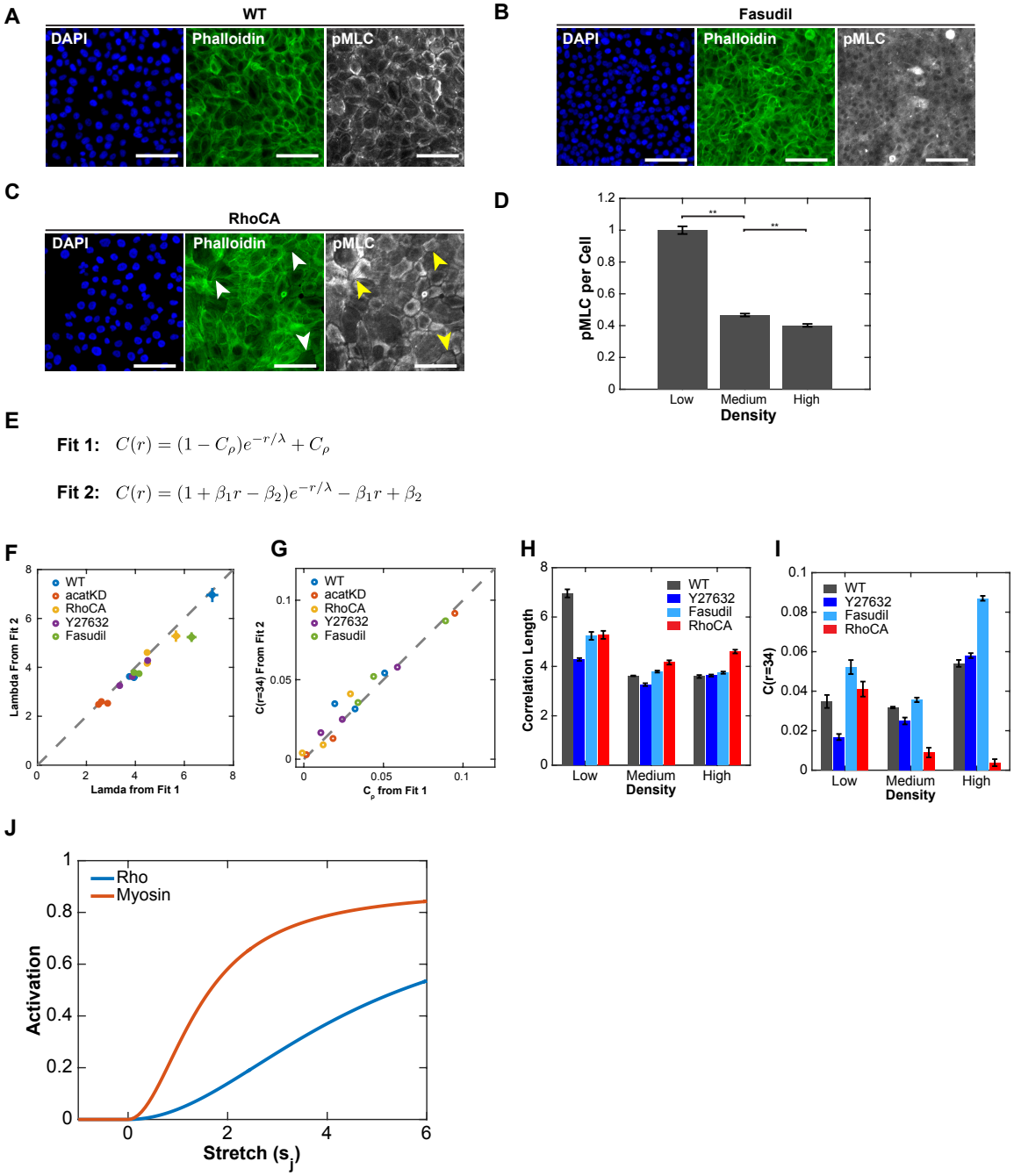
This correlation function applies spatial lag at the reference time (note the repeated subscript j on $\boldsymbol{D}$ and $\boldsymbol{P}$). Therefore, time correlations are computed between an index observation and its neighbors based on initial spatial relationships. Particle Image Velocimetry (PIV) produces velocity field data from a Eulerian perspective. Lagrangian velocity fields used in spatiotemporal computations were generated by propagating initial grid points produced by PIV forward in time using bilinear interpolation.

2.13 SUPPLEMENTAL FIGURES



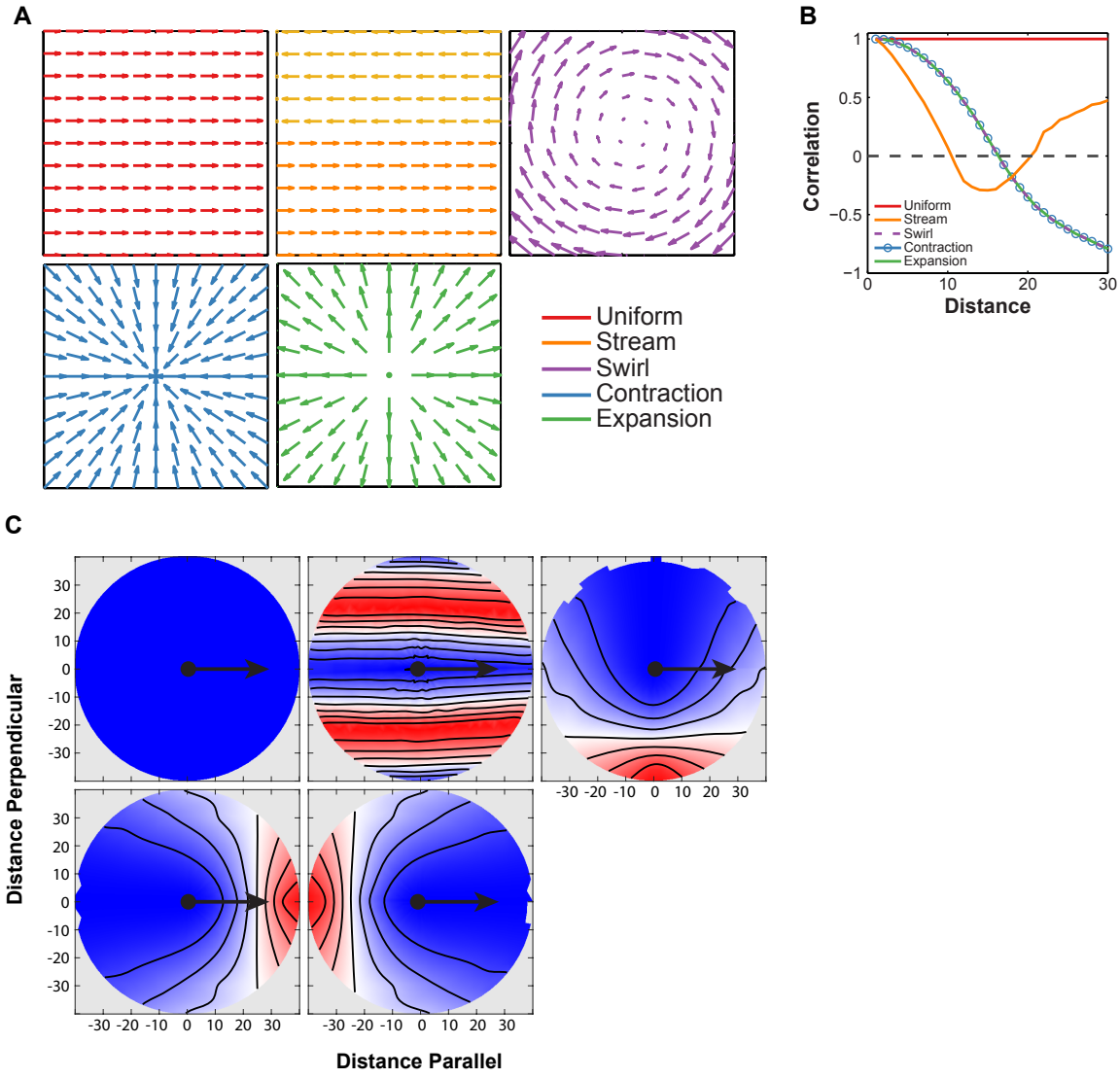
Supplemental Figure 2.1: Related to Figures 2.1 and 2.3. A. Cell density measurement diagram. Live-cell nuclear dye (left) and pseudocolored detected nuclei (right). Nuclei images were processed through adaptive thresholding and morphological operations to generate binary images of candidate nuclei. Watershed segmentation of the distance transform of this binary image was used to separate touching nuclei (right, nuclei colors assigned randomly to show distinctions).

Scale bars 100 um. B. Radial correlation curves for WT cells plated at different starting densities ($6x10^5$ and $12x10^5$ cells per 30mm dish). C. Value of correlation length (λ) obtained by fitting radial correlation curves. D. Plateau value (C_ρ) obtained by fitting radial correlation curves. E. Directional correlation plots for WT cells plated at different starting densities ($6x10^5$ and $12x10^5$ cells per 30mm dish). F. Western blot for a-catenin in WT and acatKD cells. G. Radial correlation curves for WT, acatKD, and RhoCA cells at high density over a distance of 100 cell radii. H-J. Mean Square Displacement (MSD) across cell types and densities. The MSD follows a power law form with respect to time (inset equation in H). K. The value of the exponent of this power law across cell types and densities. L. Radial correlation curves for simulations at low and high density for domains of 1000 and 3000 cells. M. Radial correlation curves computed for the first 200 time-steps for low and high-density 3000 cell simulations. Error envelopes are SEM. Each density is the mean of 12 (WT: $6x10^5$), 6 (WT: $12x10^5$), 6 (αcatKD), or 8 (RhoCA) 2.3 mm^2 regions (four 10x fields per region).

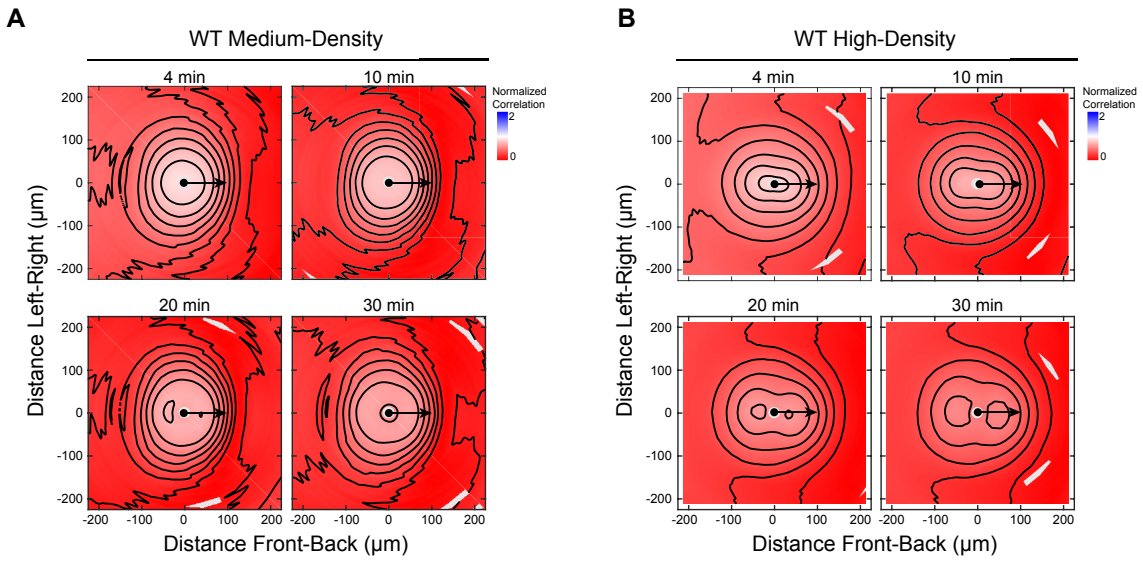


Supplemental Figure 2.2: Related to Figure 2.2. A. WT cells fixed and stained for DAPI, phalloidin, and anti-pMLC. B. WT cells treated with 30 uM of the ROCK inhibitor Fasudil for 5 hours and fixed and stained in the same manner as A. C. RhoCA cells fixed and stained in the same manner as A. Arrowheads denote the enhanced formation of stress fibers (white arrows, phalloidin channel) and corresponding increases in the intensity of pMLC staining (yellow

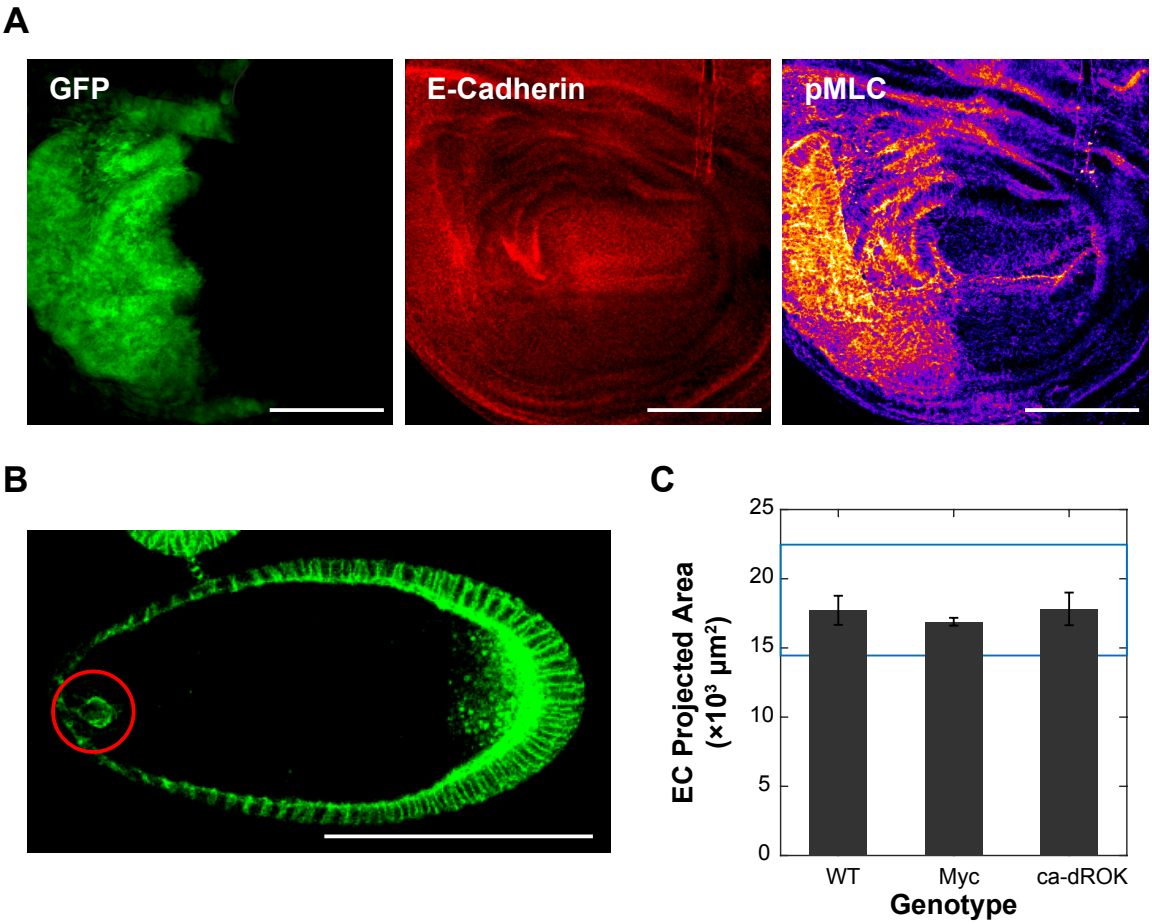
arrows, pMLC channel). D. Intensity of pMLC immunofluorescence per cell (signal divided by number of nuclei) normalized to the low-density value. Density conditions were generated in the same manner as migration experiments. E. Two parameterizations used to fit radial correlation data. The form of radial correlation function out to a radial distance of approximately 40 cell radii is an exponential decay to a constant (Fit 1), but beyond this distance, a mixed exponential-linear decay is observed (Fit 2, see Fig S1G). F. Lambda values computed from Fit 1 (using data out to 40 cell radii) plotted versus those computed from Fit 2 (using data out to 100 cell radii) show close agreement, supporting the validity of the simplified approach. G. Value of Crho from Fit 1 plotted versus the equivalent value, C(r=34), from Fit 2, showing close agreement. H. Correlation length computed using Fit 2, compare to Fig 2J. I. C(r=34) computed using Fit 2, equivalent to Crho from Fit 1. Compare to Fig 2K. J. Steady-state solutions for Rho and Myosin activation (Eq. 6 and 7). All Scale bars are 100 um. ** = $p<0.01$.



Supplemental Figure 2.3: Related to Figure 2.3. A. Example velocity fields with uniform, streaming, swirling, contracting, or expanding patterns. B. Radial-correlation computed for the velocity fields in A. C. Directional-correlation computed for the same velocity fields.



Supplemental Figure 2.4: Related to Figure 2.4. A. Spatiotemporal correlation normalized to zero delay time values for medium-density WT cells at delay times of 4, 10, 20, and 30 minutes. B. Same as A for high-density WT cells.



Supplemental Figure 2.5: Related to Figure 2.5. A. Fixed drosophila wing disk showing endogenous GFP expression, E-cadherin staining, and anti-pMLC staining. Engrailed-Gal4 drives expression of ca-dROK and GFP and is limited to half of the wing disk (GFP positive region in panel 1). B. Representative border cell delamination event. C. Projected egg chamber area during border cell delamination. All Scale bars are 100 um.

2.13 SUPPLEMENTAL VIDEOS

Supplemental videos were included in the manuscript submission. The captions are reproduced below and the videos are available upon request to the author.

Video S1. Related to Figure 2.1. Low density WT MCF10A cells move in large collectives. Video begins with a display of the first frame of a time-lapse with overlay of velocity magnitude (color intensity) and direction (color hue). The large magenta and region shows cells moving in similar directions. Time-lapse 1 hour with 2 minute frame intervals.

Video S2. Related to Figure 2.1. WT cell migration at high density using same overlay as Video S1. Patches of similar colors are much smaller and more irregularly shaped than at low density (Video S1). Time-lapse 1 hour with 2 minute frame intervals.

Video S3. Related to Figure 2.3. Low density WT MCF10A cells also exhibit regions of contraction. Video begins with the same type of overlay as before. The magenta and green regions in the horizontal center and bottom half represent opposing motions as seen in contraction. Overlay same as used in Video S1. Time-lapse 1 hour with 2 minute frame intervals.

Video S4. Related to Figure 2.3. Alpha-catenin knockdown cells move with markedly different structure, forming streams of follow-the-leader motion. Overlay same as used in Video S1.

Time-lapse 1 hour with 2 minute frame intervals.

Video S5. Related to Figure 2.6. Migration of WT, acat-IR, dMyc, and ca-dROK stage 5 egg chambers. Time-lapse 2 hours with 5 minute frame intervals.

Video S6. Related to Figure 2.6. Migration of WT and dMyc stage 8 egg chambers. Time-lapse 2 hours with 5 minute frame intervals.

Video S7. Related to Figure 2.6. Migration of WT, acat-IR, and ca-dROK stage 7 egg chambers. Time-lapse 2 hours with 5 minute frame intervals.

2.14 SUPPLEMENTAL TABLES

Cell Type	Mean Cell Radius (μm ± SEM)	Cell Density (cells/mm^2 ± SEM)
Wild Type		
Low	15.8 ± 0.3	1280 ± 20
Medium	11.1 ± 0.2	2580 ± 50
High	9.7 ± 0.1	3380 ± 30
α-cat KD		
Low	16.7 ± 0.8	1140 ± 50
Medium	12.2 ± 0.3	2140 ± 50
High	9.5 ± 0.2	3530 ± 70
RhoCA		
Low	14.1 ± 0.6	1600 ± 70
Medium	12.3 ± 0.5	2100 ± 90
High	11.2 ± 0.5	2540 ± 110
WT + Fasudil		
Low	15.0 ± 0.5	1410 ± 50
Medium	10.8 ± 0.2	2730 ± 50
High	9.0 ± 0.1	3930 ± 40
WT + Y27632		
Low	13.9 ± 0.6	1650 ± 70
Medium	11.0 ± 0.2	2630 ± 50
High	8.9 ± 0.1	4020 ± 50

Supplemental Table 2.1: Related to Figures 2.1 and 2.2. Mean cell radii and corresponding cell density across cell lines and grouped density conditions.

Parameter	Description	Value
n	Number of cells in simulation	1000
η_s	Cell-substrate friction coefficient	1
η_c	Cell-cell friction coefficient	0.1
K	Area elastic modulus	5
A_0	Preferred cell area	1
Λ	Adhesion strength	0.5
k	Contractility constant	0.8
A_ρ	Maximal activation rate for ρ	0.5
K_s	Half maximal response constant	0.1
m	Hill coefficient	2
D_ρ	Inactivation rate for ρ	0.1
A_M	Maximal M activation rate	0.5
D_M	Inactivation rate for M	0.05
t_1	T1 transformation threshold	0.07
df	Density factor	[0.5, 1, 1.5]
α	Persistence force strength	0.1
β	Persistence force decay	0.1

Supplemental Table 2.2: Related to Figures 2.1 and 2.2. Full set of parameters used to simulate migration under the biomechanical contractility model. All parameters have been normalized to be dimensionless (see Supplemental Experimental Procedures).

Model	Figure References	Parameter Changed	Value
BCC Low Adhesion	1L, 1M	Λ	5
BCC Low Friction	1L, 1M	η_c	0.01
BCC Low Contractility	1L, 1M	k	0.04
CCM	1I, 1K	Eq. 8 in Sup. Exp. Proc.	$\Gamma = k_c$ $k_c = 0.8$

Supplemental Table 2.3: Related to Figure 2.1. Modified parameters used to model alternate cell types. All parameters have been normalized to be dimensionless (see Supplemental Experimental Procedures).

CHAPTER THREE

Populations of apical-basal polarity deficient cells disrupt an otherwise normal epithelium through direct contact and field effects

3.1 INTRODUCTION

Cancer is a disease characterized by cell type heterogeneity (Gerlinger et al., 2012; Heppner and Miller, 1983; Nowell, 1976). In the majority of adult-acquired carcinomas, the disease initiates when a cell or subpopulation of cells within an otherwise normal epithelial tissue acquire a genetic mutation or change that disrupts normal regulatory control (Marusyk and Polyak, 2010). As a result, abnormal cells begin to emerge side-by-side with normal cells in a tissue.

A frequent change that arises within carcinomas is the loss of apical-basal polarity (ABP) regulation. Under normal conditions, epithelial cells maintain specific apical and basolateral domains controlled by the precise localization of protein complexes (Halaoui and McCaffrey, 2015; Igaki et al., 2006; McCaffrey et al., 2012). Three primary polarity complexes have been identified (Assémat et al., 2008). These are the Par (or aPKC), Crumbs, and Scribble complexes. The specific domains formed by these complexes allow for the proper function of epithelial tissues through control of cell division, orientation of secretory machinery, and the formation and maintenance of proper cell-cell contacts through specific junctions (Betschinger et al., 2003; Qin et al., 2005; Zhan et al., 2008). Although discussion persists on whether apical-basal polarity loss is a cause or consequence of cancer, abnormalities in the regulation of apical-basal polarity are frequently seen in tumor progression.

A primary effect of the loss of apical-basal polarity in epithelial cells is the disruption of the organization of normal cell-cell junctions. This leads to an alteration in the adhesive, force-generating, and force sensing properties of cells. The Scribble complex has been demonstrated to be important for cell-cell adhesion, and the apical Par complex plays a role in regulating actomyosin contractility at adherens junctions (Qin et al., 2005; Warner and Longmore, 2009). Additionally, polarity complexes are involved in the regulation and localization of components of the Hippo Pathway, which plays a primary role in mechanosensing (Low et al., 2014; Schroeder and Halder, 2012). These alterations are also associated with increased metastatic potential (McCaffrey et al., 2012).

The functional consequences of these cancer-related changes within the context of an otherwise normal epithelium are not fully understood. In this study, we undertake an investigation of the effect

heterogeneous loss of apical and basal polarity, focusing on the effect these losses have on epithelial reorganization and migration. We find that apical-basal polarity loss leads to cell type segregation and disruption in the migration of surrounding WT cells, but that apical loss does so in a manner distinct from basal loss. Depletion of Par6 produces a phenotype that corresponds with heterogeneity in cell-cell adhesion, while depletion of Scrib more closely matches a model system of heterogeneity in actomyosin contractility.

3.2 MATERIALS AND METHODS

Cell Culture

MCF10A cells were purchased from the ATCC and were cultured in 1:1 DMEM:F12 (Gibco) supplemented with 5% Horse Serum (Gibco), 10ug/mL Insulin (Sigma-Aldrich), 500 ng/mL Hydrocortisone (Sigma-Aldrich), 100 ng/mL Cholera Toxin (Sigma-Aldrich), 20 ng/mL EGF (Gibco), and Penicillin/Streptomycin (Gibco). 293T cells (ATCC) used for Lentiviral production were cultured in DMEM (Gibco) supplemented with 10% heat-inactivated FBS. In regular culture MCF10A cells were grown to 75% confluence, dissociated using Trypsin-EDTA 0.05% (Gibco), and split 1:5. Cells were plated at varying ABP-deficient:WT ratios for a total of 6x10^5 per 35 mm dish. Cell types were labeled by marking nuclei using a pLKO.1 lentiviral vector containing H2B-RFP or H2B-GFP (Addgene 26001 and 25999) and the resulting cell lines were sorted for uniform and high expression to aid in automated detection. To deplete α-catenin, the lentiviral pFLRu vector containing anti-α-catenin shRNA and puromycin resistance was used as described previously(Bajpai et al., 2008). MCF10A cells expressing constitutively active Rho were generated in the same manner using a QL activating mutant Rho (Coso et al., 1995) subcloned into a pFLRu vector. Puromycin (Sigma-Aldrich) was used in cell selection and maintenance at a concentration of 1.5 μg/mL.

Imaging

Timelapse imaging was performed on a Nikon Ti-E inverted epifluorescence microscope with temperature and humidity control for live cell imaging (LiveCell). Images were collected from 2.3mm^2 regions by tiling 4 fields of 10X magnification. Time-lapse imaging was performed with 2-hour duration and 7-minute frame intervals. Reference images of the dish surface were taken at each frame and field of view to correct for stage drift.

Image Analysis

Cell nuclei were detected and segmented using custom Matlab (The Mathworks, Natick MA) software. The regions of space corresponding to each cell were generated through watershed segmentation of the phase contrast image using nuclei as markers. This well approximated the cell boundaries without the need for additional labeling. Cell migration velocity fields were generated using a Matlab implementation of Lucas-Kanade Optical Flow.

Spatial Statistics

Spatial statistics are based on methods developed by Ripley (Diggle and Chetwynd, 1991; Gatrell et al., 1996; Ripley, 1976). Spatial association is measured by first computing the intra- and inter-cell-type K functions. Ripley's K function is the empirical distribution function of distances between points scaled such that $K(d) = \pi d^2$ under complete spatial randomness (CSR). The self-K function and cross-K function are respectively defined as:

$$K_{11}(s) = A\left(\sum_{i \neq j} \frac{w_{ij}\delta_{ij}}{n_1(n_1 - 1)}\right)$$

$$K_{12}(s) = A\left(\sum_{i,j} \frac{w_{ij}\delta_{ij}}{n_1 n_2}\right)$$

Where s is distance between observations, A is area, w_{ij} is a weighting matrix for edge adjustments, δ_{ij} is an indicator matrix for the distance matrix d_{ij} such that δ_{ij} takes a value of 1 if $d_{ij} \leq s$. The additional K functions K_{22} and K_{21} are similarly defined. The degree of segregation for cells of a given type is measured by the difference function $D = K_{11} - K_{12}$ and compares the relative likelihood of encountering points of the same or different type as a function of distance. The significance of this difference is assessed using random labeling and Monte-Carlo sampling techniques to develop a 95% confidence interval of D under random chance given n_1 cells of type 1 and n_2 cells of type 2. The standardized

difference reported in Fig 3.2 is $\widehat{D} = (D - D_{MC})/\sigma_{MC}$, where D_{MC} and σ_{MC} are the mean and standard deviation, respectively, obtained from Monte-Carlo sampling.

3.3 RESULTS

Control of WT/ABP-deficient interactions in heterogeneous cultures

To determine the effect of apical-basal polarity loss in the context of a normal epithelium as might be seen in cancer initiation, we devised a co-culture setup using mixtures of H2B-GFP (WT) or H2B-RFP (mutant) labeled MCF10A cells, a non-transformed human breast epithelial cell line that forms apical-basal polarized monolayers in culture (Fig 3.1A)(REF). These co-cultures spanned mutant cell fractions from 0.001 to 0.5 with single cell-type cultures serving as controls. Examining a range of mixtures allowed for the assessment of two parameters of key importance. As the fraction of mutant cells increases, the fraction of neighbors that a given WT or mutant cell has of the same or opposite type as well as the distance to the nearest cell of opposite type changes (Fig 3.1B-G). This allows for the assessment of heterotypic interactions across a range of contexts as the interaction of a mutant cell with a WT cell may be influenced by the presence or absence of additional mutant cells in the vicinity. Furthermore, interactions between the two cell types may be topological (i.e. neighbor-based) or proximity (i.e. distance-based). As the fraction of siPar6 cells increased from 1 in 1000 to 1 in 2, the mean distance from WT cells to the nearest siPar6 shifted by approximately 1.5 orders of magnitude (Fig 3.1B). Over the ratios studied, the mean distance from siPar6 cells to the nearest WT cell did not change as dramatically (Fig 3.1C). Across all mixing ratios, the number of neighbors around each cell remained similar (Fig 3.1D). Notably, the neighbor distribution changes in a different manner than the distance distribution as the ratio of siPar6 cells is increased: highlighting the importance of considering proximity-based topological effects. The distance and neighbor distributions of co-cultures with siScrib cells followed similar trends (Fig 3.1E-G).

ABP-deficiencies drive cell-type segregation

Inspection of proportions of WT versus polarity-deficient neighbors of WT cells (or of polarity-deficient cells) revealed that while the mixture of cells was even (WT fractions: 0.5176, 95% CI 0.5017-0.5335 for siPar6; 0.4932, 95% CI 0.4758-0.5106 for siScrib), the distribution of neighbor type was not equal (Fig

3.2A). In both cases of polarity deficiency, WT cells tended to have more WT neighbors than polarity deficient neighbors. Similarly, polarity deficient cells tended to have more polarity deficient neighbors than WT neighbors (Fig 3.2B). These data suggested that spatial segregation of cell types was occurring. Cell type segregation plays an important role in germ-layer formation and in the maintenance of tissue boundaries. Early experiments demonstrated that dissociated endoderm and mesoderm cells would sort by cell type when combined back together and extensive work has demonstrated that both cell-cell adhesion and interfacial tension due to cell-cell contractility can drive cell-type segregation. To determine the role of adhesion and contractility in driving cell-type segregation, we generated stable adhesion deficient and hypercontractile versions of MCF10A cells using lentiviral transduction. Adhesion was reduced by depleting α-catenin with an shRNA, and contractility was increased by adding a constitutively active version of Rho. To assess the degree of spatial association, we employed a set of spatial statistics developed by Ripley. A more detailed description can be found in the Materials and Methods section. Briefly, the degree of spatial association is assessed by comparing the distribution of inter-cell distances with the distribution expected under conditions of randomness and the significance of the difference between the observed and expected distributions is determined through Monte-Carlo methods. This procedure allows the association between cell-types to be assessed across a range of distances. The self-association of αcatKD cells and RhoCA cells was assessed at 24 and 48 hours after plating. At 24 hours, modest self-association beyond what would be expected by random chance was observed on a length-scale less than 50um for both αcatKD and RhoCA cells. At 48 hours, αcatKD cells showed a small increase in clustering, and RhoCA cells showed a large increase in clustering, extending out to 100 um (Fig 3.2B-C). Polarity deficient cells also demonstrated evidence for spatial association. siPar6 cells showed minor segregation over short length scales at 24 hours which increased in magnitude, but remained over relatively short length scales, extending out to approximately 60 um. Basal-polarity depleted cells, however, displayed strong segregation out to 60 um at 24 hours, which increased by 48 hours, extending beyond 250 um.

ABP-deficient cells alter the normal dynamics of epithelial monolayers

To determine the effect of ABP heterogeneity on the dynamics of the epithelial monolayers, we examined mean cell speed as a function of mixing ratio and stratifying by cell type within each case using a strategy analogous to solvent mixing diagrams. Each cell type moves with a characteristic mean velocity when in homogeneous monoculture (Fig 3.1A, mixing ratios 0 and 1). When mixed, if each cell maintained this characteristic speed, the observed net speed would be a weighted average of the two characteristic speeds (Fig 3.3A, grey line). This weighted-average behavior is consistent with no interactions between the two cells types. In mixtures of WT and siPar6 cells, all mixtures were observed to depart from the weighted average behavior. Mixtures of WT and siScrib cells showed similar features (Fig 3.3B). Stratifying the net behavior by cell type, at each condition revealed differences between siPar6 and siScrib. While both cell types tended to follow the net behavior for siScrib, the two types tended to split for siPar6 (Fig 3.3A and B).

To determine whether this was a result of the potential difference in physical interactions that siPar6 versus siScrib cells exerted on their WT neighbors, we compared the effect of ABP-deficiency on mean speed to model adhesion and contractility changes. Adhesion-deficient cells were observed to show the same splitting behavior by cell type. The net behavior was altered from the weighted average, but both cell types retained distinct behaviors within the system (Fig 3.3C). Mixtures of WT and RhoCA cells also departed from the weighted-average case, indicating interactions, but both cell types followed the net behavior (Fig 3.3D).

These data reinforced findings from Figure 2 suggesting that loss of apical polarity phenocopied loss of cell-cell adhesion through adherens junctions and that loss of basal polarity phenocopied heterogeneity in cell-cell contractility. Mismatches in cell-cell contractility within an epithelium are known to lead to alterations in cross-sectional 2D cell area, where cells with increasing contractility undergo constriction. Measurement of cell areas for an ABP-deficient mixing ratio of 0.5 showed

preferential constriction of WT cells in mixtures with siScrib but not siPar6. These data suggest a mismatch in contractility between WT and siScrib cells in which WT cells exert greater contractile force.

Spatial range of effects due to ABP heterogeneity

To focus on the interaction between WT and ABP-deficient cells, we measured how the impact on cell speed as a function of distance from the WT/ABP-deficient interface. Cell boundaries were estimated using a watershed-transform algorithm (see Materials and Methods). Contiguous regions of each cell type were formed and the distance from the interface into these regions was calculated (Fig 3.4A and D). The mean speed was measured using 15 um bins (approximately 1 cell radius) extending into each region. For mixtures of WT and siPar6 cells, mean speed followed a linear trend, low within siPar6 regions and increased to higher values within WT regions for a mixing ratio of 0.5 (Fig 3.4B). Mixtures of WT and siScrib cells, however, followed a different trend, with higher speeds near the interface (Fig 3.4C). A mixing ratio of 0.5 provided only a small fraction of observations that fell beyond -30 or 30 um, approximately the diameter of a cell (histograms, Fig 3.1B and C), and therefore, the effect may be a result of direct neighbor contact or a wider-reaching effect. To determine if the effect extended over a greater distance, mixing ratios of 0.1 were analyzed providing data out to 120 um into WT regions (Fig 3.4D). For siPar6 cells, mean speed again followed a linear trend with respect to distance from the interface, ranging from low within siPar6 regions to high within the WT regions. Notably, this increase in speed within WT regions extended multiple cell diameters from the interface (Fig 3.4E). Mixtures with siScrib cells followed the same trend for a mixing ratio of 0.1 as 0.5. For both cell types in the mixture, speeds were highest near the interface and dropped off with distance (Fig 3.4F). In summary, for both cell types, the effect of ABP-deficient cells on WT behavior depends on distance from the interface and extends multiple cell lengths into regions of WT cells. The changes in speed with distance for siScrib mixtures are of a smaller magnitude than for siPar6 mixtures, reflected the finding in Fig 3.3B that both WT and siScrib cells behave similarly when mixed. These data, however, suggest that although global effects exist to alter the behavior of all cells within the mixture, local effects are present at the interface.

3.4 DISCUSSION

In this work, we examined the effect of heterogeneous ABP loss in epithelial monolayers, modeling a setting of early cancer development where abnormal cells arise within a field of otherwise normal cells. We find that both apical and basal polarity loss disrupt the dynamics of normal epithelia and lead to cell type segregation, however the type of disruptions differed, suggesting that apical and basal polarity loss impacted WT neighbors through distinct mechanisms. Loss of apical polarity led to a phenotype that mimicked heterogeneity in cell-cell adhesion, while loss of basal polarity more closely matched a model system with heterogeneity in contractile force generation.

The loss of Par6, a key apical polarity complex protein, affects cell segregation and migration speed in a manner similar to adhesion deficiency caused by α-catenin depletion. Cell type segregation is more similar to that of αcatKD cells than RhoCA cells over 48 hours (Fig 3.2C and E), and siPar6 cells retain a behavior separate from neighboring co-cultured WT cells as measured by mean migration speed (Fig 3.3A and C). The apical polarity complex is important for the establishment, maintenance, and remodeling of epithelial adherens junctions (Warner and Longmore, 2009). The core proteins that provide adhesion and coupling to the actin cytoskeleton at these junctions are E-cadherin, β-catenin, and α-catenin (Knust and Bossinger, 2002). Depletion of Par6, therefore, may disrupt the proper formation of adherens junctions, providing a potential mechanism for its similarity to α-catenin loss. The loss of Scrib, however, more closely matched a setting of heterogeneity in contractility. First, the degree of cell type segregation was dramatic (Fig 3.2D and F), corresponding with mixtures of RhoCA cells; second, at a mixing ratio of 0.5, the individual migration speed of both cell types matched the net migration speed (Fig 3.3B and D); and third, cell size analysis showed preferential constriction of WT cells (Fig 3.3F), a feature seen with heterogeneity in contractility (Warner and Longmore, 2009).

We find three unique effects of heterogeneity on the migratory dynamics of epithelial cells: the net behavior of mixtures versus the expected value from a weighted average of single cell types, the

migration speed of individual cell types within a mixture relative to the net migration speed, and the role that proximity to the WT/ABP-deficient interface has on migration speed.

The migration speed in mixtures of WT and ABP-deficient cells departs from the weighted average of speeds for each cell type in monoculture. Perhaps not surprisingly, this indicated the presence of an interaction. However, the effect of apical versus basal polarity differs in two ways: first, the mixing ratio that resulted in the minimum net migration speed, and second, the uniformity of migration speed across cell types within mixtures. Both types of ABP-deficient cells migrated slower than WT cells in homogeneous monocultures, but surprisingly, mixtures of siPar6 and WT cells were slower than siPar6 cells alone. In contrast, mixtures with siScrib cells remained between the limits set by homogenous monocultures. Analyses of model heterogeneous cultures with adhesion deficiency and hypercontractility also demonstrated this difference: mixtures with αcatKD cells dropped outside of the migration speed range set by monocultures while RhoCA cells did not. One possible explanation for this is that siPar6 or αcatKD cells cannot form mature adherens junctions with neighboring WT cells and communicate forces to re-organize motion. Instead, the heterotypic interaction may lead to contact inhibition of locomotion at the boundaries between cell types (Mayor and Carmona-Fontaine, 2010). In theory, this could lead to synergistic inhibition of migration, "fencing in" contiguous islands of the same cell type and thus reducing the speed of both cell types. In Scrib depletion or when Rho is constitutively active, the mismatch in contractility may lead to a reduction in migration speed, but not a synergistic inhibition. Alternatively, the effect of ABP loss may not be transmitted to WT cells solely through cell junctions. Loss of polarity may lead to secretion of a variety of proteins that may affect cell migration, including MMPs (McCaffrey et al., 2012). While protein secretion may not explain the synergistic inhibition of migration seen in Par6 depletion or αcatKD, it still may play an underlying role in the co-culture system.

The ability to separate net migration speed into the contributions of each individual cell-type revealed further differences in how mixtures of cells behaved. In heterogeneous cultures containing αcatKD and siPar6 cells, the individual cell type migration speed differed from the net migration speed.

This suggests that while overall motion was disrupted, each cell type retained a unique behavior. In contrast, for mixtures containing siScrib or RhoCA cells, net and individual migration speeds were similar. This is consistent with siScrib and RhoCA cells influencing WT cells via a field effect that converts the behavior of WT cells beyond those in direct contact with siScrib cells. Cell-cell contractility is one mechanism by which this field effect may occur. Epithelial cells respond to external forces by generation of internal forces, and this basic property can lead to waves of contractility that spread through the epithelium (Koride et al., 2014; Serra-Picamal et al., 2012). In heterogeneous cultures, alterations in contractility at the interface between the two cell types may similarly propagate to affect distant cells.

Measuring the effect of migration speed as a function of distance from the cell-type interface provided insight into the interactions between cell types. Observations from siPar6 co-cultures are consistent with an interface of two populations of different behaviors. Within siPar6 regions, migration is slow and increases with distance into WT regions. In contrast, observations of siScrib and WT cells show unique behavior at the interface. Within regions of both cell types, migration is slow and reaches a maximum at the interface. This suggests a unique behavior arises at the interface and may be what drives the efficient cell type segregation. In both cases, however, the effects of ABP-deficiency appear to extend beyond nearest neighbor interactions. Whether this effect arises through a chain of altered neighbor interactions or a non-topological effect such as secretion of cytokines or other molecules remains to be determined.

An important consideration in the interpretation of these results is that the similarities manifested between conditions of ABP-loss and the models of cell-cell interaction changes (αcatKD and RhoCA) are based on measurements of cell segregation and effect on mean cell migration speed. Whether similarities persist in alternative measurements of epithelial dynamics, such as spatial or temporal correlation functions, or changes to the levels or localization of the molecular components governing these processes represent important areas for future work.

In conclusion, this work demonstrates that small fractions of ABP-deficient cells can have profound effects on epithelial dynamics, leading to cell type segregation and disruption of migration.

These effects act locally and extend over distances of multiple cell lengths. Furthermore, these data suggest that apical and basal polarity disruption act through different mechanisms. In the metrics used to study the state and dynamics of co-cultured epithelia, apical polarity loss more closely matches a case of heterogeneity in adhesion while basal polarity loss more closely matches heterogeneity in contractility.

3.5 FIGURES

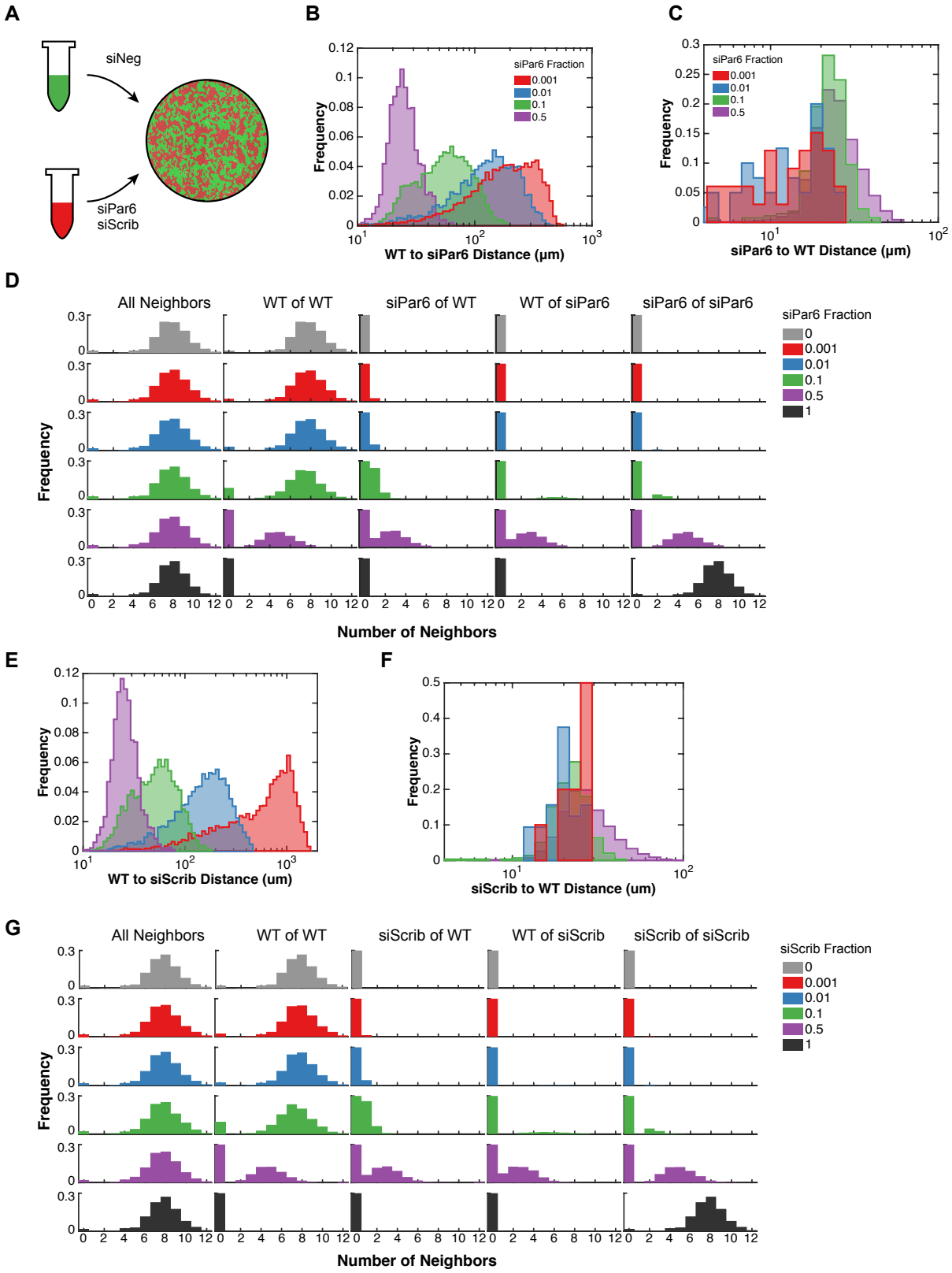


Figure 3.1: Controlled assessment of WT and polarity-deficient cell interactions through co-culture. A. Diagram of co-culture setup. WT (siNeg) cells are mixed with polarity deficient cells (siPar6 or siScrib) in various ratios. B. Distance from WT cells to the nearest siPar6 cell across mixing ratios. C. Distance from siPar6 cells to the nearest WT cell across mixing ratios. D. Neighbor type counts for siPar6 co-cultures. Rows stratify counts by mixing ratio. Columns stratify counts by a specific neighbor relationship. *All Neighbors* refers to the total number of neighbors of each cell. The remaining columns show the type of neighbors for each cell type. For example, the column *siPar6 of WT* is the number of siPar6 neighbors of WT cells. E. Equivalent to B. for mixtures of WT and siScrib cells. F. Equivalent to C. for mixtures of WT and siScrib cells. G. Equivalent to D. for mixtures of WT and siScrib cells.

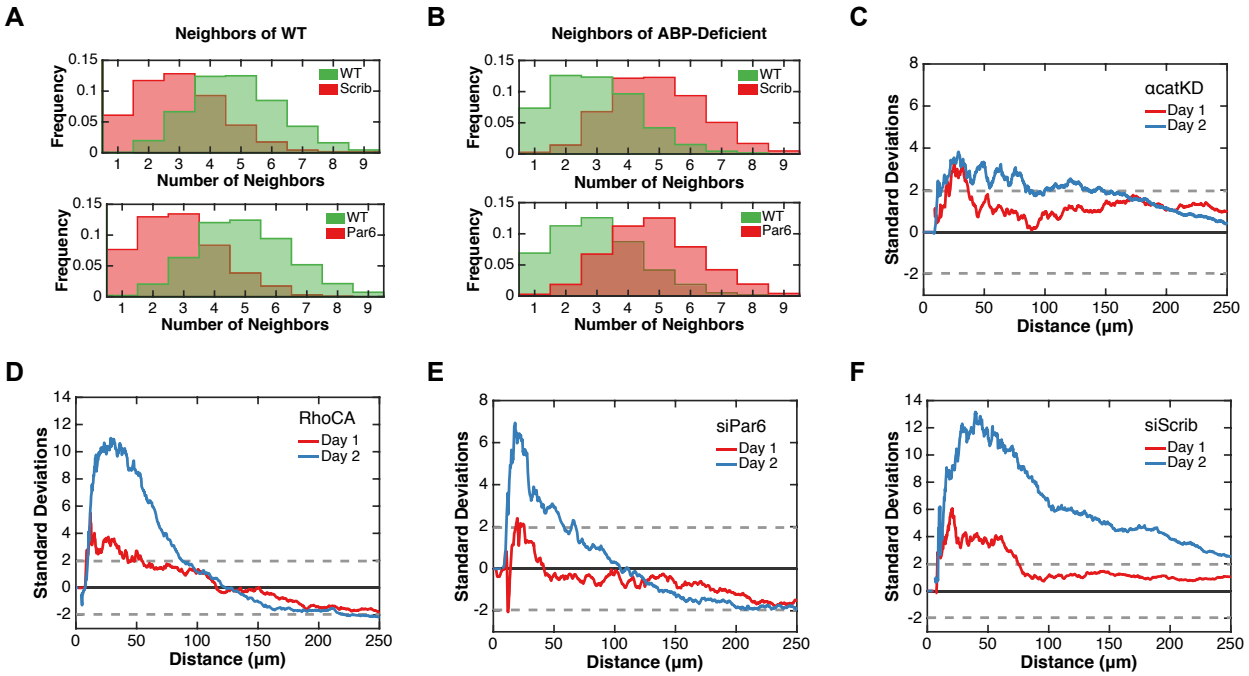


Figure 3.2: Spatial segregation of polarity-deficient cells.

A. Number of WT and ABP-deficient neighbors of WT cells at a polarity-deficient cell fraction of 0.5. B. Analogous to A where the index cell is ABP-deficient. C. Observed spatial association of αcatKD cells co-cultured with WT cells compared to random chance at 1 day and 2 day time-points. A value of 0 is expected under conditions of complete spatial randomness and the dashed grey lines highlight the 95% confidence interval obtained through Monte-Carlo sampling. For more details, see Materials and Methods. D. Equivalent to C. for co-cultures of WT and RhoCA cells. E. Equivalent to C. for co-cultures of WT and siPar6 cells. F. Equivalent to C. for co-cultures of WT and siScrib cells.

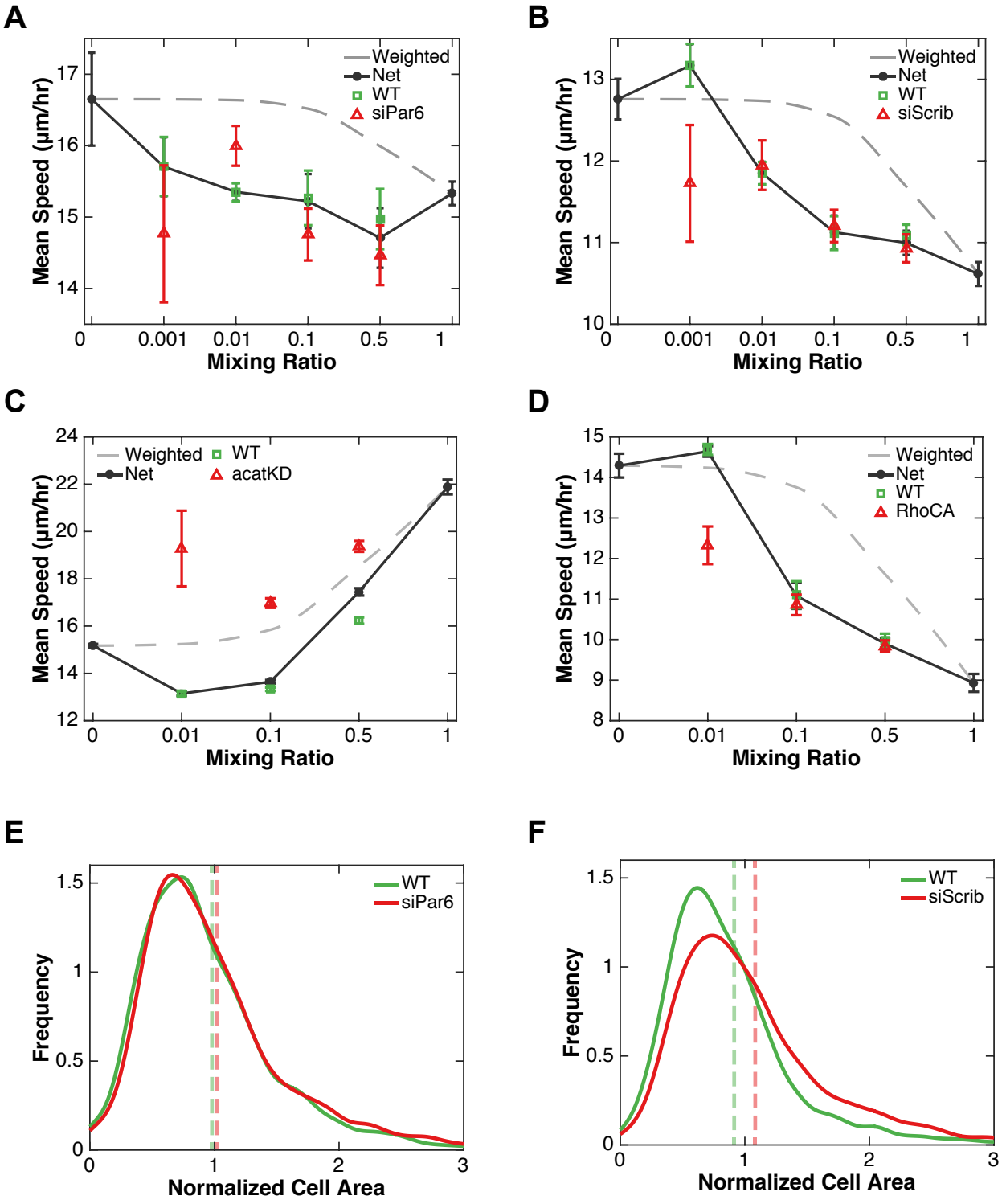


Figure 3.3: Effect of WT and polarity-deficient interactions on cell speed.

A. Mean speed mixing plot for WT and siPar6 cells. The net mean speed irrespective of cell type is shown in black. The weighted average of the mean speeds of the two cell types for each ratio is shown in

the dashed gray line. At each mixing ratio where both cell types are present, the mean speed within each population is shown in green (WT) or red (siPar6). B. Equivalent to A. for mixtures of WT and siScrib cells. C. Equivalent to A. for mixtures for WT and αcatKD cells. D. Equivalent to A. for mixtures for WT and RhoCA cells. Errorbars are SEM. E. Normalized cell area (cell area divided by mean cell area) for WT and siPar6 cells in a co-culture with a mixing ratio of 0.5. The dashed lines indicate the mean normalized cell area for each type. F. Equivalent to E. for mixtures of WT and siScrib cells.

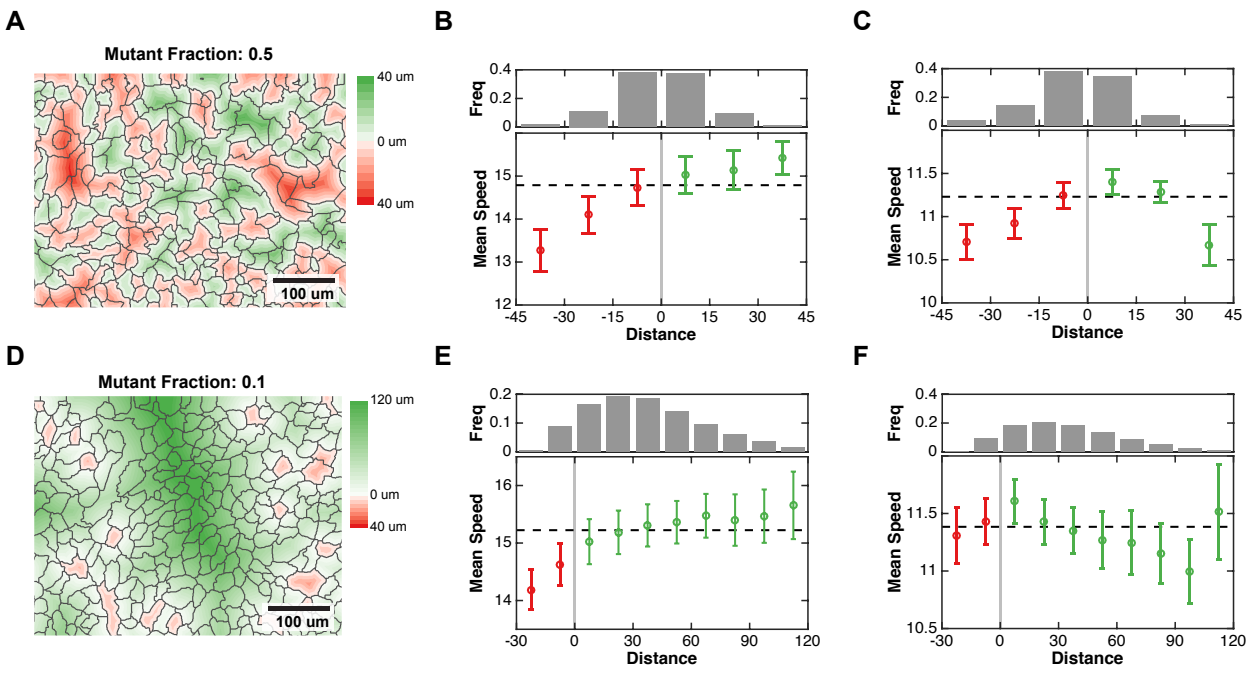


Figure 3.4: Effect of WT and polarity-deficient interactions on speed as a function of distance from cell-type border.

A. Heatmap of distance from the WT/ABP-deficient border within WT cells (positive, greens) and APB deficient cells (negative, reds) for an ABP-deficient fraction of 0.1. All cell boundaries are shown in dark grey. B. Mean speed as a function of distance from the WT/ABP-deficient border (lower plot) and the relative frequency of observations that fell within a specific distance bin (upper plot) for mixtures of WT and siPar6 cells. The cell type border is set to a distance of zero, and distances within ABP-deficient regions are negative while distances within WT regions are positive. C. Equivalent to B. for mixtures of WT and siScrib cells. D. Equivalent to A. for an ABP-deficient fraction of 0.5. E-F. Equivalent to B-C. for an ABP-deficient fraction of 0.5. Errorbars are SEM.

3.6 REFERENCES

Assémat, E., Bazellières, E., Pallesi-Pocachard, E., Le Bivic, A., and Massey-Harroche, D. (2008). Polarity complex proteins. Biochimica et Biophysica Acta (BBA)-Biomembranes *1778*, 614-630.

Bajpai, S., Correia, J., Feng, Y., Figueiredo, J., Sun, S.X., Longmore, G.D., Suriano, G., and Wirtz, D. (2008). {alpha}-Catenin mediates initial E-cadherin-dependent cell-cell recognition and subsequent bond strengthening. Proc Natl Acad Sci U S A *105*, 18331-18336.

Betschinger, J., Mechtler, K., and Knoblich, J.A. (2003). The Par complex directs asymmetric cell division by phosphorylating the cytoskeletal protein Lgl. Nature *422*, 326-330.

Coso, O.A., Chiariello, M., Yu, J.C., Teramoto, H., Crespo, P., Xu, N., Miki, T., and Gutkind, J.S. (1995). The small GTP-binding proteins Rac1 and Cdc42 regulate the activity of the JNK/SAPK signaling pathway. Cell *81*, 1137-1146.

Diggle, P.J., and Chetwynd, A.G. (1991). Second-order analysis of spatial clustering for inhomogeneous populations. Biometrics, 1155-1163.

Gatrell, A.C., Bailey, T.C., Diggle, P.J., and Rowlingson, B.S. (1996). Spatial point pattern analysis and its application in geographical epidemiology. Transactions of the Institute of British geographers, 256-274.

Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., and Tarpey, P. (2012). Intratumor heterogeneity and branched

evolution revealed by multiregion sequencing. New England journal of medicine *366*, 883-892.

Halaoui, R., and McCaffrey, L. (2015). Rewiring cell polarity signaling in cancer. Oncogene *34*, 939-950.

Heppner, G.H., and Miller, B.E. (1983). Tumor heterogeneity: biological implications and therapeutic consequences. Cancer and Metastasis Reviews *2*, 5-23.

Igaki, T., Pagliarini, R.A., and Xu, T. (2006). Loss of cell polarity drives tumor growth and invasion through JNK activation in Drosophila. Current Biology *16*, 1139-1146.

Knust, E., and Bossinger, O. (2002). Composition and formation of intercellular junctions in epithelial cells. Science *298*, 1955-1959.

Koride, S., He, L., Xiong, L.P., Lan, G., Montell, D.J., and Sun, S.X. (2014). Mechanochemical regulation of oscillatory follicle cell dynamics in the developing Drosophila egg chamber. Mol Biol Cell *25*, 3709-3716.

Low, B.C., Pan, C.Q., Shivashankar, G., Bershadsky, A., Sudol, M., and Sheetz, M. (2014). YAP/TAZ as mechanosensors and mechanotransducers in regulating organ size and tumor growth. FEBS letters *588*, 2663-2670.

Marusyk, A., and Polyak, K. (2010). Tumor heterogeneity: causes and consequences. Biochimica et Biophysica Acta (BBA)-Reviews on Cancer *1805*, 105-117.

Mayor, R., and Carmona-Fontaine, C. (2010). Keeping in touch with contact inhibition of locomotion. Trends in cell biology *20*, 319-328.

McCaffrey, L.M., Montalbano, J., Mihai, C., and Macara, I.G. (2012). Loss of the Par3 polarity protein promotes breast tumorigenesis and metastasis. Cancer cell *22*, 601-614.

Nowell, P.C. (1976). The clonal evolution of tumor cell populations. Science *194*, 23-28.

Qin, Y., Capaldo, C., Gumbiner, B.M., and Macara, I.G. (2005). The mammalian Scribble polarity protein regulates epithelial cell adhesion and migration through E-cadherin. The Journal of cell biology *171*, 1061-1071.

Ripley, B.D. (1976). The second-order analysis of stationary point processes. Journal of applied probability, 255-266.

Schroeder, M.C., and Halder, G. (2012). Regulation of the Hippo pathway by cell architecture and mechanical signals. Paper presented at: Seminars in cell & developmental biology (Elsevier).

Serra-Picamal, X., Conte, V., Vincent, R., Anon, E., Tambe, D.T., Bazellieres, E., Butler, J.P., Fredberg, J.J., and Trepat, X. (2012). Mechanical waves during tissue expansion. Nature Physics *8*, 628-634.

Warner, S.J., and Longmore, G.D. (2009). Cdc42 antagonizes Rho1 activity at adherens junctions to limit epithelial cell apical tension. The Journal of cell biology *187*, 119-133.

Zhan, L., Rosenberg, A., Bergami, K.C., Yu, M., Xuan, Z., Jaffe, A.B., Allred, C., and Muthuswamy, S.K. (2008). Deregulation of scribble promotes mammary tumorigenesis and reveals a role for cell polarity in carcinoma. Cell *135*, 865-878.

CHAPTER FOUR:

Stochastic simulation of a general reaction-transport system

4.1 ABSTRACT

Biological systems frequently contain countable and slow diffusing components, leading to fluctuations that invalidate deterministic approaches. The development of mathematical tools that account for the discrete nature of both number and spatial distribution of reacting species is vital for understanding cellular behavior and engineering biological circuits. Here we present an algorithm for an event-driven stochastic simulation of a general reaction transport process that bridges well-mixed and unmixed systems, becoming mathematically equivalent with the Gillespie algorithm in the limit of high diffusion. We develop a computational implementation of this algorithm with strategies that lead to advantageous scaling in computational time as the total number of particles in the system increases. To test this simulation method, we examine reaction and diffusion limited regimes of a bimolecular association-dissociation reaction. In the reaction limited regime where mixing occurs between individual reactions, equilibrium numbers of components match the expected values from mean field methods. In the diffusion limited regime, however, spatial correlations between newly dissociated species persist, leading rebinding events and a shift in the in the observed molecular counts. In the final part of this work, we examine how changes in enzyme efficiency can emerge from changes in diffusive mobility alone, as may result from protein complex formation.

4.2 INTRODUCTION

Biological processes are often characterized by slow-diffusing components that exist in countable numbers and reactions that depend nonlinearly on substrate concentrations and may not be at equilibrium (Huang et al., 2007; Ozbudak et al., 2004; Elf et al., 2003). In this realm, fluctuations in the number and spatial distribution of the participating species become non-trivial and can lead to functional consequences (Berg, 1978). This variation has been demonstrated to play a role in a wide variety of cellular processes, including fate in bacterial populations following viral infection or nutrient restriction (Arkin et al., 1998; Maamar et al., 2007), phenotypic variation in genetically identical populations (Elowitz et al., 2002; Blake et al., 2003, 2006), and cell differentiation in organism development (Wernet et al., 2006).

Mathematical models are critical for the understanding and prediction of biological systems, but deterministic models often break down in describing the dynamics of these biological systems (Mahmutovic et al., 2012). Molecular concentration becomes discontinuous in both domain, due to slow diffusion, and range, due to countable changes in the exact number of components. To understand the details of when and how fluctuations impact the behavior of biological systems, a variety of mathematical models have been developed to capture the discrete and stochastic nature of the systems involved (Gillespie, 1976; Hattne et al., 2005; Ander et al., 2004; Andrews et al., 2010; van Zon and Ten Wolde, 2005b; Oppelstrup et al., 2009). An early computational approach was provided by Gillespie for simulation of the chemical master equation (Gillespie, 1976). In the original work, spatial distribution of molecules was assumed to be uniform, but an extension of the model using subvolumes allowed for simulation of spatial heterogeneity provided each subvolume fulfilled the well-mixed requirement (Malek-Mansour and Houard, 1979; Ander et al., 2004; Hattne et al., 2005). A second class of simulation techniques relies on a molecular dynamics type approach, simulating the diffusive trajectories of species involved (Andrews et al., 2010). For simulations when the diffusive steps are small compared to inter-particle distance, this approaches can become inefficient. To remedy this, methods referred to as Green's Function Reaction Dynamics (GFRD)

and First-Passage Kinetic Monte Carlo (FPKMC) provided a strategy to jump forward in time to when interactions occur (van Zon and Ten Wolde, 2005b,a; Oppelstrup et al., 2009).

The ability to accurately simulate biochemical networks is critical for understanding and predicting cell and organism behaviors as well as for designing biological circuits that robustly perform specific functions. While these models present paths to examine the dynamics continuous and deterministic approaches, they retain a few disadvantages. Subvolume methods may be impacted by the selection of subvolume size and GFRD and FPKMC require careful treatment of certain spatial arrangements of molecules that, if not explicitly accounted for, can limit computational efficiency (Oppelstrup et al., 2009; Takahashi et al., 2010; Fange et al., 2010; Hepburn et al., 2012). Furthermore, processes that span timescales and length-scales may be coupled in biological systems, complicating the choices required to efficiently and accurately implement these algorithms.

Here we present the framework and implementation of an event driven stochastic simulation of a generalized reaction transport process which bridges diffusion-limited stochastic kinetics with well-mixed reaction kinetics, becoming mathematically equivalent to the Gillespie algorithm in the limit of rapid diffusion. In the first part of this work, we develop the underlying reaction probability density function and the sampling procedure to determine reaction type, location, and participating molecules. In the second part of this work, we provide an implementation for a bimolecular association-dissociation reaction and demonstrate the necessity for reaction-diffusion modeling in predicting expected values when re-binding probabilities are high. We then provide strategies for increasing computational efficiency for systems with large numbers of particles. Lastly, we examine the influence of protein complex formation on enzyme efficiency solely through changes in diffusive mobility.

4.3 SIMULATION ALGORITHM

System Definitions

Let the general system be defined as a set of M mobile species S_i that may interconvert through N reactions R_μ within a volume V. We are interested in considering systems with a countable number of members of each species and therefore define X_i as the current number of members of S_i. Individual particle identity is accounted for through a second subscript j such that S_{ij} refers to the j^{th} member of the i^{th} species, where $0 \leq j \leq X_i$.

Derivation of the Reaction Probability Density Function

The Gillespie algorithm requires the computation of the reaction probability density function, $P(\tau, \mu)$, defined as the probability at time τ that the next reaction in V will occur in the differential time interval $\delta\tau$ and will be of type μ (Gillespie, 1976). The goal of the present work is to extend this joint probability density function to include location and molecular identity by deriving:

$P(\tau, \mu, \vec{r}, S_{ij}) \equiv$ The probability at time t that the next reaction in V:

will occur in the differential interval $\delta\tau$;

and will be of type μ;

and will occur at position $\vec{r}$;

and will involve the specific molecules S_{ij}.

We first derive the univariate distribution $P(\tau)$ and then decompose it into the multivariate distribution $P(\tau, \mu, \vec{r}, S_{ij})$, which can be sampled using Monte Carlo methods. Here we focus on a bimolecular reaction between two distinct species, but the derivation can be generalized for multiple reaction types. The reaction is:

$$S_1 + S_2 \longrightarrow S_3 \tag{1}$$

Two molecules have the potential to react if they collide with proper orientation and energy. A collision will take place whenever the centers of two molecules come closer than:

$$d_{12} = (d_1 + d_2)/2, \tag{2}$$

Where d_{12} is the collision distance threshold, d_1 is the diameter of an S_1 molecule, and d_2 is the diameter of an S_2 molecule. If the average speed of an S_1 molecule relative to an S_2 molecule is $\langle v_{12} \rangle$, the S_1 molecule produces a potential collision volume over time that can be defined as:

$$\langle \delta V_{coll} \rangle = \pi {d_{12}}^2 \langle v_{12} \rangle \delta\tau \tag{3}$$

By determining the chance this collision volume is occupied by an S_2, the likelihood that the collision was successful, and the number of potential reactive collisions, Gillespie defines the probability that a reaction R_μ will occur in V in the next time interval $\delta\tau$ using the expression $h_\mu c_\mu \delta\tau$ (Gillespie, 1976). The present goal is to determine the value of this expression without making the assumption of spatial homogeneity. To begin, we define $z_\mu(\tau)\delta\tau$ as the probability that the next reaction R_μ will occur in V in the next time interval $\delta\tau$ beginning at time τ. Importantly, while $h_\mu c_\mu$ is constant with respect to time for any inter-reaction interval, z_μ may be a function of time.

The volume V can be broken into an infinite number of small sub-volumes where the next reaction could occur. Because the chance of two reactions occurring in the time interval $\delta\tau$ vanishes as $\delta\tau$ goes to zero, the probabilities of this single next reaction being in each sub-volume are mutually exclusive. This means, through the addition rule of disjoint events, that the sum of probabilities in all these sub-volumes is equal to $z_\mu \delta t$. In the limit of infinite sub-volumes, summation becomes integration, and we define a new variable, $\hat{z}_\mu$:

$$z_\mu(\tau) = \int_V \hat{z}_\mu(\tau, \vec{r}) d\vec{r} \tag{4}$$

Which is the probability of the next reaction occurring at a specific position $\vec{r}$ in the differential

interval beginning at time τ. In a general sense, $\hat{z}_\mu$ describes a conditional probability, namely:

$$\hat{z}_\mu = P(A, B) = P(A|B) \cdot P(B) \tag{5}$$

Where event A is an R_μ reaction and event B is the presence of an S_1 molecule at a location. Computation of $P(B)$ will be addressed first, followed by $P(A|B)$.

The term $P(B)$, or the chance that an S_1 molecule is at a specific location, can be defined as follows. If the location of a molecule is known at a time $t = 0$, its location at $t \geq 0$ is unknown precisely, but it is describable in terms of probability if the driving forces behind motion are known. Here we assume that thermal fluctuations drive molecular motion, yielding a Brownian diffusion process, however this method can be generalized to any source of motion that can be described in terms of an probability density function for each member of each species: $P_{S_{ij}}(\vec{r}, t)$. When multiple S_i molecules are present, a pseudo-probability density function describing probabilities of finding S_i at a point can be written:

$$\mathbb{P}_{S_i}(\vec{r}, \tau) = \sum_{j=1}^{X_i} P_{S_{ij}}(\vec{r}, \tau) \tag{6}$$

While different $P_{S_{ij}}$ are independent, they are not disjoint because describe the same region of space. Therefore, the sum rule of probabilities $(P(C \cup D) = P(C) + P(D) - P(C \cap D))$ applies for finding the union of the individual $P_{S_{ij}}$ probability distributions. The cost of neglecting the intersection term depends on the absolute probabilities of both events, and as the absolute probabilities of the distinct events approaches, the fractional error from neglecting their intersection approaches zero. Therefore, if we consider a region where all the individual $P_{S_{ij}}$ are high and ask what the probability that any S_i molecules are in that region, the integration of $\mathbb{P}_{S_i}$ over the spatial limits of that region will be a poor approximation of the true probability (integrating $\mathbb{P}_{S_i}$ over V makes this readily apparent, as a value greater than 1 is seen). If, however, the region of interest is very small, such as δV_{coll}, then the chance of individual molecules being within the region is small (computed by integrating $P_{S_{ij}}$ over the limits of the region δV_{coll}), and as $\delta\tau$ drops to zero, the chance of two particles being within the collision volume is negligible. Thus, $P(B)$ in Eq.

5, or the probability that an S_1 molecule is within an infinitesimal volume in space, can be well approximated by $\mathbb{P}_{S_1}$:

$$P(B) = \mathbb{P}_{S_1} \tag{7}$$

The first term on the right hand side of (5), $P(A|B)$, is the probability that a reaction will occur given there is an S_1 molecule at a specific location. Following the logic used by Gillespie, we break this into the chance that an S_2 molecule is within the collision volume of a given S_1 molecule and the chance that such a collision is productive. Using θ_R as the chance a collision is productive, $P(A|B)$ becomes:

$$P_{A|B}(\vec{r}) = \int_{\delta V_{coll}} \mathbb{P}_{S_2}(\vec{r}) d\vec{r} \cdot \theta_R \tag{8}$$

Where $\vec{r}$ is the location of a given S_1 molecule and δV_{coll} is the collision volume as defined in (3). The sum rule approximation employed here by the use of $\mathbb{P}_{S_2}$ is safe because the collision volume is vanishingly small. The probability of any single molecule being within it is small, so any higher order terms that result from the product of individual probabilities will be negligible. To simplify this expression, we recognize that δV_{coll} is small such that $\mathbb{P}_{S_2}$ is essentially constant throughout this domain. Thus (8) becomes:

$$P_{A|B} = \mathbb{P}_{S_2} \delta V_{coll} \cdot \theta_R \tag{9}$$

We now can re-write (5) substituting in Eqs. (3), (7), (9):

$$\hat{z}_\mu \delta t = \mathbb{P}_{S_1} \mathbb{P}_{S_2} \left(\pi {d_{12}}^2 \langle v_{12} \rangle \right) \theta_r \cdot \delta t \tag{10}$$

Integrating $\hat{z}_\mu \delta t$ with respect to its spatial domain over a volume V gives the probability that a reaction occurs in the volume, $z_\mu \delta t$, as defined in (4). The only terms that depend on space are the

pseudo-probability distributions, $\mathbb{P}_{S_1}$ and $\mathbb{P}_{S_2}$.

$$z_\mu(\tau)\delta t = \int_V \mathbb{P}_{S_1}\mathbb{P}_{S_2}\left(\pi {d_{12}}^2\langle v_{12}\rangle\right)\theta_R d\vec{r}\cdot\delta t \tag{11}$$

$$z_\mu(\tau)\delta t = \left(\pi {d_{12}}^2\langle v_{12}\rangle\right)\theta_R \int_V \mathbb{P}_{S_1}\mathbb{P}_{S_2} d\vec{r}\cdot\delta t \tag{12}$$

Defining the quantity of $z_\mu\delta t$ allows us to determine the reaction probability distribution function $P(\tau,\mu)$, the probability that the next reaction occurs at time τ and is a μ reaction. This probability can be decomposed into two components. The first components is $P_0(\tau)$, the probability that no reaction occurs over time τ. The second component is $z_\mu(\tau)d\tau$, the probability that a reaction occurs in an infinitesimal window right after τ and has already been defined in (12). Because reactions are independent, random events that occur with an underlying rate, the Poisson distribution describes their occurrences. $P_0(\tau)$ is the zero-event probability density function resulting from the inhomogeneous Poisson process governed by the time-dependent event rate $z_\mu(\tau)$. Therefore the chance that no reactions (of any type, not just μ) occur in a time interval τ is given by the following equations:

$$P_0(\tau) = P(n_R(\tau)=0) = \frac{\Lambda(\tau)^0}{0!}\exp -\Lambda(\tau) \tag{13}$$

$$\Lambda(\tau) = \int_0^\tau \left(\sum_{\nu=1}^{N} z_\nu(\xi)\right) d\xi, \tag{14}$$

Where n_R is the number of reactions and ξ is a dummy time-variable. Combining (14) and (12), we can now express $P(\tau,\mu)$ in terms of the two parts described above.

$$P(\tau,\mu) = z_\mu(\tau)\exp\left(-\int_0^\tau\left[\sum_{\nu=1}^{N} z_\nu(\xi)\right]d\xi\right) \tag{15}$$

The chance that a specific R_μ reaction happens is a function of all other possible reactions due to the summation term in the exponential. For the next reaction to be of type R_μ, the reactions $R_{\nu\neq\mu}$ and the specific reaction R_μ must not yet have occurred. The final step in computing $P(\tau)$ is to

sum over all possible reaction types:

$$P(\tau) = \sum_{\mu=1}^{N} P(\tau, \mu) = \left(\sum_{\mu=1}^{N} z_\mu(\tau) \right) \exp\left(- \int_0^\tau \left[\sum_{\nu=1}^{N} z_\nu(\xi) \right] d\xi \right) \tag{16}$$

This is the probability density function describing the likelihood of the wait-time until the next reaction. To verify that this probability density function is properly normalized over its defined domain, we must show that $\int_0^\infty P(\tau) = 1$. We recognize that (16) takes the following form:

$$P(\tau) = f'(\tau) e^{-\int_0^\tau f'(\xi) d\xi} \tag{17}$$

Where $f'(\tau) = \left(\sum_{\mu=1}^{N} z_\mu(\tau) \right)$. For the set of functions f where $f(0) = 0$ and $f(\tau > 0)$ integrating yields:

$$\int_0^\infty P(\tau) = \int_0^\infty f'(\tau) e^{-(f(\tau) - f(0))} = -e^{-(f(\tau) - f(0))} \Big|_0^\infty = 1 - e^{-(f(\infty) - f(0))} \tag{18}$$

For the set of functions where $f'(\tau) \geq 0$ and therefore $f(\tau) \geq f(0)$, (18) is bounded with the following inequality:

$$0 \leq 1 - e^{-(f(\infty) - f(0))} \leq 1 \tag{19}$$

This is true here because for all R_μ reactions, $z_\mu(\tau) \geq 0$. The circumstances where $\int_0^\infty P(\tau) \neq 1$ describe situations where there is a finite probability that no further reactions occur and provide insight on the reactive process: for a reaction to occur, $f(\infty)$ must not remain bounded and therefore f' must not decay to 0.

Gillespie Algorithm Equivalence

We are now at a point to ask if this reaction probability density function is equivalent to the Gillespie derivation in the limit of a fast diffusion rates relative to reaction rates. We begin by demonstrating the following equality:

$$z_\mu = h_\mu c_\mu \tag{20}$$

In a system governed by the limit of fast diffusion, each molecule has equal probability of being anywhere within V, so the pseudo-probability distributions defined in (6) are sums of the uniform distribution over V:

$$\mathbb{P}_{S_1} = \frac{X_1}{V}, \mathbb{P}_{S_2} = \frac{X_2}{V} \tag{21}$$

From (Gillespie, 1976), the right hand side of (20) is equal to:

$$h_\mu c_\mu \delta t = X_1 X_2 \left(V^{-1} \pi {d_{12}}^2 \langle v_{12} \rangle \delta t \right) \theta_R \tag{22}$$

Substituting the values from (21) into (12) gives:

$$z_\mu(\tau)\delta t = \left(\pi {d_{12}}^2 \langle v_{12} \rangle \right) \theta_R \int_V \left[\frac{X_1}{V} \cdot \frac{X_2}{V} \right] d\vec{r} \cdot \delta t \tag{23}$$

All the terms within the integral are constants with respect to position, therefore (23) reduces to:

$$z_\mu(\tau)\delta t = \left(\pi {d_{12}}^2 \langle v_{12} \rangle \right) \theta_R \left(X_1 X_2 V^{-1} \right) \cdot \delta t \tag{24}$$

Which is exactly equal to $h_\mu c_\mu \delta t$ in (22). To establish equivalence of $P(\tau)$ in this limit, we recognize that z_μ defined in (24) is time independent and the inhomogenous Poisson process used to derive (16) becomes a homogenous one:

$$P(\tau) = \sum_{\mu=1}^{N} z_\mu \exp\left(\sum_{\nu=1}^{N} z_\nu \tau \right) \tag{25}$$

Which, using the demonstrated relationship in (20), is equivalent to $P(\tau)$ of the Gillespie algorithm (Gillespie, 1976).

Sampling $P(\tau, \mu, \vec{r}, S_{ij})$ by decomposition of $P(\tau)$

The joint probability distribution $P(\tau, \mu, \vec{r}, S_{ij})$ may be sampled through the use of conditional probabilities in a series of steps:

1. What is the wait time τ?
2. Given a wait time τ, what reaction R_μ took place?
3. Given a wait time τ and a specific reaction R_μ, where did it take place?
4. Given a wait time τ, specific reaction R_μ, and location $\vec{r}$, which molecules S_{ij} were involved?

These questions are the four terms on the right-hand side of the following equation and may be answered by sampling each distribution using Monte Carlo methods:

$$P(\tau, \mu, \vec{r}, S_{ij}) = P(\tau)P(\mu|\tau)P(\vec{r}|\tau, \mu)P(S_{ij}|\tau, \mu, \vec{r}) \tag{26}$$

In this section, we proceed through the sampling of each distribution using a computational implementation for a simple bimolecular reaction as a guide (Fig 4.1A).

What is the wait time τ? The reaction wait-time τ can be directly obtained by inversion sampling of the probability density function $P(\tau)$ shown in Eq (16), providing the time at which the next reaction occurs (Fig 4.1B).

Given a wait time τ, what reaction R_μ took place? The conditional probability $P(\mu|\tau)$ can be expressed using (15) and (16) as:

$$P(\mu|\tau) = \frac{P(\tau, \mu)}{P(\tau)} = \frac{z_\mu(\tau)}{\sum_{\mu=1}^{N} z_\mu(\tau)} \tag{27}$$

Which is the the proportion of the total event probability that each reaction contributed. This is a discrete distribution with respect to the reaction variable μ and can be sampled to obtain the type of reaction that occurred (Fig 4.1B inset).

Given a wait time τ and a specific reaction R_μ, where did it take place? The third question is answered by sampling the probability distribution $P(\vec{r}|\tau,\mu)$, which represents the relative contribution of each point in space towards the occurrence of the reaction R_μ and is given by the following equation:

$$P(\vec{r}|\tau,\mu) = \frac{P(\tau,\mu,\vec{r})}{P(\tau,\mu)} = \frac{\hat{z}_\mu(\vec{r},\tau)}{z_\mu(\tau)} \tag{28}$$

Through use of equations (10) and (12). Additionally, because $\hat{z}_\mu$ itself is a joint probability distribution over the number of spatial dimensions in the system, further decomposition needs to be done to sample each dimension (Fig 4.1C).

Given wait time τ, specific reaction R_μ, and location $\vec{r}$, which molecules S_{ij} were involved? To answer the final question, we need to sample the probability distribution $P(S_{ij}|\tau,\mu,\vec{r})$. That is we need to answer the question of given a reaction, time and position, which of the involved molecules participated. This quantity can be expressed by the following:

$$P(S_{ij}|\tau,\mu,\vec{r}) = \frac{P_{S_{ij}}(\vec{r})}{\mathbb{P}_{S_i}(\vec{r})} \tag{29}$$

This is a discrete probability distribution where the probability of the j^{th} molecule being picked is the proportion of $\mathbb{P}$ it contributes at point $\vec{r}$. Sampling of this probability distribution is again performed using the inversion method as described in Question 2. This computation is performed for each species participating in the reaction (Fig 4.1D and E).

The reaction step has now been fully characterized. The reacting particles can be removed from the list of S_{ij} and the created particle can be added to S_{ij} at the location of the reaction, updating the quantities X_i. The final step is to update the positions of the molecules that did not react this step by sampling the individual probability density functions $P_{S_{ij}}(\vec{r})$. This completes one round of the simulation and leaves the system ready for the next (Fig 4.1F).

4.4 RESULTS

Diffusion-Influenced Bimolecular Reactions

To verify that this algorithm produces expected behavior from reaction-limited to diffusion-limited regimes, we simulated a simple bimolecular association-dissociation reaction on a two dimensional domain. In two dimensions, effects due to local concentration differences as might arise in a reaction-diffusion system are exaggerated because the probability of two given molecules reacting is much more strongly dependent on separation distance than in three dimensions (Mahmutovic 2012). The bimolecular reaction is defined as $S_1 + S_2 \rightleftarrows_{k_2}^{k_1} S_3$, where k_1 and k_2 are the microscopic rate constants for association and dissociation, respectively.

This reaction was simulated across a range of values for k_2/k_1 and diffusion constants for species S_1 and S_2. Under conditions of fast diffusion where mixing occurred between reactions, the mean count of S_1 closely matched the mean-field expected value Fig 4.2 (blue points). In the setting of slow diffusion, the effective association rate is expected to depart from the microscopic value due to the persistence of spatial correlations in molecular position created by the dissociation reaction. The dissociation of an S_3 molecule produces an S_1 molecule and an S_2 molecule separated by only a short distance. These two molecules can either diffuse away from each other or re-associate. Under conditions of fast diffusion, re-association is less likely and the temporary spatial correlation between the two reactants is more quickly lost. Under conditions of slow diffusion, re-association frequently occurs, shifting the expected values for species counts. Indeed as the diffusion constants for S_1 and S_2 molecules were lowered, the observed S_1 count shifted to the right, indicating that fewer S_1 molecules were present than expected given the rate constants alone Fig 4.2 (red and yellow points).

Efficient Computational Strategies

The proposed reaction transport algorithm has the potential to be computationally intensive, especially in regards to the generation of the pseudo-probability distributions $\mathbb{P}_{S_i}(\vec{r}, \tau)$, which requires

the summation of individual probability distributions for each member of a species. However, $P_{S_{ij}}$ evolves identically for a single species except for a shift in position, providing an alternative strategy for computation employing convolution. For particles whose positions evolve according to a general probability distribution function $f(\vec{r},t)$, the function $\mathbb{P}_{S_i}(\vec{r},\tau)$ can be re-defined as:

$$\mathbb{P}_{S_i}(\vec{r},t) = \sum_j f(\vec{r},t) * \delta(\vec{r} - \vec{r_j}) \tag{30}$$

Where δ is the Dirac delta function. This convolution may be implemented directly or through use of Fourier transforms:

$$\mathbb{P}_{S_i}(\vec{r},t) = \mathcal{F}^{-1}\left[\mathcal{F}(f(\vec{r},t)) \cdot \mathcal{F}\left[\sum_j \delta(\vec{r} - \vec{r}_j)\right]\right] \tag{31}$$

Where $\mathcal{F}$ is the Fourier transform operator. In this approach, the summation of complex functions is avoided by converting calculations to a single multiplication step in frequency-space.

To demonstrate the computational efficiency of the convolution strategy, the time required to compute $\mathbb{P}_{S_i}(\vec{r},\tau)$ for an increasing number of particles was performed. The direct summation method was compared to direct convolution and Fourier transform based convolution where the total number of particles, N, ranged from 10 to 10^4. While computation time scaled linearly with N for the direct sum method, computation time for both convolution based approaches was independent of particle number (Fig 4.3). This reflects optimum scaling that may or may not be realizable for all computational implementations of the simulation. It may be desirable to keep the number of elements in the numerical discretization constant on a per-particle level such that the numerical resolution scales with the particle number to maintain the same strength in the assumption underlying Eq (6). In this case, the number of elements in the simulated domain scales with the number of particles. Therefore the efficiency becomes bounded by that of the Fast Fourier Transform algorithm at $N\log(N)$. Regardless of which approach is taken when increasing the number of particles within a simulation, a computational efficiency that scales for N particles between a lower bound of zero-order and upper bound of $N\log(N)$ is a highly advantageous feature of this algorithm.

Diffusive Mobility and Enzyme Efficiency

The formation of protein complexes is a ubiquitous part of biological systems (Krogan et al., 2006; Levy et al., 2008). Much work has focused on understanding how conformational changes that result from these interactions can modulate intrinsic reaction rates through allostery (Kuriyan and Eisenberg, 2007), but the effect that the reduction in diffusive mobility accompanying complex formation is less well understood. The reduction in mobility may be a key feature of the system, contributing to subcellular co-localization of components, and a detailed understanding of its impact requires an approach that captures the details of reaction-diffusion processes. In this section we examine a model reaction-diffusion system motivated by understanding the effect that protein complex formation may have on enzymatic activity simply through the change in diffusivity that results from forming a macromolecular assembly.

Consider a system where a substrate is produced in a spatially restricted manner and acted on by an enzyme that is capable of forming complexes with other proteins. The reaction scheme used in the simulation is shown in Fig 4.4A. A molecule S is produced by an essentially fixed point source and diffuses away subject to degradation, creating a spatial gradient in the probability of encountering S. An enzyme A binds S and converts it to P. Alternatively, while bound to S, A can be altered in a manner that leads to a change in mobility but does not affect intrinsic enzymatic efficiency. This scenario represents a generalization of a number of specific mechanisms that could lead to spatial localization of an enzyme near a source of substrate. This may reflect a change in phosphorylation state or conformation that leads to protein complex formation, increased lipid-bilayer binding, or any alteration that reduces the diffusion rate of A. As a result, the mobility of A is reduced in a region where S is likely to be found.

This system provides a generalized scenario for examining the effect that protein complex formation has on enzymatic efficiency even if intrinsic catalytic rates remain constant. Figure 4B shows a snapshot of the enzyme positions at the conclusion of 8000 reaction time-steps for the case of no mobility reduction and a 1000-fold reduction. As expected, enzymes where mobility reduction is high are concentrated in the vicinity of the source of S, indicating successful recruitment

near the source of substrate due to the reduction in mobility. The production of P depends on the relative mobility of A^* and is higher in cases where reduction is the greatest but changes as the system evolves in time (Fig 4.4D-F). For large reductions in mobility, the production rate of P is higher than if no reduction occurs and remains so for thousands of reaction steps. In this case, although the enzymes search space less quickly, they remain in a region with an increased probability of encountering S, leading to higher overall efficiency. For intermediate drops in mobility, however, the production rate of P initially exceeds but eventually falls below the control level. Here, diffusion is not reduced enough to maintain the high degree of initial spatial correlation with the source of S to balance the effect on search-times.

4.5 DISCUSSION

In this work, we use probabilistic techniques to simulate a general reaction transport system. The approach is driven by reaction events and is capable of spanning diffusion limited to well-mixed regimes, becoming mathematically identical to the Gillespie algorithm under conditions of rapid diffusion. Using a computational implementation of this algorithm, we demonstrate that it captures equilibrium values in well-mixed systems and re-binding events when diffusion is slow. Lastly, in a model of protein complex formation, we find tradeoffs may exist between spatial localization and search-time due to macromolecular assembly. For simplicity, this algorithm will be referred to as the Gillespie Reaction Transport algorithm (GRT).

The GRT method relies on probabilistically jumping between events in order to skip details that do not change the progression of the system in a manner similar to GFRD and FPKMC methods (van Zon and Ten Wolde, 2005b,a; Oppelstrup et al., 2009). The primary difference between these methods and GRT is the definition of an event and the construction of the underlying wait time distribution. GFRD and FPKMC are event driven, but the event definition is related to chance of multi-way-interactions rather than successful reactions. In the regime of highly diffusion-limited reactions, this is equivalent, but when interactions do not guarantee reactions, these definitions differ. In this case, both GFRD and FPKMC increment through time by simulating many unsuccessful interaction events before a reaction occurs. By focusing on reactive events, the GRT algorithm avoids this potential inefficiency that may arise when a near-diffusion-limited process or even a reaction-limited process is part of the system. As a result, the GRT algorithm can simultaneously simulate reaction-limited (highlighted by its Gillespie algorithm equivalence) and diffusion-limited processes that share components. This hybrid functionality is a fundamental strength of the GRT algorithm.

To achieve this, the GRT algorithm relies on fundamental assumptions in both the analytical definition and computational implementation. The primary assumption of the analytical definition is that non-reactive collisions with labeled particles (those that have been given an identity in the

system) do not significantly alter the random-walk statistics that arise from collisions with all other unlabeled particles in the environment. This assumption may not hold if electrostatic interactions are involved that disrupt diffusive motion (Liu et al., 2005). This is a challenge not unique to this approach, but may be handled here by defining the diffusion-altering event as a separate reaction in the system to create an intermediate species for more precise treatment.

The primary assumption in the numerical implementation of the GRT algorithm is in the numerical resolution used for the pseudo-probability distribution $\mathbb{P}_S$, which approximates the chance that a single molecule resides at a position. In theory, collision volumes are vanishingly small and the chance that two molecules simultaneously fall within it approaches zero so the union probability is well approximated by summing individual probabilities and neglecting the intersection probabilities. In practice, the minimum value that individual probabilities may take depends on the resolution chosen for numerical approximation of the individual probability density functions. Therefore, to maintain the accuracy of this assumption, the numerical resolution must be kept high enough that the probability of a single particle of a given type residing in a numerical discretization of space is small such that the probability of two remains negligible.

The analysis on computational efficiency demonstrates that the algorithm exhibits efficient scaling as the number of particles increases when convolution-based methods are employed. In the best-case scenario, when the resolution of numerical implementation does not need to be changed, scaling is zero-order: additional particles do not effect the computational time for a given reaction. If the numerical resolution is scaled with the number of particles, the computational complexity then follows that of the method chosen for convolution. While there is significant start-up cost due to the required to implement the GRT algorithm, its reaction-event driven nature and this scaling are highly advantageous features as parallelized CPU and GPU computing power continues to grow.

The simple model of protein complex formation presented here demonstrates two fundamental competing factors in reaction diffusion systems: control of protein localization and effects on search times. In all cases, the enzymes A are likely to undergo a reduction in mobility where S molecules are more prevalent and do so with the same kinetics. This localization in a zone of higher levels of

S leads to increased production rates. As time progresses, enzymes that only underwent a moderate reduction in mobility begin to spread out and production drops. These enzymes are slower to search space and are no longer restricted to a region with high local concentrations of S and as a result, production falls below that of the case where no reduction in mobility occurs.

In summary, protein complex formation has an advantageous ability to increase effective enzyme efficiency through localization near a source of substrate. However, maintenance of spatial correlation with the source becomes of increased importance to avoid the disadvantageous lengthening of search-times that results from mobility reduction. This tradeoffs may be balanced through multiple mechanisms in biological systems. First, protein complex formation often increases intrinsic enzyme efficiency (Kuriyan and Eisenberg, 2007) and suggests that under certain conditions, this may be a requirement to maintain fitness. Alternatively, spatial correlation can be maintained by directly coupling the relative motions of the proteins involved through scaffolds or adaptor proteins (Good et al., 2011). These results here suggest that scaffolding of low-mobility molecular complexes that share reactants and products may be especially important. The methods proposed here provide a rigorous method for assessing these effects and the conditions where these considerations become important. which may aide in understanding the output of signaling systems or designing components of molecular circuits.

4.6 ACKNOWLEDGMENTS

We would like to acknowledge Mauricio Del Razo Sarmina and Hong Qian for their valuable discussions and constructive feedback.

4.7 REFERENCES

M. Ander, P. Beltrao, B. Di Ventura, J. Ferkinghoff-Borg, M. Foglierini, A. Kaplan, C. Lemerle, I. Tomas-Oliveira, and L. Serrano. Smartcell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks. *Syst. Biol*, 1(1):129--138, 2004.

S. S. Andrews, N. J. Addy, R. Brent, and A. P. Arkin. Detailed simulations of cell biology with smoldyn 2.1. *PLoS Comput Biol*, 6(3):e1000705, 2010. ISSN 1553-7358.

A. Arkin, J. Ross, and H. H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected escherichia coli cells. *Genetics*, 149(4):1633--1648, 1998. ISSN 0016-6731.

O. G. Berg. A model for the statistical fluctuations of protein numbers in a microbial population. *Journal of theoretical biology*, 71(4):587--603, 1978. ISSN 0022-5193.

W. J. Blake, M. Kærn, C. R. Cantor, and J. J. Collins. Noise in eukaryotic gene expression. *Nature*, 422(6932):633--637, 2003. ISSN 0028-0836.

W. J. Blake, G. Balázsi, M. A. Kohanski, F. J. Isaacs, K. F. Murphy, Y. Kuang, C. R. Cantor, D. R. Walt, and J. J. Collins. Phenotypic consequences of promoter-mediated transcriptional noise. *Molecular cell*, 24(6):853--865, 2006. ISSN 1097-2765.

J. Elf, J. Paulsson, O. G. Berg, and M. Ehrenberg. Near-critical phenomena in intracellular metabolite pools. *Biophysical journal*, 84(1):154--170, 2003. ISSN 0006-3495. URL `http://ac.els-cdn.com/S0006349503748395/1-s2.0-S0006349503748395-main.pdf?_tid=7aaa82be-60f6-11e6-b072-00000aacb35e&acdnat=1471052421_b46ab528a21c1ab24961bc0ac3d4e8e8`.

M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183--1186, 2002. ISSN 0036-8075.

D. Fange, O. G. Berg, P. Sjöberg, and J. Elf. Stochastic reaction-diffusion kinetics in the microscopic limit. *Proceedings of the National Academy of Sciences*, 107(46):19820--19825, 2010. ISSN 0027-8424.

D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4):403--434, 1976. ISSN 0021-9991.

M. C. Good, J. G. Zalatan, and W. A. Lim. Scaffold proteins: hubs for controlling the flow of cellular information. *Science*, 332(6030):680--686, 2011. ISSN 0036-8075.

J. Hattne, D. Fange, and J. Elf. Stochastic reaction-diffusion simulation with mesord. *Bioinformatics*, 21(12):2923--2924, 2005. ISSN 1367-4803.

I. Hepburn, W. Chen, S. Wils, and E. De Schutter. Steps: efficient simulation of stochastic reaction--diffusion models in realistic morphologies. *BMC systems biology*, 6(1):1, 2012. ISSN 1752-0509.

B. Huang, H. Wu, D. Bhaya, A. Grossman, S. Granier, B. K. Kobilka, and R. N. Zare. Counting low-copy number proteins in a single cell. *Science*, 315(5808):81--84, 2007. ISSN 0036-8075.

N. J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, and A. P. Tikuisis. Global landscape of protein complexes in the yeast saccharomyces cerevisiae. *Nature*, 440(7084):637--643, 2006. ISSN 0028-0836.

J. Kuriyan and D. Eisenberg. The origin of protein interactions and allostery in colocalization. *Nature*, 450(7172):983--990, 2007. ISSN 0028-0836.

E. D. Levy, E. B. Erba, C. V. Robinson, and S. A. Teichmann. Assembly reflects evolution of protein complexes. *Nature*, 453(7199):1262--1265, 2008. ISSN 0028-0836.

Y. Liu, E. Fratini, P. Baglioni, W.-R. Chen, and S.-H. Chen. Effective long-range attraction between protein molecules in solutions studied by small angle neutron scattering. *Physical review letters*, 95(11):118102, 2005.

H. Maamar, A. Raj, and D. Dubnau. Noise in gene expression determines cell fate in bacillus subtilis. *Science*, 317(5837):526--529, 2007. ISSN 0036-8075.

A. Mahmutovic, D. Fange, O. G. Berg, and J. Elf. Lost in presumption: stochastic reactions in spatial models. *Nature methods*, 9(12):1163--1166, 2012. ISSN 1548-7091. URL `http://www.nature.com/nmeth/journal/v9/n12/pdf/nmeth.2253.pdf`.

M. Malek-Mansour and J. Houard. A new approximation scheme for the study of fluctuations in nonuniform nonequilibrium systems. *Physics Letters A*, 70(5):366--368, 1979. ISSN 0375-9601.

T. Oppelstrup, V. V. Bulatov, A. Donev, M. H. Kalos, G. H. Gilmer, and B. Sadigh. First-passage kinetic monte carlo method. *Physical Review E*, 80(6):066701, 2009.

E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman, and A. Van Oudenaarden. Multistability in the lactose utilization network of escherichia coli. *Nature*, 427(6976):737--740, 2004. ISSN 0028-0836. URL `http://www.nature.com/nature/journal/v427/n6976/pdf/nature02298.pdf`.

K. Takahashi, S. Tănase-Nicola, and P. R. Ten Wolde. Spatio-temporal correlations can drastically change the response of a mapk pathway. *Proceedings of the National Academy of Sciences*, 107(6):2473--2478, 2010. ISSN 0027-8424.

J. S. van Zon and P. R. Ten Wolde. Green's-function reaction dynamics: a particle-based approach for simulating biochemical networks in time and space. *The Journal of chemical physics*, 123(23):234910, 2005a. ISSN 0021-9606.

J. S. van Zon and P. R. Ten Wolde. Simulating biochemical networks at the particle level and in time and space: Green's function reaction dynamics. *Physical review letters*, 94(12): 128103, 2005b.

M. F. Wernet, E. O. Mazzoni, A. Çelik, D. M. Duncan, I. Duncan, and C. Desplan. Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature*, 440(7081):174--180, 2006. ISSN 0028-0836.

4.8 FIGURES

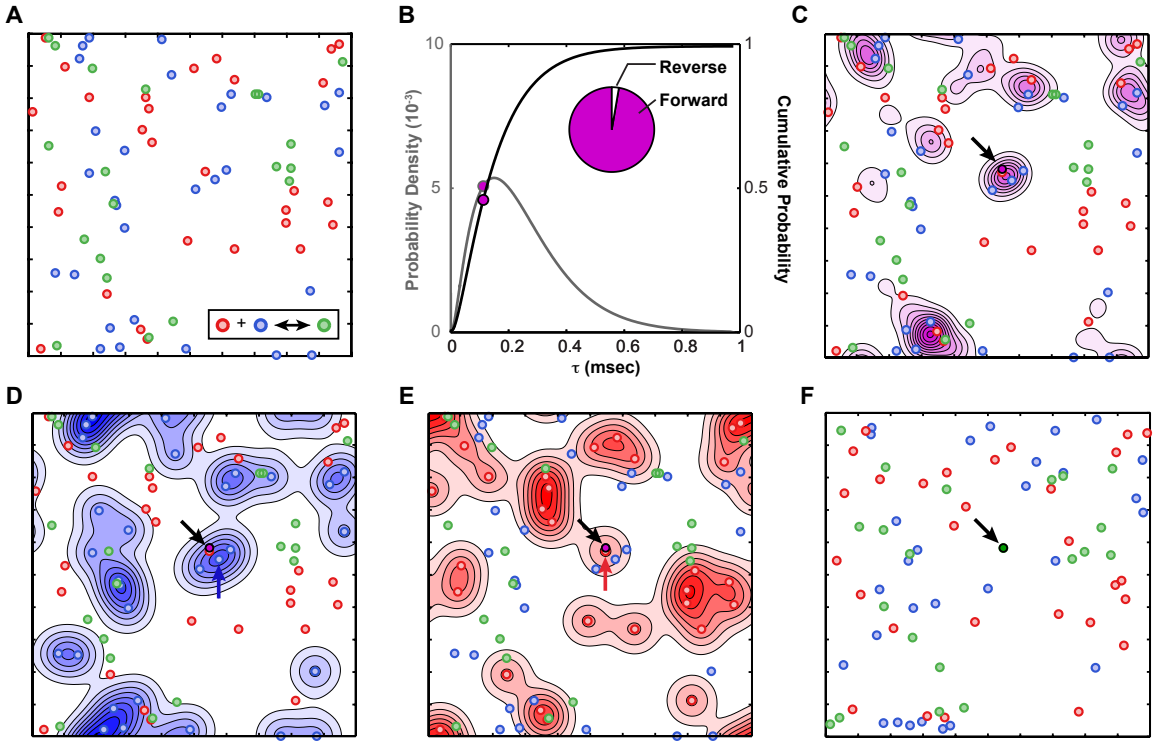


Figure 4.1: Simulation of a single reaction step.

A. Initial positions of molecules. Molecules are randomly distributed over the environment. B. Reaction probability density function, $P(\tau)$ shown in blue and the cumulative probability is shown in green. The cumulative probability distribution is sampled via the inversion method, and the resulting value of τ is the wait time (red dots). The relative probabilities of the forward and reverse reactions occurring at this wait time is shown in the inset; random sampling selects the forward reaction. C. The spatial distribution of the reaction likelihood, $\hat{z}_\mu$, and the sampled reaction location (green dot). D, E. The location density functions $\mathbb{P}_{S_{ij}}$ for S_1 and S_2 respectively. The S_1 and S_2 molecules likely to have participated are selected from these distributions. F. The updated system with removal of reacting S_1 and S_2, new S_3, and propagation of non-reacting molecules.

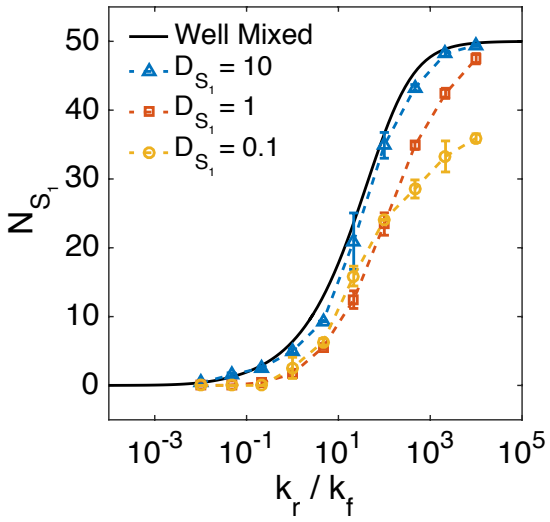


Figure 4.2: Association-dissociation reaction with re-binding effects.

Expected number of S_1 molecules computed under a mean-field approximation or simulated as a function of reactant diffusion constant. The results of the mean-field approach are shown in the black line, and simulation results are shown by the dashed lines and symbols. The diffusion constant for S_1 ranged from 10 $\mu m^2 s^{-1}$ to 0.1 $\mu m^2 s^{-1}$. Additional parameter values used in the simulation are as follows: initial S_1 = 30, initial S_2 = 30, initial S_3 = 20, $D_{S_1} = D_{S_2}$, $D_{S_3} = D_{S_1}/\sqrt{2}$, k_f= 10 $\mu m^2 s^{-1}$ and k_r is set by k_f and the ratio given by the x-axis. The simulation size was 1 μm by 1 μm, solved on a 200 by 200 point numerical domain.

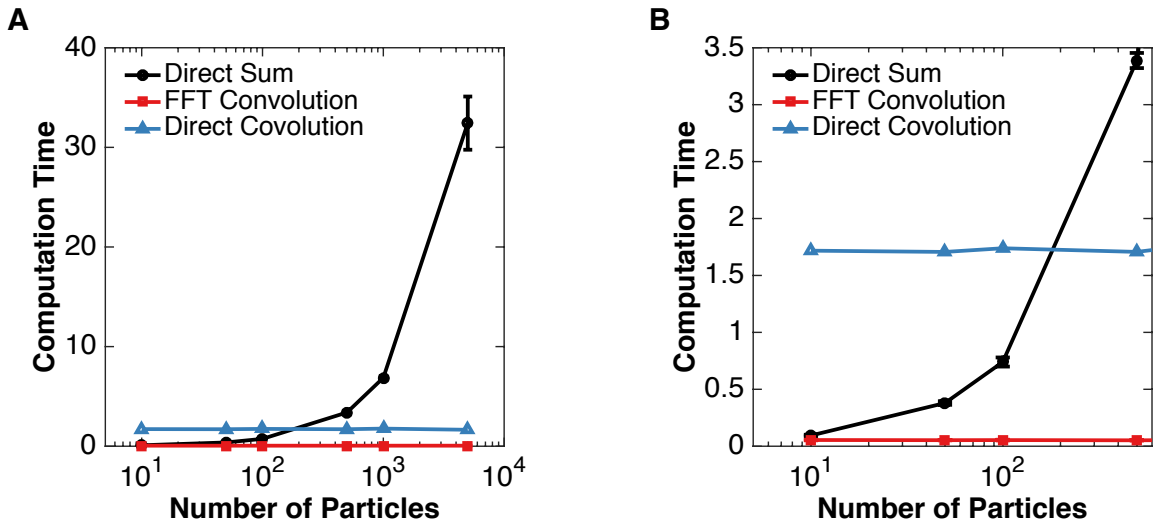


Figure 4.3: Comparison of computational efficiency for summation and convolution based approaches.

A. Time to compute $\mathbb{P}_S$ as a function of number of S molecules for summation, direct convolution, and Fourier transform convolution. B. Magnification of cases between 10 and 500 molecules to highlight relative efficiencies for small numbers of molecules. Parameter values used for testing computational efficiency were: D = 1 $\mu m^2 s^{-1}$, $\tau = 10^{-3}$ s. The domain size was 1 μm by 1 μm discretized on a 1024 by 1024 grid.

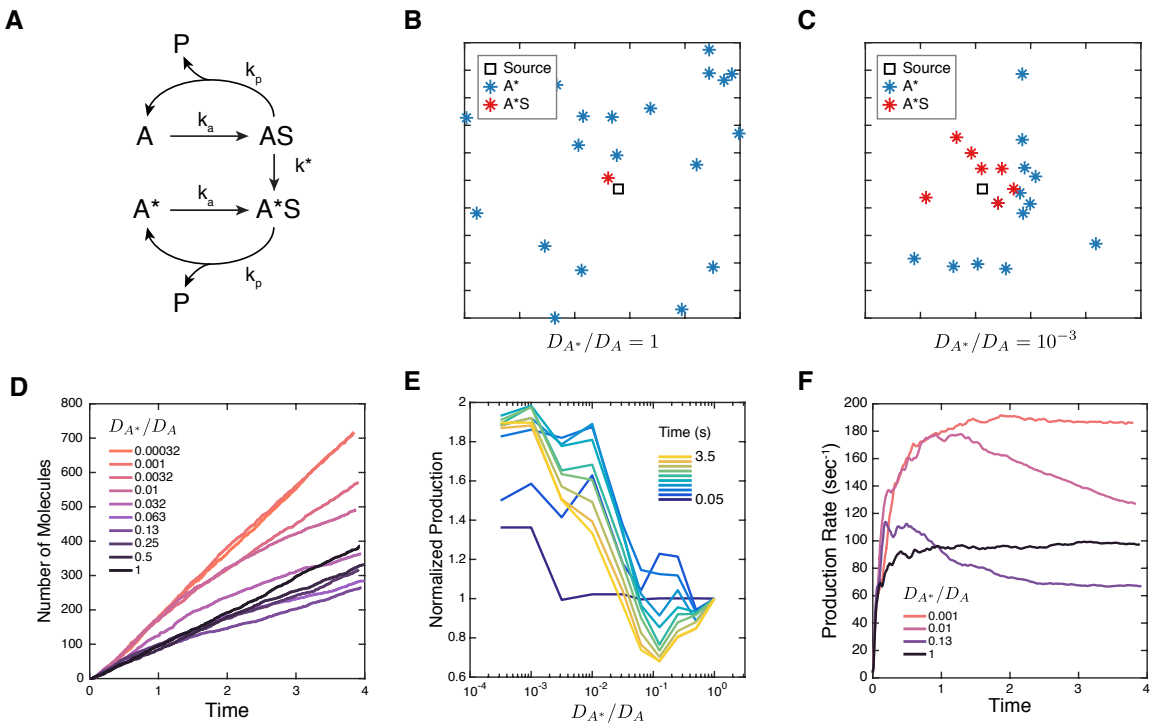


Figure 4.4: Changes to enzyme mobility alter effective production rates.

A. Diagram of reaction. Substrate S is produced by a relatively stationary point source. Enzyme A binds this substrate and converts it to a product P. Alternatively, when bound to the substrate, A can be converted to a form with lower mobility, A* with identical intrinsic catalytic ability to turn S to P. B. Snapshot of enzyme positions and occupancy after 8000 reaction time-steps with no reduction in mobility. The source of S is denoted by the black square. C. Equivalent to B but with a thousand-fold reduction in mobility. Enzymes are clustered near the substrate source and are more likely to be bound to S. D. Total number of P molecules produced as a function of time for a range of reductions in diffusivity. E. Equivalent data to D showing total production as a function of diffusion constant reduction at different times. For intermediate reductions in diffusion, production initially exceeded but then fell below the control case. F. Production rate for selected reductions in mobility spanning three orders of magnitude, computed by a sliding linear-

regression to find the slope of curves in D. For high reductions in mobility, production rate doubles the control case. For lesser reductions, production rate initially exceeds the control case but diminishes with time. Parameter values used in the simulation were: D_S = D_A= 1 μm^{2}s^{-1}, substrate production rate = 1000 s^{-1}, substrate decay rate = 300 s^{-1}, k_a = 10 μm^{2}s^{-1}, k* = 100 μm^{2}s^{-1}, k_p = 20 μm^{2}s^{-1}. The domain size was 1 μm by 1 μm discretized on a 200 by 200 grid.