AN OPTICAL FLOW BASED METHODOLOGY FOR VISUALIZING DYNAMIC SUCELLULAR ORGANIZTION DEMONSTRATED THROUGH PROFILIN AND RHOGTPASE MICRODOMAINS

APPROVED BY SUPERVISORY COMMITTEE

Konstantin Doubrovinski

Khuloud Jaqaman

Gaudenz Danuser

Satwick Rajaram

DEDICATION

I would like to thank the members of my Graduate Committee, Dr. Danuser and the rest of his highly supportive lab, my partner Charles Wilder, and my dogs Tybalt and Flo

QUANTIFIYING DYNAMIC SUCELLULAR ORGANIZTION IN 2D LIVE CELL FLUORESCENCE MOVIES

by

XUEXIA JIANG

DISSERTATION / THESIS

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY
The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

December, 2021

QUANTIFIYING DYNAMIC SUCELLULAR ORGANIZTION IN 2D LIVE CELL FLUORESCENCE MOVIES

Publication No. _____

Xuexia Jiang PhD

The University of Texas Southwestern Medical Center at Dallas, 2021

Supervising Professor: Gaudenz Danuser, PhD

Live cell imaging has enabled the collection of movies of subcellular protein dynamics at a submicron resolution. Statistical time series analysis can greatly expand our understanding of subcellular interactions in minimally perturbed systems. This was previously achieved for the leading edge of migrating cells in select cases. Importantly no strategy existed to simultaneously analyze every subcellular location. Building on existing optical flow based non-linear image registration we developed an approach to remap a migrating cell to a common cell footprint while preserving the characteristics of our signal of interest at a spatial granularity necessary for understanding micron scale biological interactions. This tool enabled us to discover that Profilin fluctuations are organized in living cells. This

organization was found to be dependent on cell polarization and actin binding capability. Expanding on this ability to query all subcellular locations, we developed a feature set and feature projection strategy to map molecular biosensor movies of Rho GTPase signaling into micron scale regions of internally consistent signaling dynamics or “microdomains”. Microdomains of GTPases match literature descriptions of signaling organization and in an optogenetic study were found to almost precisely match the perturbation footprint.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF APPENDICES

LIST OF DEFINITIONS

FRET – Förster resonance energy transfer is a physical process where proximity and specific orientation between light sensitive molecules enables the transfer of energy from an excited donor to an acceptor. In practice this is visible readout of molecular distance.

Profilin – actin-binding protein that functions as a molecular chaperone for monomeric actin and functions to accelerate in vivo actin polymerization through interacting with actin polymerizing factors. Effectively the in vivo building block of actin polymers.

Profilin R88E – mutant variant of profilin that is incapable of binding actin

Rac1 – Ras-related C3 botulinum toxin substrate 1, a small GTPase within the Ras superfamily canonically associated with lamellipodia and branched actin cytoskeleton.

RhoA – Ras homolog family member A, a small GTPase within the Ras superfamily canonically associated with stress fibers and myosin contractility.

CDC42 – Cell division control protein 42 homolog, a small GTPase within the Ras superfamily canonically associated with filopodia.

Optical flow – pattern of apparent motion in a visual scene caused by the relative motion between an observer and a scene

BBS – Blebbistatin, a cell permeable small molecule inhibitor of myosin contractility. In small doses it induces symmetry breaking or the transition of a cell from a static rounded shape to a polarized migratory shape.

CHAPTER ONE
Introduction

It was always understood that the mammalian cell is an organized environment wherein compartments facilitate each cell's unique roles. Historically this perspective on cell compartments was restricted to static structures, organelles, wherein a stable membrane or protein barrier separates chemical interactions. Recent work, however, has shown that even the cytoplasmic space is highly organized despite lacking permanent barriers and that this organization is inherently dynamic. A prime example the classic G protein signaling cascade. The earliest biochemical understanding perceived these cascades as nearly instantaneous reactions that propagate information from the extracellular space through a membrane receptor and a series of cytoplasmic proteins ultimately to potentially influence nuclear transcription. We now understand that the cellular microenvironment is a balance between activating and inactivating forces between all steps in the cascade allowing the cell to exert regulation on most steps and integrate a multitude of signals for eventual decision making. An oft ignored way the cell can exert this regulation is through changes in morphology. A reduction in local volume greatly increases the interaction likelihood between molecules due to a reduction in diffusion distance. Active motion can feed through dynamic cytoskeletal structures and Mohan et al. 2019 showed that this can relay a pro cancer proliferation signal from the cell periphery into nuclear transcription. Understanding cell behavior therefore requires us to stop considering the cell as a static reaction chamber and integrate the observable movements of the cell. This work provides a strategy to discover the organization of subcellular signaling using live cell movies.

CHAPTER TWO
Review of the Literature

Rho GTPases and Biochemical Basis of Spatiotemporal Organization

The Rho family of GTPases are small signaling proteins belonging to the Ras superfamily.[1] These small proteins (~21 kDa) transition from an GTP bound on state to a GDP bound off state under the control of Guanine Nucleotide Exchange Factors (GEFs), G-Protein Activating Proteins (GAPs), and Guanine Nucleotide Dissociation Inhibitors (GDIs).[2] There exists over 70 GEFs, over 20 GAPs, and 3 GDIs in mammals.[1] The C-terminus of Rho GTPases contains a CAAX box for attachment of membrane anchors.[3] Though the majority of Rho GTPases are cytosolic, the cytosolic pool is largely inactive due to their association with GDIs.[4] The large cytosolic pool facilitates rapid delivery of the second messengers to specific membrane sites both on the plasma membrane and on organelles. Once active the GTPases diffuse along the membrane manifold for rapid spread until they encounter inactivating GAPs. While biochemically capable of autoactivation and inactivation, the rates are entirely negligible in vivo.[3]

Alan Hall in the 1990s characterized the key members Rac1, RhoA, and Cdc42.[1] At the time there was little appreciation for spatiotemporal encoding of information in chemical second messengers. Through overexpression of constitutively active Rho GTPases, Hall assigned to Rac1 the role of directing lamellipodia, to RhoA the role forming stress fibers, and to Cdc42 the role of forming filopodia.[1] Even in these early studies it was clear that these descriptions did not capture the complexity of endogenous signaling. Hall's perturbed cells exhibited highly abnormal morphologies and his experiments showed crosstalk between

the 3 members which he described as a hierarchical relationship of Cdc42 -> Rac1 -> RhoA signaling.[1] However it was soon evident that such a simple hierarchy doesn't capture the complexity of the relationships between these molecules even in model systems.

The primary way the cell patterns the GTPase activities is through GEFs and GAPs. On the upstream level the GTPase protein structures and sequences have a high degree of homology leading to a high degree of promiscuity in the Rho GTPase regulators. For example, Vav1 is a GEF for both Rac1 and RhoA.[1] Despite this diversity in regulators, the regulator promiscuity and high degree of sequence homology between the Rho GTPases demands specific interactions to be controlled through protein localization and scaffolding factors. Recent papers have shown that cell morphology is a strong determinant of signaling dynamics within cells.[4] Increasing evidence points to membrane topologies as components in signaling networks for cell decision making. This can occur both through attenuation of diffusion constants to drive protein association kinetics or through direct recruitment of key factors in signaling pathways. GEFs contain Pleckstrin Homology (PH) and DOCK homology region 1 (DHR-1) domains which have specific affinity for certain phospholipids and allow targeting to specific membrane locations.[2] Some GEFs contain Bin-Amphiphysin-Rvs (BAR) domains which bind specific membrane curvatures (current count is 1 GEF and 7 GAPs with identified BAR domains).[2] Additionally, membrane topologies can recruit further upstream regulators such as integrin receptors or tyrosine kinase receptors.

Tyrosine kinase receptors serve as both scaffolds for Rho GTPases and their regulators as well as direct activators of GTPase-GEF interactions through protein phosphorylation. The regulators Tiam1, LARG, Vav1-3, Vsm, Dbs, RasGfr1, and Kalirin

have been found complexed with tyrosine kinase receptors.[3] Other receptors such as integrin receptors in focal adhesions also direct GTPase-GEF interactions.[4] Integrins recruit Focal Adhesion Kinase (FAK) which recruits and phosphorylates the GEF Beta-PIX to strengthen its interaction with Rac1.[4]

In addition to these fast time scale interactions (protein localization, GTP turnover, phosphorylation), the cell can also attenuate the local GTPase pool through proteasome degradation. For example, wild type RhoA can be polyubiquitinated by the E3 ligase Smurf1.[3] A MAPK pathway activity dependent degradation of RhoA has been recently identified through the E3 ligase FBXL19.[5]

Molecular Biosensors for Visualizing GTPase Activity

In the last 15 years, molecular biologists introduced biosensors of Rho GTPase capable of revealing spatially resolved activity of signaling in live cells.[2] Broadly these biosensors utilize translocation or molecular proximity to produce a fluorescence signal. For translocation biosensors a protein fragment from a GTPase downstream effector is conjugated to a fluorescent molecule.[2,6,8] The biosensor has a selective affinity for active GTPases and localizes the fluorescent molecule to sites of activity. FRET biosensors come in either monomolecular or biomolecular variants. The basis for both is molecular proximity. A FRET donor-acceptor pair (i.e. CFP and YFP) are conjugated to a GTPase or a GTPase effector binding domain.[6] The reporter GTPase is subject to the same activation and inactivation forces that influence endogenous GTPase activity. In the active form the reporter GTPase can bind the reporter effector domain and bring the FRET pairs into proximity

allowing fluorescence transfer from the donor channel into the acceptor channel. Bimolecular FRET reporters have greater influence on the behavior of the cell due to the ability of the biosensor GTPase molecule to activate endogenous effectors.[8] Monomolecular biosensors utilize a flexible linker to connect the GTPase molecule and the effector domain making the reporter binding domain the sole source of molecular interaction (endogenous effectors have a much lower likelihood of interacting with the reporter due to diffusion) at the cost of higher noise due to spurious alignment of FRET pairs.[6]

Preprocessing of Biosensor Movies for Reliable Time Series

Ideally the observed biosensor fluorescence intensities reflect the true endogenous signaling dynamics. Fluorescence light microscopy, however, produces artifacts that must be corrected for quantitative interpretations of observed signaling dynamics. The standard corrections to live cell light microscopy are flat field correction and photobleaching correction.[6,8] In brief the flat field correction comprises of a dark current and flat field image. The dark current image, a frame collected without incident light, accounts for the tendency of a camera to spontaneously record signal. The flat field image, a frame collected without the research sample in the light path, accounts for both gain and the potential for either illumination or light path induced local differences in intensity. The correction consists of subtracting the dark current from both the flat field and the research sampled followed by renormalizing the research image according to the local moving average of the dark current subtracted flat field. Photobleaching correction accounts for the fact that a certain percentage of fluorescent molecules within a research sample are destroyed by light exposure. A series

of movies with the research sample are collected and measured for the total illumination decay given time. This decay is then used to renormalize the movie frames.

In addition to microscope corrections and photobleaching, biosensors produce intensities relative to their local concentration. Bimolecular biosensors are particularly susceptible to this since both molecules can be in different concentrations. It is now typical to analyze a ratiometric image which is the intensity of biosensor activity normalized by concentration of the donor species.[6,8] For bimolecular FRET sensors this involves first normalizing the relative concentration of species as well as careful consideration of expression levels. For translocation sensors a diffuse cytoplasmic fluorescent protein of similar size can be used to measure concentration effects. While the observable intensity of the FRET donor decreases when energy is transferred to the FRET acceptor, this is a relatively small effect.

Through these corrections we are still left with an extremely high noise image. The lab's experience has been that ratiometric images from live cell movies at whole cell resolution have less immediately apparent organization than their original FRET acceptor channel counterpart. Any analysis strategies must be capable of handling such high noise and studies done prior to these corrections becoming standard should be interpreted with caution.

Anecdotal evidence of spatiotemporal regulation directing cell behavior

Largely due to the readily apparent heterogeneity in subcellular signaling and the corresponding difficulty in quantifying biosensor data, the majority of studies involving spatiotemporal regulation of Rho GTPases have been anecdotal. While many tools exist for

tackling frequency encoding of information and protein translocation, tackling both phenomena simultaneously is difficult. Thus, we have resorted to enforcing nonphysiologically relevant constraints to provide clear examples of how the Rho GTPases can be activated in a coordinated fashion to direct cell behavior.

The earliest of such studies involved wound healing in xenopus oocytes.[7] In a landmark study, the Parkhurst group demonstrated that unlike the classic dogma of Rac1 active in the front and RhoA active in the back of cells, the Rho GTPases coordinate their activity together to mediate cell behavior.[7] In the xenopus wound healing model, a laser punctures a hole in the egg damaging both the membrane and the supportive actin cortex. Membrane patching occurs rapidly through vesicle fusion, but true membrane integrity requires closure of the underlying, supportive actin cortex. This process requires the formation of a dense actin ring at the border of the damaged cortex and contraction of this ring through myosin activation. Through perturbation of RhoA using C3 isoenzyme, the group showed crosstalk between the Rho GTPases is necessary for establishing the wound healing spatiotemporal pattern as seen in a defect in Cdc42 localization.

The Klaus Hahn group published a breakthrough paper showing that GEF-H1 is responsible for localized activation of RhoA during cytokinesis.[10] While most past research in cell division focused on proper DNA segregation, there is an equally complex regulatory pathway driving the proper separation of cytosolic contents. The cell forms a cleavage furrow in between the segregated chromosomes and myosin activation leads to pinching of the membrane. GEF-H1 is localized to microtubules, released upon microtubule disassembly, and drives RhoA activation for directing myosin activity.[10] Using a RhoA

biosensor, the Hahn group showed that RhoA is specifically active around the cleavage furrow. GEF-H1 was localized to disassembling + end tips of kinetochores as seen with a phospho-antibody. This matched previous observations that GEF-H1 is released and activated upon microtubule disassembly. Disrupting the levels of GEF-H1 with siRNA led to failure of cytokinesis. This work together highlights the formation of a spatially confined, activity defined RhoA microdomain where the driving signal for reorganization of subcellular signaling is the segregation of chromosomes. Additionally, this work showed again that mechanical events and cell shape is intrinsically linked by diffusible signaling factors through specific, regulator driven protein interactions. In a follow up study that simultaneously imaged GEF-H1 activity, microtubules, and GTPase activity, the Hahn group showed that the mechanisms, GEF-H1 release from microtubule disassembly, underlying the cleavage furrow microdomain is a quantifiable component of GTPase regulation throughout the cell.[9]

The Hahn paper also showed that the RhoA signaling program for cytokinesis is mediated through delicate coordination of regulators. Previously, Ect2 was predicted to be the key GEF driving RhoA activation in cytokinesis. This prediction was based on Ect2 being a specific Rho GEF, being phosphorylated by Cdk1 (cyclin dependent kinases are key checkpoint factors determining progression of cell division), driving localization of RhoA at the cleavage furrow, and whose knockdown leads to disruption of cell division and multinucleated cells. However, upon knockdown of Ect2, elevated RhoA activity was still seen at the cleavage furrow in cells that underwent cytokinesis and failure of cytokinesis was determined to be driven by the initial failure to form and localize the cleavage furrow. This

latter observation was reinforced by the fact that catalytic domains of Ect2 were not able to rescue Ect2 depletion showing that Ect2's role in driving cytokinesis is not grossly dependent on RhoA activation but on early RhoA localization.

It was possible to extensively characterize the coordination of Rho GTPase signaling in the xenopus wound healing model and the cell division model because these events have strict spatial and temporal landmarks allowing direct comparison of different cells. The Parkhurst group matched the initial time point to laser ablation of the membrane and used the wound center as a spatial reference. The Hahn group matched the initial time point to the beginning of chromosome segregation and used the forming cleavage furrow as a spatial reference. For most biological processes, however, clear manually selectable prior location knowledge is not available.

The windowing method for subcellular time series analysis

Moving beyond the need for a temporal reference the Danuser lab used a novel windowing analysis process and cross correlation analysis to determine the coordination of Rac1, RhoA, and Cdc42 during cell protrusion.[8] Importantly the algorithm tackles the stationary phase of cell protrusion which is a continuous process of membrane oscillations in a protrusion and retraction cycle with a set average but non-rigid frequency. Under this method the cell edge serves as the "master contour" and the subcellular area is subdivided into "window" regions that are tracked across time to produce an edge-centric collection of time series.

The algorithm takes as input a movie of a cell segmented for cell space and background. The algorithm performs a distance transform D of the cell interior to the cell edge and a nearest neighbor feature transform F such that:

$$d_i = D(u_i, x_{1,2,\ldots n})$$

$$f_i = F(u_i, x_{1,2,\ldots n})$$

Where for any pixel u_i in the cell, f_i denotes the index of the closest pixel along the cell edge positions x and d_i is the distance between that cell edge position x and the u_i pixel position. These two measurements allow us to partition a subcellular region along a single depth axis. The algorithm utilizes an associated lateral distance along the cell boundary for each pixel l_i such that:

$$l_i = L(u_i) = \sum_{k=2}^{f_i} |x_k - x_{k-1}|$$

To calculate this the user arbitrarily specifies a coordinate of the starting edge position x_1. Taking these values the algorithm produces a sampling window $W_{m,p}$ where:

$$W_{m,p} = \{u_{1,2,\ldots I} \mid u_i \in \Omega \wedge b_m < d_i \leq d_{m+1} \wedge s_p < l_i \leq s_{p+1}\}$$

Here Ω is the segmented cell area, $b_{1,2,\ldots M}$ are the user-selected distances from the cell edge, and $s_{1,2,\ldots P}$ are the user-selected distances along the cell edge. Simply put this specifies that a window contains a set of pixels from the cell and for a given window that set is defined by a range of positions away from the cell edge along the "depth" axis to form a series of layers. For the lateral window boundaries along the "width" axis, the user predefines layer 1 along the cell edge using prior desired resolution knowledge. Each window in a deeper layer is

associated with a window in layer 1 and have lateral boundaries based on the associated distance l_i.

To propagate windows between time points typically the number of windows is fixed, the widths of windows are allowed to vary over time, and the depth of each window is held constant. The most reliable approach for propagating windows estimates the displacement vector of the cell edge and allows each window to contract or expand in width according to their associated edge changes over time. Importantly this method was the first systematic way of examining behaviors over time throughout the cell and was used to examine the relationships between Rac1, Cdc42, and RhoA.

It was previously thought that cell protrusion was driven by Rac1 activity for the formation of the lamellipodia with support by Cdc42 to form the filopodia. RhoA was believed to be active on the other side of the cell to produce contractile force and pull the tail of the cell towards the anchored protruding front. This dogma traces back to the original Hall papers and was established through overexpression analysis.[1] Using FRET biosensors for Rac1, RhoA, and Cdc42, this dogma was categorically disproved since activity of all three were seen in the cell's protruding leading edge.[8] This made sense also since the dynamic oscillation of the cell membrane should require both constant reorganization of the actin cytoskeleton, i.e. Rac1 and Cdc42 activity, and contraction of myosin, RhoA activity.

Using the windowing method to associate subcellular regions with the cell edge protrusion velocity, the lab was able to match the activity time series of each GTPase to the protrusion velocity time series with cross-correlation. To achieve spatial sensitivity, the windowing algorithm parcellated the membrane into half micron deep patches (the micron

deep scale was chosen since this was the expected diffusion distance of a small protein in 1s which was the imaging frequency) with width determined by the local microscale curvature of the membrane. The width determination made sense since if a patch of membrane protrudes as a unit it is likely to have the same regulation by Rho GTPases, and multiple patches of membrane in different phases of their protrusion and retraction cycles will create breaks in membrane curvature. The cross-correlation analysis revealed a tight feedback program which membrane activation from a previous protrusion event leads to RhoA activation for the next protrusion event which subsequently feedback activates Rac1 and Cdc42. This coupling not only occurs over time but also over the spatial zones near the leading edge. Most importantly this finding revealed that for the Rho GTPases and potentially for many of the signaling molecules within cells, it is the dynamics of the signaling that characterizes outcome and therefore we can no longer avoid time-series analysis.

Stochastic signal analysis for revealing subcellular relationships

The most powerful aspect of utilizing live cell movies as opposed to single time point snapshots is the potential to perform time-series analysis. By comparing the dynamic behavior of two samples over time we can infer relationships between them without the need of introducing perturbations that can fundamentally alter the behavior of interest. Importantly through multiple correlations we can infer causation.[12]

For this work and much of the literature cited the time-series were stochastic. For our purposes these stochastic series then fall into either stationary or nonstationary time-series.

Stationary series are ones that do not depend on time point of observation.[14] In practice this means that there is a constant mean and variance. With regards to biological activity this is true for processes that repeat but are not necessary in steady state or equilibrium i.e. protrusion retraction cycles. Importantly components of an observation (i.e. the high frequency band of the observed time-series) can be stationary while rest is not. As an example, the membrane in the lamellipodia oscillates in a stationary cycle with a slow outward drift for a growing structure. Stationary time series allow for reliable forecasting by describing the process through an autoregressive model (the standard regressive moving average ARMA models).[14] Nonstationary time-series analysis lacks forecasting power but can be used to compare time-series and describe their properties. A particularly powerful approach is Hilbert-Huang spectral decomposition which identifies instantaneous frequencies in the time-series and allows detection of specific events in a collection of processes.[11] This strategy revealed that while a cell's edge velocities are highly similar, the frequencies of those velocities allow the identification of distinct regions that correlate with distinct regulator activities.

Classification for time-series

Classification is a broad field whose goal is to take data samples and group them into subcategories whose members share some common property.[13] This typically assumes a very large dataset and the strategies fall under supervised, semi-supervised, and unsupervised classification.[13] All these methods require multiple measurements or "features" per observation sample. For supervised classification the goal is to assign observations to known

groupings. For unsupervised classification the goal is to discover unknown groupings in observed data. Semi-supervised is a rare approach where an algorithm is allowed to discover new groupings if they deviate significantly in measurable properties from known groupings.

Features for time-series include instantaneous frequency content, sample entropy, ARMA model coefficients, and forecasting results through various models of stochastic behavior.[15] The expectation is that these features either independently or more likely in combination capture the pertinent aspects of biological activity. If ground truth, manually labeled samples from past knowledge, is available then supervised classification algorithms can filter out non-discriminative aspects of the measurements. Unsupervised algorithms are more affected by poor measurements which create more similarity between samples. For the purposes of subcellular organization, known groupings are significantly rare that unsupervised classification is the most viable approach.

Ma et al. utilized statistical region merging to on instantaneous frequencies along both the time and space axis of membrane positions to identify regions.[11] Statistical region merging connections regions R_1 and R_2 if their similarity passes a user threshold defined by a delicacy parameter Q and maximum region size $|Rj|_{max}$. In practice it penalizes large regions and builds a hierarchical series of clustering results in which the lowest Q produces only two regions.

The Kwonmoo Lee group utilized a clustering approach to determine the relationship between edge protrusion behavior and the actin regulators arp2/3 and VASP in the edge adjacent subcellular regions.[15] Unlike edge velocities, it is possible for subcellular regions to be highly out of phase while still having the same fundamental characteristics. Therefore,

the Lee group compared the distances between samples using each sample's autocorrelation function. To cluster the resulting distance features they used a density based unsupervised clustering method which identifies high density peaks in the samples' feature space using a manually set threshold value for the minimum distance between a point and its cluster center. This algorithm was preferable since it works on high noise feature spaces where obvious gaps between data points don't exist. Effectively the user utilizes intuition to determine a cutoff for a local density that is sufficient to justify a meaningful cluster for follow up analysis and discards the rest. In this way the group extracted 4 highly internally homogenous clusters of actin regulator dynamics that largely differ based on their variance overtime.

CHAPTER THREE
Granularity Preserving Nonlinear Frame Registration for Live Cell Movies Reveals Spatiotemporal Organization of Profilin

Summary

We present an application of non-linear Image registration for the analysis of subcellular dynamics in 2D live cell movies that allows spatiotemporal analysis of extremely noisy molecular processes across the entire cell. To produce meaningful local tracking of protein dynamics in continuously deforming cells, we improved upon existing non-linear image registration by using a subcellular fiducial marker, a cell motion mask, and a topological regularization, which enforces diffeomorphism on the registration without significant loss of granularity. We demonstrate the potential of this approach in conjunction with stochastic time-series analysis through the discovery of distinct zones of coherent Profillin dynamics in symmetry-breaking U2OS cells. Further analysis of the resulting Profilin dynamics revealed strong relationships with nearby actin organization. This study thus provides a framework for extracting functional interactions between morphodynamics, protein distributions, and signaling in cells undergoing continuous shape changes.

Introduction

Time-series analysis of live cell movies can quantify functional interactions between proteins in the context of complex regulatory networks [1,2]. Utilization of these statistical tools requires registration of the locations where these interactions occur over time. For cell

biology this means following these locations through large morphological deformations such as protrusions, retractions, organelle translocations, and cytoskeletal rearrangements. Because of the difficulty in tracking subcellular locations, most analyses of live cell movies have remained incomplete and superficial, limiting the ability to test quantitative models of cell behavior.

Existing solutions to the extraction of consistent temporal information in one location can be broken down into three categories: manual sampling (i.e. kymographs), cell edge propagated sampling, and experimental approaches that limit cell deformation. Kymographs, while intuitive and easily used, are by nature incapable of simultaneously handling the entire cell and even locally tend to introduce significant artifacts to the time series as they do generally not follow cell deformation.[3] The introduction of edge propagated sampling has overcome some of these issues, in principle [1,4]. However, while registration of locations near the cell edge reveal expected interactions between proteins [1,2,5], time series at locations deeper inside the cell reveal a convolution of the real molecular dynamics with cell motion leading to uninterpretable data typically a mere few microns away from the cell edge. Experimental constraints such as micropatterning try to overcome this problem by fixing the cell footprint to a particular shape [6], but they do not necessarily limit the subcellular reorganization of molecular activities. More importantly these constraints often introduce harsh perturbations to the cell architecture, which obscure many of the natural cellular behaviors.

Medical image analysis, in particular functional brain studies, have applied non-linear image registration approaches to a similar problem of moving locations in order to extract

time-series over extended time periods.[7] Optical flow based algorithms, such as Thirion's demons, have a reputation for fast computation speed, intuitive approach, and ease of regularization for new problem constraints.[8,9,10] However, live cell imaging presents a much harder challenge due to high image noise, much larger geometric deformations, and near resolution limit of the structures of interest. Here we present a series of algorithm adaptations to Thirion's demons for the challenge of live cell imaging and demonstrate the ability to extract local time series across the entire footprint of an arbitrarily deforming cell. To demonstrate the potential of this pipeline we chose to analyze the dynamics of Profilin, which is a diffuse cytoplasmic protein and a component of the actin polymerization machinery, whose interactions with actin cannot be interpreted without registration of the live cell movies.

Results - Subcellular protein location referencing reveals spatiotemporal dynamics in live cell movies

To extract time-series from live cell movies we must first remap all frames to a common shape such that by calling any given x and y coordinate in the remapped movie we can pull out a relevant time-series. This is an ill-posed problem since we cannot directly observe all physical and chemical processes that govern the transport of a labeled protein (signal of interest) over time, including the changes in cell morphology itself. Medical brain imaging deals with fairly small shifts in the overall anatomy as well as with highly conserved organs with known landmarks to guide the remapping process. Prior work at the cellular scale that borrowed some of these ideas to track time series in live neutrophils focused on a

stereotypical set of behaviors, i.e. the formation of a immunological synapse attacking a model bead [11]. Neutrophils, whose morphology differed too much from the stereotype, were dismissed. Although in this study the authors compiled a great number of signaling time-series from morphologically diverse neutrophils, the analysis was focused on the narrow synapse region with almost rigid geometry. Moreover, it was limited to low spatial resolution, which softened the requirement to align time series with detailed variations of the synapse shape.

Nonlinear image registration comprises a large family of algorithms from simple point matching to algorithms based on optical flow and optimal transport theory.[8,12] We restricted our tests to algorithms that allow diffeomorphism constraints to enforce a 1 to 1 correspondence and neighborhood connectivity between time points. Our ideal algorithm would allow easy incorporation of known limits from diffusion of the signal of interest into morphological deformations. Optimal transport, used for the neutrophil study, defines correspondence so that overall the displacement between two time points is minimized [12]. This principle is well suited to big data problems for registering multiple cells and minimizing artifacts from interpolation operations. There is, however, no guarantee that subcellular motion between two timepoints occurs with minimal displacements. We therefore chose to modify the Therion's demons algorithm based on optical flow. Therion's demons defines correspondence using a gradient descent process using the difference between two images and the intensity gradient of the desired image shape to approximate the "diffusion" process that took place between the two images.[8] This process thus makes minimum assumptions about the data and performs well when a movie is sampled quickly relative to

the changes in morphology, a shared prerequisite for the intended goal of time-series analysis.

Our goal with this presented pipeline is to facilitate the high-content analysis of a common form of live cell data, 2D movies of single cells with fluorescently labeled proteins, in terms of identifying subcellular patterns of dynamic molecular activities. We conceptualize the cell as a dense space where the location of a molecular activity or concentration, referred to as the signal of interest, is coupled to the morphodynamics of the cell (Fig. 1a, second row). The morphodynamics is observed by a separate probe, referred to as the location fiducial (Fig. 1a, top row). Using this location fiducial, we apply nonlinear image registration techniques with the objective of matching all location fiducial images in the movie as closely as possible to a reference frame. The reference frame can be any image of the same location fiducial, including an image of another cell. Without losing the power of generalizing the pipeline, we will focus here on demonstrating the scenario where the middle frame of a movie is used as the reference frame. We then remap the corresponding signal of interest images using the deformations that produced closely matched fiducial images to compile a movie of the signal of interest with a fixed cell shape (Fig. 1a, third and fourth row). The resulting movie of the signal of interest exhibits now in a fixed cell shape, dynamics that were not coupled to changes in the location fiducial.

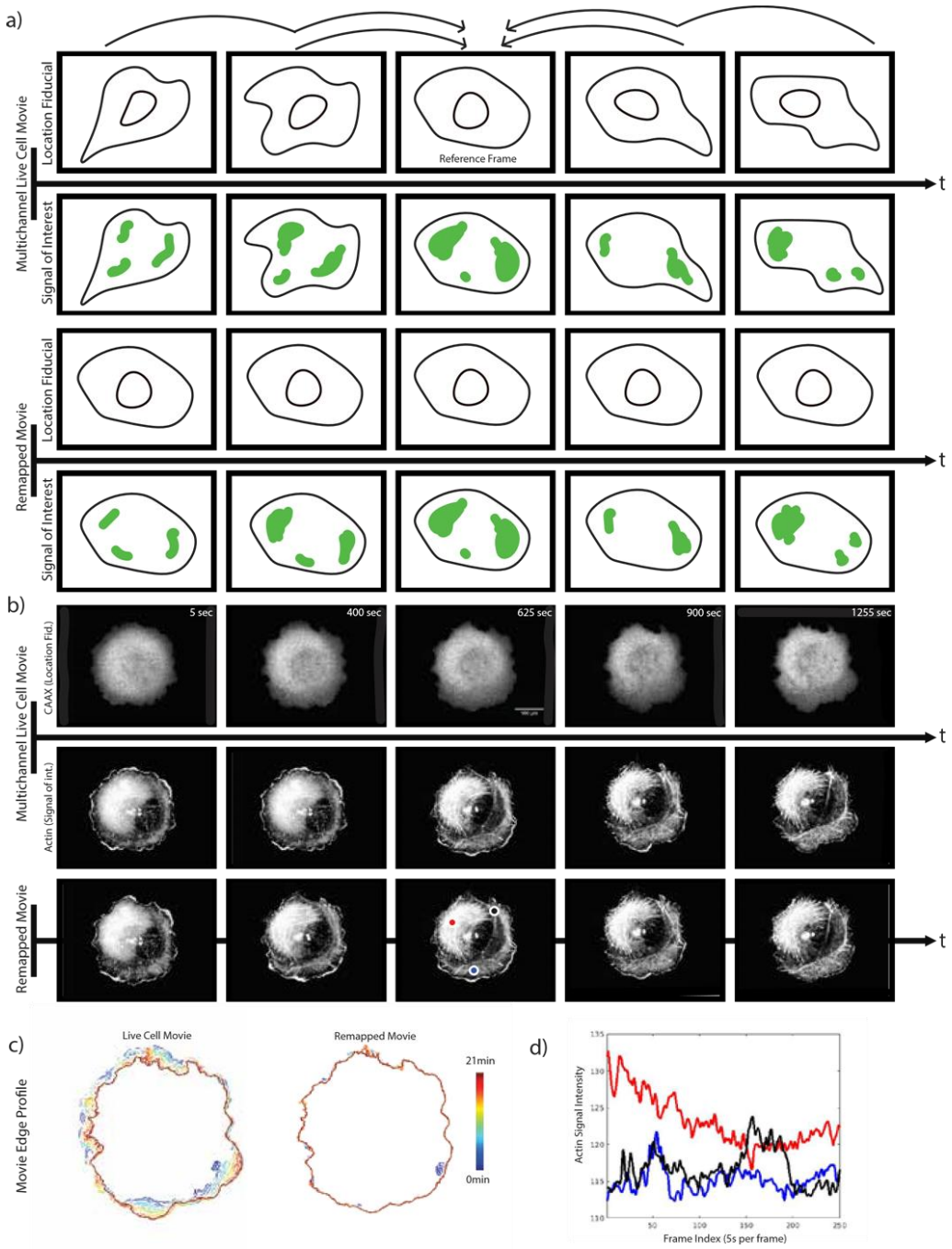

Figure 1) Fixing cell footprint through remapping movie frames allows extraction of time-series from all subcellular locations. a) our pipeline utilizes a location fiducial (top row) to remap all movie frames to the cell footprint in center reference frame. The signal of interest remains dynamic despite the remapping process. b) actin labeled movie example of the pipeline. Colored points indicate sampling positions in d. c) edge profile over time of the example movie in b. The remapping process with rare exception gives a stable cell edge. d) 3 example subcellular time-series sampled from actin movie. We extract a time series for every subcellular location.

The implemented pipeline rests on important assumptions about the data and signal of interest. There should be fast sampling of the movie such that the sampling rate is below the Nyquist limit of the biological behavior of interest. This also means that we expect small shape changes between timepoints but large shape changes can occur over the course of the movie. The cell should sit flat in 2D and should not move out of view for the duration of the analysis. While the cell can contact other cells, they should not move below or above other cells as in 2D this produces a region of high intensity in the fiducial channel that will be separated by the remapping process. Most importantly we assume that subcellular motion is faithfully represented by a location fiducial marker.

Fig. 1b introduces the pipeline on the example of a polarizing U2OS cells, i.e. over the course of approx. 20 minutes the cell changes from a rounded shape to a canonical migratory shape with a leading edge. We use Actin as the signal of interest and a CAAX membrane curvature marker as the location fiducial. The method eliminates shape deformation except for spurious artifacts in two peripheral regions of erratic ruffle formation where cell edge tracking fails (Fig 1c). After remapping, we can readily extract time-series of the signal of interest at any subcellular location (Fig 1d, position of two lamellipodia samples and a retraction fiber sample indicated in the reference frame of Fig. 1b). Of note, in this example the actin intensity is the signal of interest, and the remapping over time is accomplished based on a deformation field derived from a diffuse signal. Nonetheless, fine fibrous structure both in the front and in the back of the cell are preserved (Fig. 1b), indicating a precise estimation of the deformation field over time. Visually we can see for a stereotypical cell waviness in actin filaments indicative of the fact that, as expected, motion

of these large filaments is not identical to motion of a diffuse membrane marker. This breakdown in our assumptions and any imperfections in registration is a limit on the size scale of dynamics that we can quantitatively appreciate in our movies. The rest of this paper will attempt to quantitatively establish that our method produces time-series relevant for micron scale behaviors.

Algorithm basis for fitting deformation fields

The Thirion's demons algorithm framed image registration as a diffusion process whereby local forces, inspired by optical flow equations, pushed a moving image onto a target according to local characteristics of the images.[8,9] The algorithm takes as input a moving image (movie frames) and a target image (the reference frame) and outputs a deformation that through interpolation remaps the moving image to be as close to identical to the target image as possible. The algorithm alternates between calculating the forces based on intensity differences and spatially regularizing the forces with a Gaussian kernel. For a given coordinate, let **m** be the local intensity of the moving image and **f** be the local intensity of the target image. The local displacement **u** is given by eq1.

$$u = \frac{(m - f)\nabla f}{\alpha^2(m-f)^2 + |\nabla f|^2} \quad (1)$$

Here, $\nabla \mathbf{f}$ denotes the gradient of the target image intensity and $\boldsymbol{\alpha}$ is an optional throttling term usually set to limit the maximal displacement calculated at each iteration to 1 pixel. The matrix of these calculated displacements, **U**, is then smoothed by a Gaussian kernel $\mathbf{K_{diff}}$ and the entire process repeated **n** times until a registered image is achieved.

$$u \leftarrow K_{diff} * u \quad (2)$$

The solution for any given two images depends on the setting of $\boldsymbol{\alpha}$ and $\mathbf{K_{diff}}$, which determines the influence of small, local structures on the image matching. Since our goal was to approximate observable motion in cells, we made a series of cell biology specific adaptations to Therion's demons.

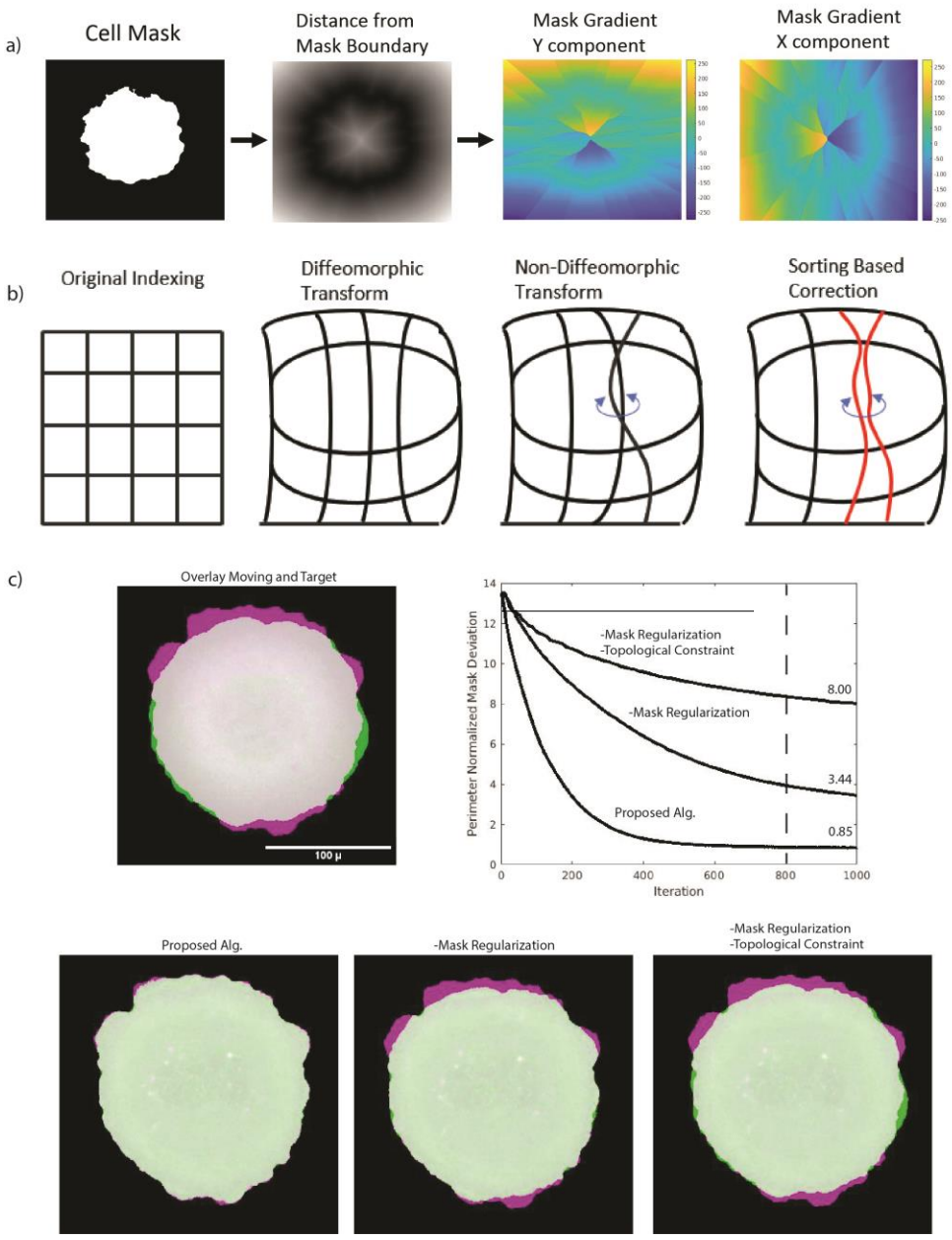

Figure 2) Basis for live cell movie specific remapping process. a) diagram demonstrating extraction of mask gradient (∇f_{mask}). From left to right (i-iii) we take an image mask and calculate the distance from the cell boundary for both the interior and exterior of the cell using a distance transform. We calculate the image gradient for this distance transform to get ∇f_{mask} in x and y for 2D images. b) From left to right (i-iv) diagrammatic illustrations of interpolation fields where edges indicate the sampling position of an input image to remap onto a target. i) no transform ii) a diffeomorphic transform iii) a non-diffeomorphic transform due to a folding in the space iv) correction via topological regularization. c) example overlays of an input cell image in (green) and a target cell image (purple) before and after remapping with and without the cell specific alg. components.

To determine the **n** iterations we picked a high motion movie in our dataset and determined the cell mask deviation over iterations (example Fig 2c) and selected **n** such that we get near pixel alignment. From prior knowledge we expect the observable motion at the cell edge to be far greater than motion in the cell interior.

To remap a sequence of frames ending at the final reference frame, we change the target image after **n** iterations to the next frame in the sequence until we fit the deformation to the reference frame after $\mathbf{n_{tot}}$ iterations (Fig. 1a). Farther time points with larger deformations receive more passes through the algorithm.

Biological data contains much higher noise and lower dynamic range than photography images for which the original framework has been developed. In the extreme case the cell exhibits a checkerboard like pattern (e.g. peripheral branched actin in Fig1b) and background that is close in intensity to the cell. This creates a scenario where the result is highly dependent on **α** and $\mathbf{K_{diff}}$ values where low values will cause the algorithm to get stuck in local minima and large values will cause the algorithm to not register small protrusions with known biological importance. To solve this, we introduced a mask regularization component where the mask is a segmentation of the cell separating the cell from background and other cells in the field of view (eq 3).

$$u = \frac{(m - f)\nabla f}{\alpha^2(m - f)^2 + |\nabla f|^2} + \frac{(m_{mask} - f_{mask})\nabla f_{mask}}{\alpha^2(m_{mask} - f_{mask})^2 + |\nabla f_{mask}|^2} \quad (3)$$

Here $\mathbf{m_{mask}}$ is the location of the moving image segmentation mask and $\mathbf{f_{mask}}$ is the location of the target image segmentation mask. We derive the mask gradient $\mathbf{\nabla f_{mask}}$ from the gradient of the map of shortest distances to the cell edge (Fig 2a). In practice mask regularization component is non-zero where the two image segmentations do not overlap and guides the

registration process to an overlay of the two cell images. Hence, the mask term moves the registration out of local minima, is independent of noise given a high-quality segmentation, and allows us to set $\mathbf{K_{diff}}$ based on the expected diffusion distance between time steps as opposed to an arbitrary number for better registrations. We then set $\boldsymbol{\alpha}$ to limit displacements to 1 pixel, but higher limits can be used to quickly test suitability of input data for this pipeline (not shown).

The original publication of Thirion's demons presented a way to enforce diffeomorphism based in Lie group theory.[8] The approach broke each displacement into a series of smaller steps based on the magnitude of the displacement and smoothed for each of these steps. This approach tends to be trapped in local minima, especially with a parameter selection meant to capture the granularity in cell data.[10] Instead, we chose to use a topological regularization through sorting. Given the 2D matrix of unsorted displacements $\mathbf{U}$, we compute $\mathbf{U_{sorted}}$ as a matrix such that:

$$U = \begin{bmatrix} u_{1,1} & \cdots & u_{1,n} \\ \vdots & \ddots & \vdots \\ u_{m,1} & \cdots & u_{m,n} \end{bmatrix} \tag{4}$$

$$\forall U_{*,n} \in U_{sorted} : U_{a,n} + a \geq U_{b,n} + b \quad \forall a > b$$

$$\forall U_{m,*} \in U_{sorted} : U_{m,c} + c \geq U_{m,d} + d \quad \forall c > d$$

To illustrate how this process works, a deformation field $\mathbf{U}$ is transformed into a matrix of indices for interpolation as per the image registration process. This can be represented in mesh format (Fig2b, panels i and ii) where square grids indicate an identity transform of the original image and non-square grids indicate either local dilation or contraction. A break in diffeomorphism occurs when indices are out of order leading to a crossing or fold in the

mesh representation (Fig2b, panel iii). Per iteration the sorting procedure corrects violations of the diffeomorphism (Fig2b, panel iv) by reordering the indices in the matrix. In practice, diffeomorphism breaks occur when either a difficult checkerboard pattern or noise continue to push optimal registration that requires crossings. These regions through our topological sorting will converge up to the step limit imposed by $\boldsymbol{\alpha}$ and $\mathbf{K_{diff}}$.

To illustrate the power of our approach we chose a U2OS cell undergoing a stereotypical isotropic spreading process and remapped 2 distant time points while setting $\mathbf{K_{diff}}$ to a non-diffusion related low value of 0.5. This condition deviates from our expectation of fast sampled changes but allows easy interpretation of the impact of our modification to the original Therion's demons and shows that our approach can handle more extreme morphology changes, which can occur in cell imaging. In combination, the motion mask and sorting based regularization permit near-pixel perfect edge alignment between two distant time points (Fig2c). Without these constraints, a typical remapping can exhibit >8 pixel per perimeter pixel deviation. This can correspond to losing an entire protrusion. Our adapted algorithm additionally requires fewer iterations to converge to an edge alignment, showing that computational efficiency is not traded for accuracy.

Evaluation of Algorithm by Image Signal Remapping

Our new approach requires an experimenter's selection of a subcellular location fiducial. Generally, this requires an additional fluorescence channel for live cell imaging besides the signal of interest. The use of a dedicated channel only for image alignment can be quite cumbersome and also increase photo-toxicity. We therefore tested the use of a down-

sampled signal of interest as a location fiducial. Since the goal of the image alignment is the extraction of informative time-series we can treat the signal of interest as a mixture of high spatial frequency signals describing local molecular activity and low spatial frequency signals describing the dynamics of subcellular molecular organization at a coarser scale. Assuming separability of the frequency bands and a known frequency cutoff, we can use the low frequency bands (down sampled images) for registration without artificially flattening the informative high frequency signals. In principle the combination of fast sampling, diffeomorphism constraint, and mask regularization limits the space of possible remapping differences. Nevertheless, the choice of location fiducial impacts the remapping results and we sought to quantify the impact of location fiducial choice.

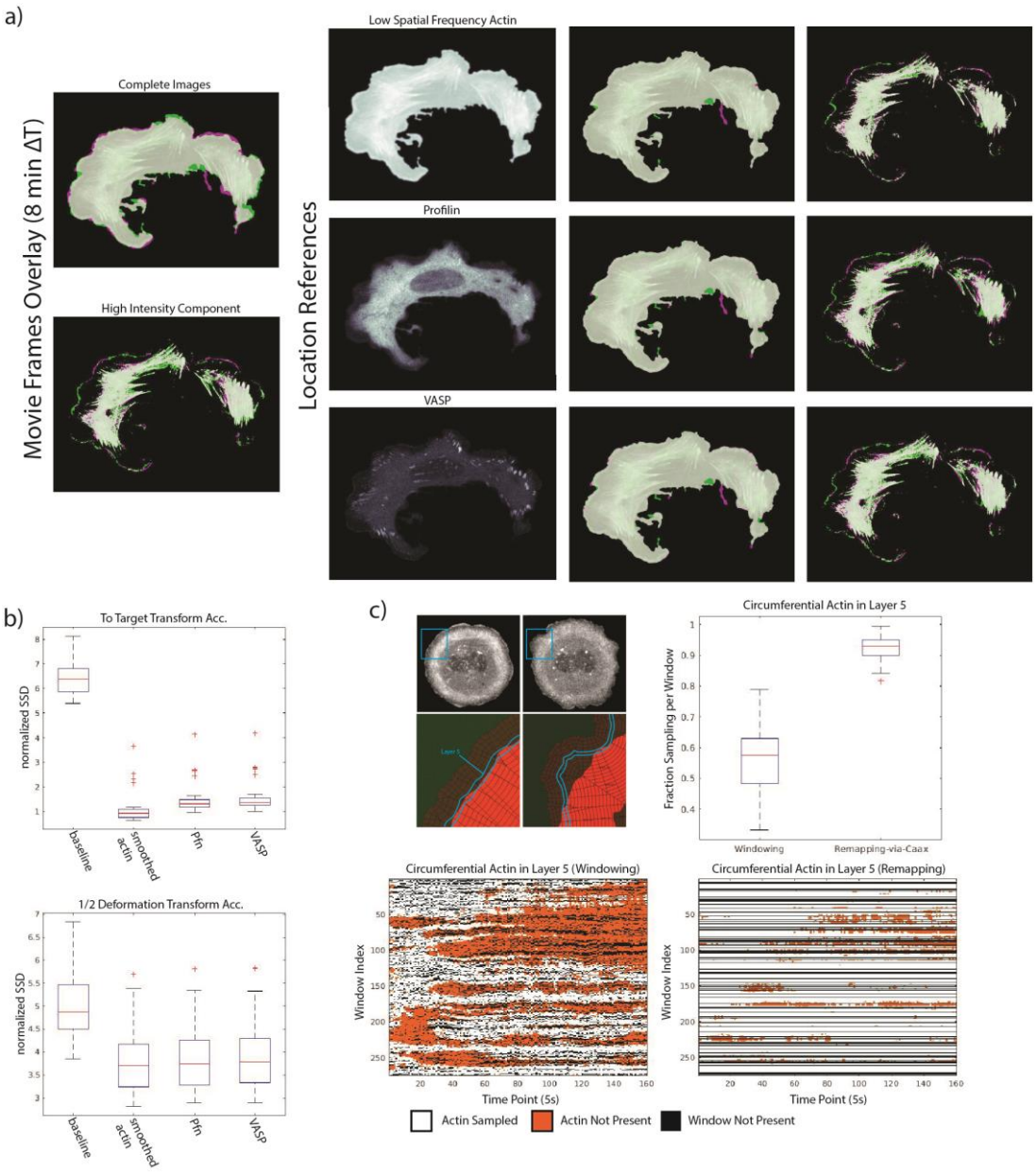

Figure 3) Impact of location fiducial on remapping accuracy. a) Movie of U2OS cell expressing labeled Actin, Profilin, and VASP. We show the overlay of the original beginning and end of a 20 min movie for both the segmented actin channel and a high intensity component highlighting the actin filaments. This was the beginning frame was remapped to the end frame footprint using 3 possible location fiducials (low freq Actin, VASP, Profilin). b) Subcellular remapping accuracy depending on the selection of location fiducial. Baseline indicates no remapping. Bottom plot shows the accuracy from our half deformation ground truth. We fit a deformation of an input frame to a target frame while skipping a middle frame. We then asked how well a half deformation remapping result matches the middle frame. c) Comparison of our pipeline in ability to capture transversal arcs (highlighted in red in cell images via threshold detection) vs published windowing approach. Bar plot shows how often any given window

Real world photography and medical image registration use landmark comparison and reconstruction of computationally distorted images to measure performance. Neither is a feasible metric for our live cell movies as we have no ground truth except for specific anecdotal examples and a simulated distortion can arbitrarily favor a particular fiducial choice. We therefore introduce the to-target transform and the half-distance transform accuracies as alternative performance measures for the remapping quality. Given a live cell movie of a particular molecules we remapped the signal of interest from moving image to target image using location fiducial deformation fields over two frames. For the to-target transform accuracy we asked how similar the remapped signal of interest is in comparison to the target. The proximity between remapped and target signal was determined by the average pixelwise squared intensity difference. This metric quantifies how much of the dynamics is captured by the location fiducial and our remapping process. For the half-distance transform accuracy, we divided the deformation field's magnitude in half and compared the thus remapped signal of interest to the skipped frame. This metric quantifies how well the approximations of a diffusion process underlying the remapping capture the real dynamics.

Fig. 3 presents these metrics for four scenarios compiled from a movie of a U2OS cell simultaneously expressing mNG-Actin, SNAP-Profilin and Halo-VASP. Profilin is a cytoplasmic binding partner of monomeric Actin and in concert with actin polymerases serves as a pacemaker for actin elongation.[14] VASP is a component of the focal adhesion complex and presents a punctate cytoplasmic signal.[13] We used Profilin, VASP, as well as down-sampled (Gaussian filtered with a sigma of 20 pixels) Actin as location fiducials to remap the raw Actin organization as the signal of interest. Like in Fig. 1, we chose again

Actin, because its dynamic and multi-factorially regulated structure, making it unlikely that any of the fiducial would fully capture the evolution over time. For a baseline we included the sum of squared intensity differences between unregistered original images over two frames. Both Profilin and VASP as location fiducials greatly outperformed the baseline deviation (Fig3b), despite the diffuse image character in the former and the relative scarce punctate pattern in the latter. Unsurprisingly the highest remapping accuracy was achieved with down sampled Actin as the location fiducial. This shows that for generating precisely stabilized images of cytoskeleton structure in a reference frame our algorithm works best using a low pass filtered copy of the original signal for alignment.

Quite unexpected, the punctate VASP pattern produced nearly as good deformation maps for signal remapping as continuously defined, diffuse Profilin distribution. Upon close inspection, we can see only small differences in the remapping of subcellular actin structures between the first and last frames of a 20 min movie.(Fig 3a) We explain this with contributions the faint and diffuse but still spatially transforming background pattern makes towards the estimation of the deformation field in between the salient puncta in the foreground. For the half-distance transform accuracy, we chose to skip only one frame since we were more likely to sample a steady state process during short time spans. Overall, the half-distance transform accuracy is worse than the to-target transform accuracy. This is mostly attributable to poor registration near the cell edge, which moves on a much faster timescale than the cell interior.

Evaluation of Algorithm for the Extraction of Subcellular Time-Series

To further establish confidence that the remapping algorithm permits extraction of meaningful time-series throughout the cell we compared the performance against our well-established cell peripheral windowing method.[1] In brief, the windowing method tracks the cell edge and divides the cell into volumes indexed by radial position and depth in layers. While the proposed new registration-based approach can handle a large variety of cell morphodynamics, the windowing strategy has largely been used against stationary membrane dynamics. We, therefore, chose a less motile U2OS cell (labeled with mNG-Actin, Halo-CAAX, SNAP-Profilin) undergoing stereotypical cell spreading (Fig 3c).

The cell periphery is delineated by distinct and persistent network of transversal actin arcs. We demarcated the arcs throughout the movie using a simple intensity threshold and hole-filling operation (Fig. 3d). We then applied the windowing algorithm to define layers of probing windows, each of which is 4 μm deep. In early time points the arcs begin ~16 μm away from the cell edge (sampled by layers 5 to 8). We chose to examine the sampling of the circumferential actin region by layer 5 in a windowed movie and in a movie where the actin signal of interest was remapped based on the CAAX membrane marker as location fiducial. In the remapped movie we applied the window positions from the reference frame to the entire movie as a stationary probing grid. Due to the fixed 1 to 1 window correspondence between layers, deeper layers exhibit drop-out events (Fig. 3d left heatmap black). Importantly, these drop-out windows do not persist over time since they are dependent on the geometry of the cell edge. In contrast, the window grid in a remapped movie has no drop-outs. The band between cell edge and transversal arcs, i.e. the lamellipodium, varies in

widths over time and also in space. Accordingly, in a sampling approach that preserves a constant distance from the edge, windows at the transition between lamellipodium and arcs alternate in the structure they sample. In contrast the proposed remapping approach accounts for the variation in subcellular structures. Indeed, while windows in layer 5 following the cell edge sample the transversal arc structure in only 58% of the windows and time points 92% of the stationary windows of layer 5 in the remapped movie sampled transversal arcs. The loss of connection to subcellular structures has been a serious limitation for many studies relying on edge tracking windows as they sample time series associated with distinct regulatory regimes.[15,16] The proposed remapping resolves this issue now for any subcellular structure that follows the diffeomorphism defined by the location fiducial.

Time-Series of Profilin reveal Spatiotemporal Organization of Dynamic Concentrations

Using windows sampled in a narrow band along the cell edge, previous work in our lab has shown that molecular signaling activities implicated in cytoskeleton regulation can be spatially partition into micro-domains of similar dynamics.[4] Equipped with a tool for sampling time series across an entire cell, we hypothesized that the same principle of spatial coherence in dynamic behavior could be applied to map out the subcellular organization of protein dynamics. Exploring this possibility, we examined the subcellular organization of Profilin.

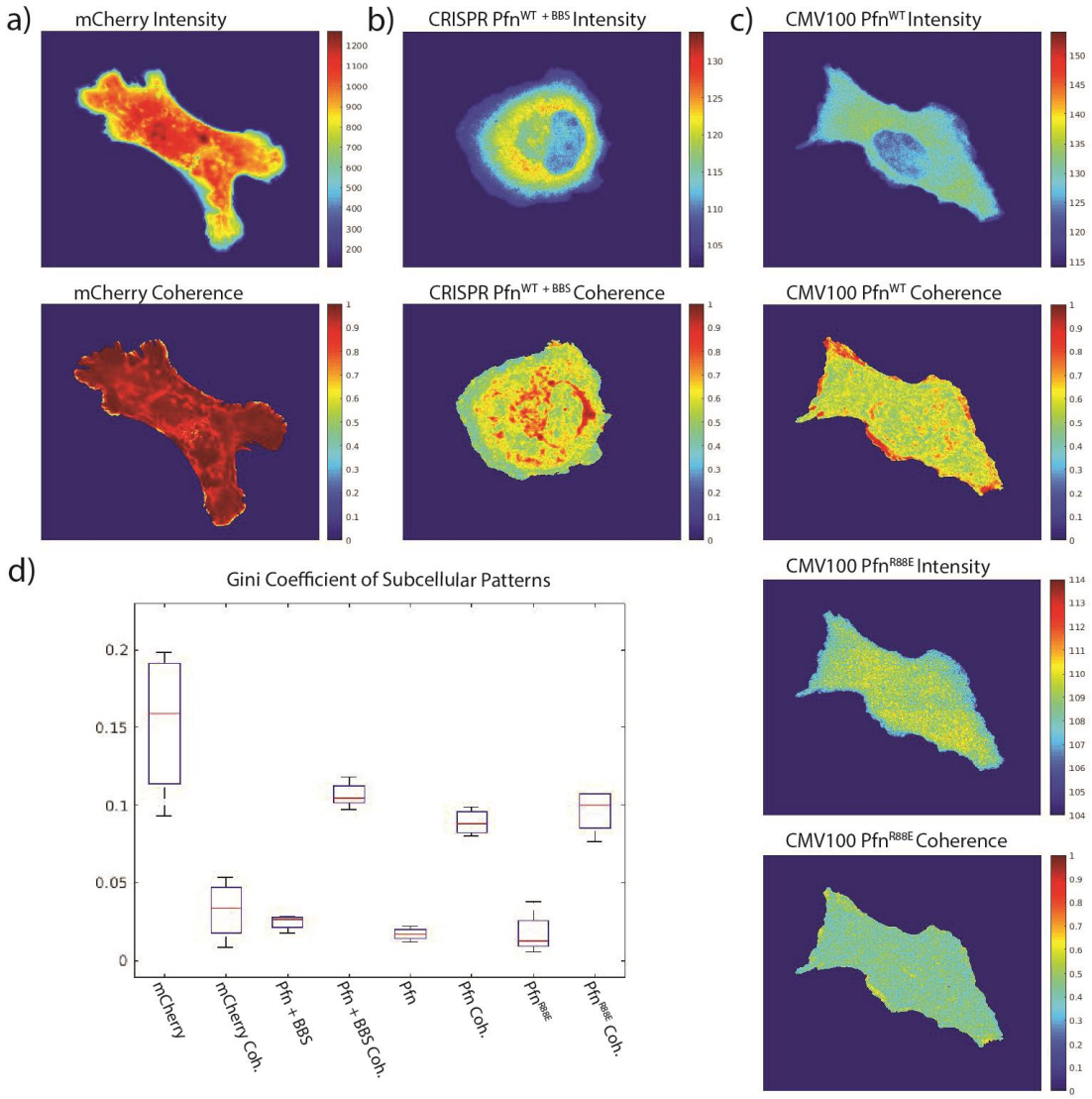

Figure 4) Profilin dynamics in live cells show patterns in local time series coherence a) Fluorescence intensity of cytoplasmic mCherry volume marker and local coherence of volume marker movie. CAAX curvature marker location fiducial. b) CRISPR SNAP-Profilin at endogenous levels and corresponding coherence. CAAX curvature marker location fiducial. c) Co-expressed WT and R88E Profilin corresponding coherence. d) Gini coefficients of observed intensity and coherence patterns. Reference time point used for intensity images. Results reveal heterogeneity increase in Profilin movies with spatiotemporal analysis despite near homogeneity in raw movies.

Profilin is a small molecule binding partner of monomeric actin and Profilin-actin is considered the physiological substrate of filament growth.[17] Numerous biochemical experiments have shown specific interactions between Profilin-actin and key cytoskeletal regulators (such as formins) suggesting that distinct actin structures would colocalize with local pools of Profilin of distinct dynamics. However, SNAP labeled via CRISPR endogenous Profilin in live cells displays a diffuse signal with no visually discernable pattern beyond intensity variations that related to the integration along the optical axis of fluorescence in variably thick cell into a 2D image (Fig 4a, top). To discover patterns of dynamics we measured in each pixel of the remapped cellular footprint the statistical coherence of the time-series within a 3x3 neighborhood (Fig. 4a, bottom). Coherence has a value between 0 and 1, where higher values indicate self-similarity among the 9 time series. This analysis resembles fluorescence correlation spectroscopy.[19] However, at the time scales of our movies, the coherence is unlikely not due to a physical property but to the cell actively maintaining such local similarity through either biochemical interactions or locally constrained morphodynamics.

We performed this analysis on 3 cell populations: First, U2OS cells expressing cytoplasmic mCherry as a volume marker in a background of mNG-Actin and Halo-CAAX. Second, U2OS cells expressing fluorescent SNAP tagged Profilin, mNG-Actin, and Halo-CAAX. These cells were further treated with the myosin-II inhibitor Blebbistatin to induce symmetry breaking throughout the movie, as described by Lomakin et al.[20] These cells allowed is to follow the changes in Profilin dynamics in response to a change in global cell morphology and allow us to demonstrate the analytical opportunities offered by the proposed

remapping. Third, we also exogenously co-expressed at very low levels EGFP- Profilin wildtype and mApple- R88E Profilin using a leaky CMV100 promoter in a SNAP-Actin background. The R88E mutation cannot bind actin.[21] We thus hypothesized that the dynamics of the mutation would follow a different coherence pattern than the wildtype. In the first 2 cases, we remapped the movies on a central reference frame using a CAAX curvature marker (images not shown) as a location fiducial. Due to expression limitations in expressing yet another tag for live cell imaging, we remapped the double Profilin labeled cells using a down-sampled actin channel set to mimic the results of CAAX based remapping.

This coherence analysis revealed drastically different organizational patterns relative to the raw Profilin signal. High coherence is observed in select sites around the cell edge, in puncta throughout the cytoplasm, and around the nucleus in both the endogenously- and exogenously- tagged Profilin. We quantified the increase in heterogeneity produced by application of the coherence operator via the Gini coefficient of statistical dispersion (Fig 4d). The Gini coefficient occupies a value range 0 to 1, where 1 indicates a signal concentrated in one pixel of the cell and 0 indicates homogeneous signal across all pixels of the cell.[22] For all three cell conditions, the coherence of the Profilin intensity is more heterogenous than the raw Profilin signal, indicating a high level of local dynamic organization of Profilin, potentially related to its roles in facilitating actin polymerization. As a control, we performed the same analysis for a cell expressing mCherry as a volume marker. In this case the intensity shows a higher level of heterogeneity than the coherence

value, because of significant variation in cell thickness. The coherence of the mCherry volume marker was homogeneously high throughout the cell.

Profilin Coherence is Related to Actin Dynamics

Since WT and R88E Profilin displayed distinct coherence patterns in the same cell, we hypothesized that WT Profilin coherence would show a stronger relationship with actin dynamics. Our coherence calculation thus far relied on time-series spanning the entire movie. To test whether Profilin coherence changes in concert with actin dynamics we computed a coherence time-series using a moving window of 20 frames, i.e. 1/10 of the length of our shortest movies. The resulting time series could then be locally correlated with a measure of the actin signal change. To match the time scales of changes in Profilin coherence and changes in actin dynamics we computed by the same moving window approach the entropy of the actin signal at every location of the cell. This signal transform extracts from the overall fairly static actin images (Fig. 5a) second-to-minute scale processes such as stress fiber growth and movements and retrograde flow (Fig. 5b).

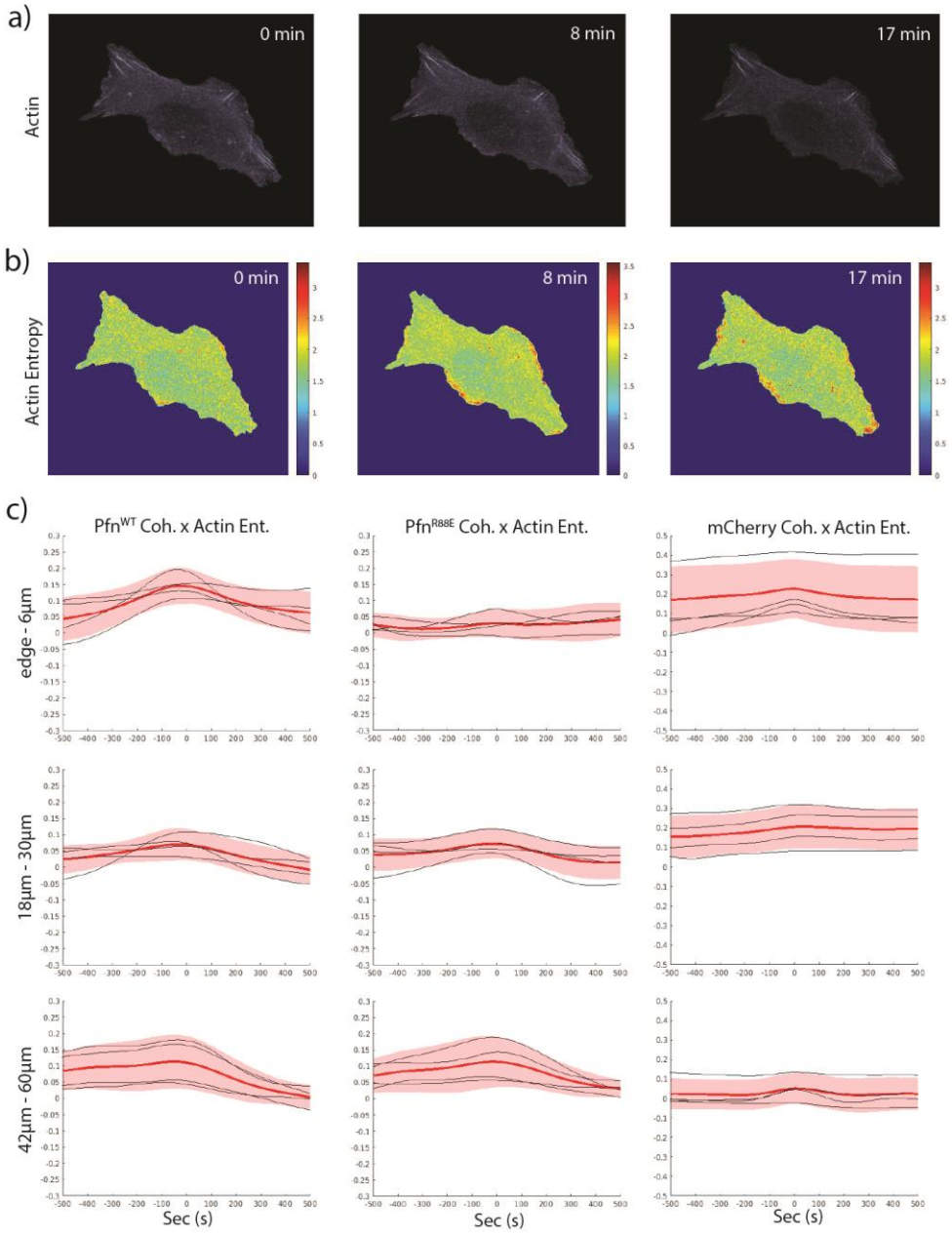

Figure 5) Co-expressed wildtype and mutant Profilin exhibit different relationship to Actin dynamics. a) example of actin behavior in 20 min movie b) from the actin dynamics in a we calculate a 20 frame moving window actin entropy along the time axis to highlight regions of high change. c) cross correlation of 20 frame moving window Profilin coherences to Actin entropy in distance from edge selected bands. We see sig differences between R88E and mut Profilin near the cell edge.

We then examined the cross correlation of Profilin coherence and actin entropy near the cell edge (edge – 6μm), around the circumferential actin network (18 – 30 μm), and around the nucleus (42 – 60 μm) since the data in Fig. 4 indicated patterns of high Profilin coherence in these positions (Fig. 5c). As expected, near the cell edge we observed higher cross correlation between actin entropy and WT Profilin when compared with R88E Profilin. The difference is less substantial in deeper regions of the cell because both actin entropy and Profilin coherence are temporally less salient. This is reinforced by the fact that the volume marker's cross correlation with actin entropy is high in the periphery but low near the nucleus where there is a persistently high volume marker coherence but rare high actin entropy events. Together, these analyses indicate how the proposed remapping of signals to a static reference cell geometry permits the application of fairly involved time series processing in order to extract subtle but significant parameters of the local interactions between molecular processes.

Profilin Coherence Correlation with Actin Dynamics is Responsive to Perturbation

To test the hypothesis that the observed relationship between Profilin coherence and actin entropy relates to Profilin's functions as modulator of actin polymerization, we analyzed cells undergoing symmetry breaking. Symmetry breaking is a process where the cell transitions from a stable rounded state to a polarized migratory state. In doing so the cell must greatly reorganize the actin network. After remapping the entire movie to a common reference frame just before the symmetry breaking event (Fig. 6a), we split the movies into before and after symmetry breaking section to compute Actin entropy (Fig. 6b) and Profilin

coherence (Fig. 6c). Comparing the spatial distribution of Actin entropy before and after induction of symmetry breaking by Blebbistatin, the high entropy zones move from the cell periphery to the front of the cell (Fig. 6b). This is consistent with the notion that the high actin turnover in frustrated lamellipodia, ruffles and transversal arcs all around the edge of an unpolarized cell is shifted during symmetry breaking towards the wide lamellipodia and lamella regions at the new cell front.[20] This reorganization of actin dynamics is paralleled by a reorganization of high Profilin coherence (Fig. 6c). Specifically, before symmetry breaking, we see high Profilin coherence right along the cell edge, on top of transversal arcs, and around the nucleus. We see background levels of coherence near the rear retraction fibers. There exists a faint but noticeable low coherence band between the leading edge and transversal arcs. Again, locally high Profilin coherence indicates high spatial coupling in the concentration fluctuations of profilin-actin complexes that are fed into actin polymers by nucleators such as formin, WASP and VASP, which promote the growth of linear actin structures.[17] Importantly we do not expect these zones to be an artifact of the remapping process as their size significantly exceeds the approximate deformation of the movie frames. The facts that R88E and WT Profilin exhibit clearly different dynamics at the cell edge (Fig5c) and the more stable cell rear does not exhibit higher coherence leads us to conclude that we observed an Actin regulator interaction driven effect. Hence, enabled by the proposed remapping algorithm for the spatial stabilization of cell footprints during dynamic processes, this analysis visualizes for the first time directly in a living cell Profilin's function and organization as a critical facilitator of actin assembly.

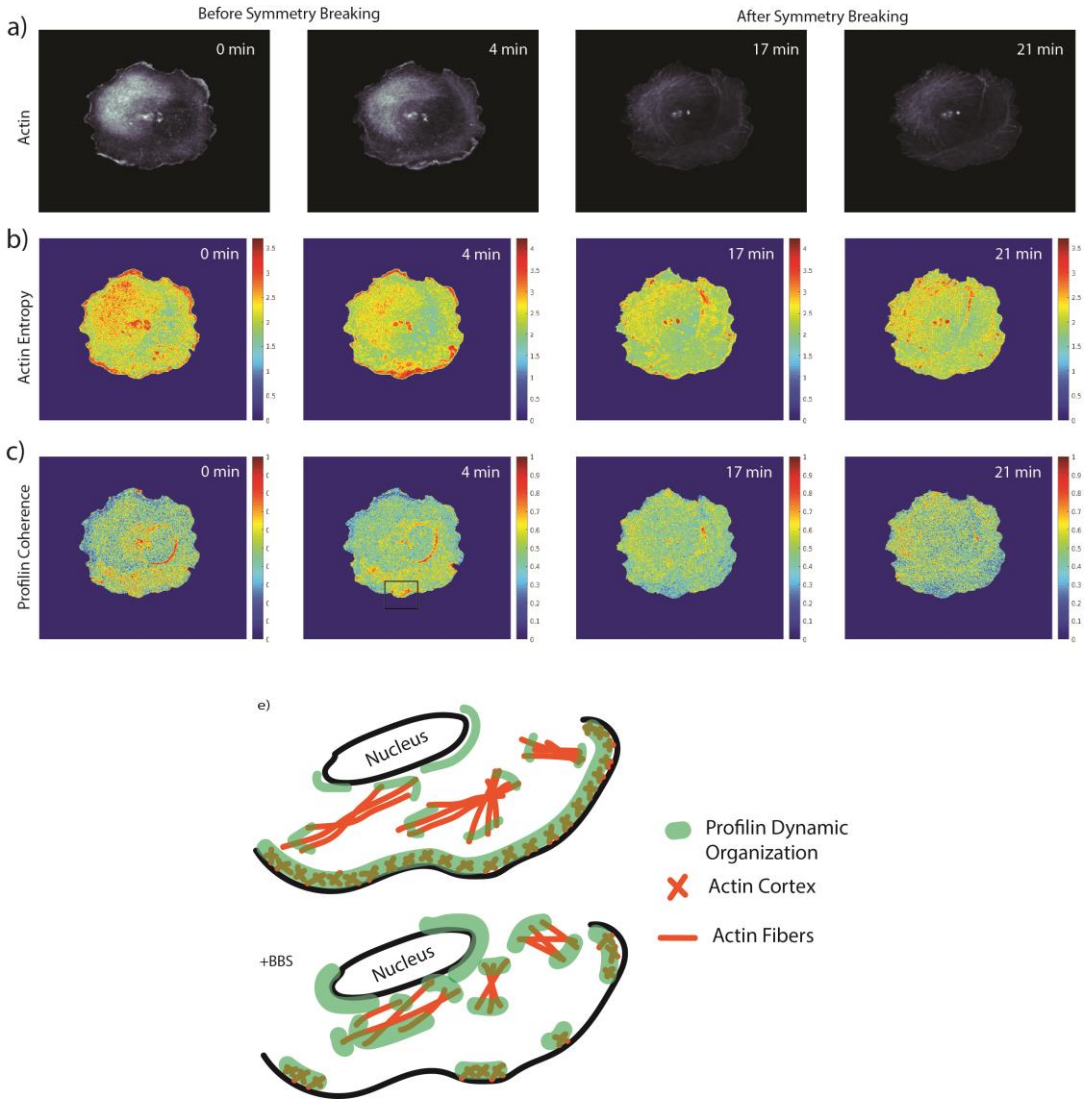

Figure 6) Blebbistatin induced symmetry breaking reveals polarization dependent organization of profilin coherence. a) example actin behavior before and after symmetry breaking b) actin entropy in 20 frame moving windows before and after symmetry breaking c) profilin coherence in 20 frame moving windows before and after symmetry breaking d) cartoon representation of profilin organization following bbs induced symmetry breaking

Discussion

In this work we implemented a non-linear image registration framework in order to analyze subcellular protein dynamics in cells undergoing substantial morphological variation throughout the observation window. The key contribution of our work to the sizable literature on non-linear registration algorithms is the capacity to handle the high noise and small structures of interest present in cell microscopy. This was accomplished by introduction of a cellular motion mask as a regularization term to the objective function governing the image mapping and by enforcing diffeomorphism in the map via a sorting of the displacement field. The former causes the map estimator to bypass local minima, the latter permits the preservation of fine grained structures during the mapping.

We demonstrated the capacity of this framework by extracting biological insight through observational studies of previously uninterpretable signals. Importantly, the algorithm supports the extraction if reliable time-series at every subcellular position and thus enables studies of the spatiotemporal organization of molecular processes. Such analyses were hitherto limited to a narrow rim along the cell boundary, where time series could be sampled by a deformable window grid following the edge motion. We illustrated these new features by examining the relationship between Profilin and Actin dynamics. Profilin lends a highly diffuse image signal with visually uninterpretable variation. Actin lends a mixture of highly structured and amorphous image signal components. As a cell undergoes morphological changes, the structured components often display complex patterns of deformation whereas the amorphous components undergo often difficult to follow flows. Because of the image registrations, we were able to transform both the Profilin and Actin

signals into secondary signals that readily revealed the spatiotemporal coupling Profilin and Actin dynamics, even during a cellular symmetry breaking event that produces large scale cell shape changes.

While designing this framework we expected the resulting time series to strongly depend on the choice of the location fiducial used for the map estimation. We were surprised that a punctate signal (VASP) permits the algorithm to remap Actin structures over time with an accuracy that is comparable to the mapping accuracy supported by a location fiducial derived from a diffuse signal like Profilin or a blurred version of the Actin signal itself. This is likely because adhesion proteins like VASP have a faint but implicitly tracked diffuse component that constrains the map estimator in the same way as Profilin and Actin signals do. Moreover, cytoskeleton structures are highly coordinated in healthy cells and the cytoplasm is a dense compartment meaning that subcellular molecular flows in general are coupled. Hence, there is some degree of tolerance in choosing a location fiducial.

Most importantly our framework opens the door to analyses of subcellular signals with dynamics that occur on the same or slower timescale as cell morphological changes. For processes much faster than cell morphological changes (i.e. cell electrical potentials and calcium signaling), microscopy has produced predictive quantitative models for in vivo signaling behavior and outcomes because cellular changes could be ignored. However, the majority of subcellular activities likely occur on a timescale that matches cell morphodynamics. While existing approaches have generated anecdotal and often visually-guided analyses of these processes in select cell regions, our framework now supports an

unbiased analysis across the entire cell. This is the starting point for robust pattern recognition in cell biological activity.

Methodology

Cell Lines

Cell lines 4 30 Human Osteosarcoma U2OS cells were a kind gift from Dr. Dick McIntosh (University of Colorado Boulder) and were maintained in DMEM supplemented with 10% fetal bovine serum 32 (Sigma; F0926-500ML) in a humidified incubator at 37 °C and 5% CO2. All cells were tested for 33 mycoplasma using a PCR-based Genlantis Mycoscope Detection Kit (MY01100). Cells were not authenticated.

Plasmids

pSpCas9n(BB)-2A-Puro (PX462), pSpCas9(BB)-2A-GFP (PX458) and pLKO.1 were from Drs. Feng Zhang and David Root, respectively (Addgene plasmids #48141, #48138 and #10878). Gene-targeting single guide RNAs (sgRNAs) were designed using CRISPor [26]. The self-cleaving donor vector pMA-tial1 was a kind gift from Dr. Tilmann Bürckstümmer [24].

mNeonGreen, SNAP, Halo, human β-actin, mouse VASP, mouse Profilin, mCherry, CAAX, DNA fragments were PCR amplified with corresponding flanking homology regions to seamlessly clone into pLVXCMV100 using HiFi Assembly to generate mNeonGreen-18-actin, SNAP-21-actin, Halo-21-tagged VASP,where 18 and 21 denotes the number of amino

acids in the linker between the tag and the tagged gene. These linkers and the terminal ends used for tagging were chosen based on our previous work.[15] The 18-linker sequence was (SGLRSGSGGGSASGGSGS) and the 21-linker sequence was (EPTTEDLYFQSDNAIAGRPRSSG). The S235 and T273 residues (corresponding to the S239/T278 residues on human VASP) in the Halo-tagged VASP were PCR mutated to aspartic acid and glutamic acid, respectively. mNeonGreen is licensed by Biotechnology & Pharmaceuticals, Inc.

Introduction of Fluorescent Probes and Considerations to Minimally Perturbing the Cells

Lentiviral constructs harboring a previously described truncated CMV promoter containing only the first 100 bps of the promoter (CMV100) (24) were used to achieve low expression of the exogenous genes. In addition, all the tagged genes contained a variable amino acid linker varying from 18-21 amino acids to minimize interference due to tagging.

Cell Culture

Cells were counted using Cellometer Auto 1000 Bright Field Cell Counter (Nexcelom). Lentiviral particles were generated using the packaging vectors psPAX2 and pMD2.G (Addgene plasmids #12260 and #12259). Infected cells were bulk sorted using FACS, or selected with Puromycin (1 µg/ml; Gibco).

Live-Cell Time-lapse Imaging of Cells

Cells were seeded on fibronectin (10 µg/ml)-coated #1.5 glass-bottom dishes and allowed to spread overnight. The following day, cells were labeled with HaloTag ligand conjugated with JF549 (0.4 – 1 µM; Promega) and/or SNAP ligand conjugated with SiR-647 (0.25 – 1 µM; NEB) for 30 minutes, and subsequently washed twice with DMEM. Prior to imaging, the media was replaced with phenol-red free DMEM supplemented with 20 mM HEPES pH7.4. Time lapse image sequences were acquired on a climate-controlled (maintained at 37°C), fully motorized Nikon Ti-Eclipse inverted microscope with Perfect Focus System, equipped with a 60×, 1.49 NA APO TIRF objective (Nikon) with an additional 1.8× tube lens (yielding a final magnification of 108×; Andor Technology), and an Andor Diskovery illuminator coupled to a Yokogawa CSU-X1 confocal spinning disk head with 100 nm pinholes. Image sequences were recorded using a scientific CMOS camera with 6.5-µm pixel size (pco.edge) at a 0.2 or 0.1 Hz frame rate.

Blebbistatin treatment to induce symmetry breaking

Cells were treated with 25 µM myosin II inhibitor blebbistatin in DMEM for 5 min prior to imaging as per Lomakin, 2015. Drug treatment was offset to allow for a complete 20 min imaging run per dish under the expectation that symmetry breaking would occur at approx. the 10 min mark. Fields were selected manually to center a single individual cell and cells were selected for a flat rounded morphology. After imaging movies were examined for spontaneous symmetry breaking near the midpoint of the movie. We utilized the first 1/3 of

the movie as before symmetry breaking samples and the last 1/3 of the movies as post symmetry breaking samples.

Image Segmentation and Preprocessing

Prior to analysis we rejected movies where the desired cell was in contact with neighbors, where the cell moved outside the field of view, and where there existed abnormal illumination due to debris or unforeseen conditions. We down sampled all movies to ½ resolution for ease of computation. Each frame of live cell movies was segmented to separate foreground and background using a gaussian filtered (σ = 2 pixels) actin channel and a single manually selected intensity cutoff for the entire movie.

Remapping Pipeline Parameters

For all published results we set $\boldsymbol{\alpha}$ = 1 for a maximum step size of 1 pixel. For sequentially transformed movies for time series analysis we set $\mathbf{n}$ = 100 iterations per time step and $\mathbf{K_{diff}}$ = 1.5 pixels where each pixel corresponds to approx. 600 nm in our down sampled images. For transforms between the first and last image frames we set $\mathbf{n}$ = 2000. For ½ deformation accuracy we set $\mathbf{n}$ = 200.

Profilin Coherence Analysis

For every subcellular location we sample its immediate neighbors for a 3x3 matrix of 9 time series. We calculated a pairwise correlation coefficient for 27 coefficients between non-

identical time-series where the correlation coefficient $\rho(A, B)$ between time series **A** and **B** is:

$$\rho(A,B) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{A_i - \mu_A}{\sigma_A}\right)\left(\frac{B_i - \mu_B}{\sigma_B}\right)$$

μ_A and σ_A are the mean and standard deviation of **A**, respectively, μ_B and σ_B are the mean and standard deviation of **B**, and **N** is the number of time points in **A** and **B**. The coherence is the mean of resulting coefficients. For edge positions we removed the out of the cell time series from the initial 3x3 matrix. We calculated coherence over the course of the entire movie and in 20 frame moving windows (1/10 length of typical movie) to calculate coherence change over time.

Actin Entropy Calculation

We calculated the Shannon information entropy for subcellular actin intensity in 20 frame moving windows (1/10 length of typical movie). For every actin intensity series **X** the entropy of said series **H(X)** is:

$$H(X) = \sum_{i=1}^{N} P(x_i)\log_2 P(x_i)$$

Where x_i is a value in series **X** and $P(x_i)$ is the probability of drawing x_i from series **X**.

CHAPTER FOUR
Time Series Features and Classification for Rho GTPase Dynamic Microdomains

Summary

Molecular biosensors for the Rho GTPases have enabled the direct second by second measurement of subcellular signaling in live migrating cells at sub-micron resolution. Observation of these biosensors has led to the conclusion that GTPases are organized within the cell in regions that the field terms microdomains to enable local control of the actin cytoskeleton in cell migration. Our lab's previous work has linked hyperactive GTPase activity to a particular morphodynamic process in melanoma cells that causes proliferation under growth signal (MAPK) inhibition. This leads to the inference that the patterning of microdomains is intricately connected to both normal and pathological cell behavior. Past work has enabled the automated detection and mapping of subcellular regions using machine learning based strategies. However, these strategies were unable to fully map the cell due to an inability to simultaneously extract time series from every subcellular location at a relevant resolution. Our lab recently developed an image analysis technique that overcomes this key limitation and now we present a novel strategy to derive a map of subcellular GTPase microdomains using extremely noisy 2D biosensor movies. We achieve this through a feature set dedicated to cataloguing events in nonstationary subcellular time series and more importantly a sample comparison strategy that combines feature distance and the known subcellular location manifold from which the features were derived. Together this allows us

to apply a commonly used unsupervised classification strategy that produces contiguous regions in the cell with internally consistent dynamics. We validate this technique using optogenetic perturbation and show that perturbation reorganizes the cell's microdomains in a light-targeting dependent manner. This technique thus serves as the basis to interrogate the organization of signaling within live cells and relate that organization to disease.

Introduction

The organization of Rho GTPases into subcellular zones of differential activity has long been appreciated; Rac1, one of the 3 canonical Rho GTPase families (Rac1, RhoA, and CDC42) maintains high activity at the leading edge of a migrating cell while low Rac1 activity exists at the trailing edge.[1] Previously our lab showed within the leading edge that Rac1 activity is temporally regulated and oscillatory with activity peaks occurring approx. 30 seconds after the initiation of membrane protrusion.[1] Both simulations of signaling cascades and optogenetic studies, which allow instantaneous, spatially resolved perturbations of subcellular signaling, have shown that changes in the topology of the regulatory network leads to changes in the frequencies of Rac1 activity.[2,3] GTPases are therefore spatiotemporally organized within the cell with the most apparent organizing principle being to pattern the actin cytoskeleton for cell migration.

Previous attempts at mapping this organization have been largely anecdotal. The Klaus Hahn group described a zone of RhoA activity corresponding to the contracting cleavage furrow during cell division.[4] The Pertz group identified zones of differential Rac1, RhoA, and CDC42 activation for stereotypical spreading cells under PDGF

stimulation.[5] In an earlier study the Hahn group identified localized zones of CDC42 activation at filopodia and the trans-golgi apparatus.[6] In a previous study we elucidated a link between cell mobility and cancer drug resistance dependent on a Rac1 dynamic in melanoma.[7] Hyperactivation of Rac1 via the P29S mutation in melanoma hijacks branched actin network assembly upon growth challenge via MAPK inhibition to massively upregulate lamellipodia formation. Through biochemical analysis we determined that the extended lamellipodial branched actin network facilitates melanoma proliferation through inactivation of tumor suppressor NF2/Merlin and therefore inferred the presence of a pathological Rac1 microdomain as the driver of drug resistance. Not only do these past approaches to GTPase coordination bear limitations in terms of objectivity and reproducibility, but these analyses are incapable of appreciating coordination of subcellular signaling.

We more recently developed an image analysis approach that takes a 2D movie of a migrating cell and fixes the cell footprint while still preserving the dynamics of a subcellular signal of interest. This technique is based on the principles of optical flow and utilizes a subcellular location fiducial to relate image pixel scale subcellular locations from one frame to the next. We demonstrated that this approach could support a location complete statistical analysis of the cell to discover that organization of Profilin fluctuations is coupled to actin dynamics. We now extend on this approach to facilitate an unsupervised machine learning process to fully map the subcellular heterogeneity of GTPase signaling into microdomains. As an additional benefit we provide the field with a strict definition of GTPase microdomains as micron scale subcellular regions with internally consistent activity dynamics .

Results) FRET biosensor donor-based referencing reveals time series of signaling activity

We utilized monomolecular FRET GTPase biosensors to collect 2D live cell movies of stereotypical cell behavior. Among biosensors the monomolecular biosensors have low signal to noise ratio (SNR) but are considerably easier to use due to a molecular balance of the donor and acceptor species and generally a reduced impact on normal cell signaling from an overexpressed biosensor (Fig1b).[5] The biosensors produce FRET donor and acceptor images which we transform into a ratiometric image of cell signaling to normalize for subcellular concentration variation. Our approach to remapping live cell movies requires the selection of a location fiducial. For monomolecular FRET biosensors this becomes a trivial choice as the donor is physically linked and therefore localized by the same forces as the readout.

We performed the remapping at full image resolution with n = 100 and K_{diff} = 1.5 with the center frame as the target footprint. As an example, we show a Rac1 movie selected for its high shape change over the course of the movie (Fig1a top row). The remapping process produces a movie with a fixed cell shape (Fig1b bottom row). We see that despite the constrained cell footprint, subcellular dynamics remain highly variable. There are some appreciable artifacts when a small region protrudes overtime since the image registration must expand a few pixels. However, these regions are warped according to the observable subcellular structure in the location fiducial. From our observations and past literature we present a diagram of expected Rac1 organization within the cell (Fig1c).

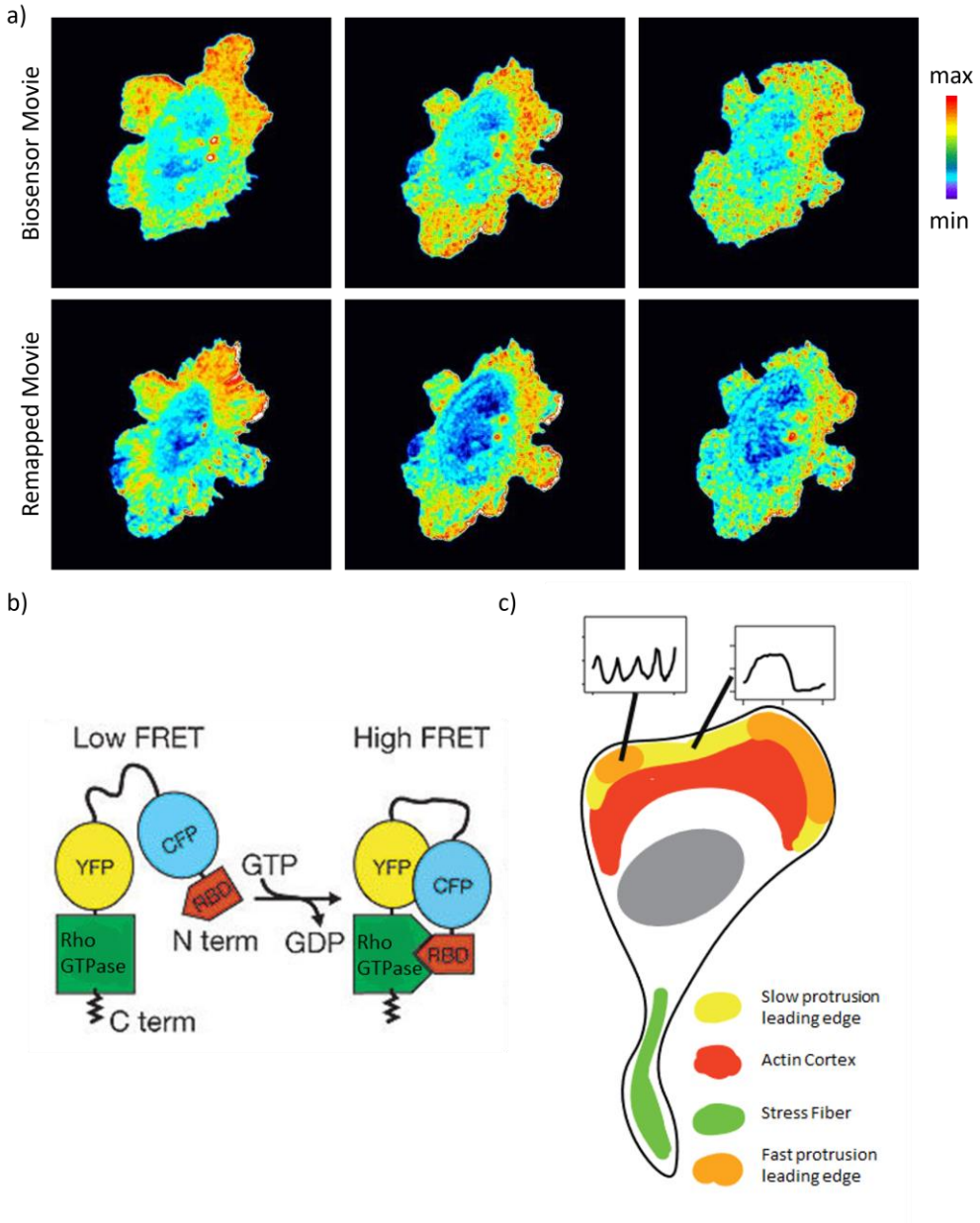

Figure 1) Remapping with FRET donor reveals subcellular activity dynamics

a) Top row: example ratiometric Rac1 readout over 20 minutes at 10 s/frame at beginning middle and end frames. Bottom row: remapped movie using CFP donor as location fiducial
b) Diagram of monomolecular biosensor construction. Upon GTP binding the GTPase binds the Rho binding domain (RBD) from a downstream effector to bring the sensor components into proximity
c) Diagram of expected patterning of Rac1 microdomains

Results) Expected size scale of Rac1 microdomains

Since the remapping process does introduce artifacts in the resulting remapped data space, we wanted to estimate the expected size scale of GTPase microdomains in an unbiased manner. To do this we measured spatial autocorrelation in movies provided by the Pertz group from their 2016 scientific reports paper.[5,6] The REF52 cells were chosen for their flat shape and therefore reduced volume effects in 2D imaging and imaged within a microfluidics chamber that introduced and washed-out platelet derived growth factor (PDGF) at prechosen time points. Unlike temporal autocorrelation along a single axis, spatial autocorrelation requires a specified spatial weighting to handle 2 dimensions. For this we chose a circular kernel and calculated the spatial autocorrelation for Rac1 movies before, during, and after treatment by PDGF for every frame in the Rac1 movies. As expected PDGF treatment, which wipes out contractility, leads to a highly spread cell shape and therefore a larger area before spatial autocorrelation drops to 0.(Fig 2) From these measurements we estimate microdomains of around 10 um in diameter for typical migrating cells.

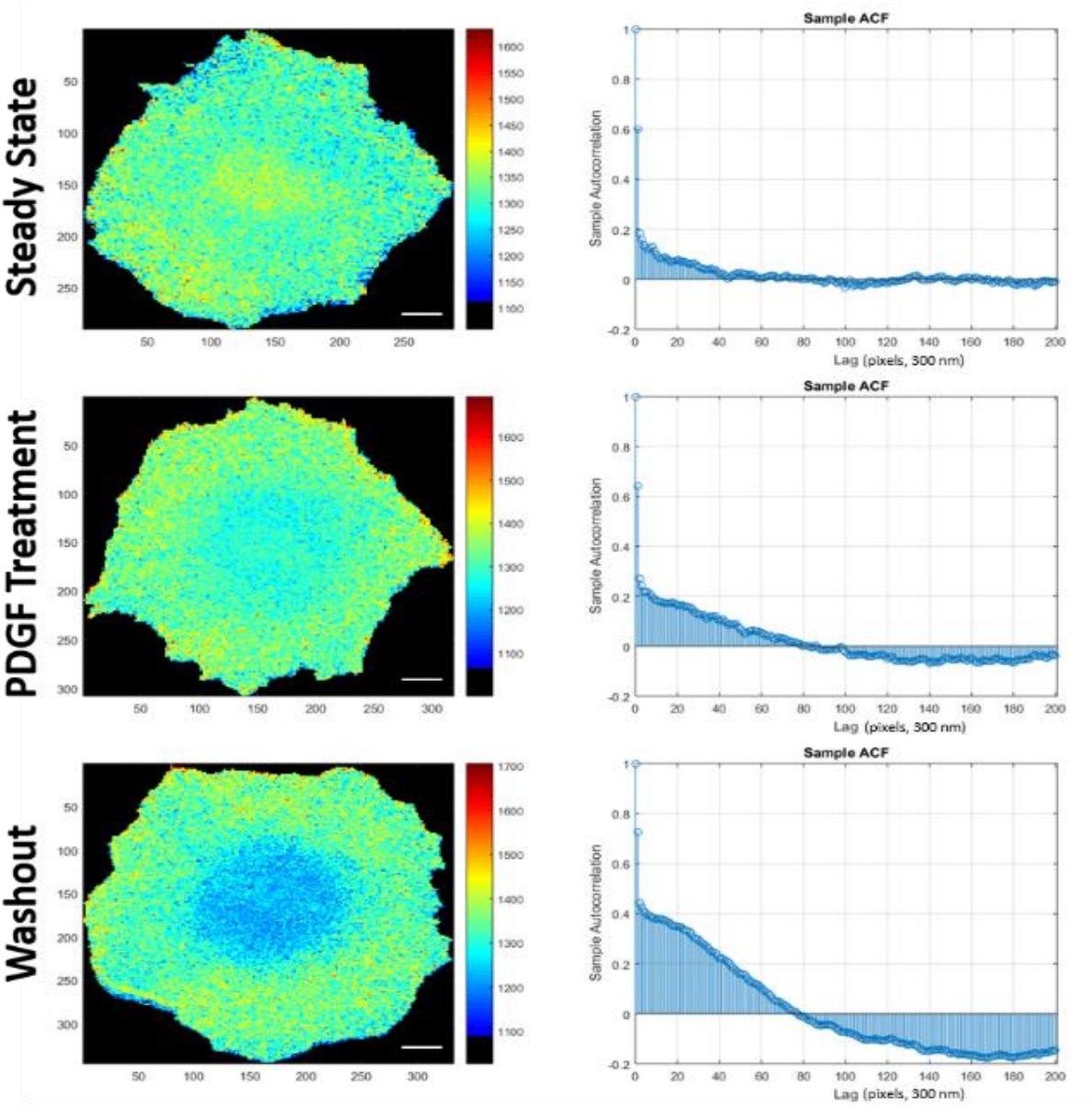

Figure 2) Example spatial autocorrelation analysis. Spatial autocorrelation was performed with sliding windows over columns of data ignoring sets that contain background. All scale bars: 10 um

Results) HCTSA feature library does not produce distinct data clusters

Unlike subcellular protein concentrations, signaling molecules can exist in on and off states which reports as an extra degree of freedom. For time series analysis this means that subcellular locations can exist out of phase while still being derived from the same biochemical process. By diffusion this is not true for protein concentrations since an increase in local volume would soon lead to increases in adjacent volumes. Because of this we don't expect simple approaches like cross correlation to be sufficient in finding microdomains.

The Jones group previously published the "highly comparative time-series analysis" HCTSA feature library which contains over 7000 features specifically for time series analysis.[8] Importantly this feature library does not assume that the analyzed time series are stationary, a requirement for our lab's previous clustering of microtubule time series using ARMA coefficients.[9,10] We first attempted to use this feature library to find cdc42 microdomains. Of the features calculated around 1000 exhibited no variation at all. A TSNE embedding (perplexity 30, exaggeration 4) revealed a highly dense space with minimal appreciable structure.(Fig 3) However, when we mapped certain dense contiguous regions in the projected feature space to the original cell landscape, we found that many produced contiguous regions. This was true for both Rac1 (Fig 3a) and CDC42 (Fig 3b). Importantly we see an expected depth dependence of these regions as well as differences between the protrusive lamella and the quiescent cell rear in polarized cells (CDC42 example Fig 3b).

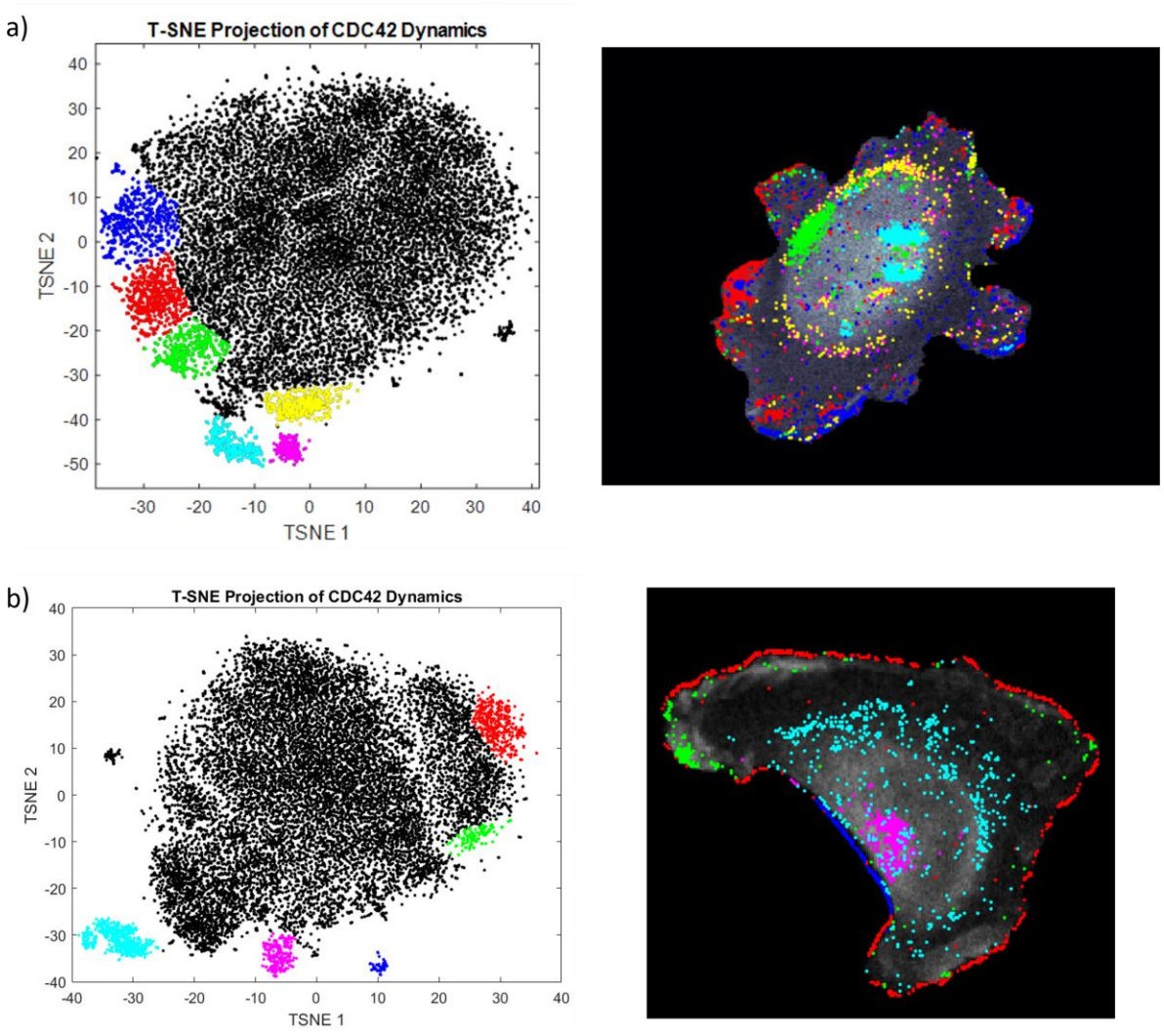

Figure 3) Potential dynamically defined microdomains using HCTSA feature library. a) Rac1 example b) CDC42 example. On the left we show a TSNE projection to 2 dimensions (perplexity 30, exaggeration 4). In colors we highlighted select regions and protected them back to their original cell locations on the right

Results) Basis for instantaneous frequency content features

A significant drawback to the HCTSA feature library was its high number of uninformative features as it was largely designed for supervised machine learning.[8] Supervised machine learning methods are capable of ignoring components of feature libraries that do not contribute to classification accuracy whereas unsupervised machine learning depends on obvious separation in high dimensional feature space.[12] Since we were unable to find a large enough collection of anecdotal microdomain examples, we sought to define a feature set more suited for subcellular signaling analysis.

Prior work from the Kwonmoo Lee group utilized instantaneous frequencies from the Hilbert-Huang spectral decomposition to identify classes of Arp2/3 and VASP dynamics.[13] Our lab also previously used Hilbert-Huang spectral decomposition to classify sectors of distinct membrane protrusion-retraction dynamics.[14] Both these studies however were limited to a few microns around the cell edge. Nevertheless we hypothesized due to the success of these past studies that comparison of instantaneous frequencies can provide a feature set suitable for GTPase dynamics while allowing for explainable classification results.

To review, the Hilbert-Huang transform (HHT) takes an input signal and outputs and time vs instantaneous frequency vs energy spectrum.[14,15] Well-behaved Hilbert transforms rely on an empirical mode decomposition (EMD), which divides the submitted time series into a finite set of component signals, referred to as intrinsic mode functions (IMFs). IMFs are signals whose number of local extrema and the number of zero-crossings either is equal to each other or at most differs by one and whose mean value of the upper

envelope defined by the local maxima and the lower envelope defined by the local minima is equal to zero. Simply put they are purely oscillatory signals with constant mean. Together the IMFs forms a complete and nearly orthogonal basis for the original signal and the Hilbert Transform is guaranteed to converge to an unbiased estimate of the instantaneous frequency spectrum of the IMF.

To generate IMFs, EMD uses the sifting procedure consisting of: i) Identifying all local extrema in the time series **X(t)**. ii) Connecting all local maxima by a cubic spline to generate an upper envelope and connecting all local minima by a cubic spline to generate a lower envelope. iii) Computing the mean $\mathbf{m_1(t)}$ of upper and lower envelopes and subtracting it from the target time series to generate a reduced series $\mathbf{h_1(t)}$.

$$X(t) - m_1(t) = h_1(t)$$

Should $\mathbf{h_1(t)}$ satisfy the 2 IMF conditions it is defined as the first IMF component $\mathbf{c_1(t)}$ else it the sifting process repeats **k** times starting with $\mathbf{h_1(t)}$ instead of **X(t)** till the IMF criteria are satisfied.

$$h_{1(1)}(t) - m_{1(2)}(t) = h_{1(2)}(t)$$

$$\vdots$$

$$h_{1(k-1)}(t) - m_{1(k)}(t) = h_{1(k)}(t)$$

Once sifting identifies the first IMF component $\mathbf{c_1(t)}$. we calculate the residual $\mathbf{r_1(t)}$ defined as:

$$r_1(t) = X(t) - c_1(t)$$

The process starts again with $\mathbf{r_1(t)}$ replacing **X(t)** to calculate the next IMF $\mathbf{c_2(t)}$ and repeats until the sifting procedure hits a tolerance stop criterion. The tolerance **s(n)** at the **n**th iteration is defined as:

$$s(n) \triangleq \frac{\|r_{n-1}(t) - r_n(t)\|_2^2}{\|r_{n-1}(t)\|_2^2}$$

If **s(n)** is less than a set threshold value of 0.2 the process stops with the **n**th IMF component giving us a series of IMFs $\mathbf{c_1(t)}$….$\mathbf{c_n(t)}$. Together the components and residual can reconstruct the original signal satisfying the criteria that:

$$X(t) = \left(\sum_{i=1}^{n} c_i(t)\right) + r_n$$

For each IMF component we perform the Hilbert transform to produce an instantaneous frequency at each time point **t**:

$$H[c_i(t)] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{c_i(\tau)}{t-\tau} d\tau$$

$$F(t) = \frac{1}{\pi} \cdot \frac{d}{dt}\left(\arctan\left(\frac{H[c_i(t)]}{c_i(t)}\right)\right)$$

$$A(t) = \sqrt{c_i^2(t) + H^2[c_i(t)]}$$

where **i = 1, …, n**. The instantaneous frequency spectrum **F(t)** is the temporal derivative of the phase change in the IMF signal $\mathbf{c_i(t)}$, which is defined by the inverse tangent function of the quotient between the Hilbert Transform of the original signal $\mathbf{H[c_i(t)]}$ and the original signal $\mathbf{c_i(t)}$. The corresponding instantaneous amplitude spectrum **A(t) is** the root of the square sum of the original signal $\mathbf{c_i(t)}$ and its Hilbert Transform.

From the Hilbert-Huang spectrum decomposition of each subcellular time series we derive a size varying set of time-frequency-energy values. Additionally, while the IMFs are ordered in decreasing frequency bands, the same indexed IMF across multiple source time series can vary in their frequency selectivity. Our prior work addressed this by fixing the number of IMFs for a given movie and rejecting data whose IMF frequency content differed too greatly. Instead, we conceptualize each time-frequency-energy triplet as an event signature akin to DNA base occurrence in a genome. We ignore the time index and collect the frequency-energy signatures in a 2D histogram of frequency-energy occurrence (Figure 4). This histogram is bounded by the minimum and maximum observed values and binned according to memory tolerance. Smoothing this histogram ($\sigma = 2$) allows better handling of rare events. We vectorized this 2D histogram to produce a 100-element feature vector that describes the occurrence of events over the observation period per subcellular location. In this way 2 subcellular locations are declared similar if they exhibit the same content of events and the events do not necessarily have to occur in the same order.

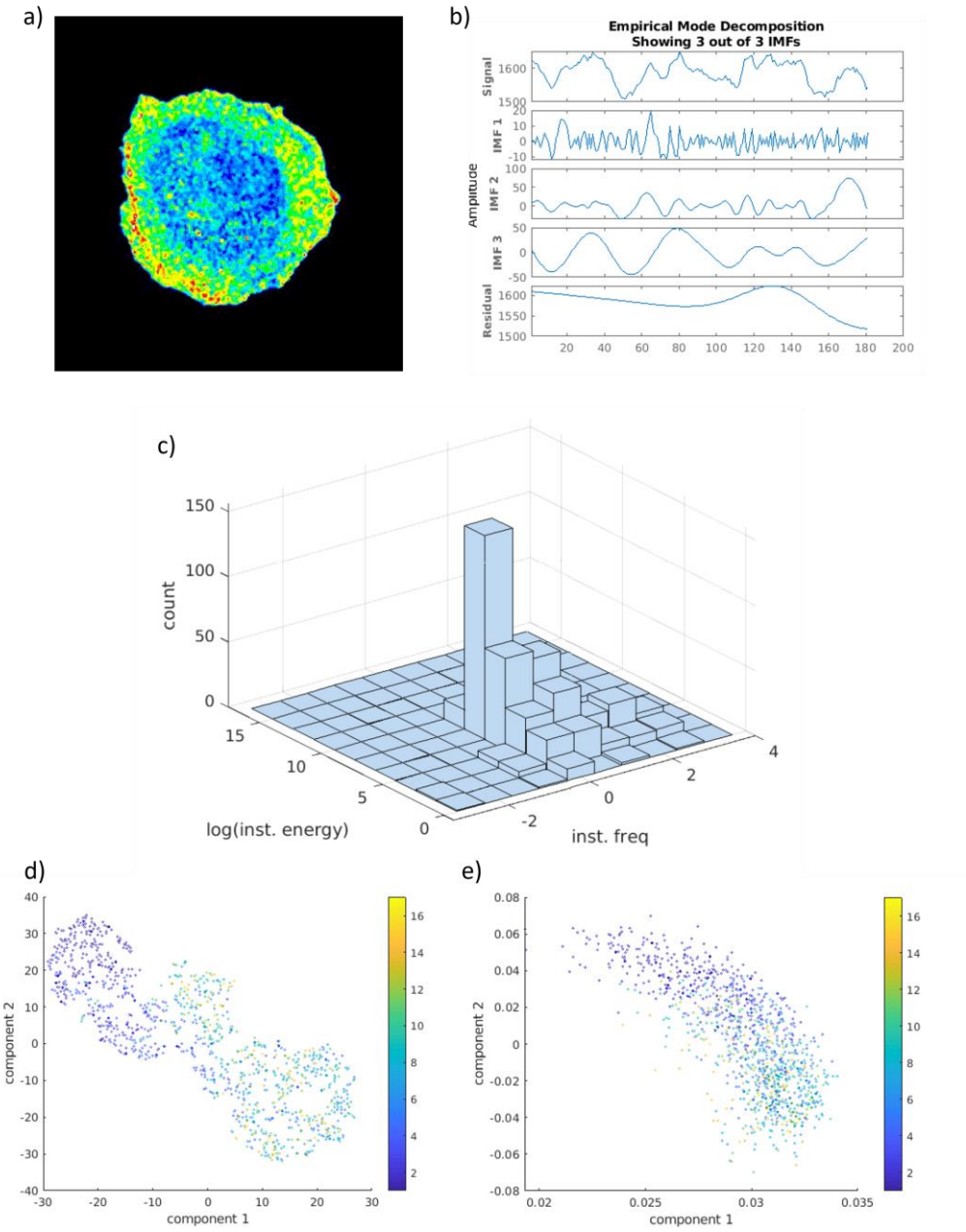

Figure 4) HHT based features real structured projections. a) example cell containing CDC42 biosensor b) example EMD results decomposing a signal (top row) into 3 IMFs. c) example 2D instantaneous frequency event occurrence histogram. d) TSNE projection of HHT based features e) first 2 PCA components of HHT features. d&e) color scale indicates distance of the source locations from the cell center.

We show in figure 4 an example of a Cdc42 biosensor movie processed using our HHT based feature set. For our movies we typically see near noise frequencies in the first IMF (Fig 4b) and high amplitude residuals due to a high level of background fluorescence. Due to the extremely high range in the amplitude axis we chose to compute the log transform of the instantaneous energy for the feature histogram. Importantly we see the vast majority of observed events at high frequencies. This was entirely expected due to the high noise present in biosensor movies. Optimistically, however, we see both a TSNE and PCA projection of the feature space reveal a clear gradient corresponding to distance away from the cell edge. (Fig 4 d & e) The TSNE projection in particular reveals potential for clusters.

Results) A cell manifold-based feature projection

Though generic approaches TSNE & PCA can reveal some structure within our feature set, we wanted a more obvious projection for high confidence clustering. One popular method that seeks to identify structure within the dataspace is spectral embedding. Classically spectral embedding is a dimensionality reduction technique that seeks to "unwrap" a high dimensional feature space according to its latent manifold.[16,17] The classical algorithm proceeds as follows: Let $\mathbf{X_1,...,X_n}$ be a set of datapoints. We want to calculate a pairwise affinity matrix $\mathbf{W}$. We define the similarity $\mathbf{W_{i,j}}$ between data points $\mathbf{X_i}$ and $\mathbf{X_j}$ as:

$$W_{i,j} = e^{\frac{\|x_i - x_j\|^2}{2\sigma^2}}$$

Where $\boldsymbol{\sigma}$ defines the size of the gaussian kernel and therefore search radius. For projection the classic approach is to utilize multidimensional scaling (MDS) which rotates and projects

the affinity matrix along its first **p** largest eigenvalue eigenvectors where **p** is the desired number of dimensions. However, any projection technique can be used including TSNE which in this case converts the affinity matrix to neighborhood probability and produces a projection that favors local structure by minimizing the Kullback–Leibler divergence between datapoints in the projected and feature space.

For comparison of subcellular time series, we know that local points are likely to be similar both through influence of diffusive upstream regulators and a degree of uncertainty due to our remapping process. Our prior success in finding zones of membrane protrusion-retraction dynamics utilized a localized a neighborhood merging process.[13] We employ a similar concept through spectral embedding.

There exist prior variants of special embedding where a geodesic distance is calculated in feature space where the distance $\mathbf{W_{i,j}}$ is the shortest path through the $\mathbf{k_{feat}}$ nearest neighbors of every sample. To do this we first construct a undirected graph **G(V,E)** where we refer to our data points $\mathbf{X_1,\ldots,X_n}$ as a set of vertices $\mathbf{V(G) \in \{v_i,\ldots,v_n\}}$ and edges **E(G)**. Let $\mathbf{D = (d_{i,j})_{j \in [1..n]}}$ be the set of pairwise distances between vertex **vi** and every other vertex. An edge exists between any two vertices $\mathbf{v_i}$ and $\mathbf{v_j}$ if the distance between them is among the $\mathbf{k_{feat}}$ shortest.

$$d_{i,j} \in \left\{ d \in D : |D \cap (-\infty, d)| \leq k_{feat} \right\}$$

If an edge connects $\mathbf{v_i}$ and $\mathbf{v_j}$ the affinity $\mathbf{W_{i,j}}$ between them is simply:

$$W_{i,j} = \left\| X_i - X_j \right\|_2$$

For Euclidean distances. If an edge doesn't connect $\mathbf{v_i}$ and $\mathbf{v_j}$ we search through the possible simple paths between the two vertices to find a path **P** that minimizes the cumulative affinity of all vertex pairs connected by the path's edges. The affinity $\mathbf{W_{i,j}}$ is then:

$$W_{i,j} = \sum_{e \in P} \|X_k - X_l\|_2 : v_k \,\&\, v_l \in e$$

Where **k** and **l** are indices of vertices connected by any edge **e** in path **P**.

Instead of "unwrapping" the features space through an approximation of the latent manifold in feature space like prior works we instead presuppose that the latent structure is the structure of the cell. For every vertex $\mathbf{v_i}$ there is a corresponding subcellular location $\mathbf{m_i}$ from which we sampled data point $\mathbf{X_i}$. An edge exists between any two vertices $\mathbf{v_i}$ and $\mathbf{v_j}$ if the subcellular location $\mathbf{m_i}$ is within a cutoff distance $\boldsymbol{\varepsilon}$ of $\mathbf{m_j}$.

Results) Clustering of cell manifold distance projection reveals microdomains

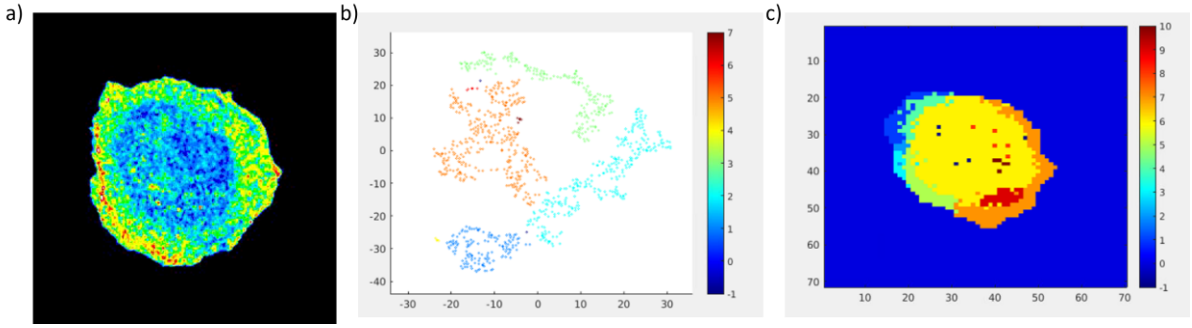

Figure 5) CDC42 microdomains
a) image of CDC42 activity in example cell
b) TSNE projection of cell manifold distance clustered with DBSCAN colored by cluster index
c) Cluster index projected back to subcellular locations

We show an example of a CDC42 cell with microdomains identified through our projection approach. In our given example we have a cell (Fig 5a) compact around the nucleus. By visual examination we could appreciate 3 regions of protrusive activity: top left, right, and bottom left of the cell. From our prior work in analyzing subcellular signaling through edge referenced windows we anticipated that at least 4 pixels of data were necessary to overcome the high noise in biosensor readouts.[1] We down sampled our remapped movies to 25% of the original resolution and spatially smoothed each frame with a gaussian kernel ($\sigma = 2$) In all likelihood this preprocessing procedure would reduce the appreciable structure in the feature space.

We calculated the pairwise cell manifold distance between every data point defining locations within 3 pixels as a neighborhood and projected the pairwise distances to 2 dimensions with TSNE (perplexity 30, exaggeration 4). This produced a highly structured feature projection (Fig 5b). This feature projection was not amenable to clustering via k-means due to its highly nonlinear structure. The typical approach to clustering such a feature space is the DBSCAN algorithm.[21] DBSCAN takes two parameters: ε and “MinPts”. We used a standard heuristic of data dimensionality + for MinPts = 3. We then searched through a reasonable range of ε to set a value of 2 which is the lowest value that allows facilitates near complete assignment of the feature space to clusters. This produced 7 identifiable regions within the feature space.

Projecting the cluster identities back into the original cell space we can see that the majority of the cell around the nucleus belongs to a single microdomain. Since all subcellular locations with the rare exceptions must be assigned to a microdomain, this is consistent with

the earliest observations of CDC42 being active largely near the protrusive cell edge.[22] Interestingly we see that the protrusive cell front at the top left is divided into two zones. This could be due to the presence of a small distinct protrusive region within a broader lamella.

Results) Optogenetic activation of Rac1 in a polarized cell causes microdomain reorganization at targeted site

To lend confidence that our identified microdomains are biologically meaningful beyond anecdotal examples we asked whether a local perturbation would reorganize the pattern of microdomains within the cell. Our collaborators in the Klaus Hahn group developed a strategy to simultaneously image a biosensor and perform localized optogenetic perturbation. Our qualitative analysis of biosensor revealed that our microdomain mapping with our chosen preprocessing tends to group entire protrusive zones into single microdomains. We therefore chose to analyze movies where the optogenetic perturbation only affects part of a protrusion.

The earliest studies of Rho GTPases argued that Rac1, a driver of branched actin cytoskeleton, and RhoA, a driver of myosin contractility in linear actin bundles, are functionally antagonistic to each other.[22] While this has been debunked as an overly simplistic interpretation of a complex interaction network between Rac1 and RhoA we hypothesized that a quantitative approach would reflect such a canonical observation.

We selected a movie of a classically polarized migrating cell with a single large lamellipodia and long tail expressing a RhoA biosensor. Halfway through the movie Rac1 was photoactivated in a zone covering half the lamellipodia. We mapped the microdomains

using the first half of the movie as our control sample and the full movie as the perturbed sample.(Fig 6)

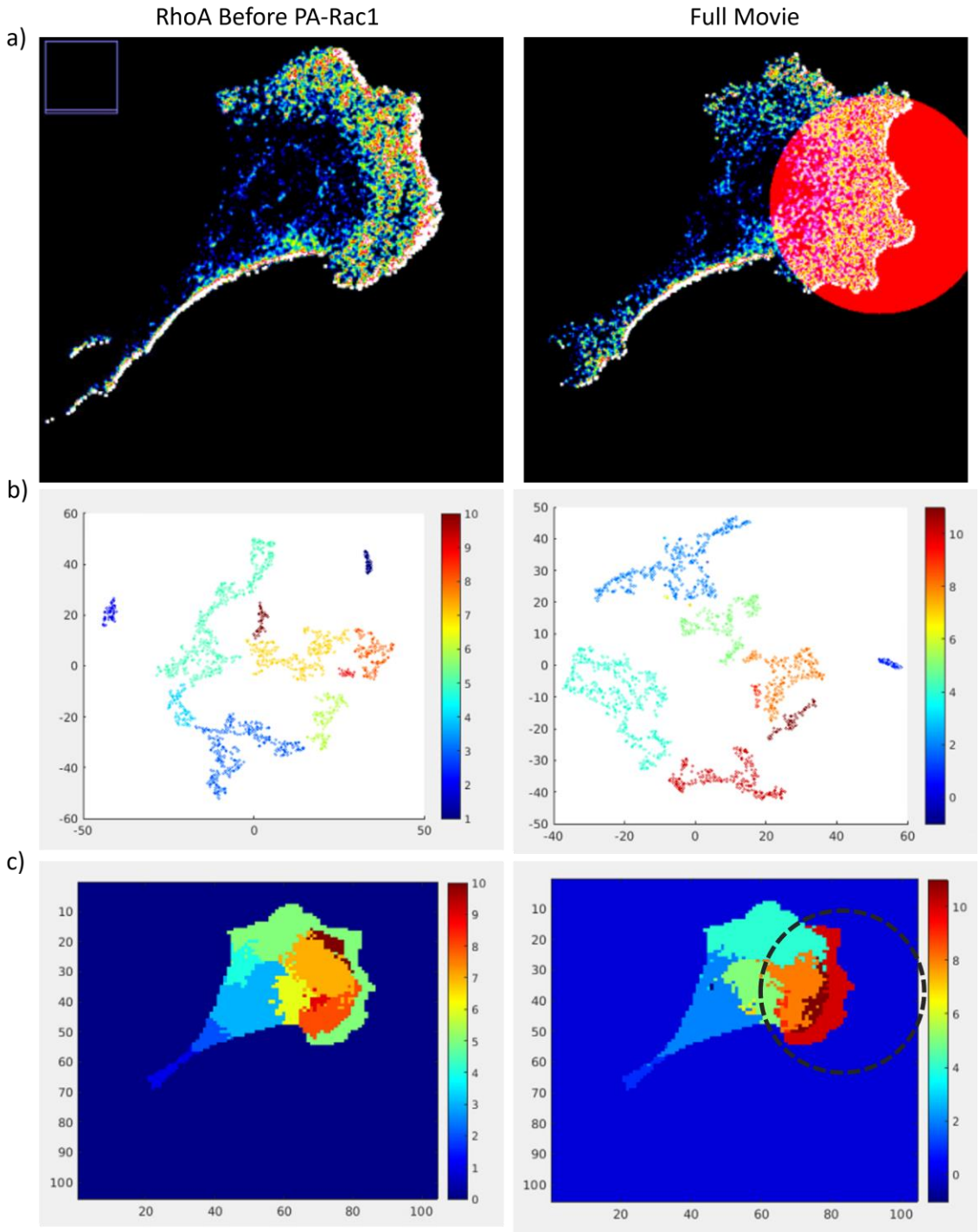

Figure 6) RhoA Microdomains Reorganized by Rac1 Photoactivation
a) image of measured RhoA activity
b) TSNE projection of cell manifold distance clustered with DBSCAN colored by cluster index
c) Cluster index projected back to subcellular locations. Dashed circle indicates area of Rac1 activation

Not only did we see a reorganization of mapped RhoA microdomains but the microdomain located at the lamellipodia was split from a single band across the entire structure into two right along the boundary of Rac1 activation.(Fig 6c) Deeper regions as well seemed to reorganize around the photoactivation but in a less obvious manner. Visually we can't appreciate any difference in intensity between the zones before and after photoactivation. In fact, the separation between the lamellipodia microdomains after Rac1 activation in the projected feature space (Fig 6b right column red and teal points) are distinct but small suggesting that Rac1 activation did not produce a dramatic effect.

We also see in this stereotypically polarized cell that our microdomain analysis reported exactly the textbook diagram of expected signaling organization despite being an entirely automated quantitative mapping. The fact that the entire lamellipodia organized as a single microdomain despite subregions undergoing distinctly out of phase protrusion-retraction cycles shows as well that our feature set is invariant to such effects as desired.

Discussion

In this work we developed a feature set and feature projection strategy specific for determining the subcellular organization of signaling visualized through noisy biosensor movies.

Our feature set is designed to be tunable to the specific characteristics of movies through adjustment of the range and resolution of our Hilbert-Huang event histogram. It improves upon existing approaches to instantaneous frequency-based classification in cell biology by allowing each time-series to be decomposed into IMFs based on independent

optimization instead of a global constraint. By conceptualizing a time series as a collection of events with specific frequency-energy signatures we allow a phase invariant comparison of dynamics and avoid a rank-based comparison of frequency bands. Our event content features are tunable to both computational capacity and prior knowledge of important frequency characteristics.

Our feature projection strategy builds upon the extensive literature in spectral embedding by “unwrapping” a high dimensional feature space according to a subcellular location adjacency manifold. This has the potential to be applicable to a wide variety of biological data for which we have a location prior. While we restricted our manifold definition to purely location adjacency, this can be further expanded to prune edges of our undirected graph based on measurable image properties. For example, we can a priori create a separation between cytoplasmic and nuclear positions by declaring that edges cannot cross a cell nucleus marker boundary. In this way we expect this to be a generalizable approach appreciating the organization of high dimensional data.

Applying these techniques gives us for the first time a full subcellular map of GTPase signaling dynamics organized into internally consistent microdomains. We were able to recapitulate textbook diagrams of subcellular organization using an objective reproducible approach. For the purposes of this paper, we presented conservative microdomain maps. Many of the microdomains we show can be further subdivided to better understand fine grain organization. We show though an optogenetic experiment that our microdomains reflect known biological organizing forces and can reveal organization not readily appreciable

through visual inspection. This study is the starting point to allow an understanding of signaling organization on par with seminal mapping of cellular organelles.

CHAPTER FIVE
Conclusions and Recommendations

Summary of the work

We present in this thesis a review of key principles in live cell signaling analysis starting with a brief history of Rho GTPases and the molecular tools that enable their direct measurement in living systems at a submicron resolution as well as the critical limitations of these tools. We briefly cover some examples of microdomains that have been described using largely qualitative descriptions and anecdotal images. Nevertheless, these anecdotal microdomains form the basis of how we understand signaling is organized within the cell and we cover examples of the limited forays into a systematic objective descriptor of these structures using statistical machine learning methodologies.

By adapting commonly used optical flow based nonlinear image registration we created a method specific for live cell movies that preserves the dynamics of a signal of interest by remapping using a separate measurable subcellular location fiducial. This is an extension as well on our labs previously published method of cell edge-based referencing of subcellular locations that was incapable of extracting interpretable subcellular time series more than a few microns away from the cell edge. We showed that our method overcomes this key limitation and allows sampling of subcellular time series from every subcellular location through our discovery of organized profilin dynamics in moving cell and that these profilin dynamics are related to actin dynamics due to their ability to bind growing actin structures likely through actin polymerizing factors.

Extending on this application we showed that our remapping method can fully analyze molecular biosensor movies of Rho GTPases. The monomolecular structure of our GTPases provided a reliable location fiducial signal. To map the organization of these live cell movies we developed an event content feature set based on instantaneous frequency-energy signatures derived from the Hilbert Huang spectrum decomposition of subcellular time series. We additionally developed a general feature projection strategy that combines feature similarity in high dimensional feature space and prior known data-based manifolds to produce highly structured projections. This feature set and projection strategy allowed us to fully map subcellular signaling into microdomains. Not only do these microdomains match textbook descriptions and expectations of GTPase organization, we, utilizing optogenetic perturbation, show that map of microdomains track almost precisely with the perturbation footprint. Hence, this thesis presents a general framework from taking a given subcellular signaling molecule of interest and revealing its organization within the cell in an objective and quantitative manner.

.

APPENDIX A
Chapter 2 Works Cited

1. Nobes CD, Hall A. Rho, rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. Cell. 1995;81(1):53-62.

2. Frydman RB, Feinstein G. Studies on porphobilinogen deaminase and uroporphyrinogen 3 cosynthase from human erythrocytes. Biochim Biophys Acta. 1974;350(2):358-73.

3. Sahai E, Marshall CJ. RHO-GTPases and cancer. Nat Rev Cancer. 2002;2(2):133-42.

4. Azoitei ML, Noh J, Marston DJ, et al. Spatiotemporal dynamics of GEF-H1 activation controlled by microtubule- and Src-mediated pathways. Journal of Cell Biology. 2019;218(9):3077-3097.

5. Wei J, Mialki RK, Dong S, et al. A new mechanism of RhoA ubiquitination and degradation: roles of SCF(FBXL19) E3 ligase and Erk2. Biochim Biophys Acta. 2013;1833(12):2757-2764.

6. Martin K, Reimann A, Fritz RD, Ryu H, Jeon NL, Pertz O. Spatio-temporal co-ordination of RhoA, Rac1 and Cdc42 activation during prototypical edge protrusion and retraction dynamics. Sci Rep. 2016;6:21901.

7. Abreu-blanco MT, Verboon JM, Parkhurst SM. Coordination of Rho family GTPase activities to orchestrate cytoskeleton responses during cell wound repair. Curr Biol. 2014;24(2):144-55.

8. Machacek M, Hodgson L, Welch C, et al. Coordination of Rho GTPase activities during cell protrusion. Nature. 2009;461(7260):99-103.

9. Azoitei ML, Noh J, Marston DJ, et al. Spatiotemporal dynamics of GEF-H1 activation controlled by microtubule- and Src-mediated pathways. Journal of Cell Biology. 2019;218(9):3077-3097.

10. Birkenfeld J, Nalbant P, Bohl BP, Pertz O, Hahn KM, Bokoch GM. Gef-h1 modulates localized rhoa activation during cytokinesis under the control of mitotic kinases. Developmental Cell. 2007;12(5):699-712.

11. Ma X, Dagliyan O, Hahn KM, Danuser G. Profiling cellular morphodynamics by spatiotemporal spectrum decomposition. PLOS Computational Biology. 2018;14(8):e1006321.

12. Noh J, Isogai T, Chi J, Bhatt K, Danuser G. Inference of Granger-Causal Relations in Molecular Systems — a Case Study of the Functional Hierarchy among Actin Regulators in Lamellipodia. Cell Biology; 2021.

13. Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer.

14. Dama F, Sinoquet C. Analysis and modeling to forecast in time series: a systematic review. arXiv:210400164 [cs]. Published online March 31, 2021.

15. Wang, C., Choi, H.J., Kim, SJ. et al. Deconvolution of subcellular protrusion heterogeneity and the underlying actin regulator dynamics from live cell imaging. Nat Commun 9, 1688 (2018). https://doi.org/10.1038/s41467-018-04030-0

APPENDIX B
Chapter 3 Works Cited

1. Machacek M., Hodgson L., Welch C., Elliott H., Pertz O., Nalbant P., Abell A., Johnson G.L., Hahn K.M., and Danuser G.. 2009. Coordination of Rho GTPase activities during cell protrusion. Nature. 461:99–103. 10.1038/nature08242

2. Azoitei ML, Noh J, Marston DJ, et al. Spatiotemporal dynamics of GEF-H1 activation controlled by microtubule- and Src-mediated pathways. Journal of Cell Biology. 2019;218(9):3077-3097.

3. Zanic, M. Measuring the effects of microtubule-associated proteins on microtubule dynamics in vitro, The Mitotic Spindle: Methods and Protocols, 47–61 (2016)

4. Ma X, Dagliyan O, Hahn KM, Danuser G. Profiling cellular morphodynamics by spatiotemporal spectrum decomposition. PLOS Computational Biology. 2018;14(8):e1006321.

5. Lee SH, Dominguez R. Regulation of actin cytoskeleton dynamics in cells. Mol Cells. 2010;29(4):311-325.

6. Tong J, Qi Y, Wang X, et al. Cell micropatterning reveals the modulatory effect of cell shape on proliferation through intracellular calcium transients. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2017;1864(12):2389-2401.

7. Gholipour A, Kehtarnavaz N, Briggs R, Devous M, Gopinath K. Brain functional localization: a survey of image registration techniques. IEEE Trans Med Imaging. 2007;26(4):427-451.

8. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage. 2009;45(1). doi:10.1016/j.neuroimage.2008.10.040

9. Wang H, Dong L, O'Daniel J, et al. Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. Phys Med Biol. 2005;50(12):2887-2905.

10. Krebs J, Delingette H, Mailhe B, Ayache N, Mansi T. Learning a probabilistic model for diffeomorphic registration. IEEE Trans Med Imaging. 2019;38(9):2165-2176.

11. Roybal KT, Buck TE, Ruan X, et al. Computational spatiotemporal analysis identifies WAVE2 and cofilin as joint regulators of costimulation-mediated T cell actin dynamics. Sci Signal. 2016;9(424):rs3-rs3.

12. Soheil Kolouri, Serim Park, Matthew Thorpe, Dejan Slepčev, Gustavo K. Rohde: "Transport-based analysis, modeling, and learning from signal and data distributions", 2016; arXiv:1609.04767.

13. Damiano-Guercio J, Kurzawa L, Mueller J, et al. Loss of Ena/VASP interferes with lamellipodium architecture, motility and integrin-dependent adhesion. eLife. 2020;9:e55351.

14. Funk J, Merino F, Venkova L, et al. Profilin and formin constitute a pacemaker system for robust actin filament growth. eLife. 2019;8:e50963.

15. Noh J, Isogai T, Chi J, Bhatt K, Danuser G. Inference of Granger-Causal Relations in Molecular Systems — a Case Study of the Functional Hierarchy among Actin Regulators in Lamellipodia. Cell Biology; 2021.

16. Martin, K., Reimann, A., Fritz, R. et al. Spatio-temporal co-ordination of RhoA, Rac1 and Cdc42 activation during prototypical edge protrusion and retraction dynamics. Sci Rep 6, 21901 (2016). https://doi.org/10.1038/srep21901

17. Pollard TD, Blanchoin L, Mullins RD. Molecular mechanisms controlling actin filament dynamics in nonmuscle cells. Annu Rev Biophys Biomol Struct. 2000;29:545-576.

18. Kamiyama D, Sekine S, Barsi-Rhyne B, Hu J, Chen B, Gilbert LA, et al. Versatile protein tagging in cells with split fluorescent protein. Nat Commun. 2016;7:11046.

19. Gunther G, Jameson DM, Aguilar J, Sánchez SA. Scanning fluorescence correlation spectroscopy comes full circle. Methods. 2018;140-141:52-61.

20. Lomakin AJ, Lee K-C, Han SJ, et al. Competition for actin between two distinct F-actin networks defines a bistable switch for cell polarization. Nat Cell Biol. 2015;17(11):1435-1445.

21. Henty-Ridilla JL, Juanes MA, Goode BL. Profilin directly promotes microtubule growth through residues mutated in amyotrophic lateral sclerosis. Current Biology. 2017;27(22):3535-3543.e4.

22. Frank A. Farris. The gini index and measures of inequality. The American Mathematical Monthly. 2010;117(10):851.

23. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR338 Cas9 system. Nat Protoc. 2013;8(11):2281-308.

24. Lackner DH, Carre A, Guzzardo PM, Banning C, Mangena R, Henley T, et al. A generic strategy for CRISPR-Cas9-mediated gene tagging. Nat Commun. 2015;6:10237.

25. Dean KM, Roudot P, Reis CR, Welf ES, Mettlen M, Fiolka R. Diagonally Scanned Light-Sheet Microscopy for Fast Volumetric Imaging of Adherent Cells. Biophys J. 2016;110(6):1456-65.

26. Concordet JP, Haeussler M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. 2018;46(W1):W242-W5.

APPENDIX C
Chapter 4 Works Cited

1. Machacek M, Hodgson L, Welch C, et al. Coordination of Rho GTPase activities during cell protrusion. Nature. 2009;461(7260):99-103.

2. Hetmanski JH, Zindy E, Schwartz JM, Caswell PT. A MAPK-Driven Feedback Loop Suppresses Rac Activity to Promote RhoA-Driven Cancer Cell Invasion. PLoS Comput Biol. 2016;12(5):e1004909.

3. Rathinam R, Berrier A, Alahari SK. Role of Rho GTPases and their regulators in cancer progression. Front Biosci (Landmark Ed). 2011;16:2561-71.

4. Birkenfeld J, Nalbant P, Bohl BP, Pertz O, Hahn KM, Bokoch GM. GEF-H1 modulates localized RhoA activation during cytokinesis under the control of mitotic kinases. Dev Cell. 2007;12(5):699-712.

5. Martin K, Reimann A, Fritz RD, Ryu H, Jeon NL, Pertz O. Spatio-temporal co-ordination of RhoA, Rac1 and Cdc42 activation during prototypical edge protrusion and retraction dynamics. Sci Rep. 2016;6:21901.

6. Huang, B., & Elsevier. (2017). Comprehensive GIS : Geographic Information System. Elsevier.

7. Müller, P.M., Rademacher, J., Bagshaw, R.D. et al. Systems analysis of RhoGEF and RhoGAP regulatory proteins reveals spatially organized RAC1 signalling from integrin adhesions. Nat Cell Biol 22, 498–511 (2020). https://doi.org/10.1038/s41556-020-0488-x

8. Fulcher BD, Jones NS. Hctsa : a computational framework for automated time-series phenotyping using massive feature extraction. Cell Systems. 2017;5(5):527-531.e3.

9. Jaqaman K, Dorn JF, Marco E, Sorger PK, Danuser G. Phenotypic clustering of yeast mutants based on kinetochore microtubule dynamics. Bioinformatics. 2007;23(13):1666-73.

10. Jaqaman K, Dorn JF, Jelson GS, Tytell JD, Sorger PK, Danuser G. Comparative autoregressive moving average analysis of kinetochore microtubule dynamics in yeast. Biophys J. 2006;91(6):2312-25.

11. van der Maaten, L.J.P.; Hinton, G.E. (Nov 2008). "Visualizing Data Using t-SNE" (PDF). Journal of Machine Learning Research. 9: 2579–2605.

12. Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer.

13. Wang, C., Choi, H.J., Kim, SJ. et al. Deconvolution of subcellular protrusion heterogeneity and the underlying actin regulator dynamics from live cell imaging. Nat Commun 9, 1688 (2018). https://doi.org/10.1038/s41467-018-04030-0

14. Ma X, Dagliyan O, Hahn KM, Danuser G. Profiling cellular morphodynamics by spatiotemporal spectrum decomposition. PLOS Computational Biology. 2018;14(8):e1006321.

15. Huang NE, Wu Z. A review on Hilbert-Huang transform: Method and its applications to geophysical studies. Rev Geophys. 2008;46(2):RG2006.

16. Tenenbaum JB. A global geometric framework for nonlinear dimensionality reduction. Science. 2000;290(5500):2319-2323.

17. for the Alzheimer's Disease Neuroimaging Initiative, Viswanath SE, Tiwari P, Lee G, Madabhushi A. Dimensionality reduction-based fusion approaches for imaging and non-imaging biomedical data: concepts, workflow, and use-cases. BMC Med Imaging. 2017;17(1):2.

18. A.Y. Ng, M.I. Jordan, Y. Weiss. On spectral clustering: analysis and an algorithm. Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, MIT Press (2001)

19. Marston DJ, Vilela M, Huh J, Ren J, Azoitei ML, Glekas G, Danuser G, Sondek J, Hahn KM. Multiplexed GTPase and GEF biosensor imaging enables network connectivity analysis. Nat Chem Biol. 2020 Aug;16(8):826-833. doi: 10.1038/s41589-020-0542-9. Epub 2020 May 18. PMID: 32424303; PMCID: PMC7388658.E.

20. W. Dijkstra. A note on two problems in connexion with graphs. Numerische Mathematik, pages 269–271, 1959.

21. Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. KDD.

22. Nobes CD, Hall A. Rho, rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. Cell. 1995;81(1):53-62