RESEARCH ARTICLE

Inferring cell state by quantitative motility analysis reveals a dynamic state system and broken detailed balance

Jacob C. Kimmel[1,2], Amy Y. Chang[1], Andrew S. Brack[2,3]*, Wallace F. Marshall[1]*

1 Dept. of Biochemistry and Biophysics, Center for Cellular Construction, University of California San Francisco, San Francisco, CA, United States of America, **2** Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, University of California San Francisco, San Francisco, CA, United States of America, **3** Dept. of Orthopedic Surgery, University of California San Francisco, San Francisco, CA, United States of America

* Andrew.Brack@ucsf.edu (ASB); Wallace.Marshall@ucsf.edu (WFM)

Abstract

Cell populations display heterogeneous and dynamic phenotypic states at multiple scales. Similar to molecular features commonly used to explore cell heterogeneity, cell behavior is a rich phenotypic space that may allow for identification of relevant cell states. Inference of cell state from cell behavior across a time course may enable the investigation of dynamics of transitions between heterogeneous cell states, a task difficult to perform with destructive molecular observations. Cell motility is one such easily observed cell behavior with known biomedical relevance. To investigate heterogenous cell states and their dynamics through the lens of cell behavior, we developed Heteromotility, a software tool to extract quantitative motility features from timelapse cell images. In mouse embryonic fibroblasts (MEFs), myoblasts, and muscle stem cells (MuSCs), Heteromotility analysis identifies multiple motility phenotypes within the population. In all three systems, the motility state identity of individual cells is dynamic. Quantification of state transitions reveals that MuSCs undergoing activation transition through progressive motility states toward the myoblast phenotype. Transition rates during MuSC activation suggest non-linear kinetics. By probability flux analysis, we find that this MuSC motility state system breaks detailed balance, while the MEF and myoblast systems do not. Balanced behavior state transitions can be captured by equilibrium formalisms, while unbalanced switching between states violates equilibrium conditions and would require an external driving force. Our data indicate that the system regulating cell behavior can be decomposed into a set of attractor states which depend on the identity of the cell, together with a set of transitions between states. These results support a conceptual view of cell populations as dynamical systems, responding to inputs from signaling pathways and generating outputs in the form of state transitions and observable motile behaviors.

Author summary

Cells in a population are not all alike. The differences between cells impact how each cell responds to stimuli, with implications for development and disease. How do cells change between these functionally important states over time? This question is difficult to answer using molecular biology, because these methods destroy cells in order to observe them, preventing us from observing the same cell more than once. Cells in culture display a diverse set of motility behaviors that can be observed non-destructively over time. We have developed software to quantify these behaviors, and used this tool to identify cell motility states and quantify the changes in cell motility states over time. Using this approach, we identify that cancerous fibroblasts behave similarly to normal fibroblasts, but prefer some behaviors to others. This suggests cancerous transformation may lead to changes between states, rather than just creating new ones. Applying this analysis to muscle stem cells, we measure the rate of change from the stem cell to progenitor state for the first time. We are also able to determine if changes in behavior state are externally driven or random, demonstrating the utility of observing cell behaviors to study changes in cells over time.

Introduction

Within any group of cells, each individual cell is not necessarily like its neighbors. These differences often have functional significance, with cells at different points along the phenotypic spectrum exhibiting distinct behavior [1]. This has been noted across a broad swath of cell biology, including in the cases of stem cell biology [2, 3] and cell geometry definition [4]. Regulatory decisions at one scale may reflect phenotypes at other scales, allowing identification of a broad "cellular state" based on a more limited set of observations. Effective classification of cancerous cell functionality based on morphology demonstrates this concept [5–7]. Since cell state determines cell function, state transitions may manifest as changes in cell behavior. Understanding the regulation of cell behavior will require understanding the nature of the cell state-space and the transitions that take place within it. Is the state space continuous or discrete? Are all state transitions equally likely, as in an equilibrium system, or do they tend to take place in a specific sequence, as in computing device? Answering such questions requires a framework for defining cell state in terms of observable behaviors. Regardless of the method used to probe cell state, it must be able to measure state in living cells at multiple time points, in order to allow state transitions to be characterized.

Recent advances in single cell assays have allowed for detailed, quantitative descriptions of individual cells at the molecular level. Single cell sequencing technologies in particular have uncovered heterogeneity at both the transcriptional and epigenetic level [8, 9]. In multiple stem cell compartments, molecular analysis of heterogeneity has revealed that not all stem cells are functionally equivalent. Within the hematopoietic [10], muscle [11], epithelial, and other [12] stem cell pools, subpopulations of cells with different functionality coexist. Stem cell heterogeneity has been demonstrated in the form of lineage-bias or differences in regenerative capacity. A similar phenomena is present in malignant tumors. Within a given tumor, some subpopulations may have higher tumorogenic potential, or increased resistance to a particular therapy [13, 14]. Understanding heterogeneity in these and other contexts is essential to building an accurate picture of cell-based systems.

In addition to defining cell states to account for population heterogeneity, we also seek to understand the transitions between states, because those transitions reveal the logic of the

cellular control system. In seminal work, Waddington introduced the conceptual model of an 'epigenetic landscape' governing cell phenotype decisions, akin to a potential energy landscape in a physical system [15, 16]. In this model, cells progress through phenotypic states by migrating continuously down the gradient of the landscape, eventually resting in a stable basin of attraction. This model implies that cell state transitions are governed by a 'potential energy' in each state, which can be estimated by the state's stability. A cell system governed by this model would display detailed balance in the absence of an external force or signal, and break detailed balance in the presence of such an external input [16, 17]. By directly measuring the dynamics of cell state transitions, we can produce an estimate of the landscape for cell behavioral phenotypes and determine if state transitions occur stochastically or are influenced by external inputs.

While existing molecular assays such as single-cell RNA-sequencing can provide detailed information about a cell population's heterogeneity, these assays are generally destructive and restricted to a single time point for analysis. Methods have been developed to infer cell state lineages from observations at a single time point [18, 19], but these methods are not able to quantify dynamics and assume that proximity in the measured variable space describes a transition relationship between states. Real-time, non-destructive assays that reveal subpopulation composition over time and observe state transitions in living cells would serve as complementary approaches to investigate cellular states and state transitions.

Cells in culture have a diverse behavioral repertoire. Each cell may exhibit motion, dynamic morphology changes [4, 20], and symmetric or asymmetric divisions, all of which can be observed by simple microscopy. These cell behaviors represent a rich phenotypic space from which many quantitative features may be extracted. Behavior also has an inherent functional relevance at the single cell level, and often a functional relevance on the system level. For example, extracellular matrix remodeling and deposition by fibroblasts is dictated by their motion [21, 22], and stem cell and progenitor migration is critical for organismal development [23]. Within each cell, behavior represents a layer of abstraction above molecular phenotypes. Hence, observably different behavioral states may serve as a proxy for distinct ensembles of underlying molecular states.

Recent work from several groups has elucidated the regulatory priciples governing cell shape definition [24, 25] and the relationships between morphological variables [26] utilizing quantitative, timelapse imaging. Profiling of static morphological information alone is sufficient to disciminate various states of cellular perturbation [27–30]. Considering temporal information, quantification of the dynamic transitions between morphological phenotypes using multiple Hidden Markov Model frameworks has allowed for increased precision in classifying both chemical and genetic cellular perturbations [20, 31]. Progenitor fate outcomes have also been predicted based on cell shape and behavior quantifications [32]. These remarkable results, achieved through quantitative exploration of cell morphology behaviors and the dynamic transitions between them, support the notion that quantitative analysis of cell behavior can allow for detection of cellular states and observation of state transitions. Modeling cells as dynamical systems in this manner has proven to not only elucidate fundamental biological principles, but to provide utility in applied contexts. Here, we extend this approach to the analysis of cell motility behaviors and consider alternative tools for extracting biological insight from observed state transitions.

Cell motility behaviors are inherently related to the morphological state of a cell [33–35]. Different motility behaviors are associated with different morphological states, and some morphological events such as protrusion extension are directly causal for motility behaviors. However, motility events involving a displacement event where the cell moves along a subtrate to a new location are not necessarily captured enitrely through morphological descriptors.

Supporting this notion, it has recently been shown that displacement motility behaviors provide a strong signal for the inference of hematopetic stem cell differentiation decisions when used in addition to morphological descriptors [36]. Therefore, we believe it is valuable to develop means of measuring motility behaviors and considering transitions between different motility phenotypes to capture the full range of information provided by cell behavior.

Cell motility is a particularly dramatic cell behavior, but it is difficult to examine quantitatively. Traditional cell motility assays rely on a binary filtration of cells based on a functional test, such as crossing a membrane barrier. Timelapse microscopy has also been appreciated for decades as a means of tracking and quantifying cell motility at the single cell level [37–40]. These classic studies demonstrate that cell motility is predictive of cellular function and that quantitative motility analysis can elucidate underlying cell state control mechanisms [39, 40]. Recent approaches to timelapse motility analysis have expanded upon these techniques to extract multi-dimensional quantitative information from individual cells [32, 41–45]. However, existing methods focus largely on speed and distance metrics of cell motility on a single arbitrarily-chosen timescale, limiting the degree of heterogeneity that can be revealed within a population.

Here, we present Heteromotility, a software tool for quantitative analysis of cell motility in timelapse images with a diverse feature set. In addition to commonly calculated features such as distance traveled, turning, and speed metrics, Heteromotility provides features that allow for comparisons to models of complex motion, such as Levy flights and fractal Brownian motion, and estimate long-term dependence within a cell's displacement distribution. This feature set creates a high-dimensional space representing the possible phenotypes of cell motility. These features may be useful for multiple downstream applications, including supervised classification of different motility phenotypes, and unsupervised definition of different motility behaviors, both of which we demonstrate here. Our tool is modular, and we also provide tools to map the high-dimensional motility feature space into a low dimensional state space, tools to quantify changes in cell motility phenotypes over time as transitions in state space, and tools to consider the stochastic or directed nature of these transitions. These tools allow the dynamics of cell behavior states to be elucidated, in addition to simply identifying heterogenous phenotypes.

We demonstrate that Heteromotility analysis is sufficient to discriminate between simulations of several models of motion. Applied to wild-type and transformed mouse embryonic fibroblasts (MEFs), Heteromotility analysis reveals a shared set of motility states between the two systems, in which transformed cells preferentially occupy a more motile state. In mouse myoblasts and MuSCs, phenotypically distinguished motility states are revealed. MuSCs display a series of progressively more motile states, suggesting states of activation. Quantifying transition dynamics within this state system over time reveals that MuSCs transition through state space toward a progressively more activated, myoblast-like motility phenotype. Viewed through the lens of myogenic activation, we are able to follow the activation dynamics of individual MuSCs for the first time. Applying probability flux analysis, we find that the MuSC motility state system breaks detailed balance, quantitatively confirming that these state transitions occur in an ordered and predictable sequence, as in a computing device.

Results

Heteromotility analysis approach

Heteromotility analyzes motility features in a set of provided motion paths, as obtained through timelapse imaging, image segmentation, and tracking (see Methods). From these paths, 79 motility features are calculated to comprise a "motility fingerprint" (S1 Fig). These features include simple metrics of speed, total and net distance traveled, the proportion of time

a cell spends moving, and the speed characteristics during that period. The linearity of motion is assessed by linear regression through all points occupied by the cell in the time series, taking Pearson's r^2 as a metric of fit. Monotonicity is also considered for the distribution of points, using Spearman's ρ^2. In some instances, cells have been proposed to have a directional bias when making turns [46]. Turn direction and magnitude features are provided that consider turns on various time intervals.

Another class of features is concerned with the directionality and persistence of motion. There are several possible ways to characterize the persistence of motion, hence we have included several metrics that allow for reasonable consideration. A cell's "progressivity," is considered as the ratio of net distance to total distance traveled. This serves as a metric of directional persistence [47]. Mean squared displacement (MSD) is considered for each cell for a variable time lag τ, and the power law exponent α is taken as a quantification of the relationship $MSD \propto \tau^{\alpha}$. Directed motion may therefore be expected to have a larger value than undirected random-walk motion or non-motion, indicating a superdiffusive behavior [48].

The distribution of a cell's displacement steps is also informative. Motion that exhibits a displacement distribution with a heavy right tail, referred to as a Levy flight, has been demonstrated to optimize the success of a random search [49]. This property has led to the Levy flight foraging hypothesis, suggesting that biological systems may perform Levy flight-like motion when searching for resources [50–52]. To assess the Levy flight-like nature of a cell's motility behavior, the Heteromotility software provides metrics of displacement kurtosis for displacement distributions on multiple time scales. Larger values of kurtosis indicate heavier distribution tails, such that higher kurtosis may indicate more Levy flight-like motion. The Heteromotility software also considers the non-Gaussian parameter α_2, for which larger values indicate a more heavily tailed, Levy flight-like distribution [53, 54]. The second and third moments of the displacement distribution are also provided as features.

If displacements are considered as a time series, the self-similarity and long range memory may provide insight into the coordination of motility behavior [39]. The Heteromotility software calculates the autocorrelation function for displacements with variable time lags τ as a metric of self-similarity. Fractal Brownian motion (fBm) describes a Gaussian process $B_H(t)$ for a time t on the continuous interval $[0, T]$ with successive displacements that are not necessarily independent. The Hurst parameter H describes the self-similarity of a fBm process, with the interval $0 < H < 0.5$ describing a process with negatively correlated successive displacements (large displacements are more likely to follow small displacements), $H = 0.5$ describing non-correlated, independent successive displacements (Brownian motion), and $0.5 < H < 1$ describing positively correlated successive displacements (large displacements are more likely to follow large displacements)(see Supp. Methods for additional detail). The Heteromotility software estimates the Hurst parameter as a metric of long range memory using Mandelbrot's rescaled range method (see Supp. Methods) [55] [56]. As Brownian motion displays a Hurst parameter $H = 0.5$, deviations from this value may indicate long range memory of a cell's displacement series and coordinated motility behavior. Each of the motility features described here concerns single cells, and none directly measure correlations in motion between cells as may be observed during collective migration. However, cells with similar behaviors, such as groups of "leader," and "follower," cells, may display correlations in single cell features.

Heteromotility features are sufficient to distinguish canonical models of motion

In principle one could devise an infinite number of motion descriptors. How can we determine whether a given set of motion features is sufficient to capture meaningful aspects of

motion? To test whether the feature set outlined above is sufficient to distinguish biologically relevant models of motion, we simulated and analyzed cell paths generated by four distinct motion models: unbiased random walks, biased random walks, Levy flights, and fractal Brownian motion with long range memory ($H = 0.9$)(see Methods for implementations). These data were initially simulated with the same mean displacement size, such that the motion models are not trivially seperable based on a parameter unrelated to their defining characteristics. This problem therefore represents the most challenging setting for separating models of these types. Large simulated data sets ($n = 20,000$) were generated with a range of track lengths. Samples with a range of sizes (members per class) were redrawn from these large populations three times per sample size/track length parameterization. We then analyzed the simulated data to determine if the feature set is sufficient to place each model in a distinct region of a joint feature space.

Visualization of high-dimensional data sets, such as the Heteromotility feature set, presents a fundamental challenge. Here we employ *t*-Stochastic Neighbor Embedding (t-SNE), a visualization method that embeds high-dimensional data into a low-dimensional map, to generate 2D projections of our high dimensional feature space [57] (see Supplemental Methods for discussion of t-SNE parameter selection). When Heteromotility feature space is visualized using t-SNE (perplexity = 50), these models of motion occupy distinct regions of feature space for a number of sample sizes and track lenghts. A sample visualization of a simulation with 1000 members per class and track lengths of 100 units is shown (Fig 1A)(see further examples, S2 Fig). Unsupervised hierarchical clustering was performed using Ward's method [58] (see Supplemental methods for discussion). Unsupervised clustering segregates the various models with high accuracy based on Heteromotility features for a range sample sizes as small as 100 members per class, and lengths as short as 50 time steps. Accuracy tends to increase with larger sample sizes and longer track lengths (Fig 1B and 1C). Clustering was also performed on simulations with varied fundamental parameters, and accuracy is likewise high (Fig 1F). Parameters for the bias magnitude of biased random walkers, random walker displacement means, the Levy flier exponent, and fractal Brownian motion Hurst parameter were all varied. 10 groupings of the four simulations with varying parameter settings were sampled and classified as above (see S2 Table for parameter settings).

Applying a supervised Random Forest classifier [59] to each of the simulated populations, we are able to discriminate the four models of motion with high accuracy (5-fold cross validation score) across a range of sample sizes and track lengths (S2 Fig) (see Supp. Methods for discussion on classification model selection). We present the performance of a sample Random Forest classifier, trained on simulations of 100 time steps with 1000 members per class (shown in Fig 1A), in a confusion matrix. A confusion matrix describes the errors made during classification in a matrix by listing the true classes as rows and the predicted classes as columns with values of the matrix cells representing the number of observations for each true class: predicted class pair (Fig 1D) [60]. In this way, values along the matrix diagonal represent correct predictions, and values off the diagonal represent incorrect predictions. As seen in the matrix, there is little confusion between Levy flights, fBm, and the random walks, but some confusion between unbiased and biased random walks.

The features important for effective classification can be determined by measuring the decrease in accuracy when each feature is removed from classification. For each of the simulated populations generated, we found the top 10 most important features based on this metric. We present the 10 features that appear most often across classifiers, and the mean decrease in accuracy associated with removal of each feature (Fig 1E). We find that the non-Gaussian parameter and other metrics of the displacement distribution, as well as metrics of self-similarity, dominate this list. This supports the notion that the non-traditional features provided by
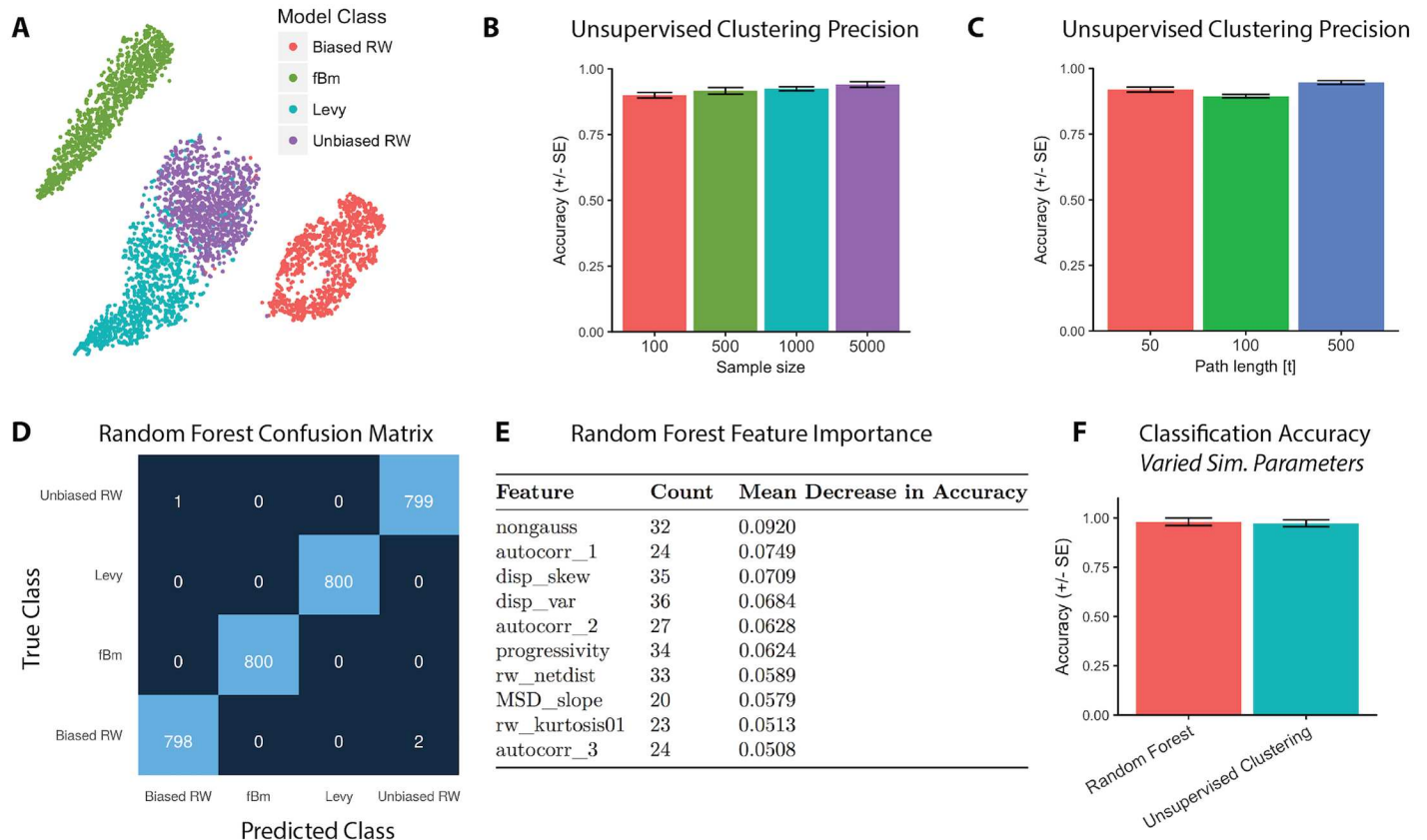
Fig 1. Simulated models of motion can be differentiated based on Heteromotility features. (A) Representative t-SNE visualization of Heteromotility feature space determined for simulated motion models (1000 members/class, 100 time steps)(perplexity = 50). (B) Unsupervised clustering accuracy for simulated data across a range of sample sizes and (C) track lengths. (D) The Confusion Matrix of a representative Random Forest classifier trained to distinguish four simulated models of motion shown in (A), mean accuracy of 99.8% (5-fold CV). (E) Features ranked by importance for Random Forest classification, where importance is determined as the decrease in accuracy when the feature is removed across all 36 simulated sample populations. The number of times the feature appears in the top 10 most important across the 36 populations is reflected in the 'Count' column. (F) Accuracy of classifying simulated motion models with varying parameters using hierarchical clustering and a Random Forest classifier. 10 populations of the four simulated motion models with varied parameters were generated and classified.

https://doi.org/10.1371/journal.pcbi.1005927.g001

Heteromotility can be useful for discriminating biologically relevant models of motion. These results demonstrate robust detection of heterogenous motility phenotypes using a rich space of motion features. We also find high accuracy Random Forest classification of simulations with varied parameters (Fig 1F). As a proof-of-concept, this analysis of simulated motion indicates that the Heteromotility feature set is sufficient to recognize different motility phenotypes where they exist.

Wild-type and transformed MEFs occupy a shared region of motility state-space

Utilizing the Heteromotility software, we are able to analyze behavior in live cells and map this behavior into a cell state space. We initially sought to determine if this cell behavior state space was continuous, with cell states existing along a spectrum, or discrete, with a limited set of states cells could adopt. To answer this question, we first employed the mouse embryonic fibroblast (MEF) culture system. Wild-type MEFs (WT MEFs) and MEFs transformed with *c-Myc* and *HRas-V12* overexpression constructs (MycRas MEFs) [61], which serve as a cancer model, were timelapse imaged for 8 hours to capture motility behavior. Images were
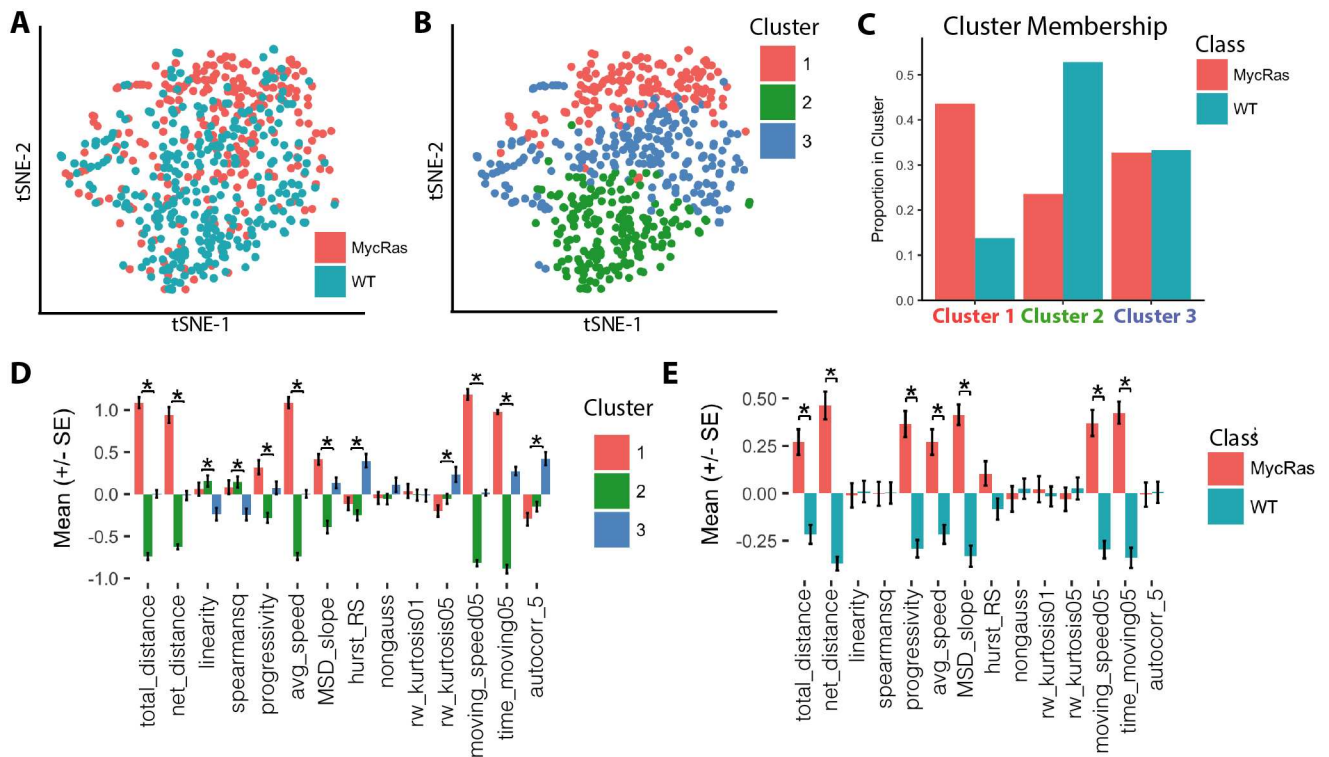
Fig 2. Wild-type and Myc/Ras transformed MEFs show differential occupancy within a shared motility state-space. (A) t-SNE visualization of wild-type (blue) and Myc/Ras transformed (red) MEFs in motility space (perplexity = 50). (B) Hierarchical clusters visualized with t-SNE (MANOVA, $p < 0.001$; Silhouette S_i = 0.21). (C) Proportion of wild-type vs. transformed cells occupying each cluster. (D) Comparison of a subset of normalized feature values between clusters (*: Holm-Bonferroni corrected $p < 0.01$ by ANOVA) and (E) wild-type and transformed MEFs (*: Holm-Bonferroni corrected $p < 0.01$ by t-test). $n > 250$ cells per condition (pooled) from six independent experiments.

https://doi.org/10.1371/journal.pcbi.1005927.g002

segmented and tracked (see Methods), with cell centroid coordinates used as the cell location for tracking. By considering cell centroids, process extensions and retractions are considered as motile behaviors, in addition to displacement behaviors that would not be captured by an analysis of morphology alone [39]. Analysis of cell paths with the Heteromotility software revealed that WT and MycRas MEFs differentially occupy different sections of a shared motility space, as visualized by t-SNE (Fig 2A). Strikingly, MycRas MEFs occupy a sub-region of wild-type motility space, rather than a unique region.

We applied hierarchical clustering (Ward's method) [58] to the motility state space generated from pooled WT and MycRas MEFs to identify heterogenous motility phenotypes (Fig 2B), considering a set of cluster validation indices to select the optimal number of clusters [62] (See Supp. Methods for complete details). Clustering was performed on the first 30 principal components of the motility feature set, as this dimensionality preserves >95% of variation for each system we study (S3 Fig). The Silhoutte value is a metric of cluster validity on the interval [-1, 1], taking into account the similarity of samples within a cluster and the difference between clusters [63]. Higher values indicate that samples within a cluster are similar and clusters are distinct. The cluster partition displays a positive Silhouette value, indicating an appropriate cluster structure. Feature mean values between clusters are also confirmed to be significantly different by multivariate analysis of variance (MANOVA) [64]. A complete set of cluster validation metrics is provided (S1 Table). This result suggests that the state space is continuous, but can be decomposed into a set of overlapping states with characteristic behavioral phenotypes. We define these motility states as heterogeneous based on significant

differenences from one other in feature space. This definition of heterogeneity is also followed for the other cell systems we investigate in this work.

WT and MycRas MEFs are found to differentially occupy different clusters within the shared set, as may be expected from their initial distributions in the space (Fig 2C). MycRas MEFs preferentially occupy Cluster 1, characterized by the highest average speed, proportion of time spent moving, and distance traveled, indicating a motile state. Conversely, WT MEFs preferentially occupy Cluster 2, characterized by the lowest average speed and time spent moving, and progressivity (net distance / total distance), indicating a less motile state (Fig 2D). A proportion of both MycRas and WT MEFs occupy Cluster 3, a population characterized by high kurtosis and progressivity, indicating a Levy flight like motile state that performs "jumping" motions. Observed visually, Cluster 2 cells move relatively little, Cluster 1 cells progress smoothly, and Cluster 3 cells exhibit erratic motion (S1 and S2 Videos). The transformation state-dependent distribution of MEFs between behavioral clusters is reproducible across all "round-robin" groupings of 4 out of the 6 biological experiments we perform (S4 Fig)(Supp. Methods). The clusters are also identifiable using only a small subset of features, or even a single feature, and the relative distribution of WT and MycRas cells within them is maintained (S5 Fig).

To estimate whether local cell density played a role in determining a cell's motility behavior, we estimate a "local cell density index" based on the sum of inverse squares of distances between a cell and its neighbors using our cell tracking data. Constructing linear regression models between this local cell density index and each of our measured motility features, we find no meaningful relationships (S3 Fig). However, we note that our tracking data purposefully does not track every cell, such that the local cell density index we estimate is an imperfect representation of local cell density. We also note that our experiments are designed to minimize cell-cell contact, such that our lack of detected cell density influence may be the simple result of sparse growth conditions failing to reach a quorum sufficient to influence motility behavior.

MycRas and WT MEF cluster preferences are statistically significant by Pearson's χ^2 test of the *transformation state × cluster* contingency table ($p < 0.0001$). Considered as a population, the MycRas MEFs demonstrate higher progressivity, mean squared displacement, time moving, average speed, and self-similarity metrics than WT MEFs. This indicates that as a population they spend more time moving in a directed manner (Fig 2E).

These results suggest that high motility and low motility states exist in both wild-type and transformed MEFs. Oncogenic transformation by *c-Myc* and *HRas-V12* may then be viewed as an input that leads a larger proportion of cells to adopt the high motility state, rather than introducing a novel phenotype unseen in WT cells. This observation is consistent with studies of motility behavior in the context of cancer, which have long suggested that increased motility is a phenotype of malignant cells [65, 66]. Functionally, the increased motility of neoplastic cells may be related to their potential for tissue invasion [67, 68].

Importantly, the motility behavior of a cancer cell population may be indicative of disease progression and outcomes [69]. We trained supervised classification models to predict if a cell was wild-type or transformed based on its motility behavior. Support vector machine (SVM) classifiers are a set of models that learn a decision boundary between classes. SVMs have shown efficacy in many problem domains [70]. SVM classifiers were trained on the top 62% features selected by ANOVA F-value in a round-robin fashion for cross-validation, training on four experiments and testing on the remaining two iteratively. On average, round-robin trained SVM classfiers are able to classify individual cells as wild-type or MycRas transformed with ∼70% accuracy based on motility features alone (S3 Fig). Parameters for feature selection and the SVM classifier were identified by a grid search (see Supp. Methods for details). While

clusters are identifiable with only a single feature, we find classification accuracy decreases if more than ∼ 62% of features are omitted (S5 Fig). The most important features for classification were estimated based on the decrease in round robin classification accuracy when a feature was removed, with larger decrease considered more important. The average moving speed, time spent moving, kurtosis, and autocorrelation are the most important features by this metric (S5 Fig).

Myoblasts display distinct motility states robust to perturbation

Our long-term goal is to understand the cell as a unit in a dynamical system, which entails not just identifying states, but also determining how various inputs from the extracellular environment may drive transitions between states. Mouse myogenic progenitor cells provide a well-characterized system in which chemical signals alter the behaviors of a mechanically-active cell type [71–74], and dynamic activation and lineage commitment processes can be manipulated [75, 76]. In light of this long-term goal, we applied Heteromotility analysis to the myogenic system to ask how myogenic cells occupy state space.

Myoblasts are the transit amplifying progenitor of the skeletal muscle, produced as the daughters of muscle stem cells [75, 76]. Myoblast motility has direct functional relevance, as muscle progenitors translocate along the muscle fiber to sites of injury during muscle regeneration [73, 77] and transverse fibers during development [78]. Primary myoblasts were time-lapse imaged for 8 hours, with and without stimulation by the growth factor FGF2. FGF2 is a known mitogen, inhibitor of differentiation, and possible chemoattractant in myogenic cells in culture [79–81]. *In vivo*, FGF2 is released during muscle injury [82, 83], promoting expansion of the myogenic progenitor pool and possibly migration to sites of injury. As FGF2 is known to elicit different functional behaviors in myogenic cells, introducing this perturbation allows us to evaluate the robustness of myoblast motility states under multiple growth conditions.

Applying hierarchical clustering, two motility clusters are detected (Fig 3A). This partitioning scheme has a positive Silhouette value [63], significantly different cluster means by MANOVA, and optimizes cluster validation metrics (S1 Table). Cluster 1 is a less motile state characterized by lower average moving speed, distance traveled, and a higher kurtosis. Conversely, Cluster 2 is characterized by a higher total distance, average speed, mean-squared displacement (MSD), and proportion of time spent moving. These characteristics indicate that myoblasts occupy a two state system with a consistently motile state (Cluster 2), and a less
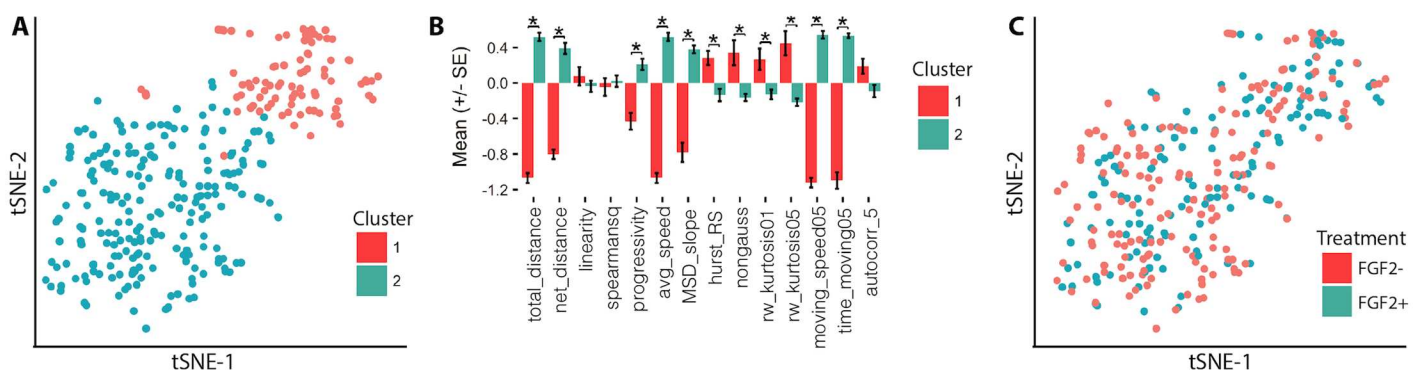
Fig 3. Myoblasts display distinct motility states, shared by both FGF2+ and FGF- conditions. (A) t-SNE visualization of hierarchical clusters in myoblast motility space (perplexity = 30) (MANOVA, $p < 0.001$; Silhouette $S_i = 0.31$). (B) Normalized feature means for motility state clusters (*: $p < 0.05$, Holm-Bonferroni corrected t-test). (C) t-SNE visualization of FGF2 treated (blue) and untreated (red) myoblasts in motility space. $n \approx 150$ cells per condition, taken from two separate animals, with total $n = 308$ cells.

https://doi.org/10.1371/journal.pcbi.1005927.g003

motile state that exhibits more Levy-flight like displacement behavior (Cluster 1). Observed in videos, cells in Cluster 1 are less motile, and display less directed motion than cells in Cluster 2 (S3 and S4 Videos). Clustergram visualizations of the myoblast clusters using multiple linkages are provided (S6 Fig).

Both FGF2 stimulated (FGF2+) and unstimulated (FGF2-) cells co-occupy both states, with no notable preference in state space induced by either condition (Fig 3C). This is confirmed quantitatively as a lack of preferential cluster occupancy between FGF2 treated and untreated cells (χ^2 test $p > 0.05$ of *FGF2-treatment × cluster* contingency table). This negative result suggests that FGF2 does not induce a notable effect on myoblast motility under these conditions.

Muscle stem cells display motility states reflecting activation

Until this point, the cells analyzed have not been undergoing any dramatic phenotypic transitions. In contrast, stem cells are specifically designed to undergo dynamic activation and differentiation processes that radically reshape cell phenotype on an hours to days long timescale. Would it be possible to observe such phenotypic transitions through the lens of cell behavior changes? We apply Heteromotility analysis to the muscle stem cell (MuSC) system during activation from quiescence in an attempt to observe dynamic transitions in cell behavior. MuSCs undergo an exit from quiescence in cell culture conditions, entering an activated state over a roughly 48 hour window, providing a model of a dynamic process where behavior state transitions would be expected [75, 84]. Heterogeneity within the MuSC population is well appreciated [11], suggesting that motility behavior may also be heterogenous during activation. Motility behavior is also relevant to MuSC function due to the physiological motility behavior of muscle progenitors during regeneration, as noted above.

Primary MuSCs were isolated from limb muscles by FACS (PI⁻/CD31⁻/CD45⁻/Sca1⁻/ VCAM⁺/α7-integrin⁺) [85] and seeded on sarcoma-derived ECM coated well plates. After 24 hours in culture, MuSCs were timelapse imaged for 8 hours in DIC. At this stage, MuSCs have begun to activate (*MyoD⁺*), but are not yet committed to differentiation (*MyoG⁺*) [84]. Visualizing hierarchical clusters (Ward's method) of MuSC motility features with t-SNE, it is apparent that multiple motility subpopulations are present (Fig 4A). We identify three distinct clusters with notably different phenotypes. This partitioning scheme has a positive Silhouette value [63], significantly different cluster means by MANOVA, and optimizes cluster validation metrics, as above (S1 Table)(see Supp. Methods for complete discussion). As in myoblasts, the clusters appear to separate based on differences in total distance, average speed, and time moving. Additionally, clusters segregate based on the linearity and progressivity of motion.

Cluster 1 is characterized by the lowest total distance, average speed, time moving, progressivity, and MSD. Observed visually, cells in Cluster 1 are immotile and appear morphologically rounded, lacking any filopodia characteristic of myogenic activation and motility (S5 Video). Clusters 2 and 3 are characterized by increasing measures of total distance, average speed, and time spent moving. Cluster 3 exhibits the highest kurtosis and non-Gaussian parameter, suggesting a jumping, Levy-flight like state. Cluster 2 exhibits the highest time moving, moving speed, progressivity, linearity, net distance, and MSD, indicating a more directed motility state. Visually, Cluster 2 cells largely move in a single direction, while Cluster 3 cells more often turn back and retrace a portion of their previous path (S6 Video)(Fig 4D)($p < 0.001$ by ANOVA for all features displayed).

MuSCs incubated with FGF2 display a similar distribution in motility space compared to untreated cells (S7 Fig). Quantitatively, FGF2 MuSCs display no motility state preference relative to untreated cells (S7 Fig)($p > 0.05$, χ^2 test of *FGF2-treatment x cluster* contingency table). On a population level, FGF2 does not appear to significantly alter any of the motility features
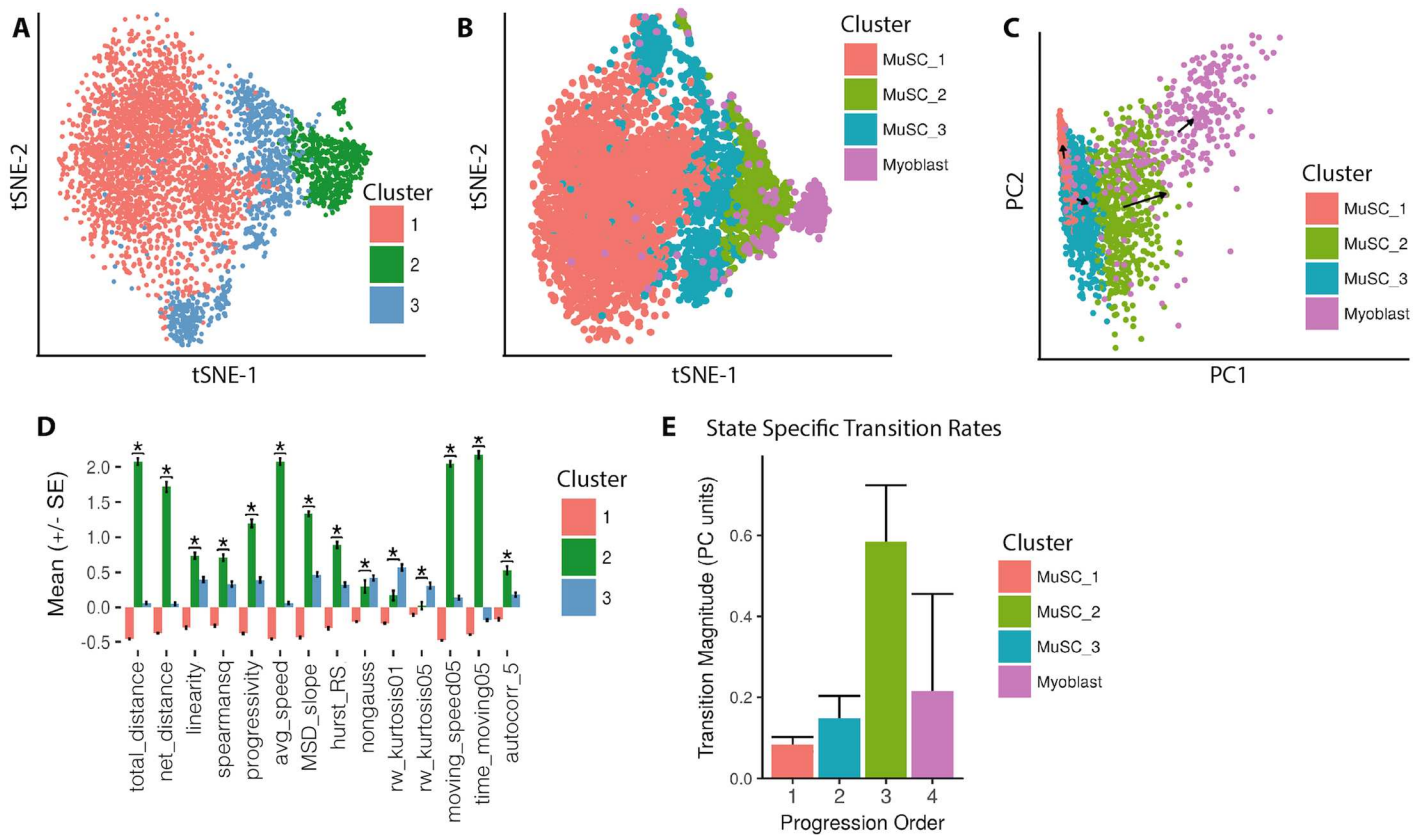
Fig 4. MuSC motility states reflect progressive myogenic activation. (A) t-SNE visualization of distinct motility states, as detected by hierarchical clustering (colors) (MANOVA, $p < 0.001$; Silhouette $S_i = 0.197$)(perplexity = 70). (B) t-SNE visualization of MuSC motility states and myoblasts in a shared space suggesting progressive myogenic motility states. (C) Average transition rate vectors for MuSC motility states (arrows, scaled for presentation) demonstrating transitions toward the myoblast motility phenotype over time in a shared PCA space. (D) Subset of normalized feature means for each motility state (*: $p < 0.05$, Holm-Bonferroni corrected ANOVA). (E) The magnitude (+/- SEM) of the mean transition vector for all cells in a given state. All data represent analysis of $n > 1800$ cells per condition pooled from three animals, with total $n = 4316$ cells.

https://doi.org/10.1371/journal.pcbi.1005927.g004

calculated by our software. This negative result suggests that FGF2 may not play a role in regulating motility in these conditions at this stage of myogenic activation.

Visualized in a shared t-SNE space with myoblasts, the order of MuSC motility states suggests a progressive motion toward the myoblast state space as total distance, time moving, and speed metrics increase (Fig 4B). This is confirmed in a linear state space using principal component analysis (PCA), in which MuSC states and myoblasts segregate primarily along the first principal component (PC1)(Fig 4C). The top 20 features contributing to PC1 in this shared space are metrics of moving speed and time spent moving, suggesting that the features which vary most between quiescent MuSCs and activated myoblasts are related to the speed and frequency of movement. These quantitative results are in line with qualitative observations, in which quiescent cells begin in a relatively immotile state and gradually increase the speed of their motility as they activate in culture. These phenotypic changes also align well with the *in vivo* role of activating MuSCs. Upon injury, MuSCs must activate from quiescence and translocate along the fiber to the site of tissue damage, necessitating more motile behavior [72, 77]. This result indicates that MuSC motility states reflect progressive states of activation toward the eventual myoblast motility state space.

We explored this possibility with pseudotime analysis, which attempts to reconstruct a dynamic cellular process from high-dimensional single cell data under the assumption that the process is ergodic and cells move in a continuous manner through feature space over time [18, 86]. Analytically, this is most commonly achieved by fitting a minimum spanning tree through the data in reduced dimensional space, interpreting the tree's longest axis as a temporal axis. Pseudotime analysis in the MuSC motility state space reflects the view that MuSC states are ordered in a progressive series (S7 Fig).

To further confirm that the MuSC motility states reflect states of activation, we quantified state transition rates and directions for cells in each MuSC state and myoblasts. Cell paths were divided into three equal length tracks ($\tau = 20$ frames) and Heteromotility features were extracted for each of these subpaths. The state of a cell during each time interval τ was defined in two dimensions as the cell's location along the first two principal components of a shared MuSC and myoblast PCA space. Transitions for each cell were calculated as the vector between sequential 2D state locations. The mean transition vector for a given cluster was calculated as the mean of all transitions made by all cells assigned to that cluster.

Visualizing transition rates as vectors originating from each cluster centroid in PCA space, it is evident that cells in each MuSC state progress toward the next state in the sequence over time (Fig 4C). The myoblast motility states are present at the end of the sequence, suggesting that MuSC states represent different stages of myogenic activation. Notably, the transition rates increase as a function of state progression, suggesting the kinetics of activation are non-linear (Fig 4F). In this way, we provide quantitative measurements of transitions between states of myogenic activation in single cells for the first time.

MuSC motility states are more dynamic than MEF and myoblast motility states

We next sought to investigate the dynamics of our motility state systems. A key benefit of using behavior features to define states is that a single cell can be subjected to a state assay at multiple time points, allowing state transitions to be detected. How long does a cell reside in a particular state? Can every state transition to every other state, or are transitions restricted to exist between particular states, as would be the case in a state automaton? The answers to such questions would help clarify the computational logic underlying cell behavior.

To visualize and quantify the dynamics of the motility state systems, we applied coarse-grained probability flux analysis (cgPFA), as implemented by Battle *et. al.* [87] (Supp. Methods). In this method, the first (PC1) and second (PC2) principal components are segmented into *k* bins, and we define each unique combination of bins on PC1 and PC2 as a unique state. To define cell state at multiple points using motility features, cell paths are segmented into τ length subpaths and motility features are extracted from each subpath. Cell state is defined for each subpath based on its position in coarse-grained PCA space, where each binned coordinate is treated as a state. Cell states are compared from one subpath to the next to quantify the dynamics of the motility state system (see Supp. Methods).

As a validation of our implementation, we performed cgPFA on simulated cell paths that varied their model of motility on a defined interval ($\tau = 20$ time points) and compared them to invariant simulations, using subpaths of the same length as the variable model's states ($\tau = 20$ time points). For each bin, a state transition rate is calculated as the vector mean of transitions from that state bin into neighboring states in the von Neumann neighborhood. These transition rates are visualized as arrows atop each state bin. Arrow direction represents the direction of transition rate and arrow length represents the rate magnitude. State bins with high transition consensus therefore display longer arrows, indicating that most cells transition in the
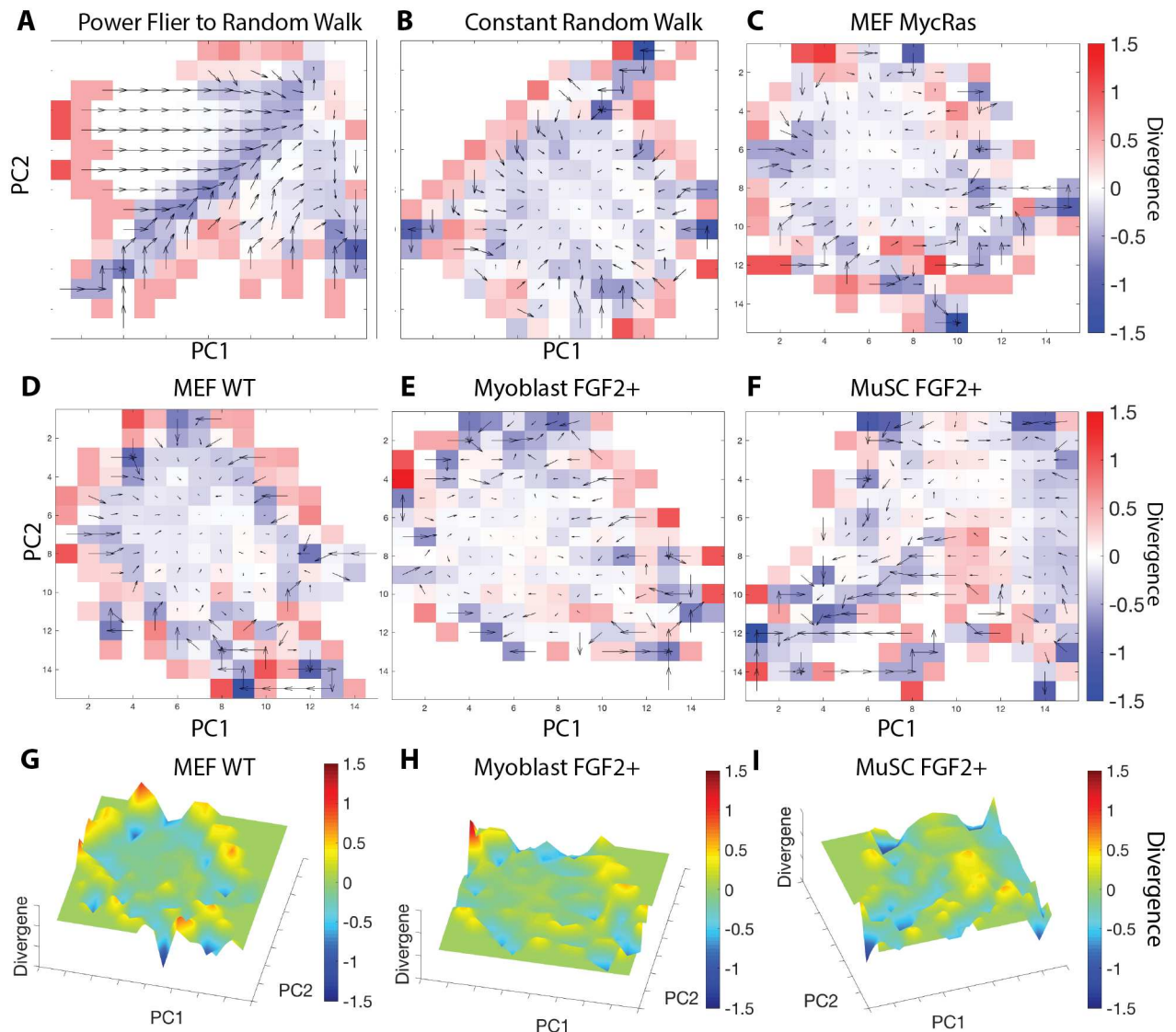
Fig 5. Coarse-grained probability flux analysis (cgPFA). cgPFA of (A) Levy flier to random walk simulations and (B) Random walk invariant simulations. cgPFA of (C) MycRas MEF, (D) WT MEF, (E) Myoblast FGF2+, and (F) MuSC FGF2+ motility states with subpaths of length $\tau = 20$ time points (130 minutes). Each unique combination of bins between PC1 and PC2 is considered as a unique state. Arrows represent transition rate vectors, calculated for each state bin as the vector mean of transitions into the neighboring states in the von Neumann neighborhood. Arrow direction represents the direction of these transition rate vectors, and arrow length represents transition rate vector magnitude. Underlying colors represent the vector divergence from that state as a metric of state stability. Positive divergence indicates cells are more likely to leave a state, while negative divergence indicates cells are more likely to enter a state. 3D representations of (G) MEF WT, (H) Myoblast FGF2+, and (I) MuSC FGF2 + motility state divergence.

same direction. As a measure of state stability, the divergence of the vector field is displayed as a heatmap. States with negative divergence may be considered metastable, with more cells entering than exiting. A Levy flight model transitioning to a random walk displays a highly directed set of state transitions (Fig 5A), as compared to a simple random walk which displays minimal directionality (Fig 5B).

Applying cgPFA to our biological systems with subpaths of length $\tau = 20$ time points (130 minutes), both WT MEF and MycRas MEF systems display no obvious state flux (Fig 5C and 5D; S8 Fig). Topographically, MEF systems display a 'basin' of metastable states. This region

has near zero divergence and low transition rates. States on the outer periphery of this metastable region have higher divergence and transition rates, indicating that these states are less stable (Fig 5C and 5D; S8 Fig). Visualizing state space divergence in three dimensions, the metastable states appear as a central valley, while the unstable states appear as peaks (Fig 5G). Myoblast motility state systems appear similar to the MEF systems by simple observation. The topology appears to be dominated by a metastable basin at the center, surrounded by unstable states on the edge of this basin. Topology is comparable between FGF2 treated and untreated conditions (Fig 5E, S8 Fig).

MuSC motility state systems appear more dynamic than MEF and myoblast systems. This transition directionality is apparent when viewing the transition vector fields, with many transition rates leading toward a metastable 'valley' and a metastable basin on the edge of state space (Fig 5F, S8 Fig). Visualized in three dimensions, these metastable regions are clearly visible as valleys in state space bordered by unstable 'ridges'. We apply these cgPFA analyses with multiple temporal and course-grained resolutions, and find that the results are qualitatively similar for multiple state definition time scales and course-grained binning schemes (S9 Fig).

To determine how long MuSCs occupy a given state, we characterized the dwell times of each occupied state in the course-grained PCA space by exponential decay curve fitting, providing a time constant for each state. In these course-grained state spaces, dwell times appear exponentially distributed with time constants ranging from $\tau_{TC} \approx 1$ to $\tau_{TC} \approx 3$ (S10 Fig). The mean dwell times for state bins range from $\tau_{dwell} = 1$ to $\tau_{dwell} \approx 2.2$, indicating that most cells transition at least once during the time course. Dwell times appear exponentially distributed, suggesting that the state transition process is memoryless on the timescales we observe. Time constants are positively correlated with the number of cells occupying a state, supporting the notion that states of high occupancy are metastable and therefore characterized by longer dwell times (S10 Fig). Topology, dwell times, and transition directionality are comparable between FGF2 treated and untreated conditions (Fig 5I, S8 and S10 Figs). These results suggest that the MuSC motility state system is more dynamic than the MEF or myoblast systems, consistent with the dynamic MuSC activation process.

MuSC motility states break detailed balance

A system with a discrete set of states at equilibrium should obey the law of detailed balance, such that each individual state transition $A \rightarrow B$ occurs at the same rate as the reverse transition $B \rightarrow A$. Systems that break detailed balance transition between states in a directed manner, such that future behavior can be predicted by current state. Biological systems frequently break equilibrium when undergoing directed processes, but confirming that detailed balance is broken in a given scenario can prove challenging [87]. A system in detailed balance would be expected to exhibit equal and opposite transition vectors with no observable pattern of transitions. In our MuSC systems, visualizing transition rates by PFA suggests that detailed balance may be broken, as shown by the directed nature of transition rate vectors (Fig 5, S8 Fig).

To confirm statistically that detailed balance is broken in these systems, we defined N-dimensional state spaces based on the first N coarse-grained principal components and performed PFA in these spaces (ND-cgPFA)(see Methods). A $2N$-dimensional matrix is generated representing every possible set of state transitions in a given N-dimensional state space. For example, a 1 dimensional state space is represented as a 2 dimensional matrix, each dimension coarse-grained into k bins (rows, columns). One dimension of this 2D space represents a cell's initial state position at one time point, and the other represents the destination state position at the next time point (as in Fig 6A and 6B). This pattern is repeated for the construction of
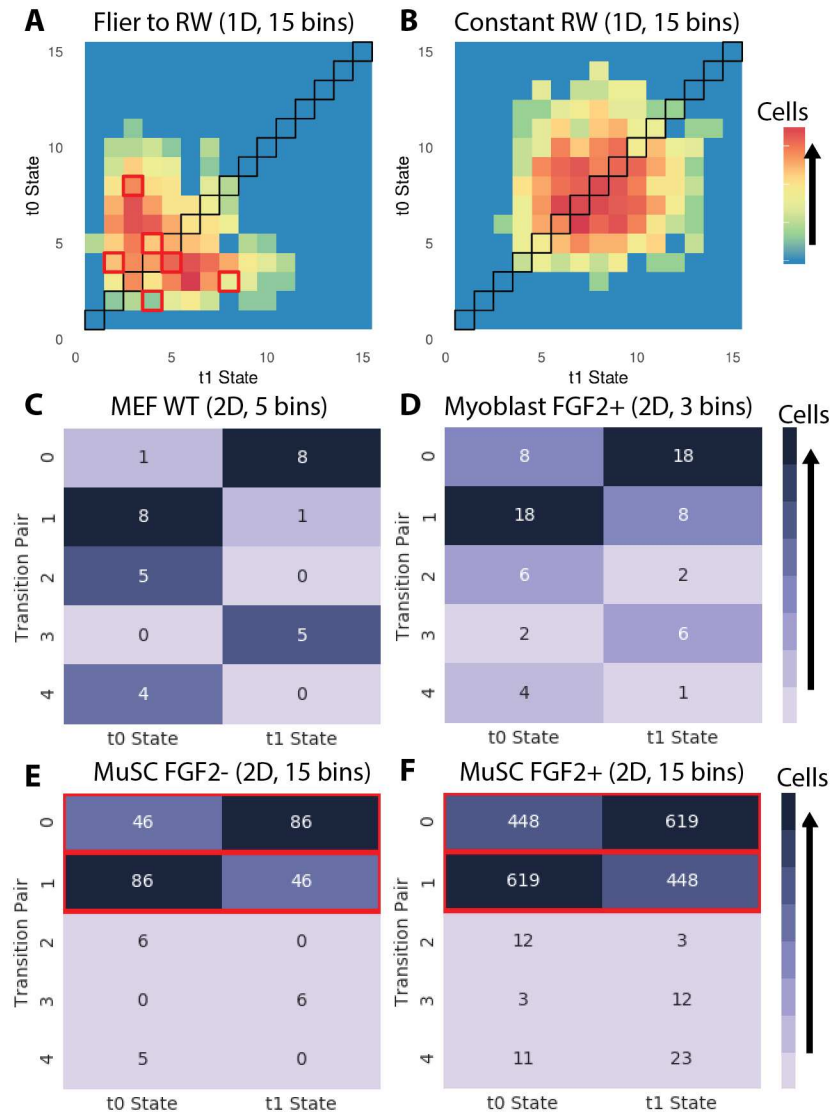
Fig 6. Analysis of state transition dynamics idicates MuSC motility states break detailed balance. One-dimensional coarse-grained PFA of (A) Simulated Levy flier transition to a random walk, (B) simulated random walk. Transition pairs from N-dimensional cgPFA displayed as heatmaps for (C) WT MEFs ($n = 312$), (D) Myoblasts (FGF2+) ($n = 150$), (E) MuSCs (FGF2-) ($n = 1838$), and (F) MuSCs (FGF2+) ($n = 2500$). Heatmaps show the five most unbalanced transitions in a system, with colors and numerical insets indicating the number of cells that transitioned from a given state at t_0 to a given state at t_1. Significantly unbalanced transitions are outlined in red ($p < 0.05$, Benjamini-Hochberg corrected binomial test). A system in detailed balance would display no unbalanced transitions.

https://doi.org/10.1371/journal.pcbi.1005927.g006

higher dimensional spaces. For instance, a 2D state space is represented as a 4D matrix, where the *first two* dimensions describe initial state, and the *next two* dimensions describe final state. The number of cells that exhibit each transition are recorded as the value of the corresponding position in the matrix (see Methods for further description).

A system in detailed balance would be expected to display symmetry for each pairwise set of transitions. To determine if each of our systems was in detailed balance, we performed ND-cgPFA at several levels of dimensional ($N \in \{1, 2, 3, 4\}$) and coarse-grained (bins $k \in \{2, 3, 5, 7, 10, 15, 20\}$) resolution. At each level of resolution, we test each of the k^{2N} possible pairwise transitions for balance by the binomial test (H_0: $p = 0.50$).

To validate this method, we performed cgPFA on simulated cell paths generated using variable and invariant models of motion. A Levy flier that transitions to a random walk shows clearly unbalanced pairwise transitions (Fig 6A) at one dimension of resolution, while an invariant random walk displays symmetry about the diagonal and balanced pairwise transitions (Fig 6B). A simple binomial test for pairwise transitions finds multiple unbalanced transition pairs in the variant model, but not the invariant model (Fig 6A and 6B).

N-dimensional matrices cannot be visualized in their totality as in the 1D case. As noted above, we tested all k^{2N} possible transitions for balance at multiple dimensional and binning resolutions for each system. Here, the five transition pairs that are most unbalanced out of the possible set of k^{2N} transitions (by *p* value of the binomial test) in a given system and state space are presented as a 5-by-2 heatmap, where columns represent the initial and final states in a transition pair. We present the dimensional and coarse-grained resolution revealing the most asymmetry for each system, as noted above each heatmap. Visualizing ND-cgPFA pairwise transitions for MEF (Fig 6C), myoblast (Fig 6D), and MuSC state systems (Fig 6E and 6F) demonstrates that transitions with some degree of unbalance exist in each of the systems. By the binomial test (Benjamini-Hochberg corrected), only the MuSC system (FGF2+ and FGF2-) displays significantly unbalanced transitions in any of the course-graining schemes tested, confirming that detailed balance is broken. Our ND-PFA test is biased toward Type II error (false negatives) rather than Type I error (false positives) for multiple reasons, providing further confidence that the detailed balance breaking we identify is valid. It is therefore possible that detailed balance is also broken in the MEF and myoblast systems and our tools are simply not sufficient to detect this asymmetry.

We performed this ND-cgPFA analysis for multiple values of the timescale parameter τ and find that these results are consistent. At each value of τ, the MuSC system demonstrates broken detailed balance and the MEF and myoblast systems do not. On a very short timescale, it is unlikely that detailed balance is broken for a motility state system, even in a dynamic system like activating MuSCs. In line with this biological prior, we find that detailed balance is *not* broken in the MuSC system when we perform ND-cgPFA with the same number of $\tau = 20$ length windows, but set them to overlap with a stride $s = 1$. In this overlapping scheme, only two time steps of difference are present between the first and final temporal window, accounting for only 13 minutes in real time in our experiments (S11 Fig).

As additional confirmation of MuSC motility state dynamism, we performed this symmetry breaking analysis using hierarchical clustering to define cell state over the whole motility feature space, rather than coarse-grained location along the principal components (S12 Fig). As with ND-cgPFA, our positive control variable model breaks detailed balance by the binomial test, while our negative control invariant models do not. MuSC systems again demonstrate greater pairwise asymmetry than MEF or myoblast systems in this test (S12 Fig).

An important question is raised when detailed balance is broken. Is the system stationary, with the same number of cells in each state over time, or non-stationary? We evaluated the stationary nature of the MuSC systems by forming a contingency table comparing state occupancy between time points for each set of length $\tau = 20$ subpaths in the dataset and find a significant difference ($p < 0.05$, χ^2 test) for each scale where detailed balance breaking is detected. These results collectively demonstrate the more dynamic nature of the MuSC system and demonstrate that the MuSC motility state system is non-stationary and breaks detailed balance.

Discussion

Cell behavior phenotypes represent an under-exploited opportunity to explore the heterogeneity of cell systems. In contrast to existing methods based on destructive molecular assays, cell

behaviors such as motility can be tracked in a single cell over time, allowing for measurement of phenotypic state transitions. Previous work to quanitfy morphological dynamics has allowed for quantification of cell state transitions in this manner [20, 24, 31]. Quantification of motility behaviors in addition to morphological dynamics may allow for detection of transitions that are not revealed by morphological dynamics alone.

Our Heteromotility analysis software provides access to one portion of this rich motility feature space for analysis. Applying analysis techniques common in other single cell assays, we demonstrate that Heteromotility features are sufficient to distinguish different motility phenotypes and provide novel insight in multiple biological contexts. In addition to detecting heterogeneity within populations, Heteromotility analysis was also useful to quantitatively describe perturbations on the population level.

In our three biological systems, overlapping but characteristic motility states are present in a continuous state space. Notably, a shared set of motility states is conserved between WT and MycRas MEFs, despite the dramatic perturbation of neoplastic transformation. A similar phenomenon of conserved phenotypic states has been described for cell shapes. Despite a large number of perturbations, multiple cell systems displayed a fairly limited set of cell morphology states, with perturbation merely altering the prevalence of these states [4, 88]. These results suggest that cell systems may have a discrete set of phenotypic states despite a much greater diversity in molecular organization, with perturbations acting largely to alter the distribution of cells within these states, rather than elicit novel behavior.

Although these state definitions are robust to perturbation in our conditions, the distribution of cells among these states appears to be dynamic. This is demonstrated by the state preferences of WT and MycRas transformed MEFs. Considering the robustness of motility state definitions, state transitions may act as a mechanism of population level phenotype change. This is not necessarily in opposition to a model of motility regulation in which cell phenotypes shift within a given state. Harkening to the subsumption architectures of robotic control systems [89], a higher level state determinant, such as oncogene expression, may be viewed as inducing a preference for the selection of behavioral states, and substates within those states, by a more direct effector mechanism, in this case the cell motility machinery. The mechanisms of state transitions and substate preference may work in synchrony to specify population level phenotypes.

Within the MEF systems transformed cells preferentially occupy the more motile state, suggesting that neoplastic transformation shifts the distribution of cells within the motility state system. This state preference is strong enough to allow for better-than-chance machine learning classification of WT and MycRas MEFs based solely on motility features. Training machine learning classifiers on quantitative motility features and ground truth patient outcomes may potentially serve as an additional cancer diagnostic metric.

Most single cell analysis methods currently used to investigate heterogeneity are not capable of assaying a single cell across a time course, but rather provide a detailed snapshot of a cell at a single time point. Real-time observation of individual cells may elucidate novel properties of a heterogenous cell system, as pioneered by analysis of morphological dynamics. Here, we combine the real-time nature of cell behavior observations with Heteromotility analysis to investigate the dynamics of motility states.

In the context of MuSCs undergoing myogenic activation, we find a set of motility behavior states progressively more similar to myoblast motility behavior. Quantifying state transitions within each of these states, we observe that cells within each state transition toward the next state in the series over time. State transition rates increase as a function of state progression, suggesting that MuSC activation is not simply a linear process. This observation may have implications for the study of MuSC heterogeneity. If the rate of phenotypic change in MuSCs

is non-linear during activation, then heterogeneity in activation kinetics between cells will be exaggerated during a 'critical period,' where the most rapidly activating cells are not only more activated at that moment, but are moving toward a fully activated state more rapidly than their less activated counterparts. This is to our knowledge the first quantification of single cell transition rates between quiescent MuSC and myoblast phenotypes during myogenic activation. It follows that similar analysis of cell behavior may elucidate transition dynamics in other contexts of cell biology.

A state system at equilibrium displays detailed balance, in which pairwise transitions between all states are equal. We check for the presence of detailed balance in our motility state systems as a method of quantitatively discerning if cells are transitioning through state space in a predictable manner. The MuSC motility state system breaks detailed balance, while the MEF and myoblast systems do not. This may reflect underlying properties of each system, as MuSCs are undergoing a dynamic activation process on the timescale of imaging, while MEFs and myoblasts are not. Broken detailed balance in the MuSC system indicates that the current MuSC state provides predictive information about the cell's future state, as cells transit state space in a predictable way. This supports our observation of progressive activation states discussed above. Analysis of detailed balance breaking by real-time cell state inference from cell behavior may be a useful approach to detect dynamic biological processes and predictable phenotypic patterns.

Moving forward, we believe that single-cell, quantitative motility analysis has multiple applications. Quantitative motility features may allow discimination of cell states by supervised classification, as demonstrated by our discrimination of neoplastic and wild-type MEFs. Supervised classification in this manner may be useful in diagnostic and biosensor applications. For example, motility metrics could be added to the cell morphology metrics commonly used to diagnose neoplasia from patient biopsy samples.

Heterogeneous cellular states within a population of cells may be detected by unsupervised clustering of motility features, similar to analyses of population heterogeneity performed using single-cell molecular assays. We demonstrate this use case by identifying heterogenous motility states in MuSCs, which correspond to heterogeneous states of stem cell activation. Unsupervised identification of heterogeneous populations is useful in understanding the behavior of cell systems, as demonstrated by findings from single-cell molecular assays. Identifying heterogeneous motility behaviors may be of particular interest in contexts where motility is directly tied to cell and tissue function, such as tissue regeneration and immunological responses.

Analysis of cell behaviors such as motility also allows for measurement of the transition rates between cell phenotypes. We demonstrate this application by measuring activation rates in the myogenic system for the first time, and identifying the presence or absence of detailed balance across biological systems. Cell state transition analysis may be applied to measure the rate of dynamic phenotypic changes, such as stem cell differentiation, immunological activation, or neoplastic transformation. Each of these applications of quantitative motility analysis may be paired with exiting approaches for quantiative analysis of morphology to provide a more complete picture of cell behavior than either approach may yield alone.

Conclusion

Our Heteromotility software provides a means of defining cell states and quantifying transitions between them by quantitative analysis of one aspect of cell behavior. We demonstrate motility behavior analysis is capable of identifying unique motility states in simulations and three biological systems. In our biological systems, these states appear to have robust definitions when perturbed, but each cell's state identify appears dynamic. Real-time, quantitative

assays of single cells such as Heteromotility analysis may reveal the dynamics of heterogenous cell systems, which cannot be accomplished by terminal molecular assays. We demonstrate this approach by showing that MuSCs transition through a progressive set of motility states toward the myoblast motility state in a predictable manner, breaking detailed balance.

Materials and methods

Ethics statement

All mice were housed at the University of California San Francisco following UCSF Institutional Use and Care of Animals Committee guidelines. Adult male C57Bl/6 mice (2-4 m.o.) were used for myogenic cell experiments. Adult female C57Bl/6 mice (2-3 m.o.) were used as mothers to derive MEFs from E13.5 embryos [90].

Cell isolation and culture

Mononuclear cells were isolated from limb muscles of adult mice as described [91]. Myoblasts were isolated by negative-selection against *CD31*, *CD45*, and *Sca1* using the EasySep Endothelial Selection kit and subsequent 5 minute preplating. Myoblasts were maintained in myogenic growth media (F10, 20% FBS, and [5 ng/mL] FGF2) on sarcoma-derived ECM coated dishes. All myoblasts used in experiments were passage 3-4. Wild-type muscle stem cells were isolated by FACS (PI-/CD31-/CD45-/Sca1-/VCAM+/α7-integrin+) and cultured as described [85]. Wild-type MEFs were isolated from E13.5 embryos as described [90]. Transformed MEFs were generated as described [61], generously donated by the authors. All MEFs were maintained in DMEM, 10% FBS, 1% Penicillin / Streptomycin.

Timelapse motility imaging

All images were performed in a temperature and CO2 controlled unit at 37°C and 5% CO_2. All experiments were performed with a 20X air objective, capturing images with a Hamamatsu C11440-22CU camera (pixel size 6.5 x 6.5 μm) using a Nikon Ti with automated XY stage. MuSCs were seeded at 500-1000 cells/well in 20 wells of an optically clear tissue culture treated, plastic bottomed 96 well plate coated with sarcoma-derived ECM (Sigma) immediately after FACS sorting. After 23 hours to allow for adaptation to culture, media was exchanged for the relevant experimental medium and MuSCs were incubated for 1 hour to adapt. MuSCs were imaged at 20X in DIC for 10 hours with a temporal resolution of 6.5 minutes / frame. Myoblasts were passaged and plated 24 hours prior to imaging at 300-500 cells/well. Media exchanges were performed 1 hr prior to imaging, in the same manner as MuSCs. Myoblast imaging was performed in the same manner as MuSCs. MEFs were similarly plated 16-20 hours prior to imaging at the same density and imaged in the same manner. For FGF2 perturbation experiments, 10 wells of an optically clear plastic bottomed 96 well plate were imaged in growth media with [5 ng/mL] FGF2, and the remaining 10 were imaged in growth media with [0 ng/mL] FGF2. Additional details are available in the Supplemental Methods.

Heteromotility analysis

DIC images were segmented using custom segmentation algorithms, optimized for each of the systems we investigated. Tracking was performed with a modified version of uTrack [92]. Code is available on the Heteromotility GitHub page. All code for the Heteromotility tool is available on the Heteromotility Github page (https://github.com/cellgeometry/heteromotility) and in the Python Package Index (PyPI). Detailed descriptions of the algorithms are provided in the Supplemental Methods.

Supporting information

S1 Text. Supplemental materials and methods.
(PDF)

S1 Fig. Heteromotility software description. (A) Heteromotility workflow diagram and (B) a table of the complete feature set.
(TIF)

S2 Fig. Clustering and visualization of simulated motion models with varied parameters.
(A) Simulated models of motion are segregated by unsupervised heirarchical clustering, as displayed in representative a heatmap of hierarchically clustered (Ward's linkage) simulated motion paths (1000 members/class, length 100 time units) based on Heteromotility features. Color labels on the left mark a sample's True Class. Effective separation of the True Classes indicates effective detection of different phenotypes by unsupervised clustering. Two-dimensional (B) PCA and (C) ICA visualization of the simulated motion paths to provide an intuition for the linearity of motility state space and performance of traditional linear dimensionality reduction techniques. Only high (30) dimensional PCA spaces are used for analysis. ICA is not used for any downstream analysis. (D) Representative t-SNE visualizations of simulated motion models with different sample sizes and track lengths, labeled with ground truth classes. Models occupy distinct regions of state space under all sample size and track length variations. (E) Representative t-SNE visualizations of simulated motion model groups with the underlying parameters for each motion model varied. Parameters for each condition shown are displayed above the t-SNE map. (F) Unsupervised clustering accuracy (Ward's linkage) as a function of parameter variations to the underlying simulations. Performance decreases as expected when parameters are set in a manner that decreases the distinctness of the models. For example, performance is lower when the bias parameter for biased random walks is set to a low value, close to an unbiased random walk, or when the fractal Brownian motion index is set to the same index displayed by a random walker (H = 0.5). Performance is high across other conditions tested.
(TIF)

S3 Fig. Comparison of variance dimensionality and local cell density relationships between cellular systems. (A) Cumulative variance explained for each dimensionality of principal component space across MuSC, MEF, and Myoblast systems. (B) Strength of relationships between our Local Cell Density Index and each of the Heteromotility features, displayed as overlapping histograms of Pearson's r^2 values for linear regression models fit between the Local Cell Density Index and each feature pairwise. No features in any system display a relationship with a meaningful effect size (max $r^2 \approx 0.03$).
(TIF)

S4 Fig. Round Robin analysis of MEF motility state distribution and classification performance. (A) Representative t-SNE visualizations, MycRas and wild-type cluster distrobutions, and selected cluster feature values for three "splits" from a Round Robin analysis. In the Round Robin analysis, analyses were performed on 4 experiments (1 MycRas, 3 wild-type) leaving 2 experiments out in an iterative fashion for each possible combination of experiments. MEF cluster distributions and properties are reproducible across all combinations. (B) Aggregate confusion matrix for SVM classifiers trained in a Round Robin fashion, such that 4 experiments were used for training and 2 for evaluation for all combinations of 4 experiments. SVMs classify MycRas and wild-type cells with 70% accuracy and do not demonstrate a prediction bias for one class over the other. (C) Distribution of Round Robin Classification Accuracies for

SVMs trained on each Round Robin split. (D) Top 10 most important features for Round Robin Classification Accuracy using SVM classifiers. Features importance is determined as the decrease in classification accuracy when an SVM is retrained without a given feature as input. The Top 10 features were selected based on the mean decrease in Round Robin Classification Accuracy across Round Robin splits. (E) t-SNE visualization of all MEF cells analyzed, labeled with their experimental origin.
(TIF)

S5 Fig. MEF analysis with reduced feature sets. (A) Round Robin classification accuracy is significantly, positively correlated to the proportion of features utilized. Points represent mean Round Robin Classification Accuracy for a given parameter set of N% of features and a value for the SVM bias parameter C. Five bias parameter values were tested in a linear distribution in the range [0.4, 0.6] around the bias parameter $C \approx 0.5$ we found for the optimal SVM by Grid Search. Reduced feature sets were selected using only the top N% of features based on ANOVA F-scores. (B) Cluster distributions and cluster feature values for clustering of MEFs with reduced feature sets. Feature sets were reduced to the top N% based on ANOVA F-scores. Behavioral clusters are still identifiable and MycRas/wild-type dependent state distribution is preserved with a single feature.
(TIF)

S6 Fig. Myoblast clustergram visualizations. Clustergrams of myoblast feature space using several hierarchical clustering linkages. The assigned cluster label for each linkage map is displayed in a color coded column on the left hand side of each heatmap. Ward's linkage was used for downstream analysis.
(TIF)

S7 Fig. MuSCs display multiple motility states, reflecting states of activation. FGF2 does not influence MuSC motility phenotypes, and MuSC motility states reflect progressive states of activation. (A) t-SNE visualization of MuSC motility space with FGF2 treated and untreated color labels. (B) Occupancy of MuSC motility states in FGF2 treated and untreated conditions. (C) Pseudotime analysis displaying a reduced dimensional representation of MuSC and myoblast motility space (DDRT) with a minimum spanning tree plotted to mark the pseudotime axis. Pseudotime analysis attempts to find a temporal axis through an ergodic process observed at a single time point by fitting a minimum spanning tree (MST) to a reduced dimensional representation of multidimensional data. The longest axis of the MST is assumed to represent the temporal axis of the ergodic process. MuSC states are clearly ordered in a progressive sequence, moving toward the myoblast phenotype over pseudotime. (D) Scatterplot of MuSC/Myoblast transition vectors, demonstrating that transition vectors are primarily along the first principal component and slightly skewed in the positive direction.
(TIF)

S8 Fig. Course-grained probability flux analysis of motility state spaces. (A) Three-dimensional representation of the MycRas MEF state divergence surface as measured by cgPFA using tau = 20 and 15 course-grained bins. Course-grained probability flux analysis (cgPFA) of (B) myoblast (FGF2-), and (C) MuSC (FGF2+) motility states with subpaths of length τ = 20 time points (130 minutes) and 15 course-grained bins per dimension. Each unique combination of bins between PC1 and PC2 is considered as a unique state. Arrows represent transition rate vectors, calculated for each state bin as the vector mean of transitions into the neighboring states in the von Neumann neighborhood. Arrow direction represents the direction of these transition rate vectors, and arrow length represents transition rate vector magnitude. Underlying colors represent the vector divergence from that state as a metric of state stability. Positive

divergence indicates cells are more likely to leave a state, while negative divergence indicates cells are more likely to enter a state. (D-I) State occupancy visualizations of the same course-grained PCA presented for cgPFA analysis. The number of cells that occupy a given state for at least one time unit is represented in the third dimension of the landscape and by the heatmap colors.
(TIF)

S9 Fig. Course-grained probability flux analysis of motility state spaces on multiple time scales and binning resolutions. Course-grained PFA analysis as demonstrated in Fig 5 and S8 Fig was performed for all parameter combinations of the temporal window size $tau \in \{20, 25, 30\}$ and binning resolution $k \in \{5, 10, 15, 20, 30\}$ across all cellular systems. Representative visualizations across these parameter ranges are presented. Both (A) MycRas and (B) wild-type MEFs retain the qualitative metastable 'basin' appearance across time scales. As binning resolution decreases below $k = 10$, the structure of the state space is obscured. At higher resolutions of k, more bins little net divergence are present. (C) Myoblast cgPFA state spaces likewise retain a metastable 'basin' appearance across time scales. (D) MuSC state spaces retain a metastable 'valley' surrounded by unstable ridges across time scales. At higher binning resolutions of k, an unstable ridge within the metastable valley becomes more apparent.
(TIF)

S10 Fig. Dwell time analysis of MuSC motility states. MuSC motility state dwell time analysis reveals rapid transitions, longer dwell times in higher occupancy states, and roughly exponentially distributed dwell times. Dwell times vs. total number of observed cells for each occupied state in course-grained PCA space for (A, B) FGF2- MuSCs, (D,E) FGF2+ MuSCs in two-dimensional PCA space, and (G, H) FGF2- MuSCs and (J, K) FGF2+ MuSCs in PCA spaces where detailed balance is broken. Dwell time distributions relative to the binned samples from a fitted exponential distribution for (C) FGF2- MuSCs and (F) FGF2+ MuSCs in two-dimensional PCA space, and (I) FGF2- MuSCs and (L) FGF2+ MuSCs in PCA space where detailed balance is broken.
(TIF)

S11 Fig. *N*-dimensional course-grained probability flux analysis across multiple time scales. ND-cgPFA as presented in Fig 6 was repeated for values of the temporal window size parameter $\tau \in \{20, 25, 30\}$. (A) The results of detailed balance breaking are robust across settings of this time scale parameter. At each time scale, the MuSC system breaks detailed balance, while the MEF and myoblast systems do not. Heatmaps display the five most unbalanced transitions for each defined cgPFA space. *tau*, course-grained bin, and stride parameters are listed above each heat map. (B) To demonstrate that detailed balance is present in the MuSC system on short time scales, we performed ND-cgPFA using the same number of temporal windows size $tau = 20$, but overlapped them with a single unit stride of $s = 1$. In this scheme, each window is only 1 time unit different than it's neighbor, such that only 2 time units of difference are present between the initial and final time window. On this short time scale, MuSC systems do not break detailed balance.
(TIF)

S12 Fig. Probability flux analysis between states defined by hierarchical clustering. Hierarchical clustering based probability flux analysis of (A) a Levy flight simulation transitioning to a random walk, (B) an invariant random walk simulation, (C) myoblasts (FGF2+), (D) myoblasts (FGF2-), (E) MycRas MEFs, (F) WT MEFS, (G) MuSCs (FGF2+), and (H) MuSCs (FGF2-). The matrix displays transitions in state space as values in a matrix. Rows of the matrix

correspond to an initial cell state (t0 state) and columns correspond to a destination state (t1 state). The value of each bin represents the number of times a state transition was observed. Symmetrical bins about the diagonal represent reciprocal pairwise transitions, with one 'forward' transition and one 'reverse' in each pair. The identity line represents "self" or non-transitions. MuSCs show a less balanced distribution than either MEFs or myoblasts by the binomial test for pairwise transition balance. Pairwise transitions breaking detailed balance are outlined in red.
(TIF)

S13 Fig. Resampling analysis of MuSC cluster partitions. (A) Representative random samples of 80% of MuSCs with a 3 cluster partition (Ward's linkage) applied. Note cluster separation along a common axis, robust to resampling. (B) Representative random samples of 80% of MuSCs with a 4 cluster partition (Ward's linkage) applied. Note separation of clusters along multiple axes in some samples, and a single axis in others. The 4 cluster partition is not robust to resampling.
(TIF)

S1 Video. Timelapse of wild-type MEF motility. Numeric labels indicate the motility state of each cell. Color coded circles indicate the positions visited by each cell over time.
(AVI)

S2 Video. Timelapse of MycRas MEF motility. Numeric labels indicate the motility state of each cell. Color coded circles indicate the positions visited by each cell over time.
(AVI)

S3 Video. Timelapse of myoblast motility. Numeric labels indicate the motility state of each cell. Color coded circles indicate the positions visited by each cell over time.
(AVI)

S4 Video. Timelapse of myoblast motility. Numeric labels indicate the motility state of each cell. Color coded circles indicate the positions visited by each cell over time.
(AVI)

S5 Video. Timelapse of MuSC motility. Numeric labels indicate the motility state of each cell. Color coded circles indicate the positions visited by each cell over time.
(AVI)

S6 Video. Timelapse of MuSC motility. Numeric labels indicate the motility state of each cell. Color coded circles indicate the positions visited by each cell over time.
(AVI)

S1 Table. Table of cluster evaluation indices calculated for each hierarchical clustering of each biological system.
(PDF)

S2 Table. Table of parameter values for groups of motion simulations analyzed with varied parameter settings.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Jacob C. Kimmel, Amy Y. Chang, Wallace F. Marshall.

Formal analysis: Jacob C. Kimmel.

Funding acquisition: Andrew S. Brack, Wallace F. Marshall.

Investigation: Jacob C. Kimmel.

Methodology: Jacob C. Kimmel, Amy Y. Chang.

Project administration: Wallace F. Marshall.

Resources: Jacob C. Kimmel, Andrew S. Brack.

Software: Jacob C. Kimmel.

Supervision: Andrew S. Brack, Wallace F. Marshall.

Validation: Jacob C. Kimmel.

Visualization: Jacob C. Kimmel.

Writing – original draft: Jacob C. Kimmel.

Writing – review & editing: Jacob C. Kimmel, Amy Y. Chang, Andrew S. Brack, Wallace F. Marshall.

References

1. Altschuler SJ, Wu LF. Cellular Heterogeneity: Do Differences Make a Difference? Cell. 2010; 141 (4):559–563.

2. Rodgers JT, King KY, Brett JO, Cromie MJ, Charville GW, Maguire KK, et al. mTORC1 controls the adaptive transition of quiescent stem cells from G0 to GAlert. Nature. 2014; 510(7505):393–396. https://doi.org/10.1038/nature13255 PMID: 24870234

3. Wang YH, Israelsen WJ, Lee D, Yu VWC, Jeanson NT, Clish CB, et al. Cell-State-Specific Metabolic Dependency in Hematopoiesis and Leukemogenesis. Cell. 2014; 158(6):1309–1323.

4. Yin Z, Sadok A, Sailem H, McCarthy A, Xia X, Li F, et al. A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes. Nat Cell Biol. 2013; 15(7):860–871. https://doi.org/10.1038/ncb2764 PMID: 23748611

5. Beck AH, Sangoi AR, Leung S, Marinelli RJ, Nielsen TO, van de Vijver MJ, et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. Science translational medicine. 2011; 3(108):1–11. https://doi.org/10.1126/scitranslmed.3002564

6. Tabesh A, Teverovskiy M, Pang HY, Kumar VP, Verbel D, Kotsianti A, et al. Multifeature prostate cancer diagnosis and Gleason grading of histological images. IEEE transactions on medical imaging. 2007; 26(10):1366–1378. https://doi.org/10.1109/TMI.2007.898536 PMID: 17948727

7. Petushi S, Garcia FU, Haber MM, Katsinis C, Tozeren A. Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer. BMC Medical Imaging. 2006; 6 (1):13–11. https://doi.org/10.1186/1471-2342-6-14

8. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, et al. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nature Biotechnology. 2015; 33(2):155–160. https://doi.org/10.1038/nbt.3102 PMID: 25599176

9. Rotem A, Ram O, Shoresh N, Sperling RA, Goren A, Weitz DA, et al. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nature Biotechnology. 2015; 33(11):1165–1172. https://doi.org/10.1038/nbt.3383 PMID: 26458175

10. Copley MR, Beer PA, Eaves CJ. Hematopoietic Stem Cell Heterogeneity Takes Center Stage. Cell Stem Cell. 2012; 10(6):690–697. https://doi.org/10.1016/j.stem.2012.05.006 PMID: 22704509

11. Tierney MT, Sacco A. Satellite Cell Heterogeneity in Skeletal Muscle Homeostasis. Trends Cell Biol. 2016; 26(6):434–444. https://doi.org/10.1016/j.tcb.2016.02.004 PMID: 26948993

12. Goodell MA, Nguyen H, Shroyer N. Somatic stem cell heterogeneity: diversity in the blood, skin and intestinal stem cell compartments. Nat Rev Mol Cell Biol. 2015; 16(5):299–309. https://doi.org/10.1038/nrm3980 PMID: 25907613

13. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. Nature. 2013; 501(7467):338–345. https://doi.org/10.1038/nature12625 PMID: 24048066

14. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. Nature. 2013; 501 (7467):328–337. https://doi.org/10.1038/nature12624 PMID: 24048065

15. Waddington CH. An Introduction to Modern Genetics. 3rd ed. Oxon (England); New York: Routledge; 1939. Available from: http://dx.doi.org/10.4324/9781315665412.

16. Huang S. The molecular and mathematical basis of Waddington's epigenetic landscape: A framework for post-Darwinian biology? Bioessays. 2011; 34(2):149–157. https://doi.org/10.1002/bies.201100031 PMID: 22102361

17. Egolf DA. Equilibrium regained: from nonequilibrium chaos to statistical mechanics. Science. 2000; 287 (5450):101–104. https://doi.org/10.1126/science.287.5450.101 PMID: 10615038

18. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature Biotechnology. 2014; 32(4):381–386. https://doi.org/10.1038/nbt.2859 PMID: 24658644

19. Samusik N, Good Z, Spitzer MH, Davis KL, Nolan GP. Automated mapping of phenotype space with single-cell data. Nature methods. 2016; 13(6):493–496. https://doi.org/10.1038/nmeth.3863 PMID: 27183440

20. Held M, Schmitz MHA, Fischer B, Walter T, Neumann B, Olma MH, et al. CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. Nature methods. 2010; 7(9):747–754. https://doi.org/10.1038/nmeth.1486 PMID: 20693996

21. Halfter W, Liverani D, Vigny M, Monard D. Deposition of extracellular matrix along the pathways of migrating fibroblasts. Cell and tissue research. 1990; 262(3):467–481. https://doi.org/10.1007/BF00305243 PMID: 1706644

22. Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. Nat Rev Mol Cell Biol. 2014; 15(12):786–801. https://doi.org/10.1038/nrm3904 PMID: 25415508

23. Laird DJ, von Andrian UH, Wagers AJ. Stem Cell Trafficking in Tissue Development, Growth, and Disease. Cell. 2008; 132(4):612–630. https://doi.org/10.1016/j.cell.2008.01.041 PMID: 18295579

24. Johnson GR, Buck TE, Sullivan DP, Rohde GK, Murphy RF. Joint modeling of cell and nuclear shape variation. Molecular Biology of the Cell. 2015; 26(22):4046–4056. https://doi.org/10.1091/mbc.E15-06-0370 PMID: 26354424

25. Cooper S, Sadok A, Bousgouni V, Bakal C. Apolar and polar transitions drive the conversion between amoeboid and mesenchymal shapes in melanoma cells. Molecular Biology of the Cell. 2015; 26 (22):4163–4170. https://doi.org/10.1091/mbc.E15-06-0382 PMID: 26310441

26. Lock JG, Mamaghani MJ, Shafqat-Abbasi H, Gong X, Tyrcha J, Strömblad S. Plasticity in the macromolecular-scale causal networks of cell migration. PLoS One. 2014; 9(2):e90593. https://doi.org/10.1371/journal.pone.0090593 PMID: 24587399

27. Kang J, Hsu CH, Wu Q, Liu S, Coster AD, Posner BA, et al. Improving drug discovery with high-content phenotypic screens by systematic selection of reporter cell lines. Nature Biotechnology. 2015; p. 1–11.

28. Loo LH, Wu LF, Altschuler SJ. Image-based multivariate profiling of drug responses from single cells. Nature methods. 2007; 4(5):445–453. PMID: 17401369

29. Rajaram S, Pavie B, Wu LF, Altschuler SJ. PhenoRipper: software for rapidly profiling microscopy images. Nature methods. 2012; 9(7):635–637. https://doi.org/10.1038/nmeth.2097 PMID: 22743764

30. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. 2006; 7(10):R100. https://doi.org/10.1186/gb-2006-7-10-r100 PMID: 17076895

31. Gordonov S, Hwang MK, Wells A, Gertler FB, Lauffenburger DA, Bathe M. Time series modeling of live-cell shape dynamics for image-based phenotypic profiling. Integrative Biology. 2016; 8(1):73–90. https://doi.org/10.1039/c5ib00283d PMID: 26658688

32. Cohen AR, Gomes FLAF, Roysam B, Cayouette M. Computational prediction of neural progenitor cell fates. Nature methods. 2010; 7(3):213–218. https://doi.org/10.1038/nmeth.1424 PMID: 20139969

33. Keren K, Pincus Z, Allen GM, Barnhart EL, Marriott G, Mogilner A, et al. Mechanism of shape determination in motile cells. Nature. 2008; 453(7194):475–480. https://doi.org/10.1038/nature06952 PMID: 18497816

34. Mitchison TJ, Cramer LP. Actin-based cell motility and cell locomotion. Cell. 1996; 84(3):371–379. https://doi.org/10.1016/S0092-8674(00)81281-7 PMID: 8608590

35. Cooper S, Sadok A, Bousgouni V, Bakal C. Apolar and polar transitions drive the conversion between amoeboid and mesenchymal shapes in melanoma cells. Molecular Biology of the Cell. 2015; 26 (22):4163–4170. https://doi.org/10.1091/mbc.E15-06-0382 PMID: 26310441

36. Buggenthin F, Buettner F, Hoppe PS, Endele M, Kroiss M, Strasser M, et al. Prospective identification of hematopoietic lineage choice by deep learning. Nature methods. 2017; 14(4):1–7. https://doi.org/10.1038/nmeth.4182

37. Thurston G, Jaggi B, Palcic B. Measurement of cell motility and morphology with an automated micro-scope system. Cytometry Part A. 1988; 9(5):411–417. https://doi.org/10.1002/cyto.990090502

38. Gail MH, Boone CW. The Locomotion of Mouse Fibroblasts in Tissue Culture. Biophysical journal. 1970; 10(10):980–993. https://doi.org/10.1016/S0006-3495(70)86347-0 PMID: 5531614

39. Dunn GA, Brown AF. A unified approach to analysing cell motility. Journal of cell science Supplement. 1987; 8:81–102. https://doi.org/10.1242/jcs.1987.Supplement_8.5 PMID: 3503898

40. Partin AW, Schoeniger JS, Mohler JL, Coffey DS. Fourier analysis of cell motility: correlation of motility with metastatic potential. Proceedings of the National Academy of Sciences of the United States of America. 1989; 86(4):1254–1258. https://doi.org/10.1073/pnas.86.4.1254 PMID: 2919174

41. Martens L, Monsieur G, Ampe C, Gevaert K, Vandekerckhove J. BMC Bioinformatics. BMC bioinfor-matics. 2006; 7(1):289–6. https://doi.org/10.1186/1471-2105-7-289 PMID: 16762054

42. Sacan A, Ferhatosmanoglu H, Coskun H. CellTrack: an open-source software for cell tracking and motility analysis. Bioinformatics. 2008; 24(14):1647–1649. https://doi.org/10.1093/bioinformatics/btn247 PMID: 18511469

43. Sebag AS, Plancade S, Raulet-Tomkiewicz C, Barouki R, Vert JP, Walter T. A generic methodological framework for studying single cell motility in high-throughput time-lapse data. Bioinformatics. 2015; 31 (12):320–328. https://doi.org/10.1093/bioinformatics/btv225

44. Cordelières FP, Petit V, Kumasaka M, Debeir O, Letort V, Gallagher SJ, et al. Automated Cell Tracking and Analysis in Phase-Contrast Videos (iTrack4U): Development of Java Software Based on Combined Mean-Shift Processes. PLoS One. 2013; 8(11):e81266–10. https://doi.org/10.1371/journal.pone.0081266 PMID: 24312283

45. Bray MA, Carpenter AE. CellProfiler Tracer: exploring and validating high-throughput, time-lapse microscopy image data. BMC bioinformatics. 2015; 16(1):368. https://doi.org/10.1186/s12859-015-0759-x PMID: 26537300

46. Xu J, Van Keymeulen A, Wakida NM, Carlton P, Berns MW, Bourne HR. Polarity reveals intrinsic cell chirality. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104 (22):9296–9300. https://doi.org/10.1073/pnas.0703153104 PMID: 17517645

47. Gorelik R, Gautreau A. Quantitative and unbiased analysis of directional persistence in cell migration. Nature Protocols. 2014; 9(8):1931–1943. https://doi.org/10.1038/nprot.2014.131 PMID: 25033209

48. Saxton MJ, Jacobson K. Single-particle tracking: applications to membrane dynamics. Annual review of biophysics and biomolecular structure. 1997; 26(1):373–399. https://doi.org/10.1146/annurev.biophys.26.1.373 PMID: 9241424

49. Viswanathan GM, Buldyrev SV, Havlin S, da Luz MG, Raposo EP, Stanley HE. Optimizing the success of random searches. Nature. 1999; 401(6756):911–914. https://doi.org/10.1038/44831 PMID: 10553906

50. Viswanathan GM, Raposo EP, da Luz MGE. Levy flights and superdiffusion in the context of biological encounters and random searches. Physics of Life Reviews. 2008; 5(3):133–150. https://doi.org/10.1016/j.plrev.2008.03.002

51. Viswanathan GM, Afanasyev V, Buldyrev SV, Murphy EJ, Prince PA, Stanley HE. Levy flight search patterns of wandering albatrosses. Nature. 1996; 381(6581):413–415. https://doi.org/10.1038/381413a0

52. Humphries NE, Queiroz N, Dyer JRM, Pade NG, Musyl MK, Schaefer KM, et al. Environmental context explains Levy and Brownian movement patterns of marine predators. Nature. 2010; 465(7301):1066–1069. https://doi.org/10.1038/nature09116 PMID: 20531470

53. Rahman A. Correlations in the Motion of Atoms in Liquid Argon. Physical Review. 1964; 136(2A):A405–A411. https://doi.org/10.1103/PhysRev.136.A405

54. Parry BR, Surovtsev IV, Cabeen MT, O'Hern CS, Dufresne ER, Jacobs-Wagner C. The Bacterial Cyto-plasm Has Glass-like Properties and Is Fluidized by Metabolic Activity. Cell. 2014; 156(1-2):183–194. https://doi.org/10.1016/j.cell.2013.11.028 PMID: 24361104

55. Mandelbrot BB, Wallis JR. Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. Water Resources Research. 1969; 5(5):967–988. https://doi.org/10.1029/WR005i005p00967

56. Bryce RM, Sprague KB. Revisiting detrended fluctuation analysis. Sci Rep. 2012; 2:1–6. https://doi.org/10.1038/srep00315

57. van der Maaten L, Hinton G. Visualizing Data using t-SNE. Journal of Machine Learning Research. 2008; 9:2579–2605.

58. Murtagh F, Legendre P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? Journal of Classification. 2014; 31(3):274–295. https://doi.org/10.1007/s00357-014-9161-z

59. Ho TK. Random decision forests. In: 3rd International Conference on Document Analysis and Recognition. IEEE Comput. Soc. Press; 1995. p. 278–282. Available from: http://ieeexplore.ieee.org/document/598994/.

60. Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters. 2006; 27(8):861–874. https://doi.org/10.1016/j.patrec.2005.10.010

61. Truitt ML, Conn CS, Shi Z, Pang X, Tokuyasu T, Coady AM, et al. Differential Requirements for eIF4E Dose in Normal Development and Cancer. Cell. 2015; 162(1):59–71. https://doi.org/10.1016/j.cell.2015.05.049 PMID: 26095252

62. Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust package for determining the number of clusters in a dataset. Journal of Statistical Software. 2014; 61(6):1–36. https://doi.org/10.18637/jss.v061.i06

63. Rousseeuw PJ. Silhouettes—a Graphical Aid to the Interpretation and Validation of Cluster-Analysis. Journal of Computational and Applied Mathematics. 1987; 20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7

64. Rencher AC, Christensen WF. Methods of Multivariate Analysis. John Wiley & Sons; 2013.

65. Enterline HT, Coman DR. The ameboid motility of human and animal neoplastic cells. Cancer. 1950; 3(6):1033–1038. https://doi.org/10.1002/1097-0142(1950)3:6%3C1033::AID-CNCR2820030612%3E3.0.CO;2-R PMID: 14783750

66. Wells A, Chao YL, Grahovac J, Wu Q. Cell motility in carcinoma metastasis as modulated by switching between epithelial and mesenchymal phenotypes. Frontiers in Bioscience. 2014; 16(1):815.

67. Verschueren H, Vandertalen I, Dewit J, Debraekeleer J, Debaetselier P. Metastatic Competence of Bw5147 T-Lymphoma Cell-Lines Is Correlated with in-Vitro Invasiveness, Motility and F-Actin Content. Journal of Leukocyte Biology. 1994; 55(4):552–556. PMID: 8145027

68. Doyle GM, Sharief Y, Mohler JL. Prediction of Metastatic Potential by Cancer Cell Motility in the Dunning R-3327 Prostatic Adenocarcinoma In Vivo Model. The Journal of Urology. 1992; 147(2):514–518. https://doi.org/10.1016/S0022-5347(17)37291-9 PMID: 1732634

69. Smith CL, Kilic O, Schiapparelli P, Guerrero-Cazares H, Kim DH, Sedora-Roman NI, et al. Migration Phenotype of Brain-Cancer Cells Predicts Patient Outcomes. Cell Reports. 2016; 15(12):2616–2624. https://doi.org/10.1016/j.celrep.2016.05.042 PMID: 27292647

70. Crammer K, Singer Y. On the algorithmic implementation of multiclass kernel-based vector machines. Journal of Machine Learning Research. 2002; 2(2):265–292.

71. Bentzinger CF, von Maltzahn J, Dumont NA, Stark DA, Wang YX, Nhan K, et al. Wnt7a stimulates myogenic stem cell motility and engraftment resulting in improved muscle strength. J Cell Biol. 2014; 205(1):97–111. https://doi.org/10.1083/jcb.201310035 PMID: 24711502

72. Siegel AL, Atchison K, Fisher KE, Davis GE, Cornelison DDW. 3D Timelapse Analysis of Muscle Satellite Cell Motility. Stem Cells. 2009; 27(10):2527–2538. https://doi.org/10.1002/stem.178 PMID: 19609936

73. Webster MT, Fan CM. c-MET Regulates Myoblast Motility and Myocyte Fusion during Adult Skeletal Muscle Regeneration. PLoS One. 2013; 8(11):e81757–16. https://doi.org/10.1371/journal.pone.0081757 PMID: 24260586

74. Bae GU, Gaio U, Yang YJ, Lee HJ, Kang JS, Krauss RS. Regulation of myoblast motility and fusion by the CXCR4-associated sialomucin, CD164. J Biol Chem. 2008; 283(13):8301–8309. https://doi.org/10.1074/jbc.M706730200 PMID: 18227060

75. Brack AS, Rando TA. Tissue-specific stem cells: lessons from the skeletal muscle satellite cell. Cell Stem Cell. 2012; 10(5):504–514. https://doi.org/10.1016/j.stem.2012.04.001 PMID: 22560074

76. Wang YX, Rudnicki MA. Satellite cells, the engines of muscle repair. Nat Rev Mol Cell Biol. 2011; 13(2):127–133. PMID: 22186952

77. Schultz E, Jaryszak DL, Valliere CR. Response of satellite cells to focal skeletal muscle injury. Muscle & nerve. 1985; 8(3):217–222. https://doi.org/10.1002/mus.880080307

78. Hughes SM, Blau HM. Migration of myoblasts across basal lamina during skeletal muscle development. Nature. 1990; 345(6273):350–353. https://doi.org/10.1038/345350a0 PMID: 2111464

79. Florini JR, Magri KA. Effects of growth factors on myogenic differentiation. The American journal of physiology. 1989; 256(4 Pt 1):C701–11. https://doi.org/10.1152/ajpcell.1989.256.4.C701 PMID: 2650560

80. Sheehan SM, Allen RE. Skeletal muscle satellite cell proliferation in response to members of the fibroblast growth factor family and hepatocyte growth factor. Journal of cellular physiology. 1999; 181 (3):499–506. https://doi.org/10.1002/(SICI)1097-4652(199912)181:3%3C499::AID-JCP14%3E3.0. CO;2-1 PMID: 10528236

81. Bischoff R. Chemotaxis of skeletal muscle satellite cells. Developmental dynamics. 1997; 208(4):505–515. https://doi.org/10.1002/(SICI)1097-0177(199704)208:4%3C505::AID-AJA6%3E3.0.CO;2-M PMID: 9097022

82. Do MKQ, Suzuki T, Gerelt B, Sato Y, Mizunoya W, Nakamura M, et al. Time-coordinated prevalence of extracellular HGF, FGF2 and TGF-β3 in crush-injured skeletal muscle. Animal Science Journal. 2012; 83(10):712–717. https://doi.org/10.1111/j.1740-0929.2012.01057.x PMID: 23035711

83. Bischoff R. A satellite cell mitogen from crushed adult muscle. Dev Biol. 1986; 115(1):140–147. https://doi.org/10.1016/0012-1606(86)90235-6 PMID: 3699242

84. Cornelison DD, Wold BJ. Single-cell analysis of regulatory gene expression in quiescent and activated mouse skeletal muscle satellite cells. Dev Biol. 1997; 191(2):270–283. https://doi.org/10.1006/dbio. 1997.8721 PMID: 9398440

85. Chakkalakal JV, Christensen J, Xiang W, Tierney MT, Boscolo FS, Sacco A, et al. Early forming label-retaining muscle stem cells require p27kip1 for maintenance of the primitive state. Development. 2014; 141(8):1649–1659. https://doi.org/10.1242/dev.100842 PMID: 24715455

86. Trapnell C. Defining cell types and states with single-cell genomics. Genome Research. 2015; 25 (10):1491–1498. https://doi.org/10.1101/gr.190595.115 PMID: 26430159

87. Battle C, Broedersz CP, Fakhri N, Geyer VF, Howard J, Schmidt CF, et al. Broken detailed balance at mesoscopic scales in active biological systems. Science (New York, NY). 2016; 352(6285):604–607. https://doi.org/10.1126/science.aac8167

88. Jones TR, Carpenter AE, Lamprecht MR, Moffat J, Silver SJ, Grenier JK, et al. Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106(6):1826–1831. https://doi.org/10.1073/pnas.0808843106 PMID: 19188593

89. Brooks RA. Intelligence without representation. Artificial Intelligence. 1991; 47(1-3):139–159. https://doi.org/10.1016/0004-3702(91)90053-M

90. E Michalska A. Isolation and propagation of mouse embryonic fibroblasts and preparation of mouse embryonic feeder layer cells. Current protocols in stem cell biology. 2007; Chapter 1:Unit1C.3. PMID: 18785164

91. Shea KL, Xiang W, LaPorta VS, Licht JD, Keller C, Basson MA, et al. Sprouty1 regulates reversible quiescence of a self-renewing adult muscle stem cell pool during regeneration. Cell Stem Cell. 2010; 6 (2):117–129. https://doi.org/10.1016/j.stem.2009.12.015 PMID: 20144785

92. Jaqaman K, Loerke D, Mettlen M, Kuwata H, Grinstein S, Schmid SL, et al. Robust single-particle tracking in live-cell time-lapse sequences. Nature methods. 2008; 5(8):695–702. https://doi.org/10.1038/nmeth.1237 PMID: 18641657