Automatic detection of spatio-temporal signaling patterns in cell collectives

Paolo Armando Gagliardi[1]* , Benjamin Grädel[1,2]* , Marc-Antoine Jacques[1,2] , Lucien Hinderling[1,2] , Pascal Ender[1,2] , Andrew R. Cohen[3] , Gerald Kastberger[4] , Olivier Pertz[1] , and Maciej Dobrzyński[1]

Increasing experimental evidence points to the physiological importance of space–time correlations in signaling of cell collectives. From wound healing to epithelial homeostasis to morphogenesis, coordinated activation of biomolecules between cells allows the collectives to perform more complex tasks and to better tackle environmental challenges. To capture this information exchange and to advance new theories of emergent phenomena, we created ARCOS, a computational method to detect and quantify collective signaling. We demonstrate ARCOS on cell and organism collectives with space–time correlations on different scales in 2D and 3D. We made a new observation that oncogenic mutations in the MAPK/ERK and PIK3CA/Akt pathways of MCF10A epithelial cells hyperstimulate intercellular ERK activity waves that are largely dependent on matrix metalloproteinase intercellular signaling. ARCOS is open-source and available as R and Python packages. It also includes a plugin for the napari image viewer to interactively quantify collective phenomena without prior programming experience.

Introduction

Cell signaling is a finely regulated process that governs fundamental aspects of a cell's life, its metabolism, responses to environmental cues, and cell fate determination. Time-lapse fluorescence microscopy has become a routine experimental modality that has greatly furthered our understanding of these processes and revealed the role of cell–cell heterogeneity and signaling dynamics (Purvis and Lahav, 2013; Albeck et al., 2013; Ryu et al., 2015). Live-cell imaging of fluorescent biosensors has also demonstrated that signaling propagates across cell collectives, a phenomenon that is becoming another important organizational principle in biology. Even though we are only beginning to understand the rules that govern the emergence of spatial correlations from single-cell signaling, their functional significance has been reported in various systems. Notably, the coordinated activation of extracellular signal-regulated kinase (ERK) in spatial clusters of cells is instrumental in the maintenance of epithelial homeostasis (Aikin et al., 2020; Gagliardi et al., 2021; Pond et al., 2022), acinar morphogenesis (Ender et al., 2022), osteoblast regeneration (De Simone et al., 2021), cell cycle progression (Hiratsuka et al., 2015), and in the coordination of collective cell migration (Aoki et al., 2017; Hino et al., 2020). Calcium waves observed in 2D monolayers trigger and facilitate the extrusion of oncogenically transformed cells from the epithelium, thus also contributing to its homeostasis (Takeuchi et al., 2020).

Waves rely on an active process in which individuals relay the signal from their neighbors (Talia and Vergassola, 2022). They are an important class of self-organizing phenomena that have emerged repeatedly at different biological scales, arguably because they can quickly transmit information at long distances throughout a community (Gelens et al., 2014; Wedlich-Söldner and Betz, 2018). At the cellular level, voltage-gated sodium channels mediate the action potential that propagates in neurons (Hodgkin et al., 1952). A rapid initial development phase in large animal eggs requires long-distance coordination of mitotic events, which can only be achieved by signaling waves (Chang and Ferrell, 2013). At the cell population level, waves rely on cell–cell communication as is the case of cyclic adenosine monophosphate waves in *Dictyostelium discoideum* (Alcantara and Monk, 1974), calcium waves in animal development, adult animal tissues, and plants (Robb-Gaspers and Thomas, 1995; Gilland et al., 1999; Choi et al., 2014), depolarization waves in the heart that control heartbeat (Noble, 1962), mechanical waves (Serra-Picamal et al., 2012), and genetic waves in the presomitic mesoderm (Tsiairis and Aulehla, 2016). Wave patterns also occur in communities of independent multicellular organisms. Giant honeybees flip their abdomens in a coordinated fashion that results in shimmering waves that repel an intruder that endangers the colony (Kastberger et al., 2014a).

[1]Institute of Cell Biology, University of Bern, Bern, Switzerland; [2]Graduate School for Cellular and Biomedical Sciences, University of Bern, Bern, Switzerland; [3]Department of Electrical and Computer Engineering, Drexel University, Philadelphia, PA, USA; [4]Institute for Biology, Karl-Franzens-University Graz, Graz, Austria.

*P.A. Gagliardi and B. Grädel contributed equally to this paper. Correspondence to Maciej Dobrzyński: maciej.dobrzynski@unibe.ch.

Despite waves being so common in biology, a general and easy-to-use computational tool to identify them and extract useful statistics for further analysis is still missing. In parallel, other fields like geographic information science and ecology have accumulated a multitude of statistical methods to analyze patterns in weather prediction, epidemiology, animal migration, and voting preferences (Moore and Carpenter, 1999; Lawson, 2006; Pfeiffer et al., 2008; Bivand et al., 2013). Here, we build upon methods for spatio-temporal clustering and analysis of point processes and develop a computational tool for the Automatic Recognition of COllective Signaling (ARCOS) events and their characterization with various statistics. The open-source code is freely available as R and Python packages, along with extensive documentation. We also developed a plugin for napari—a fast, interactive, multidimensional image viewer for Python (Sofroniew et al., 2022). The plugin provides a graphical interface to ARCOS and facilitates an intuitive discovery environment to quantify collective phenomena in time-lapse images.

In the following sections, we explain the ARCOS algorithm and apply it to several biological systems. First, we used an optogenetic actuator that activates the mitogen-activated protein kinase (MAPK) network to generate synthetic waves of ERK activity in vitro in an MCF10A epithelial monolayer. This way we could establish a ground truth to test the algorithm in a controlled setting. We then studied mammary epithelial cells expressing different oncogenic mutations. We found that ERK activity waves are exacerbated by oncogenic mutations and that the wave propagation depends on matrix metalloproteinase (MMP)–mediated intercellular communication. Further, we quantified apoptosis-induced waves of ERK activity, calcium waves in renal cells, and collective shimmering in a colony of giant honeybees. Finally, we tackled 3D geometry and identified ERK waves in mammary acini in vitro.

Results

The algorithm

To explicate the core ARCOS algorithm, let us consider a radially expanding wave, a "firework," as shown in Fig. 1 A, where we schematically depict three consecutive frames of a developing activation cluster. Such events have been observed, for example, in epithelial monolayers where a correlated activation of ERK is typically initiated by an apoptotic cell (Aikin et al., 2020; Gagliardi et al., 2021; Pond et al., 2022). In the first step, we apply a spatial clustering algorithm, *dbscan* (Ester et al., 1996; Campello et al., 2013; Hahsler et al., 2019), to every frame independently to identify pockets of active cells. The clustering requires two parameters: the search radius ε and the minimum cluster size *minClSz* within that radius. We recommend setting ε to ≈1.5× the first nearest neighbor distance (NND), and *minClSz* to the dimensionality of the problem plus one, i.e., 3 for 2D, 4 for 3D, etc. (Ester et al., 1996).

In the second step, spatial clusters identified in the first step are linked between frames based on their relative distance. The first time frame contains only a single five-cell spatial cluster #1 that becomes the seed of a new collective event (CE). The second frame also contains a single but bigger cluster #2. To link clusters

#1 and #2, we calculate the NND between cells that comprise both clusters and compare it to the search radius ε_{prev} used for tracking clusters over time. For simplicity, we set $\varepsilon_{prev} = \varepsilon$. In our case, at least one cell from cluster #2 is within ε_{prev} to cells in cluster #1. Hence, cells in cluster #2 inherit the identifier of the seed cluster #1. The third frame contains four spatial clusters #3–6. Based on NNDs between these cells and cells from the previous frame, we conclude that only clusters #3, #5, and #6 can be linked to the growing CE from the previous frame. Cells from cluster #4 are farther than ε_{prev}, and therefore they become a seed of a new CE.

Thus, the algorithm first identifies spatial clusters in individual frames and then tracks the clusters between frames based on their spatial overlap (Fig. 1 B). To illustrate how this algorithm can integrate into a real-life workflow, consider a minimal pipeline to identify collective ERK activation in a 2D MCF10A epithelium (Fig. 1 C). After acquiring multichannel time-lapse movies, images are segmented and cells are tracked over time. The resulting single-cell ERK activity time series (ERKATS) form a space–time pattern (x,y,m,t), where (x,y) are the centroid coordinates of nuclei, m is the measurement of ERK activity, and t corresponds to time. The input for ARCOS should only contain "active cells," i.e., spatial coordinates and time, (x,y,t), of cells with the measurement m greater than a threshold.

Binarization

If the input data for the first step of the analysis does not comprise active cells, a binarization of the measurement is necessary before clustering. We implemented several methods to binarize time series data in ARCOS. The most straightforward approach applies a global threshold to the measurement. However, data from time-lapse microscopy often exhibit high variability in the baseline due to different expression levels of the biosensor. Additionally, high-intensity light and long exposures can lead to photobleaching of fluorescent probes, which may cause long-term trends in fluorescence intensities. To address this, ARCOS offers detrending methods based on a running median or a fit to a polynomial (Fig. 1 D; Materials and methods).

An alternative approach to measurement binarization utilizes local indicators of spatial association (LISAs), which are a class of statistics that quantify local spatial correlation and clustering (Getis and Ord, 1992; Anselin, 1995; Naimi et al., 2019). A LISA statistic applied to individual time frames of the space–time pattern (x,y,m,t) yields a coefficient of local spatial association ρ at every coordinate (x,y). For our purposes, we are looking for pockets where the coefficient is high, which corresponds to regions with high and positively correlated single-cell ERK activity. After thresholding ρ, we obtain a space–time pattern (x,y,t) that is restricted only to cells that form pockets of high activity. We assume these clusters estimate the location of active cells and we use this subset as an input for spatial clustering with ARCOS. LISA-based binarization was performed accurately on a cluster of cells with high ERK activity in the MCF10A epithelium (Fig. 1 E).

Extracting statistics

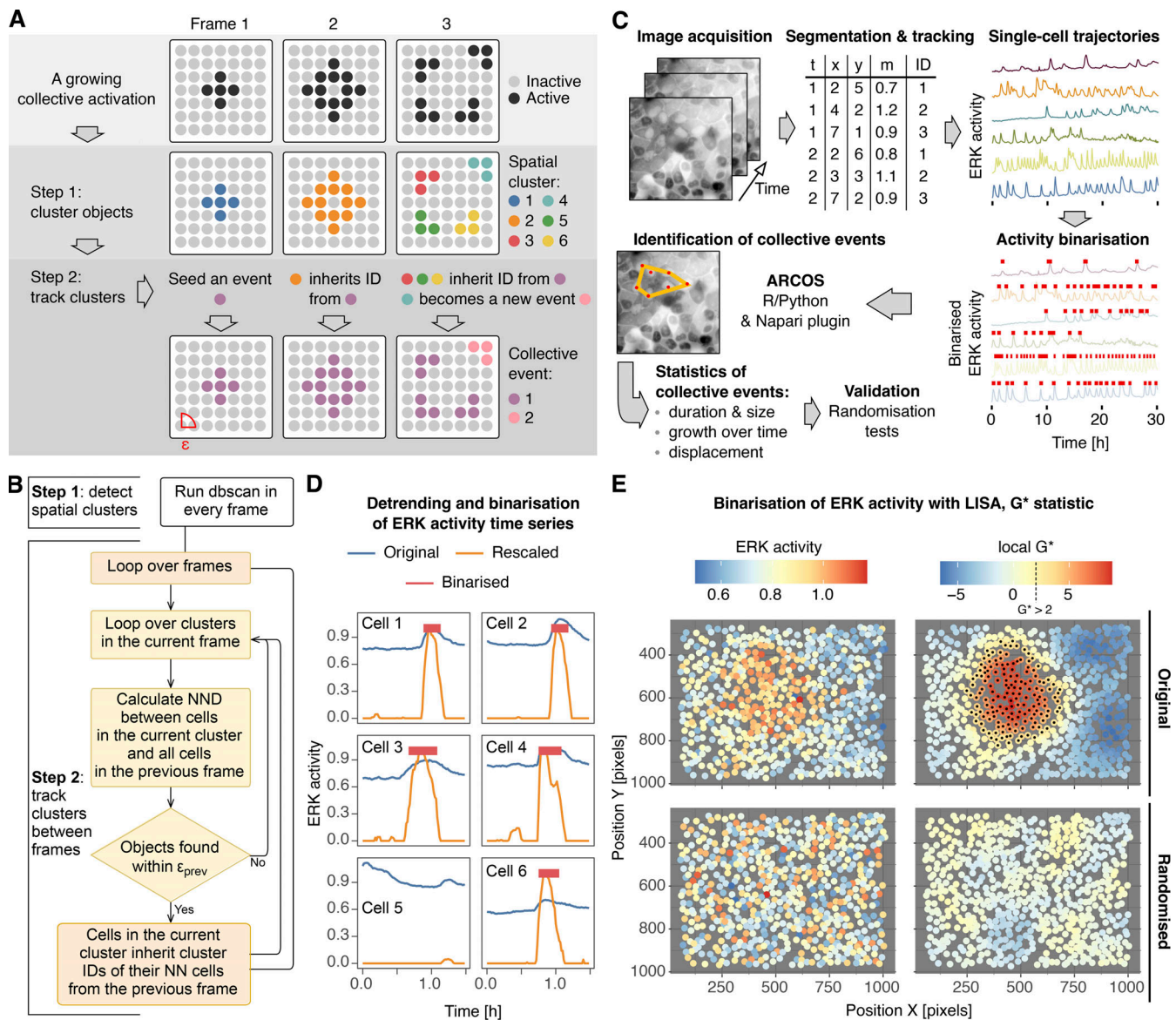ARCOS applied to binarized measurements yields spatio-temporal clusters with identifiers of participating cells. Such a

Gagliardi et al.
Automatic detection of collective signaling

Journal of Cell Biology
https://doi.org/10.1083/jcb.202207048

2 of 17

Figure 1. **The algorithm and data analysis pipeline. (A)** Demonstration of the ARCOS algorithm on a growing activity cluster. Protein activity of cells arranged in a mesh can assume two states, inactive (light gray) and active (black). In step 1, the *dbscan* algorithm spatially clusters active cells independently in each frame. In step 2, clusters are tracked between frames. The cluster in frame 1 forms a seed of a CE. The cluster in frame 2 is linked to this seed cluster because its member cells are within the neighborhood radius ε_{prev}. In frame 3, only clusters #3, #5, and #6 are linked to the previous frame's cluster. Cells in cluster #4 are too far and thus form a new seed of a CE. **(B)** Flowchart of the core algorithm for detecting CEs in time-lapse data. **(C)** Image acquisition and analysis pipeline. Segmentation and tracking software identifies cells in multichannel time-lapse movies and produces a long-format output where every row is the measurement *m* in a cell *ID* at coordinates (x,y), at time *t*. Single-cell protein activity time series are then detrended and binarized. Binary activity time series are then fed into ARCOS to identify collective activity. **(D)** Detrending and binarization of single-cell ERK activity with a running median method. Time series are first smoothed with a short-range filter, then the signal smoothed with a long-range filter is subtracted from the original. The result is rescaled and thresholded to binarize it. Red segments indicate periods of activity, which are then processed by ARCOS. **(E)** Binarization of a single frame with a LISA, G* statistic. The original frame is compared to the same frame with the measurement randomized between cells while keeping the original X-Y positions. Black dots indicate active cells with G* > 2.

dataset can be further visualized or processed to extract various statistics ([Fig. 1 C]). Our R/Python implementations offer several helper functions to calculate CEs' duration, size, growth over time, and center of mass displacement. Visualization of results is best performed in the napari plugin, where the original data can be viewed together with image segmentation results and CEs. The plugin also offers interactive plots to display the statistics

and to examine cells' measurements over time, their binarization, and their assignment to CEs ([Video 1]).

Validation

The observed statistics of CEs can be further tested against a null hypothesis that they do not differ from CEs calculated in a system with an equivalent total activity but with activations

occurring randomly in space and time. We provide several R/Python implementations, with varying assumptions about the null hypothesis, to randomize the binarized measurement and to simulate the null distribution, i.e., the sampling distribution of a statistic under the null hypothesis (Fig. S1; Materials and methods; Manly, 2007). To test a one-sided alternative hypothesis that, for example, the mean CE duration in the observed time-lapse is greater than the CE duration in randomly activating cells, we repeatedly apply any of the randomization methods and calculate the statistic for every iteration. The fraction of cases when the durations calculated from the randomized dataset are equal to or exceed the observed duration is the P value, i.e., the probability of obtaining the statistic at least as extreme as observed (given the null hypothesis is true; Davison and Hinkley, 1997; Good, 2005; Phipson and Smyth, 2010). These randomization tests of CEs' statistics quantify the "unusualness" of observed spatio-temporal correlations and thus form the validation part of our workflow.

Optogenetically induced collective ERK activation

Aside from synthetic datasets used during the algorithm development, we also wanted to test the code on a biological system, with the actual variability in protein activity, but in a controlled setting. To this end, we utilized an MCF10A cell line with a stably expressed optogenetic actuator–biosensor system described earlier (Aoki et al., 2017; Fig. 2 A). Using an OptoRAF actuator, we could induce ERK activity via the RAF-MEK-ERK pathway in individual cells with light pulses and measure ERK activity using the cytoplasmic/nuclear translocation of the ERK–KTR biosensor. Additionally, we used a digital micromirror device (DMD) to stimulate desired regions of the epithelium over time, thus generating synthetic collective ERK activity patterns that would mimic various aspects of spontaneous waves (Fig. 2 B). Notably, the ERK activity induced by OptoRAF does not propagate to neighboring cells, which allowed us to precisely control the extent of synthetic ERK activity patterns and to benchmark ARCOS.

In the first scenario, we repeatedly stimulated a growing circular region of the epithelium at 5-min intervals, starting at 11 min, with 100-ms light pulses. We observed a clear ERK activation in the stimulated regions (Fig. 2 C). ARCOS applied to detrended and binarized ERKATS identified a single, growing activity cluster induced by optogenetic stimulation (Fig. 2 C and Video 2). Additionally, several smaller CEs that lasted <5 min or comprised <20 unique cells were detected. They stemmed from spontaneous ERK activations in the epithelium and were discarded after processing.

To verify that the optogenetic stimulation was confined to desired regions, we inspected single-cell ERKATS (Fig. 2 D). Cells from the innermost region received all stimulation pulses after the 10-min mark. ERK activity increased sharply, although a few responses were delayed, and ensued only after the second or even the third pulse. We attribute this delay to biological variability and the fact that the cells are at the border of an already small activation region. Cells farther away from the center responded accordingly to pulses that they were exposed to, although a few cells were activated earlier, likely by the illumination of regions with smaller

radii. The ERK activity in cells not exposed to DMD illumination remained at the baseline.

We concluded that our protocol was sufficiently selective to control the stimulation of a growing ERK activation cluster and that ARCOS correctly identified such a synthetic wave. An intuitive visualization of that result is shown in a "spaghetti" plot (Fig. 2 E). Therein, the spatio-temporal evolution of collective ERK activity was projected onto the 2D (x,t) plane. Each trace corresponds to a single cell involved in a CE identified by the algorithm. We confirmed that the collective ERK activation grew over time and within the bounds of the DMD illumination. As a cross-check, we compared the result obtained from ARCOS with an independent quantification of the global spatial correlation per frame using Moran's I statistic (Gittleman and Kot, 1990; Bivand et al., 2013). Said statistic is commonly used in the field of geographic information science to quantify spatial phenomena. The Moran's I coefficient increased with the growing collective ERK activity region and peaked when that region was the biggest (Fig. 2 E). This result confirmed a spatially correlated cluster as detected by ARCOS.

We also tried other illumination scenarios to further challenge the algorithm. To test the tracking, we illuminated a pattern with a circle that moved around the field of view (FOV; Fig. 2, F and G; and Video 3). Another common scenario observed in the epithelium prompted us to apply a circular pattern that was slowly splitting into two detached regions (Fig. 2, H and I; and Video 4). In both cases, ARCOS correctly identified the spatio-temporal progression of the optogenetically induced ERK activity region. Due to a longer, 2 h 40 min, acquisition time, we detected additional ERK activity clusters that stemmed from apoptotic events and/or a spontaneous activation. The results from the optogenetic experiments assured us that our image analysis pipeline performed well in a controlled setting with biological variability and that we could apply it to more challenging scenarios.

Oncogenic mutations increase ERK waves in MMP/epidermal growth factor receptor (EGFR)-dependent manner

One of our earlier studies showed that apoptosis of single cells within epithelium triggers waves of ERK activity pulses (Gagliardi et al., 2021). These waves act as a survival signal that protects two to three layers of neighboring cells from further apoptosis, thus contributing to epithelial homeostasis. The ERK waves are trigger waves (Gelens et al., 2014) in which the extruding apoptotic cell activates MMP-mediated cleavage of pro-EGFR ligands. Subsequently, the EGFR activates ERK in adjacent cells (Aoki et al., 2017; Aikin et al., 2020). ERK-mediated activation of myosin contractility might then provide a mechanical input that activates ERK in farther cells through MMP/EGFR signaling. This explains how the ERK wave emerges through a mechanochemical feedback loop (Hino et al., 2020). Given those past results, we were curious how oncogenic mutations that affect the MAPK network might influence the spatio-temporal ERK activation and the molecular mechanism of CEs propagation. Therefore, we explored ERK waves in starved MCF10A cells with oncogenic heterozygous mutations, KRAS^{G12V} (Konishi et al., 2007) or PIK3CA^{H1047R} (Gustin et al., 2009), that activate the MAPK/ERK and frequently
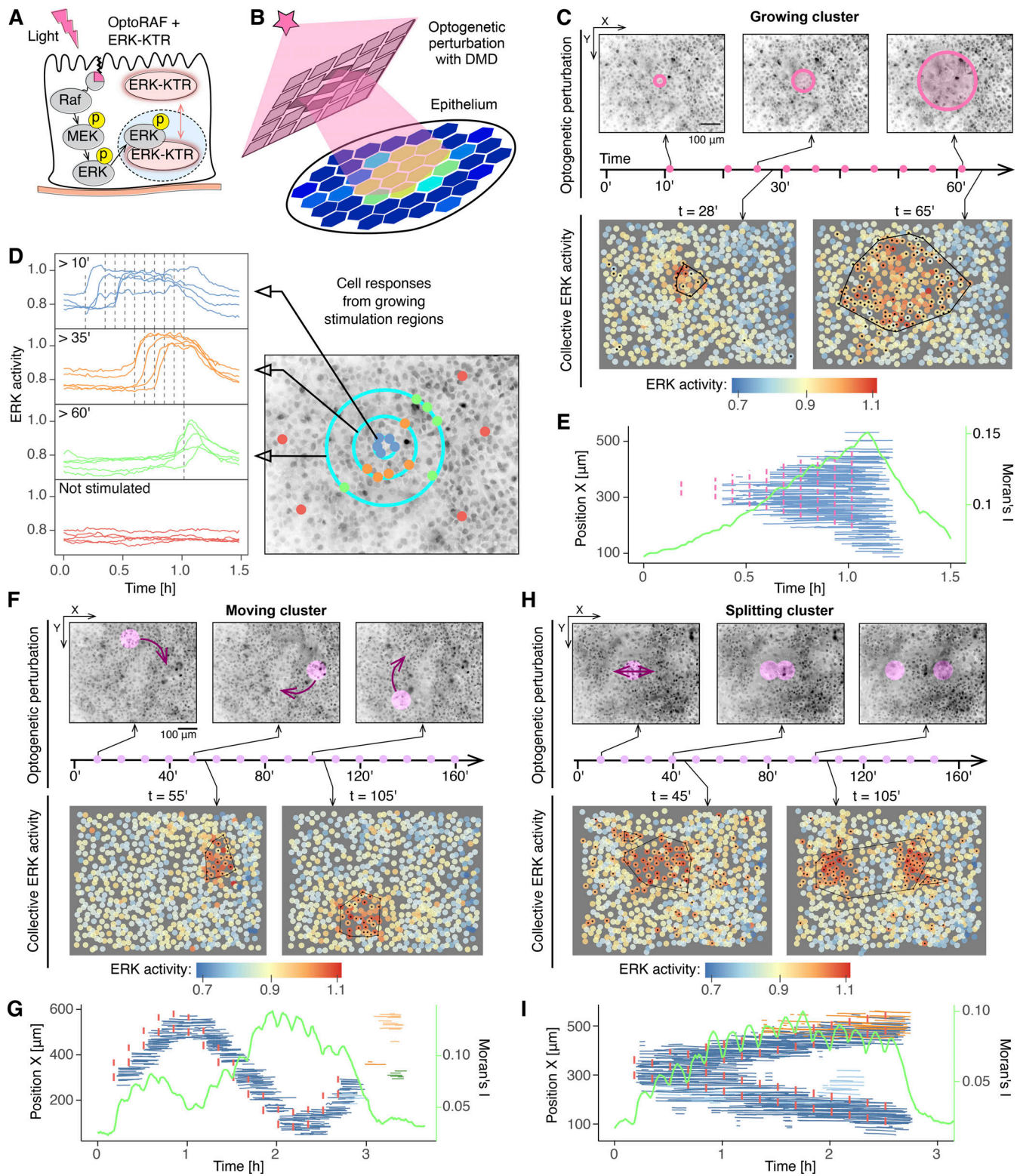
Figure 2. Testing ARCOS with optogenetically induced collective ERK activation. (A) Optogenetic actuator/biosensor circuit comprises the optogenetic actuator OptoRAF tagged with mCitrine, the ERK biosensor (ERK-KTR) tagged with mRuby2, and a nuclear marker (H2B) tagged with miRFP703. **(B)** Schematic of a spatially constrained stimulation of the 2D epithelium with a DMD. **(C)** Repeated, pulsed optogenetic stimulation of the MCF10A WT epithelium with a growing circular region using DMD. Upper row: Inverted raw images from selected time points. The circle outlines the stimulation region. Pink dots on the timeline indicate the timing of a 100-ms-long optogenetic pulse. Lower row: Quantification of ERK activity from image segmentation for selected time points. Black dots indicate active cells identified from detrended and binarized ERKATS; black lines indicate a convex hull around collective activation. **(D)** Single-cell ERK activity in different regions of the FOV from panel C. Vertical dashed lines indicate the timing of the optogenetic stimulation. Cells in the center are stimulated at all stimulation times after 10′; cells farther away are stimulated only with pulses that occur after 35′, etc. **(E)** Spaghetti plot of collective ERK

activation: an x-axis projection of single cells from panel C during their participation in a CE. Vertical dashed lines indicate the timing and position of the optogenetic stimulation. The green solid line is the Moran's I measure of global spatial autocorrelation calculated from single-cell ERKATS. **(F)** Repeated, pulsed optogenetic stimulation of a circular region moving clockwise along a circle. Legend as in panel C. **(G)** Spaghetti plot and global Moran's I of CEs from panel F. **(H and I)** Same as F and G but the optogenetic stimulation is a circular region that splits horizontally over time.

occur in cancer. In the KRAS^{G12V} mutant, loss of GTP hydrolysis leads to increased activation of the rapidly accelerated fibrosarcoma (RAF) kinase, thus activating the MAPK/ERK kinase (MEK) and ERK. The H1047R mutation in PIK3CA, the α subunit of PI3K, leads to amphiregulin expression/secretion amplification, which activates the non-mutated MAPK/ERK pathway via EGFR (Gustin et al., 2009; Young et al., 2015).

As expected, at the population level, in starved conditions, without external stimulation, the average ERK activity was higher in the mutants than in the WT cells (Fig. 3 A). Notably, the increased basal activity occurs without ERK activity waves due to apoptosis, which is negligible in the mutants. At the single-cell level, ERK activity still exhibited non-periodic pulses, although their number and shape markedly differed between the two mutant cell lines (Fig. 3 B). WT cells displayed sharp, isolated, and rare ERK pulses as observed earlier (Albeck et al., 2013). In the KRAS^{G12V} cells, individual ERK pulses were still discernible but often lasted longer and formed "massifs" of ERK activity that resulted in higher pulse frequency (Fig. 3 C). In the PIK3CA^{H1047R} cells, as already reported (Ender et al., 2022), ERK pulses retained the same shape, duration, and amplitude as in WT cells, but the frequency of these pulses was the highest (Fig. 3 C).

We then wondered how different single-cell ERK activity dynamics translated into spatial correlations in the epithelium. ERK activity was least spatially correlated in WT cells, where CEs (comprising minimum three cells and lasting at least 15 min) were sparse, short-lasting, and usually triggered by apoptotic events. In comparison, KRAS^{G12V} and PIK3CA^{H1047R} mutants exhibited a more spatially correlated ERK activity with longer and more frequent CEs (Fig. 3 D and Video 5). Given the quantification of CEs obtained from ARCOS, we quantified their duration and size, i.e., the total number of unique cells in a CE (Fig. 3 E). Even though the median statistics were similar across the cell lines, the distributions varied and confirmed our observations from spaghetti plots: small and short-lasting ERK activity waves in WT cells, and bigger and slightly longer-lasting events in the mutants (Fig. 3 F). Furthermore, the mutant cells triggered more CEs compared with the WT (Fig. 3 G).

Next, we explored the molecular mechanism of CE propagation in the presence of oncogenic mutations. We hypothesized that they were mediated by the same MMPs/EGFR intercellular communication mechanism that we identified for apoptotic waves in WT cells (Gagliardi et al., 2021). Treatment with 10 μM of the MMPs' inhibitor batimastat reduced the average ERK activity in the mutant cell lines (Fig. 3 A). WT and PIK3CA^{H1047R} ERKATS became nearly flat, with a strong increase in the fraction of cells without any ERK activity pulses (WT: 24%→65%; PIK3CA: 9%→51%). In contrast, 58% of KRAS^{G12V} cells still displayed residual pulsatile activity (Fig. 3 C). This result suggests that WT and PIK3CA^{H1047R} completely depend on intercellular

communication for the regulation of ERK activity dynamics, whereas the dependency of KRAS^{G12V} is only partial. The hampered ERK activity dynamics after MMP inhibition coincided with a strong reduction of CEs' number, size, and duration in all three cell lines. However, we observed that in KRAS^{G12V} there are more residual CEs than in WT and PIK3CA^{H1047R} (Fig. 3, D–G).

An independent quantification with the global Moran's I coefficient calculated directly from segmented time-series data showed a reduced spatial correlation of ERK activity in all cell lines after batimastat treatment (Fig. 3 H). This confirms less coordinated ERK activation, as quantified with ARCOS, and shows that not only apoptotic ERK waves but also aberrant CEs caused by oncogenic mutations are at least partly mediated by MMPs.

We then compared the observed statistics of CEs to statistics calculated in a system with the same dynamics but without spatial correlations. We performed a randomization test by shuffling positions of entire time series according to Fig. S1 B. This only affected spatial correlations between neighboring cells, while leaving the dynamics intact. After performing 1,000 iterations, we detected CEs that were shorter-lived and involved fewer cells than the observed CEs; thus the one-sided P value for CE duration and the size was <0.001 (Figs. S2 and S3). We concluded that at the significance level of 0.05, we can reject the null hypothesis that the observed statistics of CEs in three cell lines with and without batimastat are no different from those calculated from random events.

ERK waves in the NRK-52E epithelium

To further demonstrate ARCOS's versatility, we quantified ERK waves in the NRK-52E renal epithelial cells that express the EKAREV-NLS ERK sensor. The majority of the waves in this cell line are triggered by apoptosis, as reported earlier (Aoki et al., 2013). Since waves are bigger and isolated from each other, we used a LISA-based binarization of ERK activity as the preprocessing step for ARCOS. To identify pockets of active cells, we calculated the local G^*, a LISA statistic that identifies spots of activity (Naimi et al., 2019), and applied it to individual frames of the (x,y,m,t) space-time pattern from image segmentation. After thresholding the resulting statistic, we obtained the (x,y,t) pattern only with active cells, which we then used in ARCOS (Fig. 4 A and Video 6).

The spaghetti plot confirms that apoptotic events trigger ERK activity waves (Fig. 4 B). We corroborate this result with the global Moran's I indicator of spatial correlation, which indeed peaks at the time of CEs identified with ARCOS (Fig. 4 B). We also find that the median CEs' duration of 0.5 h in NRK-52E cells is comparable with all MCF10A cell lines analyzed before (0.58–0.67 h) but the events are bigger with their median size of 16 unique cells compared with five to six in
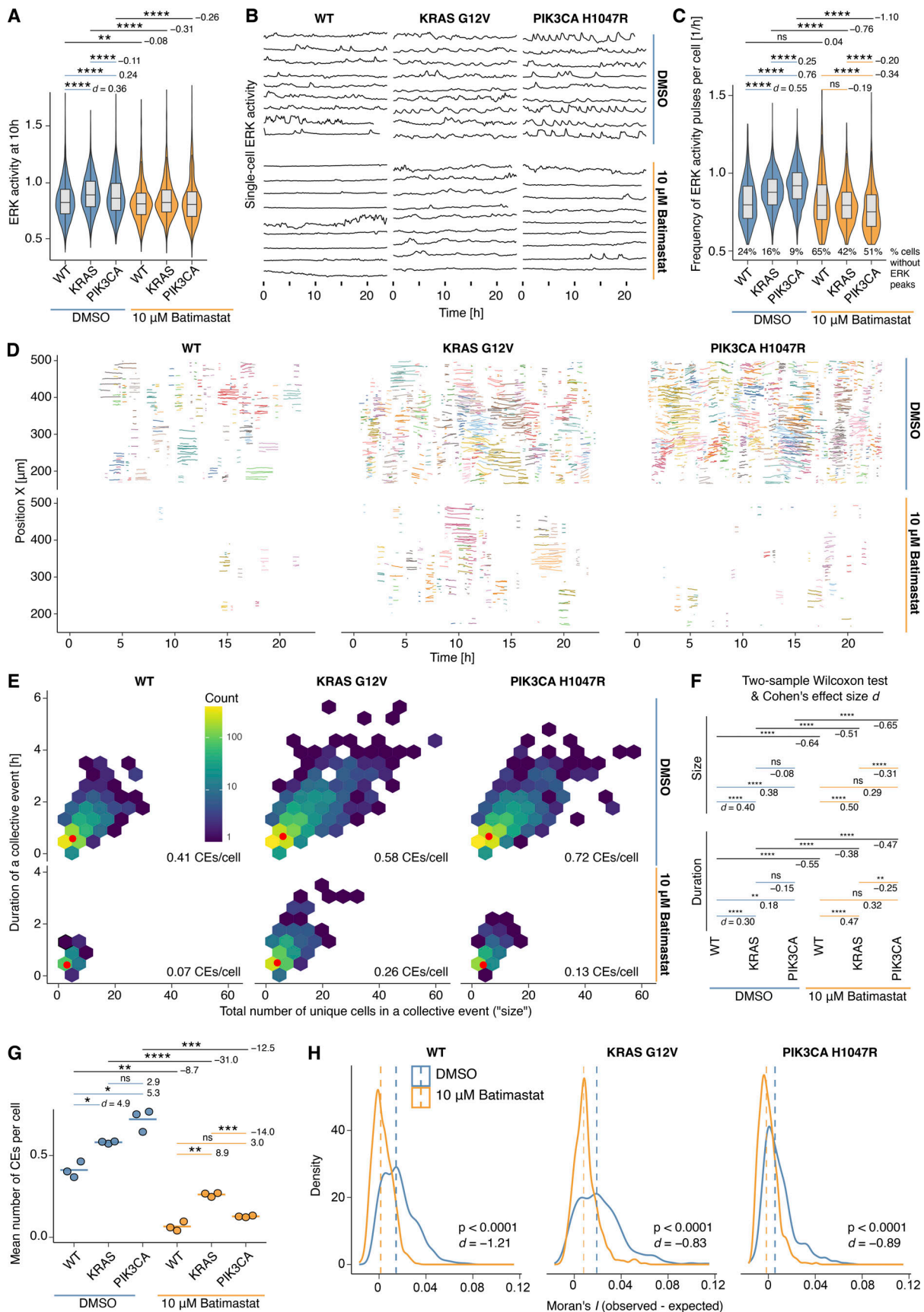
Figure 3. ERK activity in starved MCF10A WT, and KRAS[G12V] and PIK3CA[H1047R] mutant cell lines with and without the MMP inhibitor batimastat. (A) Population-averaged ERK activity at 10 h after the start of acquisition. Box plots show the median and 25th and 75th percentiles, whiskers correspond

to minimum and maximum non-outlier values; $N \approx 2,000$ cells pooled from three FOVs per cell line/treatment of a single experiment. Symbols indicate the statistical significance of a two-sample Wilcoxon rank sum test with Bonferroni correction: * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$. Numbers indicate Cohen's effect size d (Cohen, 2013). **(B)** A random sample of single-cell ERKATS from panel A. **(C)** Frequency of ERK activity pulses per cell for ERKATS from panels A and B. A pulse is a continuous period of binarized high ERK activity. Symbols and numbers as in A. **(D)** X-axis projections of single cells during their participation in CEs. Events are colored with a finite discrete palette; thus, colors repeat over time. Data from single FOVs. **(E)** Size and duration of CEs. Hexagonal tiles are colored according to the number of CEs (logarithmic scale). Red dots indicate median statistics. Data pooled from three FOVs per cell line/treatment of a single experiment. **(F)** Results of statistical tests for the size and duration of CEs for the data in E. Symbols and numbers as in A. **(G)** Each dot corresponds to the mean number of CEs in a FOV normalized to the time-averaged number of cells in that FOV. The results of unpaired two-sample t tests with Bonferroni correction are shown above the plot. Symbols and numbers as in A. **(H)** Global Moran's I calculated over time (287 time points) for each FOV. Each distribution contains values pooled from three FOVs and is based on $N = 3 \times 287$ values. Vertical dashed lines represent the median. P values obtained from the Wilcoxon rank sum test; d, Cohen's effect size.

MCF10A WT and mutants (Fig. 4 C). The one-sided P value from 1,000 iterations of the randomization test with shuffling time series' positions was <0.001 for CE duration and size (Dobrzyński, 2023).

Calcium waves in the MDCK epithelium

We then tested ARCOS on a different cell line and biosensor by analyzing calcium waves that trigger apical extrusion in the MDCK epithelium (Takeuchi et al., 2020; imaging data, courtesy
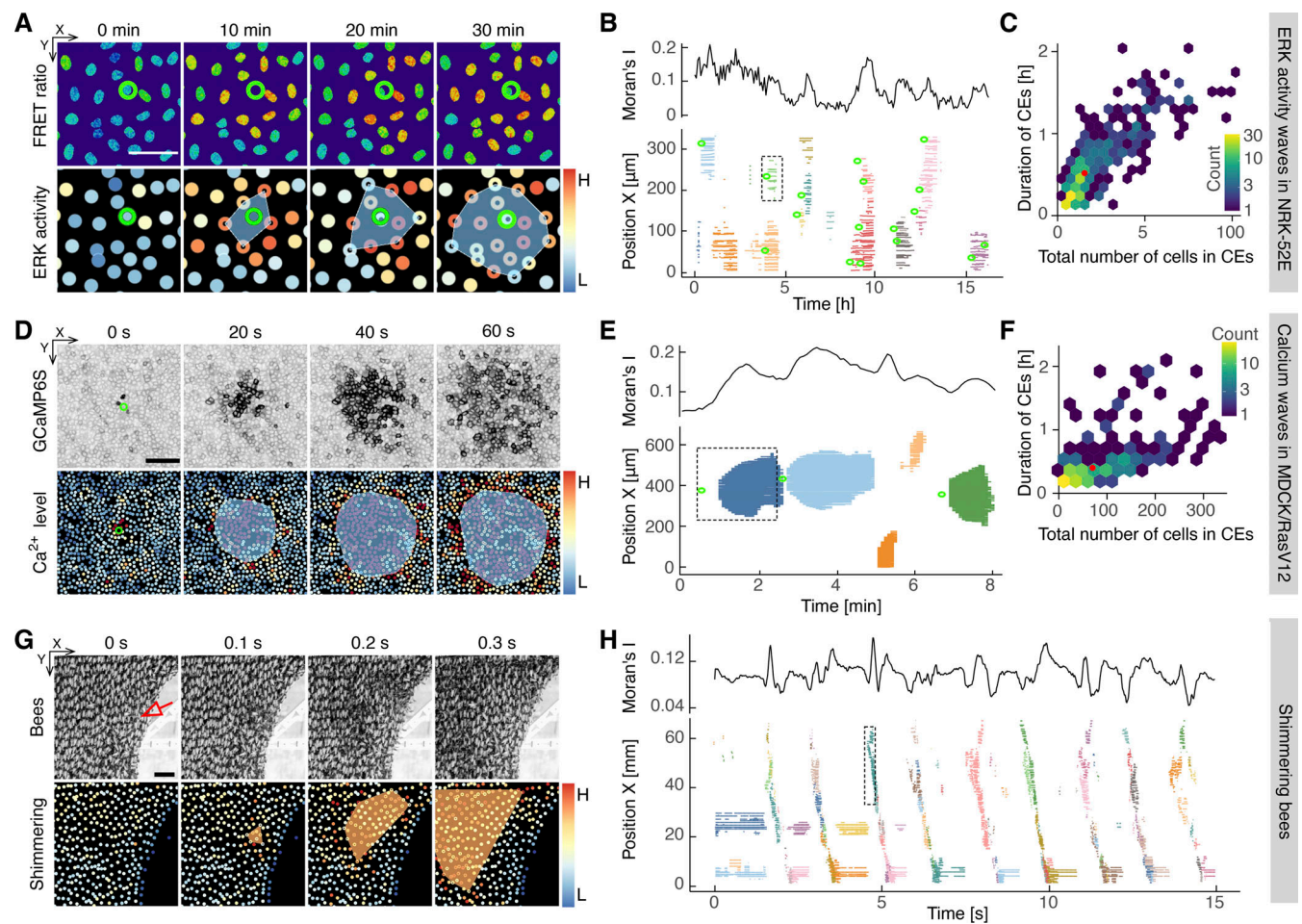

Figure 4. **ERK activity waves in the NRK-52E epithelium, calcium waves in the MDCK epithelium, and shimmering bees. (A)** Snapshots of an ERK wave in starved NRK-52E EKAREV-NLS cells induced by the apoptotic cell indicated by a green ring. The FRET ratio is calculated pixel-by-pixel from donor and acceptor images. Nuclei colored by ERK activity and a convex hull of the CE from ARCOS. Black dots indicate active cells calculated from the binarization step with the G* statistic. Scale bar, 50 μm. **(B)** CEs in the NRK-52E epithelium compared to the global Moran's I coefficient. Overlaid as green rings are manually annotated apoptotic events. **(C)** Size vs. duration of CEs. 341 CEs pooled from 11 FOVs acquired during a single experiment. The red dot indicates the median. **(D)** Snapshots of a calcium wave and its quantification in the MDCK epithelium with the apically extruded cell marked in green. Cells stably express GCaMP6S—a GFP-based intracellular calcium sensor. Scale bar, 100 μm. **(E)** CEs in the MDCK epithelium compared with the global Moran's I coefficient. Green rings are manually annotated RasV12-expressing cells that undergo apical extrusion triggered by calcium waves. **(F)** Hex-bin plot as in C containing 192 CEs from a single, 9-h-long acquisition. **(G)** Snapshots of shimmering bee waves induced by a dummy wasp presented to the colony on the right of the FOV. A leader bee that initiates the wave is indicated with a red arrow. The bottom row shows bees colored by the amount of their abdomen flipping and a convex hull of the CE identified with ARCOS. Scale bar, 50 mm. **(H)** 116 CEs from a single time-lapse of bee shimmering. **(B, E, and H)** Dashed rectangles indicate space–time sections from panels A, D, and G.

of Yasuyuki Fujita). Apical cell extrusion is an important homeostatic mechanism that eradicates harmful or suboptimal cells from the epithelium. In the experiment, the Myc-RasV12 oncogene was transiently expressed mosaically within a monolayer of MDCK cells. Due to their harmful potential, RasV12-transformed cells were apically extruded from the epithelium. Notably, before the extrusion, calcium levels acutely increased in the RasV12 cells and radially propagated to neighboring cells. Such waves have been shown to trigger and facilitate the process of cell extrusion.

Since single-cell time series of calcium levels had a well-defined baseline and lacked long-term trends, we used the G^* LISA statistic (Naimi et al., 2019) to identify pockets of active cells (Fig. 4, D and E; and Video 7). Spatial correlation calculated for each frame using global Moran's I coefficient peaks when calcium waves occur, which further strengthens our identification of CEs with ARCOS (Fig. 4 E). We find that a single collective calcium wave propagates radially from the extruded cell, can involve up to 300 cells, and can last up to 2 min (Fig. 4 F). The one-sided P-value from 1,000 iterations of the randomization test with shuffling time series' positions was <0.001 for CE duration and size (Dobrzyński, 2023).

Shimmering bees
Even though we developed ARCOS to tackle collective phenomena in cell signaling, our image processing pipeline can also quantify spatio-temporal phenomena in the context of ecology. As a demonstration, we analyzed a previously published time-lapse dataset of a giant honeybee colony (*Apis dorsata*) that collectively responds to a computer-controlled dummy wasp hovering in front of the honeybee nest (Kastberger et al., 2011). The giant honeybees responded to the wasp by flipping their abdomens in a simultaneous and cascadic way, generating a wave-like visual signal for external observers. The visual effect achieved by this repetitive collective response aims to repel the threat (Kastberger et al., 2008, 2014b). Pulses connected to these visual patterns at the nest surface also lead to vibrations within the nest, and these are suitable for informing nest mates, even on both sides of the nest, about the threat status (Kastberger et al., 2010; Kastberger et al., 2011; Radloff et al., 2011; Kastberger et al., 2012; Kastberger et al., 2013; Kastberger et al., 2014a).

As described above, shimmering waves propagate across the surface of the giant honeybee nest in response to repeated exposure to the dummy wasp presented to the colony on the right side of the FOV (Fig. 4 G). Notably, our quantification confirms an interesting aspect of the shimmering phenomenon. Each vertical streak in the spaghetti plot (Fig. 4 H) corresponds to a wave that propagates from right to left of the FOV. However, we note that each of such streaks consists of "sub-waves" as indicated by different colors within the streak. Such a shimmering wave is typically initiated by "leader" bees on the nest surface, assembled in so-called trigger centers across the nest surface (Schmelzer and Kastberger, 2009; Kastberger et al., 2010; Kastberger et al., 2014a). Video 8 confirms our observation that a single exposure to the dummy wasp induces several smaller waves consecutively triggered by the leader bees scattered across

the colony. These trigger centers were primarily arranged in the close periphery of the mouth zone of the nest, around those parts where the main locomotory activity occurs throughout daytime.

ERK waves in 3D acinus
Previously, we reported waves of ERK activity during the morphogenetic program of 3D MCF10A acini in vitro (Ender et al., 2022). One crucial stage of this process is the formation of a lumen by apoptosis of inner cells. We found that ERK waves coordinate spatial separation into two domains: the outer cell population, which survives, and the inner cell population, which undergoes apoptosis. Since the ARCOS algorithm can handle any integer dimensionality, we were curious to analyze ERK activity waves in this system. The top row of Fig. 5 A shows a single wave in a maximum projection of the raw ERK-KTR channel. This example CE starts at the surface of the acinus and propagates over time to other cells in the outer layer (Fig. 5, B and C; and Video 9). Only toward the end, some inner cells participate in the event. This observation reflects a more general trend that we reported earlier based on pooled data from 11 acini (Ender et al., 2022). Most CEs initiate and propagate at the outer layer of the acinus, thus providing a survival signal to those cells. The inner domain exhibited fewer spontaneous ERK pulses and received fewer pulses due to waves that remained predominantly in the outer layer. Such a spatio-temporal distribution of waves creates a lower survival signal in the inner domain, which contributes to a higher apoptosis rate and clearing of the lumen. The formation of hollow acini recapitulates key features of in vivo breast alveoli and makes this system an attractive model to study the effect of MAPK signaling on lumen morphogenesis.

Limitations
ARCOS is designed to identify spatial clusters and to track them over time, thus making it potentially applicable to diverse spatio-temporal phenomena. The computational cost lies mainly in clustering with *dbscan* and in the calculation of NNDs between objects in clusters, which is necessary to track CEs. For cases with many objects per cluster, such as calcium waves in Fig. 4 D, the calculation lasts longer and memory requirements increase. However, the analysis of even the longest time lapses presented here lasted several minutes on a contemporary PC. In our opinion, the main hurdle lies in the preparation of data from microscopy images. Modern machine learning–based approaches for image segmentation are robust even in the presence of a fair amount of noise (Fig. S4, A–D and G; and Video 10; Materials and methods). However, quality time series is crucial for the subsequent binarization and CE detection (Fig. S4, E, F, and H; Materials and methods).

Signal binarization may require several trials with either detrending/thresholding or LISA-based methods. In our case, the latter performed better when activity time series had well-defined baselines without long-term trends, as was the case for all datasets except MCF10A, which required detrending due to varying baselines. Furthermore, some understanding of the acquired CEs is necessary to set the search radius ε and the minimum for CE duration and size. If ε is too large, then independent CEs are merged into a single cluster; if too small, then a single CE

Gagliardi et al.
Automatic detection of collective signaling

Journal of Cell Biology 9 of 17
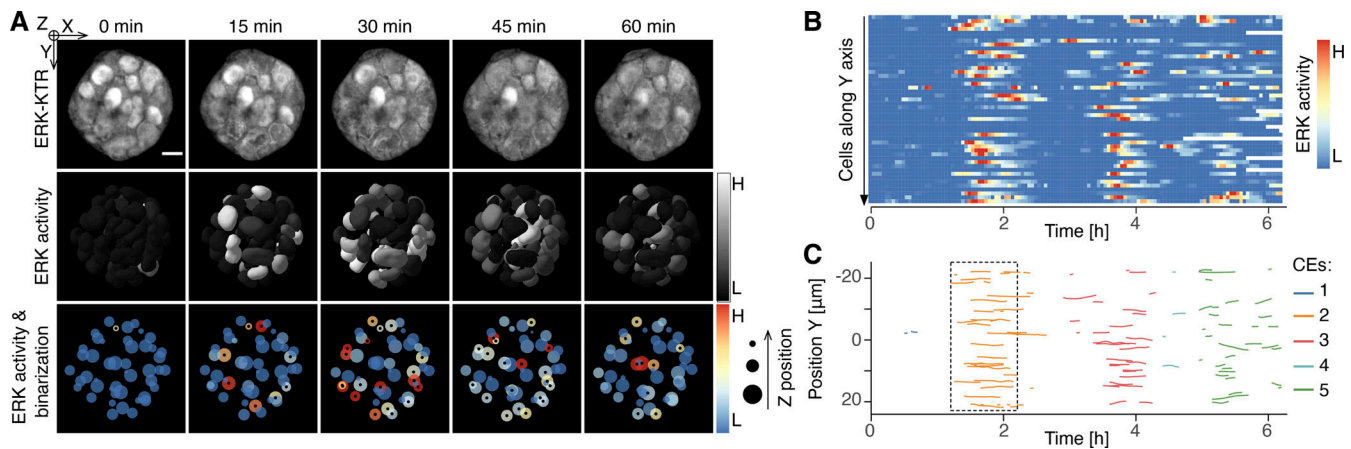https://doi.org/10.1083/jcb.202207048

Figure 5. **Quantification of collective ERK activation in a 7-d-old MCF10A WT acinus. (A)** Snapshots of a single CE. Top row—maximum projection of the ERK-KTR channel; middle row—cell nuclei shaded according to detrended and normalized ERK activity; bottom row—nuclear centroids colored by ERK activity and sized according to *z* distance. Black dots indicate active cells. Scale bar, 10 μm. **(B)** Heatmap of ERKATS ordered along the y-axis. Each row corresponds to a single cell. **(C)** Spaghetti plot of CEs during the entire acquisition. The dashed rectangle indicates the space–time section from A.

might be split into smaller events (Fig. S5, A–D; Materials and methods).

A time-lapse with randomly active cells can also lead ARCOS to identify spurious space-time correlations as CEs, especially when the total binarized activity (TBA) is high. To identify this regime, we simulated a 2D epithelium with randomly activating cells for a range of TBA. Above $TBA \approx 10\%$, due to an increased probability of neighboring cells remaining active, the algorithm identified several CEs comprising at least three cells and lasting at least three frames (Fig. S5, G). Therefore, validation with randomization tests is crucial to assess whether statistics calculated from observed data differ from the statistics obtained in a system with random activations.

Another regime when ARCOS may yield imprecise results is when CEs start overlapping. Simulation of concentrically growing events (Fig. S5 H) revealed that the statistics of collective activations agree with the ground truth up to $TBA \approx 10\%$, which is the threshold when individual events start colliding (Fig. S5, I–K). Even though ARCOS is designed to handle collisions of CEs, such situations are indiscernible above a certain threshold, making automated detection difficult. The actual percentages above which these problems emerge are only an approximate guideline and may further depend on detection thresholds for CE size and duration and on the nature of the propagating wave. In our experiments, TBA was well below the estimated overlap threshold (0.6% for calcium waves; 2% for bee shimmering, 3 and 7% for ERK activity waves in NRK-52E and MCF10A WT), although it amounted to ≈23% in untreated MCF10A mutant cell lines, which may partly explain the higher duration and size of detected CEs.

Discussion

We introduced ARCOS, a novel tool for detecting and quantifying collective spatio-temporal phenomena with a focus on cell signaling, and validated it on several biological systems with

drastically different spatio-temporal collective behaviors. We analyzed ERK activity waves in an epithelial monolayer that propagate radially on the timescale of 10–30 min, radial calcium waves on the timescale of 10–30 s, directional waves in a shimmering bee colony, and a 3D culture model of mammary acini. The open-source code provided as R and Python packages can be used in Jupyter notebooks or batch processing pipelines. Supplementary R notebooks contain full analysis of all experimental modalities presented here and provide example parameters (Dobrzyński, 2023). The napari plugin enables anyone without extensive programming knowledge to explore parameters through an intuitive graphical user interface on a platform that emerges as a de facto standard for viewing multidimensional images (Video 1).

Oncogenic mutations increase collective ERK dynamics

We unravel a new layer of complexity in cancer signaling by showing that two different oncogenic mutations known to increase ERK activity led to different, distinctive single-cell and collective responses. We focused our experiments on two widely studied and clinically relevant oncogenic mutations: KRAS[G12V] and PIK3CA[H1047R]. The former directly activates ERK by switching on RAF and MEK (Huang et al., 2021), while the latter activates ERK through a crosstalk mechanism that involves MMP-dependent amphiregulin-mediated EGFR activation (Ender et al., 2022). As observed before (Konishi et al., 2007; Gustin et al., 2009), we show that both mutations lead to increased population-averaged ERK activity in MCF10A epithelial cells. Evaluation of single-cell ERK dynamics reveals two additional levels of complexity—the oncogenic mutations modify: (1) ERK dynamics at the single-cell level; (2) the collective behavior of ERK signaling, as can be quantified using ARCOS.

Regarding ERK dynamics, we find that KRAS[G12V] cells display more ERK pulses that also switch off slower than ERK pulses in WT. In marked contrast, PIK3CA[H1047R] leads to increased ERK pulse frequency without modifying ERK pulse shape. The KRAS[G12V]-induced characteristic ERK dynamics can be

explained as follows. Active Ras triggers the tripartite RAF/MEK/ERK network with a negative feedback loop from ERK to RAF (Sturm et al., 2010; Fritsche-Guenther et al., 2011); however, it bypasses the ERK-RSK2-SOS negative feedback loop that acts upstream of the mutated Ras (Dessauges et al., 2022). This partly abrogated negative regulation explains why ERK pulses remain adaptive, albeit with slower dynamics compared with WT, rather than exhibiting sustained ERK activation. The ability of the MAPK pathway to rescale the ERK response in the presence of an activating mutation (Gillies et al., 2017) explains why cells remain sensitive to signaling inputs. In marked contrast, the PIK3CA^{H1047R} mutation activates the MMP/EGFR/amphiregulin-based crosstalk to activate the MAPK signaling network with intact feedback structures as in WT cells (Ender et al., 2022). This produces ERK pulses of identical shapes as in WT cells, albeit with a higher frequency.

Using ARCOS to quantify population-level ERK dynamics, we report that both KRAS^{G12V} and PIK3CA^{H1047R} lead to more, longer, and bigger CEs compared with WT. Because ERK waves rely on MMP-mediated cleavage of EGF ligands to propagate between cells, we used the MMP inhibitor batimastat to inhibit the propagation. As previously quantified, ERK waves vanish in WT cells upon inhibition of cell–cell communication (Gagliardi et al., 2021). We observe the same behavior for PIK3CA^{H1047R} cells. We explain this by the fact that in this cell line, the crosstalk from PI3K to ERK involves the EGFR-dependent activation by the EGF-like ligand amphiregulin, which also requires MMPs (Young et al., 2015; Ender et al., 2022). Interestingly, the number, length, and duration of CEs are only partly reduced in the KRAS^{G12V} cells treated with batimastat. We speculate that in addition to increased secretion of amphiregulin in this mutant (Minjgee et al., 2011), mechanical input is also responsible for cell–cell communication. ERK-dependent activation of myosin contractility has been implicated in ERK wave generation during collective epithelial motility, together with the MMP/EGF/EGFR system (Aoki et al., 2017; Hino et al., 2020). The increased periods of ERK activity in KRAS^{G12V} cells might exacerbate myosin activity and thus increase cell–cell communication independently of the MMP signaling system, leading to residual ERK waves upon MMP inhibition.

Concluding remarks

Our data demonstrate that the single-cell signaling dynamic does not merely result from network wiring within individual cells but also reflects emergent phenomena at the level of cell community. Disentangling these two contributions is necessary to further interpret oncogenic signaling. We showed that intercellular propagation of dynamic collective signaling plays a major role in the pathological hyperactivation of the MAPK pathway due to oncogenic mutations. Future studies of spatiotemporal dynamics in other oncogenic MAPK/Akt pathway mutants will be essential to better characterize the dependency of individual cells on intercellular communication. Quantification of these complex emergent processes might then allow for modeling signaling networks in cell communities using agent-based approaches, which will enable true understanding of

signaling processes at previously inaccessible lengths and time scales.

Materials and methods

Cell culture

WT human mammary MCF10A cells were a gift of J.S. Brugge, Harvard Medical School, Boston, MA, USA (Debnath et al., 2003). The isogenic variants of MCF10A harboring the heterozygous mutations PIK3CA^{H1047R} (Gustin et al., 2009) and KRAS^{G12V} (Konishi et al., 2007) were a gift of Ben Ho Park, Vanderbilt University Medical Center, TN, USA. NRK-52E expressing EKAREV-NLS were a gift from K. Aoki, National Institute for Basic Biology, Okazaki, Japan (Aoki et al., 2013). Cultivation of WT MCF10A cells and the isogenic derivative was carried out in growth medium composed of DMEM:F12, 5% horse serum, 20 ng/ml recombinant human EGF (Peprotech), 10 μg/ml insulin (Sigma-Aldrich/Merck), 0.5 μg/ml hydrocortisone (Sigma-Aldrich/Merck), 200 U/ml penicillin, and 200 μg/ml streptomycin. NRK-52E cells were cultured in DMEM supplemented with 10% FBS, 200 U/ml penicillin, and 200 mg/ml streptomycin.

Biosensors and optogenetic actuator

To observe nuclei and measure ERK activity, we used the PiggyBac plasmid coding for the stable nuclear marker H2B-miRFP703 (pPBbSr2-H2B-miRFP703) and the PiggyBac plasmids coding for the ERK activity biosensor ERK-KTR fused with mRuby2 (pHygro-PB-ERK-KTR-mRuby2) or mTurquoise2 (pHygro-PB-ERK-KTR-mTurquoise2). To express the CRY2-based optogenetic actuator OptoRAF and the plasma membrane anchor CIBN-KrasCT, we used a PiggyBac plasmid expressing both the proteins separated by the self-cleaving peptide P2A (pPB3.0-PuroCRY2-cRAF-mCitrine-P2A-CIBN-KrasCT). All plasmids were generated for a previous study from our laboratory (Gagliardi et al., 2021). We chose OptoRAF to locally stimulate ERK activity because, in MCF10A cells, the activation of RAF is downstream of the MMPs/EGFR-mediated cell–cell communication that propagates ERK waves. This property makes the OptoRAF system an ideal tool to generate artificial waves by precisely activating only the cells in the illuminated area and not outside.

To generate cells stably expressing the biosensors or the optogenetic actuator, we transfected the PiggyBac plasmids together with the helper plasmid expressing the transposase. The transfection of WT MCF10A cells and the mutated derivatives was carried out with FuGene (Promega) according to the manufacturer's protocol. Antibiotic selection and image-based screening were used to generate stable clones with the desired biosensors/optogenetic tools.

Live microscopy and optogenetics

MCF10A and NRK-52E cells were seeded in 5 mg/ml fibronectin (PanReac AppliChem) on 24-well #1.5 glass bottom plates (Cellvis) to reach confluence on the day of the experiment. The culture medium was replaced with imaging medium (DMEM:F12, 0.5 μg/ml hydrocortisone (Sigma-Aldrich/Merck), 200 U/ml penicillin and 200 μg/ml streptomycin, and 10 mM Hepes)

24 h before the experiments and kept for their entire duration. Image acquisition was executed with an epifluorescence Eclipse Ti inverted fluorescence microscope (Nikon) controlled by NIS-Elements (Nikon) with a Plan Apo air 20× (NA 0.8) or Plan Apo air 40× (NA 0.9) objectives and an Andor Zyla 4.2 plus camera at a 16-bit depth. Laser-based autofocus was used throughout the experiments. All the experiments were performed in a humidified chamber at 37°C and 5% CO_2. Excitation and emission of the fluorescent proteins were obtained with the following monochromatic LEDs and filters. Far red: 640 nm, ET705/72m; red: 555 nm, ET652/60m; NeonGreen: 508 nm, ET605/52; mTurquoise2: 440 nm, HQ480/40. For optogenetic experiments, cells expressing OptoRAF were kept in the dark for at least 24 h before the experiments and all preparatory steps at the microscope were carried out with red or green light (wavelength >550 nm). Stimulation of the OptoRAF was obtained with monochromatic 488 nm blue light for 100 ms at 0.3 W/cm². The stimulated areas were produced by a DMD (Andor Mosaic3) controlled by NIS Jobs based on masks prepared beforehand.

Image analysis of 2D epithelia

We used three image segmentation approaches. In the first, we used Ilastik (Berg et al., 2019) to classify pixels in the nuclear channel and build a nuclear probability mask. The classifier was trained on manual annotations of nuclei based on the H2B-mRFP703 fluorescent marker for MCF10A cells and EKAREV-NLS for NRK-52E. We used CellProfiler 3.0 (McQuin et al., 2018) to perform further analysis. A threshold-based segmentation identified nuclei from the probability masks. The nuclear segmentation masks were used to quantify the average nuclear pixel intensity in the ERK-KTR channel. To determine the area corresponding to the cytosol, we expanded the nuclei by a predefined number of pixels, but not further than the underlying fluorescence intensity in the ERK-KTR channel. We calculated the average cytosolic pixel intensity of ERK-KTR from the resulting cytosolic ring. By dividing the average pixel intensities from the cytosolic ring and the nucleus, we obtained the cytosolic/nuclear ratio that estimates the ERK kinase activity. For experiments with NRK-52E cells that express EKAREV-NLS biosensor, we calculated the fluorescence resonance energy transfer (FRET) ratio, where the FRET image is divided pixel-by-pixel by the donor image. Tracking of single cells was done on nuclear centroids in MATLAB using μ-track 2.2.1 (Jaqaman et al., 2008).

In the second approach, we used stardist (Schmidt et al., 2018) to segment nuclei, then shrank the nuclei by one pixel, and used this binary image together with the raw ERK-KTR image to segment the cytoplasm of corresponding cells using cellpose (Stringer et al., 2021; Pachitariu and Stringer, 2022). The cytoplasm was shrunk by a predefined number of pixels and mean intensities for nuclei and cytoplasm were extracted from the KTR images to subsequently calculate the KTR ratio as before.

In the third approach, we used a custom segmentation algorithm that extracts image features from the nuclear marker channel using filters from a modified VGG16 (Arsa and Susila, 2019) convolutional neural network pretrained on ImageNet (Deng et al., 2009). We then quantified nuclear and cytosolic

fluorescence intensity following the procedure described above. To track nuclei, we used a Python implementation of the linking algorithm (Crocker and Grier, 1996) provided in the trackpy library (Allan et al., 2021).

Apoptotic events were manually annotated based on nuclear shrinkage, as reported earlier (Gagliardi et al., 2021).

MCF10A acini

Live imaging of 3D mammary acini with MCF10A cells and the corresponding image analysis were executed as described earlier (Ender et al., 2022). MCF10A cells were dissociated to single cells and embedded in growth factor–reduced Matrigel (Corning) and overlaid with DMEM/F12 supplemented with 2% horse serum, 20 ng/ml recombinant human EGF, 0.5 mg/ml hydrocortisone, 10 mg/ml insulin, 200 U/ml penicillin, and 200 mg/ml streptomycin. After 3 d, EGF, insulin, and horse serum were removed from the medium. 25 mM Hepes was added to the medium before imaging.

Image acquisition was performed with an epifluorescence Eclipse Ti2 inverted fluorescence microscope (Nikon) equipped with a CSU-W1 spinning-disk confocal system (Yokogawa) and a Plan Apo VC 60× water immersion objective (NA = 1.2). Images were acquired with a Prime 95B camera at 16-bit depth (Teledyne Photometrics). 3D segmentation of nuclei and extraction of cytosolic/nuclear ERK-KTR fluorescence intensity ratio were performed using a customized version of the LEVER software (Winter et al., 2016).

Shimmering bees

Image acquisition was performed as described earlier (Kastberger et al., 2011). To segment the act of flipping the abdomen in individual bees, we used a pixel classifier from the Ilastik image processing software (Berg et al., 2019). We obtained a probability mask that could be thresholded and processed as a conventional single-cell time-lapse movie that resulted in single-bee time series data. Motion-active bees, i.e., bees that flipped their abdomen, were identified with the local G^* statistic, which we then thresholded and fed into ARCOS.

Signal binarization

Single-cell ERK activity obtained from image segmentation yields a (x,y,m,t) spatio-temporal point pattern, where (x,y) are centroid coordinates of nuclei, m is the corresponding measurement value at these coordinates (e.g., ERK activity), and t denotes time. To find CEs, we identify "active cells," i.e., cells with m above a certain threshold. We implemented three methods in the binMeas function of the ARCOS R and Python packages to identify activity periods in time series. In the simplest case (the biasMet parameter set to "none"), the measurement is rescaled globally to [0,1], smoothed with a short-term median filter to remove fast frequencies and/or incidental peaks, and binarized with a global threshold. Setting biasMet = "runmed" activates detrending, where a long-term median-smoothed measurement is subtracted from the initial time series. With biasMet = "lm," the detrending step subtracts a linear fit to the measurement. The aim of this step is to remove long-term trends due to photobleaching of fluorescent probes or to

remove permanently active cells. The subsequent thresholding of the detrended signal may lead to identification of longer pulses only during their peak activity, thus omitting the leading and the decaying phase of the pulse (Fig. 1 D).

For the LISA-based activity binarization, we used the statistics available in the R package elsa (Naimi et al., 2019). These include local Moran's I and local Geary's c (Anselin, 1995), as well as G and G^* statistics (Getis and Ord, 1992). The resulting spatial correlation coefficient is binarized with a fixed threshold, as exemplified in Fig. 1 E.

The choice of the binarization algorithm depends on the dataset and should be evaluated on a case-by-case basis. The main disadvantage of detrending with the running median is that it requires long time series to apply the median filter to estimate the baseline. The recommended window for such a filter usually amounts to a quarter of the time series' length. This condition may be difficult to satisfy when image segmentation errors produce short single-cell time series. Instead, LISA-based binarization is not contingent on long time series because frames are analyzed independently. However, the difficulty lies in the choice of a LISA statistic, e.g., local Moran's I, Geary's c, G, and G^*, for each of them measures a different aspect of the spatial pattern. For example, Moran's I measures the local deviation from the global mean, local G measures the clustering of hot/cold spots, and Geary's c measures the degree of dissimilarity of a value to its neighbor (Naimi et al., 2019). Moreover, most LISAs are not numerically bounded and are sensitive to the level of global autocorrelation, thus making them difficult to compare between systems with different overall clustering.

ARCOS algorithm

After time series binarization, the space–time point pattern (x,y,t) of active cells is passed to the clustering and tracking algorithm (Fig. 1 B). In the first step, a *dbscan* algorithm (Ester et al., 1996; Campello et al., 2013; Hahsler et al., 2019) is applied independently to every time frame t of the time-lapse to spatially cluster active cells. In the ARCOS R package, we use the implementation from the dbscan package and the Scikit-learn implementation in the Python version. Two, user-provided parameters determine this clustering: ε, the radius of the neighborhood, and minClSz, the minimum number of cells that form a cluster. The radius ε must be large enough to form clusters of neighboring active cells. We recommend setting it slightly larger than the mean first nearest neighbor distance calculated for all cells in the entire time-lapse. When ε is too small, cells are grouped into small clusters instead of one. When ε is too big, multiple CEs are joined. The minClSz parameter corresponds to the minimum number of cells that are considered a CE. Setting it to 1 allows detecting the initial stage of a CE, while risking that a sudden onset of simultaneously activating cells may be assigned to separate clusters. A bigger minClSz lets the algorithm delay the clustering until a larger event develops. A sudden appearance of multiple active cells may indicate an insufficient acquisition frequency during the experiment.

In the second step, identified clusters are tracked between the frames. The number of previous frames to link objects in the current frame depends on the parameter nPrev. Usually,

clusters between two consecutive frames are linked ($nPrev = 1$) but the number might be increased if missing frames are present in the dataset. Assuming, for simplicity, that $nPrev = 1$, the cluster tracking is performed as follows. For each cluster in t, we calculate the first NND between all cells in the current cluster and all cells in all clusters in the frame $t–1$. We use the nn2 function from the RANN package in R or the Scipy implementation of the KDTree nearest neighbor search in Python to calculate the distances. If at least a single cell in the current cluster is within a threshold distance ε_{prev} to a cell in the previous frame, cells in the current cluster inherit cluster IDs from first nearest neighbor cells in the previous frame. The ε_{prev} parameter is the threshold for detecting cluster displacements between the frames. Detection of a propagating activation wave as several consecutive CEs suggests that ε_{prev} should be increased.

Visualization with hex-bin plots

To display the relationship between the size and duration of CEs, we used hex-bin plots (Fig. 3 E; and Fig. 4, C and F), as they avoid overplotting that can occur in a conventional scatter plot when dealing with large datasets. The hexagonal binning tiles are colored according to the number of points in those bins, and we used a logarithmic color scale.

Randomization tests

Given a dataset in which we suspect that the measurement is spatio-temporally correlated, we can numerically test the hypothesis that the statistics of CEs detected by ARCOS are different from the statistics obtained in a system without space–time correlations. To achieve that, we repeatedly randomize the original binarized time-lapse data and calculate the statistics, e.g., the size or duration of CEs. The P value of a one-sided, unpaired test for the mean is then the fraction of cases when the statistic from randomized data was at least as extreme as the observed statistic. For an upper-tailed test, we have (Davison and Hinkley, 1997; Good, 2005; Phipson and Smyth, 2010):

$$p = Pr(T \geq t | H_0 \text{ is true}) = \frac{1 + \#\{t_r \geq t\}}{N_r + 1}, \quad (1)$$

where T is the test statistic, t is the observed value of the test statistic, and N_r is the number of independent randomizations from which we obtain the values t_r of the test statistic T.

We propose five randomization implementations in ARCOS R/Python packages to simulate the null distribution of CEs' statistics (Fig. S1). The methods vary regarding assumptions about the null hypothesis. The first shuffles the entire time series in space between initial spatial positions of existing cells (Fig. S1 B). This is the recommended method as it only disturbs the spatial component while conserving individual cells' measurement dynamics and the population-averaged activity over time. However, since this approach moves the entire time series to a new location, they may end up outside the initial FOV. This may happen when cells migrate, and the new initial position after shuffling is close to the FOV's border. If that is the case, we recommend other approaches: randomly shifting the measurement for every time series back and forth in time (Fig. S1 C), shuffling whole activity blocks per time series (Fig. S1 D), or shuffling individual time

points along every time series (Fig. S1 E). All three methods preserve the total activity in individual time series and leave them in their original positions but affect the onset (randomly shifting the time series) or randomize the activity dynamics (shuffling blocks or time points of activity), thus potentially disrupting correlations between the neighbors. If these methods are insufficient to rearrange the spatio-temporal pattern, e.g., for a single large collective activation, they can be combined with the last approach, where measurements are shuffled between the cells independently for every frame (Fig. S1 F). Note that such complete randomization changes the tested question, for the observed pattern is compared to a uniform distribution of activations in space and time. It is likely that the observed statistics will be sufficiently away from such a null distribution to reject the null hypothesis at a given significance level but will not be extreme enough when the time series are only shuffled in space with their dynamics intact.

The choice of the randomization method depends on the type of locality in the observed phenomenon. If cell activity is local in space but uniform in time, e.g., sustained or synchronized oscillatory activation of a cell cluster, then shuffling positions of entire trajectories or shuffling positions per time point would disrupt the spatial correlation. However, when the observed phenomenon is local in time but uniform in space, e.g., a synchronized, transient activation pulse of all cells in the FOV, position shuffling will not affect such an activity cluster. Instead, shifting trajectories in time, shuffling time points, or blocks of activity per trajectory would be required to disrupt the temporal locality.

Synthetic datasets and simulations
The implementation of the ARCOS package in R offers two functions to generate a single or multiple CEs on a 2D lattice over time. genSynthSingle2D creates a single event with predetermined cell activations over time. genSynthMultiple2D generates a random number of concentrically growing circles with a random duration at random points in space and time. Additionally, we used NetLogo, a multiagent programmable modeling environment (Wilensky, 1999), to simulate coordinated ERK activation in a 2D epithelium (Figs. S1 and S5). Configuration files to reproduce the simulations are in supplementary data (Dobrzyński, 2023).

Quantification of global correlation
In Fig. 2, E, G, and I; Fig. 3 H; and Fig. 4, B, E, and H, a global Moran's I autocorrelation coefficient was calculated according to the method described in Gittleman and Kot (1990). We applied the Moran.I function from the R package ape (Paradis and Schliep, 2019) to the (x,y,m) point pattern and calculated the coefficient independently for every time frame t. From the observed Moran's I, we subtracted the expected value of I under the null hypothesis that there was no correlation. Quantification with the global correlation coefficient is useful only for cases when a single (or a few) CE occurs in a FOV.

Sensitivity to image degradation
Imaging artifacts and experimental noise contribute to segmentation errors that can impair ARCOS analysis. To evaluate how image quality affects ARCOS, we added Gaussian white noise to

images before segmenting images with stardist (Schmidt et al., 2018), tracking with trackpy (Allan et al., 2021), and quantifying ERK activity (Fig. S4, A–D). To formalize the process of adding noise, we calculated the amount of noise needed to reach a specific signal-to-noise ratio (SNR):

$$SNR[dB] = 10\log_{10}\left(\frac{P_{Signal}}{P_{Noise}}\right), \qquad (2)$$

where P_{Signal} is the average, squared pixel intensity over the entire time-lapse. For a random Gaussian variable with $\mu = 0$, the average power is equal to its variance, hence $P_{Noise} = \sigma^2$. Solving Eq. 2 for P_{Noise} yields the required σ for a given SNR. ERK activity was quantified in the original, unperturbed image channel but with the degraded segmentation to assess the effect of segmentation degradation alone, without affecting the ERK measurement. This procedure was carried out for a range of SNR values and repeated 10 times to account for the randomness of drawing from a normal distribution.

For each time point that contained a CE, either in the original data or in the current SNR iteration, we counted the number of cells assigned correctly to an event (true positives), incorrectly to an event (false positives), or that should have been assigned but were missed (false negatives). Cells were identified by their unique track IDs. These numbers were cumulated for all time points and used to calculate:

$$precision = \frac{True\ positives}{True\ positives + False\ positives} \qquad (3)$$

and

$$recall = \frac{True\ positives}{True\ positives + False\ negatives} \qquad (4)$$

for a given SNR iteration and averaged over 10 repeats.

To test the sensitivity of CE detection on image quality, we used a time-lapse of an ERK activity wave in MDCK epithelial cells stably expressing the EKAREV-NLS FRET biosensor (Fig. S4 A and Video 10). We report data degradation using SNR expressed in decibels or dB (Eq. 2). A $SNR > 0$ dB indicates that the signal power, i.e., the average of squares of all pixel intensities over time, is greater than the power of added noise; $SNR < 0$ dB is when the power of noise dominates the power of the original data. The higher the ratio, the cleaner the signal, with $+\infty$ corresponding to original, measured images.

For a chosen set of segmentation parameters in stardist (Schmidt et al., 2018), we observed that nuclei segmentation was severely affected only when $SNR < -10$ dB (Fig. S4 B). Below that SNR, higher noise in images resulted in severely deteriorated time series (Fig. S4 C) to a point when ARCOS could no longer detect CEs (Fig. S4 D). The number of false negatives, i.e., cells that were not assigned to true CEs increased sharply for $SNR < -10$ dB, which corresponds to noise power being 10× the power of the original signal. The number of false positives, i.e., cells that were erroneously assigned to clusters, increased only slightly at $SNR < -10$ dB (Fig. S4 G). Overall, the image segmentation and cell tracking pipeline can tolerate noisy images for $SNR > -10$ dB before recall (Eq. 4) or the number of relevant items retrieved decreases (Fig. S4 G). The code to reproduce this

Gagliardi et al.
Automatic detection of collective signaling

Journal of Cell Biology 14 of 17
https://doi.org/10.1083/jcb.202207048

analysis is available in the accompanying data repository (Dobrzyński, 2023).

Sensitivity to data degradation

To study the effect of noise on ARCOS analysis, we added Gaussian white noise with an increasing amplitude to single-cell ERK activity from the segmentation of the wave in MDCK cells (Fig. S4, E and F). ARCOS was then run on the newly generated dataset and compared with the original output as above. This was carried out for a range of SNRs and repeated 10 times.

CEs were faithfully detected only down to $SNR \approx 20$ dB (Fig. S4 F), which corresponds to the power of the original time series being $100\times$ higher than the power of the added Gaussian noise. For lower SNR, the number of false negatives soared, which caused the recall to drop sharply (Fig. S4 H). We conclude that for the parameters in our pipeline, CE detection is more sensitive to degradation of time series rather than images.

Sensitivity to parameters

The core ARCOS algorithm clusters active cells in individual frames of the time-lapse and then tracks the clusters. The key parameter for spatial clustering is the neighborhood radius, ε. It should be greater than the first NND between active cells to let the *dbscan* algorithm form clusters. If too large, nearby independent CEs are included in a single cluster; too small, and a single CE is identified as several smaller events. We investigated the quality of event detection as a function of ε by simulating a synthetic time-lapse of 200 frames in NetLogo. We simulated 69 radially expanding, collective activations randomly spaced on a 2D hexagonal mesh, like ERK waves observed in the MCF10A epithelium (Fig. S5 A). The simulation established a ground truth with a known number (69 events), size (37 unique cells), and duration (9 time frames) of CEs. The events were identified correctly for a range of $\varepsilon \in [1.5, 4] \times$ NND (Fig. S5, B–F), bar several small errors due to overlaps between simulated events. For $\varepsilon < 1.5$, cells were not clustered together and formed their own clusters; for $\varepsilon > 4$, the algorithm identified large clusters that comprised many independent events. These results hint at a general rule for processing spatial datasets: ε ought to be slightly larger than the NND, but care must be taken to set it low enough to avoid bundling nearby events.

Online supplemental material

Fig. S1 presents various randomization techniques used to evaluate the accuracy of the ARCOS output. Fig. S2 depicts statistics derived from randomization tests of collective events detected in wild-type, KRAS[G12V], and PIK3CA[H1047R] cell lines. Fig. S3 depicts the same statistics as in Fig. S2 but applied to batimastat treatment on the same cell lines. Fig. S4 explores the sensitivity of ARCOS to image or time-series degradation. Fig. S5 portrays the algorithm's sensitivity to the search radius ε and the total binarized activity. Video 1 shows the features implemented in the ARCOS napari plugin. Video 2 displays the detection of a growing circular region generated in cells with the OptoRAF/ERK-KTR actuator/biosensor circuit. Video 3 displays a moving circle generated in cells with the same technique. Video 4 displays a splitting circular region generated in cells

with the same technique. Video 5 shows ARCOS analysis of collective ERK activity waves in MCF10A WT, KRAS[G12V], and PIK3CA[H1047R] epithelial cells. Video 6 shows collective ERK activity waves in starved NRK-52E epithelial cells detected with ARCOS. Video 7 displays collective calcium waves identified by ARCOS in starved MDCK epithelial cells expressing the calcium reporter GCaMP6S. Video 8 demonstrates the ability of ARCOS to analyze shimmering waves in a giant honeybee colony. Video 9 showcases collective 3D ERK activity waves, identified by ARCOS in a mammary acinus in vitro. Video 10 demonstrates ARCOS's sensitivity to iterative image degradation with additive Gaussian white noise.

Data availability

The data, R notebooks to reproduce the plots and randomization tests, configurations of simulations, and Jupyter notebooks to show the results in the napari image viewer were uploaded to Mendeley Data (Dobrzyński, 2023). ARCOS can be downloaded from https://arcos.gitbook.io as an R or a python package or a napari plugin together with extensive documentation.

References

Aikin, T.J., A.F. Peterson, M.J. Pokrass, H.R. Clark, and S. Regot. 2020. MAPK activity dynamics regulate non-cell autonomous effects of oncogene expression. *Elife*. 9:e60541. https://doi.org/10.7554/eLife.60541

Albeck, J.G., G.B. Mills, and J.S. Brugge. 2013. Frequency-modulated pulses of ERK activity transmit quantitative proliferation signals. *Mol. Cell*.49: 249–261. https://doi.org/10.1016/j.molcel.2012.11.002

Alcantara, F., and M. Monk. 1974. Signal propagation during aggregation in the slime mould *Dictyostelium discoideum*. *J. Gen. Microbiol*. 85:321–334. https://doi.org/10.1099/00221287-85-2-321

Allan, D.B., T. Caswell, N.C. Keim, C.M. van der Wel, and R.W. Verweij. 2021. soft-matter/trackpy: Trackpy v0.5.0. *Zenodo*. https://doi.org/10.5281/ZENODO.4682814

Anselin, L. 1995. Local indicators of spatial association-LISA. *Geogr. Anal*. 27: 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Aoki, K., Y. Kondo, H. Naoki, T. Hiratsuka, R.E. Itoh, and M. Matsuda. 2017. Propagating wave of ERK activation orients collective cell migration. *Dev. Cell*. 43:305–317.e5. https://doi.org/10.1016/j.devcel.2017.10.016

Aoki, K., Y. Kumagai, A. Sakurai, N. Komatsu, Y. Fujita, C. Shionyu, and M. Matsuda. 2013. Stochastic ERK activation induced by noise and cell-to-cell propagation regulates cell density-dependent proliferation. *Mol. Cell*. 52:529–540. https://doi.org/10.1016/j.molcel.2013.09.015

Arsa, D.M.S., and A.A.N.H. Susila. 2019. VGG16 in Batik classification based on random forest. In 2019 International Conference on Information Management and Technology (ICIMTech). IEEE, Jakarta/Bali, Indonesia. 295–299.

Berg, S., D. Kutra, T. Kroeger, C.N. Straehle, B.X. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy, et al. 2019. ilastik: Interactive machine learning for (bio)image analysis. *Nat. Methods*. 16:1226–1232. https://doi.org/10.1038/s41592-019-0582-9

Bivand, R.S., E.J. Pebesma, and V. Gómez-Rubio. 2013. Applied Spatial Data Analysis with R. Second edition. Springer, New York, NY.

Campello, R.J.G.B., D. Moulavi, and J. Sander. 2013. Density-based clustering based on hierarchical density estimates. In Advances in Knowledge Discovery and Data Mining. J. Pei, V.S. Tseng, L. Cao, H. Motoda, and G. Xu, editors. Springer Berlin Heidelberg, Berlin, Heidelberg. 160–172. https://doi.org/10.1007/978-3-642-37456-2_14

Chang, J.B., and J.E. Ferrell Jr. 2013. Mitotic trigger waves and the spatial coordination of the *Xenopus* cell cycle. *Nature*. 500:603–607. https://doi.org/10.1038/nature12321

Choi, W.-G., M. Toyota, S.-H. Kim, R. Hilleary, and S. Gilroy. 2014. Salt stress-induced Ca²⁺ waves are associated with rapid, long-distance root-to-shoot signaling in plants. *Proc. Natl. Acad. Sci. USA*. 111:6497–6502. https://doi.org/10.1073/pnas.1319955111

Cohen, J. 2013. Statistical Power Analysis for the Behavioral Sciences. Zero edition. Routledge, England, UK.

Crocker, J.C., and D.G. Grier. 1996. Methods of digital video microscopy for colloidal studies. *J. Colloid Interf. Sci*. 179:298–310. https://doi.org/10.1006/jcis.1996.0217

Davison, A.C., and D.V. Hinkley. 1997. Bootstrap Methods and Their Application. First edition. Cambridge University Press, Cambridge, England.

De Simone, A., M.N. Evanitsky, L. Hayden, B.D. Cox, J. Wang, V.A. Tornini, J. Ou, A. Chao, K.D. Poss, and S. Di Talia. 2021. Control of osteoblast regeneration by a train of Erk activity waves. *Nature*. 590:129–133. https://doi.org/10.1038/s41586-020-03085-8

Debnath, J., S.K. Muthuswamy, and J.S. Brugge. 2003. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods*. 30:256–268. https://doi.org/10.1016/S1046-2023(03)00032-X

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and F.-F. Li. 2009. ImageNet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Miami, FL. 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Dessauges, C., J. Mikelson, M. Dobrzyński, M.-A. Jacques, A. Frismantiene, P.A. Gagliardi, M. Khammash, and O. Pertz. 2022. Optogenetic actuator: ERK biosensor circuits identify MAPK network nodes that shape ERK dynamics. *Mol. Syst. Biol*. 18:e10670. https://doi.org/10.15252/msb.202110670

Dobrzyński, M. 2023. ARCOS, a computational approach for automatic recognition of collective signalling. *Mendeley Data*. https://doi.org/10.17632/8GCNCG6ZKT.3

Ender, P., P.A. Gagliardi, M. Dobrzyński, A. Frismantiene, C. Dessauges, T. Höhener, M.-A. Jacques, A.R. Cohen, and O. Pertz. 2022. Spatiotemporal control of ERK pulse frequency coordinates fate decisions during mammary acinar morphogenesis. *Dev. Cell*. 57:2153–2167.e6. https://doi.org/10.1016/j.devcel.2022.08.008

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu. 1996.A density-based algorithm for discovering clusters in large spatial databases with noise. AAAI Press, Washington, USA. 226–231.

Fritsche-Guenther, R., F. Witzel, A. Sieber, R. Herr, N. Schmidt, S. Braun, T. Brummer, C. Sers, and N. Blüthgen. 2011. Strong negative feedback from Erk to Raf confers robustness to MAPK signalling. *Mol. Syst. Biol*. 7: 489. https://doi.org/10.1038/msb.2011.27

Gagliardi, P.A., M. Dobrzyński, M.-A. Jacques, C. Dessauges, P. Ender, Y. Blum, R.M. Hughes, A.R. Cohen, and O. Pertz. 2021. Collective ERK/Akt activity waves orchestrate epithelial homeostasis by driving apoptosis-induced survival. *Dev. Cell*. 56:1712–1726.e6. https://doi.org/10.1016/j.devcel.2021.05.007

Gelens, L., G.A. Anderson, and J.E. Ferrell Jr. 2014. Spatial trigger waves: Positive feedback gets you a long way. *MBoC*. 25:3486–3493. https://doi.org/10.1091/mbc.e14-08-1306

Getis, A., and J.K. Ord. 1992. The analysis of spatial association by use of distance statistics. *Geogr. Anal*. 24:189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x

Gilland, E., A.L. Miller, E. Karplus, R. Baker, and S.E. Webb. 1999. Imaging of multicellular large-scale rhythmic calcium waves during zebrafish gastrulation. *Proc. Natl. Acad. Sci. USA*. 96:157–161. https://doi.org/10.1073/pnas.96.1.157

Gillies, T.E., M. Pargett, M. Minguet, A.E. Davies, and J.G. Albeck. 2017. Linear integration of ERK activity predominates over persistence detection in Fra-1 regulation. *Cell Syst*. 5:549–563.e5. https://doi.org/10.1016/j.cels.2017.10.019

Gittleman, J.L., and M. Kot. 1990. Adaptation: Statistics and a null model for estimating phylogenetic effects. *Syst. Zool*. 39:227–241. https://doi.org/10.2307/2992183

Good, P. 2005. Permutation, Parametric and Bootstrap Tests of Hypotheses. Springer-Verlag, New York.

Gustin, J.P., B. Karakas, M.B. Weiss, A.M. Abukhdeir, J. Lauring, J.P. Garay, D. Cosgrove, A. Tamaki, H. Konishi, Y. Konishi, et al. 2009. Knockin of mutant PIK3CA activates multiple oncogenic pathways. *Proc. Natl. Acad. Sci. USA*. 106:2835–2840. https://doi.org/10.1073/pnas.0813351106

Hahsler, M., M. Piekenbrock, and D. Doran. 2019. Dbscan : Fast density-based clustering with R. J. *J. Stat. Softw*. 91. https://doi.org/10.18637/jss.v091.i01

Hino, N., L. Rossetti, A. Marín-Llauradó, K. Aoki, X. Trepat, M. Matsuda, and T. Hirashima. 2020. ERK-mediated mechanochemical waves direct collective cell polarization. *Dev. Cell*. 53:646–660.e8. https://doi.org/10.1016/j.devcel.2020.05.011

Hiratsuka, T., Y. Fujita, H. Naoki, K. Aoki, Y. Kamioka, and M. Matsuda. 2015. Intercellular propagation of extracellular signal-regulated kinase activation revealed by in vivo imaging of mouse skin. *Elife*. 4:e05178. https://doi.org/10.7554/eLife.05178

Hodgkin, A.L., A.F. Huxley, and B. Katz. 1952. Measurement of current-voltage relations in the membrane of the giant axon of Loligo. *J. Physiol*. 116:424–448. https://doi.org/10.1113/jphysiol.1952.sp004716

Huang, L., Z. Guo, F. Wang, and L. Fu. 2021. KRAS mutation: From undruggable to druggable in cancer. *Signal Transduct. Target. Ther*. 6:386. https://doi.org/10.1038/s41392-021-00780-4

Jaqaman, K., D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S.L. Schmid, and G. Danuser. 2008. Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods*. 5:695–702. https://doi.org/10.1038/nmeth.1237

Kastberger, G., T. Hoetzl, M. Maurer, I. Kranner, S. Weiss, and F. Weihmann. 2014a. Speeding up social waves. Propagation mechanisms of shimmering in giant honeybees. *PLoS One*. 9:e86315. https://doi.org/10.1371/journal.pone.0086315

Kastberger, G., M. Maurer, F. Weihmann, M. Ruether, T. Hoetzl, I. Kranner, and H. Bischof. 2011. Stereoscopic motion analysis in densely packed clusters: 3D analysis of the shimmering behaviour in giant honey bees. *Front. Zool.* 8:3. https://doi.org/10.1186/1742-9994-8-3

Kastberger, G., E. Schmelzer, and I. Kranner. 2008. Social waves in giant honeybees repel hornets. *PLoS One.* 3:e3141. https://doi.org/10.1371/journal.pone.0003141

Kastberger, G., F. Weihmann, and T. Hoetzl. 2010. Complex social waves of giant honeybees provoked by a dummy wasp support the special-agent hypothesis. *Commun. Integr. Biol.* 3:179–180. https://doi.org/10.4161/cib.3.2.10809

Kastberger, G., F. Weihmann, and T. Hoetzl. 2013. Social waves in giant honeybees (Apis dorsata) elicit nest vibrations. *Naturwissenschaften.* 100:595–609. https://doi.org/10.1007/s00114-013-1056-z

Kastberger, G., F. Weihmann, T. Hoetzl, S.E. Weiss, M. Maurer, and I. Kranner. 2012. How to join a wave: Decision-making processes in shimmering behavior of giant honeybees (Apis dorsata). *PLoS One.* 7: e36736. https://doi.org/10.1371/journal.pone.0036736

Kastberger, G., F. Weihmann, M. Zierler, and T. Hötzl. 2014b. Giant honeybees (Apis dorsata) mob wasps away from the nest by directed visual patterns. *Naturwissenschaften.* 101:861–873. https://doi.org/10.1007/s00114-014-1220-0

Konishi, H., B. Karakas, A.M. Abukhdeir, J. Lauring, J.P. Gustin, J.P. Garay, Y. Konishi, E. Gallmeier, K.E. Bachman, and B.H. Park. 2007. Knock-in of mutant K-ras in nontumorigenic human epithelial cells as a new model for studying K-ras mediated transformation. *Cancer Res.* 67:8460–8467. https://doi.org/10.1158/0008-5472.CAN-07-0108

Lawson, A. 2006. Statistical Methods in Spatial Epidemiology. Second edition. Wiley, Chichester, UK; Hoboken, NJ. 398 pp.

Manly, B.F.J. 2007. Randomization, Bootstrap, and Monte Carlo Methods in Biology. Third edition. Chapman & Hall/CRC, Boca Raton, FL. 455 pp.

McQuin, C., A. Goodman, V. Chernyshev, L. Kamentsky, B.A. Cimini, K.W. Karhohs, M. Doan, L. Ding, S.M. Rafelski, D. Thirstrup, et al. 2018. CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* 16:e2005970. https://doi.org/10.1371/journal.pbio.2005970

Minjgee, M., M. Toulany, R. Kehlbach, K. Giehl, and H.P. Rodemann. 2011. K-RAS(V12) induces autocrine production of EGFR ligands and mediates radioresistance through EGFR-dependent akt signaling and activation of DNA-PKcs. *Int. J. Radiat. Oncol. Biol. Phys.* 81:1506–1514. https://doi.org/10.1016/j.ijrobp.2011.05.057

Moore, D.A., and T.E. Carpenter. 1999. Spatial analytical methods and geographic information systems: Use in health research and epidemiology. *Epidemiol. Rev.* 21:143–161. https://doi.org/10.1093/oxfordjournals.epirev.a017993

Naimi, B., N.A.S. Hamm, T.A. Groen, A.K. Skidmore, A.G. Toxopeus, and S. Alibakhshi. 2019. ELSA: Entropy-based local indicator of spatial association. *Spat. Stat.* 29:66–88. https://doi.org/10.1016/j.spasta.2018.10.001

Noble, D. 1962. A modification of the Hodgkin--Huxley equations applicable to Purkinje fibre action and pace-maker potentials. *J. Physiol.* 160: 317–352. https://doi.org/10.1113/jphysiol.1962.sp006849

Pachitariu, M., and C. Stringer. 2022. Cellpose 2.0: How to train your own model. *Nat. Methods.* 19:1634–1641. https://doi.org/10.1038/s41592-022-01663-4

Paradis, E., and K. Schliep. 2019. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics.* 35:526–528. https://doi.org/10.1093/bioinformatics/bty633

Pfeiffer, D.U., T.P. Robinson, M. Stevenson, K.B. Stevens, D.J. Rogers, and A.C.A. Clements. 2008. Spatial Analysis in Epidemiology. Oxford University Press, Oxford, UK.

Phipson, B., and G.K. Smyth. 2010. Permutation P values should never be zero: Calculating exact P values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol.* 9:e39. https://doi.org/10.2202/1544-6115.1585

Pond, K.W., O. Alkhimenok, J. Chakrabarti, Y. Zavros, C.A. Thorne, and A.L. Paek. 2022. Live-cell imaging in human colonic monolayers reveals Erk waves limit the stem cell compartment to maintain epithelial homeostasis. *Elife.* 11:e78837. https://doi.org/10.7554/eLife.78837

Purvis, J.E., and G. Lahav. 2013. Encoding and decoding cellular information through signaling dynamics. *Cell.* 152:945–956. https://doi.org/10.1016/j.cell.2013.02.005

Radloff, S.E., H.R. Hepburn, and M.S. Engel. 2011. The asian species of Apis. In Honeybees of Asia. H.R. Hepburn, and S.E. Radloff, editors. Springer Berlin Heidelberg, Berlin, Heidelberg. 1–22.

Robb-Gaspers, L.D., and A.P. Thomas. 1995. Coordination of Ca²⁺ signaling by intercellular propagation of Ca²⁺ waves in the intact liver. *J. Biol. Chem.* 270:8102–8107. https://doi.org/10.1074/jbc.270.14.8102

Ryu, H., M. Chung, M. Dobrzyński, D. Fey, Y. Blum, S.S. Lee, M. Peter, B.N. Kholodenko, N.L. Jeon, and O. Pertz. 2015. Frequency modulation of ERK activation dynamics rewires cell fate. *Mol. Syst. Biol.* 11:838. https://doi.org/10.15252/msb.20156458

Schmelzer, E., and G. Kastberger. 2009. "Special agents" trigger social waves in giant honeybees (Apis dorsata). *Naturwissenschaften.* 96:1431–1441. https://doi.org/10.1007/s00114-009-0605-y

Schmidt, U., M. Weigert, C. Broaddus, and G. Myers. 2018. Cell detection with star-convex polygons. Medical Image Computing and Computer Assisted Intervention: MICCAI 2018: 21st International Conference, September 16–20, 2018, Granada, Spain. 265–273.

Serra-Picamal, X., V. Conte, R. Vincent, E. Anon, D.T. Tambe, E. Bazellieres, J.P. Butler, J.J. Fredberg, and X. Trepat. 2012. Mechanical waves during tissue expansion. *Nat. Phys.* 8:628–634. https://doi.org/10.1038/nphys2355

Sofroniew, N., T. Lambert, K. Evans, J. Nunez-Iglesias, G. Bokota, M. Bussonnier, G. Peña-Castellanos, P. Winston, K. Yamauchi, D.D. Pop, et al. 2022. napari/napari: 0.4.15rc1. *Zenodo.* https://doi.org/10.5281/ZENODO.3555620

Stringer, C., T. Wang, M. Michaelos, and M. Pachitariu. 2021. Cellpose: A generalist algorithm for cellular segmentation. *Nat. Methods.* 18: 100–106. https://doi.org/10.1038/s41592-020-01018-x

Sturm, O.E., R. Orton, J. Grindlay, M. Birtwistle, V. Vyshemirsky, D. Gilbert, M. Calder, A. Pitt, B. Kholodenko, and W. Kolch. 2010. The mammalian MAPK/ERK pathway exhibits properties of a negative feedback amplifier. *Sci. Signal.* 3:ra90. https://doi.org/10.1126/scisignal.2001212

Takeuchi, Y., R. Narumi, R. Akiyama, E. Vitiello, T. Shirai, N. Tanimura, K. Kuromiya, S. Ishikawa, M. Kajita, M. Tada, et al. 2020. Calcium wave promotes cell extrusion. *Curr. Biol.* 30:670–681.e6. https://doi.org/10.1016/j.cub.2019.11.089

Talia, S.D., and M. Vergassola. 2022. Waves in embryonic development. *Annu. Rev. Biophys.* 51:327–353. https://doi.org/10.1146/annurev-biophys-111521-102500

Tsiairis, C.D., and A. Aulehla. 2016. Self-Organization of embryonic genetic oscillators into spatiotemporal wave patterns. *Cell.* 164:656–667. https://doi.org/10.1016/j.cell.2016.01.028

Wedlich-Söldner, R., and T. Betz. 2018. Self-organization: The fundament of cell biology. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373:20170103. https://doi.org/10.1098/rstb.2017.0103

Wilensky, U. 1999. NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Winter, M., W. Mankowski, E. Wait, S. Temple, and A.R. Cohen. 2016. LEVER: Software tools for segmentation, tracking and lineaging of proliferating cells. *Bioinformatics.* 32:3530–3531. https://doi.org/10.1093/bioinformatics/btw406

Young, C.D., L.J. Zimmerman, D. Hoshino, L. Formisano, A.B. Hanker, M.L. Gatza, M.M. Morrison, P.D. Moore, C.A. Whitwell, B. Dave, et al. 2015. Activating PIK3CA mutations induce an epidermal growth factor receptor (EGFR)/Extracellular signal-regulated kinase (ERK) paracrine signaling axis in basal-like breast cancer. *Mol. Cell. Proteomics.* 14: 1959–1976. https://doi.org/10.1074/mcp.M115.049783
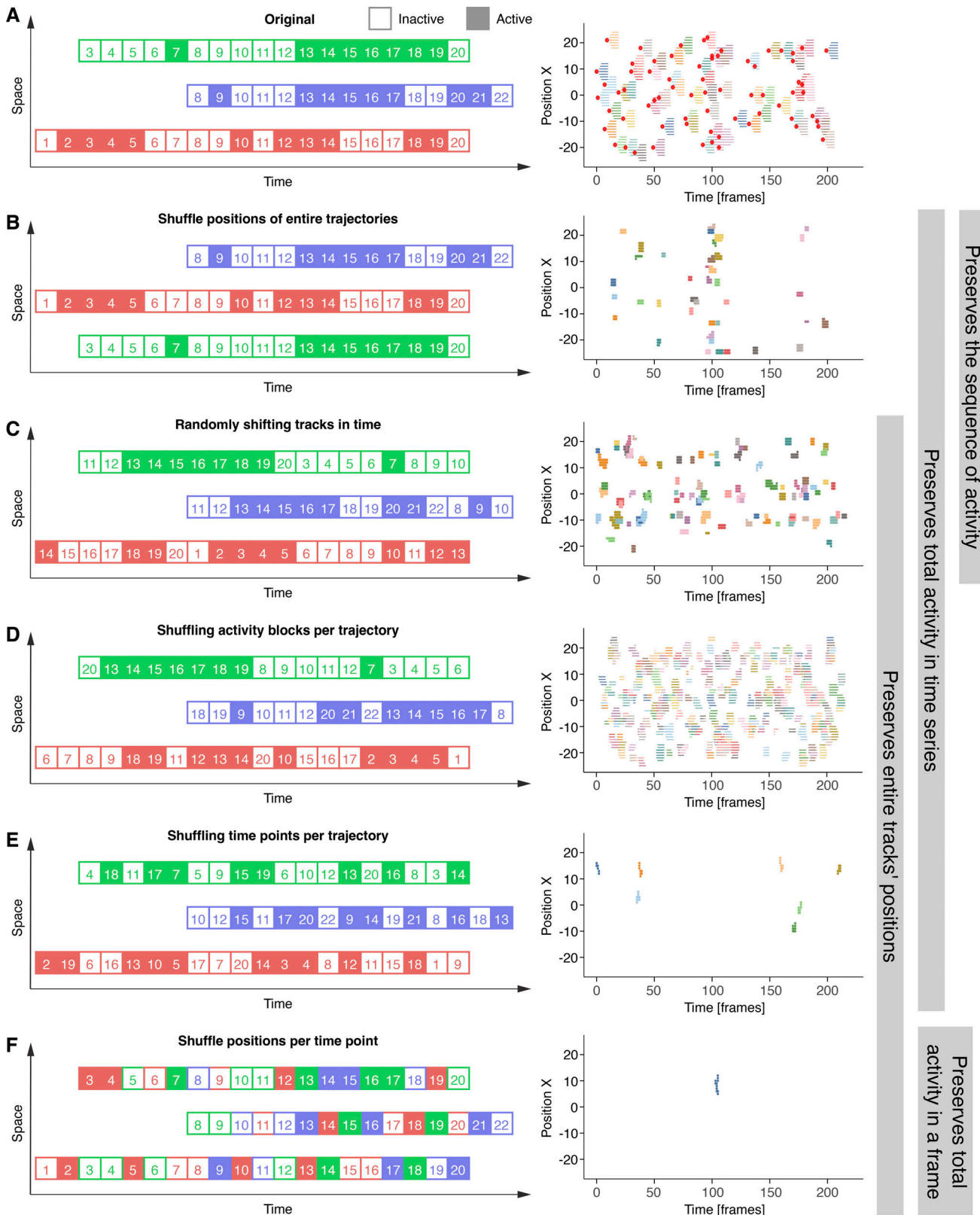
Figure S1. **Measurement randomization methods.** Left: Three sample binarized time series with blocks of cell activity indicated by shaded squares. Right: Spaghetti plots of collective events simulated in NetLogo. **(A)** The original dataset. **(B)** Positions of entire trajectories are shuffled between cells while preserving the original dynamics. **(C)** Time series are independently and randomly shifted back and forth in time while preserving positions of entire tracks. The original dynamics are preserved, although the onset changes. **(D)** Blocks of activity are shuffled in time independently for every trajectory while preserving positions of entire tracks. **(E)** Individual time points are shuffled independently for every trajectory while preserving positions of entire tracks. **(F)** The activity is shuffled between cells independently in every time frame. Positions of entire tracks remain intact.
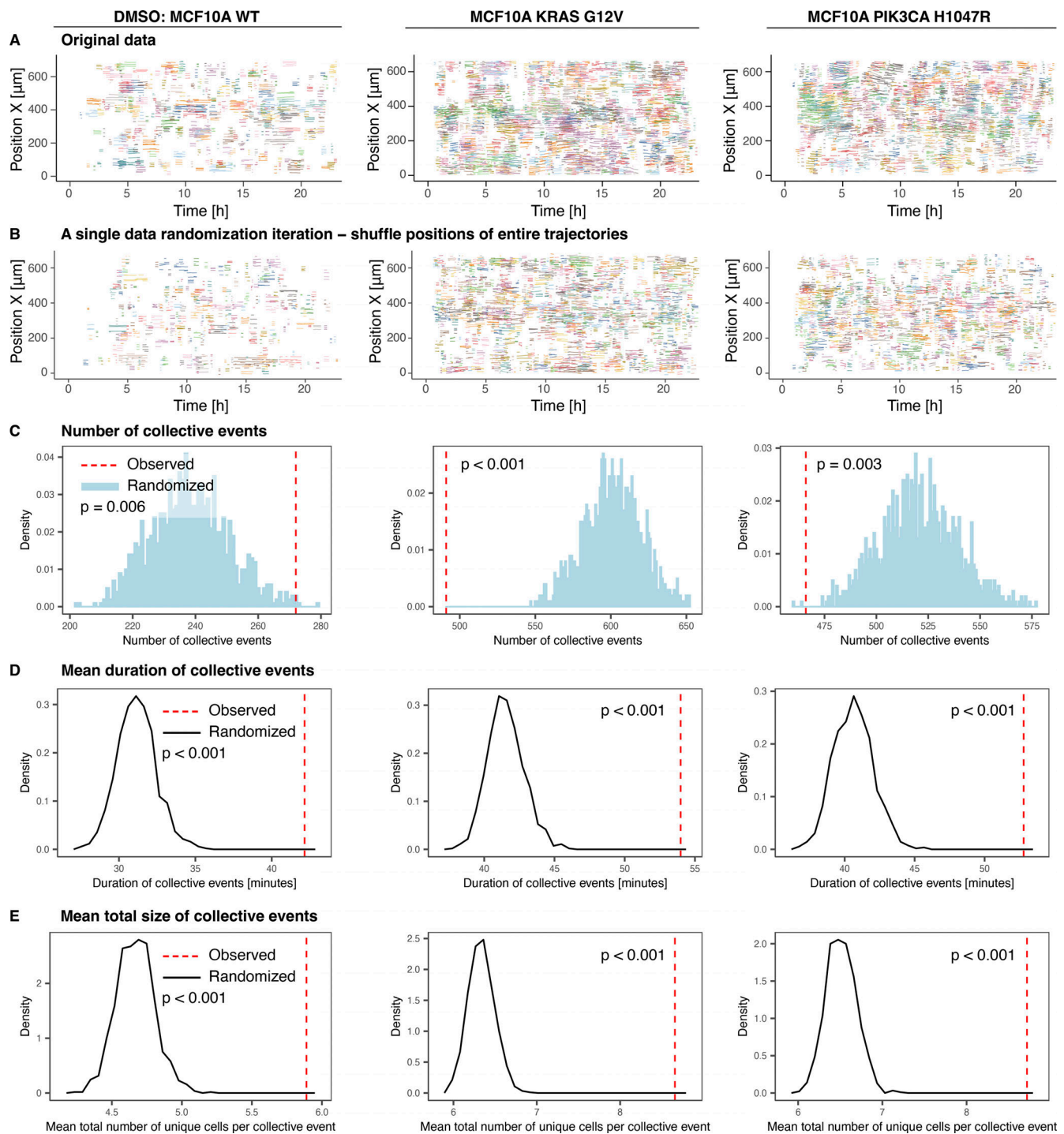
Figure S2. **Statistics of CEs from 1,000 randomization iterations for untreated and starved MCF10A WT, and KRAS[G12V] and PIK3CA[H1047R] mutant cell lines. (A)** Spaghetti plots of CEs identified in the original data from Fig. 3. **(B)** Spaghetti plots of CEs identified in a single iteration of randomization, where positions of binarized time series were shuffled according to Fig. S1 B. **(C–E)** Distributions of statistics calculated from 1,000 randomization iterations compared to the observed statistic indicated by the vertical dashed line: the number of CEs (C), the mean duration of CEs (D), the mean total number of unique cells in a CE, i.e., the "size" of a CE (E). The P value was calculated according to Eq. 1 as the fraction of cases when the statistic was at least as extreme as observed.
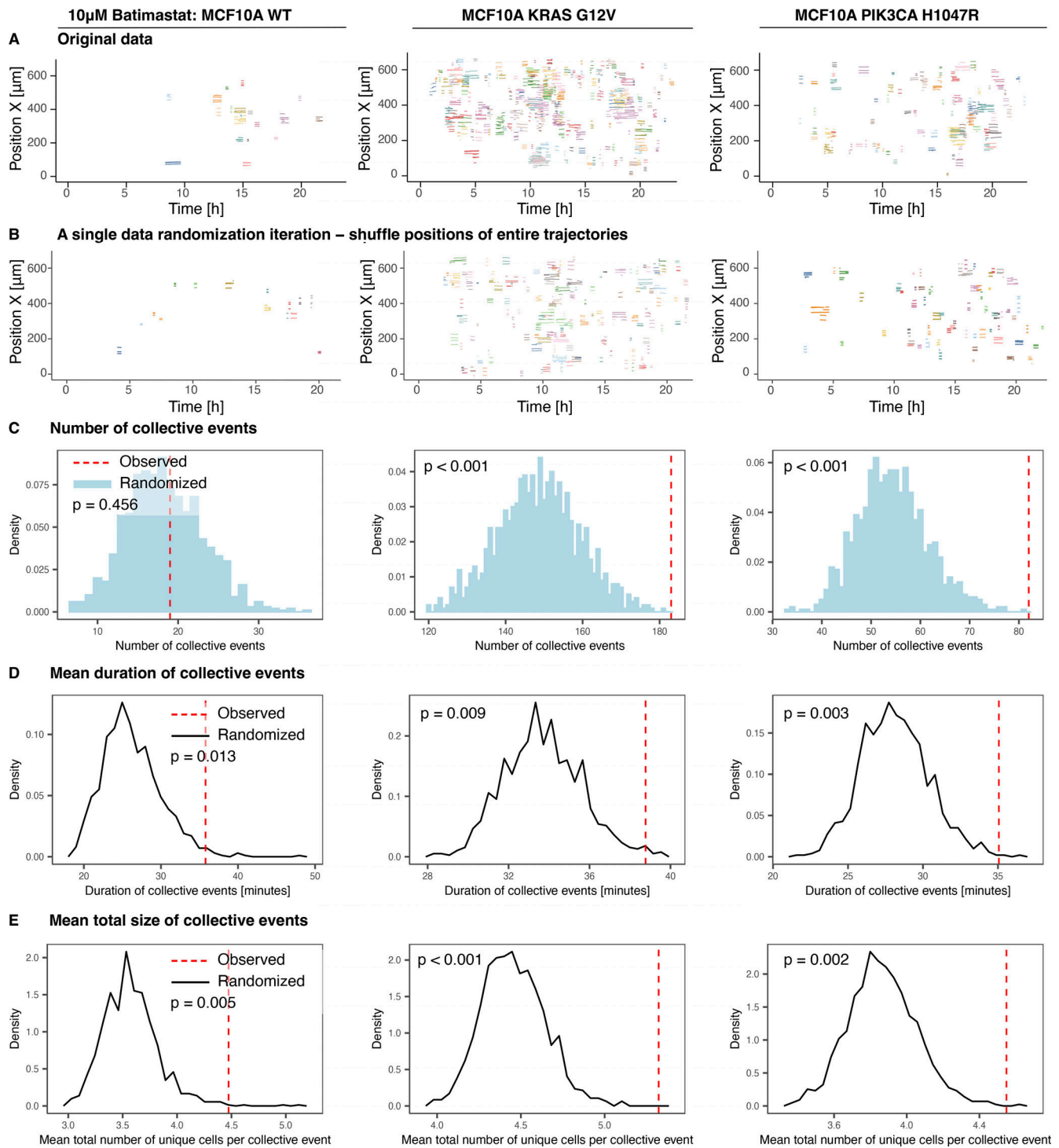
Figure S3.   **Statistics of CEs from 1,000 randomization iterations for starved MCF10A WT, and KRAS[G12V] and PIK3CA[H1047R] mutant cell lines treated with 10 µM batimastat. (A–E)** As in Fig. S2.
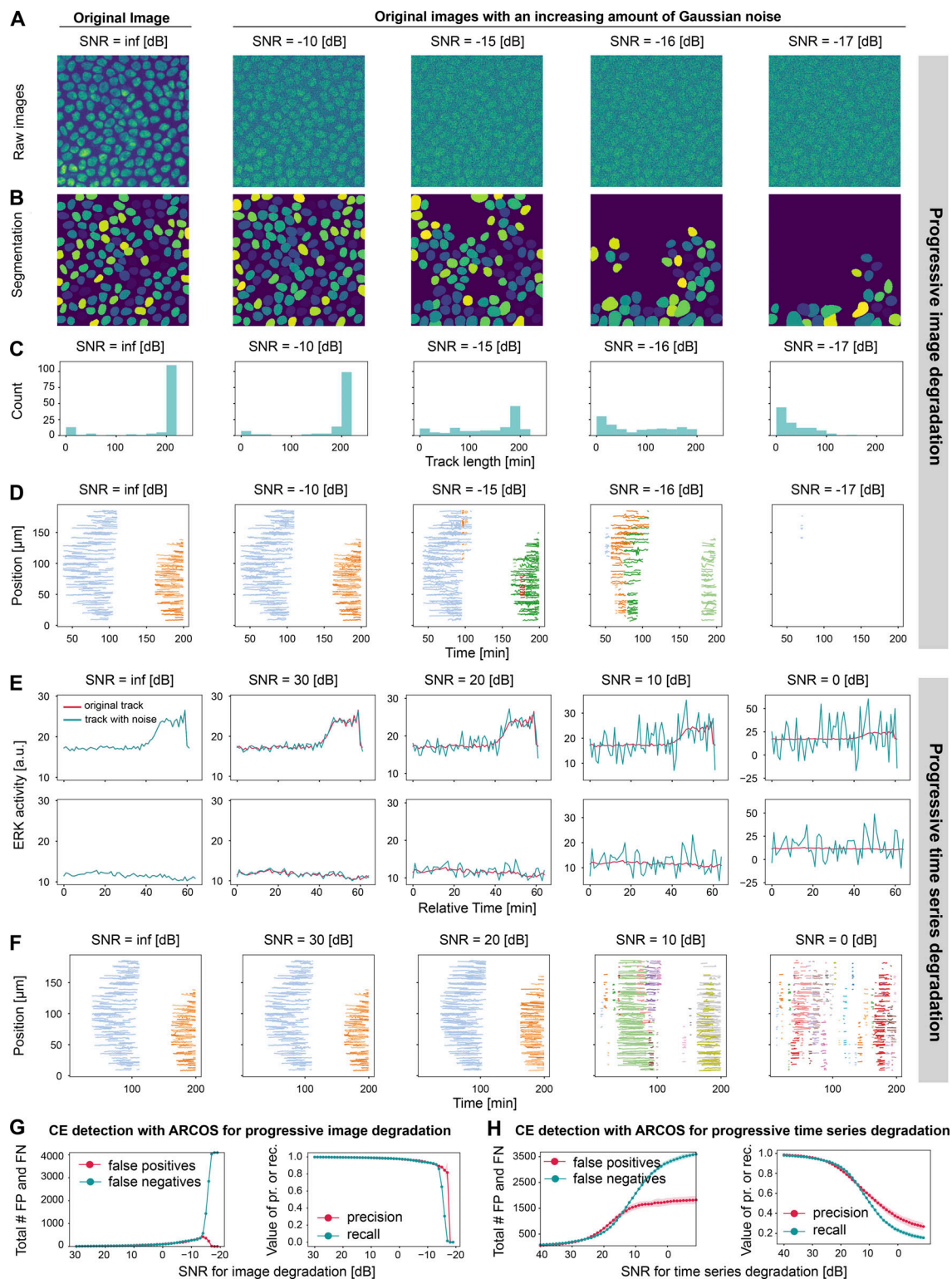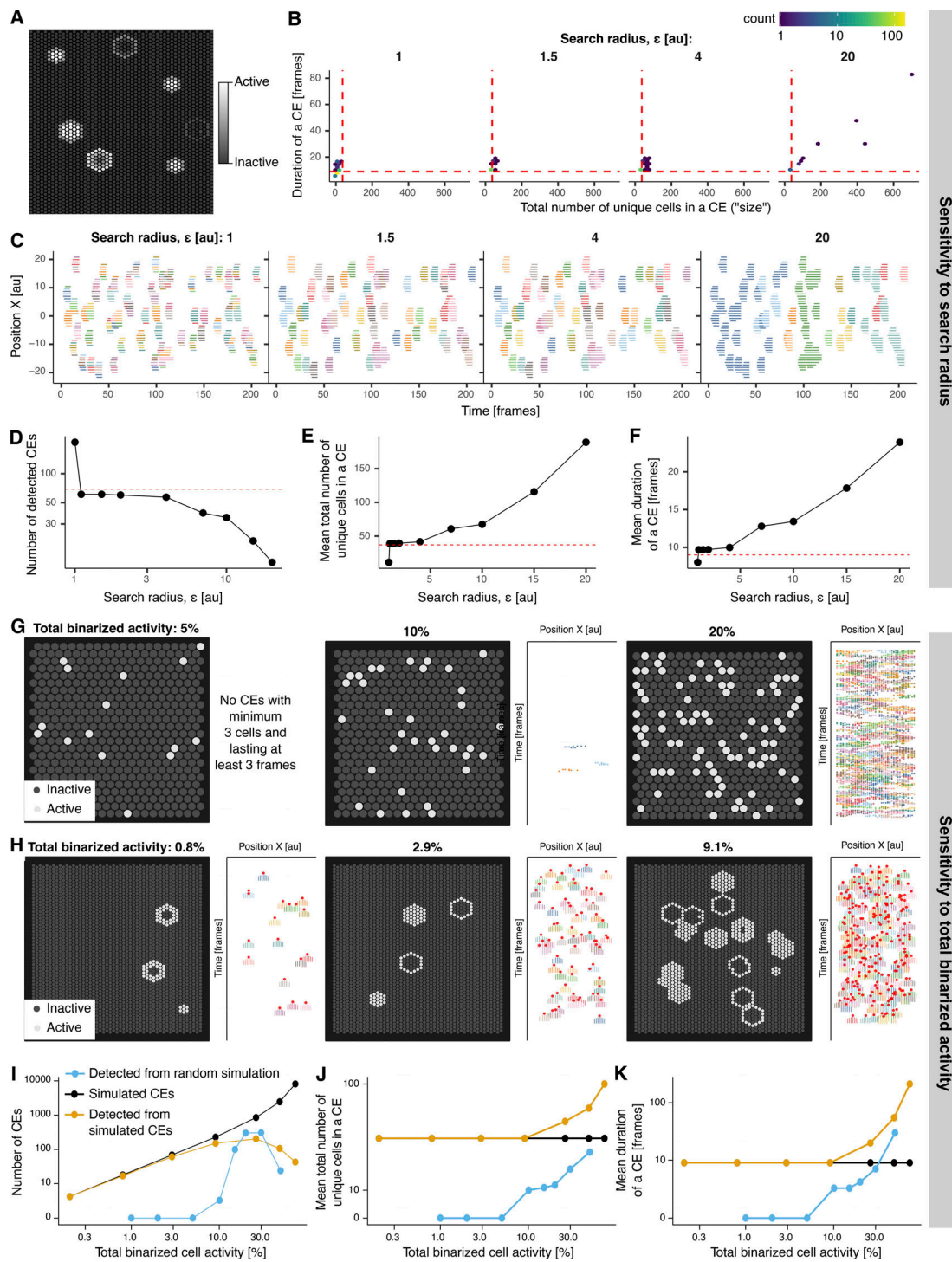
Figure S4. **Evaluation of ARCOS sensitivity to iterative image and time series degradation with additive Gaussian white noise. (A)** Sample images of the sum of donor and acceptor channels of the EKAREV-NLS FRET biosensor in MDCK cells with an increasing amount of Gaussian white noise. The amount of added noise corresponds to a target SNR in dB according to Eq. 2 (Materials and methods). **(B)** Segmentation with the stardist python package. **(C)** Track-length histograms of respective iterations. **(D)** Spaghetti plots depicting CEs detected at a specific SNR. **(E)** Sample single-cell ERK trajectories of MDCK cells measured using the EKAREV-NLS FRET biosensor. An increasing amount of Gaussian white noise was added to the time series. The noise amplitude is calculated according to Eq. 2 (Materials and methods) to reach a target SNR in dB. **(F)** Spaghetti plots depicting CEs detected at a specific SNR. **(G)** Total number of false positives (FP), false negatives (FN), and precision and recall for detecting CEs with ARCOS for progressive image degradation from panels A–D. Calculated by checking if a specific cell was detected as part of a CE at a specific time point. **(H)** Same as panel G but for progressive time series degradation from panels E and F. **(G and H)** Shaded regions represent ±1 SD of 10 random noise additions.

Gagliardi et al.
Automatic detection of collective signaling

Journal of Cell Biology   S5
https://doi.org/10.1083/jcb.202207048

Figure S5. **Sensitivity of CE identification to search radius length and TBA. (A)** A single frame from a NetLogo simulation of collective cell activation on a 2D hexagonal lattice with distance a = 1. Cells assume a gradient of activity levels from inactive (dark gray) to active (light gray). We simulated 69 activation events placed randomly in space and time. Each event lasts nine time frames and comprises 37 cells. **(B)** Size vs. duration of detected CEs for a range of search radius ε. The minimum CE detection size was three cells, with a minimum duration of three frames. Red dashed lines indicate the simulated size and duration. **(C)** Spaghetti plots of CEs detected in simulations. **(D–F)** The number of CEs detected in the simulation, their size, and duration. Red dashed lines indicate the true properties of simulated CEs, i.e., 69 events, 37 cells, 9 frames. **(G)** Single frames of simulated random cell activations on a 2D hexagonal lattice for a range of TBA fractions. Cells assume active or inactive states; active cells are uniformly distributed in space and time. Spaghetti plots visualize CEs detected in simulations. ARCOS parameters are: ε = 1.2× the lattice distance, the minimum cluster size is three cells, and the minimum cluster duration is three frames. **(H)** Single frames from NetLogo simulations of collective cell activation on a 2D hexagonal lattice for a range of TBA fractions. Each simulated CE involved 37 unique cells and lasted nine frames. Spaghetti plots underneath the frames visualize CEs detected in simulations. ARCOS parameters are: ε = 1.5× the lattice distance, the minimum cluster size is three cells, and the minimum cluster duration is three frames. **(I–K)** Comparison of statistics of CEs detected in a simulation of random activations in panel G, and CEs simulated and detected in a simulation in panel H.

Gagliardi et al.
Automatic detection of collective signaling

Journal of Cell Biology   S6
https://doi.org/10.1083/jcb.202207048

Video 1. **Video demonstration of features implemented in the interactive napari plugin.** Showcases loading and preprocessing data, choosing parameters to detect CEs, and visualizing them in interactive plots.

Video 2. **Starved MCF10A WT epithelial cells expressing OptoRAF/ERK-KTR actuator/biosensor circuit are stimulated optogenetically with a growing circular region using DMD.** Left panel: Raw images from the ERK-KTR channel; right panel: nuclear centroids color-coded according to ERK-KTR cytoplasmic/nuclear ratio. Black dots—active cells identified in the binarization step; shaded convex hulls—CEs identified by ARCOS. The stimulation region is indicated in magenta. 1 movie frame corresponds to 1 min.

Video 3. **Starved MCF10A WT epithelial cells expressing OptoRAF/ERK-KTR actuator/biosensor circuit are stimulated optogenetically with a moving circular region using DMD.** Left panel: Raw images from the ERK-KTR channel; right panel: nuclear centroids color-coded according to ERK-KTR cytoplasmic/nuclear ratio. Black dots—active cells identified in the binarization step; shaded convex hulls—CEs identified by ARCOS. The stimulation region is indicated in magenta. 1 movie frame corresponds to 1 min.

Video 4. **Starved MCF10A WT epithelial cells expressing OptoRAF/ERK-KTR actuator/biosensor circuit are stimulated optogenetically with a splitting circular region using DMD.** Left panel: Raw images from the ERK-KTR channel; right panel: nuclear centroids color-coded according to ERK-KTR cytoplasmic/nuclear ratio. Black dots—active cells identified in the binarization step; shaded convex hulls—CEs identified by ARCOS. The stimulation region is indicated in magenta. 1 movie frame corresponds to 1 min.

Video 5. **Collective ERK activity waves in starved MCF10A WT, KRAS^{G12V}, and PIK3CA^{H1047R} epithelial cells expressing ERK-KTR biosensor.** Results of the ARCOS analysis are overlaid on raw images from the ERK-KTR channel. Each CE is indicated with a distinct color from a palette, reused over time. 1 movie frame corresponds to 5 min.

Video 6. **Collective ERK activity waves in starved NRK-52E epithelial cells expressing ERK activity reporter EKAREV-NLS.** Left panel: Nuclei; right panel: nuclear centroids color-coded according to FRET ratio. Green circles—apoptotic cells; black dots—active cells identified in the binarization step; shaded convex hulls—CEs identified by ARCOS. 1 movie frame corresponds to 5 min.

Video 7. **Collective calcium waves in starved MDCK epithelial cells expressing calcium reporter GCaMP6S.** Left panel: Raw images from the GCaMP6S channel; right panel: nuclear centroids color-coded according to ERK-KTR cytoplasmic/nuclear ratio. Green circles—apically extruded cells; black dots—active cells identified in the binarization step; shaded convex hulls—CEs identified by ARCOS. 1 movie frame corresponds to 5 s.

Video 8. **Bee shimmering in the giant honeybee colony.** Left panel: Raw images of the bee colony; right panel: bee centroids color-coded according to the amount of shimmering. Black dots—"active bees" identified in the binarization step; shaded convex hulls—CEs identified by ARCOS. 1 movie frame corresponds to 1/60th of a second.

Video 9. **Collective ERK activity waves in stage 2 MCF10A WT acinus expressing the ERK-KTR biosensor.** From left to right: (1) maximum Z projection of the ERK-KTR channel; (2) 3D rendering of cell nuclei volumes with lever.js software, where gray shades indicate low/high ERK activities; (3) nuclear centroids color-coded according to CEs identified with ARCOS. The size of the nuclei in the right panel corresponds to their Z position. (4) Spaghetti plot of collective events. 1 movie frame corresponds to 3 min.

Video 10.   **ARCOS sensitivity to iterative image degradation with additive Gaussian white noise.** Top left panel: Collective ERK activity in an MDCK epithelium expressing the EKAREV-NLS FRET biosensor. Shown are nuclei color coded according to the FRET ratio. Warmer colors indicate higher ERK activity. Top right panels: Sum of donor and acceptor channels with white Gaussian noise. Black outlines indicate segmentation of the nuclei. Active cells are marked with a black dot, whereas cells marked by a colored dot belong to respective CEs. Colored shapes indicate convex hulls of respective CEs. SNR label represents how much noise was added to the corresponding image. Bottom left panel: Timestamp; bottom right panels: animated spaghetti plots for the corresponding SNR iterations above. Colors correspond to different CEs. 1 movie frame corresponds to 1 min.