Tissue-wide coordination of calcium signaling regulates the epithelial stem cell pool during homeostasis

Jessica L Moore*[1], Feng Gao*[1], Catherine Matte-Martone[1], Shuangshuang Du[1], Elizabeth Lathrop[1], Smirthy Ganesan[1], Lin Shao[2], Dhananjay Bhaskar[1], Andy Cox[1], Caroline Hendry[1], Bastian Rieck[3,4], Smita Krishnaswamy†[1,5], and Valentina Greco†[1,6]

[1]Department of Genetics, Yale University School of Medicine, New Haven, CT 06510, USA
[2]Department of Neuroscience, Yale University School of Medicine, New Haven, CT 06536, USA
[3]Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland
[4]SIB Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
[5]Department of Computer Science, Yale University, New Haven, CT 06510, USA
[6]Departments of Cell Biology and Dermatology, Yale Stem Cell Center, Yale Cancer Center, Yale University School of Medicine, New Haven, CT 06510, USA

October 12, 2021

Abstract

Regenerative processes in the mammalian skin require coordinated cell-cell communication. Ca^{2+} signaling can coordinate tissue-level responses in developing and wounded epithelia in tissue explants and invertebrates[1, 2, 3, 4]. However, its role in the homeostatic, regenerative basal layer of the skin epithelium is unknown due to significant challenges in studying signaling dynamics in a spatially complex tissue context in live mice. Here we combine *in vivo* imaging of dynamic Ca^{2+} signaling at the single cell level across thousands of cells with a novel computational approach, Geometric Scattering Trajectory Homology (GSTH). GSTH models Ca^{2+} as signals over a cell adjacency graph and uses a multi-level wavelet-like transform (called a scattering transform) to extract signaling patterns from our high dimensional *in vivo* datasets. We discover local Ca^{2+} signaling patterns are orchestrated so that signals flow in a coordinated and directed manner across the tissue, distinct from topographically uncoordinated Ca^{2+} signaling in excitatory tissues. Directed Ca^{2+} signaling is regulated by the major gap junction protein in the epidermal stem cell layer, Connexin 43 (Cx43). Cx43 gap junctions are dissociated as cells progress through the cell cycle out of G1 and play an essential role in the progression of stem cells from G2 towards mitosis. Finally, G2 cells display related signaling patterns and are essential for tissue-level signaling coordination. Together, our results provide insight into how such a ubiquitous signaling pathway regulates highly specific behaviors and outcomes at a tissue-wide level to maintain proper homeostasis.

Introduction

Each day our bodies make and lose billions of cells[5, 6, 7]. This regenerative capacity is based on the ability to orchestrate fate decisions within an actively cycling stem cell pool, resulting in a balanced production of new stem cells (by division) and loss of cells (by delamination or apoptosis). It has been shown that this balance is carried out locally within the stratified skin epithelium, with cells leaving G1 and progressing through S and G2 to divide in response to a neighboring cell's exit from the basal layer into the above differentiated layer[8]; however, how these locally coordinated

*These authors contributed equally.
†Correspondence to: Smita Krishnaswamy, Email: smita.krishnaswamy@yale.edu, and Valentina Greco, Email: valentina.greco@yale.edu

neighborhoods of cells are integrated across the entire basal layer to globally maintain homeostasis is poorly understood. Tight coordination of these regenerative processes imply underlying communication among these cells.

Ca^{2+} signaling offers an attractive mechanism for cell-to-cell communication across a tissue because of its established role as a signal integrator in developing and wounded epithelial tissue in invertebrates [2, 3, 1, 9] and during epidermal differentiation in the skin[10, 11, 12, 13, 14, 15, 16, 17]. In epithelial tissues, Ca^{2+} signaling is mediated primarily by gap junctions: assemblies of cell-cell channels composed of connexin hexamers[18]. Whether dynamic Ca^{2+} signaling plays an important role *in vivo* has been challenging to interrogate since standard methods of tissue isolation and fixation obscure the spatiotemporal dynamics of the molecular states of constituent cells. Hence, the possible role of Ca^{2+} signaling activity within the basal stem cell layer remains largely uncharacterized[19].

Here, we adapt a two-photon microscopy system[20, 21] to the use of Ca^{2+}-sensor mice[22] in order to visualize Ca^{2+} signaling activity *in vivo* in the mammalian epidermis and to investigate local patterns of dynamic Ca^{2+} signaling across the basal layer. However, the representation and quantitative analysis of Ca^{2+} signaling data at a tissue scale remains a significant challenge. Existing methods in bioinformatics are not set up to globally capture the spatiotemporal patterns of complex signals. Computational methods developed for signaling pattern analysis often use principal component analysis to identify cell assemblies (i.e., clusters of cells with similar dynamics) or rely on manual detection of signaling patterns[23, 24, 3]. To address this, we developed a novel data-driven computational analysis framework named Geometric Scattering Trajectory Homology (GSTH). GSTH is based on a combination of graph signal processing (to capture signaling patterns over the tissue), data geometry (to represent the dynamic trajectory), and topology (to quantitatively characterize the trajectory). GSTH facilitates exploration of signaling patterns by allowing comparisons of global dynamics between experimental conditions, as well as within the dynamics of a single experiment. The geometric scattering transform step of GSTH can also be adapted to learn representations of cells, which we then visualize with PHATE (a visualization method that preserves trajectory and manifold information)[25] to compare signaling dynamics of thousands of individual cells within a single experiment.

Our analyses of the spatiotemporal characteristics of Ca^{2+} signaling in thousands of cells reveal local patterns that are active across the cycling epidermal stem cell pool. Further, GSTH uncovers directed and coordinated spread of Ca^{2+} signaling across the epithelial basal layer, in contrast to the neuronal visual cortex. Loss of Cx43, a major mediator of Ca^{2+} signaling in the epidermis[18, 26], disrupts this coordination and leads to defects in the mitotic capacity of the stem cell pool. Cx43 localization at cell junctions peaks in the G1 stage of the cell cycle and progressively diminishes as cells advance towards mitosis, leading us to investigate the link between Cx43-mediated Ca^{2+} signaling and cell cycle regulation. We find that all G2 cells have related Ca^{2+} signaling patterns. Finally, G2 cells are essential for the coordinated and directed spread of Ca^{2+} signaling across the basal layer.

Results

Epidermal basal cells display local and dynamic patterns of Ca^{2+} signaling

Under homeostatic conditions, epidermal stem cells either progress through the cell cycle towards division or differentiate and exit into the suprabasal layer **(Figure A1A)**. To determine the dynamics of Ca^{2+} signaling during these transitions we generated mice with a Ca^{2+}-sensor expressed in all epidermal cells expressing Keratin 14 (K14-Cre; Rosa-CAG-LSL-GCaMP6s) and combined this with a nuclear marker (K14-H2BmCherry). Live imaging of the mouse ear skin[20, 21] revealed highly variable levels of participation from the basal cells, with 43.1 $\pm$ 21.4 % of cells signaling within a given 30 minute timeframe (**Movie 1, Figure 1A**). We next quantified Ca^{2+} signaling in a large (500 µm by 500 µm) region, encompassing about 2,500 epidermal basal cells, over a period of 24 hours to determine whether homeostatic Ca^{2+} signaling is restricted to specific locations in the basal layer or whether it is shared across all basal cells (**Figure A1B**). Comparison of active Ca^{2+} signaling across the 24 hour time period revealed that Ca^{2+} signaling is not spatially persistent (**Figure 1B**) but rather changes regionally with time. To quantify this, we measured the degree to which the pixels with thresholded Ca^{2+}-sensor fluorescence intensity overlapped at 0hr and 24hr and calculated image correlation coefficients. The image correlation coefficients over 24hrs (0.1411 $\pm$ 0.06416) were significantly lower than coefficients generated by comparing the same region at 0hr and 0.5hr, just 30 minutes later (0.3723 $\pm$ 0.0064) (Figure **1C, A1C**). Together, these results demonstrate how intercellular Ca^{2+} signaling is dynamic across tissue domains and occurs ubiquitously throughout the basal layer.

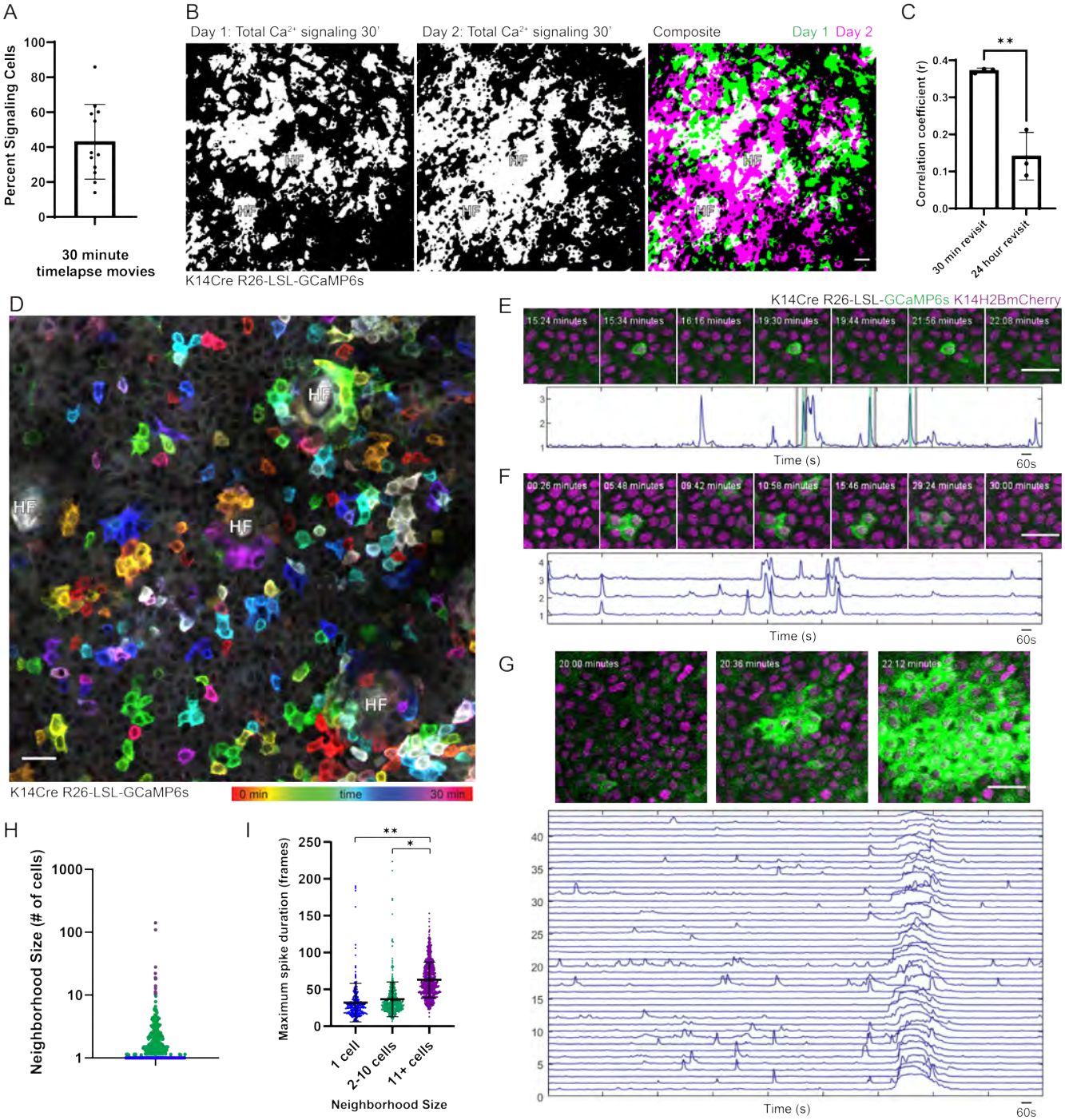

Figure 1: **Epidermal basal cells display local patterns of Ca^{2+} signaling, with highly active regions changing as the tissue turns over. (A)** Percent of epidermal basal cells spiking at least once during 30-minute recording of Ca^{2+} signaling in a live Ca^{2+}-sensor mouse (K14-Cre; Rosa26-CAG-LSL-GCaMP6s; K14-H2BmCherry). N = 12 thirty-minute movies from 6 mice. **(B)** Max intensity projection of all optical sections of a 30-minute timelapse at 0- and 24-hours of the same region of the epidermis. To the right, composite image of the same region at 0- (green) and 24-hours (magenta), where white indicates overlapping regions of Ca^{2+} signaling. Transverse views of the infundibulum of hair follicles marked with HF. Scale bars: 25 μm. **(C)** Correlation coefficient quantification of pixel overlap of Ca^{2+} signaling during 30-minute timelapses from revisits of the basal layer taken at 30-minute and 24-hour timepoints. ** = $p < 0.01$, Student's t test. N = 3 mice. **(D)** 30-minute timelapse video of the epidermal basal layer showing a diversity of spatiotemporal signaling patterns. Color scale represents time. Transverse views of the hair follicle infundibulum marked with HF. Scale bars: 25 μm. **(E)** Region of the basal layer where a single cell spikes repeatedly over 30 minutes of imaging (K14-Cre; Rosa26-CAG-LSL-GCaMP6s; K14-H2BmCherry). Below is normalized fluorescence intensity plotted over the duration of 30-minute timelapse. Black and green bars indicate timepoints corresponding to the snapshots above. Scale bars: 25 μm. **(F)** Region of the basal layer where a cluster of three cells spike repeatedly over 30 minutes of imaging (K14-Cre; Rosa26-CAG-LSL-GCaMP6s; K14-H2BmCherry). Below is normalized fluorescence intensity plotted over the 30-minute timelapse for each of the three spiking cells. Scale bars: 25 μm. **(G)** Region of the basal layer where a large group of cells exhibit an intercellular Ca^{2+} wave (ICW) (K14-Cre; Rosa26-CAG-LSL-GCaMP6s; K14-H2BmCherry). Beneath is normalized fluorescence intensity plotted over the 30-minute timelapse for 43 of the cells involved in the ICW. Scale bars: 25 μm. **(H)** Neighborhood sizes of cells with spatiotemporally localized Ca^{2+} signaling from 30-minute timelapse videos of the epidermal basal layer in a live anesthetized Ca^{2+}-sensor mouse. Purple, blue, and green dots represent the three different spatial patterns of Ca^{2+} signaling. N = 6 thirty-minute timelapse movies from 3 mice. **(I)** Maximal spike duration (maximal number of frames between the start and end of individual Ca^{2+} events) for three different spatial patterns of Ca^{2+} signaling. * P = 0.0213, ** P = 0.0056, Nested One-way ANOVA, N = 6 homeostatic thirty-minute timelapse movies.

To understand how Ca^{2+} dynamics are orchestrated on a shorter timescale within a field of connected epithelial cells, we again imaged the same large region of the basal epidermis for 30 minutes and temporally color coded each frame of the timelapse movie to simultaneously visualize all the Ca^{2+} signaling patterns (**Figure 1D**). We observed distinct spatiotemporal patterns of Ca^{2+} signaling: in some cases single cells spiked quickly in isolation, whereas in other cases neighborhoods of cells spiked simultaneously or in a propagating wave. This is consistent with previous reports of Ca^{2+} signaling in the basal layer that show three distinct patterns of Ca^{2+} signaling[19]; however, these patterns have not yet been extensively quantified. Therefore, we developed an automated segmentation and peak identification pipeline in MATLAB adapted from existing platforms including CaImAn[23, 27] that allowed us to simultaneously analyze the Ca^{2+} signaling from each individual cell and identify timepoints with Ca^{2+} transients. To understand the connectivity of signaling cells, we defined a graph in which nodes were connected if the cells they represented were direct neighbors (within 1 μm of each other) and spiked within 10 seconds of their neighbor.

Using this computational analysis, we showed that the epidermal stem cell pool displays a balance of intracellular and intercellular Ca^{2+} signaling, with stereotyped spatial and temporal patterns. Most events we detected were either single cells that spiked in isolation from their neighboring cells (65.88 $\pm$ 2.65%; **Figure 1E, 1H**) or spatiotemporally clustered transients across 2 or more neighboring cells (31.27 $\pm$ 1.96%; **Figure 1F**). We also observed relatively rare Ca^{2+} signaling waves that occurred across hundreds of cells (**Figure 1G**). We found that Ca^{2+} transients in larger neighborhoods of cells persist longer than in single cells or small neighborhoods (**Figure 1I**). The dynamic nature of these intercellular signaling events is limited to epidermal basal cells and does not characterize the directly above differentiated suprabasal layer (**Figure A1D**). These data characterize local patterns of Ca^{2+} signaling that manifest across the basal layer, with regions of signaling changing as the basal epithelium turns over.

Geometric Scattering Trajectory Homology (GSTH): a computational framework to represent information flow across tissue

In the above analyses, we identified mainly local signaling events limited to neighborhoods of 10 or fewer cells. Tissue homeostasis in the skin takes place across many dimensions, including tissue wide, and must integrate coordination across all local neighborhoods of cells. Understanding information flow at the tissue-wide scale represents a formidable challenge due to both the spatiotemporal dynamics of the signaling itself as well as the cellular complexity inherent to the tissue (i.e., number of cells, three dimensional organization, etc.). To address this, we developed a method called GSTH - *Graph Scattering Trajectory Homology*, which captures spatial and temporal patterns of signaling.

Frequency-domain descriptions of signaling To motivate GSTH, consider the problem of representing a signal on a set of cells (here arranged in planar spatial patterns). If we simply describe the signals as a vector of values on an indexing of cells, then we could not compare signaling patterns from different tissues, as specific cellular coordinates are not matched between tissues. Therefore, the signaling description has to be invariant to permutations in cell indexing, shifts in the signal, and even the number of cells. To address this issue, in classic signal processing, researchers use frequency domain descriptions, such as the Fourier transform (FT), which describe the periodicity rather than time/space-specificity of signals. Recently, with the prevalence of graph-structured data, there is an emerging field of graph signal processing[28]. Researchers in this field have invented the analogous graph Fourier transform (GFT)[29] for descriptions of signals on graphs. However, the GFT (and the FT) is usually only suitable for describing signals with global periodic patterns. More localized signaling patterns can be described with a wavelet transform. Here, to describe signals on planar arrangements of cells, we use a graph wavelet transform[30], which can capture both localized and global patterns.

Multiscale descriptions However, one scale of wavelet transforms is not sufficient to capture all the invariances we need in the signals. For instance, even if signals were being compared on the exact same cells, two signaling patterns could be similar but just shifted by one cell, or similar overall but different cell to cell. To capture a broader notion of similarity, we essentially look at signaling at all scales of granularity. To achieve this, we do not simply use raw signals but also versions of these signals that are diffused to different scales. These diffusions are a natural part of a particular type of graph wavelet, called a *diffusion wavelet*[30], which we employ in GSTH.

Permutation-invariant descriptions GSTH uses multiple scales of diffusion wavelets, whose coefficients are locally averaged in a *geometric scattering transform*. Specifically, the original geometric scattering transform uses statistical moments of wavelet coefficients over all cells to achieve permutation invariant descriptions of the signaling patterns (see [31] for permutation invariance results). Depending on the level of permutation invariance needed, one can also locally average wavelet coefficients as in [32].

Signaling patterns in time Not only are our signals patterned in space, but they are also patterned in time. To understand signaling dynamics, we look at a time series of geometric scattering transforms in a dimensionality reduction visualization called PHATE. PHATE, due to its ability to preserve manifold distances in low dimensions, reveals the time trajectory of the signaling. Finally, to develop a descriptor of the entire spatiotemporal signaling pattern, we convert the PHATE time trajectory into a quantitative descriptor called a persistence diagram[33]. This kind of descriptor was developed in the field of topological data analysis and uses abstract topological features that appear in the trajectory as points are merged with one another using a coarse-graining operation. Persistence diagrams calculate how many holes, voids, and components feature in the trajectory at each level of coarse-graining. Moreover, persistence diagrams between entire dynamic trajectories can be readily compared via well-defined Wasserstein distances[34], which we employ here. A more precise description of the steps of GSTH is given below.

Geometric Scattering Trajectory Homology (GSTH): Algorithmic Overview

GSTH chiefly consists of four steps (**Figure 2A**), which are explained below:

Step 1: The first step of GSTH consists of defining a graph on which the Ca^{2+} is regarded as a static node signal for each timepoint. For the epithelial cells, we defined a nearest-neighbors graph $G = \{V, E\}$ with each vertex $v_i \in V$ being a cell, and an edge $(v_j, v_k) \in E$ if the cells v_i and v_j are spatially adjacent. The connectivity of this cell graph G can be described by its adjacency matrix A, where $A_{ij} = 1$ if v_i and v_j are connected (i.e., $(v_j, v_k) \in E$) and 0 otherwise. The vertex degrees are collected in a degree matrix D, with $D_{ii} = \sum_{j=1}^{i} A_{ij}$. The Ca^{2+} signaling level of the cell is regarded as a *signal* x defined on the graph, i.e., each vertex v_i in the graph has an activation value $x(v_i, t)$ representing the level of Ca^{2+} signaling at time t. From here on, this graph is referred to as the cellular graph.

Step 2: The second step is to derive a mathematical descriptor of the cellular signaling pattern defined in Step 1. We utilize the geometric scattering transform, which is a generalization of scattering transforms to graphs, to embed each time point. The geometric scattering transform consists of a cascade of graph wavelet transforms followed by a nonlinear modulus operation applied to graph signals and is built in a multi layer (or multi order) architecture (Figure 2), each extracting finer descriptions of signals. Graph wavelets, in turn, are defined using a diffusion operator (or equivalently a random walk operator) on the graph and are essentially a difference between signal diffusions of different time steps on the graph. Thus, we first define a diffusion operator $\mathbf{R} = \frac{1}{2}(I + AD^{-1})$, where A and D are the graph adjacency matrix and degree matrix, respectively, from Step 1 and I is the identity matrix. This diffusion operator computes the probability of a lazy random walk starting from any vertex ending in another vertex in t steps, where each step (from one vertex to another) has a probability proportional to its adjacency and inversely to the degree. Hence, the t-th power of $\mathbf{R}$, $\mathbf{R}^t$, represents the probability distribution after t steps. Based on this operator $\mathbf{R}$, we can then define a graph wavelets of different scales as follows[30]:

$$\mathbf{\Psi}_0 = \mathbf{I} - \mathbf{R}\,, \quad \mathbf{\Psi}_j = \mathbf{R}^{2^{j-1}} - \mathbf{R}^{2^j} = \mathbf{R}^{2^{j-1}}(\mathbf{I} - \mathbf{R}^{2^{j-1}})\,, \quad j \geq 1\,. \tag{1}$$

By using multiple wavelet scales $\mathbf{\Psi}j_1, \mathbf{\Psi}j_2, \mathbf{\Psi}j_3 \ldots$, the operators can thus be used to extract multilevel signal information from the graph. We then calculate the scattering transform $S(X)$ at different orders, which yields a collection of scattering coefficients $S(X(t))$ as overall cellular signaling pattern embeddings for timepoint t. Specifically, the zeroth-order scattering transform calculates the local averaging of raw signals $X(t_i)$ at vertex/cell v_ℓ and is obtained by:

$$S_0(X(v_\ell, t_i)) = \mathbf{R}^{2^J}\mathbf{x}(v_\ell, t_i). \tag{2}$$

The resulting scattering coefficients $S_0(X(v_\ell, t_i))$ for every vertex/cell at timepoint t_i are then concatenated to form the zeroth-order scattering coefficients $S_0(X(t_i))$ for timepoint t_i. While the diffusion operator $\mathbf{R}$ provides local averaging of neighboring cell patterns, it also suppresses high frequency information. We can further calculate the first-order scattering

coefficients to retrieve finer description of these high frequency information. This is achieved by applying graph wavelets described above:

$$S_1(X(j, v_\ell, t_i)) = \mathbf{R}^{2^J} |\mathbf{\Psi}_j \mathbf{x}(v_\ell, t_i)|,\ 1 \leq j \leq J, \tag{3}$$

Similarly, the zeroth-order and first-order scattering coefficients can be further complemented by the second-order scattering coefficients:

$$S_2(X(j, j', v_\ell, t_i)) = \mathbf{R}^{2^J} |\mathbf{\Psi}_{j'} |\mathbf{\Psi}_j \mathbf{x}(v_\ell, t_i)||, 1 \leq j < j' \leq J \tag{4}$$

Finally, we concatenate the zeroth-order, first-order and second-order scattering coefficients to form the timepoint embedding $S(X(t_i))$ for timepoint t_i. More details of the timepoint embedding calculation can be found in Computational Methods.

Step 3: In the third step, we reduce the descriptors of signaling patterns of each timepoint $S(X(t))$ from Step 2 into a low-dimensional trajectory E, where $E = \{e_{t1}, e_{t2}, \ldots, e_{tn}\}$ with each e_i corresponding to a low-dimensional PHATE embedding of timepoint t_i. This is achieved by applying PHATE, a dimensionality reduction method that captures both local and global nonlinear structure by constructing a diffusion geometry[25]. The advantage of PHATE over other methods is that it retains trajectory structure and global distances as opposed to stochastic neighbor embeddings, such as UMAP and t-SNE, which tend to shatter trajectory structure. Once we apply PHATE, we can reduce the high-dimensional scattering coefficients $S(X(t))$ for each timepoint into the 3-D PHATE embedding coordinates $E(t) = (E_1(X_t), E_2(X_t), E_3(X_t))$. The collection of these timepoints forms a low-dimensional trajectory $E = \{E(t_1), E(t_2)..., E(t_n)\}$, allowing visualization of the temporal dynamics of cells signaling.

Step 4: Finally, to quantify features of the trajectories E obtained in Step 3 (e.g., to compare the existence of loops or holes in the trajectories), we quantify their shape using persistent homology, a topological data analysis method. In persistent homology calculations, data is gradually coarse-grained by merging nearby points and at each level of granularity, a graph (or simplicial complex in higher dimensions) of the data is quantified by counting the number of connected components, cycles, and potentially higher-dimensional "holes". The gradual coarse-graining is referred to as "filtration". Here, we use a filtration method known as the Vietoris-Rips filtration, where we create connections between two points e_i and e_j in the trajectory if the points are closer than a threshold ϵ, measured using Euclidean distance on the embedding. The threshold ϵ is gradually increased, ranging from 0 to ∞, until all points are connected in a fully-connected graph. We visualize the results in two ways: first, we calculate a persistence diagram Q that tracks the birth and death of each topological feature that occurs during the filtration process, i.e., it contains a point at (ϵ_i, ϵ_j) for each connected component (for instance) that is created at threshold ϵ_i and destroyed (i.e., merged) at threshold ϵ_j (See methods for more details). Persisent homology is still under-explored in data science but offers a way of quantifying general shape features of datasets. For instance, we can differentiate between smoother trajectories (which will only contain large cycles appearing at later granularities, i.e., higher values of ϵ) versus rougher trajectories (where cycles can get created and destroyed at smaller values of ϵ). Besides, we can also quantify the differences between persistence diagrams by calculating the Wasserstein distances between two diagrams[34], which can translate to distances between signaling diagrams.

A second way to visualize the results is through the so-called Betti curves that are associated with a persistence diagram Q, resulting in a simple summary curve $B(Q, q)$ for the persistence diagram in dimension q. More precisely, the Betti curve of dimension q of a diagram Q refers to the sequence of Betti numbers, i.e., active topological features, of dimension q in Q, evaluated for each threshold ϵ. It is a useful descriptor for numerous machine learning tasks[35]. Intuitively, the Betti numbers represent the number of q-dimensional holes in a topological space. Therefore, the Betti curve characterizes the connectivity of $VR_s(E)$ and, by extension, of the Ca^{2+} signaling data.

Cellular Embeddings using GSTH In addition to the time trajectory, we also create embeddings of the cells (henceforth referred to as the cellular embeddings), based on all points in time, to compare the participation of individual cells in the Ca^{2+} signaling patterns within the same experiment. We utilize the same cellular graph as described in Step 1. However, unlike before where we considered the signals from all the cells at each individual time point, we now consider the signals $x_i = x_{i,1}, x_{i,2}, \ldots, x_{i,t}$, which are defined on each cell i across all timepoints as features. Hence, the generated cellular embedding can reflect the Ca^{2+} signaling pattern of an individual cell across all timepoints. Similarly, we first create the diffusion operator R and graph wavelet $\mathbf{\Psi}$ as described in GSTH Step 2: $R = \frac{1}{2}(I + AD^{-1})$ and $\mathbf{\Psi}_j = \mathbf{R}^{2^{j-1}} - \mathbf{R}^{2^j}$. We can then collect wavelet coefficients at each vertex following Equations 8, 9 and 10. These wavelet coefficients represent the patterns from the cell itself and also incorporate larger scale signaling patterns by considering neighboring cells at multiple scales. By concatenating these wavelet coefficients at every timepoint for each cell, we have a cellular embedding

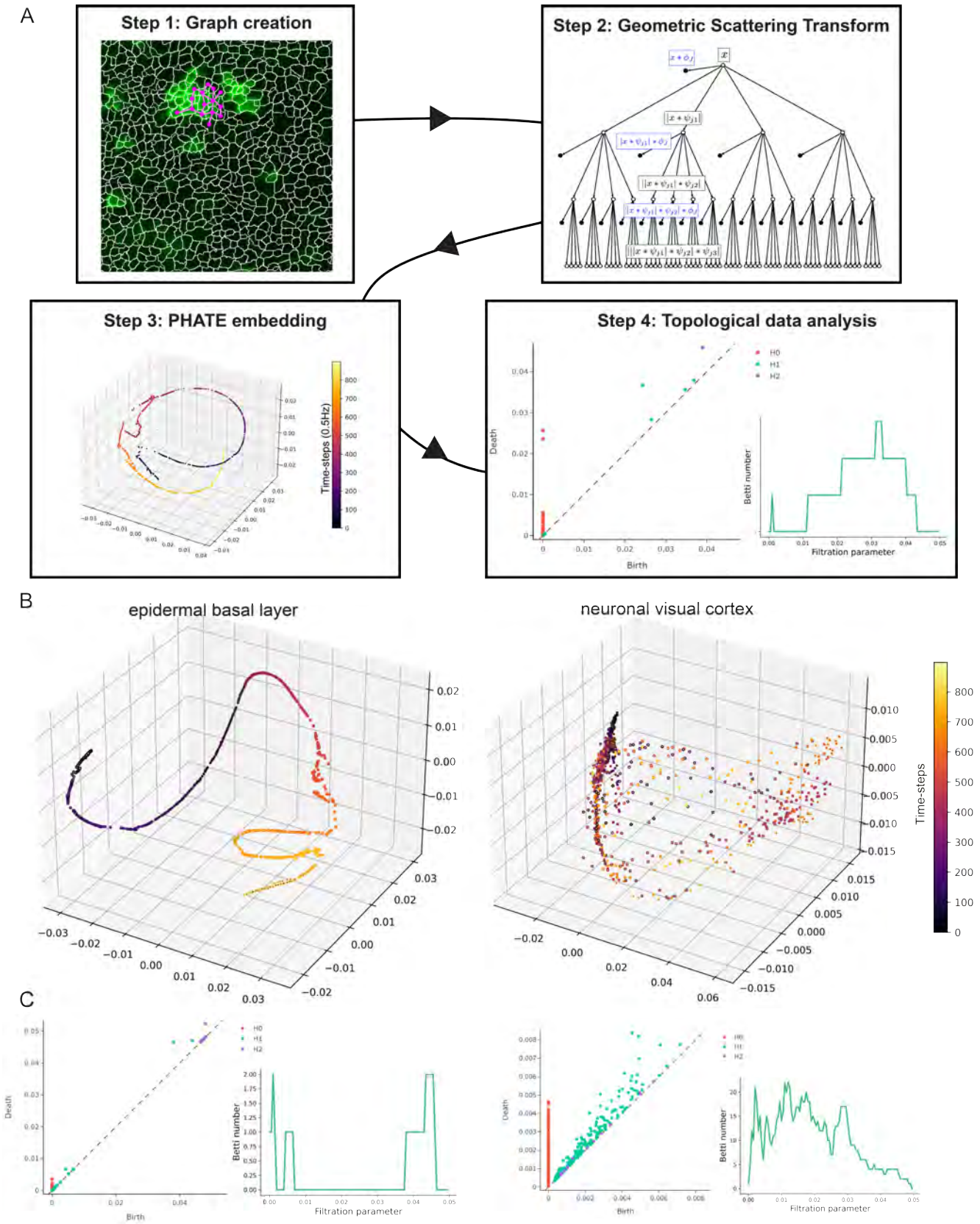

Figure 2: **GSTH analysis reveals directed and coordinated Ca^{2+} signaling patterns across the basal epithelium. (A)** GSTH workflow. Step 1: The cellular graph is created based on spatial adjacency (shown superimposed on segmented image of basal layer Ca^{2+} signaling). Step 2: Time point embeddings are created using the geometric scattering transform. Step 3: PHATE (dimensionality reduction method) visualization of signaling time trajectory. Step 4: Topological data analysis via persistence diagrams of each trajectory ($H0$: connected components, $H1$: loops, $H2$: voids) and featurized representations of the diagram with Betti curves. **(B)** Representative PHATE visualization of Ca^{2+} signaling in the homeostatic basal epithelial layer (left) versus Ca^{2+} signaling in the neuronal visual cortex (right). **(C)** Corresponding persistence diagrams of Ca^{2+} signaling time trajectories in the homeostatic basal epithelial layer versus in the neuronal visual cortex. Each point corresponds to a topological feature in the trajectory, which appears at a certain birth time and disappears at a death time. As an example, green points represent $H1$ features that correspond to the formation of loops in the trajectory while purple dots represent $H2$ features that correspond to the formation of voids. The further they are from the respective diagonals, the longer they exist, i.e., the larger their persistence. To the right, examples of corresponding Betti curves of $H1$ loop features.

for each cell that encompasses Ca^{2+} signaling information from that cell and its neighbors across all timepoints. We can finally apply PHATE to generate low-dimensional PHATE embeddings of the cells. Further, this embedding allows us to cluster cells by their participation in signaling dynamics.

Validation of GSTH using synthetic datasets

To better understand GSTH and the signaling patterns it captures, we applied GSTH to three synthetic datasets, each simulating different signal diffusion scenarios. We compared the results obtained from GSTH with four other approaches representing ablations of key steps in GSTH: (1) applying PHATE directly to the raw signals (thus ablating scattering), (2) using the scattering transform to generate time step embeddings but visualizing with PCA, (3) using the scattering transform to generate time step embeddings but visualizing with t-SNE, and (4) using the same time step embeddings but visualizing with UMAP (**Figure A2**).

The first dataset (Testcase 1 in Methods) aimed to test GSTH's stability to small perturbations, as well as its ability to capture signal diffusion dynamics on the graph. To achieve this goal, we simulated a wave of Ca^{2+} signaling and a slightly perturbed wave. After applying GSTH, time points with perturbed and original signal values overlapped, demonstrating that the scattering transform and PHATE are invariant to small degrees of noise (**Figure A2A**). Additionally, the smooth trajectory demonstrated that the scattering transform and PHATE of GSTH can effectively capture the signal propagation on the graph. In contrast, using PHATE directly on the raw input signals resulted in condensed timepoints (**Figure A2C**). Applying PCA (**Figure A2D**) resulted in a more dispersed trajectory, and applying t-SNE (**Figure A2E**) and UMAP (**Figure A2F**) both led to overlapping timepoints, all failing to reflect the true propagation of signals on the graph.

With the second synthetic dataset (Testcase 2), we aimed to compare each method's ability to retrieve signal diffusion dynamics on the graph under more complex conditions. We hence simulated two different waves with similar initiating signaling events. Then, as the signal diffused on the graph, each of the waves eventually diffused to different patterns. After applying GSTH (**Figure A2H**), time points from the two signal sources formed two branches with their starting points near each other in PHATE coordinates. Thus, from one end to the next this was akin to a signal condensing and then diffusing again. As expected, this created a loop-like structure in the PHATE graph. However, directly applying PHATE to the raw signals (**Figure A2J**) resulted in multiple scattered points separate from the main trajectory and was not able to fully capture or distinguish the signals. Furthermore, applying PCA (**Figure A2K**), t-SNE (**Figure A2L**), and UMAP (**Figure A2M**) to the scattering coefficients failed to form loop-like structures to reflect the similar initiation of the waves.

Finally, for the third synthetic dataset (Testcase 3), we aimed to simulate the propagation of signals similar to those observed in the basal epithelial layer to better understand what types of signaling patterns GSTH can capture. Therefore, we first simulated the diffusion of signals on the graph, following by single cells spiking. With GSTH (**Figure A2O**), we observed time points forming smooth trajectories at first (reflecting the propagation of signals on the graph), then disrupted trajectories (corresponding to random spiking of single cells). In comparison, using PHATE directly on the input signals (**Figure A2Q**), PCA on the generated scattering coefficients (**Figure A2R**) or UMAP on the scattering coefficients (**Figure A2T**) led to very condensed trajectories (for the initial stage when signals are diffusing on the graph), making it hard to identify the inner dynamics. Finally, using t-SNE on the scattering coefficients (**Figure A2S**) generated more scattered clustering that failed to capture the propagation of signals on the graph.

One additional validation of our GSTH approach is based on the observation of similar signaling patterns observed within experimental groups and consistently different patterns between experimental groups, reflected in the time trajectories. To quantify this and compare across experimental groups, we calculated Wasserstein distances among persistence diagrams from multiple experimental groups (**Figure 5G**). We found small distances between duplicates from each group and larger distances across different experimental groups, demonstrating GSTH is able to capture homeostatic and perturbed signaling patterns across many samples.

In addition, we conducted ablation studies of our cellular embedding method. Specifically, we compared our cellular embedding method with directly applying PHATE to the raw signals for two more synthetic datasets (Testcase 4 and Testcase 5 in Methods). For Testcase 4, we created signals for each cell so that cells contained one of two major types of Ca^{2+} signaling observed in the experiment: single cells spiking or clustered signaling. We showed that our cellular embedding method successfully distinguished these two signaling patterns and is aware of the graph

structure (**Figure A3A**), whereas applying PHATE directly to the raw signals (**Figure A3B**) failed to reveal this, as single spiking cells were condensed (showing less information about graph distances). In Testcase 5, we simulated two intercellular Ca^{2+} waves resulting from two Dirac signals diffused on the graph. Both methods revealed that cells from the two similar waves overlapped due to similar signaling patterns, however applying PHATE alone failed to incorporate the graph structure information, as shown by condensed points despite differences in graph distance (**Figure A3C, A3D**).

Comparison between the visual cortex and basal epithelium reveals directed and coordinated signaling

Having validated the GSTH algorithms and computational pipeline, we applied it to 30 minute timelapses (900 time steps) of homeostatic Ca^{2+} signaling in the epidermal layer. As described in the GSTH overview, we first constructed a nearest-neighbors cellular graph, generated scattering coefficients for each time point, and then used PHATE to visualize the time trajectory. Our analyses revealed smooth PHATE trajectories between time points, showing that Ca^{2+} signals steadily diffuse to neighbors in a directed and coordinated manner (**Figure 2B, A4A**). To determine how these PHATE trajectories from the epidermis compared to other tissues that participate in Ca^{2+} signaling, we next turned to a classic example of Ca^{2+} signaling in the nervous system and applied GSTH to previously published recordings of functional Ca^{2+} signaling from 10,000 neurons of the primary visual cortex. Spontaneous activity from the primary visual cortex has been shown to not be organized topographically during the resting state[36, 37]. Neurons can be connected via long processes, and so we used correlation between neurons' Ca^{2+} signals to create a neuronal graph (instead of the nearest-neighbors graph built for epidermal cells). We then followed the same steps to generate scattering coefficients for each time point. Our analysis revealed markedly discontinuous, rougher time trajectories with PHATE, indicating less spatially and temporally coordinated signaling across the tissue and more abrupt changes in signaling patterns over time (**Figure 2B, A4A**).

We then also visualized persistence diagrams and Betti curves of $H1$ features for both datasets. The persistence diagrams were markedly different, with many features appearing and disappearing at all scales, revealing a complex data geometry for the neuronal dataset. By contrast, the persistence diagram of the epidermal basal layer had a few noise features that quickly disappeared and then only large scale loops, revealing looping dynamics appeared much later in the persistence diagrams. Further, a prominent $H2$ feature or void (like the inside of a hollow ball) appeared at a late persistence stage in the epidermal dynamics, revealing an area of the state space that was not entered in these dynamics. By contrast, the neuronal persistence diagram had several low-persistence $H2$ features that appeared and disappeared quickly, revealing more complex topological features in the neuronal dataset (**Figure 2C**). Thus we find that Ca^{2+} signaling is directed and coordinated across the epidermal basal layer, demonstrating the spatial and temporal connectivity of cells in the basal epithelium, despite the tissue's continuous and rapid turnover during normal homeostasis.

GSTH generalizes to other biological systems To establish the generality of the GSTH method, we studied whether GSTH could differentiate between two different neuronal dynamics (which would represent a more subtle difference than comparisons between epidermal and neuronal signaling). We thus applied GSTH to an additional published visual cortex dataset, this time stimulated with 32 natural image stimuli, to see whether we would detect differences when compared with the spontaneous signaling in the visual cortex. While we observed some similarity in signaling patterns (in that they both were not as organized in time as the epidermal basal cells), the stimulated neuronal dataset displayed a much narrower, lower dimensional state space (locally some points with similar colors were near each other), contrasting with the dispersed PHATE plots from the unstimulated visual cortex datasets (**Figure A5A**). We then calculated persistence diagrams and Betti curves to investigate topological differences in signaling patterns of the stimulated and unstimulated visual cortex (**Figure A5B**. We observed that the spontaneous neuronal signaling trajectory contained more $H2$ features, representing voids in the trajectories. This comparison demonstrates the applicability of GSTH in a variety of systems to detect differences in global signaling patterns.

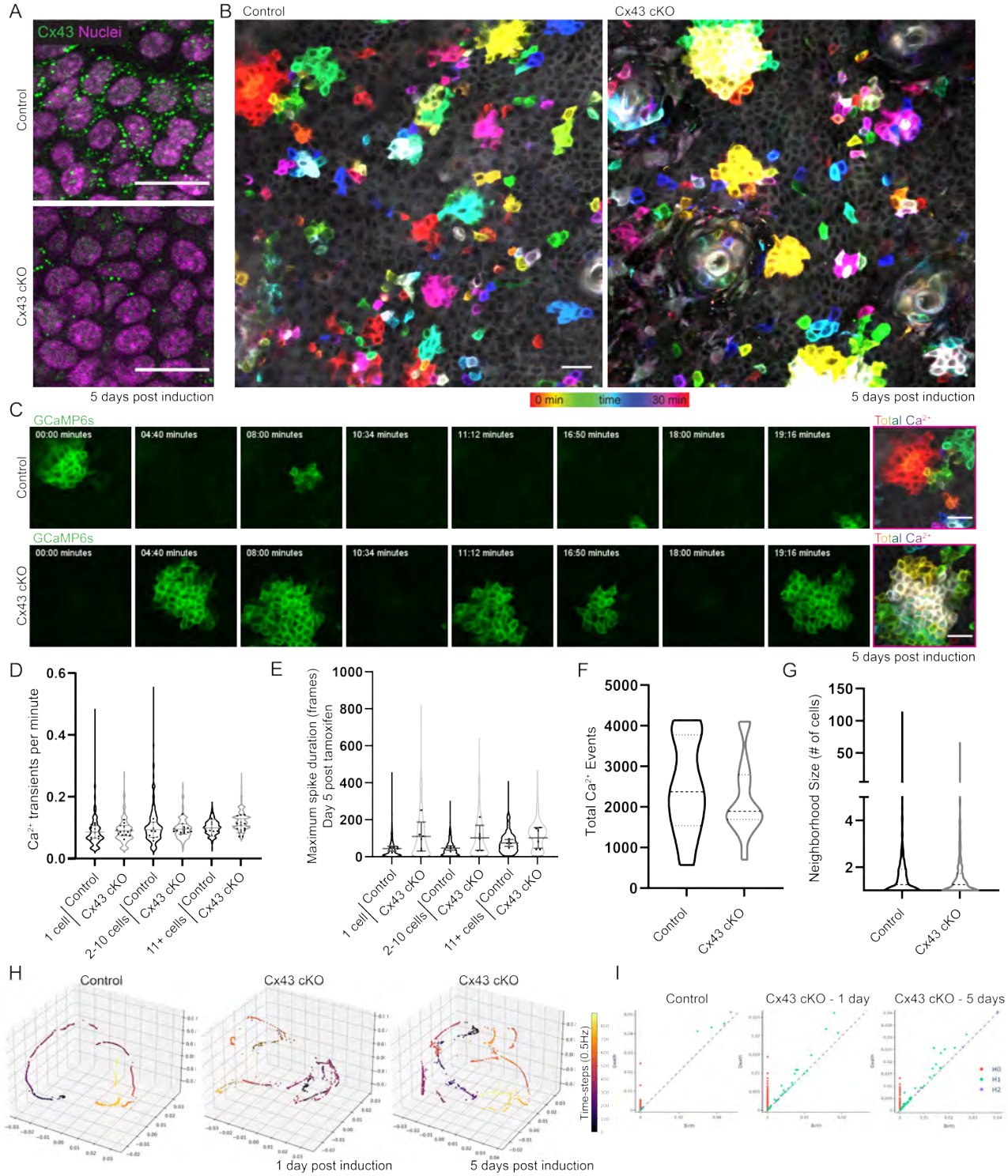

Figure 3: **Cx43 modulates homeostatic intercellular Ca^{2+} signaling dynamics in epidermal basal layer. (A)** Immunofluorescence staining of basal layer from K14-CreER; $Cx43^{+/+}$ and K14-CreER; $Cx43^{fl/fl}$ mice 5 days post-tamoxifen induction. Cx43 in green and DAPI (nuclei) in magenta. Scale bar: 25 µm. **(B)** Max intensity projection of 30-minute timelapse videos of the basal layer of control and Cx43 cKO Ca^{2+}-sensor mice 5 days post-induction (Rosa26-CAG-GCaMP6s; K14-CreER; $Cx43^{+/+}$ and Rosa26-CAG-GCaMP6s; K14-CreER; $Cx43^{fl/fl}$). Color scale represents time. Repeated signaling manifests as white signal (the sum of colors). Scale bars: 25 µm. **(C)** Time-course of clustered signaling from 30-minute videos of the basal layer of control and Cx43 cKO Ca^{2+}-sensor mice 5 days post-induction. Last image on right is max intensity projection with time represented by a color scale. Scale bars: 25 µm. **(D)** Ca^{2+} transients per minute per cell for three patterns of Ca^{2+} signaling (1 cell, 2-10 cells, or 11+ cells) in control versus Cx43 cKO Ca^{2+}-sensor mice. NS for 1 cell, $P = 0.0139$ for 2-10 cells, $P < 0.0001$ for 11+ cells comparison, Mann-Whitney test. **(E)** Maximal spike duration of Ca^{2+} transients per cell for three patterns of Ca^{2+} signaling (1 cell, 2-10 cells, or 11+ cells) in control versus Cx43 cKO mice 5 days post-tamoxifen induction. $P < 0.0001$, Mann-Whitney test. **(F)** Total number of Ca^{2+} signaling events in control versus Cx43 cKO mice. $N = 11$ (control) and 14 (Cx43 cKO) thirty-minute timelapse movies from at least 3 mice per condition. **(G)** Average neighborhood size of signaling in control versus Cx43 cKO mice. $N = 11$ (control) and 14 (Cx43 cKO) thirty-minute timelapse movies from at least 3 mice per condition. **(H)** Representative PHATE visualization of Ca^{2+} signaling in the Cx43 cKO versus control basal layer shows disruption of smooth, directed and coordinated patterns of signaling in mice 1 and 5 days after loss of Cx43. **(I)** Representative persistence diagrams (H0: connected components, H1: loops, H2: voids) for control and Cx43 cKO mice (Rosa26-CAG-GCaMP6s; K14-CreER; $Cx43^{+/+}$ and Rosa26-CAG-GCaMP6s; K14-CreER; $Cx43^{fl/fl}$) 1 and 5 days post-induction. H1 features from Cx43 cKO mice appear later in time and have a longer persistence.

Cx43 orchestrates directed and coordinated intercellular Ca^{2+} signaling in the epidermal basal layer.

Given Ca^{2+} signaling spreads from neighbor to neighbor in a directed and coordinated manner, we next wanted to understand the molecular modulators of these dynamics. Gap junctions are the primary mediators of Ca^{2+} signaling in epithelial tissues, directly linking the cytoplasm of neighboring cells[38, 39]. Connexin43 (Cx43) is the most ubiquitously expressed component of gap junctions in the skin, is highly expressed in the epidermal basal layer, and has been shown to propagate Ca^{2+} waves across epithelial cells *in vitro*[18, 40, 41, 42, 43]. To address whether Cx43 plays a role in homeostatic Ca^{2+} signaling patterns and their function in the skin, we crossed Cx43 conditional knockout mice (cKO) with a germline recombined Ca^{2+}-sensor line (K14CreER; $Cx43^{fl/fl}$; Rosa26-CAG-GCaMP6s and K14CreER; $Cx43^{+/+}$; Rosa26-CAG-GCaMP6s littermate controls) and performed live timelapse imaging at 1-, 5-, and 7-days post-induction. We first confirmed loss of Cx43 protein expression within 5 days of recombination (**Figure 3A**). While loss of Cx43 abolishes Cx43 gap junctions, it does not completely abolish all gap junctions, as detected by immunofluorescence whole-mount staining for Connexin31 (**Figure A7A**).

To simultaneously visualize all Ca^{2+} signaling events in the basal layer over time we temporally color coded each frame of the timelapse movie and combined these into a single image (**Figure 3B**). Our analyses revealed that loss of Cx43 altered spatiotemporal characteristics of signaling patterns compared to control. We observed neighborhoods of Ca^{2+} signaling oscillating repeatedly within the 30-minute window, in contrast to more dispersed clustered signaling in littermate controls, and cells in neighborhoods of more than 11 cells displayed increased transient frequency (**Figure 3B, 3C, 3D, Movie 2**). We also quantified the duration of the longest Ca^{2+} transient per cell, which we termed maximal spike duration. Loss of Cx43 resulted in a longer maximal spike duration for transients of Ca^{2+} signaling across all neighborhood sizes, most dramatically in single cells and small neighborhoods of 2 to 10 cells (**Figure 3E**). The average total Ca^{2+} events, the distribution of neighborhood sizes, and percent of cells participating in Ca^{2+} signaling were all unchanged (**Figure 3F, 3G, A7B**). Disruption of local patterns of Ca^{2+} signaling demonstrate a modulatory role for Cx43 and prompted us to question whether loss of Cx43 affects signaling dynamics at a tissue-wide level.

To address this, we applied GSTH to map Ca^{2+} signaling across the basal layer over a 30 minute time period. We observed a striking loss of smooth, coordinated Ca^{2+} signaling time trajectories in the epidermal basal layer upon loss of Cx43 compared to littermate control mice (**Figure 3H**). Instead, Ca^{2+} signaling trajectories in the Cx43 mutant mice appeared scattered and rougher (**Figure 3H, A6A**), showing more rapid changes of signals over the graph and less connected neighborhoods of intercellular signaling. This was also reflected by the persistence diagrams (**Figure 3I**). In these diagrams, $H0$ features represent connected components in the trajectory, $H1$ features represent loops, and $H2$ represent voids. There were fewer $H1$ features from the persistence diagrams of the control group, and many of them were created and died at later stages. This indicates that signaling dynamics in the control group were less disjointed and formed smoother trajectories. The reasoning is that, if there are deviations from the main trajectories, then these can form persistence features that appear earlier because they create loops at low thresholds of point connection. By contrast, smooth trajectories only create large scale loops appearing later in the persistence diagram. Most $H1$ features in the Cx43 cKO group appeared and disappeared at an earlier stage, thus these features were short-lived.

The differences in the persistence diagrams represent different topological features in the underlying PHATE trajectories, and therefore different Ca^{2+} signaling patterns after the loss of Cx43. These differences were further revealed through topological descriptors such as Betti curves, which we depict for $H1$ features. The Betti curves for the control group show that all loops are formed and closed at later thresholds, while those in the Cx43 cKO exhibit loops that emerged at earlier stages (**Figure A6B**). This also demonstrates different signaling patterns in the Cx43 cKO compared to the control. Perturbed trajectories across the basal layer were evident as early as one day after loss of Cx43, suggesting a direct role for Cx43 in Ca^{2+} signaling regulation. Taken together, our results reveal an emerging role for Cx43 in the modulation of Ca^{2+} signaling patterns, with disrupted tissue-wide signaling flow (**Figure 3H, A6A**).

Cx43 mediates basal layer cell cycle progression

To further understand the consequences of losing Cx43 expression, we investigated the regenerative behaviors of cells in the basal layer. Basal stem cells are constantly balancing between two behaviors - proliferation and differentiation. To determine whether loss of Cx43 affects these behaviors, we stained for the mitotic marker phospho-histone H3 (pH3) and for the early differentiation marker Keratin10 (K10)[44] in the Cx43 cKO mice 5 days post-induction. We observed a drop

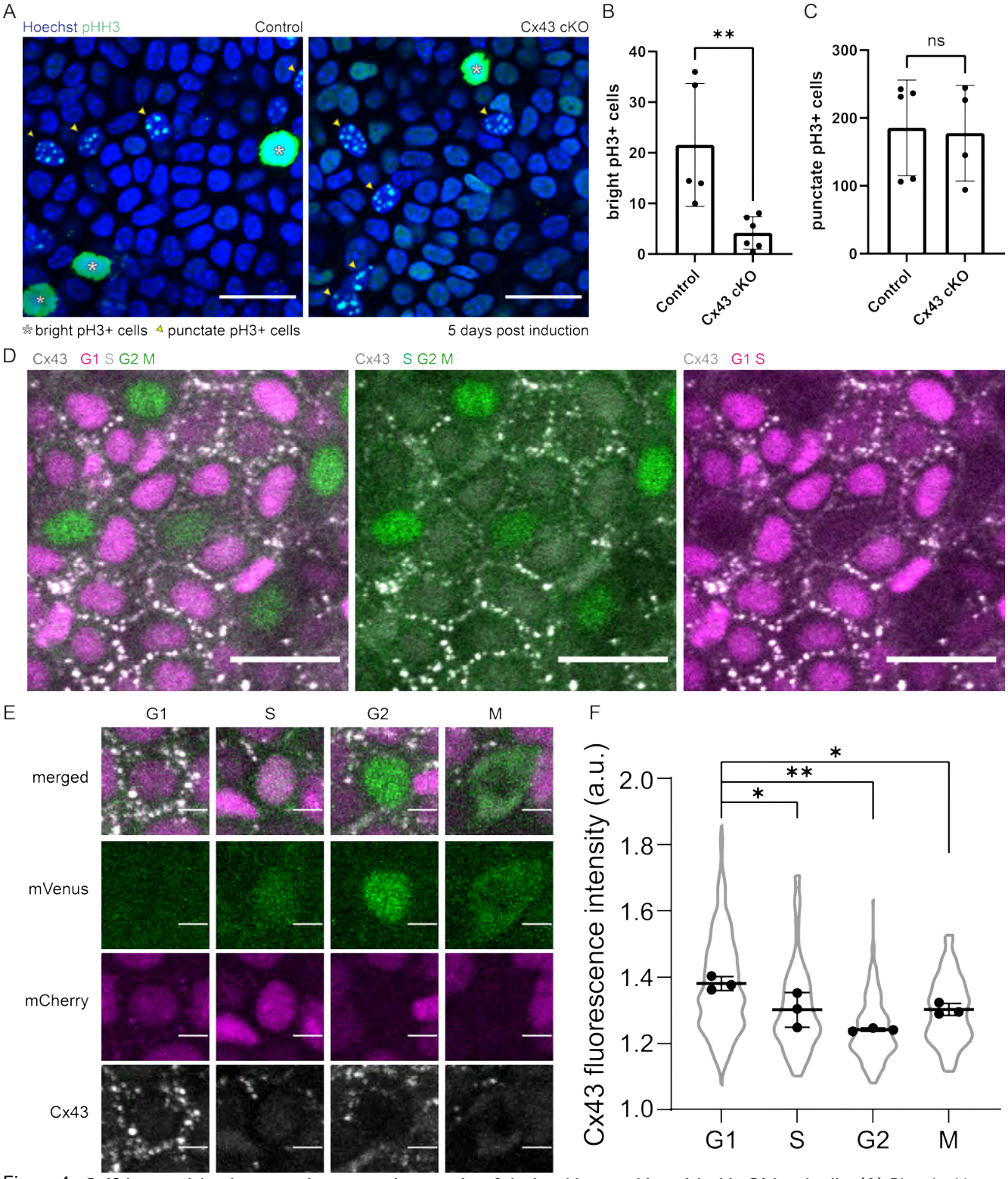

Figure 4: **Cx43 is essential to homeostatic regenerative capacity of the basal layer and is enriched in G1 basal cells. (A)** Phospho-histone H3 (pH3) immunofluorescence staining in control versus Cx43 cKO mice (K14-CreER Cx43$^{+/+}$ and K14-CreER Cx43$^{fl/fl}$) 5 days post-tamoxifen induction. Scale bars: 25 µm. **(B)** Quantification of number of bright pH3 positive cells per 250 mm^2 region 5 days post-tamoxifen induction, marking mitotic cells. ** P <0.01, Student's t test. N = 5 control and 6 Cx43 cKO mice. **(C)** Quantification of punctate pH3 positive cells per 250 mm^2 region in control versus Cx43 cKO mice 5 days post-tamoxifen induction, marking late G2 cells. N = 5 control and 4 Cx43 cKO mice. **(D)** Cx43 immunofluorescence staining, shown in white, in Rosa26p-Fucci2 mice, where G1 and S cells are mCherry$^+$ (magenta) and S, G2, and M cells are mVenus$^+$ (green). Scale bars: 25 µm. **(E)** Insets of G1, S, G2 and M cells in Rosa26p-Fucci2 mice with Cx43 immunofluorescence staining shown in white. Scale bars: 5 µm. **(F)** Quantification of Cx43 mean fluorescence intensity at the borders of G1, S, G2 and M cells in Rosa26p-Fucci2 mice. * P <0.05, ** P <0.01, One-way ANOVA, N = 3 control and 3 Cx43 cKO mice.

in bright pH3 staining (marking mitotic basal cells) in Cx43 cKO mice compared to littermate controls (**Figure 4A, 4B**), while punctate pH3 staining (small dots of staining across the nucleus, marking cells in the G2 stage of the cell cycle) was unchanged 5 days after loss of Cx43 (**Figure 4C**). This demonstrated that Cx43 is required for the progression of cells from G2 to mitosis. We did not see any change in the percent of K10 positive basal cells, indicating that similar numbers of basal cells prepare to exit the basal layer on their differentiation trajectory despite a loss of Cx43, and showing a disruption in the balance of proliferation and differentiation behaviors (**Figure A7C, A7D**). Additionally, we observed no noticeable change in basal cell density or overall epidermal thickness **(Figure A7E, A7F)** - two indicators of overall tissue homeostasis.

Our findings thus far demonstrate a decoupling between cell cycle progression and differentiation upon loss of Cx43, since entrance into mitosis, not differentiation, is perturbed. To further investigate this, we examined Cx43 expression at different stages of the cell cycle in the homeostatic epidermis of cell cycle reporter (Fucci2) mice. Interestingly, we found enrichment of Cx43 gap junction plaques in G1 cells versus S, G2, or M cells (**Figure 4D, 4E, 4F**). Together, these results demonstrate that Cx43-dependent coordinated patterns of Ca^{2+} signaling play a role in coordination of homeostatic cell behaviors in the basal epidermis and that Cx43 gap junctions are dissociated as cells progress out of G1 and through the cell cycle.

G2 cells are essential in mediating directed, coordinated patterns of signaling

The differential Cx43 expression throughout the cell cycle and the stall in mitosis that we observed led us to interrogate the relationship between Ca^{2+} signaling and the cell cycle in the epidermal basal layer. To investigate this, we sought to determine whether all basal epidermal cells were equally competent to signal with their neighbors throughout all stages of their cell cycles or if Ca^{2+} signaling would be more important during specific stages of the cell cycle, as shown in other systems[45, 46, 47]. Further investigation of our data revealed that cells undergoing mitosis do not display competence to participate in intercellular Ca^{2+} signaling, rarely showing any cytosolic Ca^{2+} signaling (**Figure 5A, Movie 3**). To delve deeper into cells' competence in propagating Ca^{2+} signals during mitosis, we treated mice with the drug demecolcine to enrich for mitotic basal cells. Consistent with our observations in homeostatic tissue, enriching for mitotic cells severely dampened Ca^{2+} signals and abrogated intercellular signaling (**Figure 5E, A8A, Movie 4**).

To investigate characteristics of Ca^{2+} signaling in other stages of the cell cycle, we used the Ca^{2+}-sensor combined with the Fucci cell cycle reporter that fluorescently labels G1 and S cells in red[48]. We observed clusters of Ca^{2+} signaling propagating across cells of G1, S, and G2 cell cycle stages (**Figure 5B**). Overall Ca^{2+} signaling in G1 and S cells versus G2 cells based on nuclear red signal revealed similar overall competencies to participate in signaling, regardless of G1, S, or G2 cell cycle stage (**Figure A8B**). Further, the proportion of G1 and S versus G2 cells competent to participate in single, small and large clustered signaling was the same across all neighborhood sizes (**Figure A8C, A8D**).

To investigate Ca^{2+} signaling patterns in basal cells across cell cycle stages in an unbiased manner, we used the cell embeddings based on wavelet coefficients computed during GSTH (see GSTH Algorithmic Overview). The cell embeddings from the geometric scattering transform reflected the overall participation of a cell in signaling over the whole timecourse. This includes the cell itself signaling or being close to (or far from) signaling cells. Clustering these embeddings using spectral clustering[49] on the diffusion operator computed by PHATE allows us to visualize which cells share similar Ca^{2+} signaling patterns and are part of similar neighborhoods of signaling. We can then color each point in the PHATE plot based on its cell cycle stage to see G1, S, or G2 cells and their Ca^{2+} signaling patterns in 3-D space. In four out of five of the PHATE visualizations, G2 cells clustered together, showing related Ca^{2+} signaling patterns (**Figure 5C, A9A**). This was in contrast to G1 and S cells, which were highly dispersed across the PHATE plots, indicating heterogeneous patterns of signaling. These results demonstrate that G2 cells display Ca^{2+} signaling patterns that are more similar to each other in spatial and temporal dimensions than G1 or S cells.

We next set out to understand the relevance of cell cycle phase to the homeostatic characteristics of Ca^{2+} signaling we discovered. Thus, we first interrogated how Ca^{2+} signaling might manifest in a layer of cells enriched for the G1 state, using the Keratin-14 promoter to induce the cell cycle inhibitor Cdkn1b (p27) and combining this with a constitutive Ca^{2+}-sensor (K14rtTA; pTRE-Cdkn1b; Rosa26-CAG-GCaMP6s). Through timelapse recordings, we observed no change in the spatial patterns of Ca^{2+} signaling upon cell cycle inhibition (**Figure A8F)**; however, our analyses revealed a temporal disruption such that neighborhoods larger than 11 cells showed a trend towards a shorter maximal spike duration (**Figure A8E**). Additionally, we noticed a marked decrease in the amount of total Ca^{2+} signaling in these G1-enriched

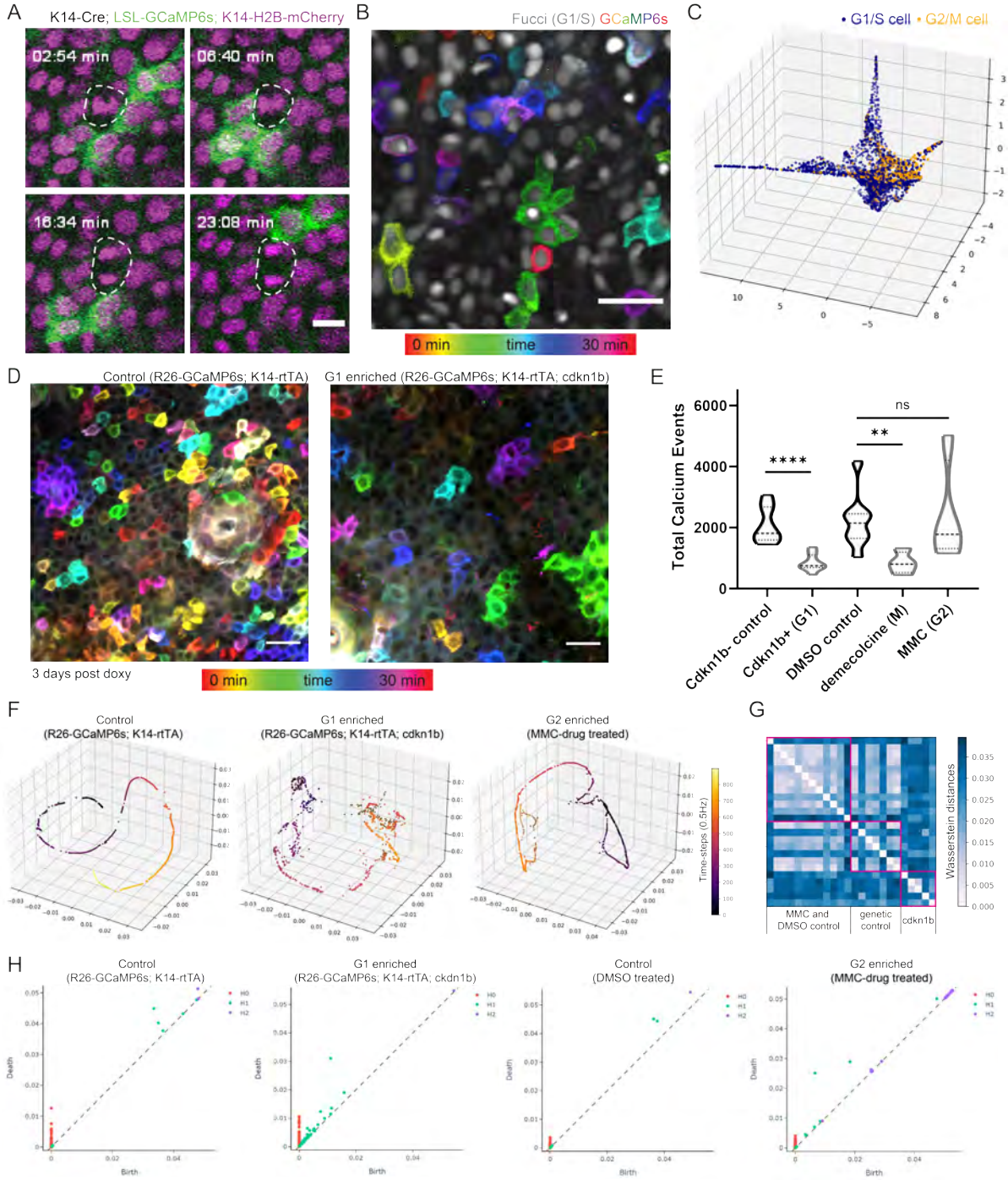

Figure 5: **G2 cells are essential in mediating directed, coordinated patterns of signaling. (A)** Representative example of mitotic cell with lack of Ca^{2+} signaling in live mice (K14-Cre; Rosa26-CAG-LSL-GCaMP6s; K14-H2BmCherry). Nuclei in magenta and intracellular Ca^{2+} levels represented by green fluorescence intensity. Mitosing cells were identified via the visual appearance of mitotic spindles in the nuclear signal. Scale bar: 10 μm. **(B)** Max intensity projection of 30-minute timelapse video of the epidermal basal layer of Ca^{2+}-sensor mouse with a cell cycle reporter (K14-Cre; Rosa26-CAG-LSL-GCaMP6s; Rosa26p-Fucci2). Only the mCherry-hCdt1 expression is visible and is shown in grayscale in the image. Time represented as color scale. Scale bar: 25 μm. **(C)** Representative PHATE plots of cellular embeddings. Each dot represents a single cell; its position in space represents similarity of its Ca^{2+} signaling characteristics with other cells in space; each cell/node is colored by its cell cycle state based on nuclear Fucci signal. **(D)** Max intensity projection of 30-minute timelapse videos of the basal layer of control and G1-enriched Ca^{2+}-sensor mice (Rosa26-CAG-GCaMP6s; K14rtTA and Rosa26-CAG-GCaMP6s; K14rtTA; cdkn1b). Time represented by color scale. Mice imaged 3 days post-induction. Scale bars: 25 μm. **(E)** Total Ca^{2+} events identified in G1-enriched cdkn1b mice, G2-enriched MMC-treated mice, and mitosis-enriched demecolcine-treated mice versus genetic or drug-treated controls. **** $P < 0.0001$, Student's t test, $N = 8$ control and 9 cdkn1b+ movies, ** $P < 0.01$, One-way ANOVA, $N = 8$ DMSO control, 4 demecolcine, and 4 MMC movies. **(F)** PHATE time trajectory visualization of Ca^{2+} signaling in the cdkn1b+ G1-enriched basal layer and MMC-treated G2-enriched basal layer versus control basal layer shows disruption of smooth, directed and coordinated patterns of signaling in G1-enriched mice, whereas the G2-enriched basal layer maintains homeostatic patterns. **(G)** Wasserstein distances from the persistence diagrams of G2-enriched MMC, DMSO control, G1-enriched Cdkn1b, and genetic control mice trajectories show similarity across G2-enriched MMC and control mice and differences when compared to G1-enriched Cdkn1b mice. **H** Representative persistence diagrams (H0: connected components, H1: loops, H2: voids) for G1 and G2 enriched conditions (R26-GCaMP6s; K14-rtTA; cdkn1b 3 days post-induction and MMC drug 2 days post-treatment). The two control groups show similar patterns. The persistence diagram from G1-enriched group has more H1 features/loops and the persistence diagram from G2-enriched group has more H2 features/voids

Cdkn1b mice compared to littermate controls (**Figure 5D, 5E, Movie 5**). Comparison between G1-enriched and control cells at a population level via our GSTH analysis also revealed different patterns between the two groups (**Figure 5F, A10A**). The PHATE trajectories for the control group were smooth, consistent with wildtype tissue, suggesting the change of signals over the graph was generally steady; however, trajectories for the Cdkn1b positive group of G1-enriched cells showed different and more scattered patterns with many loops and holes in the trajectories. Again these were also reflected in the persistence diagrams (based on the PHATE trajectories) and the corresponding Betti curves, where $H1$ features were shorter in the control group. Additionally, Betti curves for the control group show that all loops are formed and closed at later thresholds. These data demonstrate that a homogeneous layer of G1-enriched cells is not able to carry out globally coordinated Ca^{2+} signaling.

We next took advantage of Mitomycin C (MMC) drug treatment to enrich for cells in the G2 cell cycle. In MMC-treated mice, we observed normal Ca^{2+} signaling, consistent with DMSO vehicle-treated controls (**Figure 5E, A8G, Movie 6**). We also analyzed the tissue-wide patterns of Ca^{2+} signaling using GSTH and observed smooth PHATE trajectories (**Figure 5F, A10B**). We were also able to quantify Wasserstein distances between the persistence homology plots of MMC, Cdkn1b, and control populations to quantitatively compare topology of the PHATE plots (**Figure 5G, 5H**). Wasserstein or *earth mover's distances* offer a powerful method to quantify differences between sets based on cost of displacement from one configuration to another. A key advantage of the persistence diagram description of signaling is that Wasserstein distances are well-studied in this context[50]. The Wasserstein distances within the MMC G2-enriched and DMSO control groups were small, suggesting similar PHATE trajectories (hence similar signaling patterns). However, the Wassertein distances between the Cdkn1b G1-enriched and genetic control groups are large, indicating different signaling patterns. By comparing Ca^{2+} signaling in enriched populations of cells in G1, G2, or mitosis, we demonstrated that G2 cells are essential at the tissue-wide level to maintain directed and coordinated patterns of Ca^{2+} signaling, while all G2 cells also displayed similar patterns of signaling during homeostatic adult skin regeneration.

Discussion

Functional *in vivo* studies of Ca^{2+} signaling, especially in non-excitatory tissues, have been limited until recently. Development of genetically encoded Ca^{2+} reporters[22] and new live imaging methods[51, 52, 53] have expanded the realm of possibility. In this study, we built on our two-photon microscopy setup to enable fast live imaging of Ca^{2+} signaling dynamics in the adult mammalian skin epidermis. We applied segmentation methods to be able to analyze thousands of active and inactive populations of epithelial basal cells simultaneously in the skin of live mice, creating rich datasets in which to investigate the regulation and role of Ca^{2+} signaling in homeostasis. We wondered if slowly remodeling dermal components[54, 55, 56] of the skin might be signaling to the basal layer and creating persistent hotspots of activity or whether the observed Ca^{2+} signaling dynamics were characteristic across the remodeling basal layer. Our daily revisits of the same region of the tissue demonstrated that Ca^{2+} signaling is not constricted to regional domains but rather a property shared across the remodeling basal epithelium.

As the signaling dynamics field advances, experimentalists are generating higher dimensional data that is more difficult to interpret. We address these issues by first introducing methods to define neighborhoods of connected signaling and quantify signaling dynamics of different local neighborhood sizes. Additionally, we develop a high throughput, unsupervised machine learning pipeline called GSTH to investigate signaling networks on a global scale and compare across experimental conditions. GSTH incorporates many domains of machine learning, including graph signal processing (to capture patterns of Ca^{2+} spread across the tissue), data geometry (to visualize and represent the dynamic trajectory), and topology (to quantitatively characterize trajectories). We found spatiotemporally coordinated patterns of signaling, as compared to signaling patterns in the neuronal visual cortex, hinting at the unique cohesiveness of the epithelial basal layer, despite its continuous self-renewal (balanced via cells exiting from the basal layer during differentiation and others proliferating to compensate). Our work establishes a way to investigate tissue-wide signaling networks across thousands of cells in many different conditions and tissue types, opening up the possibility of dissecting the functional roles of spatiotemporally diverse patterns of signaling dynamics.

Through these approaches, we discovered that Cx43 modulates intercellular Ca^{2+} signaling within the basal layer and is necessary for the directed, coordinated signaling patterns seen in homeostatic, wildtype tissue. We observed a very rapid disruption in the global coordination of signaling patterns upon loss of Cx43, establishing a distinct regulatory role for Cx43 in the spread of Ca^{2+} across an epithelium. This demonstrates the complexity of Ca^{2+} signaling in this compartment, where many molecular regulators may come together to determine patterns of Ca^{2+} signaling. Recent work

in the upper differentiated layers of the epidermis showed a similar role for voltage-gated ion channels in regulating the temporal dynamics of Ca^{2+} signaling[57].

Our results demonstrate that Cx43 gap junctions are dissociated as cells exit G1 and progress through the cell cycle, creating a centralized mechanism for modulating the coordination of Ca^{2+} flow across thousands of cells and balancing cell cycles in the stem cell pool. Despite differentiation rates remaining stable after loss of Cx43-dependent Ca^{2+} signaling, stem cells across the basal layer were not able to compensate for neighbor loss at rates seen during homeostasis, demonstrating a disruption of the global behavioral coordination that has been shown to characterize this layer[8] and revealing a role for Cx43-mediated Ca^{2+} signaling in maintenance of homeostasis.

While Ca^{2+} signaling has been linked to cell proliferation[58, 59], its role during cell cycle progression is not well understood. We found that when cells are in G2, they display Ca^{2+} signaling patterns that are more similar to each other across spatial and temporal dimensions than when they are in G1. These G2 cells are essential to tissue-level signaling coordination. In contrast to the historical paradigm that implicates Ca^{2+} signaling in mitosis, we find that mitotic cells do not participate in Ca^{2+} signaling. Our study represents the first time Ca^{2+} signaling and cell cycle have been studied in conjunction *in vivo* and sheds light on the interplay between signaling dynamics and the regulation of cell cycle coordination in a stem cell pool. Together, our results provide insight into how complex signaling pathways are regulated and interpreted at a tissue-wide level to maintain proper homeostasis.

Experimental Methods

Mice and experimental conditions

K14-Cre[60] mice were obtained from E. Fuchs (Rockefeller University). R26p-Fucci2[48] mice were obtained from S. Aizawa (RIKEN). K14-H2BmCherry mice were generated in the laboratory and described previously[61]. $Cx43^{fl/fl}$[62, 63], Rosa26-CAG-LSL-GCaMP6s[64], mTmG[65], Sox2-Cre[66], K14-CreER[60], K14-rtTA[67], tetO-Cdkn1b[68] mice were obtained from The Jackson Laboratory. Germline recombined Rosa26-CAG-GCaMP6s mice were generated by crossing Rosa26-CAG-LSL-GCaMP6s to Sox2-Cre mice. To block the cell cycle progression of epithelial cells during G1, Rosa26-CAG-GCaMP6s mice were mated with K14-rtTA; tetO-Cdkn1b mice (Rosa26-CAG-GCaMP6s; K14-rtTA; tetO-Cdkn1b) and given doxycycline ($1\ mg \cdot ml^{-1}$) in potable water with 1% sucrose between P45 and P60. Doxycycline treatment was sustained until imaging was performed three days later. Siblings without the tetO-Cdkn1b allele (Rosa26-CAG-GCaMP6s; K14-rtTA) were used as controls. Mice from experimental and control groups were randomly selected from either sex for live imaging experiments. No blinding was done. All procedures involving animal subjects were performed under the approval of the Institutional Animal Care and Use Committee (IACUC) of the Yale School of Medicine.

In vivo imaging

Imaging procedures were adapted from those previously described[20, 21]. All imaging was performed in distal regions of the ear skin during prolonged telogen, with hair removed using depilatory cream (Nair) at least 2 days before the start of each experiment. Mice were anaesthetized using an isoflurane chamber and then transferred to the imaging stage and maintained on anesthesia throughout the course of the experiment with vaporized isoflurane delivered by a nose cone (1.25% in oxygen and air). Mice were placed on a warming pad during imaging. The ear was mounted on a custom-made stage and a glass coverslip was placed directly against it. Image stacks were acquired with a LaVision TriM Scope II (LaVision Biotec) laser scanning microscope equipped with a tunable Two-photon Vision II Ti:Sapphire (Coherent) Ti:Sapphire laser and tunable Two-photon Chameleon Discovery Ti:Sapphire laser (Coherent) and Imspector Pro (LaVision Biotec, v.7.0.129.0). To acquire serial optical sections, a laser beam ($940nm$, $1120nm$ for mice and whole-mount staining) was focused through a 20x or 40x water-immersion lens (NA 1.0 and 1.1 respectively; Zeiss) and scanned with a field of view of $500\ \mu m^2$ or $304\ \mu m^2$, respectively at 800 Hz or through a 25x water-immersion lens (NA 1.0; Nikon) and scanned with a field of view of $486\ \mu m^2$ at 800 Hz. Z-stacks were acquired in 0.5–3 µm steps to image a total depth of up to 100 µm of tissue. To visualize large areas, 2–64 tiles of optical fields were imaged using a motorized stage to automatically acquire sequential fields of view. Visualization of collagen was achieved via the second harmonic signal at $940nm$. For all time-lapse movies, the live mouse remained anesthetized for the length of the experiment and

serial optical sections were captured at intervals of 2 seconds. For revisits, the same region of live mouse skin was imaged across intervals of multiple days. Anatomical features and patterns of hair follicles and collagen were used as landmarks for finding the same skin location (see Figure S1B).

Image Analysis

Raw image stacks were imported into FIJI (ImageJ, NIH) for analysis. Individual optical planes, summed or max Z stacks of sequential optical sections were used to assemble figures. To prepare movies where the nuclear signal bleached over the course of the timelapse, we used the Fiji Bleach Correction plugin[69], specifying the Simple Ratio Method.

Segmentation of actively signaling cells was performed using the CaImAn MATLAB package as previously described[27]. In order to segment all cells in the field of view, including non-flashing cells, we used part of the MATLAB package from Romano et al[23], a watershed segmentation method. From the raw fluorescence values for each segmented cell, we used peak finding in MATLAB (version R2018b) and then fit Gaussian curves to each peak to be able to quantify spike duration, peak intensity, frequency of flashing, etc. To quantify neighborhood size of clustered signaling, we created a graph for each timelapse, where each node represented one segmented, spiking cell. We connected nodes that represented cells spiking directly adjacent to one another (spatial neighbors) within 10 seconds of each other (temporally correlated). We then counted the number of connected nodes to quantify the size of each signaling neighborhood.

Whole-mount staining

Ear tissue was incubated epidermis side up in $5\ mg \cdot ml^{-1}$ Dispase II solution (Sigma, 4942078001) at 37 °C for 15 min, and epidermis was separated from dermis using forceps. The epidermis was fixed in 4% paraformaldehyde in PBS for 15 min at room temperature, washed and blocked with 0.2% Triton X-100, 5% normal donkey serum, 1% BSA in PBS. The samples were then incubated with primary antibodies for 12 h at 4 degrees and with secondary antibodies for approximately 2 hours at room temperature. Primary antibodies used were as follows: purified mouse anti-Connexin 43, C-terminal, clone P4G9 (1:100, Sigma, MABT901), rabbit anti-Connexin 30.3 polyclonal antibody (1:100, ThermoFisher, 40-0900), rabbit anti-Connexin 26 polyclonal antibody (1:100, ThermoFisher, 71-0500), rabbit anti-Connexin 31 polyclonal antibody (1:100, ThermoFisher, 36-5100), guinea pig anti-K10 (1:200; Progen, GP-K10), rabbit anti-pH3 (1:300; Millipore, 06-570). All secondary antibodies used were raised in a donkey host and were conjugated to AlexaFluor 488, 568, or 647 (Thermofisher). Some tissue was then incubated with Hoechst 33342 (Becton Dickinson; H3570, 1:500) for 15 min, then washed with blocking solution. Finally, the tissue was mounted with Vectashield Anti-fade mounting medium (Vector Laboratories) or SlowFade™ Diamond Antifade Mountant (ThermoFisher) and a $\#1.5$ coverslip and imaged on a LaVision TriM Scope II as described in 'In vivo imaging'.

Tamoxifen Induction

To induce expression of membrane-GFP and/or loss of Cx43 expression, K14-CreER; Cx43$^{fl/fl}$; mTmG mice or K14Cre-ER; Cx43$^{fl/fl}$ mice were given three doses of Tamoxifen ($20mg/kg$ body weight in corn oil) 3, 4, and 5 days before imaging or tissue collection by intraperitoneal injection. In order to observe phenotypes of total loss of Cx43 just one day after recombination, we also topically applied 0.01 mg (Z)-4-Hydroxytamoxifen (4-OHT) in an ethanol-Vaseline slurry to the ear of Rosa26-CAG-GCaMP6s; K14CreER; Cx43$^{fl/fl}$ or Rosa26-CAG-GCaMP6s; K14CreER; Cx43$^{+/+}$ mice one day before the start of imaging.

Topical drug treatment

To stall cells as they transition from S to G2 phase of their cell cycles, Mitomycin C (MMC)[70] was delivered topically to the ear skin. MMC was dissolved in a $15\ mg \cdot ml^{-1}$ stock solution in dimethyl sulfoxide (DMSO) and then diluted 100 times in 100% petroleum jelly (Vaseline; final concentration is $150\ mg \cdot ml^{-1}$). One hundred micrograms of the mixture of the MMC and the petroleum jelly was spread evenly on the ear 1 and 2 days before imaging. A mixture of 100% DMSO

in petroleum jelly (1:100) was used as a vehicle control. Demecolcine was used to block microtubule polymerization[71]. Colcemid was dissolved to $25\ mg \cdot ml^{-1}$ stock solution in DMSO and delivered as described for the MMC treatment.

Statistics and reproducibility

Biostatistical analyses were performed using GraphPad Prism (version 9.2) software (GraphPad Inc., La Jolla, CA). Statistical comparisons were made using an unpaired two-tailed Student's t test, Mann-Whitney test, or the one-way analysis of variance (ANOVA) with multiple comparison's test. Differences between the groups were considered significant at $P < 0.05$, and the data are presented as means $\pm$ standard deviation unless otherwise noted.

Computational Methods

Diffusion Geometry

A useful assumption in representation learning is that high dimensional data originates from an intrinsic low dimensional manifold that is mapped via nonlinear functions to observable high dimensional measurements; this is commonly referred to as the manifold assumption. Formally, let $\mathcal{M}^d$ be a hidden d-dimensional manifold that is only observable via a collection of $n \gg d$ nonlinear functions $f_1, \ldots, f_n : \mathcal{M}^d \to \mathbb{R}$ that enable its immersion in a high dimensional ambient space as $F(\mathcal{M}^d) = \{\mathbf{f}(z) = (f_1(z), \ldots, f_n(z))^T : z \in \mathcal{M}^d\} \subseteq \mathbb{R}^n$ from which data is collected. Conversely, given a dataset $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^n$ of high dimensional observations, manifold learning methods assume data points originate from a sampling $Z = \{z_i\}_{i=1}^N \in \mathcal{M}^d$ of the underlying manifold via $x_i = \mathbf{f}(z_i)$, $i = 1, \ldots, n$, and aim to learn a low dimensional intrinsic representation that approximates the manifold geometry of $\mathcal{M}^d$.

To learn a manifold geometry from collected data, scientists often use the diffusion maps construction of [30] that uses diffusion coordinates to provide a natural global coordinate system derived from eigenfunctions of the heat kernel, or equivalently the Laplace-Beltrami operator, over manifold geometries. This construction starts by considering local similarities defined via a kernel $\mathcal{K}(x, y)$, $x, y \in F(\mathcal{M}^d)$, that captures local neighborhoods in the data. We note that a popular choice for $\mathcal{K}$ is the Gaussian kernel $\exp(-\|x - y\|^2/\sigma)$, where $\sigma > 0$ is interpreted as a user-configurable neighborhood size. However, such neighborhoods encode sampling density information together with local geometric information. To construct a diffusion geometry that is robust to sampling density variations, we use an anisotropic kernel

$$\mathcal{K}(x, y) = \frac{\mathcal{G}(x, y)}{\|\mathcal{G}(x, \cdot)\|_1^\alpha \|\mathcal{G}(y, \cdot)\|_1^\alpha}, \text{ where } \quad \mathcal{G}(x, y) = e^{-\frac{\|x-y\|^2}{\sigma}}, \tag{5}$$

as proposed in [30], where $0 \leq \alpha \leq 1$ controls the separation of geometry from density, with $\alpha = 0$ yielding the classic Gaussian kernel, and $\alpha = 1$ completely removing density and providing a geometric equivalent to uniform sampling of the underlying manifold. Next, the similarities encoded by $\mathcal{K}$ are normalized to define transition probabilities $p(x, y) = \frac{\mathcal{K}(x,y)}{\|\mathcal{K}(x,\cdot)\|_1}$ that are organized in an $N \times N$ row stochastic matrix

$$\mathbf{P}_{ij} = p(x_i, x_j) \tag{6}$$

that describes a Markovian diffusion process over the intrinsic geometry of the data. Finally, a diffusion map [30] is defined by taking the eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ and (corresponding) eigenvectors $\{\phi_j\}_{j=1}^N$ of $\mathbf{P}$, and mapping each data point $x_i \in X$ to an N dimensional vector $\Phi_t(x_i) = [\lambda_1^t \phi_1(x_i), \ldots, \lambda_N^t \phi_N(x_i)]^T$, where t represents a diffusion-time (i.e., number of transitions considered in the diffusion process). In general, as t increases, most of the eigenvalues λ_j^t, $j = 1, \ldots, N$, become negligible, and thus truncated diffusion map coordinates can be used for dimensionality reduction [30].

Note that thus far we have described the diffusion operator construction in the case when datapoints are sampled from a high dimensional space. However, in some cases the connectivity structure of the datapoints may be more apparent then their ambient dimensions. This is true for cases where the data already *comes as a graph* in the case of social networks or

protein interactions. In our case too, the data can be easily turned into a connectivity structure on the basis of spatial adjacency. In these cases, the distance computation is not necessary and one can simply start with the adjacency or connectivity structure.

Cellular Graphs and Graph Signals

We represent the imaged tissue as a graph $G = \{V, E\}$, consisting of nodes $v_i \in V$ and edges $(v_j, v_k) \in E$, where each node v_i represents a cell and a pair of nodes v_j and v_k is connected with an edge based on a predefined criterion. For epithelial cells, we connect nodes that are spatially adjacent as the flow of signals is thought to be between spatially proximal cells. On the other hand, neurons can have long processes that are often hard to image, and therefore we use correlation between neurons' Ca^{2+} signals to connect the neuronal graph. Finally, the connectivity of graph G can be described by its adjacency matrix $\mathbf{A}$, where $\mathbf{A}_{ij} = 1$ if v_i and v_j are connected and 0 otherwise. The degree of each vertex is defined as a diagonal matrix $\mathbf{D}$, where $\mathbf{D}_{ii} = \sum_{j=1}^{i} \mathbf{A}_{ij}$.

Graph signals can associate with each node or edge in a graph. In the Ca^{2+} signaling data, the signals associated with cell v_i is the Ca^{2+} fluorescence intensity at each timestep t. Since every cell has a related Ca^{2+} signal, this signal $X(v_i, t)$ is defined over the whole graph for timestep t.

Geometric scattering for timepoint embeddings

The geometric scattering transform is an unsupervised method for generating embeddings for graph-structured data[31]. It is constructed by applying a cascade of graph wavelet transforms followed by a nonlinear modulus operation such as an absolute value nonlinearity[31, 72]. Graph wavelets are designed based on the diffusion operator (lazy random walks) $\mathbf{R} = \frac{1}{2}\left(\mathbf{I} + \mathbf{A}\mathbf{D}^{-1}\right)$ over a graph, i.e.,

$$\mathbf{\Psi}_0 = \mathbf{I} - \mathbf{R}\,, \quad \mathbf{\Psi}_j = \mathbf{R}^{2^{j-1}} - \mathbf{R}^{2^j} = \mathbf{R}^{2^{j-1}}(\mathbf{I} - \mathbf{R}^{2^{j-1}})\,, \quad j \geq 1\,. \tag{7}$$

The multi-scale nature of graph wavelets allows the geometric scattering transform to traverse the entire graph in one layer, which provides both local and global graph features. Summation of the signal responses is used to obtain invariant graph-level features. Since the summation operation could suppress the high frequency information, it could be complemented by using higher order summary statistics of signal x. Due to the iteration of applying graph wavelets followed by a nonlinaer modulus operation, as shown in Figure 2, geometric scattering transforms can be constructed as in a multi layer (or multi order) architecture. Specifically, the zeroth-order scattering coefficients are calculated by taking statistical moments of the summation of signals, and the first order features are obtained by applying a graph wavelet, which aggregates multiscale information of the graph. Second-order geometric scattering features can further augment first order features by iterating the graph wavelet and absolute value transforms. The collection of graph scattering features provides a rich set of multiscale invariants of the graph G and can be used under both supervised and unsupervised settings for graph embedding.

For a signal $X(t_i) = [X(v_1, t_i), X(v_2, t_i), \ldots, X(v_m, t_i)]$ we compute the zeroth-order scattering coefficients for each vertex/cell for timepoint t_i as follows:

$$S_0(X(v_\ell, t_i)) = \mathbf{R}^{2^J}\mathbf{x}(v_\ell, t_i). \tag{8}$$

The diffusion operator (lazy random walks) $\mathbf{R}$ here works as a low pass filter as shown in [31] and provides local averaging of neighboring cell patterns[73]. Unlike the summation operator that averages all vertex information and suppresses the high frequency information and hence has to be retrieved by higher order statistical moments, this retains finer description of cell/vertex embeddings. Then by concatenating the wavelet coefficients for each cell/vertex at timepoint t_i, we can obtain the corresponding timepoint embedding $S_0(X(t_i))$ for timepoint t_i. Finally, the timepoint embedding for N timepoints can be calculated and the resulting $S_0(X(t)) = \{S_0(X(t_0)), S_0(X(t_1)), \ldots, S_0(X(t_n))\}$ is a feature matrix of dimension $N \times M$, where N is the number of timepoints and M is the number of cells. We hence obtain the zeroth-order scattering coefficients for the N timepoints. The scattering transform here is a result of local averaging of wavelet coefficients.

As in [31], the zeroth-order scattering features can be augmented by first-order scattering features by applying graph wavelets and extracting finer description of high frequency response of a signal $X(t_i)$. Specifically, the first-order scattering coefficients for each time point at each vertex/cell are calculated as

$$S_1(X(j, v_\ell, t_i)) = \mathbf{R}^{2^J} |\mathbf{\Psi}_j \mathbf{x}(v_\ell, t_i)|,\ 1 \leq j \leq J, \tag{9}$$

The value $\mathbf{\Psi}_j \mathbf{x}(v_\ell, t_i)$ aggregates the signal information $\mathbf{x}(v_m, t_i)$ from the vertices v_m that are within 2^j steps of v_ℓ. It responds to sharp transitions or oscillations of the signal $\mathbf{x}$ within the neighborhood of v_ℓ with radius 2^j (in terms of the graph path distance). By concatenating all the vertex/cell embeddings, we can obtain the first order scattering coefficients $S_1(X(t_i))$ for timepoint t_i.

Finally, the second-order scattering coefficients can be obtained by further applying graph wavelets and extract even finer description of high frequency response of the signal $X(t_i)$:

$$S_2(X(j, j', v_\ell, t_i)) = \mathbf{R}^{2^J} |\mathbf{\Psi}_{j'} |\mathbf{\Psi}_j \mathbf{x}(v_\ell, t_i)||, 1 \leq j < j' \leq J \tag{10}$$

The above calculations are conducted for each timepoint and a total of N timepoints. The first-order and second-order scattering transform will generate a feature matrix of shape $N \times (M \times J)$ and $N \times (M \times \frac{J \times (J-1)}{2}))$, respectively, as timepoint embeddings for the N timepoints. Finally, the zeroth-order, first-order and second-order scattering coefficients were combined together as the embeddings for each time point $S(X(t_i))$. The scale of the wavelet J was selected based on the diameter of graphs, and the number of scattering coefficients generated depended on the graph sizes.

PHATE

PHATE is a dimensionality reduction method that captures both local and global nonlinear structure through constructing a diffusion geometry[25]. It computes the diffusion operator as in Equation 6. However, rather than eigendecomposing this operator to find new coordinates, PHATE creates a new distance matrix from $\mathbf{P}$ by defining an M-divergence between datapoints, called *potential distance* as $\mathbf{ID}_{i,j} = \| \log \mathbf{P}^t_{i,:} - \log \mathbf{P}^t_{j,:} \|_2$ between corresponding t-step diffusion probability distributions of the two points.

The advantage of this step is that the information theoretic distance between probabilities emphasizes differences in lower probabilities (corresponding to distant points) as well as high probabilities (corresponding to neighbors), and therefore globally contextualizes the point. The resulting information distance matrix $\mathbf{ID}$ is finally embedded into a low dimensional (2D or 3D) space by metric multidimensional scaling (MDS), and makes it possible to visualize intrinsic geometric information from data. In [25], authors demonstrate that PHATE performs better than all compared methods including diffusion maps and UMAP in preserving *denoised manifold affinity (DeMAP)* in low dimensions, and in particular excels at preserving trajectory structures without shattering.

PHATE trajectories of timepoint embeddings

The time point embeddings $S(X(t_i))$ from geometric scattering form a matrix of dimensions $T \times M$, where T is the number of time points in the data and M is the number of scattering coefficients for each time point. We can visualize these embeddings by applying PHATE. Following our previous description of PHATE, we calculated a distance matrix $\mathbf{D} = \|S(X(t_i)) - S(X(t_j)\|_2$ based on the Euclidean distance between time point embeddings and applied an α-decaying kernel K with a locally-adaptive bandwidth $\epsilon_{k,i}$ corresponding to the k-NN distance of the i-th data point to generate an affinity matrix $\mathbf{W}$ as well as the diffusion operator $\mathbf{P}$. The elements of $\mathbf{W}$ are given by:

$$\mathbf{W}_{i,j} = K_{k,\alpha}(i,j) = \frac{1}{2} \exp\left(-\left(\frac{\mathbf{D}_{i,j}}{\epsilon_{k,i}}\right)^\alpha\right) + \frac{1}{2} \exp\left(-\left(\frac{\mathbf{D}_{i,j}}{\epsilon_{k,j}}\right)^\alpha\right)$$

The decaying factor α regulates the decay rate of the kernel (smaller $\alpha \Rightarrow$ kernel with lighter tails), $\alpha = 2$ corresponding to the Gaussian. The diffusion operator $\mathbf{P}$ can then be obtained by calculating the row-sum of the affinity matrix $\mathbf{W}$ with element $\mathbf{P}_{i,j}$ giving the probability of moving from the i-th to the j-th data point in one time step. The global structure

of the data can be further learned through calculating the tth power of the diffusion operator $\mathbf{P}$, which propagates affinity of the data through diffusion up to a scale of t. The optimal value t for diffusion is automatically chosen to be the knee point of the von Neumann entropy of $\mathbf{P}^t$. This diffusion operator is then log scale transformed and converted to a potential distance matrix $ID((X))$ which is embedded by MDS to result in 3-D PHATE embedding coordinates $E(t) = (E_1(X(t)), E_2(X(t)), E_3(X(t)))$ for each time point t, and point cloud $E = \{E(t_1), E(t_2), \dots E(t_n)\}$.

The 3D coordinates enable visualization of the trajectory, which reflects the time-varying patterns of Ca^{2+} fluorescence data. Thus neighbors in the PHATE embedded trajectories indicate similar signaling patterns even if they occur at distal timepoints. In fact many of the dynamics we notice have loopiness or circularity which motivates the use of topology in the next section.

Persistent homology and topological data analysis

Topological data analysis (TDA) refers to techniques for understanding complex datasets by their topological features, i.e., their connectivity[74]. Here we focus on the topological features of a data graph where the simplest set of topological features are given by the number of connected components b_0 and the number of cycles b_1, respectively. Such counts, also known as the Betti numbers, are coarse graph descriptors that are invariant under graph isomorphisms. Their expressivity is increased by considering a function $f\colon V \times V \to \mathbb{R}$ on the vertices of a graph $G = (V, E)$ with vertex set V and edge set E. Since V has finite cardinality, so does the image $\operatorname{im} f$, i.e., $\operatorname{im} f = \{w_1, w_2, \dots, w_n\}$. Without loss of generality, we assume that $w_1 \leq \dots \leq w_n$. We write G_i for the subgraph induced by filtering according to w_i, such that the edges satisfy $E_i := \{(u, v) \in E \mid \|u - v\|_2^2 \leq w_i\}$. The subgraphs G_i satisfy a nesting property, as $G_1 \subseteq G_2 \subseteq \dots \subseteq G_n$. When analyzing a point cloud, the vertices of each G_i arise from spatial coordinates for the data and w_i constitutes a distance threshold between points, such that G_n is a fully-connected graph, containing all the vertices from V. This is commonly known as the Vietoris-Rips (VR) filtration.

It is then possible to calculate topological features alongside this *filtration* of graphs, tracking their appearance and disappearance as the graph grows. If a topological feature is created in G_i, but destroyed in G_j (it might be destroyed because two connected components merge, for instance), we represent this by storing the point (w_i, w_j) in the *persistence diagram* $\mathcal{D}_f$ associated to G. Another simple descriptor is given by the Betti curve of dimension d of a diagram $\mathcal{D}$, which refers to the sequence of Betti numbers of dimension d in $\mathcal{D}$, evaluated for each threshold w_i.

Persistent homology analysis of PHATE trajectories

In this study, to obtain an invariant characterization of the generated PHATE trajectories E for topological data analysis, we calculated their persistent homology. Specifically, we calculated the persistent homology of E via a Vietoris–Rips filtration $VR_s(E)$. The Vietoris–Rips complex of E is defined as the filtered simplicial complex that contains a subset of E as a simplex if and only if all pairwise distances in the subset are less than or equal to s, i.e., $VR_s(E) = \{\{n_0, ..., n_m\} \mid \forall i, j \; d(i, j) \leq s\}$. We noted here that we could also use the potential distance ID from PHATE, however we directly used the PHATE coordinates and the Euclidean distance for simplicity.

As described above, from $VR_s(E)$, we obtain a set of persistence diagrams Q consisting of birth-death-dimension triples $[b, d, q]$ that describe multiscale topological features of E. Each such point corresponds to a topological feature in the trajectory, which appears at a certain birth time and disappears at a death time. Note that the times are supposed to be understood with respect to the parameter s from above. A point's distance from the diagonal therefore represents the prominence or the eponymous *persistence* of the associated topological feature; higher values indicate that the feature occurs over a large scale, thus increasing its significance. We further calculated the associated Betti curves for each Q, resulting in a simple summary curve $B(Q, q)$ for the qth dimension consisting of the number of points (b_i, d_i) in Q such that $b_i \leq s < d_i$. The Betti curve characterizes the connectivity of $VR_s(E)$ and, by extension, of the Ca^{2+} fluorescence data.

Synthetic dataset for timepoint embeddings

To validate the utility of our method, we first tested it on three synthetic datasets we created, which simulated different signal diffusing scenarios.

We took a graph G created from one of our Ca^{2+} signaling samples with 1867 vertices (cells) and used a normalized graph Laplacian L to diffuse a Dirac signal x defined on node i, where $x_i = 1$ and 0 elsewhere. We diffused the signal over the graph for 300 steps via:

$$x_t = L^t x, t = 1, 2, \ldots, t \tag{11}$$

This resulted in a series of signals $X = \{x_1, x_2, \ldots, x_t\}$ of 300 timesteps, each more diffused than the previous.

Synthetic testcase 1 We first added normalized random noise $\epsilon \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = 0$ and $\sigma = 0.001$ to X and obtained perturbed signals $X_{perturbed}$, with $X_{perturbed} = X + \epsilon$. The individual instances of $X_{perturbed}$ are thus similar but not exactly the same as the original signals. We then combined X and $X_{perturbed}$ to form a new 600-step series of signals.

Synthetic testcase 2 Next we created another signal x' similar to the previously defined Dirac signal x centered on node i. This new signal x' is centered on both node i and node j. In other words, $x'_i = 1, x'_j = 1$ and 0 otherwise. Therefore, the diffusion of this new signal x' on the graph initially was similar to signal x, but eventually diffused to different patterns. We also diffused this signal x' for 300 steps and obtained another signal X'. As in the previous testcase, we combined X and X' to form another series of signals of 600 steps.

Synthetic testcase 3 Finally, we took the first 50 timesteps from X, then starting from timestep 51, we created new signals for each timestep. Specifically, we first removed all signals defined on each cell, then 100 cells were randomly picked to choose one of three signals ($signal1 = sin(\frac{\pi}{20})x$, $signal2 = sin(\frac{\pi}{10})x$, $signal3 = sin(\frac{\pi}{5})x$) to spike for a random interval of 11 timesteps during a total of 550 timesteps. These signals were only defined on each cell and not diffused to other cells. Finally, we combined the 50 timesteps from X with the newly generated signals and formed a series of 600 timesteps.

Comparison of GSTH with PHATE, t-SNE, PCA and UMAP

We compare application of the proposed GSTH method on the three synthetic datasets with approaches that ablate or replace steps in the GSTH method. In particular, we test:

- Applying PHATE directly on the raw input signals to obtain time-trajectories—without the use of the geometric scattering transform.
- Applying PCA on the generated scattering coefficients instead of PHATE.
- Applying t-SNE on the generated scattering coefficients instead of PHATE.
- Applying UMAP on the generated scattering coefficients instead of PHATE.

For the synthetic testcase 1, we aim to compare the approaches for their stability to small perturbations as well as their ability to retrieve signal diffusion dynamics on the graph. As shown in Figure A2A, after applying GSTH, time points with perturbed signals overlapped with time points with original signals, showing scattering transform and PHATE are invariant to small degrees of noise. The smooth trajectory also reflects that the scattering transform and PHATE of GSTH can effectively capture the signal propagation on the graph. By contrast, directly using PHATE on the raw input signals will result in the condensed timepoints in Figure A2C, thus failing to retrieve the dynamics. While applying PCA (**Figure A2D**) and t-SNE (**Figure A2E**) on the generated scattering coefficients can retrieve the dynamics to some extent, Figure A2D shows a more disrupted trajectory and the trajectory from Figure A2E overlaps with itself. Similarly,

applying UMAP (**Figure A2F**) on the generated scattering coefficients also led to overlapping timepoints. All these methods thus failed to reflect the propagation of signals on the graph.

For the second synthetic dataset, we further compare the ability of different approaches to retrieve signal diffusion dynamics on the graph under a more complex scenario. For GSTH (**Figure A2H**) time points from two signal sources formed two branches with their starting points near each other in PHATE coordinates. Thus from one end to the next this is akin to a signal condensing and then diffusing again. As expected, this creates a loop-like structure in the PHATE graph. However, directly applying PHATE on the raw signals (**Figure A2J**) results in multiple scattered points separated from the main trajectory, demonstrating that using PHATE only is not able to fully capture and distinguish the signals. Furthermore, although applying PCA on the scattering coefficients (**Figure A2K**) generates two separate trajectories, they fail to form the loop-like structure as with using GSTH. Applying t-SNE (**Figure A2L**) and UMAP (**Figure A2M**) on the generated scattering coefficients also failed to form loop-like structures.

Finally, for the third synthetic dataset, we aim to simulate the propagation of signals similar to that observed in epithelial cells. This will help us to better understand what types of signals GSTH can capture. The propagation of signals among cells can reflect the communication within groups of cells (corresponding to the clusters of Ca^{2+} transients among epithelial cells using Ca^{2+}-sensor imaging), while each cell itself can also generate independent signals without communicating with other cells (corresponding to single cells spiking using Ca^{2+} fluorescence imaging). As in Figure A2O for GSTH, the time points formed smooth trajectories at first simulating the propagation of signals on the graph, then disrupted trajectories corresponding to random spiking of single cells. In comparison, using PHATE directly on the raw input signals (**Figure A2Q**) results in more dense clustering (for the initial stage when the signal is just diffusing on the graph) and using t-SNE on the scattering coefficients generates more scattered clustering, making it hard to identify the inner dynamics (**Figure A2S**). Although applying PCA (**Figure A2R**) and UMAP (**Figure A2T**) on the scattering coefficients can reflect part of the dynamics, they also generate very condensed trajectories for the early stage, when the signal is simply diffusing on the graph.

In addition, we computed the Wasserstein distances between the persistence diagram from our GSTH method and persistence diagrams from other methods using the three synthetic datasets (Fig. A2G, A2N, A2U). The Wasserstein distance between two persistence diagrams Q_1 and Q_2 is calculated as follows:

$$W_m(Q_1, Q_2) = [\inf_{\eta: Q_1 \to Q_2} \sum_{q_1 \in Q_1} ||q_1 - \eta(q_1)||_\infty^m]^{\frac{1}{m}} \tag{12}$$

where η is a bijection from Q_1 to Q_2. Specifically, we consider the diagonal Δ of persistence diagrams to have infinite multiplicity, i.e. points can be matched to the diagonal. We used Eirene to compute Wasserstein distances, which utilizes the Hungarian algorithm to find the optimal matching. It has been shown that persistence diagrams are stable[33]. Hence by calculating the Wasserstein distances between two persistence diagrams, we can quantify the differences of persistence diagrams. We showed that the persistence diagrams from our GSTH method are different from persistence diagrams produced using other methods ablating different parts of GSTH and visualized the distances with heatmaps. The Wasserstein distances were further calculated for different experimental groups to compare persistence diagrams (MMC drug treatment, Cdkn1b overexpression, and control groups). This allowed us to quantify the similarities within experimental groups and differences across multiple experimental conditions.

Overall, we show that GSTH can effectively capture signal diffusion dynamics on the graph and is stable to small perturbation of signals. These valuable characteristics of GSTH make it possible to investigate the complex spatiotemporal Ca^{2+} signaling patterns and hence eventually enable us to explore the underlying relationships between Ca^{2+} signaling and specific cell behaviors.

Cellular Embedding

We can also generate embeddings for individual cell to explore their signaling patterns across time. For each cell, we then consider the signals $X(v_l) = [X(v_l, t_1), X(v_l, t_2), \ldots, X(v_l, t_n)]$, which are defined on cell v_l across all timepoints as features. We utilize the same diffusion operator R and graph wavelets $\mathbf{\Psi}$ defined as in the timepoint embeddings to learn cellular embeddings. Following the calculations in Equation 8, 9 and 10, we can obtain the wavelet coefficients at each vertex/cell. We then concatenate the coefficients of cells across all timepoints to form the cellular embeddings. The cellular embeddings give us description of cell patterns along time, capturing patterns from the cell itself as well as incorporating larger scale signaling patterns by considering neighboring cells at multiple scales.

Synthetic Dataset for Cell Embedding

Similarly to using synthetic datasets to understand GSTH and the timepoint embeddings, we also created datasets to test our cell embedding methods. Since there are mainly two types of Ca^{2+} signaling patterns observed (single cells spiking and clustered signaling), we aimed to simulate these patterns in the synthetic dataset. Therefore, we created two datasets:

Synthetic testcase 4 This dataset contains both types of Ca^{2+} signaling. Specifically, to simulate cells that belong to clustered intercellular Ca^{2+} signaling, we again diffused a Dirac signal x defined on node i using graph Laplacian L for 20 timesteps as:

$$x_t = L^t x, t = 1, 2, \ldots, t \tag{13}$$

Then for all other cells that do not belong to the Ca^{2+} wave defined above, we defined a single non-diffusing signal on it for a time interval of 5 steps as

$$\hat{x}_t = sin(\frac{1}{20}\pi t) \tag{14}$$

Synthetic testcase 5 In the second dataset we consider two intercellular Ca^{2+} waves resulting from two Dirac signals diffused on the graph. The two Dirac signals defined on two different nodes i and j were diffused for 20 timesteps (following Equation 13). This will result in two flashing waves with similar patterns.

Comparison of Cell Embedding with PHATE using Synthetic Dataset

We first applied geometric scattering to generate wavelet coefficients and then visualized them using PHATE. We colored each data point (each representing a cell in the dataset) with a color scale representing the graph distance to the center cell of each wave, where the Dirac signal was defined initially. We finally compared the results to visualizations where we directly applied PHATE to raw cell signals. For the first synthetic dataset containing both single cell spiking and intercellular Ca^{2+} waves, the geometric scattering transform together with PHATE can clearly distinguish the two types of signaling cells (**Figure A3A, A3C**), with single spiking cells distributed further from the center of flashing waves. In contrast, directly applying PHATE to the cells' raw signals failed to reveal this pattern, as single spiking cells were condensed, suggesting that PHATE alone is not able to distinguish the graph structure and hence showed less information about graph distances. For the second synthetic dataset containing two similar waves (**Figure A3B, A3D**), cells from the two waves show many overlaps. This is as expected, since although the cells have different spatial locations, their signals are all from the diffusion of Dirac signals on the graph. Thus, they also share similar patterns. Although applying PHATE alone to this dataset also demonstrated similar overlap patterns, the trajectory it generated still failed to reflect the changing graph distance that reflects the graph structure.

Data and code availability

The source code can be downloaded from https://github.com/krishnaswamylab/GSTH.

References

[1] Y. Ohno and J. M. Otaki, "Spontaneous long-range calcium waves in developing butterfly wings," *BMC Developmental Biology*, vol. 15, p. 17, 3 2015.

[2] S. Restrepo and K. Basler, "Drosophila wing imaginal discs respond to mechanical injury via slow InsP 3 R-mediated intercellular calcium waves," *Nature Communications*, vol. 7, pp. 1–9, 8 2016.

[3] R. Balaji, C. Bielmeier, H. Harz, J. Bates, C. Stadler, A. Hildebrand, and A.-K. Classen, "Calcium spikes, waves and oscillations in a large, patterned epithelial tissue," *Scientific Reports*, vol. 7, p. 42786, 2 2017.

[4] C. A, T. CL, L. E, B. DD, and M. TM, "Decreased Calcium-Sensing Receptor Expression Controls Calcium Signaling and Cell-To-Cell Adhesion Defects in Aged Skin," *The Journal of investigative dermatology*, 2021.

[5] S. H. Orkin and L. I. Zon, "Hematopoiesis: An Evolving Paradigm for Stem Cell Biology," 2 2008.

[6] E. Fuchs, "Finding One's Niche in the Skin," *Cell Stem Cell*, vol. 4, pp. 499–502, 6 2009.

[7] R. Sender and R. Milo, "The distribution of cellular turnover in the human body," *Nature Medicine*, vol. 27, pp. 45–48, 1 2021.

[8] K. R. Mesa, K. Kawaguchi, K. Cockburn, D. Gonzalez, J. Boucher, T. Xin, A. M. Klein, and V. Greco, "Homeostatic Epidermal Stem Cell Self-Renewal Is Driven by Local Differentiation," *Cell Stem Cell*, vol. 23, pp. 677–686, 11 2018.

[9] M. A. Peters, T. Teramoto, J. Q. White, K. Iwasaki, and E. M. Jorgensen, "A Calcium Wave Mediated by Gap Junctions Coordinates a Rhythmic Behavior in C. elegans," *Current Biology*, vol. 17, no. 18, pp. 1601–1608, 2007.

[10] M. J. Behne, S. Sanchez, N. P. Barry, N. Kirschner, W. Meyer, T. M. Mauro, I. Moll, and E. Gratton, "Major translocation of calcium upon epidermal barrier insult: imaging and quantiWcation via FLIM/Fourier vector analysis," *Arch Dermatol Res*, vol. 303, pp. 103–115, 2011.

[11] B. Darbellay, L. Barnes, W. H. Boehncke, J. H. Saurat, and G. Kaya, "Reversal of murine epidermal atrophy by topical modulation of calcium signaling," *Journal of Investigative Dermatology*, vol. 134, pp. 1599–1608, 6 2014.

[12] J. Kumamoto, M. Goto, M. Nagayama, and M. Denda, "Real-time imaging of human epidermal calcium dynamics in response to point laser stimulation," *Journal of Dermatological Science*, vol. 86, pp. 13–20, 4 2017.

[13] S. Koizumi, K. Fujishita, K. Inoue, Y. Shigemoto-Mogami, M. Tsuda, and K. Inoue, "Ca2+ waves in keratinocytes are transmitted to sensory neurons: The involvement of extracellular ATP and P2Y2 receptor activation," *Biochemical Journal*, vol. 380, pp. 329–338, 6 2004.

[14] A. Celli, D. S. Mackenzie, D. S. Crumrine, C. L. Tu, M. Hupe, D. D. Bikle, P. M. Elias, and T. M. Mauro, "Endoplasmic reticulum Ca2+ depletion activates XBP1 and controls terminal differentiation in keratinocytes and epidermis," *British Journal of Dermatology*, vol. 164, pp. 16–25, 1 2011.

[15] A. Celli, D. S. Mackenzie, Y. Zhai, C.-L. Tu, D. D. Bikle, W. M. Holleran, Y. Uchida, and T. M. Mauro, "SERCA2-Controlled Ca2+-Dependent Keratinocyte Adhesion and Differentiation Is Mediated via the Sphingolipid Pathway: A Therapeutic Target for Darier's Disease," *Journal of Investigative Dermatology*, vol. 132, pp. 1188–1195, 4 2012.

[16] S. H. Yuspa, A. E. Kilkenny, P. M. Steinert, and D. R. Roop, "Expression of murine epidermal differentiation markers is tightly regulated by restricted extracellular calcium concentrations in vitro.," *The Journal of cell biology*, vol. 109, pp. 1207–17, 9 1989.

[17] P. M. Elias, S. K. Ahn, M. Denda, B. E. Brown, D. Crumrine, L. K. Kimutai, L. Kömüves, S. H. Lee, and K. R. Feingold, "Modulations in Epidermal Calcium Regulate the Expression of Differentiation-Specific Markers," *Journal of Investigative Dermatology*, vol. 119, pp. 1128–1136, 11 2002.

[18] T. Toyofuku, M. Yabuki, K. Otsu, T. Kuzuya, M. Hori, and M. Tada, "Intercellular calcium signaling via gap junction in connexin-43- transfected cells," *Journal of Biological Chemistry*, vol. 273, pp. 1519–1528, 1 1998.

[19] T. Murata, T. Honda, G. Egawa, Y. Yamamoto, R. Ichijo, F. Toyoshima, T. Dainichi, and K. Kabashima, "Transient elevation of cytoplasmic calcium ion concentration at a single cell level precedes morphological changes of epidermal keratinocytes during cornification," *Scientific Reports*, vol. 8, p. 6610, 12 2018.

[20] C. M. Pineda, S. Park, K. R. Mesa, M. Wolfel, D. G. Gonzalez, A. M. Haberman, P. Rompolas, and V. Greco, "Intravital imaging of hair follicle regeneration in the mouse," *Nature Protocols*, vol. 10, pp. 1116–1130, 6 2015.

[21] P. Rompolas, E. R. Deschene, G. Zito, D. G. Gonzalez, I. Saotome, A. M. Haberman, and V. Greco, "Live imaging of stem cell and progeny behaviour in physiological hair-follicle regeneration," *Nature*, vol. 487, no. 7408, pp. 496–499, 2012.

[22] T.-W. Chen, T. J. Wardill, Y. Sun, S. R. Pulver, S. L. Renninger, A. Baohan, E. R. Schreiter, R. A. Kerr, M. B. Orger, V. Jayaraman, L. L. Looger, K. Svoboda, and D. S. Kim, "Ultrasensitive fluorescent proteins for imaging neuronal activity," *Nature*, vol. 499, pp. 295–300, 7 2013.

[23] S. A. Romano, V. Pérez-Schuster, A. Jouary, J. Boulanger-Weill, A. Candeo, T. Pietri, and G. Sumbre, "An integrated calcium imaging processing toolbox for the analysis of neuronal population dynamics," *PLoS Computational Biology*, vol. 13, 6 2017.

[24] T. Hiratsuka, I. Bordeu, G. Pruessner, and F. M. Watt, "Regulation of ERK basal and pulsatile activity control proliferation and exit from the stem cell compartment in mammalian epidermis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, pp. 17796–17807, 7 2020.

[25] K. R. Moon, D. van Dijk, Z. Wang, S. Gigante, D. B. Burkhardt, W. S. Chen, K. Yim, A. van den Elzen, M. J. Hirn, R. R. Coifman, N. B. Ivanova, G. Wolf, and S. Krishnaswamy, "Visualizing structure and transitions in high-dimensional biological data," *Nature Biotechnology*, vol. 37, pp. 1482–1492, Dec. 2019.

[26] N. Halidi, F. Alonso, J. M. Burt, J. L. Bény, J. A. Haefliger, and J. J. Meister, "Intercellular calcium waves in primary cultured rat mesenteric smooth muscle cells are mediated by connexin43," *Cell Communication and Adhesion*, vol. 19, pp. 25–37, 4 2012.

[27] A. Giovannucci, J. Friedrich, P. Gunn, J. Kalfon, B. L. Brown, S. A. Koay, J. Taxidis, F. Najafi, J. L. Gauthier, P. Zhou, B. S. Khakh, D. W. Tank, D. B. Chklovskii, and E. A. Pnevmatikakis, "CaImAn an open source tool for scalable calcium imaging data analysis," *eLife*, vol. 8, 1 2019.

[28] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, pp. 83–98, 2013.

[29] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Graph fourier transform," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6167–6170, 2013.

[30] R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 53–94, 2006.

[31] F. Gao, G. Wolf, and M. Hirn, "Geometric scattering for graph data analysis," in *Proceedings of the 36th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.), vol. 97 of *Proceedings of Machine Learning Research*, pp. 2122–2131, PMLR, 09–15 Jun 2019.

[32] F. Gao, G. Wolf, and M. Hirn, "Geometric Scattering for Graph Data Analysis," in *Proceedings of the 36th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.), vol. 97 of *Proceedings of Machine Learning Research*, pp. 2122–2131, PMLR, 2019.

[33] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, "Stability of persistence diagrams," *Discrete & computational geometry*, vol. 37, no. 1, pp. 103–120, 2007.

[34] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and Y. Mileyko, "Lipschitz functions have L_p-stable persistence," *Foundations of Computational Mathematics*, vol. 10, pp. 127–139, Apr. 2010.

[35] B. Rieck, F. Sadlo, and H. Leitte, "Topological machine learning with persistence indicator functions," in *Topological Methods in Data Analysis and Visualization V* (H. Carr, I. Fujishiro, F. Sadlo, and S. Takahashi, eds.), pp. 87–101, Cham, Switzerland: Springer, 2020.

[36] C. Stringer, M. Pachitariu, C. B. Reddy, M. Carandini, and K. D. Harris, "Recordings of ten thousand neurons in visual cortex during spontaneous behaviors," 2018.

[37] C. Stringer, M. Pachitariu, N. Steinmetz, C. B. Reddy, M. Carandini, and K. D. Harris, "Spontaneous behaviors drive multidimensional, brainwide activity," *Science*, vol. 364, 4 2019.

[38] J. M. Churko and D. W. Laird, "Gap Junctions," *Cellular Domains*, pp. 339–347, 2011.

[39] D. W. Laird, "The gap junction proteome and its relationship to disease," *Trends in Cell Biology*, vol. 20, pp. 92–101, 2 2010.

[40] C. A. Scott, D. Tattersall, E. A. O'Toole, and D. P. Kelsell, "Connexins in epidermal homeostasis and skin disease," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1818, pp. 1952–1961, 8 2012.

[41] P. E. Martin and M. van Steensel, "Connexins and skin disease: insights into the role of beta connexins in skin homeostasis," *Cell and Tissue Research*, vol. 360, no. 3, pp. 645–658, 2015.

[42] M. Kretz, K. Maass, and K. Willecke, "Expression and function of connexins in the epidermis, analyzed with transgenic mouse mutants," *European Journal of Cell Biology*, vol. 83, pp. 647–654, 1 2004.

[43] M. M. Lurtz and C. F. Louis, "Intracellular calcium regulation of connexin43," *American Journal of Physiology-Cell Physiology*, vol. 293, pp. C1806–C1813, 12 2007.

[44] K. Cockburn, K. Annusver, S. Ganesan, K. R. Mesa, K. Kawaguchi, M. Kasper, and V. Greco, "Gradual differentiation uncoupled from cell cycle exit generates heterogeneity in the epidermal stem cell layer," *bioRxiv*, 2021.

[45] K. A. Skelding, J. A. Rostas, and N. M. Verrills, "Controlling the cell cycle: The role of calcium/calmodulin-stimulated protein kinases I and II," *Cell Cycle*, vol. 10, pp. 631–639, feb 2011.

[46] T. Capiod, "Cell proliferation, calcium influx and calcium channels," *Biochimie*, vol. 93, pp. 2075–2079, 12 2011.

[47] M. J. Rosendo-Pineda, J. J. Vicente, O. Vivas, J. Pacheco, A. Loza-Huerta, A. Sampieri, L. Wordeman, C. Moreno, and L. Vaca, "Phosphorylation of NMDA receptors by cyclin B/CDK1 modulates calcium dynamics and mitosis," *Communications Biology*, vol. 3, no. 1, pp. 1–13, 2020.

[48] T. Abe, A. Sakaue-Sawano, H. Kiyonari, G. Shioi, K. I. Inoue, T. Horiuchi, K. Nakao, A. Miyawaki, S. Aizawa, and T. Fujimori, "Visualization of cell cycle in mouse embryos with Fucci2 reporter directed by Rosa26 promoter," *Development (Cambridge)*, vol. 140, no. 1, pp. 237–246, 2013.

[49] U. von Luxburg, "A tutorial on spectral clustering," 2007.

[50] G. Peyré and M. Cuturi, "Computational optimal transport," 2020.

[51] P. Rompolas, K. R. Mesa, K. Kawaguchi, S. Park, D. Gonzalez, S. Brown, J. Boucher, A. M. Klein, and V. Greco, "Spatiotemporal coordination of stem cell commitment during epidermal homeostasis," *Science*, vol. 352, no. 6292, pp. 1471–1474, 2016.

[52] F. B. Shipley, N. Dani, H. Xu, C. Deister, J. Cui, J. P. Head, C. Sadegh, R. M. Fame, M. L. Shannon, V. I. Flores, T. Kishkovich, E. Jang, E. M. Klein, G. P. Goldey, K. He, Y. Zhang, M. J. Holtzman, T. Kirchhausen, C. Wyart, C. I. Moore, M. L. Andermann, and M. Lehtinen, "Tracking Calcium Dynamics and Immune Surveillance at the Choroid Plexus Blood-Cerebrospinal Fluid Interface," *SSRN Electronic Journal*, pp. 1–17, 2020.

[53] A. J. Stevenson, G. Vanwalleghem, T. A. Stewart, N. D. Condon, B. Lloyd-Lewis, N. Marino, J. W. Putney, E. K. Scott, A. D. Ewing, and F. M. Davis, "Multiscale imaging of basal cell dynamics in the functionally mature mammary gland," *Proceedings of the National Academy of Sciences*, vol. 117, pp. 26822–26832, 10 2020.

[54] E. Marsh, D. G. Gonzalez, E. A. Lathrop, J. Boucher, and V. Greco, "Positional Stability and Membrane Occupancy Define Skin Fibroblast Homeostasis In Vivo," *Cell*, vol. 175, pp. 1620–1633, 11 2018.

[55] D. Peña-Jimenez, S. Fontenete, D. Megias, C. Fustero-Torre, O. Graña-Castro, D. Castellana, R. Loewe, and M. Perez-Moreno, " Lymphatic vessels interact dynamically with the hair follicle stem cell niche during skin regeneration in vivo ," *The EMBO Journal*, vol. 38, pp. 1–17, 10 2019.

[56] Y. C. Hsu, L. Li, and E. Fuchs, "Emerging interactions between skin stem cells and their niches," 2014.

[57] T. Matsui, N. Kadono-Maekubo, Y. Suzuki, Y. Furuichi, K. Shiraga, H. Sasaki, A. Ishida, S. Takahashi, T. Okada, K. Toyooka, J. Sharif, T. Abe, H. Kiyonari, M. Tominaga, A. Miyawaki, and M. Amagai, "A unique mode of keratinocyte death requires intracellular acidification," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, no. 17, pp. 1–10, 2021.

[58] S. A. Gudipaty, J. Lindblom, P. D. Loftus, M. J. Redd, K. Edes, C. F. Davey, V. Krishnegowda, and J. Rosenblatt, "Mechanical stretch triggers rapid epithelial cell division through Piezo1.," *Nature*, vol. 543, pp. 118–121, mar 2017.

[59] H. Deng, A. A. Gerencser, and H. Jasper, "Signal integration by Ca 2+ regulates intestinal stem-cell activity," *Nature*, vol. 528, no. 7581, pp. 212–217, 2015.

[60] V. Vasioukhin, L. Degenstein, B. Wise, and E. Fuchs, "The magical touch: Genome targeting in epidermal stem cells induced by tamoxifen application to mouse skin," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, pp. 8551–8556, 7 1999.

[61] K. R. Mesa, P. Rompolas, G. Zito, P. Myung, T. Y. Sun, S. Brown, D. G. Gonzalez, K. B. Blagoev, A. M. Haberman, and V. Greco, "Niche-induced cell death and epithelial phagocytosis regulate hair follicle stem cell pool.," *Nature*, vol. 522, no. 7554, pp. 94–97, 2015.

[62] M. R. Calera, H. L. Topley, Y. Liao, B. R. Duling, D. L. Paul, and D. A. Goodenough, "Connexin43 is required for production of the aqueous humor in the murine eye," *Journal of Cell Science*, vol. 119, pp. 4510–4519, 11 2006.

[63] Y. Liao, K. H. Day, D. N. Damon, and B. R. Duling, "Endothelial cell-specific knockout of connexin 43 causes hypotension and bradycardia in mice," *Proceedings of the National Academy of Sciences*, vol. 98, pp. 9989–9994, 8 2001.

[64] L. Madisen, A. R. Garner, D. Shimaoka, A. S. Chuong, N. C. Klapoetke, L. Li, A. van der Bourg, Y. Niino, L. Egolf, C. Monetti, H. Gu, M. Mills, A. Cheng, B. Tasic, T. N. Nguyen, S. M. Sunkin, A. Benucci, A. Nagy, A. Miyawaki, F. Helmchen, R. M. Empson, T. Knöpfel, E. S. Boyden, R. C. Reid, M. Carandini, and H. Zeng, "Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance," *Neuron*, vol. 85, pp. 942–958, 3 2015.

[65] M. D. Muzumdar, B. Tasic, K. Miyamichi, L. Li, and L. Luo, "A global double-fluorescent Cre reporter mouse," *genesis*, vol. 45, pp. 593–605, 9 2007.

[66] S. Hayashi, P. Lewis, L. Pevny, and A. P. McMahon, "Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain," *Mechanisms of Development*, vol. 119, 12 2002.

[67] W. Xie, L. T. Chow, A. J. Paterson, E. Chin, and J. E. Kudlow, "Conditional expression of the ErbB2 oncogene elicits reversible hyperplasia in stratified epithelia and up-regulation of TGFα expression in transgenic mice," *Oncogene*, vol. 18, pp. 3593–3607, 6 1999.

[68] S. C. Pruitt, A. Freeland, M. E. Rusiniak, D. Kunnev, and G. K. Cady, "Cdkn1b overexpression in adult mice alters the balance between genome and tissue ageing," *Nature Communications*, vol. 4, 10 2013.

[69] K. Miura, "Bleach correction imagej plugin for compensating the photobleaching of time-lapse sequences [version 1; peer review: 4 approved, 1 approved with reservations]," *F1000Research*, vol. 9, no. 1494, 2020.

[70] M. Tomasz, "Mitomycin C: small, fast and deadly (but very selective)," *Chemistry & Biology*, vol. 2, pp. 575–579, sep 1995.

[71] C. Rieder and R. Palazzo, "Colcemid and the mitotic cycle," *Journal of Cell Science*, vol. 102, pp. 387–392, 07 1992.

[72] F. Gama, A. Ribeiro, and J. Bruna, "Diffusion scattering transforms on graphs," in *International Conference on Learning Representations*, 2019.

[73] M. Perlmutter, F. Gao, G. Wolf, and M. Hirn, "Geometric wavelet scattering networks on compact Riemannian manifolds," in *Proceedings of The First Mathematical and Scientific Machine Learning Conference* (J. Lu and R. Ward, eds.), vol. 107 of *Proceedings of Machine Learning Research*, pp. 570–604, PMLR, 20–24 Jul 2020.

[74] G. Carlsson, "Topology and data," *Bulletin of the American Mathematical Society*, vol. 46, no. 2, pp. 255–308, 2009.

Acknowledgements: We thank all members of the Greco and Krishnaswamy labs for critical feedback on the manuscript. We thank S. Aizawa for the Rosa26p-Fucci2 mice. Finally, we'd like to thank Rachael Norris for her expertise and input on the manuscript.

Funding: This work is supported by an HHMI Scholar award and NIH grants number 1R01AR063663-01, 1R01AR067755-01A1, 1DP1AG066590-01 and R01AR072668 (VG). J.M. was supported by the Lo Graduate Fellowship for Excellence in Stem Cell Research and NIH grants number 2T32GM007499-41A1 and 5T32HD007149-40.

Author contributions: J.M., F.G., S.K. and V.G designed experiments. J.M. performed 2-photon imaging, whole mount staining, mouse genetics, and image analysis. F.G. performed data analysis and statistical modeling. C.M.-M. performed

2-photon imaging and image analysis. S.D. and E.L. performed whole mount staining and image analysis. S.G. performed mouse genetics. L.S. assisted with image analysis. D.B. and B.R. assisted with statistical modeling. J.M., F.G., A.C., C.H., S.K., and V.G wrote the manuscript with input throughout from S.D., C.M.-M., and B.R.. **Competing interests:** The authors declare no competing financial interests.

Data availability: All data from this study are available from the authors on request. The MATLAB and python scripts for the image analysis will be available on request. The source code for GSTH and the cell embeddings can be downloaded from https://github.com/krishnaswamylab/GSTH.

1 Appendix

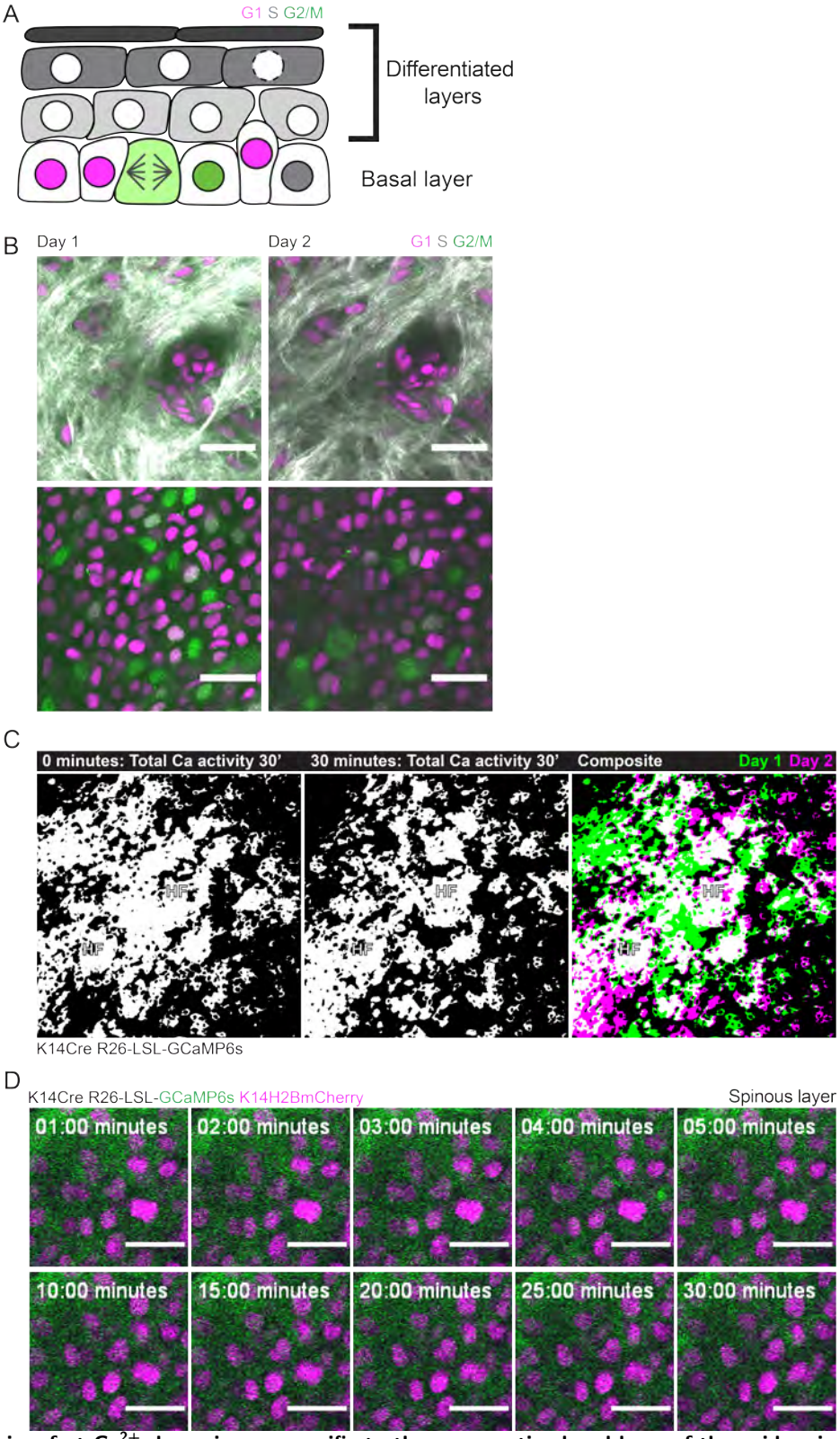

Figure A1: **Pervasive, fast Ca^{2+} dynamics are specific to the regenerative basal layer of the epidermis. (A)** Schematic of the layers of the interfollicular epidermis, with basal cells' cell cycle state represented by nuclear color. **(B)** Revisit of the same region of a mouse at 0 and 24 hours. Top panel show dermis and bottom panel shows epidermal basal layer. Collagen is in white, G1 cells are in magenta, S cells are double positive for magenta and green and shown in gray, G2 and M cells are in green. Scale bars: 25 µm. **(C)** Maximum intensity projection of all optical sections of a 30-minute time-lapse at 0- (green) and 30-minutes (magenta) of the same region of the epidermis. To the right, composite image of the same region at 0- and 30-minutes, where white indicates overlapping regions of Ca^{2+} activity. Transverse views of the top of the infundibulum region of hair follicles marked with HF. **(D)** Region of the spinous layer over 30 minutes of imaging (K14-Cre; Rosa26-CAG-LSL-GCaMP6s; K14-H2BmCherry) with nuclei in red and Ca^{2+} reporter in green.

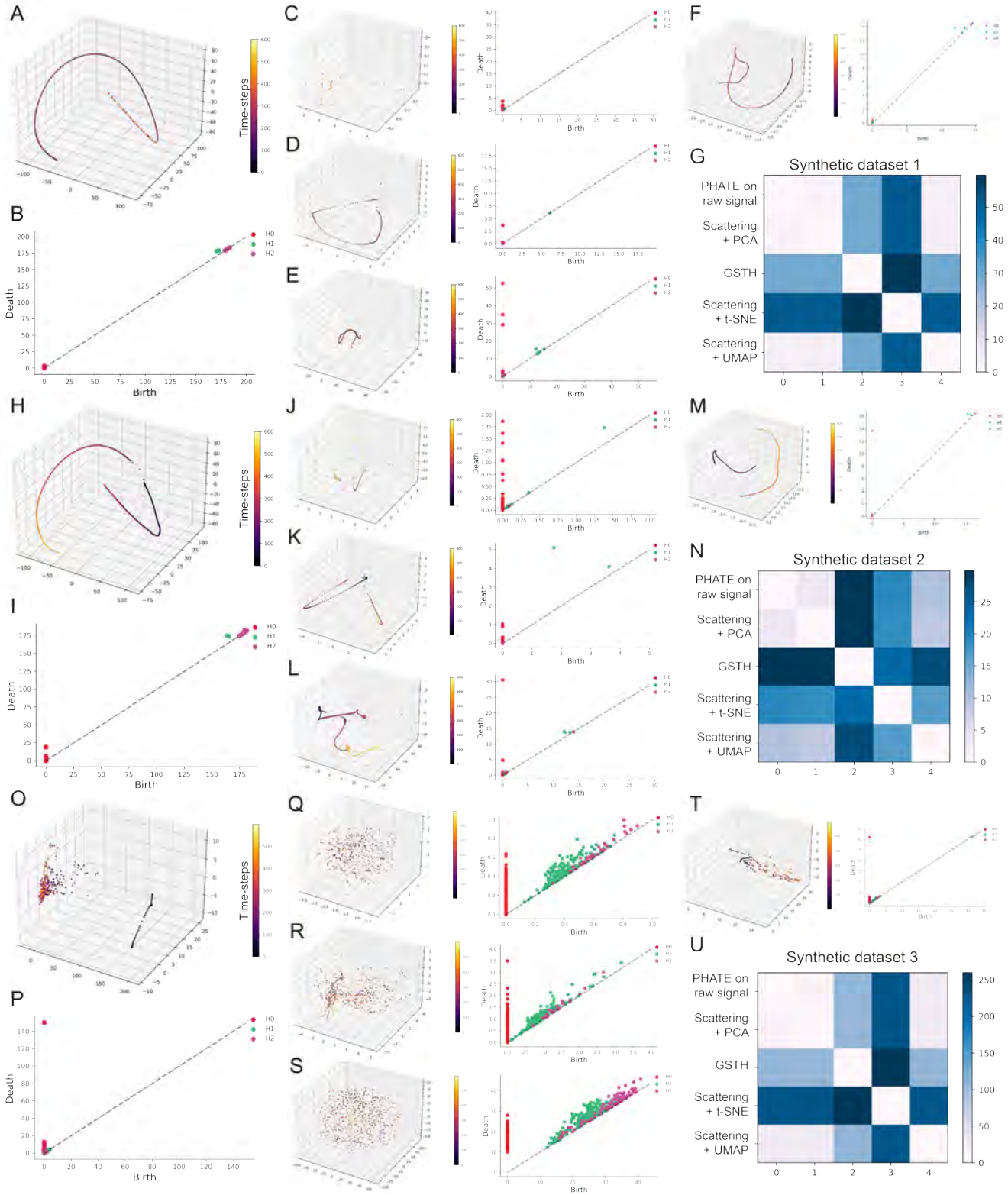

Figure A2: PHATE visualization **(A, H, O)** and persistence homology **(B, I, P)** on synthetic data using GSTH and comparison with (1) directly applying PHATE on the input signals **(C, J, Q)**; (2) PCA on generated scattering coefficients **(D, K, R)**; (3) t-SNE on generated scattering coefficients **(E, L, S)**; (4) UMAP on generated scattering coefficients **(F, M, T)**. Finally, Wasserstein distances from the persistence diagrams of each methodology for each of the 3 synthetic datasets **(G, N, and U)**

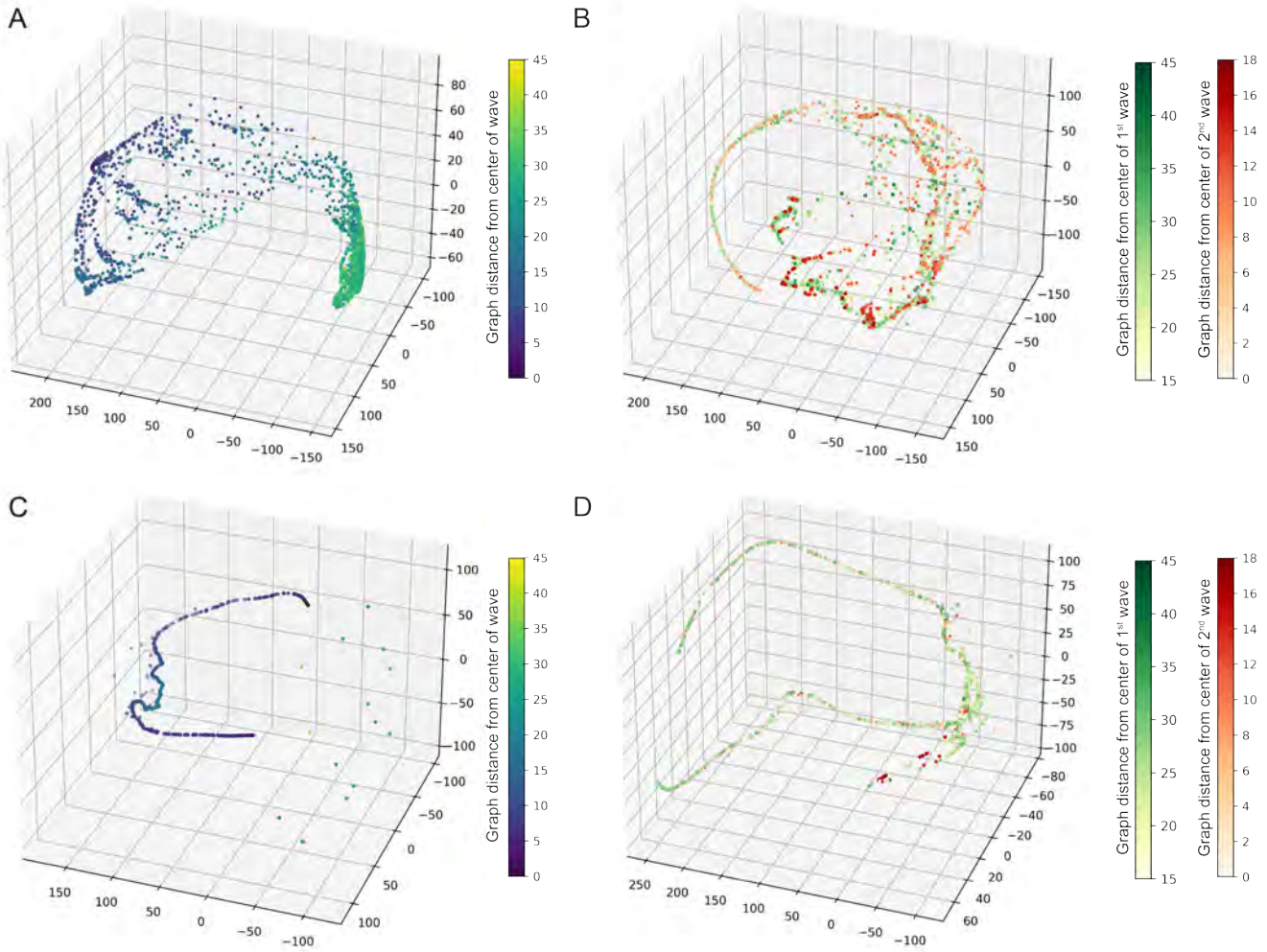

Figure A3: **(A)** PHATE visualization of cellular embeddings generated with geometric scattering transform for synthetic dataset 4. **(B)** PHATE visualization of cellular embeddings generated with geometric scattering transform for synthetic dataset 5. **(C)** PHATE visualization alone of cells with raw signals for synthetic dataset 4. **(D)** PHATE visualization alone of cells with raw signals for synthetic dataset 5.

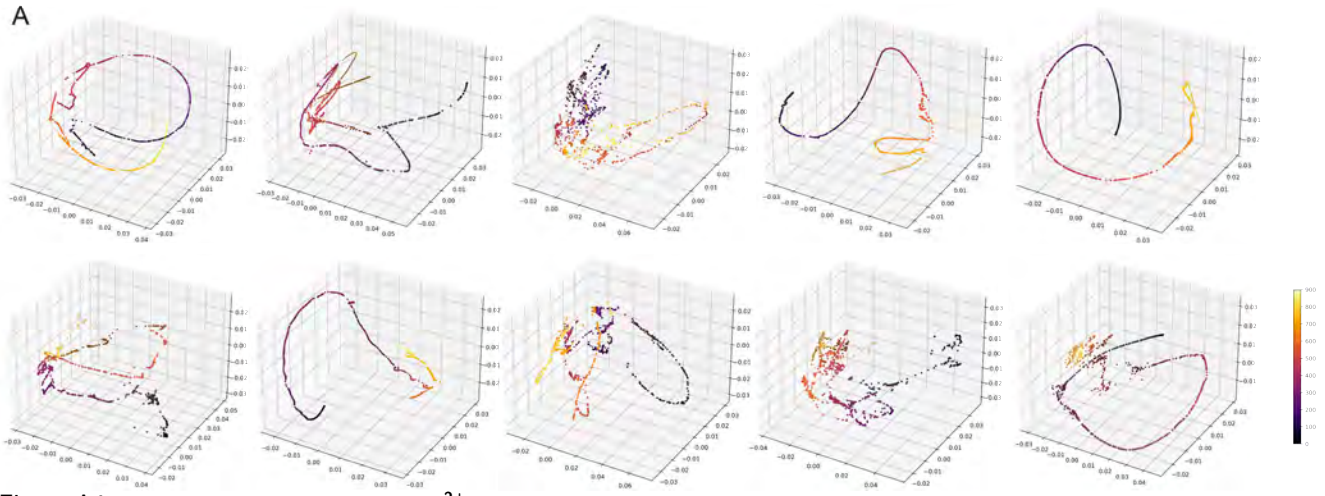

Figure A4: **Unsupervised modeling of Ca^{2+} signaling patterns reveals smooth, directed signaling in the homeostatic basal epidermis. (A)** PHATE visualizations of Ca^{2+} signaling time trajectories in the homeostatic basal epithelial layer from 30-minute time-lapse movies show mainly smooth trajectories.

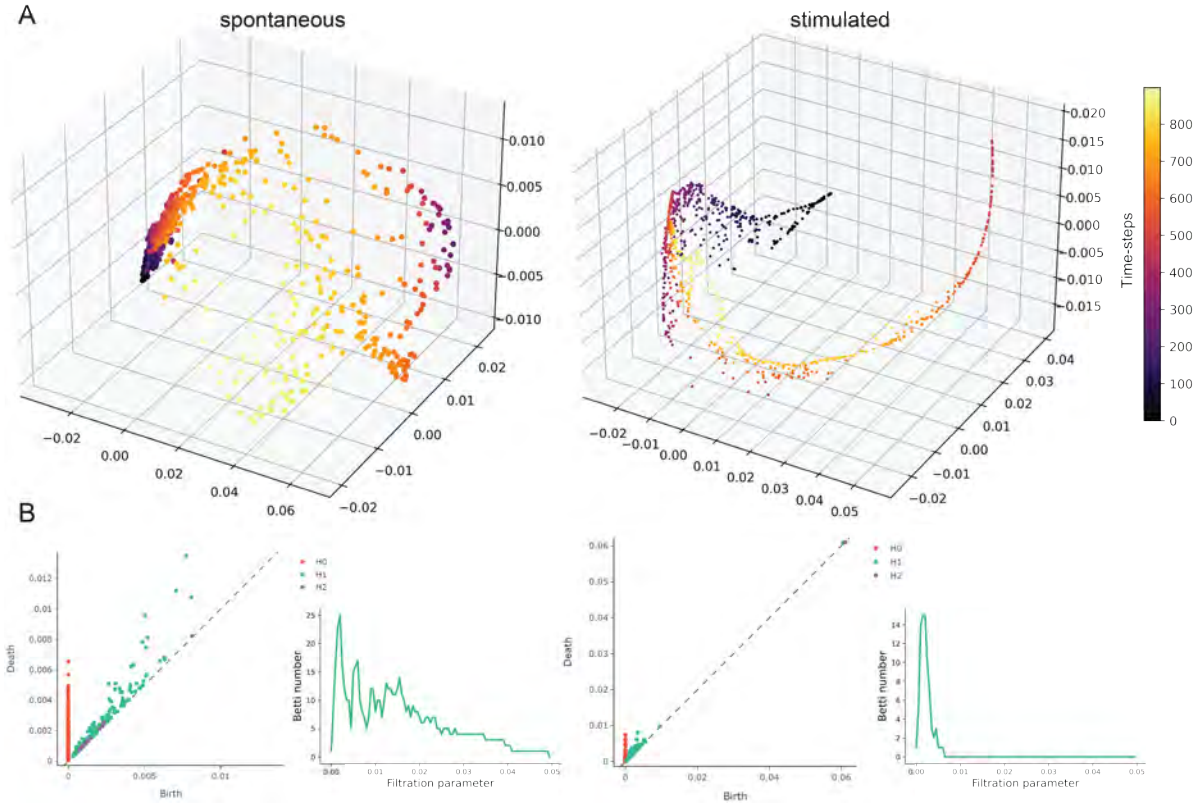

Figure A5: **(A)** PHATE visualization of Ca^{2+} signaling patterns in visual cortex with spontaneous or stimulated neuronal activity. **(B)** Representative persistence diagrams (H0: connected components, H1: loops, H2: voids) and Betti curves of H1 features for spontaneous and stimulated neurons of the visual cortex. The persistence diagram for spontaneous activity has a rich collection of H2 features/voids, which is less common in stimulated activity. In addition, the H1 feature/loops from the spontaneous activity shows longer persistence

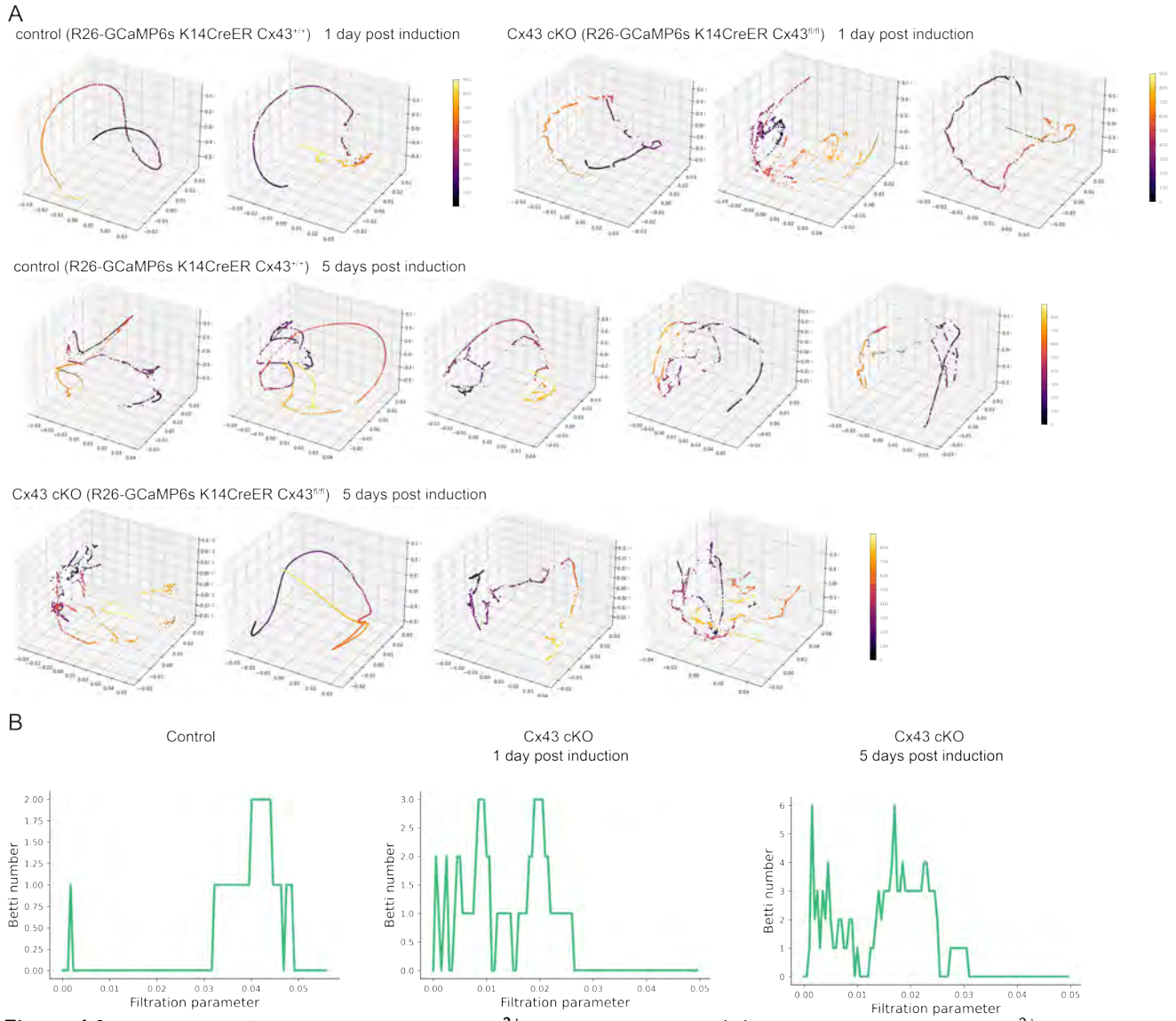

Figure A6: **Loss of Cx43 disrupts coordinated Ca^{2+} signaling patterns (A)** PHATE visualization of Ca^{2+} signaling time trajectories in the Cx43 conditional knockout versus control basal layer shows disruption of smooth, directed and coordinated patterns of signaling in mice 1 and 5 days after loss of Cx43. **(B)** Representative Betti curves of H1 features for control and Cx43 cKO mice (Rosa26-CAG-GCaMP6s; K14-CreER; $Cx43^{+/+}$ and Rosa26-CAG-GCaMP6s; K14-CreER; $Cx43^{fl/fl}$) 1 and 5 days post-tamoxifen induction.

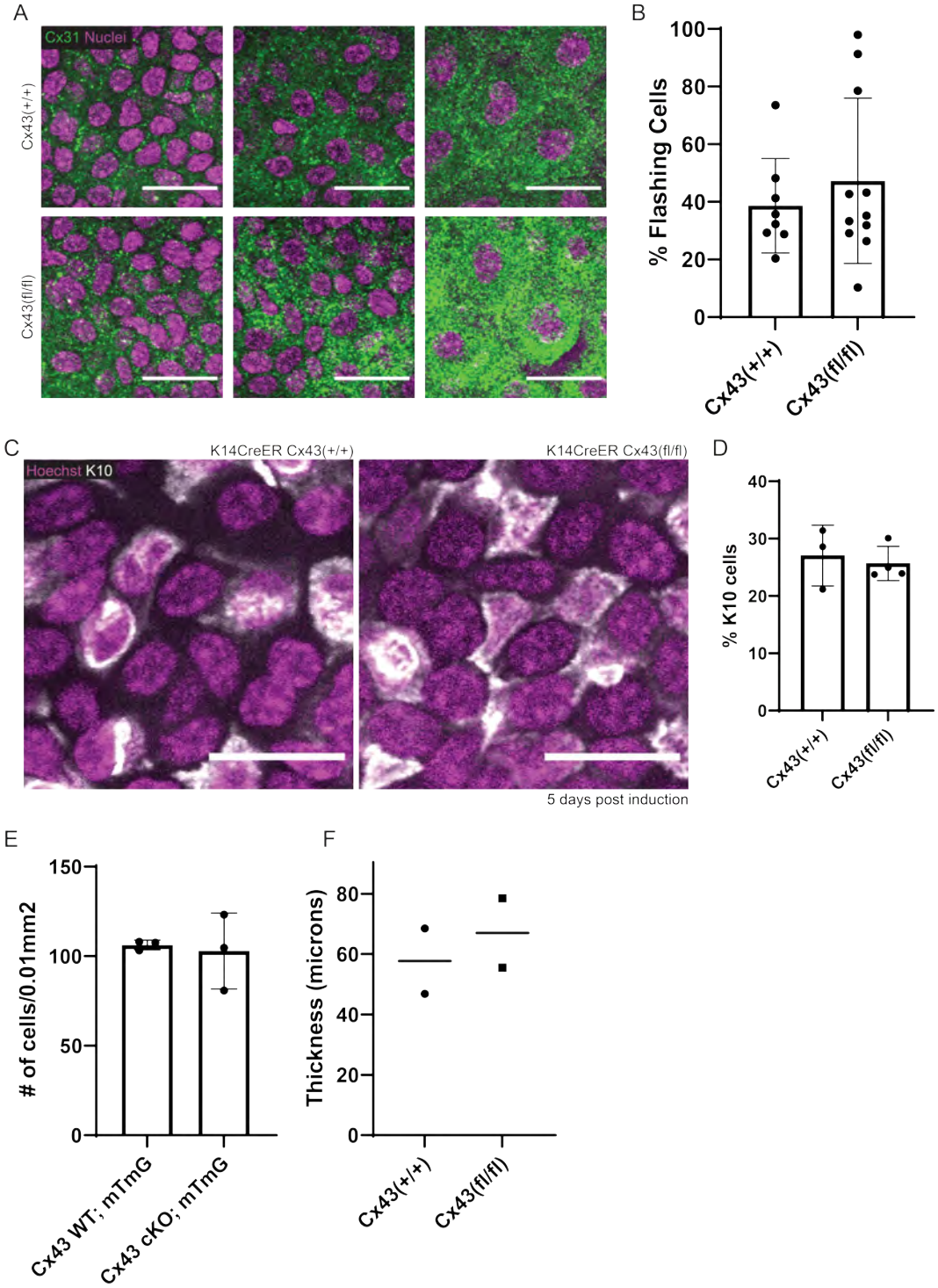

Figure A7: **Gap junctions are maintained in the absence of Cx43. (A)** Immunofluorescence staining of epidermal basal(left), spinous(middle), and granular(right) layer from K14-CreER; Cx43$^{+/+}$ and K14-CreER; Cx43$^{fl/fl}$ mice 5 days post tamoxifen induction, with staining for Cx31 in green and Hoechst marking nuclei in magenta. Scale bar: 25 μm. **(B)** Quantification of percent of flashing cells in control versus Cx43 cKO mice 1, 5, and 7 days post-tamoxifen induction. N = 11 (control) and 14 (Cx43 cKO) thirty-minute time-lapse movies from at least 3 mice per condition. **(C)** K10 immunofluorescence staining in control versus Cx43 cKO mice (K14-CreER Cx43$^{+/+}$ and K14-CreER Cx43$^{fl/fl}$ respectively) 5 days post-tamoxifen induction. Scale bars: 25 μm. **(D)** Quantification of K10 positive basal cells as a percentage of total basal cells 5 days post-tamoxifen induction in Cx43 cKO and control mice, indicating cells that are beginning differentiation. NS, Student's t test. N = four 10 mm^2 regions per mouse, 3 mice per experimental group. **(E)** Quantification of average cell density in control versus Cx43 cKO mice (K14-CreER Cx43$^{+/+}$ and K14-CreER Cx43$^{fl/fl}$ respectively) 5 days post-tamoxifen induction. N = six 10 mm^2 regions per mouse, 3 mice per experimental group. **(F)** Quantification of epidermal thickness in control versus Cx43 cKO mice (K14-CreER Cx43$^{+/+}$ and K14-CreER Cx43$^{fl/fl}$ respectively) 5 days post-tamoxifen induction. N = six 10 mm^2 regions per mouse, 2 mice per experimental group.

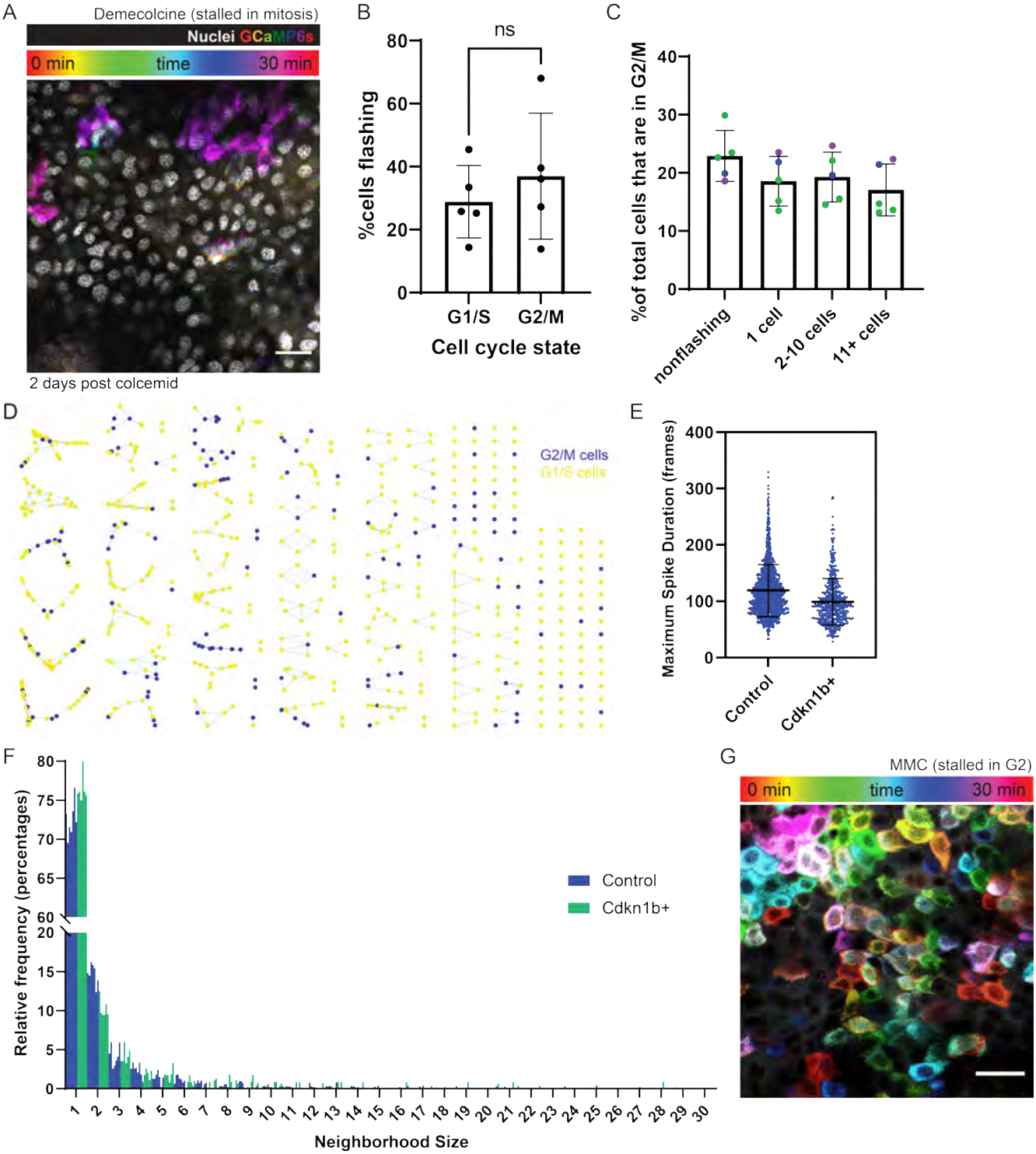

Figure A8: **(A)** Maximum intensity projection of a 30-minute time-lapse video of the epidermal basal layer of a live Ca^{2+} reporter mouse two days after treatment with demecolcine stalling cells in mitosis (K14-Cre; Rosa26-CAG-LSL-GCaMP6s; K14-H2BmCherry). Ca^{2+} sensor fluorescence over time is represented as a color scale and nuclei are shown in white. Scale bar: 25 μm. **(B)** Percent of G1/S versus G2/M cells flashing over the course of 30 minutes. N = 5 thirty-minute time-lapse movies from 3 individual mice. **(C)** Percent of G2/M cells in groups of non-flashing, single flashing, small clusters, and large clusters of flashing cells based on mCherry-hCdt1 expression. N = 5 thirty-minute time-lapse movies from 3 individual mice. **(D)** "Neighborhoods" of spatiotemporally connected Ca^{2+} signaling colored by cell cycle stage (blue = G1/S and yellow = G2/M) from a representative 30-minute time-lapse movie. **(E)** Maximal spike duration (maximum number of frames between the start and end of individual Ca^{2+} events) in control versus G1-stalled cdkn1b+ mice. Bars denote mean and error bars represent SD. N = 9 control and 8 cdkn1b+ thirty-minute time-lapse movies from ¿3 mice per condition. **(F)** Histogram showing relative frequency of different neighborhood sizes of spatiotemporally connected Ca^{2+} signaling from 30-minute time-lapses of cdkn1b+ G1-stalled basal layers (green) versus control (blue) in Ca^{2+} sensor mice. **(G)** Cropped regions from the basal layer of a representative thirty-minute time-lapse from mice treated with MMC to stall cells in S/G2.

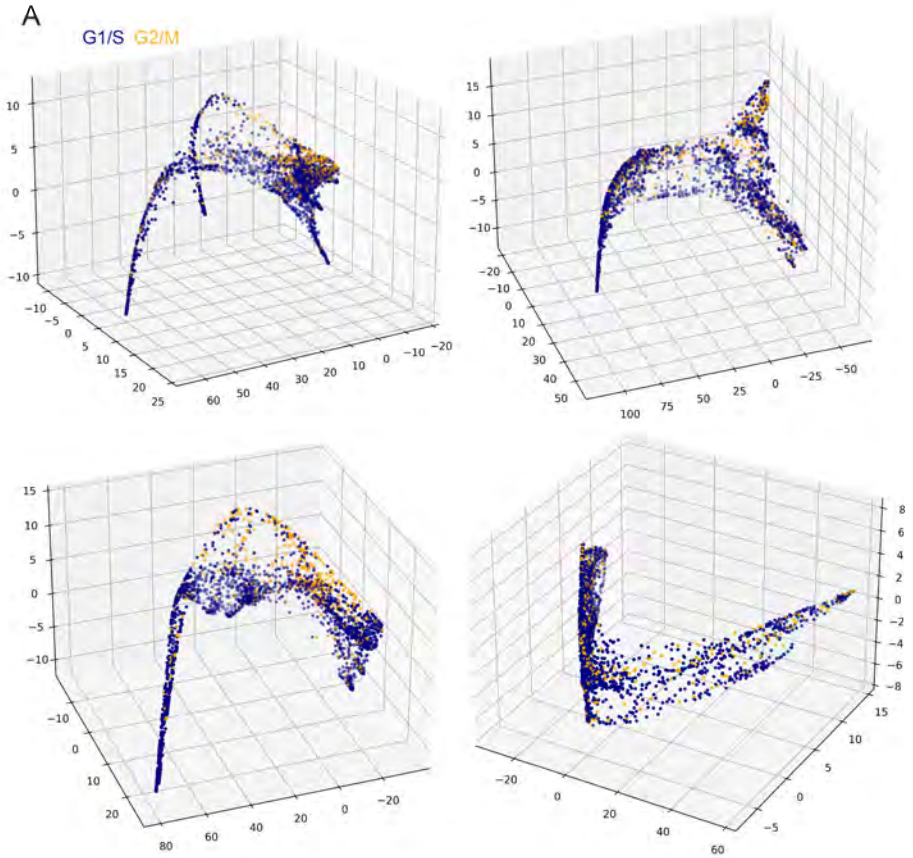

Figure A9: **(A)** PHATE visualization of cell clustering of Ca^{2+} signaling patterns, where each dot represents a single cell; its position in space represents how similar its Ca^{2+} signaling is to other cells in space; each cell or node is colored by its cell cycle state based on nuclear Fucci2 signal.

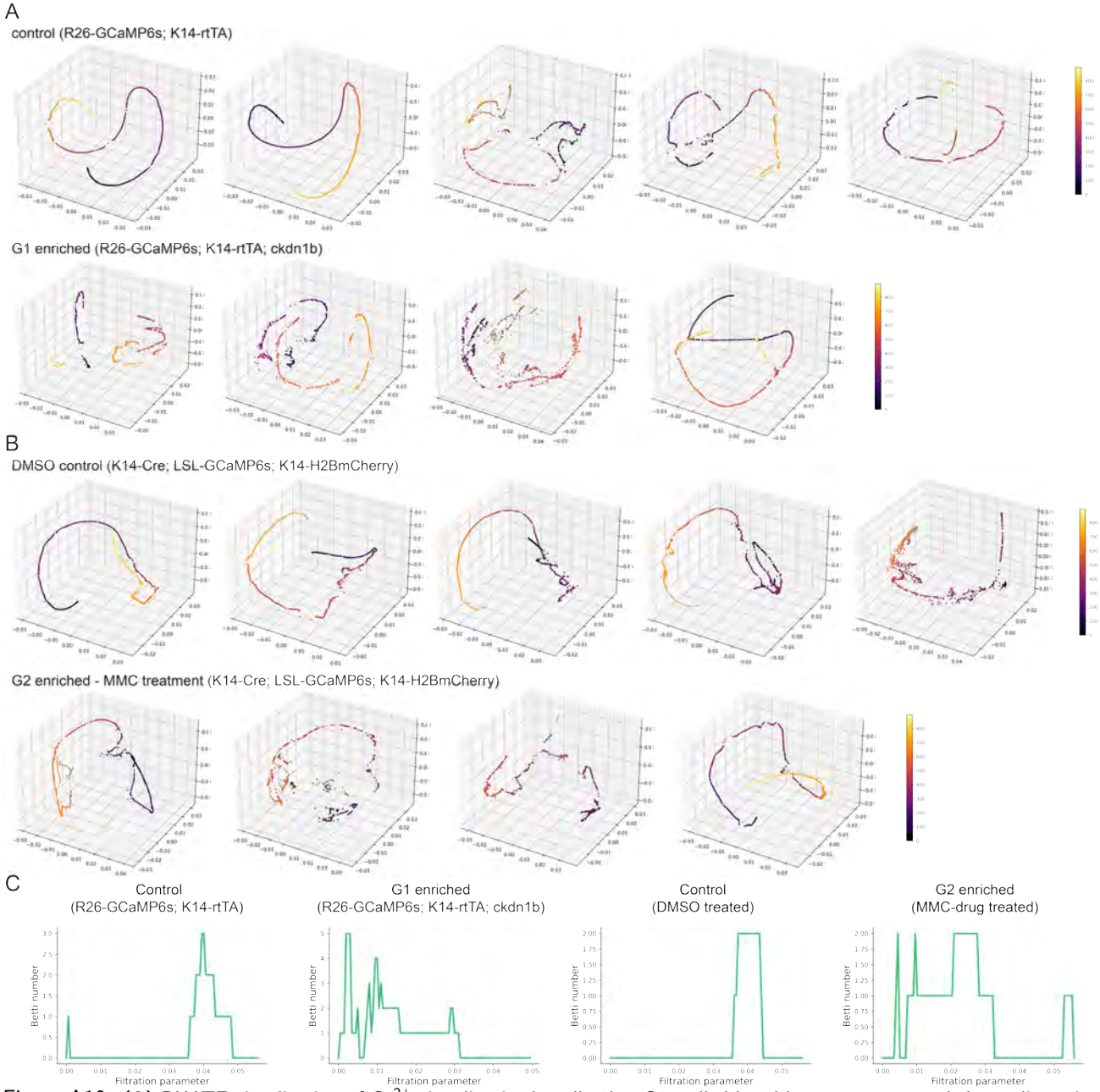

Figure A10: **(A)** PHATE visualization of Ca^{2+} signaling in the cdkn1b+ G1-stalled basal layer versus control shows disruption of smooth, directed and coordinated patterns of signaling. **(B)** PHATE visualization of Ca^{2+} signaling in the MMC-treated G2-stalled basal layer versus control shows smooth, directed and coordinated patterns of signaling. **(C)** Representative Betti curves of $H1$ features for G1 and G2 enriched conditions (R26-GCaMP6s; K14-rtTA; cdkn1b 3 days after doxycycline treatment and MMC drug 2 days after treatment respectively)