

EXPERIMENTAL AND ANALYTICAL
TECHNIQUES FOR STUDYING
MECHANOTRANSDUCTION IN ARTICULAR
CARTILAGE

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Jingyang Zheng

August 2023

© 2023 Jingyang Zheng
ALL RIGHTS RESERVED

EXPERIMENTAL AND ANALYTICAL TECHNIQUES FOR STUDYING
MECHANOTRANSDUCTION IN ARTICULAR CARTILAGE

Jingyang Zheng, Ph.D.

Cornell University 2023

Tissues are often complex heterogeneous systems, where individual cells coordinate responses to internal and external stimuli. To fully understand cell function in such environments, it is critical to connect the behavior of individual cells with the full tissue-scale response. Articular cartilage is one such system where the complex extracellular matrix and heterogeneous cell responses make it difficult to understand how chondrocytes respond to injury-inducing strain. In this thesis I will explore the methods I developed to map the spatiotemporal behaviors of chondrocytes in articular cartilage after impact. These methods combine microscopy techniques with large scale data analysis, making use of a custom-programmed decision tree algorithm, supervised time series classifiers, and unsupervised clustering via a variational autoencoder to identify and categorize cell phenotypes. Time series data collected from thousands of chondrocytes *in situ* during and after impact allow me to probe responses through the lenses of calcium signaling, mitochondrial polarity, and nuclear membrane permeability.

This thesis serves to outline the experimental and analytical methods developed to probe cellular response to external stimuli. While the focus is on method development, I will touch on some of the biological implications of impact on articular cartilage, and identify questions and hypotheses generated through the experiments conducted for this work.

BIOGRAPHICAL SKETCH

Jingyang Zheng was born on April 22nd, 1995 in Beijing, China, and spent her first few years living on the beautiful Tsinghua University campus. She moved to Minnesota at age five, attending Brimhall Elementary School, Roseville Area Middle School, and Roseville Area High School.

Jingyang completed her undergraduate studies at the University of Minnesota - Twin Cities, where she completed a Bachelor of Science in Physics *summa cum laude* and a Bachelor of Science in Chemistry, as well as a minor in Astrophysics. During her time at the U of M, she worked with Prof. Jim Kakalios to complete her Honors Thesis on electronic transport in silicon nanocrystals. She also worked with Prof. William Irvine at the University of Chicago for a summer project studying interactions in crack propagation on Gaussian surfaces. To complete the trifecta of studying basic sciences, she joined Prof. Itai Cohen's group at Cornell University to study mechanotransduction in articular cartilage for her PhD.

Jingyang also enjoys reading, baking, cooking, hiking, and video games. She finds particular joy in tempting Itai with baked goods.

To Mama, Laolao, and Laoye, for their love and support,
To all of my teachers, who encouraged me through the years,
and
To Eric, my light.

ACKNOWLEDGEMENTS

My time here at Cornell working on this PhD has been incredibly rewarding and a fun learning experience. So many people have helped me along the way and truly enriched my life, and I would like to thank them here. First, I would like to thank my advisor Itai Cohen for being an amazing mentor. His unwavering confidence and enthusiasm in both my research and my abilities as a scientist have helped me grow immensely. He has taught me so much about how to effectively communicate my research and how to approach problems. My time in his lab has shaped my growth as a scientist and person, and I truly appreciate his support throughout my time here. I would also like to thank Larry Bonassar for welcoming me into his lab and teaching me so much about biomedical engineering and cartilage. His perspectives on my work and science in general were an invaluable resource. In addition, I would like to thank Michelle Delco for teaching me so much about biology and working with cartilage. Her excitement about my progress was truly encouraging and motivational. Finally, I would like to thank Chris Myers and Lisa Fortier, who were incredibly helpful in providing guidance and expertise along the way.

I have been lucky to work with so many wonderful members of the Cohen lab, who made this PhD an unforgettable experience. Lena Bartell introduced me to cartilage and provided so much mentorship and guidance as I tried to follow in her footsteps. My time on the cartilage project would not have been the same without the friendship and support of Thomas Wyse Jackson, who made lab work great fun and was always willing to talk about anything—it was wonderful to grow as scientists and individuals together. Sam Whitehead, Yunus Kinkhabwala, and Han Kheng Teoh helped me learn so much about anything and everything to do with coding or making plots, and were always so incred-

ibly helpful. Meera Ramaswamy set the best example of who I aspire to be and let me constantly distract her with questions about science and cat videos. The office would not have been as fun without Edward Ong, who always made time for insightful conversations about science and life. Baris Bircan was always fun to talk to and was an amazing listener. Finally, I really appreciate Japheth Omonira for allowing me to practice my mentoring skills and for taking on this difficult project.

To my collaborators in the Bonassar and Delco labs, who guided me as a physicist joining in the world of biomedical engineering, thank you so much. I could not have done any of my experiments without the help of Becka Irwin and Liz Feeney teaching me so many lab skills. And thanks to Sean Kim, Sera Lopez, Leigh Slyker, Steven Ayala, Caroline Thompson, Alicia Matavosian, Marianne Lintz, and Matthew Thomas for all of their assistance.

So many people among the Physics and LASSP staff have helped me throughout this process. Many thanks to Kacey Bray Acquilano and Sara Edelman for answering my endless questions, Robert Sprankle for helping me with every single order I needed to make since I was locked out of the eShop, Nate Ellis for teaching and helping me with everything in the machine shop, and Craig Wiggers, Barry Robinson, and Lea Reagan.

I could not have made it this far without the help of my mentors and undergraduate advisors who believed in me when I wanted to switch to a physics major in my junior year—Jeremy Mans for helping me sort through the classes, Paul Crowell for his incredible advice on both my first experimental physics project and in everything I needed to apply to graduate school, and Jim Kakalios for his patience and generosity in guiding me through my Honors Thesis and teaching me how to communicate about my work. I would also like to thank

Mike Gorman for his amazing advice and being wonderful person to work for in my internship at 3M.

I am truly grateful for all of my friends, especially Ariana Ray and Divya Gadkari for their constant support and always being available to chat and listen, and Melissa Bosch, Colin Bundschu, Michelle Wang, and Kitty for making my time here amazingly fun. I feel so gratified to have friends back home in Minnesota who made time for our friendship despite the distance—Jan Ohm, Rena Wang, Travis Tran, and Raymond Zhang. I would like to thank Kyle Evans and Grant Kvendru for always being fun to play video games with, and the best DnD group, Michael Roberts, Chris Lawler, Nathan Ng, Daniel Sun, and David Steinberg.

My family: Mama, Laolao, and Laoye, have loved and supported me from the start. I could not have made it this far without their guidance and encouragement. Many thanks and treats for Mibao, the cutest and softest little kitty, who has never failed to make me smile. And I cannot thank Eric enough for his constant love and support, for bringing so much joy and laughter into my life, and for making each day brighter.

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Acknowledgements	v
Table of Contents	viii
List of Tables	x
List of Figures	xi
1 Introduction	1
1.1 Introduction to Articular Cartilage	2
2 STRAINS: A big data method for classifying cellular response to stimuli at the tissue scale	6
2.1 Abstract	6
2.2 Introduction	7
2.3 Results	10
2.4 Discussion	24
2.5 Materials and Methods	26
2.6 Supplementary materials	37
2.6.1 Strain-dependent cellular response	37
2.6.2 Decision Tree Methodology	40
2.6.3 Tracking and Intensity Analysis Code	40
2.6.4 MATLAB GUI	44
2.6.5 Time Series Classification Code	45
2.6.6 Example Application of STRAINS to Worm Neuron Data	45
3 Variational Autoencoder for Clustering Cartilage Data	48
3.1 Abstract	48
3.2 Introduction	49
3.3 Results	51
3.3.1 Principal Components of the VAE Generate Valuable Insights	52
3.3.2 Clustering Post-Impact Data into Distinct Behavior Phenotypes	58
3.3.3 Correlating Peracute Calcium Response with VAE Generated Clusters	63
3.4 Discussion	65
3.5 Methods	66
3.5.1 Impacting articular cartilage tissue	67
3.5.2 Extracting time series data of chondrocyte behavior	68
3.5.3 VAE Structure	68
3.5.4 Training details	70
3.5.5 Principal Component Analysis	71
3.5.6 Hierarchical Clustering	71

4	Conclusion	73
4.1	Conclusions	73
4.2	Future Directions	74
4.2.1	Mitoprotective Therapy	74
4.2.2	Changing Impact Magnitude	75
4.2.3	Cell-to-cell Signaling and Gap Junctions	76
4.2.4	Combining STRAINS and VAE with other techniques . . .	76
4.2.5	Using Automated Algorithms to Improve Decision Tree Structure	77
4.2.6	TRPV4 and Piezo 1/2	77
A	Imaging Chondrocytes in Cartilage Impact	79
A.1	Introduction	79
A.2	Experiments	80
A.2.1	Dissection and Culture	80
A.2.2	Staining	81
A.2.3	Using the Impactor	83
A.2.4	Microscopes	84
A.2.5	Strain Mapping with the High Speed Camera	87
A.3	Impactor Calibration and Maintenance	88
B	STRAINS and Other Supporting Code	90
B.1	Code Breakdown	90
B.2	Image Setup	90
B.3	Tracking	91
B.4	STRAINS GUI	93
B.4.1	GUI Operation	93
B.5	Feature Extraction	95
B.6	Decision Tree	96
B.7	Time Series Classification	97
B.8	VAE Clustering	98
	Bibliography	100

LIST OF TABLES

2.1	Category based accuracy of decision tree classification	22
2.2	Time series classification methods with details on classifier types and their corresponding scores	23

LIST OF FIGURES

2.1	<p>Sample processing, imaging, tracking, and intensity analysis procedure a) Biopsy punches of condylar articular cartilage are collected and bisected. b) Hemicylinders are cultured and stained for calcium concentration, mitochondrial polarization, and nuclear membrane permeability (stand in for cell death). c) Samples are mounted side-by-side to the back plate of impactor. One half of the sample impacted once, with the other serving as an unimpacted control. d) Imaging occurs during impact at site 1 and after impact at sites 1-5. e) Individual cells are tracked through all time points and their stain intensities extracted. f) Temporal patterns of strain intensities are classified by STRAINS into distinct categories of cell behaviors, which are identified and mapped onto tissue location.</p>	11
2.2	<p>Timelapse of cartilage response Images are shown at the impact site (region 1), with the impact location labeled with an orange arrow. a) High speed imaging of calcium concentration in the minute after impact showed a wave of calcium uptake in cells moving outward from the immediate impact site. Inset shows the area experiencing mostly shear strain, with a more muted and delayed response in comparison with areas experiencing compressive strain. (Caption continued on next page)</p>	14
2.2	<p>(cont.) b) Long term imaging of all three stains (mitochondrial polarization, calcium concentration, nuclear membrane permeability). As time progressed, mitochondrial polarization and calcium concentration slowly decreased while nuclear membrane permeability progressively increased with greatest intensity within 400 μm of the impact site.</p>	15
2.3	<p>A MATLAB GUI for image processing Shown on the left is a panel of controls for video selection and scanning, single cell analysis, and multi cell analysis. Shown in the center is the video player which allows for choosing specific frames or time points using a slider. The GUI allows selecting and displaying data for either individual cells or regions within the frame. On the right are the resulting plots showing channel intensities. The plots can also be clicked on to highlight cells of interest within a group (which will draw a circle around the cell in the image). In addition, time points can be selected to generate cross-sectional histograms for each stain. A full video detailing GUI functionality can be found in the SI.</p>	16
2.4	<p>Cell behavior categories and their distributions (Caption continued on next page)</p>	18

2.4	(cont.) For each category, a representative time series for a single cell (top, with blue representing nuclear membrane permeability, green representing calcium concentration, and red representing mitochondrial polarity) along with a heat map of cell counts indicating relative frequency at different locations are shown. Colorbar maxima vary between categories to resolve spatial distributions for categories with fewer cells. Regions 1-3 are on the impacted sample, with 1 as the site of impact. Regions 4-5 are on the control sample. a) The nuclear membrane permeability starts out elevated and decays. b) The nuclear membrane permeability increases within 30 min of impact and plateaus or continues increasing. c) The nuclear membrane permeability increases within 30 min of impact and then decays. d) The nuclear membrane permeability increases after 30 min after impact but no prior signaling event is observed. e) The nuclear membrane permeability increases despite the mitochondria being polarized. f) The nuclear membrane permeability increases after the calcium concentration drops after a period of sustained elevation. g) Transient calcium signaling is observed after the nuclear membrane permeability has already increased. h) The nuclear membrane permeability increases after calcium transient(s). i) Calcium transient(s) are observed with no increase in nuclear membrane permeability. j) No calcium transients or increases in nuclear membrane permeability. k) All three signals are very low. i) The nuclear membrane permeability increases multiple times.	19
2.5	The full decision tree algorithm for our system. Each cell time series is run through the tree separately. Grey boxes indicate final sorted categories.	41
2.6	Example screen capture of modified GUI on worm GCaMP data for selecting single cells within the video.	46
2.7	Example screen capture of modified GUI on worm GCaMP data for selecting groups of cells within the video.	46

3.1	<p>Sample processing, imaging, tracking, and VAE structure. a) Biopsy punches of condylar articular cartilage are collected, bisected, and halved. Hemicylinders are stained for calcium concentration, mitochondrial polarization, and nuclear membrane permeability (cell death) via Calbryte 520 AM, Sytox Blue, and Tetramethylrhodamine, respectively. b) Paired hemicylinders are glued side-by-side to the back plate of a confocal-mounted impactor. One hemicylinder is impacted (site 1) while the other serves as control. Imaging occurs during impact at site 1 and after impact at sites 1-6. c) Individual cells are tracked through all time points. Stain intensities are extracted and manually sorted (for one sample, to provide comparison baseline for VAE analysis). d) Structure of the Variational Autoencoder (VAE). The first portion of the encoder consists of three 1D convolution layers and a reshape operation that is subsequently fed into fully connected layers of a dense neural network which outputs the mean and diagonal covariance vectors of the compressed latent space representation. A sample z is drawn from the d-dimensional Gaussian and sent through the decoder to obtain a reconstructed intensity profile. The decoder has the same architecture as the encoder, with the exception that the 1D convolution layers are replaced with 1D transposed convolution layers. e) Example reconstruction of cell intensity data. Light colored lines represent the input data and dark colored dashed lines represent the reconstruction.</p>	53
3.2	<p>Principal components and associated spatial heatmaps for three impact experiments. Left: example reconstructed plots showing the variation along the first twelve principal components, varying from $\alpha = -3$ to $\alpha = 3$. Right: Spatial maps of all cells and their associated α values for each PC, where α varies from -3 (purple) to 3 (yellow). Impact location shown with orange arrow. The experiment in the rightmost column was subject to larger impact strain than the other two experiments (left and center). . .</p>	55

3.3	Example branching of the hierarchical clustering tree. Left: Overall agglomerative hierarchical clustering, with example cut-off line for 14 clusters shown in black. Each vertical line is a cluster. The x-axis shows cell counts and the y-axis represents distance between the clusters represented by the symmetrized Kullback-Liebnner divergence. Right: Expanded view of gray shaded box. Clusters break down into smaller groups with more detail when the total number of clusters is increased. Bolded lines represent cluster averages, and thin lines represent individual cells within a cluster. Red represents mitochondrial polarity, green calcium concentration, and blue nuclear membrane permeability (cell death). Four levels are shown, showing the specificity of clusters when there are 3, 5, 9, and 14 total clusters. Letters are used to name clusters for ease of identification and have no specific meaning.	57
3.4	All fourteen clusters from one example data set, organized in two columns. Each cluster is shown in three panels: the spatial distribution of cells within the cluster (cells represented as orange dots), the timeseries plots of cells within the cluster (with means as bolded lines), and the UMAP distribution of cells within the cluster (cells represented as orange dots). The clusters do not have a specific ordering.	61
3.5	Impact calcium signatures associated with VAE-produced clusters. Differentiation between healthy cells and damaged or dying cells can be seen from clustering on post-impact long term data. Each cluster has three panels: the timeseries of all cells in the cluster for longer-term imaging (3 hours) which includes all three color channels, the timeseries of the impact imaging for that cluster (1 minute) which includes only calcium concentration, and the spatial distribution of cells in said cluster which are shown as orange dots. In timeseries plots, bolded lines represent means while thin lines represent individual cells within the cluster. The exact impact location is indicated with an orange arrow. Left: clusters showing functional/healthy cells. Right clusters showing damaged/dying cells.	64

CHAPTER 1

INTRODUCTION

When we study things in nature, we often need to delve into studying the coordination within a system. How do the fibers of a cartilage network coordinate to produce its mechanical properties? How do colloidal particles form networks of particle configurations? How do fruit flies coordinate their flight reactions between muscles and neurons? How do cartilage cells coordinate their responses to normal and injurious strain?

To investigate any of these questions, and more specifically the last, we must incorporate knowledge and techniques from many different fields. This thesis lies at the intersection of physics, biology, biomedical engineering, and machine learning, requiring the coordination of all of these disciplines to exist, just like how chondrocytes coordinate in their daily function. And even more, this thesis would not exist without input from all of the different projects in the Cohen lab: the particle tracking from the colloids projects, the coding and computational knowledge from the fruit fly projects, and support of the depth of knowledge developed in the preceding cartilage projects. Combining everything I have learned from other fields and within this lab, I coalesce that knowledge into developing methods for studying coordinated cellular behaviors in articular cartilage.

This thesis presents two methods of investigating mechanotransduction and cell signaling in articular cartilage. Chapter 2 outlines and demonstrates the experimental and analytical techniques for studying cell signaling in articular cartilage. The calcium signaling, mitochondrial polarity, and nuclear membrane permeability of chondrocytes are imaged following an injury-inducing

impact. Cells within these images are tracked over the entire course of the experiment and their individual behaviors extracted as time series. The GUI, feature extraction, decision tree, and time series classification methods are used alongside initial manual classification to establish and categorize chondrocyte behavior phenotypes. Chapter 3 expands upon the analyses conducted in the previous chapter by implementing unsupervised machine learning with a Variational Autoencoder (VAE). The VAE learns the most important features of the chondrocyte timeseries data, which is used to cluster the cells based on their behavior. Details of experimental techniques and analysis codes are documented in the appendices, Chapter A and Chapter B.

By using these methods, we are able to shine a light (laser) on the processes occurring in cartilage tissue both in normal function and after injury.

1.1 Introduction to Articular Cartilage

Osteoarthritis (OA) is a joint disease characterized by pain and disability which affects up to 10% of adults in the United States. OA that develops secondary to injury is referred to as Post-Traumatic OA (PTOA). In both cases, the disease is marked by damaged articular cartilage. Despite its prevalence in the population, there is no effective treatment for OA and the pathogenesis of disease after cartilage injury is not fully elucidated.

Articular cartilage is present in diarthrodial joints and functions as a smooth, low-friction tissue that facilitates load-bearing and lubrication for joint articulation. Subject to constant strain, the mechanical properties of cartilage are important for its normal function. Articular cartilage consists of chondrocytes embed-

ded in a complex pericellular (PCM) and extracellular matrix system (ECM), which gives cartilage its mechanical properties. Mature cartilage lacks blood vessels, so all necessary compounds for sustaining chondrocyte function must diffuse into the tissue. This limits the ability of cartilage to repair after injury, and contributes to tissue degradation after injury [1].

In both normal tissue function and injurious loading, chondrocytes are affected by a myriad of biochemical and mechanical stimuli. Mechanotransduction is the process by which mechanical stimuli are converted into biochemical responses to elicit cellular responses through signaling pathways. These pathways are important mediators for promoting anabolic responses to strain. Physiologic loading causes chondrocytes to synthesize matrix proteins and maintain the tissue, but superphysiologic strains can lead to cell death and matrix degradation [2].

The complex system of cartilage impact produces a wide array of responses, since cells are subject to a range of both compression and shear strains. When cartilage is strained, many coupled external stimuli affect chondrocytes. The extracellular matrix deforms differently based on depth, where the superficial zone absorbs more of the strain [3, 4]. Stress and strain on the network physically deform extracellular matrix and chondrocytes. Fluid flow through the tissue can cause streaming currents and osmotic and hydrostatic pressure changes. Cell potential also changes under mechanical stress. Heterogeneous strain fields within the cell can also cause different mechanotransduction responses, so it is important to both capture and explain the different types of signaling within the impacted tissue.

One of the most important mediators of chondrocyte mechanotransduction

is calcium. Ca^{2+} ions are employed as a 'second messenger' in many different tissue systems. In cartilage, Ca^{2+} is largely present in the ECM, with almost no Ca^{2+} in the cytoplasm and some Ca^{2+} in the endoplasmic reticulum. Calcium signaling in chondrocytes plays an important role in dictating cellular response to strain. There are a wide variety of calcium channels present on the cell membrane of chondrocytes which respond to stimuli ranging from strain to voltage to ligand binding, etc. Specifically, channels such as TRPV4 and Piezo 1/2 have been established as mechanotransducers of physiologic and hyperphysiologic strains. However, the specific details of how these channels act *in situ*, the connections between peracute signaling and longer-term outcomes, and how these signals coordinate responses between cells is not fully known.

Previous studies have demonstrated the local strain thresholds for cell death and that mitochondrial depolarization leads to cell death [5, 6]. Importantly, mitochondria (MT) have been shown to be mediators of early cellular responses to damage. Not only do mitochondria buffer Ca^{2+} ions, but they also play an important role in attenuating oxidative stress and the process of apoptosis. Ca^{2+} is necessary for normal ion transport function, but too much can cause mitochondrial depolarization.

This thesis expands upon these studies by incorporating the ability to measure cell signals and specific behaviors over time in order to expand our ability to probe mechanotransduction in cartilage. The three signals demonstrated in the methods illustrated in the following chapters—calcium concentration, mitochondrial polarity, and nuclear membrane permeability, were selected to capture calcium-based mechanotransduction responses alongside cell health measures. Mitochondrial polarity serves to indicate homeostatic function, while

nuclear membrane permeability is used as a measure of cell death. Notably unlike standard cell death measures, permeability of the nuclear membrane is not binary. Nuances in nuclear membrane permeability illustrate potentially different mechanisms of cell death, and serve to highlight the utility of collecting time series data

The goal of the methods outlined in this thesis is to establish a system where the spatiotemporal behaviors of thousands of chondrocytes within a sample of cartilage are mapped. The data extracted by these methods can be used to explore the relationship between strain and specific signaling pathways, connect between the millisecond calcium response to the hours-long overall cellular behaviors, and obtain a detailed spatial map to probe how signals travel through the tissue. Further, upon establishing these behaviors, our methods allow us to quickly classify and cluster cells in order to effectively parse thousands of cells simultaneously and iterate upon the specific stimuli used to probe cell responses and test hypotheses about mechanotransduction.

CHAPTER 2

STRAINS: A BIG DATA METHOD FOR CLASSIFYING CELLULAR RESPONSE TO STIMULI AT THE TISSUE SCALE

Zheng, J., Wyse Jackson, T., Fortier, L. A., Bonassar, L. J., Delco, M. L., & Cohen, I., *PLOS ONE* 17 no. 12 (2022).

2.1 Abstract

Cellular response to stimulation governs tissue scale processes ranging from growth and development to maintaining tissue health and initiating disease. To determine how cells coordinate their response to such stimuli, it is necessary to simultaneously track and measure the spatiotemporal distribution of their behaviors throughout the tissue. Here, we report on a novel SpatioTemporal Response Analysis *IN Situ* (STRAINS) tool that uses fluorescent micrographs, cell tracking, and machine learning to measure such behavioral distributions. STRAINS is broadly applicable to any tissue where fluorescence can be used to indicate changes in cell behavior. For illustration, we use STRAINS to simultaneously analyze the mechanotransduction response of 5000 chondrocytes—over 20 million data points—in cartilage during the 50 ms to 4 hours after the tissue was subjected to local mechanical injury, known to initiate osteoarthritis. We find that chondrocytes exhibit a range of mechanobiological responses indicating activation of distinct biochemical pathways with clear spatial patterns related to the induced local strains during impact. These results illustrate the power of this approach.

2.2 Introduction

To sustain tissue function, cells must coordinate their response to external stimuli. In mechanically sensitive tissues, for example, cells sense their environment using cytoskeletal elements, ion channels, and other mechanisms to initiate such coordinated behaviors [7, 8]. Changes in fluid pressure in vascular systems affect mechanosensitive ion channels, driving cell migration and muscle development [9]. Altered mechanosensing in cancer cells makes them unable to sense stiffness, potentially playing a role in metastasis, migration, and disease progression [10]. And, in cartilage, tendon, and bone, cell mechanosensing pathways regulate growth and development during normal function or promote disease during aberrant loading [11, 12, 13]. Pioneering studies furthered our understanding of which mechanotransduction pathways are activated in *single* cells in response to various perturbations [14]. Studies using pillar arrays [15, 16, 17], traction force microscopy [18], magnetic tweezers [19], or optical traps [20] for example, have demonstrated the role of substrate rigidity in stem cell differentiation [21], the alignment of cellular microfilaments in the direction of force [22], and the highly varying force profiles of migrating cells [23, 14]. Such cellular responses must be coordinated at the *tissue scale* to sustain mechanical function, direct resources to regions in need of repair, or initiate healing [24, 25, 26]. This coordination, however, remains poorly understood because few techniques are available for imaging, analyzing, and sorting the *in situ* collective response of thousands of cells over thousands of time points throughout the tissue.

Here, we introduce a SpatioTemporal Response Analysis IN Situ (STRAINS) tool that uses new experimental methods and a big data analysis technique to

investigate tissue scale coordination of the cellular responses. STRAINS tracks thousands of cells within tissue during and after an applied stimulation, extracts their individual fluorescence traces, and analyzes their spatiotemporal behavior patterns. This technique makes use of newly developed protocols to stain and image processes such as Ca^{2+} signaling, mitochondrial depolarization, and nuclear membrane permeability *in situ* over sub-second to hour time scales. The advances we report here entail tracking responses in thousands of cells that are moving, visualizing millions of data points with an intuitive graphical user interface (GUI), and using new custom sorting and machine learning algorithms to classify and map a wide range of cellular behaviors throughout the tissue.

We demonstrate the utility of this approach by using STRAINS to investigate the complex relationships between mechanical strain and chondrocyte responses in articular cartilage, identifying distinct patterns of cell behaviors and mapping their spatiotemporal distribution. Macroscale joint injury, specifically rapid cartilage overloading, is known to precipitate osteoarthritis.

For example, previous work has demonstrated that articular impact injury triggers tissue scale catabolic responses *in situ* and *in vivo* [5, 27, 6, 28, 29]. During impact, chondrocytes use mechanosensors like integrins [30, 31], the primary cilium [32, 33, 34, 35], and various mechanosensitive ion channels [36, 37, 38, 15, 39] to convert mechanical signals into biochemical responses ranging from the synthesis of extracellular matrix proteins for maintaining tissue integrity to apoptosis and matrix degradation.

Signaling within and between cells in cartilage post impact occurs on multiple timescales. Within seconds, activation of mechanosensors on the cellular membrane enables calcium and other force-sensitive signaling [40, 36]. In the

ensuing hours, the initial cellular response affects mitochondrial polarization [6], cell viability [27], and subsequent signaling cascade, leading to distinct outcomes based on the initial local strain experienced by the cell. By measuring both short (sub-second to second) and long (minutes to hours) term signaling, STRAINS enabled us to make connections between signaling events and paint a fuller picture of the signaling landscape after injury.

Additionally, the complex structure of articular cartilage matrix causes local strain within the tissue to vary with location and depth, which in turn can lead to different cellular behaviors. For example, cells directly below the impact site primarily experience compression. In contrast, cells to the sides of the impact site experience greater shear stresses. In single cells, these distinct mechanical deformation modes are known to trigger different responses in the chondrocytes [41]. Whether cells maintain these individual behaviors based solely on the local deformations they experience or coordinate their response more globally is poorly understood.

Collectively, cartilage's depth-dependent spatial heterogeneity, the complex load distribution within the tissue during impact, and the broad range of timescales for chondrocyte responses makes it an ideal tissue for showcasing the power of this method. Importantly, however, STRAINS can be applied to any system where it is relevant to study the collective spatiotemporal response of large numbers of cells to external stimuli.

2.3 Results

Experimental system and *in situ* imaging procedure

We have developed a microscale impacting system to assess the real-time multichannel cellular response to mechanical stimulus. A custom-built confocal-mounted impactor was used to injure and image fluorescently stained bovine cartilage samples [5]. Specifically, 6 mm plugs were sterilely extracted from the condyles of neonatal bovids (Fig. 3.1a). Each plug was bisected into two hemicylinders, cultured for stabilization [28] and stained for simultaneous measurement of Ca^{2+} concentration (Calbryte 520 AM), mitochondrial depolarization (tetramethylrhodamine, TMRM), and nuclear membrane permeability (Sytox Blue) (Fig. 3.1b, and Methods). To image the tissue response, two hemicylinders from the same plug were glued onto the fixed backplate of the confocal-mounted impactor (Fig. 3.1c). One sample was used as a control while the second sample was impacted to induce injury. The impactor was calibrated to deliver to the articular surface a 5-10 ms impact with a peak stress of ~ 1 MPa, which produced superphysiologic strains and strain rates encompassing the wide range of strains observed in joints with traumatic injury within a small field of view [27, 6]. Strain calibration was conducted with a high speed camera (v7.1, Vision Research), as described in Bartell et. al., 2015 [5] and Henak et. al., 2017 [42]. To create visual texture for measuring strain fields, the cut surface of the impacted sample was coated with fluorescent polystyrene microspheres (2 μm carboxylate particles). During impact, cartilage deformation was recorded using the high speed camera at 1000 frames per second. Digital image correlation was used to track the deformation and calculate strain fields.

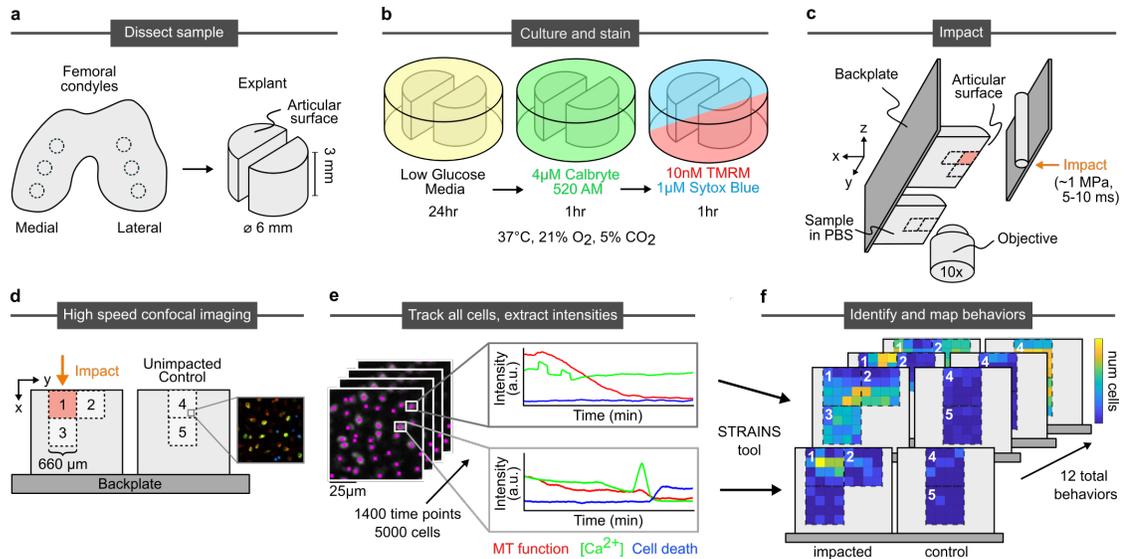


Figure 2.1: Sample processing, imaging, tracking, and intensity analysis procedure a) Biopsy punches of condylar articular cartilage are collected and bisected. b) Hemicylinders are cultured and stained for calcium concentration, mitochondrial polarization, and nuclear membrane permeability (stand in for cell death). c) Samples are mounted side-by-side to the back plate of impactor. One half of the sample impacted once, with the other serving as an unimpacted control. d) Imaging occurs during impact at site 1 and after impact at sites 1-5. e) Individual cells are tracked through all time points and their stain intensities extracted. f) Temporal patterns of strain intensities are classified by STRAINS into distinct categories of cell behaviors, which are identified and mapped onto tissue location.

This entire apparatus was loaded onto a fast confocal microscope that enabled imaging of the cells throughout the impact and relaxation process. Each region corresponding to the field of view for our 10X objective was $660\ \mu\text{m} \times 660\ \mu\text{m}$ in size. We assessed multiple regions to understand the influence of a wide range of local tissue strains on the behaviors of cells (Fig. 3.1d). On the impacted sample, we imaged the entirety of the impact site (region 1) at the surface of the tissue from $60\ \mu\text{m}$ above the articular surface to $600\ \mu\text{m}$ below the surface (superficial zone of articular cartilage), lateral to and at the same depth as the impact site (region 2), also at the surface of the tissue, and directly below the impact site (region 3), in the region extending $600\ \mu\text{m}$ to $1260\ \mu\text{m}$ below

the surface (middle zone of articular cartilage). On the unimpacted control, we imaged two sites (regions 4 and 5) at the same tissue depths (superficial and middle zones) as the impacted sample in order to compare cells of similar phenotype. From this imaging process, we obtained the fluorescent intensities of each cell (expanded image in Fig. 3.1d). For region 1 we imaged the Ca^{2+} response at 40 frames per second over 1.5 min during and immediately after the impact. Subsequently, we imaged all 5 regions and all three color channels every 10 seconds over a 4 hour period. Collectively, we obtained the time-dependent fluorescent response of each channel for ~ 5000 cells corresponding to ~ 20 million data points (Fig. 3.1e). Each cell exhibited a pattern of intensity responses with time for the three fluorescent channels. Once classified, the location and frequency of these distinct temporal response patterns within the tissue were mapped (Fig. 3.1f).

Strain-dependent cellular response

The strain field resulting from impact and the associated cell response had complex behaviors that varied spatiotemporally. Specifically, we observed complex patterns in the immediate post-impact Ca^{2+} response and hours-long cellular responses for all three measured signals. In the milliseconds after impact trauma, increased calcium concentration can be observed in cells proximal to the impact site, in the superficial region. However, on the seconds timescale, we observed differences in total intensity between chondrocytes experiencing shear and compression within this region (Green in Fig. 2.2a). On longer time scales, we found that mitochondrial polarity rapidly diminished at the impact site in the minutes after injury, with calcium concentration following the same pattern but with

some cells exhibiting transients on the scale of minutes (SI Video). Conversely, nuclear membrane permeability initially showed a very low intensity throughout the region and reached higher intensities in a fraction of the cells in regions extending up to 400 μm below the impact site on a time scale of hours. Consistent with the short time calcium response, this pattern of cell death did not extend to areas of the tissue which experienced primarily shear strains.

Collectively, these distinct spatiotemporal patterns of cell response indicated that multiple mechanobiological pathways may have been activated in response to local strain. Developing an understanding of how such processes are related requires identifying distinct cellular signatures and mapping out where in the tissue they are localized. To obtain these maps, however, we must first identify each cell, track its movement and multi-channel fluorescence response over time (Fig. 3.1e), and classify its cellular signature (Fig. 3.1f).

Enhanced particle tracking captures behaviors of moving cells

In order to measure tissue-level cellular behaviors, we must first track each cell individually, over time and through movement, to obtain the fluorescence intensities of each channel. To track the cells, we summed the intensities from all three fluorescent channels at each time step such that the composite image showed bright isolated regions corresponding to the cells. We then applied a modified version of Crocker and Grier's particle tracking algorithm to filter the image and obtain trajectories for each cell centroid (Fig. 3.1e) [43]. Using this method, we obtained the fluorescent time traces of over 5000 cells for the five imaged regions yielding over 20 million measurements.

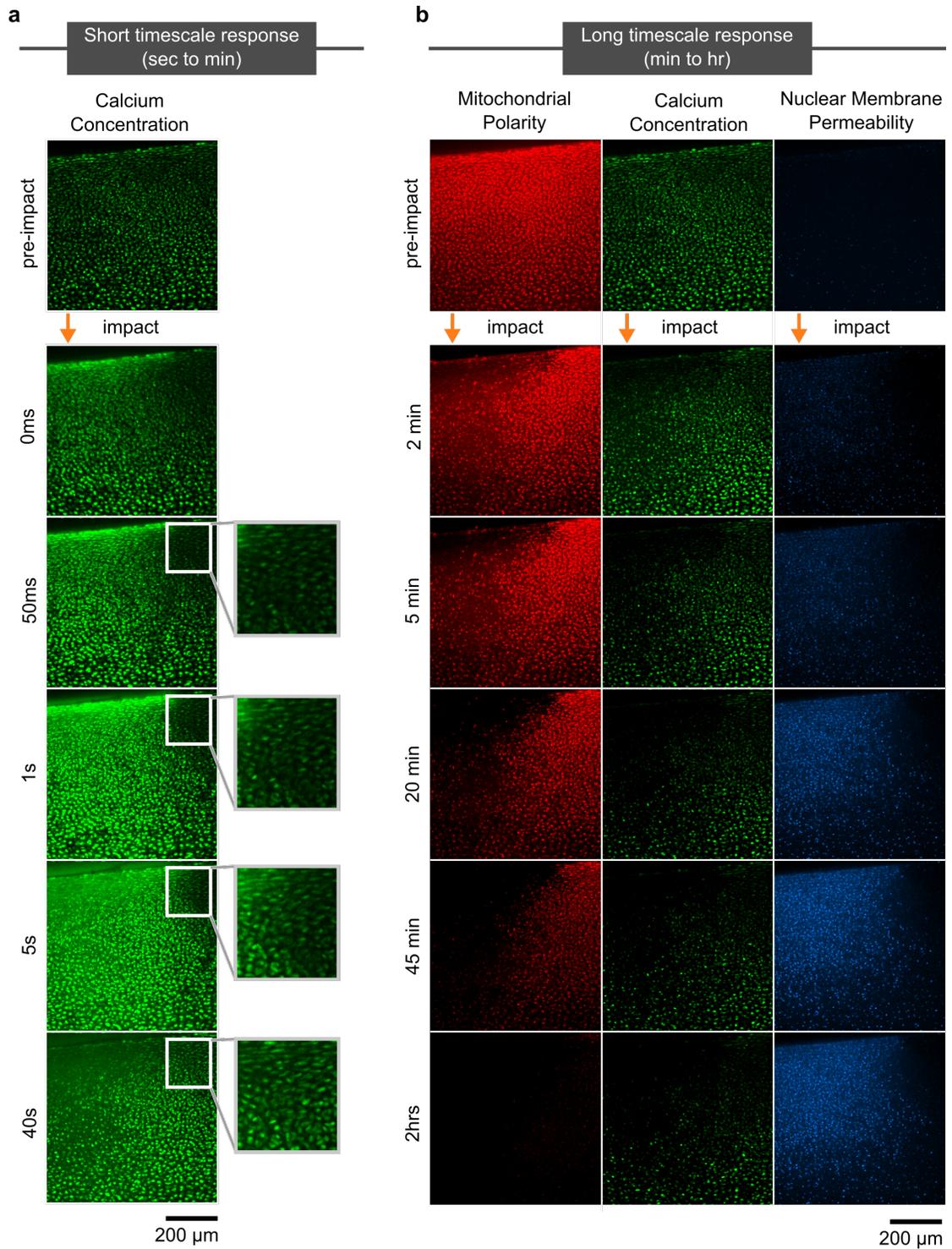


Figure 2.2: Timelapse of cartilage response Images are shown at the impact site (region 1), with the impact location labeled with an orange arrow. a) High speed imaging of calcium concentration in the minute after impact showed a wave of calcium uptake in cells moving outward from the immediate impact site. Inset shows the area experiencing mostly shear strain, with a more muted and delayed response in comparison with areas experiencing compressive strain. (Caption continued on next page)

Figure 2.2: (cont.) b) Long term imaging of all three stains (mitochondrial polarization, calcium concentration, nuclear membrane permeability). As time progressed, mitochondrial polarization and calcium concentration slowly decreased while nuclear membrane permeability progressively increased with greatest intensity within 400 μm of the impact site.

Importantly, the sheer scale of data made it prohibitively difficult to use typical statistical analyses or cell tracking tools, which are limited in tracking capabilities and unable to pick out specific time series characteristics and interactions between multiple different fluorescent channels tracking different components of cellular function. In order to interpret our data, we needed to categorize cellular signatures, relate each cells' response to its location within the tissue, and determine whether the cellular signals from multiple cells were spatiotemporally clustered.

A MATLAB graphical user interface enables identification of cell behaviors

To address this challenge, we built a graphical user interface (GUI) in MATLAB to allow researchers to directly make comparisons between images and time series signatures (Fig. 2.3). A built-in video player allows the user to scan through and select image frames for analysis. Individual cells within any frame can be selected by clicking on the image, entering the cell ID (the number assigned by the particle tracking algorithm), or providing its x-y pixel coordinate (the program will find the nearest cell). The program then plots the three color intensity versus time curves for that cell on the right side of the GUI. To enable the analysis of multiple cells, our GUI allows a user to select a rectangular region

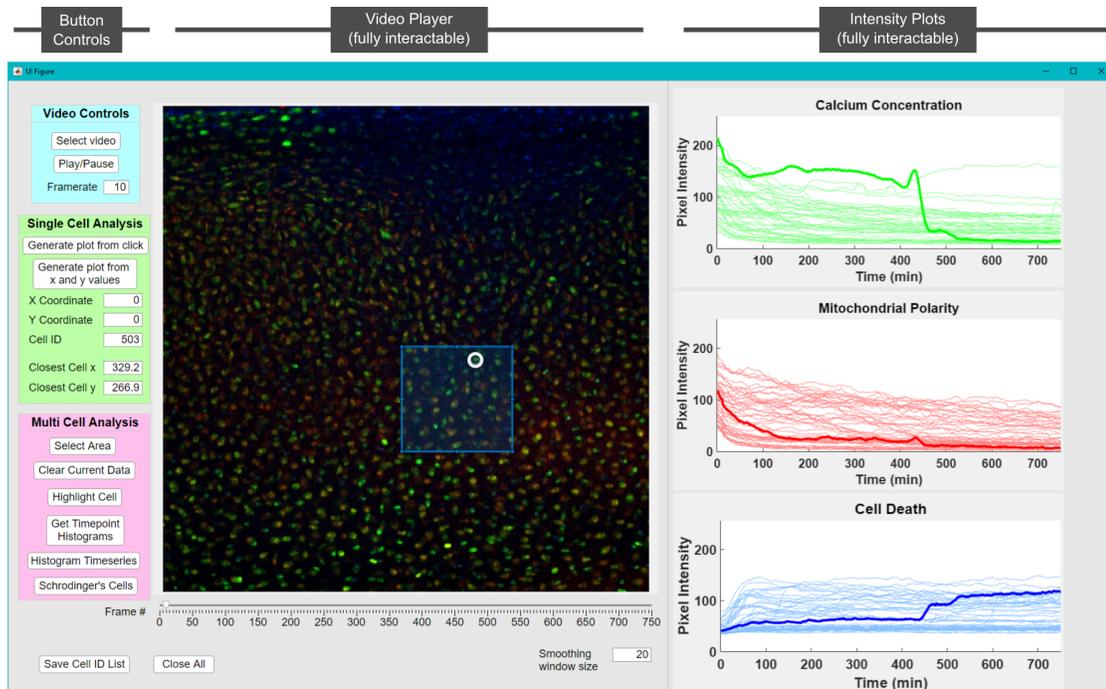


Figure 2.3: A MATLAB GUI for image processing. Shown on the left is a panel of controls for video selection and scanning, single cell analysis, and multi cell analysis. Shown in the center is the video player which allows for choosing specific frames or time points using a slider. The GUI allows selecting and displaying data for either individual cells or regions within the frame. On the right are the resulting plots showing channel intensities. The plots can also be clicked on to highlight cells of interest within a group (which will draw a circle around the cell in the image). In addition, time points can be selected to generate cross-sectional histograms for each stain. A full video detailing GUI functionality can be found in the SI.

within any frame, and the program will plot the fluorescence curves for all the cells within the region. Directly clicking on a time series in one of the plots will bold the selected line in each color and circle the selected cell within the image (Fig. 2.3). Finally, the cell ID of all observed cells can be saved to a text file before exiting the program for record keeping. This GUI allowed for targeted investigations of cellular behaviors in different regions of the tissue.

Using this GUI, we found specific repeating patterns in the intensity curves related to observed peaks (transients), intensity jumps, decay time scales,

plateaus, and the temporal locations of such features. For example, cell death was identified as a sudden increase in nuclear membrane permeability (see for example bold blue channel in Fig. 2.3). This behavior often followed a peak in the calcium concentration (green channel in Fig. 2.3). We also used the GUI to distinguish between seemingly similar curves. While numerous cells showed a rapid increase in nuclear membrane permeability within half an hour of impact, in some cells this signal plateaued and remained high, while in others it slowly decayed to a lower plateau. These differences in time series shape are subtle but distinct. Collectively, this GUI and the analysis features it enabled provided a pathway for sorting the millions of data points in an intuitive fashion, enabling the user to quickly identify categories of cell behaviors and develop an intuition for where each behavior tends to localize.

Making use of the GUI analysis features, we identified twelve distinct behaviors across all cells within an impacted tissue and mapped their location relative to the impact site (Fig. 2.4). We observed eight different behaviors where chondrocytes showed a high level or a rapid increase in their nuclear membrane permeability, likely related to cell death (Fig. 2.4a-g, l). Additionally, two behaviors were related to distinct calcium transients (Fig. 2.4h, i). Finally, we identified two behaviors where cells maintained low nuclear membrane permeability throughout the experiment (Fig. 2.4j, k).

Importantly, the precision with which we measure the cell intensities and correlations in cell signatures allowed for distinguishing between these behaviors, even when the accompanying curves were quite similar. For example, in some cells the calcium transients were not associated with an increase in nuclear membrane permeability (Fig. 2.4i) while in others, the calcium transient oc-

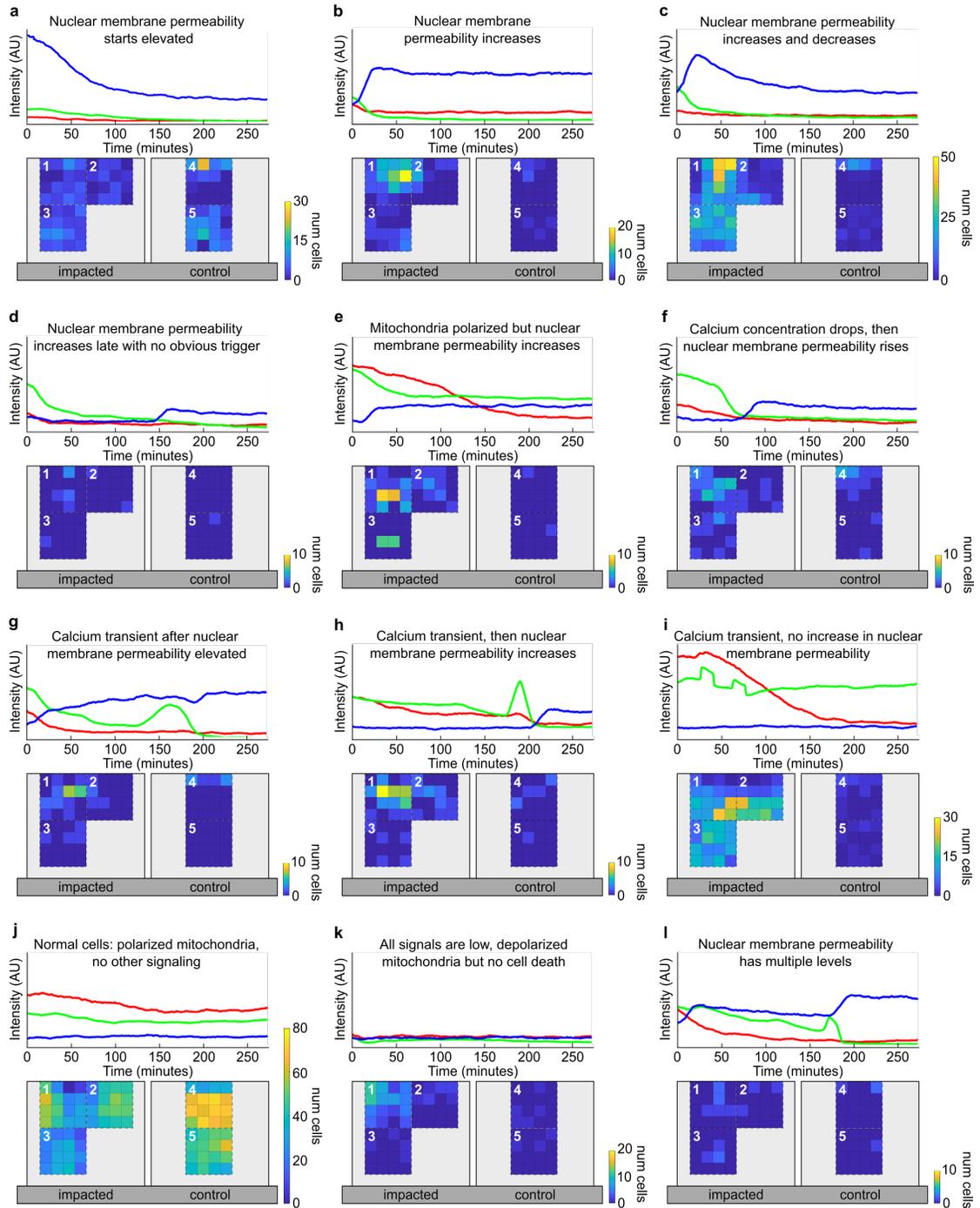


Figure 2.4: Cell behavior categories and their distributions (Caption continued on next page)

Figure 2.4: (cont.) For each category, a representative time series for a single cell (top, with blue representing nuclear membrane permeability, green representing calcium concentration, and red representing mitochondrial polarity) along with a heat map of cell counts indicating relative frequency at different locations are shown. Colorbar maxima vary between categories to resolve spatial distributions for categories with fewer cells. Regions 1-3 are on the impacted sample, with 1 as the site of impact. Regions 4-5 are on the control sample. a) The nuclear membrane permeability starts out elevated and decays. b) The nuclear membrane permeability increases within 30 min of impact and plateaus or continues increasing. c) The nuclear membrane permeability increases within 30 min of impact and then decays. d) The nuclear membrane permeability increases after 30 min after impact but no prior signaling event is observed. e) The nuclear membrane permeability increases despite the mitochondria being polarized. f) The nuclear membrane permeability increases after the calcium concentration drops after a period of sustained elevation. g) Transient calcium signaling is observed after the nuclear membrane permeability has already increased. h) The nuclear membrane permeability increases after calcium transient(s). i) Calcium transient(s) are observed with no increase in nuclear membrane permeability. j) No calcium transients or increases in nuclear membrane permeability. k) All three signals are very low. i) The nuclear membrane permeability increases multiple times.

curred immediately before a rise in nuclear membrane permeability (Fig. 2.4h) or at some time after the nuclear membrane permeability rose (Fig. 2.4g). The differences in how these two channels interacted, along with the distinct spatial distributions associated with each behavior, indicated that separate biochemical processes associated with activation of calcium channels may have taken place, illustrating the power of our approach for generating a comprehensive map of the tissue scale multi-channel cellular response to an applied perturbation. By quickly identifying and spatiotemporally mapping behaviors ranging from normal to abnormal, our approach allows us to focus on interesting and puzzling behaviors for future experimentation.

Implementing automated sorting algorithms to identify cell behavior and category

While this analysis framework is clearly very useful for identifying distinct categories of behaviors, it required extensive manual sorting, which is cumbersome for the scale of data acquired from our technique. This burden, however, was lightened by implementing an augmented strategy which combined the manually sorted categories with custom feature extraction algorithms and supervised machine learning to quickly classify thousands of cellular responses in subsequent experiments. While such algorithms are invariably somewhat system specific, it is nevertheless instructive to illustrate their implementation in our system.

Programmed feature extraction of time series characteristics

Our first approach to implement cellular behavior classification employed a custom feature extraction and decision tree algorithm. In our system, sudden signaling events like calcium transients (peaks) or sharp changes to cell nuclear membrane permeability (change points) played an important role in dictating which category a cell belongs to. These events, however, can occur at random times during the imaging process, making it difficult to search for specific features using automated machine learning methods like clustering or classification. Here, we made use of peak and change point detection algorithms in MATLAB alongside extraction of basic time series statistics (minima, maxima, mean, variance, range, etc.) to find the identifying features of each behavior category. The relationship between these specific features and other time series character-

istics could then be used to create a “fingerprint” for each behavior, which could be searched for en masse.

Once these features were identified, the fluorescence response for each cell was categorized using a decision tree. The tree started by establishing if a cell has a certain feature, such as a step in the nuclear membrane permeability channel or a calcium transient and then branches to more specific criteria based on the relationships between them. For example, if a cell had a changepoint in nuclear membrane permeability, then the tree moved to more specific criteria such as whether the fluorescence in this channel rose and fell or remained high.

In order to distinguish between these possibilities, we searched for peaks in the blue nuclear membrane permeability signal. If the nuclear membrane permeability showed a peak, then the cell was classified as shown in Fig. 2.4c. A similar process was applied to classify all twelve manually identified behaviors (See Methods).

Using this decision tree, we found that the best classified categories reached ~85% accuracy, defined as the true positives divided by the total number of cells manually identified for that category (Table 2.1). Using this decision tree we were able to correctly classify the vast majority of cells (4331 out of 5347) with a total accuracy of 81%. Most importantly, while such decision trees must be determined individually for each system, once established, they can be used repeatedly and with high fidelity.

Table 2.1: **Category based accuracy of decision tree classification**

Category	Accuracy	Total Cells
Nuclear membrane permeability starts elevated (Fig. 2.4a)	0.85	313
Nuclear membrane permeability increases (Fig. 2.4b)	0.56	151
Nuclear membrane permeability increases and decreases (Fig. 2.4c)	0.84	748
Nuclear membrane permeability increases late (Fig. 2.4d)	0.18	11
Mitochondria polarized but nuclear membrane permeability increases (Fig. 2.4e)	0.06	63
Calcium concentration drops, then nuclear membrane permeability increases (Fig. 2.4f)	0.62	60
Nuclear membrane permeability elevated, but calcium transient occurs afterwards (Fig. 2.4g)	0.44	41
Calcium transient, then nuclear membrane permeability increases (Fig. 2.4h)	0.73	73
Calcium transient, then no changes to nuclear membrane permeability (Fig. 2.4i)	0.85	524
Normal cells, polarized mitochondria (Fig. 2.4j)	0.84	3244
All signals low, depolarized mitochondria (Fig. 2.4k)	0.34	101
Nuclear membrane permeability has multiple levels (Fig. 2.4l)	0	18

Time series classification of chondrocyte signaling using sktime library

A second strategy we used to identify cell signatures was time series classification, a type of supervised machine learning where a model system learns to assign labels to objects based on training examples. In our system, relevant information for our data was embedded as a multi-channel time series, so we made use of multivariate time series classifiers from the Python sktime library [44, 45]. Here, we randomly split the data into training and testing sets and opti-

mized the classifiers over a range of parameters. For example, for the Canonical Interval Forest (CIF) classifier [46], we found that the accuracy plateaued once we used more than 1000 estimators. Similarly, for the RandOm Convolutional KErnel Transform (ROCKET) classifier we found that the accuracy plateaued when we used more than 100000 kernels [47]. For detailed list of parameters, see Methods. Finally, we determined what percentage of cells were labeled correctly by the classifier (Table 2.2).

We found that classifiers worked better at detecting certain categories. Our system was an imbalanced multi-label classification problem where there were multiple categories of behaviors with unevenly distributed numbers of objects between categories. Since classifiers are much better trained on categories with more cells, the heavily populated behavior categories were identified with much higher accuracy. Also, when cells had specific features that were not temporally consistent, (e.g. calcium transients or changes in nuclear membrane permeability that occurred in varying numbers, with different amplitudes, and at varying time points) classifiers struggled because no two signals were exactly the same. Consequently, we found that higher scores could be achieved when we split the cells with calcium transients from the data set and classified them separately. When combined with the decision tree algorithm described above, these supervised time series classifiers achieved accuracy values of up to 89%.

Table 2.2: Time series classification methods with details on classifier types and their corresponding scores

Classifier	Accuracy
Canonical Interval Forest (CIF) [46]	0.889
Diverse Representation Canonical Interval Forest (DrCIF) [46]	0.887
RandOm Convolutional KErnel Transform (ROCKET) [47]	0.852
Arsenal Ensemble [46]	0.854

These results speak to the vast potential for automated sorting in future studies [48]. In particular, since the algorithms used here were only recently developed, it is likely that as new more powerful classifiers become available such supervised machine learning approaches will produce greater sorting accuracy. One could even imagine future implementations where unsupervised machine learning is used to extract the most impactful features of the data and cluster cell behaviors with minimal human effort. More broadly, these results demonstrated that a strategy combining automated feature extraction with various machine learning techniques could effectively sort complex cellular data in a streamlined and automated process.

2.4 Discussion

Using STRAINS to monitor cellular behaviors *in situ*, categorize them, and determine where in the tissue they occur enabled novel observations about mechanotransduction in articular cartilage that could not have been obtained from single cell experiments. For example, high levels of compressive strain and transiently high hydrostatic pressure are known to dominate close to the impact site [49] and previous work in our group has demonstrated that microscale local strain exceeding 8% causes cell death [27]. However, by using STRAINS to continuously collect cell behavior data after impact, we were able to observe nine categories of behavior associated with increased nuclear membrane permeability, which suggested different pathways to cell death (Fig. 2.4a-h,l). Here, the nuclear membrane permeability increased and decreased (Fig. 2.4c), or increased and plateaued (Fig. 2.4c), indicative of cell death due to superphysiologic strain. Further from the impact site we observed multiple behaviors re-

lated to elevated nuclear membrane permeability, but with additional signals suggesting other biological mechanisms at work. For example, a subset of cells displayed multiple levels of nuclear membrane permeability (Fig. 2.41), which may reflect multiple inputs or different stages of cell death processes. Collectively, such results open the door to analyzing how cellular responses are coordinated at the tissue scale. These empirical observations do not drive a concrete understanding without further experimentation, but provide the foundation for establishing how cellular behaviors change in future experiments aimed at probing specific ion channels, mitochondrial function, senescence, and other processes.

More broadly, STRAINS is customizable for analyzing many other tissue systems, and scales well for large numbers of cells. In particular, while the data shown here relate to investigating mechanotransduction in articular cartilage, the described techniques can be applied to any tissue scale system where cell response can be quantified using fluorescence. For example, traumatic brain injury is caused by large mechanical forces on brain tissue [50], with cellular mechanotransduction playing an important role in pathology [51]. Similar staining protocols could enable application of STRAINS to this system. In cardiac tissue, the role of mechanotransduction in determining cardiac myocyte behavior has been studied in single cells, but STRAINS can be applied to address these questions on a tissue scale [52]. In tumors, drug diffusion is hampered by various tissue-scale complications [53]. STRAINS can be used alongside fluorescence labeling and deep tissue imaging techniques [54] to spatiotemporally assess diffusion and drug delivery in tumors. Collectively, these examples speak to the potential for implementing STRAINS to comprehensively study signal transduction *in situ* on the tissue scale for a wide range of systems.

Furthermore, STRAINS aligns with new techniques in multiplexed imaging and large-scale omics data collection in the push for spatially-resolved cell data. Recently developed methods such as PASTE can produce full tissue-scale renderings of transcriptomic data, enabling identification of gene expression and cell type within tissues [55]. Similarly, techniques like IBEX [56] or Cell DIVE [57] make use of immunofluorescent imaging to detect protein-level spatial organization of cells and tissues [58]. Further, spatially-resolved isotope tracking has recently been used to quantify metabolic activity in various tissues [59]. While most of these techniques capture data at a single time point, STRAINS enables real-time *in situ*, nondestructive spatiotemporal mapping and analysis of cell behavior in response to dynamic stimuli. Integrating STRAINS with such techniques would allow us to probe how any stimulus affects coordinated cellular responses on the milliseconds-to-hours timescale, resulting in patterns of, for example, gene expression, protein synthesis, energy utilization, or ultimately cell and tissue fate. As a new tool capable of simultaneously tracking multiple responses of thousands of individual cells and analyzing patterns of cellular behaviors, STRAINS provides insights into how events are coordinated in complex biological systems. By combining time histories of cellular responses with spatial maps of behavioral distributions, we have demonstrated that STRAINS can effectively make use of large datasets to study signal transduction and cell fate in the context of tissue injury and disease.

2.5 Materials and Methods

The method consists of three main components to comprehensively study chondrocyte responses to strain: *in situ* fast confocal imaging, cell tracking and in-

tensity extraction, and cell signal analysis.

Impact-induced trauma to articular cartilage explants *in situ*

Dissection

Samples were sterilely dissected from the femoral condyles of neonatal bovinds obtained from a local abbatoir (Gold Medal Packing, Rome, NY) within 24 hours of sacrifice. Cylindrical explants (6mm diameter x 3mm depth) were extracted with a biopsy punch (Fig. 3.1a) and cultured for 24 hours at 37°C, 21% O₂, and 5% CO₂ in low glucose media containing phenol-free Dulbecco's modified Eagle's medium containing 1% fetal bovine serum, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) 0.025 mL/mL, penicillin 100U/mL, streptomycin 100U/mL, and 2.5mM glucose (Fig. 3.1b). Special care was taken to ensure that the surface of the tissue was cut perpendicular to the depth of the cylinder to maintain uniformity of the strain field during mechanical testing.

Bovine synovial fluid (abbatoir derived, Lampire Biologics, Pipersville, PA) was applied to the joint surface to ensure smooth cutting and to lower the shear forces applied onto the tissue by the biopsy punch, to preserve as many chondrocytes as possible. Similarly, synovial fluid was applied to the blade of extra sharp razors used to bisect the sample. Samples were trimmed to 3mm thickness and bisected in a custom built stainless steel cutting jig, where the cartilage was submerged in a warmed PBS bath during the entire cutting process.

Staining

In order to measure cellular signaling and mitochondrial activity during impact-induced trauma, the tissue is fluorescently labeled with 3-color assay: (a) Calbryte 520 AM, a intracellular calcium flux assay to observe cellular calcium signaling (4 μ M, 1 hr incubation at 37°C), (b) tetramethylrhodamine, a mitochondrial membrane potential indicator to observe mitochondrial polarity (TMRM, 10nM, 1hr incubation at 37° C), and (c) Sytox Blue, a nucleic acid stain used to identify dead cell nuclei (1 μ M, 1hr incubation at 37° C) (Fig. 3.1b). The stains selected for the assay can be modified to reflect parameters of interest in the study.

Cylinders were bisected and mounted side by side on the back plate of a previously described confocal-mounted impactor, with the deep zone of the tissue adhered to the backplate of the impactor with superglue. Samples were submerged in a bath of Dulbecco's Phosphate Buffered Saline (DPBS) with Sytox Blue to ensure that cells dying over the course of the experiment were labeled (Fig. 3.1c). Previous experiments have demonstrated the efficacy of the impactor system in inducing mitochondrial dysfunction and cell death.

Injury

The impacting device, which has been described and validated in previous experiments [5, 6], delivers an energy-controlled impact using a spring-loaded piston with an impacting tip consisting of an 0.8mm diameter stainless steel rod. The half cylinders are positioned so that one is centered on the impactor tip and the other is left as an unimpacted control. The impact lasts 5-10 ms, and

produces a peak stress of approximately 1 MPa (Fig. 3.1c). This replicates a superphysiologic loading rate which is characteristic of cartilage injury. However, it is not designed directly to replicate a specific loading pattern, but instead to expose the tissue within field of view of the microscope objective to a wide range of strains, allowing us to directly make connections between the mechanics of impact and injury to cellular responses.

Confocal imaging

The impactor was mounted to the stage of an inverted, spinning disk confocal microscope (3i Marianas) with a 10x objective, which allowed the capture of a $660\mu\text{m} \times 660\mu\text{m}$ (512×512 pixel) area. The site of impact, along with three surrounding locations (two adjacent sites on the articular surface and one adjacent site into the depth of the tissue), and two locations on the control sample were captured. Together, these imaging sites combine to capture up to 1.2mm into the depth of the tissue, and 1.8mm laterally surrounding the impact site and up to a similar depth in the unimpacted control. Z-focus was centered on a depth $30\mu\text{m}$ away from the cut surface of the tissue to avoid imaging chondrocytes damaged during the sample preparation process. Additional sites can also be used, with limitations set by the scanning speed of the confocal microscope (Fig. 3.1d).

In order to observe the peracute timecourse of events surrounding impact, we image continuously during and immediately following impact at a rate high enough to capture the cellular dynamics of injury. Cell calcium concentration is imaged in the green channel with 25ms exposure, a fast enough imaging rate to observe sub-second changes in chondrocyte calcium (Fig. 2.2a).

A three-color staining assay is used to monitor the subsequent effects of impact in the longer minutes to hours timescale. Images were collected every 20 seconds sequentially at each site, with Calbryte 520 AM (green; 488nm excitation/499-553nm detection), TMRM (red; 561nm/563-735nm), and Sytox Blue (blue; 405nm/414-479nm). A slower rate of imaging is used to allow for longer exposure with the weaker red and blue signals, while simultaneously minimizing the effect of photobleaching (Fig. 2.2b).

Modified particle tracking

While the original Crocker and Grier algorithm [43] assumed a relatively constant intensity for the tracked particles, our cell fluorescence signal could vary significantly. Thus, to maintain a coherent track when fluorescence levels fluctuated, we interpolated between successfully tracked frames. We found that linear interpolation was reasonable when the distance between tracked frames was small, typically less than a cell diameter. Finally, once the tissue was relaxed and cells were nearly static in their position, we continued to measure fluorescence over long periods of time even when cells were not visible. This procedure captured tracks for $\sim 96\%$ of the cells uniformly distributed over the entire tissue for time scales ranging from milliseconds to hours post impact.

The coordinates of each cell centroid in each frame were used to extract intensity data for each channel. In order to account for fluctuations in imaging and inhomogeneities within a cell, we averaged over a 3x3 pixel region surrounding the centroid pixel. The size of this region was chosen to ensure that data collection in one cell did not overlap with data collection in neighboring

cells, or extend into the extracellular matrix. We then extracted the fluorescence intensity of each channel for each tracked cell over the entire experiment.

Fluorescent intensity validation

Multiple controls were included for each of the fluorescent stains, in order to provide relevant comparisons with experimental data. Calcium staining controls were conducted with EGTA, a calcium chelator, and Thapsigargin, an endoplasmic reticulum calcium chelator, in 0mM Ca_2^+ media as a minimum, and with 10mM Ca_2^+ media as a maximum. Mitochondrial polarity controls were conducted with FCCP in the media, which fully depolarizes mitochondria, as a minimum, and incubation with oligomycin, which hyperpolarizes mitochondria, as a maximum. Nuclear membrane permeability/cell viability controls were conducted with ethanol in the media as a minimum, and incubation with P188, a membrane stabilizer, as a maximum.

Background subtraction for confocal images

After tracking cells within the tissue, background subtraction is implemented. Impact-induced cartilage trauma can cause formerly cell-localized stains to leak into the extracellular matrix, locally increasing the background intensity of certain areas of tissue for some time. This causes the background to be both non-uniform throughout the imaging frame and also changing with time. We address this problem by dividing each image into an 8-by-8 grid and subtracting the mean value of the twenty lowest pixel values within each region. For a 512-

by-512 pixel image, this corresponds to the twenty lowest pixel values out of 4096, ensuring that no pixels within cells would be accounted for in this background subtraction.

When necessary, we used a moving average to smooth the data to remove high-frequency noise arising from fluctuations in the confocal images. The size of the moving average window is chosen to adequately remove high frequency noise without disturbing the shape or size of features of interest in the time series. For the time series data in our system, a window size of 20 frames is optimal.

Details of feature extraction

To find peaks in the timeseries, we used the MATLAB FINDPEAKS function with slight modification. The 'minpeakprominence' parameter was set to 3, and a width-to-prominence ratio of 10 was used to filter out extremely wide peaks not considered calcium transients. An additional shape parameter was established by selecting for peaks with a set width to prominence ratio in order to filter out extremely broad peaks that should not be considered transients. Once the peak detection parameters were established for one data set, they could be used to identify peaks in subsequent data sets.

We identified sudden changes in the fluorescence timeseries data using the MATLAB FINDCHANGEPTS function. Here, we adjusted the residuals such that the fitting parameters identified either one or two changepoints per "step" in intensity, and additionally filtered for changepoints where the slopes before and after the identified point were significantly different, to distinguish from

baseline drift. Similarly to setting the peak detection parameters, once change-point detection parameters were established, they too could be used to identify steps in the fluorescence data of subsequent data sets. Notably, these features are highly customizable with user-adjusted inputs, and can easily be adapted to other systems.

Full decision tree methodology

A visualization of the decision tree algorithm for our system is available in the Supplementary Information. In this algorithm, individual cells are run through the tree one-by-one. We start with one of the most conditional arguments. In the following description, all numbers have arbitrary intensity units. Cells where the nuclear membrane permeability starts high (Fig. 2.4a) are categorized based on their maximum blue value (greater than 7) and frame at which blue reaches maximum (less than 7). Then, the tree splits between cells that have blue changepoints and cells that do not have blue changepoints.

For cells that have blue changepoints, we ask if there are more than three changepoints to determine cells where the nuclear membrane permeability has multiple levels (Fig. 2.4l). For cells with fewer blue changepoints, we determine if the mean difference between red and blue timeseries is greater than 20 (red being higher). If true, then the cell has polarized mitochondria but the nuclear membrane permeability still increases (Fig. 2.4e). If false, then we determine if the cell has green peaks. If the green peaks are before the blue changepoint, then the calcium transient occurs and the nuclear membrane permeability increases (Fig. 2.4h) and if they are after the blue changepoint, then the cell has cal-

cium transients after the nuclear membrane permeability is elevated (Fig. 2.4g). For cells that have no green peaks, then we determine whether or not the blue changepoint occurs immediately after impact. Cells where the maximum value occurs after frame 120 and the blue changepoint after frame 100 are considered to have a late increase in nuclear membrane permeability, not immediately from impact. If true and there is a green changepoint, then the calcium concentration drops and the nuclear membrane permeability increases (Fig. 2.4f). If true and there is not green changepoint, then the nuclear membrane permeability increases late with no obvious trigger (Fig. 2.4d). If the cell increases in nuclear membrane permeability immediately after impact, then we determine if it has blue peaks, which indicate an increase and decrease in nuclear membrane permeability (Fig. 2.4c). If there are no peaks, and the range after the blue maximum is lower than 8, then the cell's nuclear membrane permeability increases and plateaus (Fig. 2.4b). All remaining cells in this branch of the tree (has blue changepoints) are categorized as nuclear membrane permeability starts elevated (Fig. 2.4a).

For cells that do not have blue changepoints, we first determine if the cell has green peaks. If yes, then the cell has physiologic calcium transients (Fig. 2.4i). If no, we find cells where the range of blue values is greater than 10, indicating that the nuclear membrane permeability rises (Fig. 2.4b). Then, if the range of blue values is lower than 10, we determine whether the mean of red values after frame 50 is lower than 6. This frame value is chosen to represent a significant enough time after impact where cells have stabilized to their state. For cells with a high red mean, they are normal cells with polarized mitochondria (Fig. 2.4j), and cells with a low red mean have all low signals and depolarized mitochondria (Fig. 2.4k).

Finally, in developing this decision tree we were able to make more informed decisions about how we manually classified cells. For example, the normal cells (Fig. 2.4j), and the all low signal cells (Fig. 2.4k) exist on a continuum. Both categories feature no calcium signaling and no elevated nuclear membrane permeability, with the sole distinction between the two being the intensity of mitochondrial polarity. While hand-sorted cells were judged to the best of our human efforts, there was no easily defined cutoff between the two categories. In programming the decision tree, we specifically defined a numerical cutoff for mean values of fluorescence which provided a more quantitative method of separating the two categories. By iteratively designing the decision tree and updating the manual classifications, we were able to produce more accurate classifications.

Time Series Classifier Parameters

Each classifier in the sktime library has its own model and training parameters. We optimized the parameters for our dataset by testing a range of values for each parameter. We found that while increasing the number of estimators or kernels may very slightly increase accuracy, the run time of training and testing these models increased significantly. We also tested other available parameters for these models, but have only listed relevant ones which noticeably affected accuracy.

For the Canonical Interval Forest (CIF) classifier, we used $n_{\text{estimators}} = 200$ and $n_{\text{intervals}} = 100$. For the Diverse Representation Canonical Interval Forest (DrCIF) classifier, we used $n_{\text{estimators}} = 200$ and $n_{\text{intervals}} = 40$. For the RandOm

Convolutional Kernal Transform (ROCKET) Classifier, we used $n_{\text{kernel}} = 50000$.
Finally, for the Arsenal Ensemble Classifier, we used $n_{\text{kernel}} = 10000$ and $n_{\text{estimators}} = 50$.

2.6 Supplementary materials

2.6.1 Strain-dependent cellular response

The strain field resulting from impact and the associated cell response had complex behaviors that varied spatiotemporally. For calcium, in the milliseconds after impact trauma (Green in Fig. 2a), increased concentrations can be seen in the cells proximal to the impact site. By one second, we observed spatially sporadic increases in calcium deeper in the imaged region. After a few seconds, the cells damaged at the impact site begin to fade, suggestive of death, while cells in the deeper regions exhibited a more uniform increase in calcium. Subsequently, the calcium concentration in cells below the impact slowly decayed. These differing calcium responses in cells that experienced various amounts of compression suggest that distinct mechanotransduction pathways may have been activated depending on mechanics governing the local strain field and the phenotype of each cell, which is known to vary with depth in cartilage [60, 61, 62].

Interestingly, in addition to this depth dependent response, we observed a different temporal pattern of calcium concentration in the region laterally adjacent to the impact (region indicated by orange boundary). Cells in this region near the articular surface primarily experienced shear strain during impact and showed increases in calcium on the tens of seconds time scale, with maximum intensities that were almost an order of magnitude lower. These data suggest that shear and compressive strains may trigger different mechanotransduction pathways, consistent with findings in cell-agarose constructs [41].

Such complex spatiotemporal patterns are also exhibited on longer time

scales in all three measured channels (Fig. 2b): mitochondrial polarity (Red), calcium concentration (Green), and nuclear membrane permeability (Blue). We found that mitochondrial polarity rapidly diminished at the impact site in the minutes after injury. In the surrounding regions, we observed a slow decay in mitochondrial polarity over the course of multiple hours. Calcium concentration largely followed the same pattern, with some cells which exhibited transients on the scale of minutes (See supplementary Video 2). Conversely, nuclear membrane permeability initially showed a very low intensity throughout the region and reached higher intensities in a fraction of the cells in regions extending down to 400 μ m below the impact site on a time scale of hours. Finally, consistent with the short time calcium response, this pattern of cell death did not extend to areas of the tissue which experienced primarily shear strains.

Collectively, these distinct spatiotemporal patterns of cell response indicated that multiple mechanobiological pathways may have been activated in response to local strain. Cells at the impact site most likely died immediately due to membrane rupture caused by extremely high strains [6]. Cells further away from impact, which are subjected to lower strain intensities, may have activated different signaling pathways such as physiologic calcium signaling which led to normal tissue response (i.e. TRPV4 pathway [15, 40]) or superphysiologic calcium signaling which led to apoptosis (i.e. Piezo 1/2 [63, 36, 37]) each of which produced cellular signatures with characteristic combinations of fluorescence intensity curves. Developing an understanding of how such processes are related requires identifying distinct cellular signatures and mapping out where in the tissue they are localized. To obtain these maps, however, we must first identify each cell, track its movement and multi-channel fluorescence response over time (Fig.1e), and classify its cellular signature (Fig.1f).

Manual sorting of fluorescent intensity analysis in impacted articular cartilage shows spatially distinct cell behaviors

Below the impact site, we observed that many chondrocytes showed a high level or a rapid increase in their nuclear membrane permeability and low mitochondrial polarity, which most likely indicated that they were undergoing cell death. Interestingly, we observed different shapes in this channel. In some cells, nuclear membrane permeability was immediately elevated following impact and stayed high (Fig.4a). These cells were primarily localized in the 150 μ m just below the impact site. A second group showed a rapid increase in permeability in the 20 min following impact and maintained this intensity throughout the observation window (Fig.4b). These cells were distributed in a region extending 400 μ m below the impact site. A third group showed a similar rapid increase in nuclear membrane permeability in the first 20 min following impact but then decayed in this channel to an intermediate level still higher than baseline. These cells could be found up to 1 mm below the impact site (Fig.4c). Additionally, in some cells down to 450 μ m below the impact site we observed a late rise in nuclear membrane permeability with no obvious trigger (Fig.4d), while in others we observed a rapid rise in nuclear membrane permeability despite the cell maintaining high mitochondrial polarity (Fig.4e).

There were a number of distinct behaviors associated with changes in cell calcium concentration. In the region located 150 - 400 μ m below the impact site, one group of cells showed a rapid drop in calcium concentration followed by a rapid increase in nuclear membrane permeability (Fig.4f). Another group showed calcium transients in cells where the nuclear membrane permeability was already elevated (Fig.4g). A third group of cells exhibited one or more su-

perphysiologic calcium signals with mostly rounded peaks, and a subsequent increase in nuclear membrane permeability indicating cell death (Fig.4h). In regions greater than 400 μm below the impact site there was a group of cells that exhibited one or more physiologic calcium signals with mostly square peaks, maintained relatively high mitochondrial polarity, and showed no change to the nuclear membrane permeability (Fig.4i).

Finally, we observed two additional groups of cells that maintained low nuclear membrane permeability throughout the experiment. In the first group the mitochondria remained polarized (Fig.4j), while in the second all three signals were low (Fig. 4k). Both of these groups were evenly distributed throughout the impacted samples and the controls.

2.6.2 Decision Tree Methodology

See Fig. 3.3 below for a graphical description of the decision tree logic.

2.6.3 Tracking and Intensity Analysis Code

This set of codes requires Crocker Grier particle tracking code found here: <https://site.physics.georgetown.edu/matlab/>, and `export_fig` from the MATLAB Fileshare found here: https://www.mathworks.com/matlabcentral/fileexchange/23629-export_fig, as well as the MATLAB Signal Processing Toolbox.

The tracking code does not necessarily re-assign the same cell ID to the same

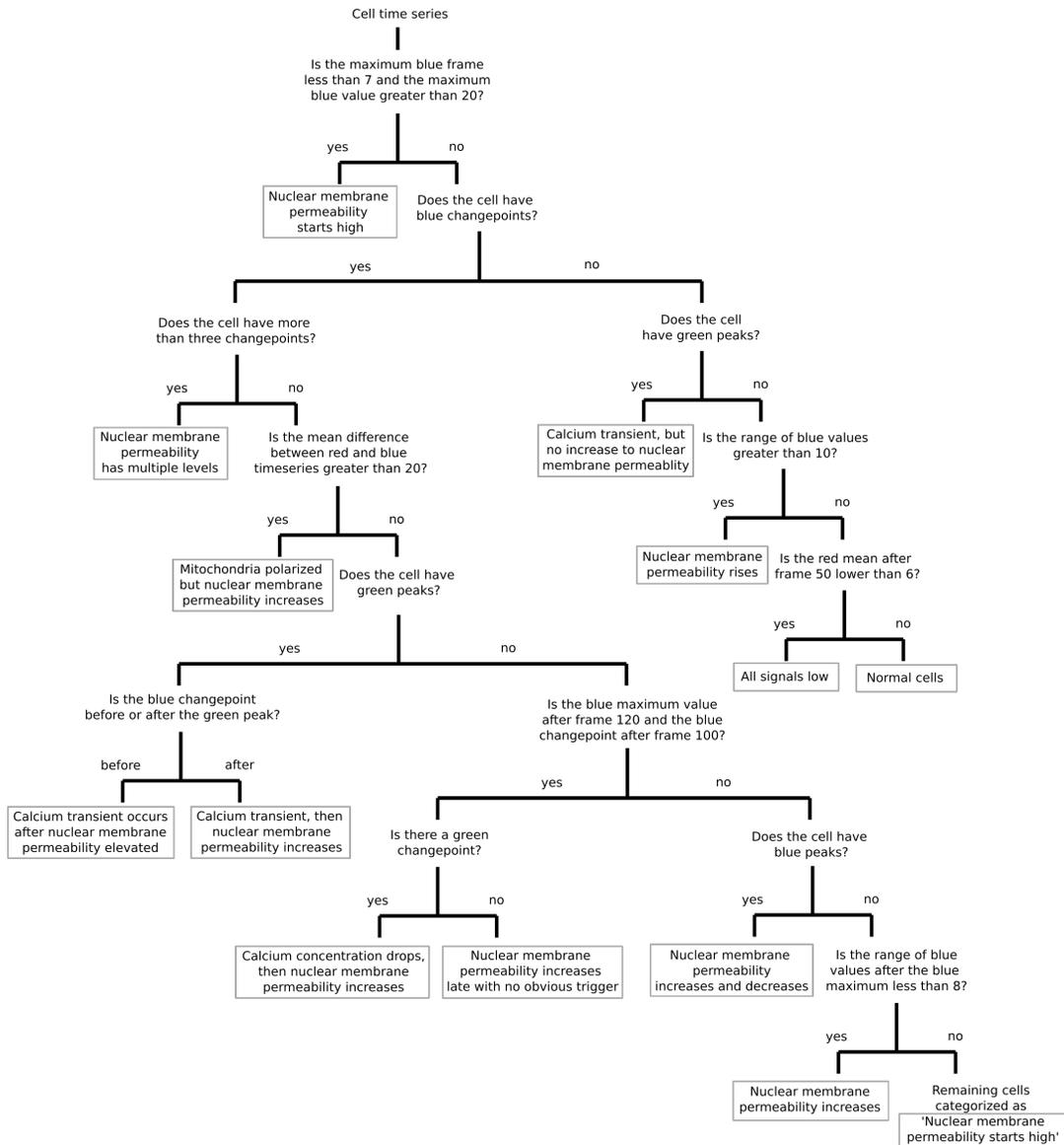


Figure 2.5: The full decision tree algorithm for our system. Each cell time series is run through the tree separately. Grey boxes indicate final sorted categories.

cell each time it is run, due to the nature of the way the Crocker and Grier code is written. Because of this, different sample data is provided for some of the different steps here. This is because the manually-sorted data is not necessarily going to have the same cells with the same ID when the user runs the sample data.

The post-impact input files for this code should be in RGB format .tif files. ImageJ or Fiji can be used to convert from other image formats. Ensure that the colors are in the correct channels (for example, Slidebook likes to swap R and B). The impact input file should be an 8-bit .tif file.

For cells that are not moving, the video can be registered onto the first frame. This will allow the user to connect tracking between the impact and post-impact videos (and only works at the impact site). MATLAB's `fitgeotrans` and `imwarp` functions are used to accomplish this. An example snippet of code is provided, but this is not used in the final data analysis so is not integrated into the code as a whole.

To run the codes provided, ensure that the MATLAB directory includes all files within the folder. Start with `all_tracking_function_calls`. This code will require the user to fill in the various parameters associated with the timing of the images, where the images are saved, alongside the positions that the user wishes to track. Impact and post-impact tracking parameters for the Crocker and Grier algorithm must be filled in, with example values saved within the code itself. Feature extraction parameters should also be filled in (for detailed explanation see the code). Once all parameters are filled, the code can be run. Input parameters for each of the called functions can be found in the functions. Most of the functions require a folder and date to designate where the images are stored and where the output files should be sent to. `TrackImpact`, `TrackPostImpact`, and `FeatureExtraction` have print parameters, which can be set to 'on' or 'off' to determine whether or not the tracking images or feature extraction images are printed. These images are used to help optimize the parameters for tracking or feature extraction. The first time the code is run, the should be

set to 'on' in order to ensure that the tracking is working correctly, and that the peak detection looks correct. Future iterations of the code with optimal parameters do not require printing the images. The code takes significantly longer to run if printing all images, but will be faster on a computer with a better GPU (MATLAB renders faster).

Once `all_tracking_function_calls` is run, the tracking and feature extraction is complete. The user is given a choice between the `sorting_function_calls_manual` and `sorting_function_calls_nomanual` codes. For the first time running the data, the user should use the GUI (detailed below) to identify the categories of all of the cells. These categories are then used to label and sort all of the cells. Sorting can be easily accomplished by creating a copy of all intensity curves, then using the first section of `sorting_function_calls_manual` to make folders corresponding to all of the input categories. `ManualDataCompilation` will then scrape file-names from within each of these folders to add them to a structure with the correct labels. From here, both of the sorting codes will run `CellAttributes` to organize features for each time series and `DecisionTree` to categorize them, finally using `SplitPeaksDataCompilation` to format the data into shape for time series classification.

The decision tree was custom programmed for our system. After `CellAttributes` is run in the main script, the features are used to categorize cells into their respective categories. The parameters of the tree must be changed for each new system, with the overall structure remaining quite similar. For other systems, new categories may be identified, and used to determine the if-else statements building the tree.

2.6.4 MATLAB GUI

The GUI requires MATLAB 2019a or newer. If other code has been run that has changed figure defaults (like code above), it is recommended that a new instance of MATLAB is opened or the command `'reset(groot);'` is run before starting the GUI to avoid figure sizing errors. A video detailing how to use the GUI is included, along with example data.

The GUI has several main functions: to observe cells and groups of cells within the video and plot their individual fluorescence intensities, and to look at how these intensities are distributed. First, pre-processed data is loaded in using the interactive menu and Windows explorer dialog. The pre-processed data consists of the video, tracked intensities of each cell, and location of each cell, with the same naming convention (see example data). Intensity data is taken from `'pos_()_intensity'` and `'pos_()_locs'` .mat files. Then, the video player in the center can be used to either play at a specified framerate or scroll through all of the frames within the video. Individual cells can be clicked on within the video to display their intensities on the plot side of the GUI. The Cell ID (given by the particle tracking code) or x-y coordinates can be used to find cells as well.

Groups of cells can be selected using the rectangular selection tool. When using this functionality, the plots on the right side of the GUI can be interacted with. First, click within the plot area on one of the three colored plots (but not directly on a line). Then, a cursor will appear. Click on a point within the graph. After this process is completed, `'highlight cell'` can be clicked on to circle the cell within the image, `'Timepoint Histograms'` can be clicked on to produce a pop-up with histograms of all three colors at the selected time, and `'Cell Populations'` can be clicked on to produce a pop-up with the selected area split into three

populations based on the intensity of the blue curve at the selected time. At any point, 'Clear Current Data' can be used to start over. At the end of analysis, 'Save Cell ID List' will pop out a Windows dialog to save the list of cells that were interacted with within this session of the GUI.

2.6.5 Time Series Classification Code

The hand-sorted data above is used to train several time series classifiers. Data is loaded as a MATLAB .mat file, which is converted to a dataframe. All of the classifier functions are located within `classifier_functions.py`. In order to train models, labeled data is used, alongside the `trainCIF`, `trainDrCIF`, `trainROCKET`, and `trainArsenal` functions. These functions will save the models, alongside the accuracy of the model tested on a subset of the data. To use the trained models, the `loadCIF`, `loadDrCIF`, `loadROCKET`, and `loadArsenal` functions can be used. This will label new data. The model will not be trained exactly the same every time, due to random seeding. However, the accuracies should be very similar.

Dependencies of this code are listed within the functions file. Sktime can be found at <https://www.sktime.org/en/latest/>.

2.6.6 Example Application of STRAINS to Worm Neuron Data

We tested our system on data obtained from Wen, et. al. *eLife*, 2021 [64]. Using GCaMP and calcium dynamics data from Fig. 5 in that publication, we applied our modified particle tracking to the video of a moving worm brain. Our GUI was modified to automatically accommodate the new video dimensions,

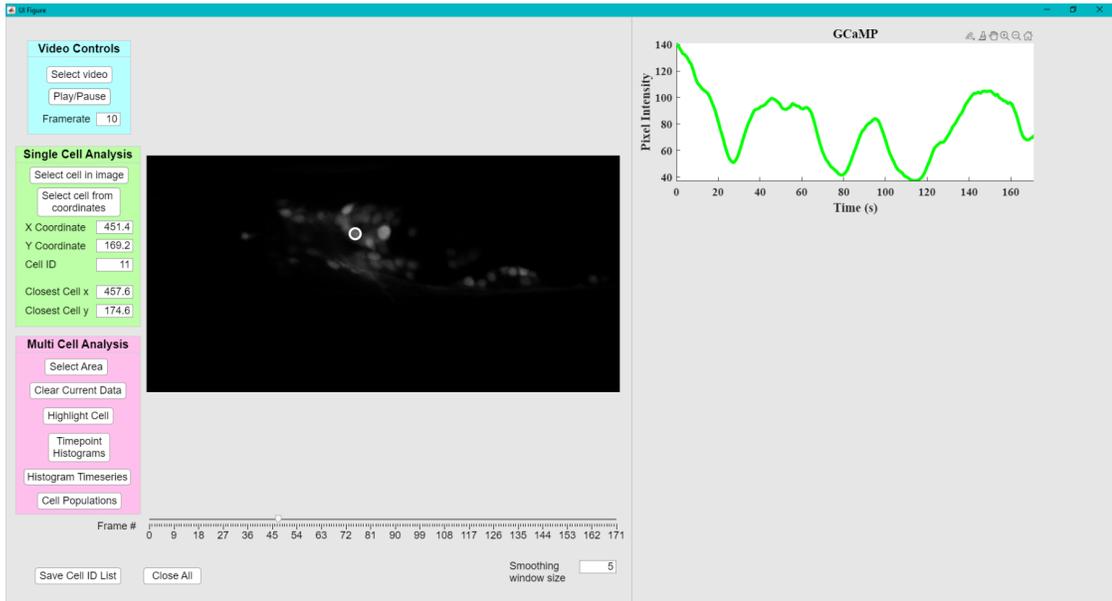


Figure 2.6: Example screen capture of modified GUI on worm GCaMP data for selecting single cells within the video.

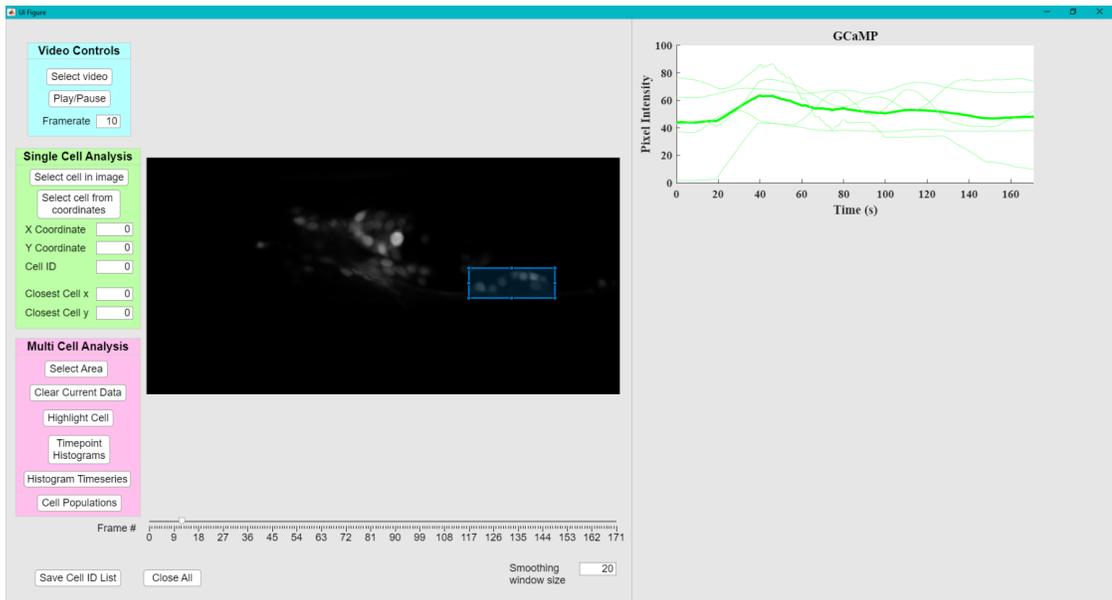


Figure 2.7: Example screen capture of modified GUI on worm GCaMP data for selecting groups of cells within the video.

tracking, and single channel data (Fig. 2.6 and Fig. 2.7). Based on manual classification, unmodified time series classification was trained and tested on calcium dynamics data, reaching classification accuracy of up to 91.2% for all four classifiers, CIF, DrCIF, Arsenal, and ROCKET (similar accuracy is obtained for all classifiers due to small population size of data - 137 cells total).

CHAPTER 3

VARIATIONAL AUTOENCODER FOR CLUSTERING CARTILAGE DATA

Zheng, J., Teoh, H.K., Delco, M., Bonassar, L. J., & Cohen, I., *Manuscript in preparation*

3.1 Abstract

In various biological systems, analyzing how cell behaviors are coordinated over time would enable a deeper understanding of tissue-scale response to physiologic or superphysiologic stimuli. Such data is necessary for establishing both normal tissue function and the sequence of events surrounding injury. However, collecting and analyzing these large datasets presents a challenge—such systems are time consuming to process and the overwhelming scale of data makes it difficult to parse overall behaviors. This problem calls for an analysis technique that can quickly provide an overview of the groups present in the entire system and also produce meaningful categorization of cell behaviors. Here, we demonstrate the application of an unsupervised method—the Variational Autoencoder (VAE)—to learn the features of cells in cartilage tissue after impact-induced injury and identify meaningful clusters of chondrocyte behavior. We quickly generated new insights into the spatial distribution of specific cell behavior phenotypes and connected specific peracute calcium signaling timeseries with long term cellular outcomes, demonstrating the value of the VAE technique.

3.2 Introduction

Exploring how cells coordinate their behaviors in response to stimuli is important for understanding tissue function in health and disease. Tissues are complex systems where many factors such as spatial location and structural mechanics can affect how mechanical and biochemical signal transduction between cells and dictate the subsequent cellular response. Methods have been developed to approach this problem through imaging the behaviors of individual cells over time [39, 65, 66, 67, 68, 69] or taking snapshots of pooled cell populations [38, 70, 71]. However, fully addressing this problem can be difficult. In order to make specific conclusions about how coordinated cellular behaviors are affected by external factors, the spatial and temporal behaviors for each of thousands of cells within a tissue must be collected, processed, and interpreted for many iterations of experiments. Experiments with different stimuli must then be compared in order to determine how behaviors may change. This process amounts to an overwhelming amount of data for analysis using well-established methods.

Recently, we developed a technique that combines real-time *in situ* imaging of cartilage tissue during tissue impact with supervised learning techniques to establish and probe specific behaviors within the system (STRAINS) [72]. Our method enables the detailed spatiotemporal analysis of individual cell behaviors and the classification of specific cell responses and phenotypes. However, such process requires initial manual classification, which is initially time consuming for large datasets and may miss subtle differences between cells due to human error.

An unsupervised machine learning method that would allow the simultaneous analysis of thousands of time series alongside other factors such as location or strain would enable the fast and unbiased categorization of cellular behaviors. We propose a promising new technique using a Variational Autoencoder (VAE) to learn the basic behaviors of a tissue system in an unsupervised manner and make use of this learning to cluster cells by behavior.

The Variational Autoencoder (VAE) is a recently developed probabilistic generative neural network [73]. Comprised of an encoder, a decoder, and a loss function, the VAE iterates upon itself to learn how to accurately reproduce input data, and map that data to a latent space. The latent space of the VAE is connected (two points in the space that are close together give similar decoded results) and complete (all points in the latent space give meaningful information upon decoding). This feature of the VAE latent space both prevents overfitting and allows for generation of new data. As an unsupervised machine learning tool that can easily process large amounts of data, VAEs have been applied to a wide range of fields, from reconstructing complexities in many-body physics to anomaly detection in industrial robots [74].

In biomedical research, VAEs have been more commonly applied as a diagnostic tool for image classification in MRIs and other medical imaging [75], for the classification of tumors and image segmentation [76, 77], multi-omics data integration, and even molecule and protein design [78]. The processing capability of VAEs have led to its use in single cell analysis techniques have focused on specific biomarkers in separate cells [79] or single cell transcriptome profiling [80] and in time dependent biological signals in spectrographic data with the clustering and analysis of budgie and mouse vocalizations [81].

Here, we propose to expand the use of a VAE to analyze cellular response in articular cartilage, where tissue-scale data of thousands of cells subject to stimuli is recorded over space and time for multiple cellular functions via confocal imaging. The VAE is highly suited for processing this complex spatiotemporal dataset; cartilage consists of chondrocytes within a heterogeneous extracellular matrix with mechanical properties and cellular phenotypes that vary with depth [1]. We use the VAE to accurately reconstruct time series cell data and at the same time, learn a compressed representation of the cellular profile which is subsequently used to cluster via the agglomerative hierarchical clustering method. These clusters are used to identify the phenotypes of behaviors present in the cartilage. Additionally, the VAE reduces complex high-dimensional time series data to a compressed representation, which allows us to identify the dominant effects in the system from the principal components in the latent space [82]. Finally, we use the generated clusters to correlate between sub-second cellular responses and long-term cellular outcomes, to build a fuller picture of cell response after injury. This application of the VAE is a demonstration of an effective tool for the large-scale analysis of spatiotemporal data alongside meaningful clustering for biological insights.

3.3 Results

We trained the VAE on multiple sets of cartilage data. As shown in Fig. 3.1 a-c, articular cartilage samples were loaded onto a confocal-mounted impactor, side-by-side with a control. The sample was impacted, and videos were taken of the impact and the longer-term cell processes in the hours after impact. All of the cells at multiple locations around the impact site and similar sites on the control

sample were tracked over time, and the calcium concentration, mitochondrial polarity, and nuclear membrane permeability data was extracted as timeseries, as described in previous work [72]. This is repeated for multiple experiments.

We input these timeseries into the VAE, which we then iterate until the reconstructions match the input data. Once we are confident that the VAE effectively encapsulates the information contained within all of the cell timeseries, we use it as an effective tool to quickly cluster cell time series data to obtain useful insight into what behaviors are present in the system. This is done in several ways: the latent space of the VAE shows us information about the defining axes upon which the system varies, the clusters generated allow us to speedily learn about which behaviors are present in the system, the breakdown of the cluster tree teaches us about how similar clusters are, and clustering the post-impact data allows us to discern interesting patterns in the impact data itself and learn about how the two timeseries are related.

3.3.1 Principal Components of the VAE Generate Valuable Insights

The VAE process itself can provide us with useful information about our system. The principal components (PCs) of the latent space generated by the VAE show us the dominant behaviors that comprise the system. These form an orthonormal basis upon which we can construct any of our cell behaviors—each cell is composed of some linear combination of all of the PCs. This process identifies the most basic behaviors, the building blocks of all cell behaviors in the system. Different samples may produce different clusters, but the basis on which these

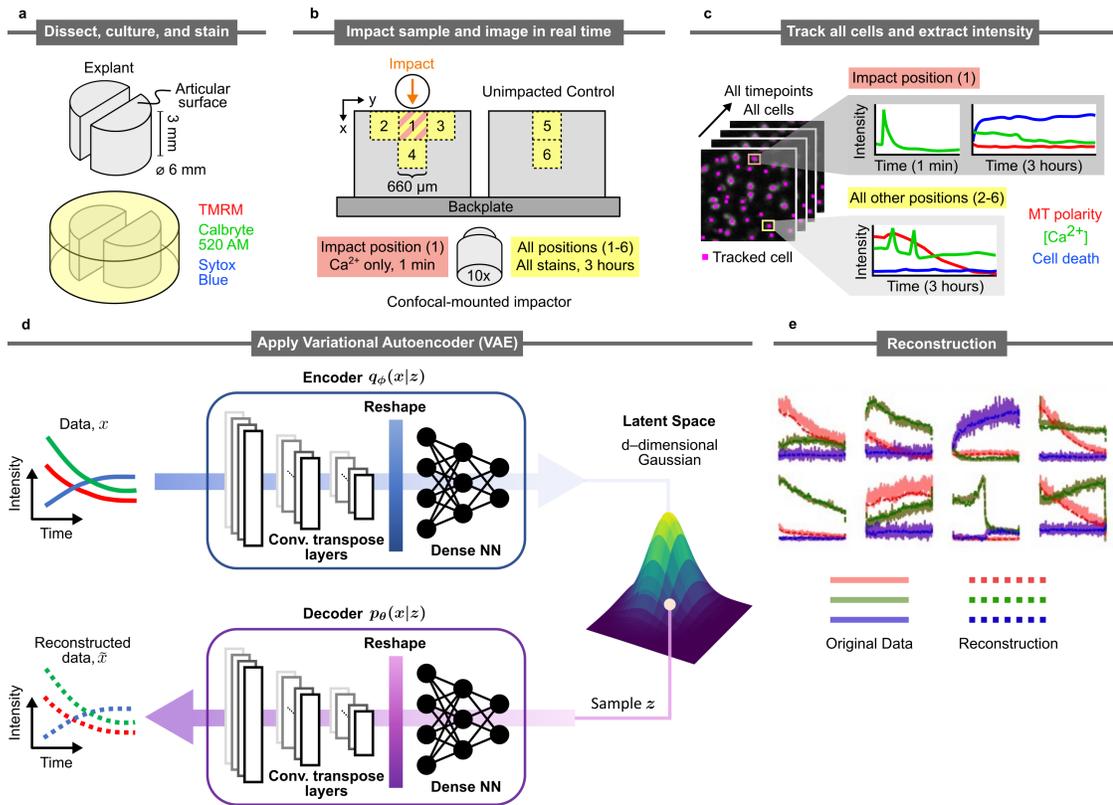


Figure 3.1: Sample processing, imaging, tracking, and VAE structure. a) Biopsy punches of condylar articular cartilage are collected, bisected, and halved. Hemicylinders are stained for calcium concentration, mitochondrial polarization, and nuclear membrane permeability (cell death) via Calbryte 520 AM, Sytox Blue, and Tetramethylrhodamine, respectively. b) Paired hemicylinders are glued side-by-side to the back plate of a confocal-mounted impactor. One hemicylinder is impacted (site 1) while the other serves as control. Imaging occurs during impact at site 1 and after impact at sites 1-6. c) Individual cells are tracked through all time points. Stain intensities are extracted and manually sorted (for one sample, to provide comparison baseline for VAE analysis). d) Structure of the Variational Autoencoder (VAE). The first portion of the encoder consists of three 1D convolution layers and a reshape operation that is subsequently fed into fully connected layers of a dense neural network which outputs the mean and diagonal covariance vectors of the compressed latent space representation. A sample z is drawn from the d -dimensional Gaussian and sent through the decoder to obtain a reconstructed intensity profile. The decoder has the same architecture as the encoder, with the exception that the 1D convolution layers are replaced with 1D transposed convolution layers. e) Example reconstruction of cell intensity data. Light colored lines represent the input data and dark colored dashed lines represent the reconstruction.

cell fingerprints are constructed are the same.

In this system, 75% of the cumulative explained variance can be accounted for in the first four PCs and 90% in the first eight PCs. The three color timeseries data is used without normalization because the absolute intensity value carries important information. Thus, more PCs are required to fully account for the explained variance than some other systems due to the broader range of possible timeseries combinations.

Examples of Principal Components (PCs) can be seen in Fig. 3.2, for three different datasets. Fig. 3.2a plots the first four PCs as they vary from $\alpha = -3$ to $\alpha = +3$ with all other PCs at $\alpha = 0$. The corresponding heatmaps for each of those PCs is seen on the right in Fig. 3.2b, with each cell represented by a dot of color varying from purple $\alpha \approx -3$ to pink $\alpha \approx 0$ to yellow $\alpha \approx +3$. From these heatmaps, we can see that the PCs show strong spatial correlations.

For example, PC0 varies from a cell that dies due to impact at $\alpha = -3$ to a cell with decaying but elevated calcium concentration and decaying mitochondrial polarity at $\alpha = +3$. This PC covers the spectrum between cell death due to impact and cells with slowly decaying function (as expected in an *ex vivo* setup). Notably, cells at the exact impact location (shown with orange arrows in Fig. 3.2b) and some cells below the impact show strongly negative values in this PC, whereas cells to the side of impact and on the control sample show strongly positive values in this PC.

PC1 demonstrated a different trend, with negative values associated with low calcium signal and high but decaying mitochondrial polarity and high values associated with decaying calcium signal and slightly elevated nuclear mem-

brane permeability. Cells with high values in this PC were generally found at the impact site or below. PC2 captured a different type of cell death/dysfunction at high values, where all three signals decayed. This was most prevalent in the dataset on the right, at and below the impact location. Specifically, this dataset experienced the highest strain, indicating that this PC may be positively correlated with strain. PC3 represented no cell death/dysfunction, varying from normally decaying signals to cells with higher calcium activity, which was more often seen on the impacted sample. More examples of principal components can be seen in Fig. 3.3.

Subsequent PCs each contributed less to describing the explained variance ratio, which is obvious from the trend of heatmaps becoming more uniformly pink ($\alpha \approx 0$), there are some notable features that the VAE was able to learn. PC5, PC9, and PC10 all showed an increase in calcium concentration followed by cell death, but at uniquely different timepoints. PC6, PC7, and PC8 showed late cell death also at unique timepoints. These behaviors were known to be less common within the population of cells observed, but were still captured by the VAE during its learning process.

By observing the relative importance and spatial distributions of these PCs, we find that while the dominant behavior of cell death is quite obvious, there are three distinct modes: Instantaneous cell death—highly elevated nuclear membrane permeability signal with minimal other signals, cell death/dysfunction—somewhat elevated nuclear membrane permeability signal with elevated calcium signal, cell death/dysfunction—highly elevated but decaying nuclear membrane permeability signal with decaying other signals. Notably, these are three of the most distinctive features used in manual labeling, which the VAE

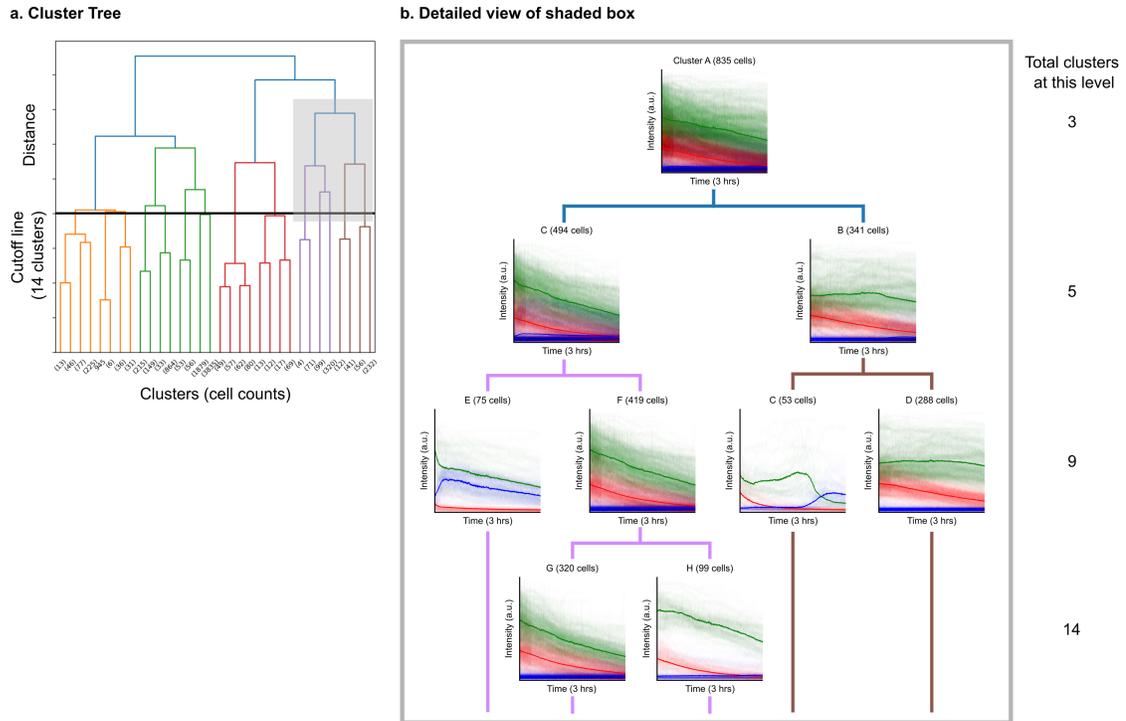


Figure 3.3: Example branching of the hierarchical clustering tree. Left: Overall agglomerative hierarchical clustering, with example cutoff line for 14 clusters shown in black. Each vertical line is a cluster. The x-axis shows cell counts and the y-axis represents distance between the clusters represented by the symmetrized Kullback-Liebner divergence. Right: Expanded view of gray shaded box. Clusters break down into smaller groups with more detail when the total number of clusters is increased. Bolded lines represent cluster averages, and thin lines represent individual cells within a cluster. Red represents mitochondrial polarity, green calcium concentration, and blue nuclear membrane permeability (cell death). Four levels are shown, showing the specificity of clusters when there are 3, 5, 9, and 14 total clusters. Letters are used to name clusters for ease of identification and have no specific meaning.

distinguished quickly without the need for time-consuming analysis.

3.3.2 Clustering Post-Impact Data into Distinct Behavior Phenotypes

Once the VAE learns how to accurately reconstruct cell timeseries data, we use the latent features to determine how different cells are from each other. We use the symmetrized Kullback–Leibler divergence, which measures differences between probability distributions, to quantify distances based on latent variables of each cell. Using this metric, we perform agglomerative hierarchical clustering to cluster our cells, where similar cells are grouped together until clusters form. This process builds up a tree, whereby we can set our desired distance cutoff to select for any number of clusters (Fig. 3.3a).

Starting from a small number of clusters, we get some separation between behaviors, but each cluster still contains too many cells and overlapping behaviors. As we increase the number of clusters, we see separation in behaviors. Eventually, the process produces clusters that are too small to be meaningful. Mapping this out produces a hierarchical clustering tree (Fig. 3.3a). Moving down the cluster tree allows us to determine when it becomes relevant to stop decreasing cluster size. This allows us to look at what number of clusters is relevant to the system. The way that the clusters split up tells us which types of behaviors the VAE considers to be more similar, informing us that we should to look more closely at how to similar clusters differ.

For example, Fig. 3.3a shows the overall hierarchical clustering tree for this dataset and demonstrates an example cutoff distance that produces fourteen total clusters. The shaded gray box highlights one branch of the cluster tree and is expanded in Fig. 3.3b to show the timeseries within each cluster. We

determined fourteen clusters to be a reasonable cutoff, as eleven clusters still produced clusters that appeared to encompass multiple behaviors, and eighteen resulted in too many clusters with very small cell counts.

At three total clusters, we see that cluster A is largely comprised of cells with decaying calcium concentration and decaying mitochondrial polarity with lower overall intensity. As we increase in specificity to five total clusters, we see cluster A split into cluster B, which looks largely similar to cluster A, and cluster C, which has a overall higher mitochondrial polarity signal and a flat calcium concentration. Moving to nine clusters, we see cluster B split into D and E. Cluster D is very different, showing dead/damaged cells with elevated and decaying nuclear membrane permeability and calcium concentration, whereas cluster E again looks similar to its parent cluster. Cluster C splits, and we see that cluster F is very unique, pulling out cells where there is a calcium transient at roughly two hours into imaging, after which the cell dies. Again, cluster G looks similar to its parent cluster. Finally, at fourteen total clusters, we see cluster E split, this time into two clusters (H and I) where both the calcium concentration and mitochondrial polarity decay, but have different overall intensity values.

Interestingly, we found that the dead/dysfunctional cells in this data set are not separated at a low cluster count and instead split from clusters that look like largely normal cells. The clusters with increased nuclear membrane permeability all appeared at the 9 or 14 cluster level. However, for a different dataset with higher total strain and more cell death trained on the same VAE, we found that the dead/dysfunctional cells were split into different clusters at the 3 cluster level. By doing this exercise, we can gain a sense of the similarity measure

for specific cell behaviors, and their resistance or willingness to be clustered together. These distances between clusters can be used to guide which specific biological factors should be changed in future experiments, or which types of behaviors should be specially probed.

In Fig. 3.4, one dataset is shown broken down into fourteen clusters. For each cluster, the spatial distribution of cells on the cartilage tissue, the timeseries for all of the cells within the cluster, and the UMAP visualization of cells is shown, with cells represented as orange dots. The UMAP embedding visualization includes all of the cells from multiple experiments that were simultaneously used to train our VAE (shown in gray). Notably, the UMAP embedding shows two big groups of cells, with the left group correlated with cells where the nuclear membrane permeability rises at any point during the experiment, and the right group correlates with cells with no nuclear membrane permeability.

In this dataset, the majority of cells at the 14 total clusters level are grouped into cluster 1, which has low and decaying mitochondrial polarity and low calcium and nuclear membrane permeability signals. These cells are present throughout the sample. Clusters 2-9 all showed low nuclear membrane permeability, but different configurations of calcium and mitochondrial polarity decay. When classifying by hand, these cells were all grouped within one phenotype, due to the difficulty of assigning cutoffs (rather arbitrary). By clustering with the VAE, we were able to separately identify cells with constant mitochondrial polarization, cells where the mitochondrial polarization decreased over time, and separate out groups where the calcium concentration was higher/lower than the mitochondrial polarization with varying relationships to the mitochondrial polarization. This produced a more detailed analysis of differences be-

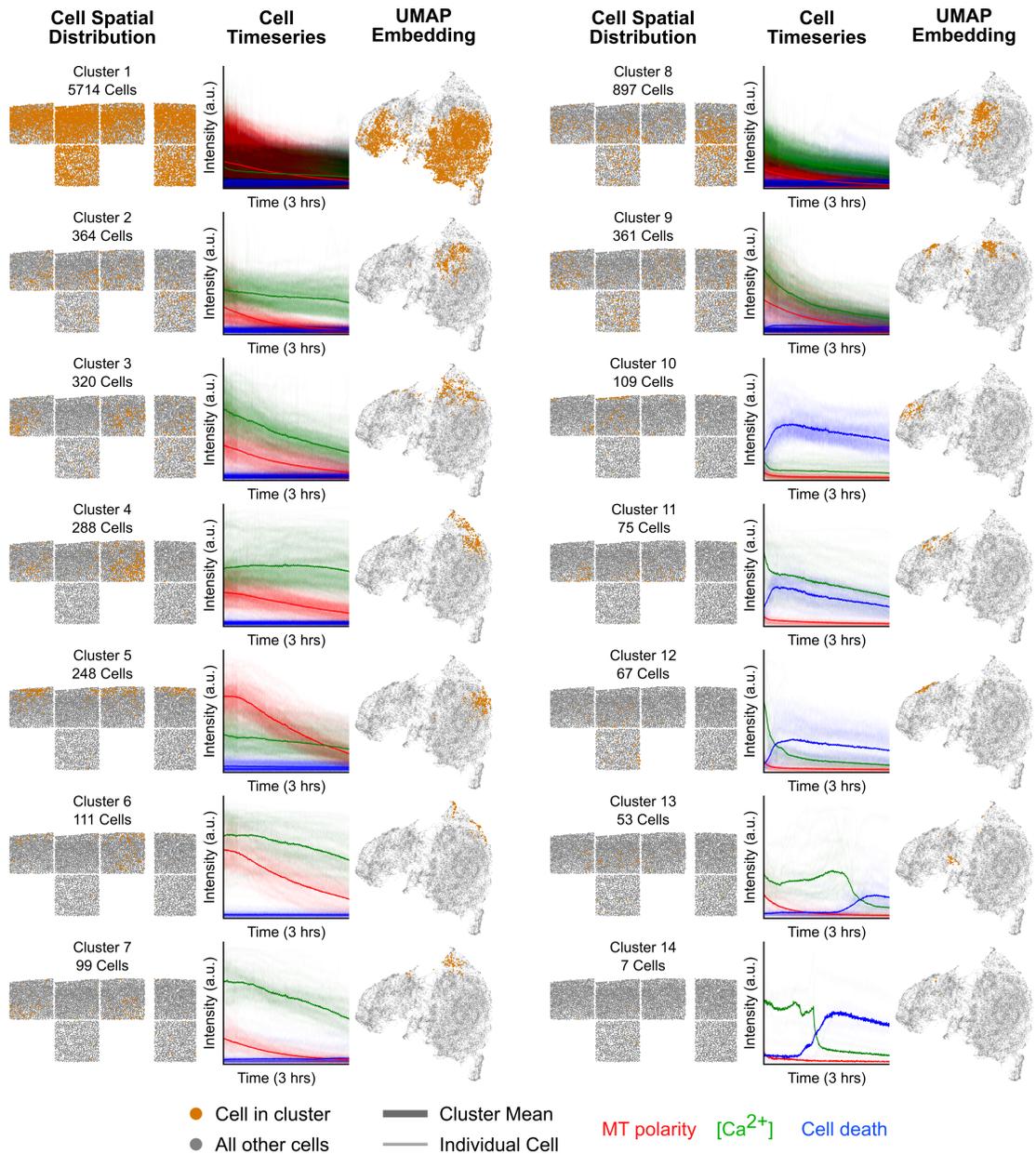


Figure 3.4: All fourteen clusters from one example data set, organized in two columns. Each cluster is shown in three panels: the spatial distribution of cells within the cluster (cells represented as orange dots), the timeseries plots of cells within the cluster (with means as bolded lines), and the UMAP distribution of cells within the cluster (cells represented as orange dots). The clusters do not have a specific ordering.

tween viable cells, which will help to further our understanding of beneficial processes in subsequent experiments.

Clusters 10-14 all showed cells with increases in nuclear membrane permeability, but in different ways. Cluster 10 highlights cells mostly at and directly below the impact site that have a large increase in nuclear membrane permeability at the very beginning of imaging then decayed, with no other signals. Cluster 11 showed similar timeseries, but the cells were more spread out to the sides of the impact site, and all of the cells showed high but decaying calcium concentration. Cluster 12 also showed cells that died near the beginning of image, but with a much more rapid calcium decay and less nuclear membrane permeability decay, spread around the impacted sample. Clusters 13 and 14 showed much later cell death after a calcium transient, with cluster 13 cells dying roughly 2 hours into imaging and cluster 14 cells at 1.3 hours.

The differences between these two clusters highlights the utility of the VAE in picking out small clusters that are very different than other behaviors that may not be picked out by hand but are in fact distinct clusters. The two different timepoints of these cell death events may indicate different methods or triggers of cell death (for example, there were many more cells in cluster 13, including a few present on the control). The ability of the VAE to generate this distinction demonstrates its value as a fast and effective way to generate cell behavior clusters in any stage of analysis.

3.3.3 Correlating Peracute Calcium Response with VAE Generated Clusters

By clustering the post-impact data and plotting the corresponding impact clusters, we were able to quickly identify distinct differences in immediate calcium signaling post-impact between cells that remain viable and cells that display necrotic behavior (Fig. 3.5).

Cells at the impact location are tracked for both immediate impact calcium signaling (1 minute) as well as the longer term post-impact imaging (several hours). Using the post-impact clusters generated by the VAE, we constructed the corresponding clusters for impact calcium signaling. From this, we noticed that the shape of the initial calcium signal is highly correlated with cell death. Cells that show sharp calcium peaks within milliseconds of impact with a subsequent exponential decay were associated with cells that died within the first hour after impact (Fig. 3.5b). Among those groups, the cells that showed higher overall calcium intensity and faster decay of the calcium signal were associated with higher nuclear membrane permeability in the hours after impact. These cells were generally localized directly below the impact site.

Conversely, cells that demonstrated broad peaks or plateaus of calcium concentration after impact (reaching lower overall intensity compared to peaks in the previous categories) were more often associated with cells that did not die after impact (Fig. 3.5a). These cells were generally located away from or to the side of the impact site. While these associations were hinted at from directly viewing videos, they were not fully established until viewed using these VAE generated clusters. Additional analysis could be conducted to further distin-

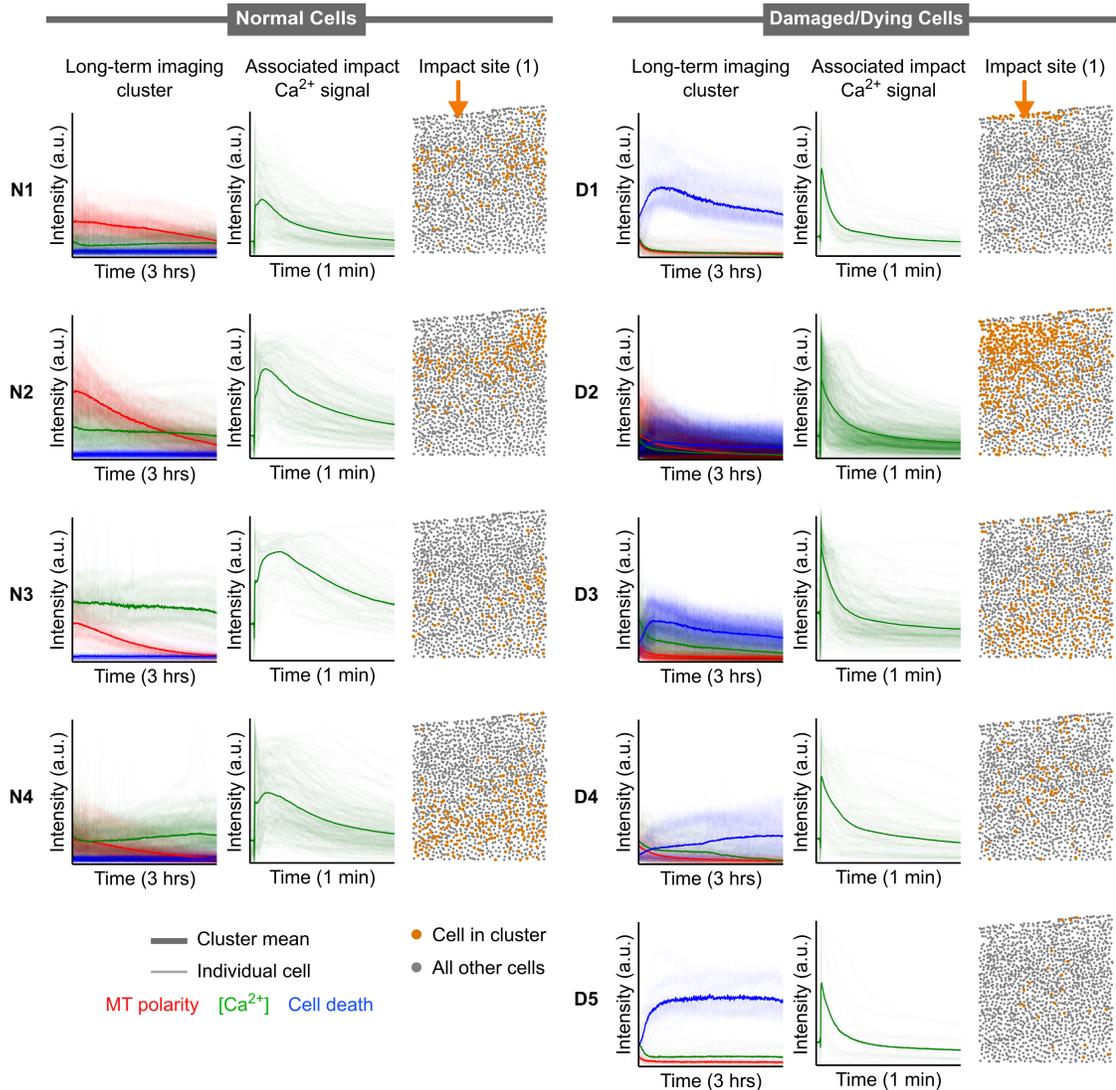


Figure 3.5: Impact calcium signatures associated with VAE-produced clusters. Differentiation between healthy cells and damaged or dying cells can be seen from clustering on post-impact long term data. Each cluster has three panels: the timeseries of all cells in the cluster for longer-term imaging (3 hours) which includes all three color channels, the timeseries of the impact imaging for that cluster (1 minute) which includes only calcium concentration, and the spatial distribution of cells in said cluster which are shown as orange dots. In time-series plots, bolded lines represent means while thin lines represent individual cells within the cluster. The exact impact location is indicated with an orange arrow. Left: clusters showing functional/healthy cells. Right clusters showing damaged/dying cells.

guish between the decay time of the calcium signal and the sKL divergence, the distance metric used to establish clusters.

3.4 Discussion

This novel application of a VAE to analyzing tissue scale spatiotemporal cellular behavior data (confirms, emphasizes, validates) the utility of this unsupervised learning method. The VAE streamlines analysis by quickly processing large amounts of high dimensional data in multi-channel fluorescence microscopy videos with minimal (oversight). This enables us to rapidly iterate through experimental inputs and test hypotheses by changing individual factors and observing changes to clusters or identifying varying spatial distributions.

Using our VAE, we produced clusters that allowed us to identify the types of behaviors occurring in our system, ranging from various types of normal behaviors to distinct patterns of cell death. We also used it to quickly identify correlations between immediate impact and post-impact imaging by using the clusters produced by the VAE.

The VAE method has demonstrated several advantages over a traditional manual labeling system or supervised machine learning. The VAE is capable of working with noisy data, eliminating the need to do extensive background correction. The VAE is fast: once input parameters are set up, no additional work is necessary in order to produce meaningful results. Additionally, the VAE can work with multi-channel and non-normalized data, allowing us to take into consideration the importance of both absolute intensity and relationships between channels in dictating cellular behaviors.

This implementation of the VAE is easily adaptable to analyze any tissue system where behaviors of cells are captured over time, not only from fluorescence microscopy. The method is not confined to any specific geometry, has no limitations on dataset size, and can work with time series data of any dimension. The VAE can be trained on any number of samples while still producing unique clusters for individual samples, enabling us to quickly look at sample to sample variation while simultaneously gaining a sense of the dominant overarching behavioral themes with the principal components.

Finally, the VAE can be combined with other methods in order to comprehensively analyze cellular data in a tissue system, specifically with recently developed techniques focusing on spatially-resolved cell data (cite). Additionally, the VAE is compatible with our previously published supervised learning/decision tree method, STRAINS, and works as a complementary technique for high-throughput analysis of spatiotemporal cell data. Further, samples are preserved after imaging, allowing for post-imaging analysis of gene expression, protein synthesis, cell metabolic activity, etc. By combining our VAE method with these analyses, a fuller picture of tissue-scale behaviors can be created.

3.5 Methods

The method detailed in this manuscript focuses on the VAE. Data acquisition methods for this dataset in cartilage can be found in our previous work [72]. However, relevant descriptions for the experimental procedures are included here.

3.5.1 Impacting articular cartilage tissue

Sterilely-dissected 6mm cylindrical explants of articular cartilage were cultured and stained for calcium concentration (Calbryte 520AM), mitochondrial polarity (tetramethylrhodamine, TMRM), and nuclear membrane permeability (Sytox Blue). Selected stains were chosen to reflect relevant parameters for this study, but can be readily modified. Samples were bisected and attached to the back plate of a confocal-mounted impactor while submerged in a bath of PBS (Dulbecco's Phosphate Buffered Saline) and Sytox Blue stain (to capture cells dying over the course of the experiment). One half of the sample was impacted, and the other half served as control. The impactor delivers an energy-controlled impact using a spring-loaded piston, producing a peak stress of ~ 1 MPa over 5-10ms and replicating a superphysiologic loading rate which is known to induce tissue damage in cartilage.

The impactor was mounted to the stage of an inverted spinning disk confocal microscope (3i Marianas) for the course of the experiment. Using a 10x objective, the region of impact, the regions lateral to and below the impact, and two regions of corresponding depth on the control sample were imaged. Each region was $660\mu\text{m} \times 660\mu\text{m}$ (512×512 pixels) in size. During the impact itself and the following one minute, the region of impact is imaged at approximately 40 frames per second. Longer term imaging was continuously conducted at all sites, with ~ 12 seconds between frames. The imaging depth was selected to be approximately $30\mu\text{m}$ below the cut surface of cartilage in order to avoid imaging chondrocytes damaged during sample handling.

3.5.2 Extracting time series data of chondrocyte behavior

Cells in the images are tracked using a modified version of the Crocker and Grier algorithm [43]. As described in previous work, the color channels are summed and linear interpolation and static extrapolation are used to obtain cell centroid locations [72]. These centroid values are then used to obtain a small sub-cellular surrounding region over which fluorescence values are averaged. This procedure is used to produce fluorescence traces for all three channels for each cell. Additional moving-average smoothing is used to reduce high-frequency noise, with a window size small enough to not affect features within the timeseries.

After tracking is complete, additional background subtraction is implemented. During and after cartilage impact, cell-localized stains leak into the extracellular matrix and eventually dissipate. However, this phenomenon locally increases the background intensity for some frames of the video, causing a non-uniform background both spatially and temporally. This background is removed by generating a grid of 8x8 subsets within each image, and subtracting the mean of the twenty lowest non-zero pixel values within each subset. The timeseries data for each cell is then plotted in its own image for ease of analysis and organized into dataframes for use in the VAE.

3.5.3 VAE Structure

The variational auto-encoder (VAE) architecture employed in our analysis is as illustrated in Fig. 3.1d. Specifically, we parameterized the encoder as a stack of 1D convolution layers, whose operation derives interesting features from

shorter data segments. The derived features are then mapped into latent means and latent diagonal co-variance vectors that describe a d -dimensional normal distribution, $q_\phi(z|x) = \mathcal{N}(\mu_\phi(x), \sigma_\phi(x))$. In the VAE framework, $q_\phi(z|x)$ acts as an approximation for the true posterior probability $p_\theta(z|x)$. The first portion of the encoder consists of three 1D convolution layers with kernel of size 3, stride of size 2 and padding of size 1. The input and output channels are i) $3 \rightarrow 4$, ii) $4 \rightarrow 8$ and iii) $8 \rightarrow 16$ respectively. This is followed by a reshape operation that turns the transformed data into one dimensional array that is subsequently fed into fully connected layers which are used to compute the mean and diagonal co-variance vectors of size d . The input and output channels are i) $1488 \rightarrow 128$, ii) $256 \rightarrow 128$ and iii) $128 \rightarrow d$ respectively. A sample z is then drawn from the d -dimensional Gaussian and sent through the decoder to obtain a reconstructed intensity profile. The decoder has the same architecture as the encoder, with the exception that the 1D convolution layers are replaced with 1D transposed convolution layers. Batch normalization is applied on the convolution and transposed convolution layers. Non-linearity is introduced with RELU activation function on each layer.

One can then use drawn samples from the encoder distribution to reconstruct the data via the decoder, whose layers are the inverse operations of the encoder. In training the VAE, one wishes to maximize the log-likelihood of generating real data $\log p_\theta(x)$, and at the same time minimizing the information loss when the encoder distribution $q_\phi(z|x)$ is used to approximate the true posterior distribution, $p_\theta(z|x)$. The information loss can be quantified via the KL divergence, $D_{\text{KL}}(q_\phi(z|x)||p_\theta(z|x))$.

$$\mathcal{L}_{\text{VAE}} = \log p_\theta(x) - D_{\text{KL}}(q_\phi(z|x)||p_\theta(z|x)). \quad (3.1)$$

In practice, this optimization is achieved by minimizing the evidence lower

bound objective (ELBO),

$$\mathcal{L}_{\text{ELBO}} = -\mathbb{E}_{z \sim q_{\phi}(z|x)} \log p_{\theta}(x|z) + D_{\text{KL}}(q_{\phi}(z|x) || p_{\theta}(z)) \quad (3.2)$$

which consists of two terms: 1) expected log-likelihood of the decoder distribution, which minimizes the prediction error of the reconstruction, and 2) a regularization term that seeks to minimize the difference between the encoder distribution $q_{\phi}(z|x)$ and the prior distribution, $p(z)$. Here, we assume the prior distribution over the latent features z to be unit Gaussian, $\mathcal{N}(0, \mathbb{I})$.

3.5.4 Training details

To create training and test dataset for our VAE, we first smooth the cell intensity profile with a window size of 10 and down-sampled the time series to a length of 750. We then randomly selected 80% of the data as our training set and the remaining 20% as our validation set.

The VAE is implemented using PyTorch (v1.1.0) [83], where we set the latent dimension d to be 32 and applied batch normalization on the convolution and transposed convolution layers. The nonlinear RELU activation function is acted upon each layer. We train the VAE for a maximum of 100 epochs with a batch size of 32 using the Adam optimizer with a learning rate of 0.001. In order to make the network learn robust representations [84, 85], we also injected a unit Gaussian noise $\mathcal{N}(0,0.01)$ to the cell data in our training process.

3.5.5 Principal Component Analysis

Beyond simply clustering, we can also learn how cellular responses are encoded within the latent representation. We performed principal component analysis (PCA) on the latent mean vectors $\vec{\mu}_z$ generated via the VAE encoder. PCA reduces the dimensionality of our data by projecting the data onto a new set of axes, with each subsequent axis capturing less of the variation. This allows us to determine the first few orthonormal axes \vec{e}_i within the latent space that captures the most variation of the cell's latent mean vectors, i.e. $\vec{\mu}_z \approx \alpha_1 \vec{e}_1 + \alpha_2 \vec{e}_2 + \dots$, where $\alpha_i = \langle \vec{z}, \vec{e}_i \rangle$ is the dot product between $\vec{\mu}_z$ and \vec{e}_i . We can then learn about what cellular profile is encoded within each \vec{e}_i . This can be done by sampling along \vec{e}_i by tuning α_i and keeping the rest of α_j s zero. The cellular profile encoded is recovered by feeding these sampled latent representations through the decoder.

3.5.6 Hierarchical Clustering

To cluster the cell data in an unsupervised manner, we made use of agglomerative hierarchical clustering technique. This technique works by treating each cell data as an individual cluster. At each iteration, similar clusters are then merged until a predefined clusters are formed. We performed the clustering using the latent representation of cell data obtained from the VAE. As the clustering technique requires a metric to quantify the differences between latent representation, we use symmetrized Kullback-Liebler divergence (sKL), a distance measure that quantifies the dissimilarity between two distributions. This choice is motivated from the fact that the latent representation consists of a pair of 32 di-

mensional vectors (mean and diagonal variance) that describe a 32-dimensional Gaussian distribution.

CHAPTER 4

CONCLUSION

4.1 Conclusions

The two methods outlined in this thesis provide a new way of studying mechanotransduction in articular cartilage. By incorporating big data analysis with experimental techniques, we enable future experiments to study changes to chondrocyte phenotypes under different experimental conditions with new stimuli.

In Chapter 2, we demonstrate that the specific behaviors of articular chondrocytes can be captured using fast confocal microscopy for the milliseconds and hours timescales after impact. Cellular fluorescence from these images can be tracked, linked between immediate and long term, and effectively categorized based on their specific behavior patterns. The shape and interaction between three tracked values provide enough information to be classified using a manually-defined decision tree algorithm and train a number of time series classifiers via supervised learning. This technique enables the processing of thousands of cells within an experiment, opening up new possibilities for understanding both individual and coordinated cellular responses to strain.

In addition, we show that the chondrocyte timeseries data obtained via STRAINS can be analyzed using an unsupervised learning technique in Chapter 3. The VAE is a probabilistic neural network that has been applied across many fields. This work demonstrates that a VAE is capable of accurately reconstructing cell behavior data with multiple channels and that the learning process of the VAE generates principal components of the latent space with meaningful

information. Additionally, this is the first application of such a neural network for clustering chondrocyte timeseries data to quickly generate clusters of cell behaviors which are comparable to and surpass manual classification. The VAE method both

Together, these two methods provide the framework for a wide variety of studies probing chondrocyte response to stimuli. The multiple classification and clustering methods enable the analysis of thousands of cells per experiment and streamline the identification of behaviors. The phenotypes identified through these experiments raise a number of questions about chondrocyte responses and the relationship between specific timeseries events, generating a wide variety of new hypotheses to test.

4.2 Future Directions

Based on the methods outlined in this thesis, a number of future projects can be envisioned.

4.2.1 Mitoprotective Therapy

Previous work has shown that mitochondrial depolarization after impact leads to cell death and that mitoprotective compounds like SS-31 can mitigate both mitochondrial depolarization and cell death [6]. SS-31 is a small peptide that acts to stabilize the inner mitochondrial membrane and prevent apoptosis through inhibition of caspase-3. By targeting the cardiolipin/cytochrome C complex in the inner mitochondrial membrane, both membrane curvature and

membrane surface area can be preserved, preventing cytochrome C release and apoptosis. Additionally, SS-31 has been shown to improve mitochondrial bioenergetics and reduce the generation of reactive oxygen species [86, 87, 88, 89].

While the processes behind SS-31's mechanism of action are well defined, how the distribution of cell phenotypes as defined by STRAINS and clustered by the VAE changes when SS-31 is introduced into the system is not yet known. Mitochondria are known to buffer Ca^{2+} in the tissue during injury. SS-31 affects cell signaling and reduces cell death. By testing SS-31 mitoprotection using STRAINS, we will be able to further understand how cell responses after injury are coordinated and mitigated through changes in the spatiotemporal distribution of phenotypes.

4.2.2 Changing Impact Magnitude

Articular cartilage has depth-dependent heterogeneity, where the superficial zone has much lower moduli in both shear and compression. This layer is known to play a protective sacrificial role and significantly alters impact-induced strain fields. Current experiments have also identified cells in the middle zone of tissue that experience much lower strain from impact but show similar phenotypes of cell death to those at the surface. Impacting samples with different peak strain magnitudes while analyzing changes to phenotypes with STRAINS or the VAE would enable the decoupling of zonal effects and strain-induced effects.

Additionally, preliminary evidence has shown that chondrocytes respond differently to compression and shear. Experiments performed in the above

chapters show that cells subject to high shear strains do not die at the same rate as cells experiencing compression, despite similar overall strain magnitude. Altered peak strain from impact or alterations to impact geometry could be used to further probe cellular response to shear strain.

4.2.3 Cell-to-cell Signaling and Gap Junctions

The ability to map cell behaviors onto specific locations allows us explicitly probe interactions between spatially adjacent cells. Preliminary use of the GUI and additional analyses have looked at the affect of calcium signaling in one cell on the likelihood of calcium signaling or cell death in neighboring cells. While the results proved inconclusive, the method could be combined with various channel activators or inhibitors to probe signaling between neighboring cells.

Specifically, chondrocytes have been demonstrated to communicate through intercellular connections—gap junctions[90, 91]. Chondrocytes were shown to exchange nutrients to maintain homeostasis. It would be interesting to probe these connections in response to impact.

4.2.4 Combining STRAINS and VAE with other techniques

As outlined in the conclusions of both chapters, the experimental techniques demonstrated in both methods is highly compatible with other analysis techniques. Confocal microscopy of the cartilage tissue is non-destructive and the samples can be preserved post-experiment. This opens up the possibility of combining our methods with techniques to measure gene expression, protein

synthesis, cell metabolics, and others.

4.2.5 Using Automated Algorithms to Improve Decision Tree Structure

Many different methods of setting up a decision tree exist. This thesis uses a very traditional set of if/else statements based on specific time series features that I determined to be relevant. It is possible that an unsupervised technique to develop a decision tree will improve accuracy.

4.2.6 TRPV4 and Piezo 1/2

One of the most direct next steps using this system is to probe changes in calcium signaling when blocking or activating specific ion channels. TRPV4 has been demonstrated to be a mechanotransducer of physiologic strain in cartilage [15, 92, 40]. Preliminary experiments blocking TRPV4 with HC-067047 have shown that TRPV4 inhibition may be affecting the strain threshold of an impact-induced calcium signaling wave. Lower strain impacts with TRPV4 inhibition demonstrate a significantly slowed initial spreading wave of calcium signaling with almost no subsequent cell death. Activating TRPV4 with GSK101 may be beneficial when affecting chemical signaling, but preliminary experiments with similar concentration [15] show almost full mitochondrial depolarization before impact and increased cell death. Further experiments tuning the concentrations of agonists and antagonists will be useful to look more specifically into how strain affects TRPV4 signaling.

Piezo 1/2 have been demonstrated to be mechanotransducers of hyperphysiologic strain and inhibition of these channels has been demonstrated to preserve cell viability [63, 37, 36]. Activating or inhibiting these channels with the STRAINS method would give more detailed maps of potential strain thresholds for cell death and provide a potentially interesting look into how cell behavior categories change.

APPENDIX A
IMAGING CHONDROCYTES IN CARTILAGE IMPACT

A.1 Introduction

The confocal-mounted impactor used throughout these experiments was developed by Bartell, Delco, Bonassar, and Cohen [5]. This impactor is uniquely capable of enabling real-time imaging of cartilage during the impact itself. When used in conjunction with a high speed camera, the impactor can be used to map strain fields in cartilage during an impact.

The impactor has been used to demonstrate that the threshold of cell death is 8% microscale strain norm [5] and that SS-31, a mitoprotective compound, can prevent mitochondrial depolarization and apoptosis after impact injury [6]. Designed as an instrument to probe chondrocyte behavior, the impacting tip is a 0.8mm diameter rod, which impacts to induce a peak stress of ~ 1 MPa over 5 to 10 ms. This geometry produces a wide range of strains which can be observed within the field of view of a 10x objective, but does not mimic injury-like hyperphysiologic strains. To observe the full strain field during impact, a 5x objective should be used.

A.2 Experiments

A.2.1 Dissection and Culture

Standard sterile dissection practices should be used for these experiments. However, extra care must be taken to ensure that the cartilage tissue never dries during any part of the process. Once the joint capsule is opened, if the joint itself has plenty of synovial fluid, then a sterile transfer pipette can be used to coat the cartilage in its own synovial fluid. More often than not, the joint will have very little synovial fluid, and PBS should be used instead. Always err on the side of caution.

The dissection of plugs from the cartilage should be done on the dissection bench for these experiments—placing cut cartilage into a conical can cause avoidable cell death at the surface. When cutting, ensure that there is synovial fluid on the biopsy punch. Additionally, if the tissue is thick, make sure that the plastic end of the biopsy punch *never* touches the surface of the tissue—i.e. do not push the biopsy punch to its full extent, or the cells at the surface will be dead. Use fresh scalpels to cut around the biopsy punches. Use the scalpel blade to push the punch up out of the tissue, then gently nudge the side of the punch with the blade to fully push it out enough to grab with flat forceps. Place the plugs in a conical with PBS (with 1% penicillin/streptomycin). Use extra PBS in the conicals to store explants, so that they do not fall as far when dropped into the conical.

Transfer everything into the biosafety cabinet. Set up two petri dishes with PBS, and gently swish the explants or use a sterile transfer pipette to gently

squeeze PBS at the explants to remove excess synovial fluid from cutting. Use a 12 well culture plate with 2.5mL media (see below). The plugs should be placed gently into the dish with the articular surface up (the cut surface touching the plate). There should be enough media to cover the explants. If there is not, use a razor to cut off part of the bottom surface. If there is excess blood in the tissue, let them sit in the PBS to gently rinse for ~ 5 min. Following these protocols, the surface death of chondrocytes should be incredibly minimal.

These experiments use extra low glucose media (0.45g/L). This was determined previously (cite lena/michelle) to allow for changes in mitochondrial polarity to be observed. However, the low glucose concentration can lead to cells in the center of the explant dying after a few days, or having very low signal while on the microscope. If necessary, increase glucose as needed. There cannot be any phenol in the media, since it may interfere with imaging.

Per 100mL of (no glucose, no glutamate, no pheno, no pyruvate DMEM), use 0.045g glucose, 2mL Sodium Pyruvate, 2.5mL HEPES, 1mL FBS, 1mL Penicillin/Streptomycin, 1mL L-Glutamine, 1mL Amino Acid Master Mix. The Amino Acid Master Mix is made by combining 500mg Ascorbic acid and 300mg α -Ketoglutaric acid with 100mL L-Glutamine, which can then be aliquoted and frozen at -20°C .

A.2.2 Staining

Cartilage is stained with Calbryte 520 AM for calcium concentration, Tetramethylrhodamine (TMRM) for mitochondrial polarity, and Sytox Blue for nuclear membrane permeability (stains DNA). Staining must occur immediately before

imaging, in the incubator. Chondrocytes take up much more of the stain while incubating, and this also minimizes the amount of time that the tissue is a room temperature prior to an experiment. Room temperature cartilage tends towards mitochondrial depolarization and reduced cell activity (including calcium transients). Prior to use, individual aliquots of stain are produced to minimize freeze-thaw cycles which will affect stain viability. The same media used for culture is also used for staining.

Cartilage samples should be cut with the cutting cube as guide and using extra-sharp stainless steel razors. During the cutting process, the cube should be wetted with PBS and the blade and cartilage coated with synovial fluid. The cut samples should be rinsed in PBS prior to staining—synovial fluid prevents the stain from diffusing into the tissue correctly. Always stain at least two well-cut samples, since there are many stages of the experiment where things can go wrong (cartilage glued incorrectly, cut surface not flat enough, chondrocytes in explant are dead before impact, etc).

Calbryte 520AM should be stained first, since it is partitioned into the chondrocytes and the excess is washed away. TMRM and Sytox Blue are stained simultaneously afterwards. Any additional probes (i.e. TRPV4 inhibitor HC-067047) should be incubated for one hour in media after all stains. Each staining or agonist/antagonist step takes one hour. Store in a warmed 50mL conical of media for transport. At least 30mL of Sytox Blue in PBS should also be prepared for the impactor well.

Selecting Stains

When choosing stains for chondrocytes, several things should be considered: the ability of the stain to partition into cells and then be washed away (Calbryte AM), whether or not there will be overlapping emission/excitation wavelengths, and compatibility with available lasers. Most importantly, stains should not affect other cellular functions being measured.

The concentration of staining may need to be increased when adjusting for explants. Too much of a stain like TMRM, however, can affect mitochondrial function. So adjust as needed, carefully. Additional methods for improving staining include: ensuring that the imaging surface is facing up in the culture dish, increasing staining time, staining in the incubator, and combining with compounds that facilitate stain uptake (this will be unique to specific stains).

Most cell stains have a stated washing time—this ensures that excess stain is removed from the tissue before imaging and reduces background intensity. The washing time can be used for other staining. Finally, some stains, agonists, and antagonists should be included in the impactor well while imaging to allow for uptake during the experiment.

A.2.3 Using the Impactor

Fill impactor well with Sytox Blue solution. Set spring spacing and prime impactor tip. Gently remove both halves of the plug from the conical and set onto a glass slide that has plenty of PBS on it. Dry the deep zone surface of the plugs with a kimwipe. Put glue onto the impactor backplate and use the tweezers or

spatula to push the plugs into the glue. Keep the plugs as flat as possible. Set glue using Zip-Kicker pipetted on the entire sample, and rinse with PBS. Line up impact tip with the center of one of the plugs. Screw backplate onto impactor, making sure that backplate is aligned so that the force from the impact will not move it (the slots are wider than the screw posts).

Remove all parts—impactor tip, back plate, back plate base, and any screws. Use razor blade to scrape sample and glue off of the back plate. Clean with soap/water, making sure to remove all traces of PBS (to prevent corrosion). Use a transfer pipette to suck out as much of the PBS as possible, then use a kimwipe to wipe out the inside of the well until dry. Use the pipette to transfer water inside the well to rinse, then remove and wipe dry again. If using compounds that may transfer easily, spray the inside of the well with isopropanol and wipe dry. Place impactor back in box, on top of the foam and close the top of the box.

A.2.4 Microscopes

The 3i confocal must first be enabled through the CCMR FOM system. To properly turn on the microscope, make sure that the computer is on (this is not controlled through FOM), the laserstack is turned on and cycle the shutter key—all keys should be in the ON position. Open Slidebook (no piezo) and load jz848.3color imaging parameters. Find the impact tip and move to the sample. Offset the sample slightly so that the impact can be captured clearly. The impact tip should be near the bottom of the screen. Cover the impactor well with the lid of the cryopro box to block light. Set location 1 to be centered on the impact, and the surrounding locations as 2-4. Set locations 5-6 to be on the control sample,

one at the surface and one adjoining. Set any additional locations as needed. During impact, image 1000 frames only of the green calcium signal at 25ms (or less?) exposure and 1000 intensification. Set other exposures as needed. Set wait interval as needed. Turn off the light in the room. When done, save the slidebook file and upload it to Google Drive (upload speeds on Box are very slow).

Open the slidebook software and load in the correct imaging parameters. When consistently using the 3i for similar experiments, save the parameters needed to speed up setup. Unless otherwise directed, the camera should be at max intensification and the laser power and exposure as low as possible while still giving reasonable images. This reduces photobleaching and light toxicity.

When using the impactor on the 3i microscope, the head of the microscope should be tilted back so that the impactor will fit. Use the 10x objective and ensure that it is in loading position (minimum z). Add the lens extenders before loading the objective onto the microscope. Then, slot the impactor into the microscope stage. When doing this, 3 clicks should be heard, and the impactor should be nudged to ensure that it is fully locked into the stage. Center the objective under the cartilage samples, and move to the correct z focus for imaging. Check that all of the stains are visible, and that there is not excess cell death. If the sample is not viable, switch to the backup sample. During imaging, ensure that the imaging depth should be $\sim 30\mu\text{m}$ into the tissue from the cut surface, to avoid imaging cells that died due to the cutting and handling process. Additionally, a cardboard box (usually the lid of a cryopro microtube box) is placed over the impacting well to block out all of the light.

Using the x-y controls (with the laser ON), find the location of the impact-

ing tip. To ensure that the impact is centered on the cartilage tissue, adjust the impacting tip location to the configuration shown below, and remove the box to allow for better viewing. Once complete, put the box back on the impactor. It should stay on the impactor for the rest of the experiment.

Then, move back over to the cartilage sample and save six locations in a configuration seen in Fig. ??b. When finding locations, z-positioning may need to be adjusted depending on quality of cut surface and gluing. Additional depths may be used, depending on the age of the tissue and the presence of blood vessels. Ensure that the edge of the surface is close to the edge of the image. Then, used saved parameters to take 10 frames of each location. This is the pre-imaging control. For active samples with lots of calcium signaling, this should capture enough baseline signaling to allow for calibration.

The impact is capture using the 'stream to disk' feature in the focus window of slidebook. Impact videos capture only the calcium signal (green) channel. The lowest reasonable framerate for an experiment like this is roughly 40fps, with an exposure time of 25ms. The laser power should be increased to compensate, but not too much as to saturate the video upon impact. Roughly 2000 frames should be just over 1 minute. Before triggering the capture, ensure that the image is centered on the impact site, and that the camera is on and the shutter open. Start the video before impacting, and make sure that it is recording. Once the impact video is complete, quickly swap to pre-saved parameters in the capture window and start the post-impact capture. The time between impact and post-impact videos should be minimized. Capture the post-impact videos for 2-3 hours.

The impactor can also be used on the Zeiss LSM 880 (inverted). The usage

process is largely the same, but care must be taken in selecting which stains should be imaged in which channels, since overlapping absorption/emission spectra may affect the final data.

A.2.5 Strain Mapping with the High Speed Camera

Maps of chondrocyte strain can be produced using Ncorr (Matlab digital image correlation function) and videos from a high speed camera. The high speed camera is set up on the COOL confocal and used with the mercury (Hg) arc lamp. The Hg lamp requires special bulbs that must be replaced after 200 hours. The digital display shows how many hours have elapsed, but a written logbook is helpful to prevent any errors. To use the lamp, it must first warm up for at least half an hour. Additionally, keep the lamp on for at least one hour before turning off, and keep the lamp off for at least one hour before turning on. This prevents issues related to sudden temperature fluctuations.

Normal cartilage samples are used (bisected 6mm plugs). The cells do not need to be alive, so no culturing is necessary. To make the sample visible, we use 2 μ m fluorescent microspheres from Polysciences, as detailed in Lena's paper. The microspheres are allowed to dry onto the cut surface of the sample for roughly ten minutes before the sample is loaded in the impactor.

To ensure that the entire strain map of impact is captured, a 5x objective is necessary. Additionally, when focusing on the impact, know that the sample will bulge out of focus during the impact itself, so ensure that the focus is on a z-plane slightly below where the microspheres are in focus. Videos are captured on the group laptop using phantom camera software.

Strain mapping is conducted using Ncorr, a package for 2D digital image correlation in MATLAB that can be downloaded from www.ncorr.com/. The Ncorr manual has detailed instructions on its use. For this system, use the smallest region size that still produces a flat strain surface for calculation.

A.3 Impactor Calibration and Maintenance

The force on the back plate of the impactor is measured with a load cell. A roughly T-shaped piece of aluminum is used to attach the load cell in a configuration that allows it to fit in the impactor.

If the glass plate of the impactor well breaks, it can be fixed using epoxy and a large glass slide. Remove the screws attaching the well to the main body of the impactor, remove the impacting rod section, and soak the entire thing in boiling water to help loosen the epoxy.

The well of the impactor is attached using two-part epoxy. The glass window is a large glass slide. There are several grooves in the metal impactor parts where epoxy should be placed for everything to be attached properly. Do not use too much epoxy, or the parts will not screw on properly. I have found that the faster setting epoxy (in small packets as opposed to the tubes) removes more easily, with no observable downside besides less working time.

The impacting tip must be oiled every once in a while, with 3-in-1 oil (available with the rest of the impactor parts). This is most helpful to lubricate the trigger mechanism. Notably, when pressing the trigger button, do not press all the way in. I have marked a position with black sharpie that allows the impact

to happen with the least friction, which increases experimental consistency.

APPENDIX B

STRAINS AND OTHER SUPPORTING CODE

B.1 Code Breakdown

The setup and analysis codes are broken down into a few components. Initial image setup is accomplished in ImageJ/Fiji. The tracking, data formatting, feature extraction, and decision tree codes are written in Matlab. Timeseries figure generation codes, as well as the STRAINS GUI, are written in Matlab. Time series classification via supervised machine learning and the VAE are realized in python.

B.2 Image Setup

Images saved from the 3i Marianas spinning disk confocal are in the .sld format. Download the program from the website to extract the images to .tiff files. Images from the COOL and HOT (Zeiss LSM 5/7 Live) microscopes are saved as .lsm files which can be converted to .tiff in Fiji.

Images from the i880 are saved as .czi files which can be read by Fiji if the Bio-Formats extension is installed

To speed up image setup, two macros have been written for Fiji (ImageJ): image setup and impact setup. To ensure that they are properly installed in Fiji, they should be added to the startup_macros.iji file in the Fiji installation folder. Once images are exported from the Slidebook file and opened in Fiji, the macros

will output the correct images for use in the subsequent Matlab codes. The save folder for these macros may need to be modified to work with your own file system.

B.3 Tracking

Cell tracking uses the Crocker and Grier particle tracking algorithm in conjunction with supporting code. After image setup is completed, the workflow is as follows:

1. The folder, date of experiment, and positions are input into any of the `***tracking_function_call` codes.
2. The relevant time information is manually input into the tracking function call, based on the information from the output text files when exporting from slidebook.
3. Impact tracking occurs with the parameters as listed in the `***tracking_function_call` codes. This calls the `'TrackImpact'` function.
4. Post-impact tracking occurs with the parameters as listed in the `***tracking_function_call` codes. This calls the `'TrackPostImpact'` function.

`TrackImpact` and `TrackPostImpact` use the Crocker-Grier algorithm, which can be found at <https://site.physics.georgetown.edu/matlab/>. The tracking is run on the sum of all three color channels. Both of these functions will run the tracking and output timeseries data, as well as generate any relevant plots to the tracking process. The tracking algorithm used will produce the centroid of each cell, around which a 3x3 grid of pixels is averaged to produce

the "intensity" of that cell at that time point. Additional background subtraction is performed, where the 512x512 image is broken up into an 8x8 grid. The lowest 20 (this should be optimized to the image) nonzero pixel values within each subsection are averaged to find the background within this region, which is then subtracted out of this subsection.

Tracking at the impact site can occur one of two ways: the impact video can be concatenated with the post-impact video at position one (often called pos1_full) and the entire video tracked, or the impact and pos1 videos are tracked separately and linked afterwards. There are drawbacks to either method. Tracking with the longer video often means that more cells are lost and not tracked properly. This is especially true if the post-impact video contains movement. Conversely, linking the two videos loses all of the cells that are not tracked in both videos, and can lead to some issues where the impact data of one cell is linked to the post-impact data of another cell. The best course of action is to usually run both techniques and see which one is better. Always check the output tracking images to make sure that the tracking is reasonable. The linking function works by finding cells that are within a certain set Euclidean distance between the last impact frame and the first post-impact frame.

Once the cells are tracked and intensities extracted, the timeseries plots for each cell are saved into the folder containing the cell images. The plots are often generated using a moving average filter to remove high frequency noise (for plotting purposes only). This window size can be tuned, but should be somewhere between 5 and 20 frames, in order to reduce noise but keep relevant signals. These timeseries plots should be examined to ensure that background subtraction was reasonable and that there are not a lot of obvious "jumps" in the

data that are a sign of bad tracking. The timeseries plots are used for manual labeling of categories.

B.4 STRAINS GUI

The STRAINS GUI is programmed in the Matlab App Designer. There are two versions, one for pre-Matlab 2019 and earlier, and one for Matlab 2019 and later.

For the pre-Matlab 2019 the `UIFigureWindowButtonDown` function is used to record clicks. This is a workaround because `ginput` and other click detection mechanisms do not work. The Matlab 2019 and later version of the GUI uses `drawpoint` to record clicks. When possible, use the newer version, which is much more robust and is more intuitive to use.

B.4.1 GUI Operation

The GUI front panel consists of three main regions: controls, the video player, and plots. This is the user interface, where the plots and image display area are fully interactable.

The back end of the GUI is composed of two main components: functions and callbacks. Here, functions are snippets of code that we expect to use repeatedly and are defined in the methods section, which are mostly for plotting. Callbacks are the code segments that are triggered when a user interacts with something in the interface (i.e. push a button, change a number, move a slider, etc). The individual functions and callbacks are all commented with specific de-

tails about how they work. Additionally, there is a startup function that runs automatically when the GUI is opened. This controls how the GUI looks at startup, and cannot be modified. Each function and callback is labeled with its specific purpose, and the MATLAB appdesigner debugger can be used to iterate through how each line of code works. Finally, properties that are used throughout the GUI are initialized at the top.

The GUI has several main features:

1. Loading and playing a cell video
2. Plotting an individual cell from the video (the GUI always finds the closest cell)
 - (a) By clicking on a cell in the video
 - (b) By entering coordinates of a cell
 - (c) By entering the Cell ID (from particle tracking)
3. Plotting an individual cell from the video, including the impact curve (this requires the user to upload a specific dataset that has both the impact and post-impact timeseries from each cell)
4. Plotting all of the cells within a selected rectangular region (with mean in bold, color channels split)
 - (a) A specific timeseries can be clicked on in the plot to circle the cell on the video and bold the selected timeseries on the plot
 - (b) A specific time can be selected on the plots to produce a histogram of all three colors

(c) If a time has been selected, a mixed-model Gaussian can be fit on the blue channel to split the data into low, medium, and high blue signals at that timepoint.

5. Cell IDs of selected cells can be saved for future reference

B.5 Feature Extraction

The feature extraction method uses various MATLAB functions to identify important shapelets in the timeseries. The current iteration finds peaks and changepoints in the green and blue channels. Peak detection uses MATLAB's `findpeaks` function to record peak intensity, time, width, prominence, and the width-to-prominence ratio (manually calculated to help with later filtering). The detection method uses minimum peak prominence (above baseline). Changepoints are locations in the timeseries where the derivative is high. These are detected using MATLAB's `findchangepoints` function, with linear thresholding as the method. Changepoint location is recorded. The difference in slopes before and after a changepoint are calculated manually, with slope differences less than a set amount eliminated. This allows the changepoint detection to find small changepoints, but removes ones where the slope is changing very gradually.

Peak detection is used to find calcium transients in the green channel and cells where the nuclear membrane permeability rises and falls in the blue channel. Changepoint detection is used to find places where the calcium concentration rises or falls suddenly in the green channel, and where the nuclear membrane permeability rises suddenly in the blue channel. The functions include

an optional parameter to plot detected peaks and changepoints for manual validation. To speed up normal function, plots should be turned off. All detected peaks and changepoints with their respective statistics are recorded in structures.

These structures are then incorporated into one large structure containing all of the features of each cell through the CellAttributes function. This function reads in all of the peak and changepoint data alongside the cell number and location as well as calculating statistics including: red and blue ranges, red and blue means, red and blue maximum values (along with maximum value frame), red means for the last 100 frames, and the difference between red and blue means. This final cell attributes structure is then passed to the decision tree.

B.6 Decision Tree

The decision tree makes use of peaks, changepoints, and additional timeseries statistics to decide in which category a cell should belong. These categories are established by previous manual labeling. The structure of the decision tree itself was optimized through multiple iterations in order to determine which cutoffs were more important, which features the system was better at identifying, the ordering of questions to ask, etc. The decision tree itself is a complex series of if/else statements. A full diagram of the tree can be found in Chap. 2. The decision tree itself consists of several branches of if/else statements as it iterates through the cell attributes structure in the order of the index. Each statement is commented with the output that it would produce. Individual parameters

are hard-coded into the tree itself, but can be modified easily. Additional code at the end of the DecisionTree function checks accuracy compared to manually sorted labels.

B.7 Time Series Classification

As noted in Chap. 2, the basis of the time series classification uses the sktime library in python. There are multiple models used in this system: two that are interval-based (CIF and DrCIF) and two that are kernel-based (ROCKET and Arsenal). Interval-based methods are generally more accurate, but kernel-based methods are generally faster. Only multivariate methods are used, where each channel corresponds to one dimension. The models highlighted here are the ones that produced the best accuracy when testing on the hand-sorted dataset. When using these models for repeated experiments, use all of the classifiers (and possible new classifiers) and allow them to vote on the correct category to get a final label. Sktime has recently updated with newer classifiers, including a convolutional neural network that works with multivariate time series, that may be relevant to include.

Univariate classifiers can also be used, where the three color channels are input as one timeseries. Generally, the accuracy of univariate classifiers is lower, but a much wider array of univariate classifiers are available. When reformatting data for univariate classification, concatenating the data (i.e. red, then green, then blue) produced the best accuracy.

The code is split into two sections: the functions and the function calls. Classifier_functions.py includes all of the functions for loading and reformatting

data from MATLAB, training classifier models, and loading classifier models. The training functions include splitting data into training and testing sets with a random seed. To use the functions, implement some code similar to what can be found in `all_classifier_function_calls.py`, which will set up proper directories for storing the classified data, dump and load for the models, and inputting parameters for each classifier model.

When using these classifiers, split the data into two sets: one that includes calcium transients detected in MATLAB with peak detection, and one that does not include calcium transients. The classifiers do not work well with randomly occurring transients. The two sets can be combined together at the end after classification.

The parameters for these models can be adjusted. The specific parameters accepted in the current training functions were the ones found to impact the training accuracy the most, but training functions can easily be rewritten to include additional parameters. These parameters can be found in the `sktime` documentation. I have found that the `sktime` github is easier to access and more up-to-date than the website sometimes, but `sktime.net` has some good example implementations. Python library dependencies are noted in the functions.

B.8 VAE Clustering

There are several MATLAB codes used to restructure data for use with the Variational Autoencoder. These codes also process the cell location into a global location for the experiment, instead of local pixel location within a single imaging position. Locations and timeseries for all cells in order of position and cell

ID are saved to a .mat file, which is then imported in python. The rest of the VAE code is implemented in python, consisting of the cell_script, which imports all of the data and calls the VAE functions, the loaders, the wrapper, the encoder, and the decoder.

BIBLIOGRAPHY

- [1] A. J. Sophia Fox, A. Bedi and S. A. Rodeo, "The Basic Science of Articular Cartilage", *Sports Health* **1** (Nov., 2009) 461–468.
- [2] M. M. Knight, Z. Bomzon, E. Kimmel, A. M. Sharma, D. A. Lee and D. L. Bader, "Chondrocyte Deformation Induces Mitochondrial Distortion and Heterogeneous Intracellular Strain Fields", *Biomechanics and Modeling in Mechanobiology* **5** (Mar., 2006) 180.
- [3] M. R. Buckley, J. P. Gleghorn, L. J. Bonassar and I. Cohen, "Mapping the depth dependence of shear properties in articular cartilage", *Journal of Biomechanics* **41** (Aug., 2008) 2430–2437.
- [4] J. L. Silverberg, S. Dillavou, L. Bonassar and I. Cohen, "Anatomic variation of depth-dependent mechanical properties in neonatal bovine articular cartilage", *Journal of Orthopaedic Research* **31** (May, 2013) 686–691.
- [5] L. R. Bartell, L. A. Fortier, L. J. Bonassar and I. Cohen, "Measuring microscale strain fields in articular cartilage during rapid impact reveals thresholds for chondrocyte death and a protective role for the superficial layer", *Journal of Biomechanics* **48** (Sept., 2015) 3440–3446.
- [6] L. R. Bartell, L. A. Fortier, L. J. Bonassar, H. H. Szeto, I. Cohen and M. L. Delco, "Mitoprotective therapy prevents rapid, strain-dependent mitochondrial dysfunction after articular cartilage injury", *Journal of Orthopaedic Research* **38** (2020) 1257–1267.
- [7] D. E. Ingber, "Cellular mechanotransduction: putting all the pieces together again", *The FASEB Journal* **20** (2006) 811–827.
- [8] J. H. C. Wang and B. P. Thampatty, "An Introductory Review of Cell Mechanobiology", *Biomechanics and Modeling in Mechanobiology* **5** (Mar., 2006) 1–16.
- [9] G. Perbal and D. Driss-Ecole, "Mechanotransduction in gravisensing cells", *Trends in Plant Science* **8** (Oct., 2003) 498–504.
- [10] L. Chin, Y. Xia, D. E. Discher and P. A. Janmey, "Mechanotransduction in cancer", *Current Opinion in Chemical Engineering* **11** (Feb., 2016) 77–84.

- [11] J. H. C. Wang, "Mechanobiology of tendon", *Journal of Biomechanics* **39** (Jan., 2006) 1563–1582.
- [12] D. R. Carter, G. S. Beaupré, N. J. Giori and J. A. Helms, "Mechanobiology of Skeletal Regeneration.", *Clinical Orthopaedics and Related Research (1976-2007)* **355** (Oct., 1998) S41.
- [13] M. L. Delco and L. J. Bonassar, "Targeting calcium-related mechanotransduction in early OA", *Nature Reviews Rheumatology* **17** (Aug., 2021) 445–446.
- [14] T. Iskratsch, H. Wolfenson and M. P. Sheetz, "Appreciating force and shape — the rise of mechanotransduction in cell biology", *Nature Reviews Molecular Cell Biology* **15** (Dec., 2014) 825–833.
- [15] C. J. O’Conor, H. A. Leddy, H. C. Benefield, W. B. Liedtke and F. Guilak, "TRPV4-mediated mechanotransduction regulates the metabolic response of chondrocytes to dynamic loading", *Proceedings of the National Academy of Sciences* **111** (Jan., 2014) 1316–1321.
- [16] N. Q. Balaban, U. S. Schwarz, D. Riveline, P. Goichberg, G. Tzur, I. Sabanay et al., "Force and focal adhesion assembly: a close relationship studied using elastic micropatterned substrates", *Nature Cell Biology* **3** (May, 2001) 466–472.
- [17] J. L. Tan, J. Tien, D. M. Pirone, D. S. Gray, K. Bhadriraju and C. S. Chen, "Cells lying on a bed of microneedles: An approach to isolate mechanical force", *Proceedings of the National Academy of Sciences* **100** (Feb., 2003) 1484–1489.
- [18] T. Oliver, K. Jacobson and M. Dembo, "Traction forces in locomoting cells", *Cell Motility* **31** (1995) 225–240.
- [19] R. Zhao, T. Boudou, W.-G. Wang, C. S. Chen and D. H. Reich, "Decoupling Cell and Matrix Mechanics in Engineered Microtissues Using Magnetically Actuated Microcantilevers", *Advanced Materials* **25** (2013) 1699–1705.
- [20] D. Choquet, D. P. Felsenfeld and M. P. Sheetz, "Extracellular Matrix Rigidity Causes Strengthening of Integrin–Cytoskeleton Linkages", *Cell* **88** (Jan., 1997) 39–48.

- [21] A. J. Engler, S. Sen, H. L. Sweeney and D. E. Discher, "Matrix Elasticity Directs Stem Cell Lineage Specification", *Cell* **126** (Aug., 2006) 677–689.
- [22] J. Kolega, "Effects of mechanical tension on protrusive activity and microfilament and intermediate filament organization in an epidermal epithelium moving in culture.", *Journal of Cell Biology* **102** (Apr., 1986) 1400–1411.
- [23] N. C. Gauthier, M. A. Fardin, P. Roca-Cusachs and M. P. Sheetz, "Temporary increase in plasma membrane tension coordinates the activation of exocytosis and contraction during cell spreading", *Proceedings of the National Academy of Sciences* **108** (Aug., 2011) 14467–14472.
- [24] D. E. Jaalouk and J. Lammerding, "Mechanotransduction gone awry", *Nature Reviews Molecular Cell Biology* **10** (Jan., 2009) 63–73.
- [25] T. Hodgkinson, D. C. Kelly, C. M. Curtin and F. J. O'Brien, "Mechanosignalling in cartilage: an emerging target for the treatment of osteoarthritis", *Nature Reviews Rheumatology* (Dec., 2021) 1–18.
- [26] J. Eyckmans, T. Boudou, X. Yu and C. Chen, "A Hitchhiker's Guide to Mechanobiology", *Developmental Cell* **21** (July, 2011) 35–47.
- [27] L. R. Bartell, M. C. Xu, L. J. Bonassar and I. Cohen, "Local and global measurements show that damage initiation in articular cartilage is inhibited by the surface layer and has significant rate dependence", *Journal of Biomechanics* **72** (Apr., 2018) 63–70.
- [28] M. L. Delco, E. D. Bonnevie, L. J. Bonassar and L. A. Fortier, "Mitochondrial dysfunction is an acute response of articular chondrocytes to mechanical injury", *Journal of Orthopaedic Research* **36** (2018) 739–750.
- [29] M. L. Delco, M. Goodale, J. F. Talts, S. L. Pownder, M. F. Koff, A. D. Miller et al., "Integrin 101-Selected Mesenchymal Stem Cells Mitigate the Progression of Osteoarthritis in an Equine Talar Impact Model", *The American Journal of Sports Medicine* **48** (Mar., 2020) 612–623.
- [30] J. D. Humphries, M. R. Chastney, J. A. Askari and M. J. Humphries, "Signal transduction via integrin adhesion complexes", *Current Opinion in Cell Biology* **56** (Feb., 2019) 14–21.

- [31] R. F. Loeser, "Integrins and chondrocyte–matrix interactions in articular cartilage", *Matrix Biology* **39** (Oct., 2014) 11–16.
- [32] R. Ruhlen and K. Marberry, "The chondrocyte primary cilium", *Osteoarthritis and Cartilage* **22** (Aug., 2014) 1071–1076.
- [33] M. M. Knight, S. R. McGlashan, M. Garcia, C. G. Jensen and C. A. Poole, "Articular chondrocytes express connexin 43 hemichannels and P2 receptors – a putative mechanoreceptor complex involving the primary cilium?", *Journal of Anatomy* **214** (2009) 275–283.
- [34] A. Subramanian, G. Budhiraja and N. Sahu, "Chondrocyte primary cilium is mechanosensitive and responds to low-intensity-ultrasound by altering its length and orientation", *The International Journal of Biochemistry & Cell Biology* **91** (Oct., 2017) 60–64.
- [35] A. K. T. Wann, N. Zuo, C. J. Haycraft, C. G. Jensen, C. A. Poole, S. R. McGlashan et al., "Primary cilia mediate mechanotransduction through control of ATP-induced Ca²⁺ signaling in compressed chondrocytes", *The FASEB Journal* **26** (Jan., 2012) 1663–1671.
- [36] W. Lee, H. A. Leddy, Y. Chen, S. H. Lee, N. A. Zelenski, A. L. McNulty et al., "Synergy between Piezo1 and Piezo2 channels confers high-strain mechanosensitivity to articular cartilage", *Proceedings of the National Academy of Sciences* **111** (Nov., 2014) E5114–E5122.
- [37] B. Coste, J. Mathur, M. Schmidt, T. J. Earley, S. Ranade, M. J. Petrus et al., "Piezo1 and Piezo2 Are Essential Components of Distinct Mechanically Activated Cation Channels", *Science* **330** (Oct., 2010) 55–60.
- [38] W. Lee, R. J. Nims, A. Savadipour, Q. Zhang, H. A. Leddy, F. Liu et al., "Inflammatory signaling sensitizes Piezo1 mechanotransduction in articular chondrocytes as a pathogenic feed-forward mechanism in osteoarthritis", *Proceedings of the National Academy of Sciences* **118** (Mar., 2021) .
- [39] Lv Mengxi, Zhou Yilu, Chen Xingyu, Han Lin, Wang Liyun and Lu X. Lucas, "Calcium signaling of in situ chondrocytes in articular cartilage under compressive loading: Roles of calcium sources and cell membrane ion channels", *Journal of Orthopaedic Research* **36** (Nov., 2017) 730–738.
- [40] M. Rocio Servin-Vences, M. Moroni, G. R. Lewin and K. Poole, "Direct measurement of TRPV4 and PIEZO1 activity reveals multiple

mechanotransduction pathways in chondrocytes”, *eLife* **6** (Jan., 2017) e21074.

- [41] H. D. Welhaven, C. N. McCutchen and R. K. June, “A comparison of shear- and compression-induced mechanotransduction in SW1353 chondrocytes”, preprint, *Bioengineering*, May, 2021. 10.1101/2021.05.25.445657.
- [42] C. R. Henak, L. R. Bartell, I. Cohen and L. J. Bonassar, “Multiscale Strain as a Predictor of Impact-Induced Fissuring in Articular Cartilage”, *Journal of Biomechanical Engineering* **139** (Jan., 2017) .
- [43] J. C. Crocker and D. G. Grier, “Methods of Digital Video Microscopy for Colloidal Studies”, *Journal of Colloid and Interface Science* **179** (Apr., 1996) 298–310.
- [44] M. Löning, A. Bagnall, S. Ganesh and V. Kazakov, “sktime: A Unified Interface for Machine Learning with Time Series”, .
- [45] A. Amidon, “Sktime: a Unified Python Library for Time Series Machine Learning”, Sept., 2020.
- [46] M. Middlehurst, J. Large, M. Flynn, J. Lines, A. Bostrom and A. Bagnall, “HIVE-COTE 2.0: a new meta ensemble for time series classification”, *arXiv:2104.07551 [cs]* (Apr., 2021) .
- [47] A. Dempster, F. Petitjean and G. I. Webb, “ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels”, *Data Mining and Knowledge Discovery* **34** (Sept., 2020) 1454–1495.
- [48] L. Ye and E. Keogh, “Time series shapelets: a new primitive for data mining”, in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '09, (New York, NY, USA), pp. 947–956, Association for Computing Machinery, June, 2009. DOI.
- [49] Y.-J. Kim, L. J. Bonassar and A. J. Grodzinsky, “The role of cartilage streaming potential, fluid flow and pressure in the stimulation of chondrocyte biosynthesis during dynamic compression”, *Journal of Biomechanics* **28** (Sept., 1995) 1055–1066.
- [50] C. E. Keating and D. K. Cullen, “Mechanosensation in traumatic brain injury”, *Neurobiology of Disease* **148** (Jan., 2021) 105210.

- [51] M. A. Hemphill, S. Dauth, C. J. Yu, B. E. Dabiri and K. K. Parker, "Traumatic Brain Injury and the Neuronal Microenvironment: A Potential Role for Neuropathological Mechanotransduction", *Neuron* **85** (Mar., 2015) 1177–1192.
- [52] J. Lammerding, R. D. Kamm and R. T. Lee, "Mechanotransduction in Cardiac Myocytes", *Annals of the New York Academy of Sciences* **1015** (2004) 53–70.
- [53] A. I. Minchinton and I. F. Tannock, "Drug penetration in solid tumours", *Nature Reviews Cancer* **6** (Aug., 2006) 583–592.
- [54] G. Alexandrakis, E. B. Brown, R. T. Tong, T. D. McKee, R. B. Campbell, Y. Boucher et al., "Two-photon fluorescence correlation microscopy reveals the two-phase nature of transport in tumors", *Nature Medicine* **10** (Feb., 2004) 203–207.
- [55] R. Zeira, M. Land, A. Strzalkowski and B. J. Raphael, "Alignment and integration of spatial transcriptomics data", *Nature Methods* **19** (May, 2022) 567–575.
- [56] A. J. Radtke, E. Kandov, B. Lowekamp, E. Speranza, C. J. Chu, A. Gola et al., "IBEX: A versatile multiplex optical imaging approach for deep phenotyping and spatial analysis of cells in complex tissues", *Proceedings of the National Academy of Sciences* **117** (Dec., 2020) 33455–33465.
- [57] M. J. Gerdes, C. J. Sevinsky, A. Sood, S. Adak, M. O. Bello, A. Bordwell et al., "Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue", *Proceedings of the National Academy of Sciences* **110** (July, 2013) 11982–11987.
- [58] J. W. Hickey, E. K. Neumann, A. J. Radtke, J. M. Camarillo, R. T. Beuschel, A. Albanese et al., "Spatial mapping of protein composition and tissue organization: a primer for multiplexed antibody-based imaging", *Nature Methods* **19** (Mar., 2022) 284–295.
- [59] L. Wang, X. Xing, X. Zeng, S. R. Jackson, T. TeSlaa, O. Al-Dalahmah et al., "Spatially resolved isotope tracing reveals tissue metabolic activity", *Nature Methods* **19** (Feb., 2022) 223–230.
- [60] J. B. Choi, I. Youn, L. Cao, H. A. Leddy, C. L. Gilchrist, L. A. Setton et al., "Zonal changes in the three-dimensional morphology of the chondron

under compression: The relationship among cellular, pericellular, and extracellular deformation in articular cartilage”, *Journal of Biomechanics* **40** (Jan., 2007) 2596–2603.

- [61] I. Youn, J. B. Choi, L. Cao, L. A. Setton and F. Guilak, “Zonal variations in the three-dimensional morphology of the chondron measured in situ using confocal microscopy”, *Osteoarthritis and Cartilage* **14** (Sept., 2006) 889–897.
- [62] F. Guilak, A. Ratcliffe and V. C. Mow, “Chondrocyte deformation and local tissue strain in articular cartilage: A confocal microscopy study”, *Journal of Orthopaedic Research* **13** 410–421.
- [63] W. Lee, F. Guilak and W. Liedtke, “Chapter Ten - Role of Piezo Channels in Joint Health and Injury”, in *Current Topics in Membranes* (P. A. Gottlieb, ed.), vol. 79 of *Piezo Channels*, pp. 263–273. Academic Press, Jan., 2017. DOI.
- [64] C. Wen, T. Miura, V. Voleti, K. Yamaguchi, M. Tsutsumi, K. Yamamoto et al., “3DeeCellTracker, a deep learning-based pipeline for segmenting and tracking cells in 3D time lapse images”, *eLife* **10** (Mar., 2021) e59187.
- [65] B. Pinguan-Murphy, M. El-Azzeh, D. L. Bader and M. M. Knight, “Cyclic compression of chondrocytes modulates a purinergic calcium signalling pathway in a strain rate- and frequency-dependent manner”, *Journal of Cellular Physiology* **209** (2006) 389–397.
- [66] K. L. Ellefsen, J. R. Holt, A. C. Chang, J. L. Nourse, J. Arulmoli, A. H. Mekhdjian et al., “Myosin-II mediated traction forces evoke localized Piezo1-dependent Ca²⁺ flickers”, *Communications Biology* **2** (Aug., 2019) 1–13.
- [67] D. G. Spiller, C. D. Wood, D. A. Rand and M. R. H. White, “Measurement of single-cell dynamics”, *Nature* **465** (June, 2010) 736–745.
- [68] R. Trouillon, M. K. Passarelli, J. Wang, M. E. Kurczy and A. G. Ewing, “Chemical Analysis of Single Cells”, *Analytical Chemistry* **85** (Jan., 2013) 522–542.
- [69] S. Lindström and H. Andersson-Svahn, “Overview of single- cell analyses: microdevices and applications”, *Lab on a Chip* **10** (2010) 3363–3372.

- [70] C. Vilanova and M. Porcar, “Are multi-omics enough?”, *Nature Microbiology* **1** (July, 2016) 1–2.
- [71] W. Li, R. N. Germain and M. Y. Gerner, “Multiplex, quantitative cellular analysis in large tissue volumes with clearing-enhanced 3D microscopy (Ce3D)”, *Proceedings of the National Academy of Sciences* **114** (Aug., 2017) E7321–E7330.
- [72] J. Zheng, T. W. Jackson, L. A. Fortier, L. J. Bonassar, M. L. Delco and I. Cohen, “STRAINS: A Big Data Method for Classifying Cellular Response to Stimuli at the Tissue Scale”, June, 2022. 10.1101/2022.06.12.495830.
- [73] D. P. Kingma and M. Welling, “Auto-Encoding Variational Bayes”, *arXiv:1312.6114 [cs, stat]* (May, 2014) .
- [74] I. A. Luchnikov, A. Ryzhov, P.-J. Stas, S. N. Filippov and H. Ouerdane, “Variational Autoencoder Reconstruction of Complex Many-Body Physics”, *Entropy* **21** (Nov., 2019) 1091.
- [75] K. Han, H. Wen, J. Shi, K.-H. Lu, Y. Zhang, D. Fu et al., “Variational autoencoder: An unsupervised model for encoding and decoding fMRI activity in visual cortex”, *NeuroImage* **198** (Sept., 2019) 125–136.
- [76] C. He, S. Zhu, X. Wu, J. Zhou, Y. Chen, X. Qian et al., “Accurate Tumor Subtype Detection with Raman Spectroscopy via Variational Autoencoder and Machine Learning”, *ACS Omega* **7** (Mar., 2022) 10458–10468.
- [77] D. Ai, Y. Wang, X. Li and H. Pan, “Colorectal Cancer Prediction Based on Weighted Gene Co-Expression Network Analysis and Variational Auto-Encoder”, *Biomolecules* **10** (Sept., 2020) 1207.
- [78] R. Wei and A. Mahmood, “Recent Advances in Variational Autoencoders With Representation Learning for Biomedical Informatics: A Survey”, *IEEE Access* **9** (2021) 4939–4956.
- [79] L. Ternes, M. Dane, S. Gross, M. Labrie, G. Mills, J. Gray et al., “A multi-encoder variational autoencoder controls multiple transformational features in single-cell image analysis”, *Communications Biology* **5** (Mar., 2022) 1–10.
- [80] S. Oller-Moreno, K. Kloiber, P. Machart and S. Bonn, “Algorithmic

advances in machine learning for single-cell expression analysis", *Current Opinion in Systems Biology* **25** (Mar., 2021) 27–33.

- [81] J. Goffinet, S. Brudner, R. Mooney and J. Pearson, "Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires", *eLife* **10** (May, 2021) e67855.
- [82] H. K. Teoh, K. N. Quinn, J. Kent-Dobias, C. B. Clement, Q. Xu and J. P. Sethna, "Visualizing probabilistic models in Minkowski space with intensive symmetrized Kullback-Leibler embedding", *Physical Review Research* **2** (Aug., 2020) 033221.
- [83] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito et al., "Automatic differentiation in PyTorch", .
- [84] B. Poole, J. Sohl-Dickstein and S. Ganguli, "Analyzing noise in autoencoders and deep networks", June, 2014. 10.48550/arXiv.1406.1831.
- [85] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio and P.-A. Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion", .
- [86] H. H. Szeto and S. Liu, "Cardiolipin-targeted peptides rejuvenate mitochondrial function, remodel mitochondria, and promote tissue regeneration during aging", *Archives of Biochemistry and Biophysics* **660** (Dec., 2018) 137–148.
- [87] H. H. Szeto, "Cell-permeable, mitochondrial-targeted, peptide antioxidants", *The AAPS Journal* **8** (June, 2006) E277–E283.
- [88] H. H. Szeto, "First-in-class cardiolipin-protective compound as a therapeutic agent to restore mitochondrial bioenergetics", *British Journal of Pharmacology* **171** (2014) 2029–2050.
- [89] H. H. Szeto, "Mitochondria-targeted peptide antioxidants: Novel neuroprotective agents", *The AAPS Journal* **8** (Sept., 2006) E521–E531.
- [90] M. D. Mayan, R. Gago-Fuentes, P. Carpintero-Fernandez, P. Fernandez-Puente, P. Filgueira-Fernandez, N. Goyanes et al., "Articular chondrocyte network mediated by gap junctions: role in metabolic cartilage homeostasis", *Annals of the Rheumatic Diseases* **74** (Jan., 2015) 275–284.

- [91] M. D. Mayan, P. Carpintero-Fernandez, R. Gago-Fuentes, O. Martinez-de Ilarduya, H.-Z. Wang, V. Valiunas et al., "Human Articular Chondrocytes Express Multiple Gap Junction Proteins: Differential Expression of Connexins in Normal and Osteoarthritic Cartilage", *The American Journal of Pathology* **182** (Apr., 2013) 1337–1346.
- [92] C. L. Gilchrist, H. A. Leddy, L. Kaye, N. D. Case, K. E. Rothenberg, D. Little et al., "TRPV4-mediated calcium signaling in mesenchymal stem cells regulates aligned collagen matrix formation and vinculin tension", *Proceedings of the National Academy of Sciences* **116** (Feb., 2019) 1992–1997.