The Genetic and Mechanical Origins of Tumor Heterogeneity

By

Curtis Tucker Schunk

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biomedical Engineering

December 17, 2022

Nashville, Tennessee

Approved:

Cynthia Reinhart-King, Ph.D.

Xin (Maizie) Zhou, Ph.D.

ACKNOWLEDGMENTS

I would like to thank Dr. Reinhart-King for all the guidance and mentoring over the past two years. I came to Vanderbilt University having never worked in a lab, and Dr. Reinhart-King was willing to provide me with research experience that I desperately needed. Dr. Reinhart-King offered insightful ideas that assisted my research and pointed me in the right direction many times. Dr. Reinhart-King allowed me to gain irreplaceable biology knowledge and advised me on many issues that I ran into. Dr. Reinhart-King has been a fantastic mentor and advisor, and I was able to grow a lot in my time at Vanderbilt University due to her help.

I would like to thank Dr. Zhou for all her help and for being a wonderful mentor. I asked about joining Dr. Zhou's lab having no previous bioinformatics experience, and she graciously allowed me to. She allowed me to solve interesting bioinformatic problems and helped me gain skills in a field that I am very interested in. Dr. Zhou has been an invaluable resource during my time at Vanderbilt University and I am very grateful for help and knowledge she has provided me. I would not be near as prepared for my future career without her help.

I would like to thank all the members of both the Reinhart-King Lab and the Maizie Zhou Lab for all their help and support throughout my time at Vanderbilt University. I constantly felt like I was surrounded by people smarter than me, and that has helped my growth astronomically. They were able to provide insight many of the times that I was stuck. They also provided me an outstanding support network at Vanderbilt University, and many friendships that will continue past my time at Vanderbilt University.

I would like to thank my family for all the support that they have provided me over the last two years at Vanderbilt University and my educational journey as a whole. They have always been able to offer advice and provide support in situations where I greatly needed it. No matter the problem I had, they have always given assistance and I would not be where I am without them.

I would finally like to thank my partner Allison for all the support and encouragement that she has provided. She was always willing to pick up my slack whether I was too busy or stressed to work on anything other than my academic work. She has been a large force of encouragement during this journey. I would like to also thank my cat Dakota for always being there and providing a morale boost whenever it was needed.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

The Genetic Origins of Tumor Heterogeneity

1.1 Introduction

Cancer is a disease in which millions of patients are diagnosed yearly in the United States alone [1]. Near 40% of all people will be diagnosed with a form of cancer in their lifetime [1]. On top of this, cancer is also typically very difficult to treat. The heterogeneous nature of tumors is often what makes cancer treatment so challenging. In an age where we are able to produce treatment options for many diseases, sometimes in a one-size-fits-all manner, the treatment of cancer continues to be elusive. With a less complicated disease, cell types can be targeted and eliminated, but with cancer, there are many different cell types and targeting these cells with treatments can pose a large challenge [2]. Because of the heterogeneity of cancer between different patients, different patients must be treated in different ways. Different forms of cancer treatment include localized therapies and systemic therapies [3]. Localized therapies can be forms of surgery, radiation therapy, chemical and heat ablation, and cryotherapy [3]. Systemic therapies can include chemotherapy, hormonal therapy, immunotherapy, and targeted therapy [3]. These treatment options can be used in tandem or alone, all depending on the nature of the cancer in the individual and the recommendation by the patient's oncologist. The difference in cancers root from the fact that cancer can be caused in many different manners, from many different pathways. This not only results in differences between patients, but also differences within the same tumor of one patient. The intratumor heterogeneity in a patient is a reason why so many different treatment options may be required for a single individual.

Tumor heterogeneity is apparent in all aspects of the tumor microenvironment [2]. Due to mutations and unnatural signaling in tumors, there are typically regions of different cell types and varying microenvironment structure. The heterogeneity of the tumor microenvironment, which can include changes in gene expression, tissue stiffness, and vasculature, are associated with metastasis, decreased patient survival, and affected ability to deliver drugs to the tumor [4]. Much of the heterogeneity in cancer is a result of genetic differences in cells across the area of a tumor. Genetic mutations are common in cancer [5]. Mutations of the actual genetic code and mutations

of the epigenomics of cells lead to altered gene expression [2, 5]. When there is altered gene expression, there can be a reduced or increased amount of signaling proteins, receptors, and other important proteins. This can lead to certain cell types behaving in a way that is much different than how they normally would in their native tissue [6]. Genetic mutations can affect the natural ability of cells to protect against rapid replications or can suppress the ability of the immune system to fight the growing tumor [6]. Because of this issue, different methods have been created to better understand the genetic differences across tissue.

Spatial transcriptomics is an exciting new way to look at genetic heterogeneity. Spatial transcriptomics datasets contain gene expression data at many points over the area of a tissue section [7]. To obtain spatial transcriptomics data, different spots across a tissue section are RNA sequenced [7]. Each sequenced spot contains multiple cells in the same spatial location, with the number of cells per spot varying depending on the method used. The spots are typically barcoded with an amino acid sequence which allows later identification to associate the sequenced data with spatial locations [8]. After RNA sequencing is completed, analysis is performed to obtain the gene expression for each spot. Spatial transcriptomics data consists of the gene expression data of various spots across the tissue section, and the coordinate locations of these spots. Spatial transcriptomics is being used to help find the genetic causes of tissue heterogeneity [9]. Tissue heterogeneity is a hallmark of cancer; however, forms of heterogeneity can be also seen in normal tissue as well. Many types of tissue, such as brain tissue, contain some heterogeneity, showing different cell types and cell functions across an area [10]. Because the heterogeneity in brain tissue follows expected patterns, looking at these tissues with spatial transcriptomics can help us test methods that can be used to understand the heterogeneous behavior of tumors [9].

There has been a recent surge of published tools that use spatial transcriptomics to find genetic similarities between spots and assign spots into clusters [11-15]. There are multiple assumptions, based on biology, that many of these tools use to help their ability to accurately assign spots to clusters. Some of the more important assumptions are that cells of similar type are likely to have similar gene expression and are likely to be spatially close together [16]. Using these assumptions along with machine learning techniques, it becomes possible to break a section of tissue into multiple regions, or clusters. After spots are assigned to clusters, analysis tools like differential gene expression analysis and gene ontology (GO) analysis can be used to understand the genetic and biological differences among clusters [17]. For cancer, having the ability to identify

the different groups of cells in a tumor and their characteristics could allow better diagnosis without the need for a histologist, or could be used to help verify the results found by a histologist [18]. Understanding the differences between these different regions could help form better treatment options for patients affected by this disease.

When dealing with large datasets, like ones containing spatial transcriptomics data, there are certain steps that need to be taken to remove less important aspects of the data. Spatial transcriptomics datasets often contain the expression values for 20,000 genes, for each spot. For many datasets, there are thousands of spots. Performing clustering on a dataset with 200,000 or more values can lead to inaccurate results. Dimension reduction is an important step to increase accuracy [19]. A useful form of dimension reduction is removing genes that are not likely to be different among different cell types. A common approach for this is finding genes that are highly variable across the tumor section and using only these genes to run clustering analysis [19]. However, there are multiple issues with this mostly stemming from the logistical problems that arise when performing RNA sequencing. It is not completely uncommon for a cell which RNA sequencing is performed upon to have inaccurate gene expression measured [20]. At a given time cells can have fluctuations in their mRNA, caused by cells being in different parts of the cell cycle [21]. This can lead to highly variable genes representing genes that are likely to have variable measurements of RNA, instead of genes that are different among distinct cell types in the tissue. Even with the known inaccuracies, gene expression data is still useful. However, it is less useful when performing dimension reduction based on its spot variability. There is a need for methods that can better identify genes that are varying in different cell types.

Another issue with clustering tools in spatial transcriptomics is the difficulty to confirm the accuracy of clusters. There are many spatial transcriptomics datasets that are publicly available for download; however, many of these datasets do not have a gold standard. When dealing with the clustering of datasets, a gold standard is the clustering assignment which is accepted to be reliable and accurate [22]. When a tool assigns spots to clusters, the gold standard would be used to test the accuracy of the tool. Many gold standards are created by experts who manually look at expression data and stained images of tissue sections to determine which spots should be grouped together. Because this is such an essential step in validating the ability of a machine learning tool, finding new ways to create gold standards would allow many more datasets to be used for testing purposes.

In this study, we look at different tools that can be used to better understand the heterogeneity of tissue, with a focus on cancer. We look at using the Gini coefficient of genes to improve upon the current standard of dimension reduction to allow tools to assign spots to clusters more accurately. We look at different published spatial transcriptomics clustering tools and compare them against each other to understand which tool can produce the most accurate results. We look at using biomarkers to identify the accuracy of cluster assignments when there is no gold standard. Finally, we look at using differential gene expression analysis and GO analysis to better understand the genetic and biological differences between clusters.

1.2 Results

1.2.1 Identifying Spatially Varying Genes using the Gini Coefficient

A new approach for addressing the problem of dimension reduction has been recently discovered [23]. This approach is finding the Gini coefficient of genes and using genes with a large Gini coefficient for clustering. The Gini coefficient is used to find inequality in a dataset and is often used to measure income distribution [24]. However, the Gini coefficient can also be used to find highly spatially varying genes across a tissue section, meaning genes that are likely to be expressed in one specific region of the tissue. Finding spatially varying genes can be very powerful as there is an assumption with most clustering tools that cells of similar type are likely close together [16].

Prior to finding the Gini coefficient of genes in spatial transcriptomics data, multiple steps must first be performed. The spots must be put into groups, with spots that are spatially close grouped together. Also, an average of the gene expression of these grouped spots must be found. This is done to find the average gene expression of a spatial location of the tissue section. After this, the Gini coefficient can be found using a line of perfect equality and the Lorenz curve [25]. The line of perfect equality represents the distribution of the gene if it is evenly distributed among the tissue. The Lorenz curve represents the actual cumulative distribution of the gene among the tissue. To find the Gini coefficient of each gene, the area between the line of perfect equality and the Lorenz curve is found and is divided by the area under the line of perfect equality. Using this approach, spatially varying genes can be found and used to help better identify clusters of different cell types in tissue sections.

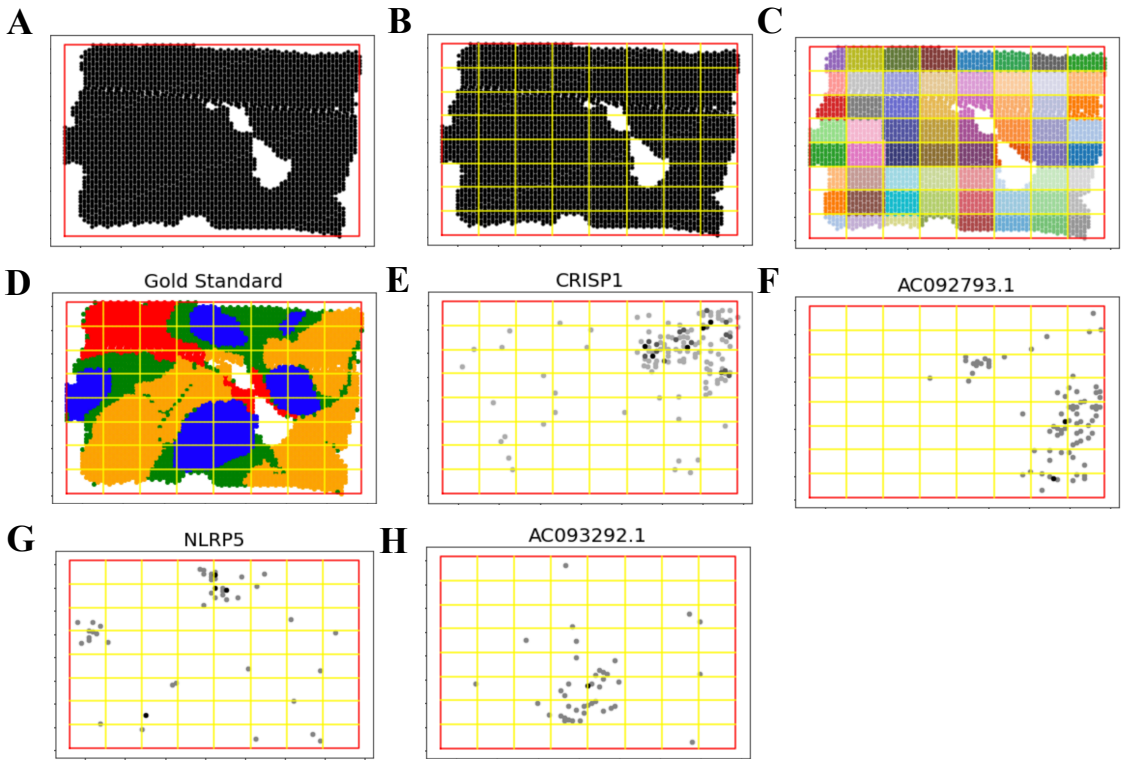

Figure 1. The steps of finding the Gini coefficient of genes and the expression pattern of genes with a high Gini coefficient. (A) Creating a rectangular outline around the spatial locations of the spots. (B) Dividing the rectangular outline into subsections. (C) Assigning the spots to the subsections. (D) The gold standard, or ideal cluster assignment of the data, containing healthy tissue in red, invasive tissue in yellow, surrounding tumor tissue in green, and cancerous tissue in blue. (E-H) The expression pattern for CRISP1, AC092793.1, NLRP5, and AC093292.1.

I created an algorithm in Python to find the Gini coefficient of genes using spatial transcriptomics data. To perform this method, the algorithm begins by finding a rectangular outline around all the spatial coordinates of the spots (Figure 1A). Rectangular subsections within the outline are then formed (Figure 1B). The size of the subsections is determined based on the total number of spots in the dataset. If there are more spots, there will be more subsections with a smaller area. All the spots in one individual subsection will be assigned to the same group (Figure 1C). The average gene expression for all spots in each group is then found. Using this data and the previously mentioned equation, the Gini coefficient of each gene is found. Genes with Gini coefficients over a certain limit (usually 0.5) can then be kept and used to run clustering analysis. This method is able to drastically reduce the number of genes analyzed by clustering tools. For

example, in a dataset obtained from mouse brain tissue, 500 spatially varying genes were found out of 15,000 total genes [23].

The validity of this method was proved using a spatial transcriptomics dataset from a tissue section of a breast cancer tumor. Figure 1D shows the gold standard of the data, containing marked healthy and cancerous tissue. This was found by a histologist analyzing the gene expression and staining of the tissue. The expression pattern of certain genes with high Gini coefficients is also shown. It can be seen that genes with high Gini coefficients do correlate with the gold standard. CRISP1 and AC09273.1 (Figure 1E/1F) both correlate with the invasive region of the tissue. NLRP5 and AC093292.1 (Figure 1G/1H) both correlate with the cancerous region of the tissue. This shows how using genes with a high Gini coefficient can be used to help assign spots into clusters. Through using this form of dimension reduction to improve clustering tools, there is a hope that certain portions of a tumor section can be denoted without the need of a histologist.

1.2.2 Benchmarking Spatial Transcriptomics Clustering Tools

Various published spatial transcriptomics clustering tools were also analyzed to better understand their ability to detect heterogeneity in tissue. The clustering tools BayesSpace, SEDR, SpaGCN, and STAGATE were all recently proposed [12-15]. To test these clustering tools, datasets obtained from mouse brain tissue sections were used. Brain tissue contains heterogeneity that follows expected patterns, so using these datasets is a good way to benchmark the clustering

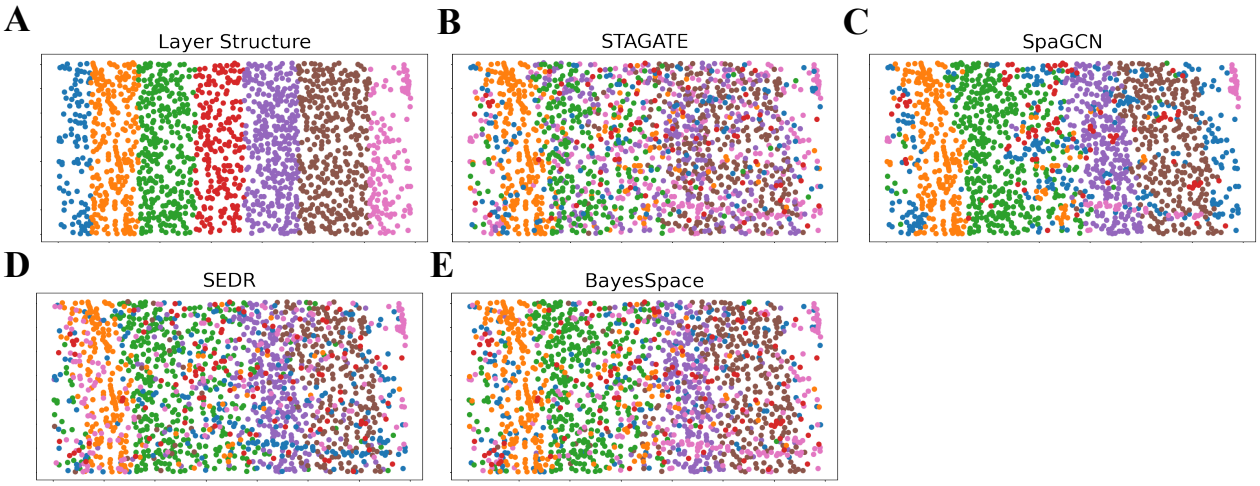

Figure 2. Cluster assignments from different tools for the visual cortex dataset. (A) The actual layer structure of the visual cortex. (B-E) The cluster assignments made from STAGATE, SpaGCN, SEDR, and BayesSpace.

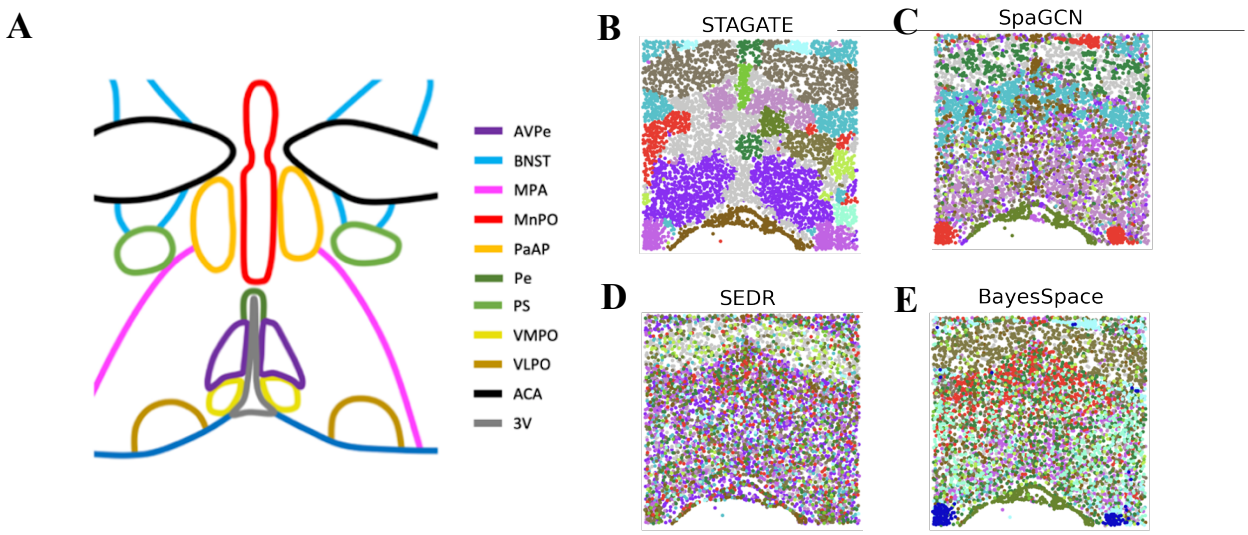

Figure 3. Cluster assignments from different tools for the hypothalamic nucleus dataset. (A) The regions of the hypothalamic nucleus. AvPe, anteroventral periventricular nucleus; BNST, bed nucleus of the stria terminalis; MPA, medial preoptic area; MnPO, median preoptic nucleus; PaAP, paraventricular hypothalamic nucleus; Pe, periventricular hypothalamic nucleus; PS, parastrial nucleus; VMPO, ventromedial preoptic nucleus; VLPO, ventrolateral preoptic nucleus; ACA, anterior commissure; 3V, third ventricle. (B-E) The cluster assignments made from STAGATE, SpaGCN, SEDR, and BayesSpace.

tools opposed to using tumor section datasets which are known to have unpredictable heterogeneous gene expression.

The first dataset tested contained spatial transcriptomics data of the visual cortex region of the mouse brain. This dataset does contain a gold standard, which is an annotated layer structure of the visual cortex (Figure 2A). Using this gold standard, the Adjusted Rand Index (ARI) of the cluster assignments from the various tools were calculated to test the accuracy of the tools. We found that BayesSpace's cluster assignments had a resulting ARI of 0.26 (Figure 2E), SEDR's cluster assignments had an ARI of 0.24 (Figure 2D), SpaGCN's cluster assignments had the highest ARI out of the tools tested at 0.41 (Figure 2C), and STAGATE's cluster assignments had the lowest ARI out of the tools we tested at 0.17 (Figure 2B). With this dataset, SpaGCN is the most successful at correctly denoting which spots should be assigned to which clusters, with BayesSpace and SEDR close behind.

The second dataset tested contained spatial transcriptomics data of the mouse major hypothalamic nucleus region of the brain. There is no gold standard for this dataset, so the ARI of

each tool cannot be calculated. However, there is a figure that shows which spatial regions correlate with known sections (Figure 3A). This figure can be treated as a gold standard and used to help verify the cluster assignments. Using this dataset, we found that BayesSpace (Figure 3E) and SEDR (Figure 3D) had cluster assignments that were the least accurate at identifying the known sections of the hypothalamic nucleus. SpaGCN was the second best at correctly identifying the known sections (Figure 3C). STAGATE was the best at identifying these known sections (Figure 3B). These are surprising results, as STAGATE was the worst tool when using the last dataset and the best tool for this dataset.

The final dataset contained spatial transcriptomics data of the anterior and posterior regions of the sagittal plane of the mouse brain. Once again, there is no gold standard for this dataset. However, there is an image of the H&E stain of the tissue section. The H&E stain can be treated as a gold standard and can help identify if the different clusters correlate with different cell types (Figure 4A). It was found that SEDR (Figure 4D) was the worst at identifying cell types in the dataset. This tool created cluster assignments that covered large portions of the tissue. SEDR was not able to distinguish some of the smaller and more difficult to identify domains, and its cluster

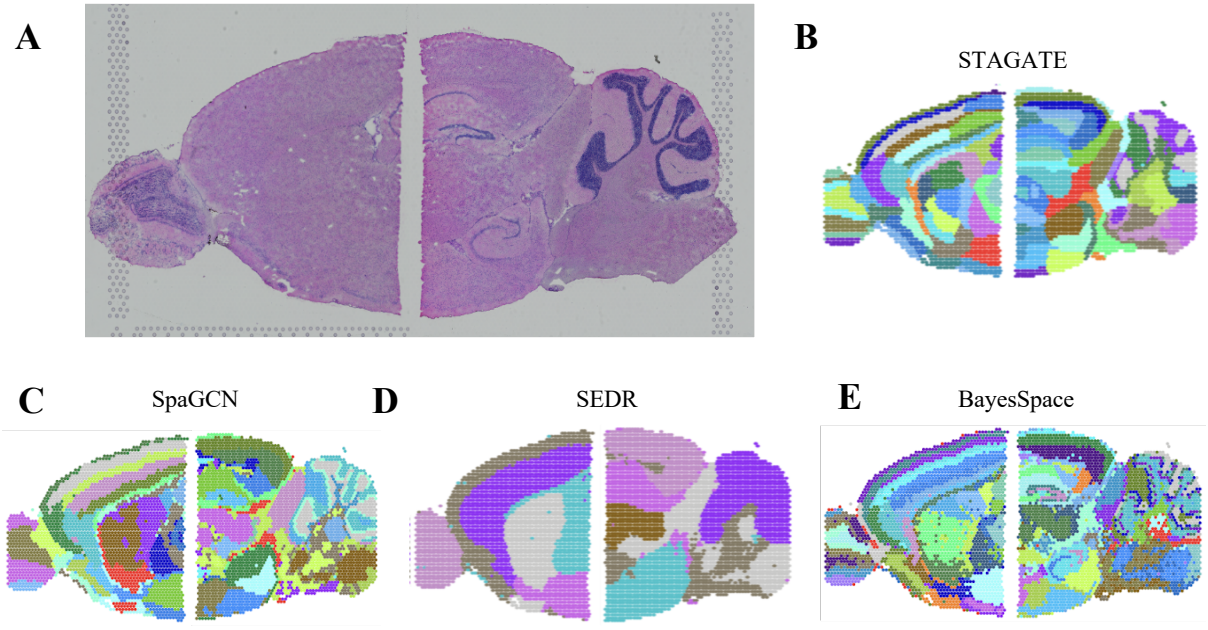

Figure 4. Cluster assignments from different tools for the sagittal plane dataset. (A) The H&E stain of the sagittal plane. (B-E) The cluster assignments made from STAGATE, SpaGCN, SEDR, and BayesSpace.

assignments did not correlate well with the sections of different cell types seen in the H&E stain. STAGATE (Figure 4B), BayesSpace (Figure 4E), and SpaGCN (Figure 4C) were all fairly equal in their cluster assignments. The accuracy of the cluster assignments can be seen when looking at the cerebellum region of the mouse brain, which is the darker, wavy region in the upper right of the H&E stain. Due to the size and shape of this region, it can be difficult to distinguish for clustering tools. STAGATE, BayesSpace, and SpaGCN were all able to correctly identify this region. These three tools were also able to correctly identify the olfactory region, which is the darker region seen in the far left of the H&E stain. Although comparing the H&E stain to cluster assignments does not allow us to find quantitative results, we are still able to qualitatively understand how well these tools can assign clusters.

Overall, when using these three datasets, SpaGCN had the best performance. While the other three tools each had poor performance when using one of the datasets, SpaGCN was able to constantly create cluster assignments that correlated with different cell types in the tissue sections. While a cancer dataset was not used in this benchmarking, it can be assumed that SpaGCN would also be the best tool at deciphering different regions in a tumor section.

1.2.3 Confirming Cluster Assignments Using Biomarkers

It was discovered that when using the dataset of the sagittal plane of the mouse brain, biomarkers for different regions of the mouse brain could be used to verify cluster assignments [23]. This could be a useful tool for confirming cluster assignments for benchmarking purposes in the future when there is not an ideal cluster assignment created for the dataset. Using the Protein Atlas, different marker genes for regions in the mouse brain were identified [26]. In Figure 5, the expression pattern for some of these marker genes are shown. The gene Gng4 is a marker of the olfactory bulb in the mouse brain (Figure 5B). The gene S100a5 is a marker for a different region of the olfactory bulb in the mouse brain (Figure 5C). The olfactory bulb is in the far-left location of this dataset. Both marker genes can be used to further confirm the cluster assignments created by SpaGCN (Figure 5A). The Gng4 expression pattern is in the same region as the greenish-brown cluster created by SpaGCN on the left side of the figure. The S100a5 expression pattern is in the same region as the blue cluster created by SpaGCN on the bottom left side of the figure. In the

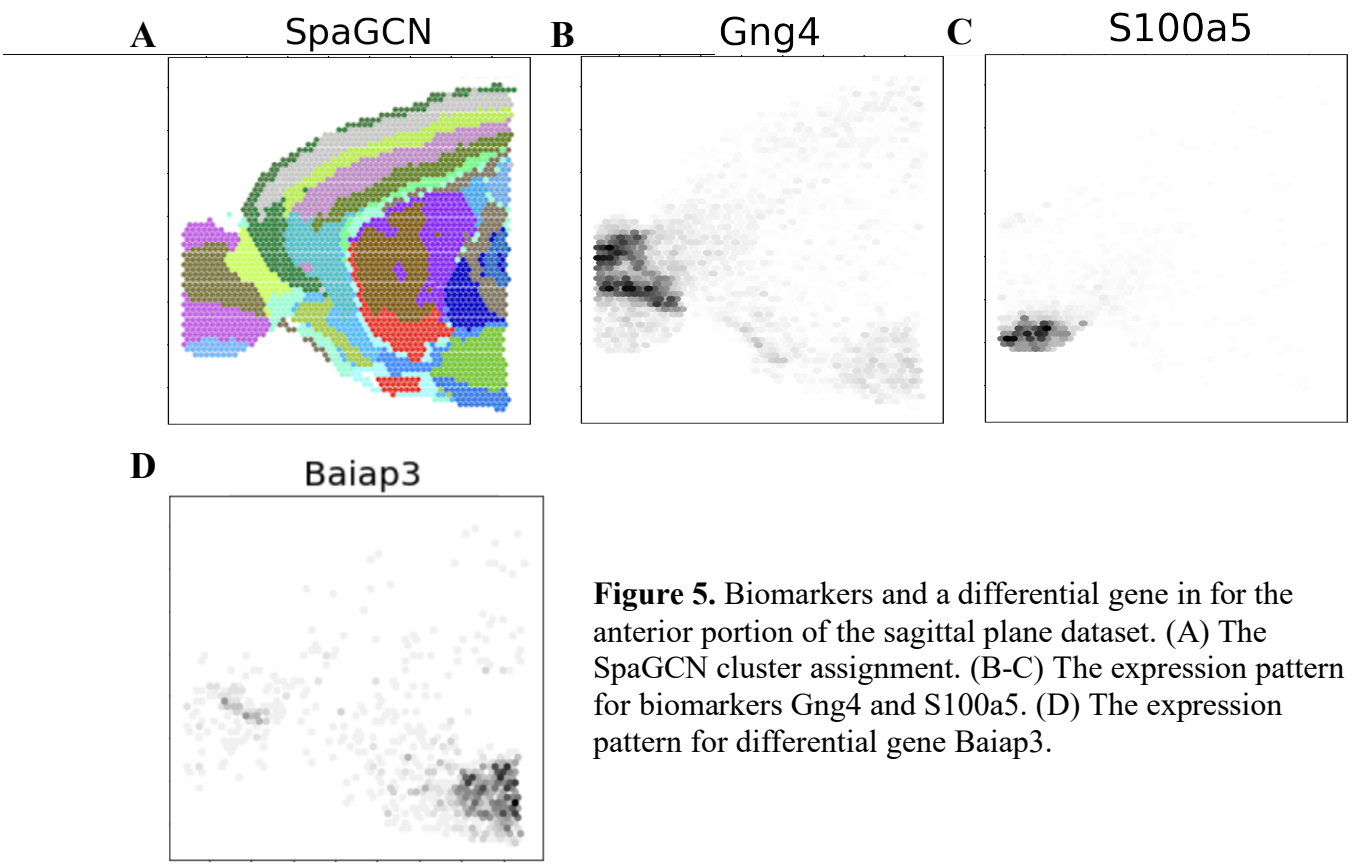

Figure 5. Biomarkers and a differential gene in for the anterior portion of the sagittal plane dataset. (A) The SpaGCN cluster assignment. (B-C) The expression pattern for biomarkers Gng4 and S100a5. (D) The expression pattern for differential gene Baiap3.

future, marker genes could be used to create a gold standard so that the quantitative accuracy of clustering tools can be found. That this is a useful method for creating a gold standard in a dataset like this, where there are known sections that have known marker genes. However, this method would be less useful when used on a cancer dataset, due to the lack of clear regions in a tumor section and the lack of known marker genes.

1.2.4 Finding the Gene Expression and Biological Differences Between Clusters

Once we were confident with the established clusters, differential gene expression analysis and gene ontology (GO) analysis were performed to better understand the genetic and biological differences among tissue.

Differential gene expression analysis is used to find the genes that are most different between different clusters. This can be used to find biomarkers and to run GO analysis. Most tools that run differential gene expression analysis are based in R. None of them could be integrated with other popular bioinformatic software in Python, so I reimplemented a new differential gene

expression analysis algorithm in Python. This algorithm uses a Mann-Whitney U-Test to find the genes that are most differentially expressed among clusters. Once again, this tool needs to be used after clusters are already established. The ability of this tool to create insightful biological results was shown using the dataset from the sagittal plane of a mouse brain. For this application, only the anterior portion of the data was used. For the bottom right green cluster that was found using SpaGCN (Figure 5A), the 'Baiap3' gene was found to be most differentially expressed. From looking at the expression pattern for 'Baiap3', the gene is in fact heavily expressed in the bottom right green cluster (Figure 5D). This cluster appears to correlate with the hypothalamus region of the mouse brain. Via Protein Atlas, it was found that 'Baiap3' is a marker for the hypothalamus region of the mouse brain. This confirms the ability of this method to find differential genes among clusters, as well as the ability of the method to discover biomarkers for regions of tissue after clustering is performed. If cluster assignments are trusted, this can be a powerful method at discovering biomarkers.

The second reason that the differential gene expression analysis algorithm was created was to perform GO analysis. GO analysis is a form of analysis that finds the biological functions

Table 1. The number of cell cycles correctly identified out of four, the percent of GO terms correlated with the correct cell cycle phase, and the percent of GO terms correlated with the incorrect cell cycle phase for each combination of clustering tools and GO analysis tools.

Clustering Tool	GO Analysis Tool	Cell Cycle Phases Identified	% of Terms Correlated with Cell Cycle Phase	% of Terms Correlated with Different Cell Cycle Phase
CCST	GO Term Finder	4	70%	3%
CCST	GO Resource	4	65%	5%
SpaGCN	GO Term Finder	4	53%	6%
CCST	g:Profiler	4	53%	5%
BayesSpace	GO Resource	4	40%	3%
SpaGCN	GO Resource	4	40%	8%
STAGATE	GO Term Finder	3	54%	11%
SEDR	GO Term Finder	3	53%	13%
STAGATE	GO Resource	3	53%	8%
SEDR	GO Resource	3	48%	13%
SEDR	g:Profiler	3	45%	10%
BayesSpace	GO Term Finder	3	44%	0%
SpaGCN	g:Profiler	3	35%	5%
STAGATE	g:Profiler	3	35%	10%
BayesSpace	g:Profiler	3	30%	3%

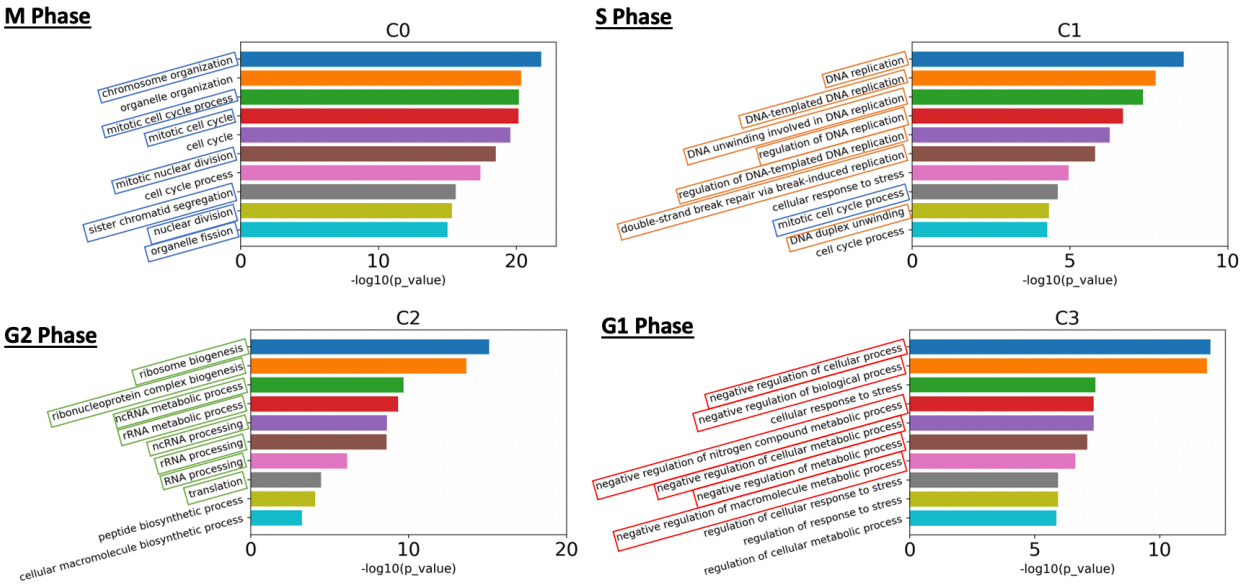

Figure 6. The top ten most significant GO terms found from the top 200 differential genes in the human osteosarcoma cells dataset. The four clusters were found using CCST. Each cluster had its differential genes found and its GO terms found. The GO Term Finder tool was used to find the GO terms.

associated with the differentially expressed genes. This is used to understand the biological differences between clusters. This tool can be extremely beneficial when looking at a dataset from a tumor section because the biological differences amongst tumor tissue need to be understood to understand how processes like metastasis may be increased in certain regions of the tissue.

There are many different GO analysis tools; however, it was found that some of the best performing and simplest to run are web-based. These tools include GO Term Finder, GO Resource, and g:Profiler [27-29]. We ran all three tools to understand their consistency. GO analysis was performed by breaking a dataset into clusters, finding the differentially expressed genes in each cluster, and running GO analysis with these differentially expressed genes. It has been shown that in datasets containing one cell type, clustering tools can be used to assign cells into clusters based on their phase in the cell cycle [11]. GO analysis can then be performed on the differential genes of each cluster, resulting in GO terms associated with the different phases. The resulting GO terms can be used to identify the phase of the clusters [11]. This approach was used to test the consistency of the GO analysis tools. A dataset containing only human osteosarcoma cells was found and the dataset was broken up into four clusters using various clustering tools. Different clustering tools were used to test the ability of the GO analysis tools when receiving slightly different inputs [11].

From previous studies, we know that the clustering tool CCST can accurately assign cells in different phases of the cell cycle into separate clusters, so we tested this tool along with other clustering tools [11]. Four clusters were found to represent the four phases of the cell cycle: gap 1 (G1), synthesis (S), gap 2 (G2) and mitosis (M). The differential genes of the four clusters were then found and run through the three mentioned GO analysis tools. In Table 1, we can see the results of the GO analysis when using different clustering tools and different GO analysis tools. The results were sorted by how many clusters, out of 4 total clusters, were able to have their phase of the cell cycle identified and then by the percent of resulting GO terms that were associated with the correct phase of the cell cycle. The combination of clustering tool and GO analysis tool that was best able to determine the cell cycle phase was CCST and GO Term Finder (Table 1). The most significant GO terms found with GO Term Finder from the four clusters found with CCST are shown in Figure 6. Terms outlined with blue are associated with the M phase, orange with the S phase, green with the G2 phase, and G1 with the red phase. This combination was able to successfully cluster cells into the four separate phases and identify the phases of clusters. We found that the different GO tools are consistent; however, GO Term Finder and GO Resource are able to provide slightly more accurate results. We discovered from these results that the clustering tool

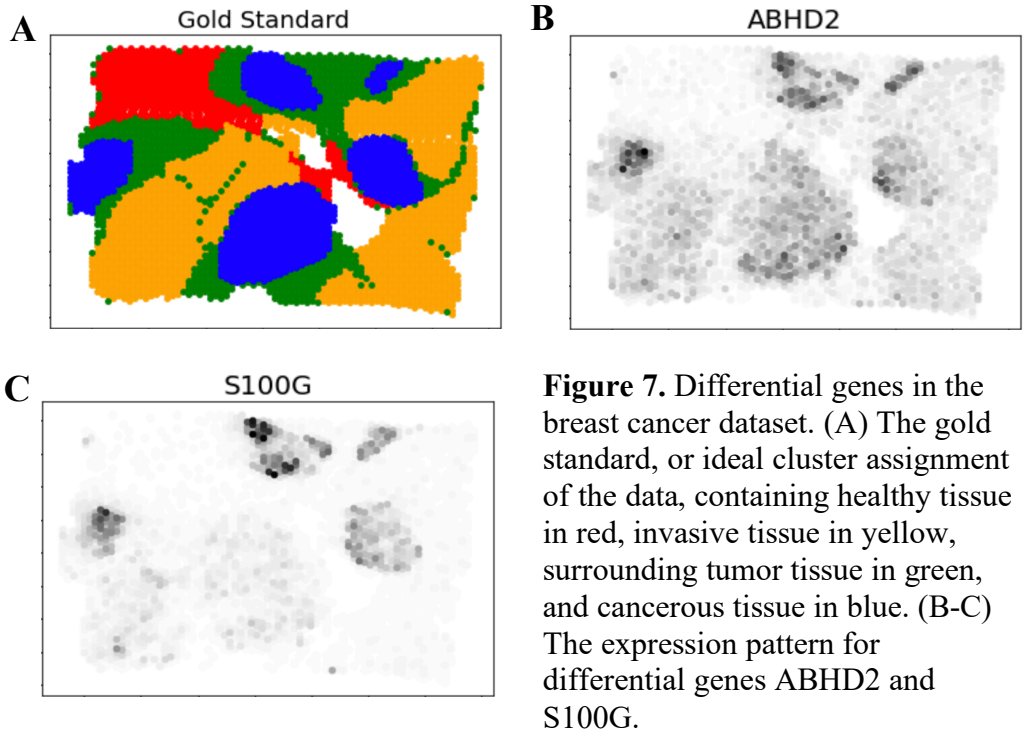

Figure 7. Differential genes in the breast cancer dataset. (A) The gold standard, or ideal cluster assignment of the data, containing healthy tissue in red, invasive tissue in yellow, surrounding tumor tissue in green, and cancerous tissue in blue. (B-C) The expression pattern for differential genes ABHD2 and S100G.

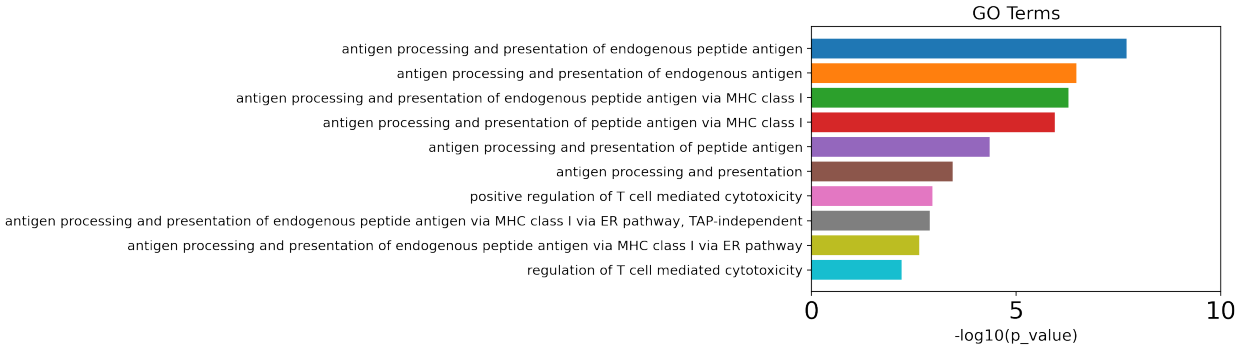

Figure 8. The top ten most significant GO terms found from the top 200 differential genes in the tumor containing cluster of the breast cancer dataset. The GO Term Finder tool was used to find these GO terms.

plays a much larger role in the resulting GO terms than the GO analysis tool. Clusters created by CCST always had all four phases identified while clusters created by SEDR always only had three of the phases identified. These results show the importance of having clusters that you can trust prior to performing GO analysis.

We explored differential gene expression analysis and GO analysis further to show their usefulness on a cancer dataset. Differential gene expression analysis was performed on the breast cancer dataset previously used, to discover potential biomarkers. The region of the breast cancer section assigned to be tumor containing was compared to the other regions of the tissue (Figure 7A). We found the two hundred genes that were most highly expressed in the tumor regions. Two of the genes highly expressed in the tumor regions are ‘ABHD2’ (Figure 7B) and ‘S100G’ (Figure 7C). It was discovered that these two highly expressed genes have been shown to play a role in cancer. ABHD2 has been previously found to promote prostate cancer growth and resistance to chemotherapy [30]. S100G has been found to be involved in the development and progression of many different cancer types, including breast cancer [31]. The results of differential gene expression analysis could be used to find patient-specific biomarkers and neoantigens for targeted therapy. We then ran the two hundred highest differentially expressed genes through GO analysis, using GO Term Finder. The ten most significant GO terms are shown in Figure 8. Using GO analysis, we found that the tumor regions of the tissue are very involved in processes associated with antigen presentation and T-cell mediated cytotoxicity. This gives us a better insight into what is biologically happening within the tumor. Tumor antigens are getting presented to elicit a response from the immune system. T-cell mediated cytotoxicity is likely working to kill the tumor

cells. Results like this from GO analysis could be used in patients to better understand how their body is responding to the tumor.

Through these experiments differential gene expression analysis has been shown to be effective at finding biomarkers. GO analysis has been shown to effectively determine the biological functions of different clusters. These tools have the possibility of helping our ability to find important differences among various heterogeneous regions of tissue.

1.3 Discussion

Through this study, it was shown how spatial transcriptomics datasets and different tools can be used to better understand the heterogeneity of gene expression in tissue. Different aspects of a clustering pipeline were used. We looked at using the Gini coefficient to help create clusters, the performance of different clustering tools, biomarkers that can be used to confirm cluster assignments, and analyses that can be conducted after clusters are created to understand the differences between clusters.

As stated previously, dimension reduction of spatial transcriptomics data is very important prior to performing clustering. Ideally, genes that are likely to be different among different cell types would be the only genes used for clustering. We have shown that the Gini coefficient is a good way to find genes with high spatial variety. It was also shown how spatially varying genes are likely to vary between different cell types. Using the tool that we created, we are able to find the Gini coefficients of genes in spatial transcriptomics datasets. Looking at a dataset from breast cancer tissue, genes that were spatially varying had expression patterns that correlated with different regions of the tissue. This shows how genes with high Gini coefficients can be used to create more accurate cluster assignments. We also discovered that this method could be used to help find biomarkers for certain regions of tissue.

We were able to benchmark the performance of four different clustering tools by running each tool on three different datasets. We were able to test three Python-based tools and one R-based tool. As each tool works in slightly different ways, it is helpful to find out which method can cluster data with the most accuracy. We found that depending on the dataset used, each tool had varying performance. For example, STAGATE was the worst clustering tool for the first dataset and the best clustering tool for the second dataset. Overall, we showed that out of the four tools tested, SpaGCN was the most effective at creating accurate clusters.

For many spatial transcriptomics datasets, there is no gold standard to test the accuracy of clustering tools. We found that when using certain datasets, biomarkers for regions of the tissue can be used to validate clustering assignments. This was shown by using biomarkers for the sagittal plane of the mouse brain. Biomarkers for different regions of the mouse brain were found using the Protein Atlas and correlated with assigned clusters [26]. Given that there are known biomarkers for regions of the tissue used, this approach will allow many more spatial transcriptomics datasets to have the accuracy of clustering assignments tested.

Finally, we looked at different methods to understand the differences among tissue regions. We showed how differential gene expression can effectively find the biomarkers of clusters. We then showed how after differential gene expression is performed, GO analysis can find the biological differences of clusters. We used three different web-based GO analysis tools to determine the consistency of the tools, finding that the clusters assignments were more important than the GO analysis tool used. We then finally tested the ability of these two analyses on the breast cancer dataset. We found that two of the upregulated genes in the cancerous regions are involved in cancer development, and these differential genes can give us a better idea of the characteristics of this cancer. This approach could be used to find neoantigens to create antigen targets for treatments like CAR T-cell therapy [32]. Using a GO analysis tool, we were able to pin down the biological differences in the cancerous region compared to the other regions. We found that the cancerous region has high antigen presentation and T-cell mediated cytotoxicity. The GO analysis tells us that this patient's immune system is actively working, likely fighting the cancer.

Overall, we were able to show the usefulness of spatial transcriptomics data, improvements that can be made to clustering tools, tools that can be used to discover biomarkers, and tools that can be used to discover differences among regions of tissue. These findings could be used to further help cancer treatments.

1.4 Methods

1.4.1 Gini Coefficient

The Gini coefficient of genes was found by first grouping spatially close spots together. The spots were grouped together by making rectangular subsections, covering the area of the data, and assigning all spots contained in one subsection to the same group. The number of subsections was determined by the number of spots in the dataset. The number of subsections was equal to the

square root of the number of spots, rounded to the nearest whole number. After groups were found, the average gene expression for all spots in a group, for each gene, were found. The Gini coefficient of each gene was then found by finding the line of perfect equality and the Lorenz curve for each gene. The Gini coefficient was calculated using the formula: *Gini Coefficient = A/(A+B)*, where *A* is the area in between the line of perfect equality and the Lorenz curve and *B* is the area under the Lorenz curve [33].

1.4.2 Clustering Tools

Different clustering tools were run to test their performance and to understand how their performance varied when clustering different datasets. The clustering tools BayesSpace, SEDR, SpaGCN, and STAGATE were all run [12-15]. The tools were run according to their published documentation, following the recommended pipeline and default parameters.

BayesSpace GitHub: https://github.com/edward130603/BayesSpace [12].

SEDR GitHub: https://github.com/JinmiaoChenLab/SEDR [13].

SpaGCN GitHub: https://github.com/jianhuupenn/SpaGCN [14].

STAGATE GitHub: https://github.com/zhanglabtools/STAGATE [15].

1.4.3 Differential Gene Expression Analysis

Differential gene expression analysis was run by comparing the gene expression in one cluster to the gene expression in all other clusters. Gene expression data for spots were placed into two separate groups. One group contained the gene expression data for the spots in the cluster to be analyzed. The other group contained the gene expression data for all other spots. To find differential genes, a Mann-Whitney U-Test was run in Python using the *mannwhitneyu* function in the *stats* package of *SciPy*. When running gene ontology (GO) analysis, only genes that were upregulated in the analyzed cluster were used.

1.4.4 Dataset Availability

The datasets used in this study are available from their original authors. All the datasets used are spatial transcriptomics datasets. The breast cancer dataset is available on the website of 10X Genomics (https://www.10xgenomics.com/). The “Human Breast Cancer (Block A Section 1)” was the dataset used. The tissue annotations for this dataset were provided by SEDR [13]. The mouse visual cortex dataset is available on the STARmap website

(https://www.starmapresources.com/data) [34]. The mouse hypothalamic nucleus dataset is available via the authors of the original article (https://datadryad.org/stash/dataset/doi:10.5061/dryad.8t8s248) [35]. The mouse sagittal plane anterior and posterior datasets are available on the website of 10X Genomics (https://www.10xgenomics.com/).

CHAPTER 2

The Mechanical Origins of Tumor Heterogeneity

2.1 Introduction

In cancer, cells often behave in unexpected ways, resulting in the creation of tumor subclones and heterogeneity within the tumor. Every type of cancer contains an assorted population of cells with differences found in their genetic, epigenetic, and phenotypic traits [36]. The heterogeneity of tumors contributes to worse patient prognosis and treatment failures. Intratumor heterogeneity is an identified driver of metastasis and is associated with decreased patient survival [4]. The intratumor heterogeneity is caused by a variety of processes including mutations during DNA replication, heterogeneity in the physical aspects of the tumor microenvironment, and abnormal signaling interactions between cells [36]. The heterogeneity in cancer needs to be better understood.

Intratumor heterogeneity is often shown in many aspects of the cancer phenotype, including cell morphology, gene expression, metabolism, motility, proliferation, angiogenic behavior, and metastatic potential [37]. There can be profound consequences resulting from these heterogeneous phenotypes. As an example, heterogeneity of angiogenic behavior can greatly affect the way that nutrients are delivered and waste is removed from the microenvironment, affecting the ability of immune cells to reach and survive in the tumor microenvironment [38]. Heterogeneity can be caused by many different processes, including through genetic mutations. Genetic mutations are common in cancer cells and due to mutations, cancer cells often have a different number of gene copies when compared to a regular genome [37]. These copy number variants (CNVs) in cancer cells can include decreases in the number of tumor suppressor genes, like TP53 which makes the p53 protein, or increases in the number of proto-oncogenes in a cell, like KDR which is the receptor for vascular endothelial growth factor (VEGF) [39, 40]. CNVs in cancer cells can cause all the alterations that we see in the cancer phenotype, including altered metabolism and angiogenic behavior. In tumors, there is also heterogeneity associated with the mechanical aspects of the microenvironment [41]. Cell subtypes seen in cancer, such as cancer associated fibroblasts (CAFs), can create a diverse extracellular matrix (ECM) in tumors by

depositing different ECM proteins like fibronectin [42]. ECM protein deposition and organization can alter tumor stiffness. Cells change their physiological behavior due to the mechanical forces that they sense, so an abnormal mechanical microenvironment can cause cells to behave in unnatural ways. A mechanically heterogeneous microenvironment can affect cellular energetics and angiogenesis [38, 41].

Tumors are typically stiffer than healthy tissue due to multiple reasons, including increased collagen deposition and cross-linking [43]. The stiffening of tissue during cancer affects many physiological aspects of cells, helping contribute to tumor heterogeneity and the formation of irregular vasculature. During cancer, clear increases of matrix stiffness can be seen in many different tissue types. Normal breast tissue has a stiffness between 0.4 kPa and 2 kPa, while breast carcinoma has much higher stiffness between 4 kPa and 12 kPa. Normal brain tissue has a stiffness ranging from 0.3 kPa - 0.5 kPa, and brain glioblastoma has a stiffness between 7 kPa and 26 kPa [38]. It is has been shown that tumor stiffness is a promoter of irregular vasculature. Angiogenic outgrowth, invasion, branching, and porosity of vasculature have all been observed to increase with increasing matrix stiffness [44]. The mechanisms remain unclear, but recent work suggests matrix stiffening causes higher endothelial cell contractility, disrupting endothelial barrier activity [45]. This is important as endothelial cell are the cells that create vasculature [38].

There is a need to better understand the cause of irregular vasculature in cancer. This upregulated, abnormal, and unorganized vasculature is a hallmark of cancer and cancer development [38]. The irregular vasculature can affect the delivery of nutrients to the tumor microenvironment, furthering the heterogeneity in tumors [38]. The heterogeneity of tumors and irregular vasculature of the tumor microenvironment lead to increases in tumor growth and metastasis. The irregular vasculature also results in leaky and dilated blood vessels [38]. This contributes to a reduced ability to deliver drugs to tumors [38]. Because the abnormal vasculature structure can both help cancer progress and hinder different treatment options, there is a large need to further understand the causes. Finding a connection between the behavior of endothelial cells in the tumor microenvironment is essential to further understanding this phenomenon.

Traction forces and metabolic activity are important components that affect the tumor microenvironment and can affect the physiological aspects of endothelial cells. Traction forces are important components in cell adhesion, cell spreading, migration, and ECM remodeling [46-48]. Metabolic activity is associated with cell migration and proliferation. Both components are heavily

involved in angiogenesis. Due to stiffer substrates, endothelial cells have been observed to have increased traction forces [49]. Cell-generated traction forces are needed for blood vessel formation; however, when traction forces are too high, the permeability of vasculature is raised, and leaky blood vessels are formed [38, 50, 51]. Altered metabolism is also observed in endothelial cells when stiffness is varied [52, 53]. The metabolism of endothelial cells has been observed to regulate angiogenesis, so altered metabolism can result in unnatural angiogenic tendencies [54].

Traction forces are generated by stress fibers and actin filaments, which allow cells to attach to their substrate through focal adhesions [55]. In endothelial cells, traction forces and migration are initiated by actin polymerization, allowing the cell to pull on its environment [48]. Many of the studies first quantifying these relationships come from the mid-2000's [46, 47]. Although there have been some improvements, the method used then, traction force microscopy (TFM), is still an effective method for measuring the traction forces that cells cause. TFM was first developed as a method in 1996 [56]. TFM allows for the quantification of the traction forces generated by cells in 2D, by measuring the deformation of a cell's substrate [48]. The current method of conducting TFM involves seeding cells on polyacrylamide (PA) gel substrates which contain fluorescent beads [48]. PA gels are very useful due the ability to tune their stiffness by adjusting the ratio of acrylamide to bis-acrylamide in their solution [48]. To measure TFM, cells are seeded on PA gels, an image is taken of the cell and the fluorescent beads embedded in the PA gel, the cell is removed, and a final image is taken of the fluorescent beads [48]. As the cells pull on the substrate, the position of the fluorescent beads will be deformed due to traction forces on the substrate. This bead movement can be used to determine the traction forces that cells create [48]. Traction forces have been shown to correlate with many aspects of the tumor microenvironment. There has been work in our lab showing a positive correlation with increasing cellular traction force and increasing metastatic potential in cancer cells [57]. In endothelial cells, traction forces have also been shown to play a large part in cell-cell adhesion and network formation [55]. If traction forces are increased, vasculature becomes leaky, and if traction force is inhibited, endothelial cell networks are disrupted [55].

Cellular metabolism, which was once difficult to study in individual cells, has been explored more recently due to the prevalence of new measurement tools. One of these tools, the PercevalHR probe, has often been used for this purpose. PercevalHR contains a combination of a fluorescent protein and a protein GlnK1, which binds to active ATP and ADP [58]. Binding of

each ATP and ADP produces a fluorescent signal, which allows the ATP/ADP ratio of cells to be calculated as a metric of the metabolic activity of cells [58]. Cells use their intracellular energy through the dephosphorization of ATP into ADP [59]. If cells have higher energy demands, they need a higher supply of ATP [59]. So, a higher ATP/ADP ratio of a cell means that that cell has higher energetic demands than a cell with a lower ATP/ADP ratio. The PercevalHR probe has been used by our lab to show many changes that occur in cells' metabolic activity due to being in the tumor microenvironment, such as showing that cancer cells often choose the least costly energetic path during migration [59].

It is known that cell traction forces are transmitted to the substrate through stress fibers and focal adhesions, and the formation of more polar actin stress fibers and an increase in focal adhesions have been observed in cells with increasing substrate stiffness [60-62]. Furthermore, it is known that stress fiber-mediated contractility requires ATP. From this, a relationship can be drawn between substrate stiffness, cellular traction forces, and cell metabolism, all contributing to irregular vasculature in the tumor microenvironment.

In this study, we used human umbilical vein endothelial cells (HUVECs) to measure endothelial cell traction forces and metabolic activity on substrates of varying stiffness, to better understand the relationship of these two cellular properties in association to matrix stiffness. Using PA gels, substrates for cell culture of tunable stiffness were created. We seeded HUVECs on these substrates, then measured cellular traction forces and cellular ATP/ADP ratios. Different methods were also used to further understand the mechanism behind the changes of these properties with changing matrix stiffness. Because previous studies show increased cellular traction forces on stiffer substrates and traction forces require energy, we hypothesized that endothelial traction forces and metabolic activity would both increase due to increasing substrate stiffness [49]. However, our data indicates increased traction forces and a biphasic trend of metabolism in endothelial cells due to increasing substrate stiffness. Overall, this data suggests that higher matrix stiffness may increase a cell's likelihood of producing irregular vasculature.

2.2 Results

To understand the relationship between the properties of endothelial cells in relation to substrate stiffness, HUVECs were cultured on substrates of varying stiffnesses. Unless otherwise stated, the HUVECs were measured in a single-cell state. PA gels of stiffness 0.5 kPa, 1 kPa, 2.5 kPa, 5 kPa, 10 kPa, and 20 kPa were used as the substrate for cells.

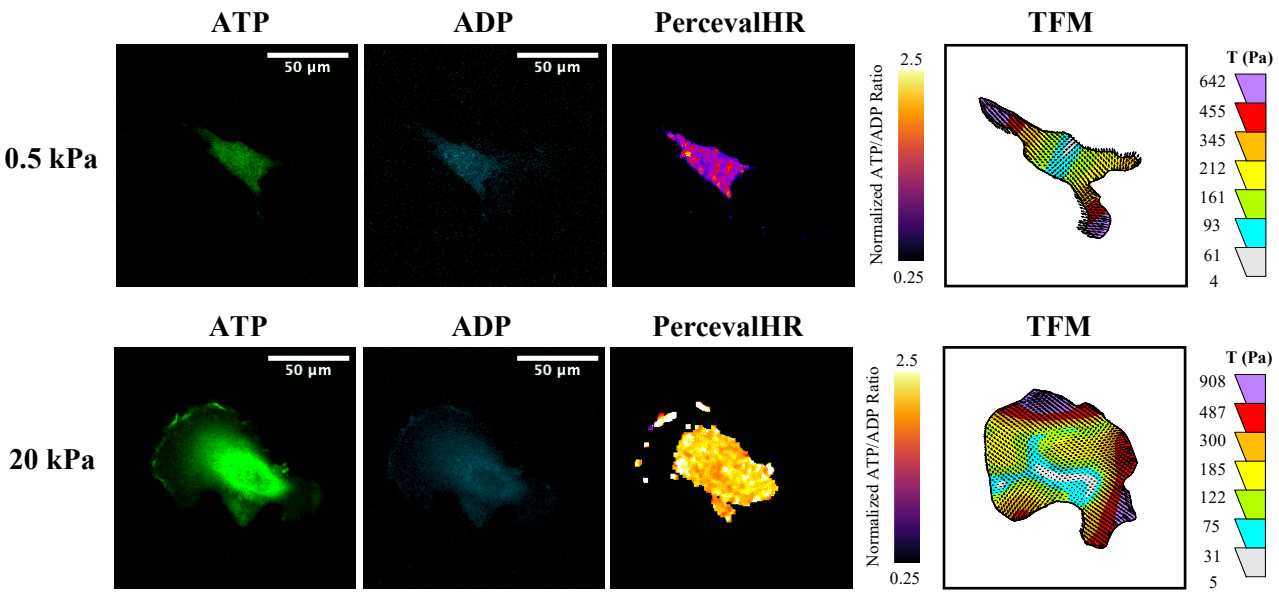

Figure 9. HUVECs cultured for 24 hours on PA gels of stiffnesses 0.5 kPa and 20 kPa. Confocal microscope images of the fluorescent ATP and ADP channels of the PercevalHR probe are shown. The heatmaps of the ATP/ADP ratio are shown. The traction force maps are also shown.

2.2.1 Substrate Stiffness Alters the Physiological Properties of HUVECs

To investigate the change in a cell's behavior due to substrate stiffness, the traction forces and ATP/ADP ratio of cells on PA gels of varying stiffnesses were measured using traction force microscopy and the PercevalHR probe. These measurements were executed in the same experiment, with each cell having both measured to allow for direct comparisons on a cell-by-cell basis. Figure 9 shows microscope images obtained of cells on 0.5 kPa and 20 kPa substrates. In this figure. an image of the fluorescent ATP channel, an image of the fluorescent ADP channel, a heatmap of the ATP/ADP ratio, and the traction force diagram are shown for each cell.

It was found that the traction force of cells increases as cells are on increasing stiffness substrates (Figure 10A). Cell spreading also increases as cells are on increasing stiffness substrates (Figure 10B). However, there is a linear relationship with traction force and cell spreading, with larger cells generally applying more traction force (Figure 10C). Because cells are typically larger on stiffer substrates and larger cells typically apply higher traction forces, traction stress was used moving forward to measure the forces that cells apply to their environment. This is to allow cells on stiffer substrates to be fairly compared to cells on softer substrates. Traction stress is calculated as the traction force of a cell divided by the area of a cell. HUVECs on substrates of increasing

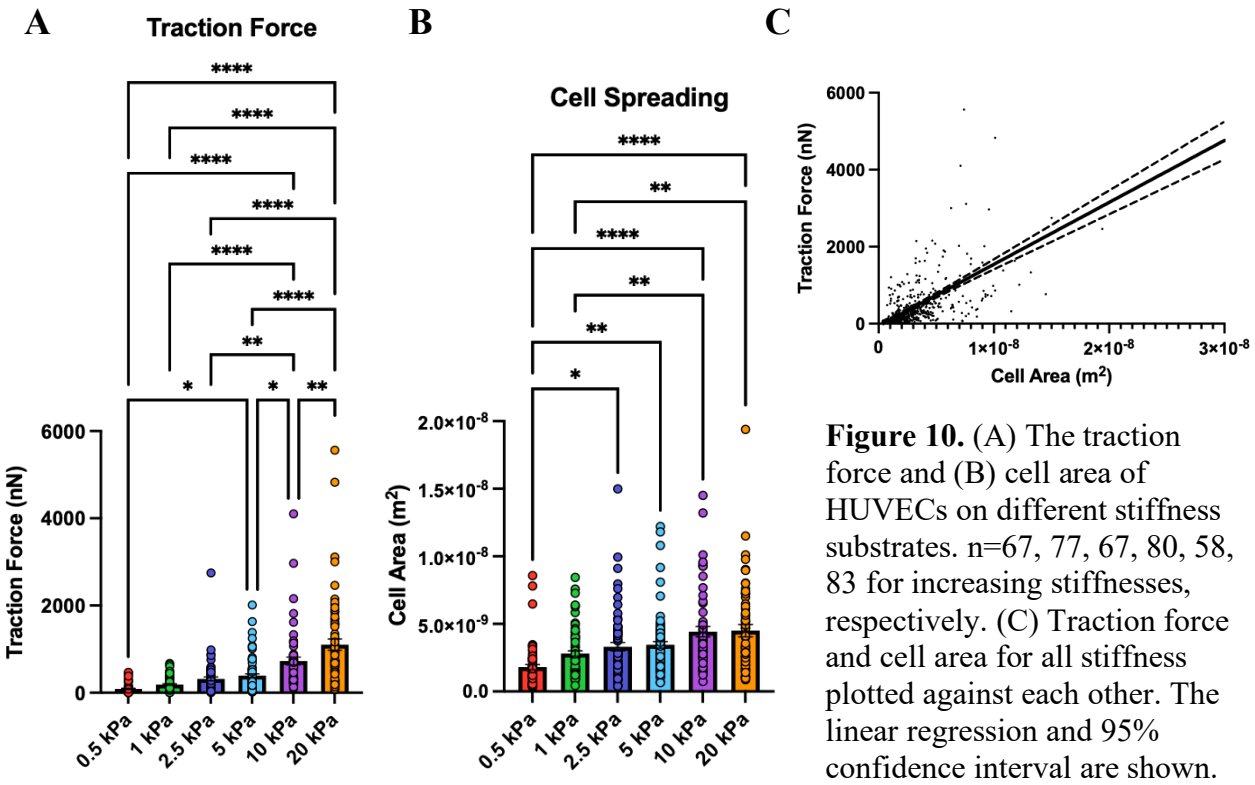

Figure 10. (A) The traction force and (B) cell area of HUVECs on different stiffness substrates. n=67, 77, 67, 80, 58, 83 for increasing stiffnesses, respectively. (C) Traction force and cell area for all stiffness plotted against each other. The linear regression and 95% confidence interval are shown.

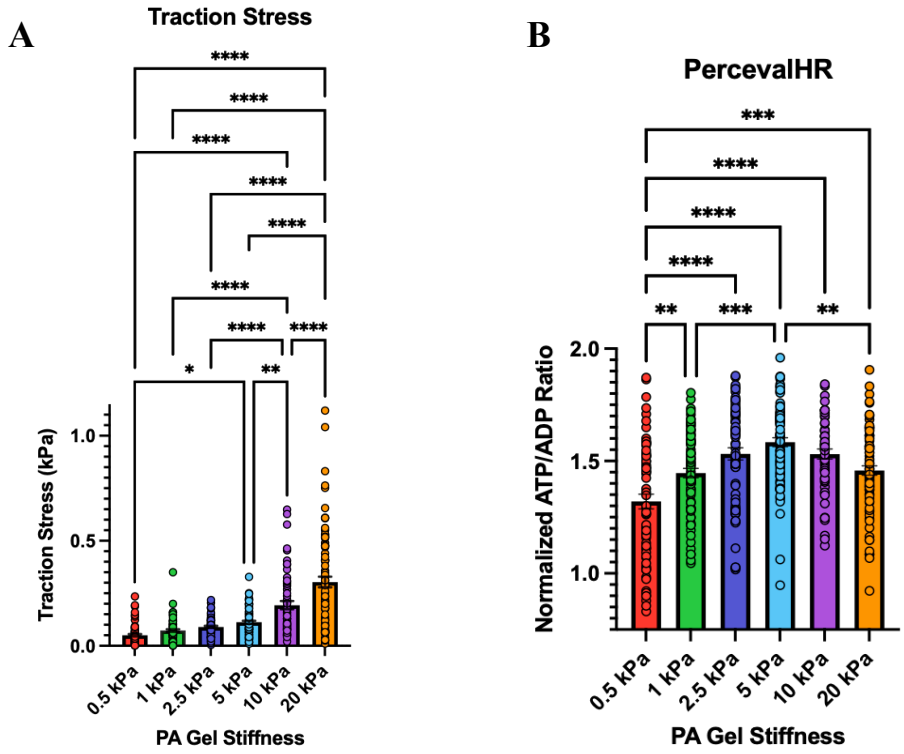

Figure 11. (A) The traction stress and (B) ATP/ADP ratio of HUVECs on different stiffness substrates. n=67, 77, 67, 80, 58, 83 for increasing stiffnesses, respectively.

stiffness have increasing traction stresses (Figure 11A). This follows the results of previous studies using endothelial cells [49].

Based on measurements of ATP using the PercevalHR probe, there is a biphasic trend of cell energetics as a function of substrate stiffness (Figure 11B). Since cells have higher traction stresses on stiffer substrates, and cells require energy to apply traction stresses, we hypothesized that cellular ATP/ADP ratio would increase as cells are on increasing stiffnesses. So, we expected that higher traction stresses would correlate with higher ATP/ADP ratios. From 0.5 kPa and going to 5 kPa, the ATP/ADP ratio of cells increases, following the expected trend. However, from 5 kPa to 20 kPa, the ATP/ADP ratio of cells decreases. This is the opposite of the expected trend and the opposite of the trend observed over softer stiffnesses. This unexpected biphasic trend has led to further investigation to discover its causes.

2.2.2 Relationship of Altered Properties

To further investigate the relationship between traction stress and cellular ATP/ADP ratio, the efficiency of cells was calculated. Efficiency represents the amount of energy that cells require to apply similar traction stresses on their environment. Efficiency was calculated by dividing the traction stress of individual cells by the associated ATP/ADP ratio. When analyzing the efficiency of cells on varying stiffness substrates, it was found that there is no significant difference between the efficiency of cells on 0.5-5 kPa substrates; however, whenever cells reach the stiffest substrates, 10 and 20 kPa, their efficiency starts to significantly

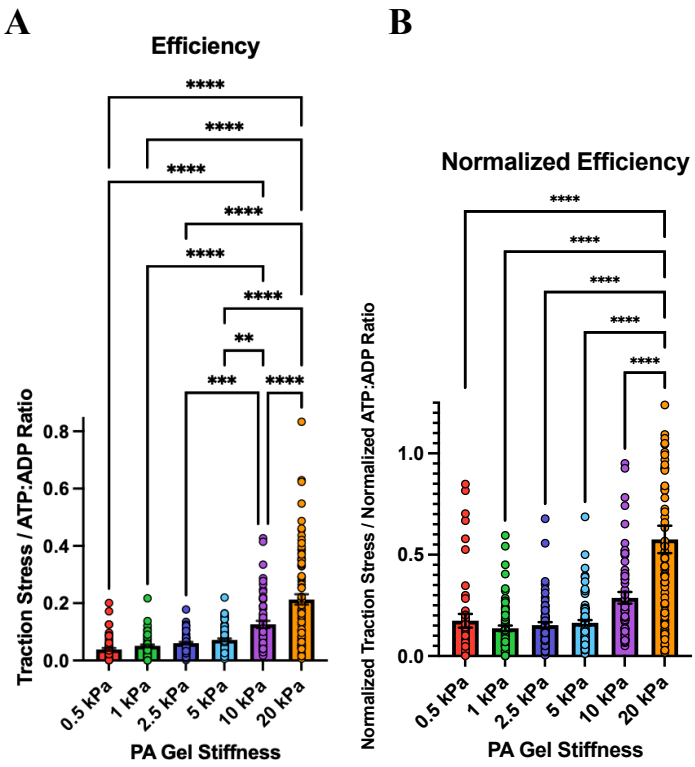

Figure 12. (A) The efficiency and (B) normalized efficiency of HUVECs on different stiffness substrates. n = 67, 77, 67, 80, 58, 83 for increasing stiffnesses, respectively.

increase (Figure 12A). This metric is important as it may explain why cells on 10 and 20 kPa substrates are able to create more traction stresses than cells on 5 kPa substrates, while having a lower ATP/ADP ratio.

Since traction stress measurements contain more variability compared to ATP/ADP ratio measurements, it is possible the traction measurements are overly weighted in the efficiency calculations. To account for this possibility, normalized efficiency was calculated by normalizing traction stress and ATP/ADP ratio separately. To normalize values in each dataset, the lowest value of each was set to zero and the highest value was normalized to one. The normalized traction stresses were then divided by the normalized ATP/ADP ratios to calculate normalized efficiency. Normalized efficiency calculations demonstrated similar trends to unnormalized efficiency calculations (Figure 12B).

2.2.3 Further Exploring Altered Energetics

Given that the PercevalHR probe indicated that there is a biphasic relationship between ATP/ADP and substrate stiffness on an individual cell basis, we sought to expand this work using an ATPlite assay to measure overall ATP in a cell population. After culturing cells on varying stiffness substrates, the cells were lysed, the lyse solution for each stiffness was placed into 12 different wells of a 96 well plate, and the ATP content of each well was measured using an ATPlite assay. This assay measured the ATP content of many cells which were cultured together. The number of cells lysed for each stiffness were counted and verified to be similar, to ensure that there was not an altered number of cells on different stiffness substrates. The relative amount of ATP in cells for each stiffness is represented as the luminescence of each well after performing the ATPlite assay. Our data indicates there is also a biphasic trend that emerges in bulk ATP levels of cells cultured on various stiffness substrates (Figure 13A). Cells cultured on the softest substrates, 0.5 - 2.5 kPa, all have similar levels of ATP. Cells cultured on increasing stiffnesses, 5 - 10 kPa, exhibit increased ATP. On the stiffest substrate, 20 kPa, the ATP level slightly decreases. This is similar to the biphasic trend observed for ATP/ADP ratio; however, in this instance the largest ATP was found in cells cultured on 10 kPa substrates. This is opposed to the largest ATP/ADP ratio, which was observed in cells cultured on 5 kPa substrates. This supports that there is a biphasic trend of energetics when HUVECs are on substrates of increasing stiffness. The reason for this biphasic trend is still not apparent; however, this result confirms a decrease in energetics in HUVECs on the stiffest substrates. These results also indicate there are differences in the two methods used to

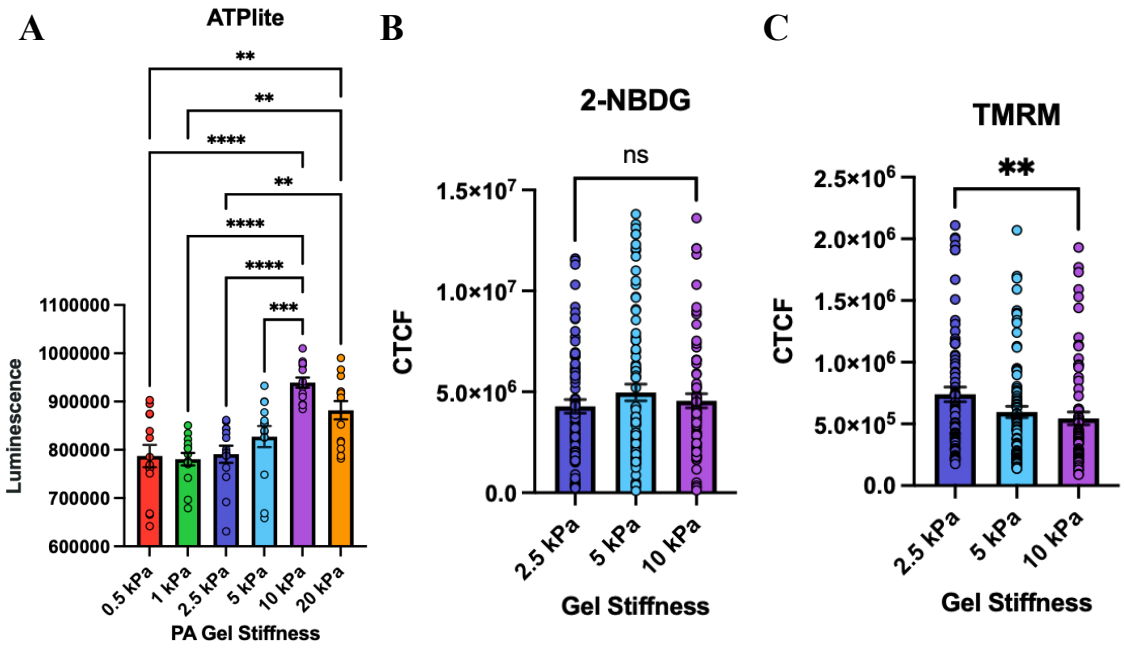

Figure 13. (A) The relative ATP level of HUVECs on different stiffness substrates. For each stiffness, n=14. (B-C) The fluorescence of 2-NBDG and TMRM of HUVECs on different stiffness substrates. n = 74, 85, 70 for increasing stiffnesses, respectively.

measure energetics. PercevalHR measures the ATP/ADP ratio of cells and ATPlite measures the ATP level of cells. It is likely the slight difference in measurements between the two methods is due to changing ADP levels, which are not measured by ATPlite.

To further investigate the relationship of cellular energetics and substrate stiffness, another method was attempted, using 2-NBDG and TMRM. 2-NBDG is fluorescent glucose analog, which can be used as a measure of the relative glycolic rate of cells. TMRM is a fluorescent mitochondrial membrane potential indicator, which can be used to find the relative rate of oxidative phosphorylation (OXPHOS) in cells. Using both methods together, the relative level of both major metabolic pathways can be found for individual cells. These methods were used on cells on substrates of stiffnesses 2.5 kPa, 5 kPa, and 10 kPa. Using 2-NBDG, it was found that the relative glycolic rate of cells is not affected by substrate stiffness (Figure 13B). Based on TMRM, as cells are on increasingly stiff substrates, their OXPHOS rate decreases (Figure 13C). However, TMRM does have the tendency to wash away over time, affecting the ability to obtain accurate results. During the experiment this was not accounted for, and TMRM was imaged in the order of cells on 2.5 kPa gels first, cells on 5 kPa gels second, and cells on 10 kPa gels last. TMRM may have

washed away in the cells imaged later, causing this significant difference in results. Because of this, the current results are not thorough enough to draw conclusions from, and this experiment needs to be repeated.

Seahorse XF analyzers can measure glycolic and OXPHOS rates of cells with more accuracy than 2-NBDG and TMRM. An issue with this method is that Seahorse XF analyzers use specific plastic well plates, which are very difficult to polymerize PA gels in. The difficulty is because PA gels' polymerization is inhibited in the presence of oxygen, and there is a lot of oxygen on the surface of plastic well plates. One previous study was found where the authors polymerized PA gels inside of a Seahorse XF plate; however, after contacting the authors it was found that the method to do this was difficult and hard to replicate [63]. It was then attempted to find a different method to polymerize PA gels inside of this plastic plate. After trying numerous methods involving plate treatment, increased initiator concentration, the use of a UV activated adhesive, using riboflavin as a secondary initiator, and using oxygen scavengers; a method to polymerize PA gels inside of the Seahorse plate was finally discovered. This method uses an increased initiator concentration along with sodium sulfite, an oxygen scavenger. The protocol for PA gel polymerization inside of a Seahorse is available in the Appendix. However, this method only worked on 10 kPa and 20 kPa PA gels, removing the benefit, as we would like other softer stiffnesses to be tested as well. Because of this, the ATPlite was the sole experiment to verify the biphasic trend found using the PercevalHR probe.

2.2.4 Determining the Mechanism of Action

Further experiments were performed to determine the mechanism by which a cell changes its physiological properties due to substrate stiffness.

To further explore the relationship of HUVEC behavior due to varying stiffness substrates, PercevalHR HUVECs were cultured to form a monolayer. Cells on stiffer substrates have been noted to have larger and more mature focal adhesions [64]. Cells in a monolayer have increased adhesions due to cell-to-cell interactions. Because of this, it was thought that cells in a monolayer on softer substrates may behave similarly to single-cells on stiffer substrates. Due to the way TFM works, traction forces could not be measured for cells in a monolayer. So, the ATP/ADP ratios of HUVECs in the monolayer were measured and compared to single-cell HUVECs. Through this experiment, it was found that HUVECs in a monolayer had a significantly lower ATP/ADP ratio

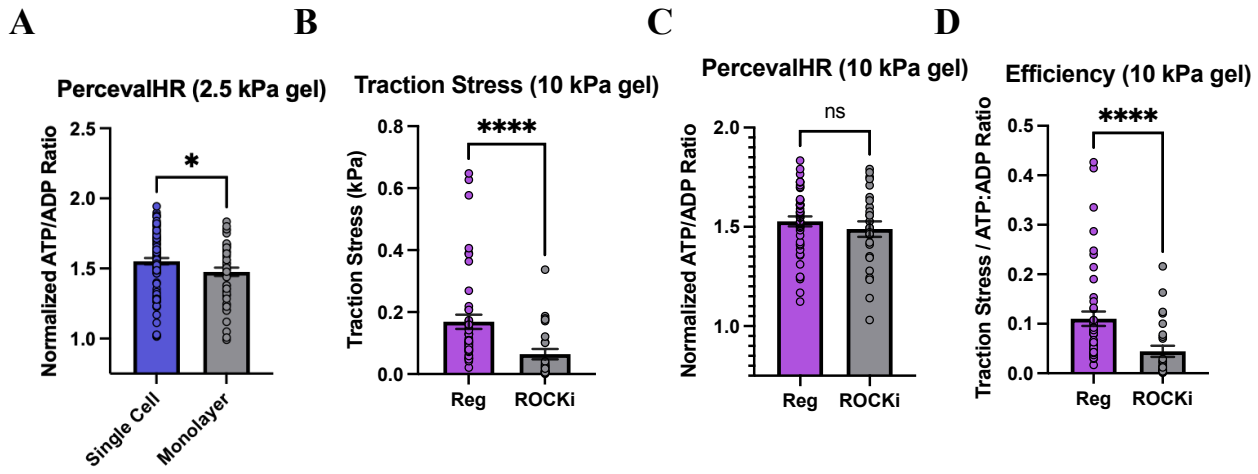

Figure 14. (A) The ATP/ADP ratio of HUVECs single-cell and in a monolayer, on 2.5 kPa stiffness substrates. Single-cell, n=93; monolayer, n = 49. (B) The traction stress, (C) ATP/ADP ratio, and (D) efficiency of HUVECs nontreated and treated with ROCK inhibitor. Nontreated, n = 58; treated, n = 26.

when compared to single cells in isolation (Figure 14A). This supports the hypothesis that cells in a monolayer on softer gels will have a similar behavior as cells on stiffer gels. The cells in a monolayer on 2.5 kPa gels had a similar ATP/ADP ratio to single-cells on 20 kPa gels. This experiment was only performed for cells on 2.5 kPa PA gels and would need to be expanded to include other stiffnesses to provide more definitive conclusions.

The final experiment to further understand the mechanisms of varying ATP/ADP ratio and efficiency of cells on varying stiffness substrates was treating the cells with an inhibitor of Rho-associated protein kinase (ROCK), Y-27632. ROCK critical for pathways involving cell spreading, migration, and traction force generation [59, 65]. Inhibiting ROCK majorly disrupts these functions. There is typically an increase in cell spreading, migration, and traction force generation for cells on stiffer substrates, and it was thought that inhibiting ROCK for cells on stiffer substrates would induce the cells to behave similarly to cells on softer substrates. Because of this, it was expected that inhibiting ROCK in cells would decrease traction stress and efficiency, while increasing the ATP/ADP ratio, which are all characteristics measured of cells on softer substrates. Treatment with ROCK inhibitor for HUVECs on a 10 kPa substrate was able to significantly decrease HUVECs' traction stress (Figure 14B), while not altering HUVECs' ATP/ADP ratio (Figure 14C). This resulted in a significantly decreased efficiency (Figure 14D). This supported the hypothesis that treatment with ROCK inhibitor would decrease a cell's traction stress and

efficiency. However, the full hypothesis was not supported as the ATP/ADP ratio was not measured to increase. Although, because the ATP/ADP ratio trend is biphasic, it is difficult to confirm with ATP/ADP ratio that the treatment did not make cells behave as if they were on a softer substrate. With this treatment on a 10 kPa substrate, the cells could be behaving similar to untreated cells on a 2.5 kPa substrate. Compared to regular HUVECs on 10 kPa, HUVECs on 2.5 kPa have lower traction stresses, lower efficiency, and similar ATP/ADP ratio, all of which were measured here. Because ROCK is important for so many pathways, we decided to pursue other experiments in the future that will have less of a major effect on the cell instead of repeating this experiment for cells on other stiffnesses. This was because we were not sure if the inhibition of ROCK changed traction stress and efficiency due to other pathways being affected.

2.3 Discussion

2.3.1 Substrate Stiffness Alters the Physiological Properties of HUVECs

Through these experiments we have found many physiological trends of HUVECs in relation to substrate stiffness. We found that cellular traction stress increases as cells are on substrates of increasing stiffness, which follows findings in other studies using endothelial cells and other cell types [49, 57]. This study does expand on the stiffnesses that were used in other studies, so it is interesting to see that the trend remains the same when a wider range of stiffnesses are used.

Whenever cells are on gels between 0.5-5 kPa, we have an expected trend of ATP/ADP ratio. As stated earlier, cells require energy to apply traction stress on their environment. It would be assumed that cells on stiffer substrates, which apply higher traction stresses, would require more energy to apply higher traction stress. This is what was found when looking at cells on stiffnesses from 0.5-5 kPa. As the substrate gets stiffer, energetics increase. However, when cells are on 10 and 20 kPa gels, the trend reverses. From 5-20 kPa as substrate stiffness increases, traction stress is increased, and ATP/ADP ratio is decreased. Biphasic trends of cell behavior in relation to substrate stiffness are not entirely new. For certain cell types, biphasic trends have been found in cell spreading in relation to substrate stiffness [57]. Although this is not identical to the trend noticed here, it shows that cells can have an ideal substrate stiffness where their physiological properties are at a maximum. It is likely that for each different cell type, the stiffness to maximize

certain cellular processes would be different. Repeating this experiment using a different type of endothelial cell would further this thought and provide more context into what is occurring.

It was considered that other cellular processes that are not involved in the generation of traction forces could be accounting for the difference of ATP/ADP ratio for cells on different stiffness substrates. There are many processes in cells which require ATP, that do not involve traction force generation. Some of these processes are cell proliferation, intracellular signaling, and DNA/RNA synthesis [66]. However, differences in proliferation according to substrate stiffness were not observed when performing these experiments. Biphasic trends in these processes have also not been observed in relation to substrate stiffness, showing that this is unlikely the cause of varying energetics.

The biphasic trend of ATP/ADP ratio leads to a hypothesis that cells pull on their environment more efficiently at these higher stiffnesses (10 and 20 kPa), reducing the required energy to generate traction stress. Actin has been found to be more polarized and show higher organization in cells on higher stiffness substrates [67]. This leads to more stable stress fibers, more order in the cytoskeletal network, and increased cellular traction forces [67]. As actin is a main component in the cell involved with generating traction forces, the increased actin organization may lead to an increased ability of cells to generate traction forces, with more efficiency. Focal adhesions have also been found to be more mature and larger in cells on higher stiffness substrates [64]. Focal adhesions are the way that cells attach to their environment. Focusing on the relationship between focal adhesions and the actin cytoskeleton, focal adhesions act as the connection of the cell to the environment and the cytoskeleton is what enables the cell to create a force to transmit. An enhanced relationship between actin and focal adhesions would allow cells to be more efficient in traction force generation.

2.3.2 Relationship of Altered Properties

The ability of cells to create traction stresses more efficiently, on stiffer substates, is further shown by efficiency. It was shown by using efficiency that cells on stiffer substrates do have an increased ability to create higher traction stress, with similar ATP/ADP ratios. As stated previously, cells use their actin cytoskeleton, in addition with their focal adhesions to help pull on their environment. Cells on stiffer substrates are better able to apply traction stress to their environment with a similar amount of energy, and actin and focal adhesion structure is likely the reason for this.

2.3.3 Further Exploring Altered Energetics

We wanted to verify the energetic trends observed with the PercevalHR probe, so we used ATPlite to measure the ATP levels of cells on different stiffness substrates. We found that there is also a biphasic trend of cellular ATP levels in relation to substrate stiffness. This confirms the findings when using PercevalHR, that there is a biphasic relationship between cellular energetics and substrate stiffness. As stated earlier, this trend is very interesting and unexpected, and the mechanisms of it need to be explored further.

The trend of ATP levels in cells was slightly different from the trend of ATP/ADP ratios. The trend of ATP seems to slightly lag behind the ATP/ADP trend, as the highest level of ATP observed was when cells were on 10 kPa substrates and the highest ATP/ADP ratio was observed when cells were on 5 kPa substrates. The origin of the difference is likely from the varying metabolic measuring methods. The energetics of cells are often difficult to determine, as there are many cellular processes which create and use energy. The difference in measurements is likely due to changes in ADP levels, which are not measured with ATPlite. Also, both methods fail to measure how AMP levels are altered. ATP is hydrolyzed into ADP to create energy, and ADP can further be hydrolyzed into AMP [66]. While the hydrolyzation of ADP into AMP is a process used by cells much less frequently, it still affects the energetics of cells. Both methods have shown the ability to quantify metabolic activity within cells [68, 69]. However, as these methods work in different ways, it is difficult to do a simple one-to-one comparison between them.

When using ATPlite, the bulk ATP levels of all cells inside of a well are measured. It would be useful to use a method where the ATP level of each individual cell could be quantified, so that the relationship between an individual cell's traction stress and ATP level could be found, just like what was able to be done with ATP/ADP ratios. This could further verify the trends found when using PercevalHR.

Other methods used to verify the energetic trends either need to be repeated, in the case of 2-NBDG/TMRM, or were not able to be used with the stiffnesses needed, in the case of the Seahorse XF analyzer. 2-NBDG/TMRM can be repeated to help verify the results found and to determine which metabolic pathway is being affected by substrate stiffness. Although the Seahorse XF analyzer cannot be used for this experiment, a novel method of creating PA gels inside of a Seahorse XF plate was discovered. This method can be used by our lab in the future to further the connection between cell metabolism and substrate stiffness.

2.3.4 Determining the Mechanism of Action

Finally, we explored the mechanism that creates higher cell efficiency on higher stiffness substrates. Because cells on stiffer substrates show more developed focal adhesions and actin cytoskeletons, we enhanced or disrupted these features in an attempt to increase or decrease cell efficiency.

Cells on stiffer substrates have larger and more mature focal adhesions [64]. It is thought the larger surface area of focal adhesions help contribute to increased efficiency. As a way to create a similar focal adhesion structure for cells on softer substrates, cells were cultured into a monolayer. Cells in a monolayer are more adhered to their environment, due to cell-to-cell adhesions, which should allow a similar level of force transmission while using a lower amount of energy. As stated previously, cells in a monolayer cannot have their traction forces measured because TFM must be conducted in a single-cell manner, where the bead displacement of a single cell is measured. This means that the efficiency of these cells can also not be calculated. Because of this, the ATP/ADP level of these cells was measured, to see if the cells in a monolayer required less energy to produce their traction stresses compared to single-cells. We found that cells in a monolayer do have a lower ATP/ADP ratio compared to single-cells. This confirms the hypothesis that more advanced or additional focal adhesions help contribute to a lower cellular energy level to produce traction forces. When cells are in a monolayer, there are also changes in cell-to-cell signaling. It was considered when performing this experiment that these changes could affect the energetics of cells. Because of this, other experiments were performed to confirm the results. This experiment was only performed on one substrate stiffness, 2.5 kPa. It would be interesting to repeat the experiment on other stiffness substrates to further understand how more cell attachment affects the efficiency of cells. It would also be interesting to measure cellular traction forces of cells in a monolayer, to get an actual measurement of their efficiency. While TFM cannot be used for this, there are other methods of measuring cellular traction force that could possibly be used in future experiments.

Cells were treated with an inhibitor of ROCK in an attempt to disrupt the efficiency of cells. In this experiment we were still trying to alter the efficiency of cells; however, here we wanted to decrease efficiency. In the monolayer experiment we attempted to increase efficiency. ROCK is known to help coordinate cytoskeleton reorganization, and inhibiting ROCK can disrupt the cytoskeletal organization and prevent cell polarization [70]. ROCK treatment was used to

disrupt the organization of actin, to decrease the efficiency of cells on stiffer substrates. We found that inhibiting ROCK does decrease the efficiency of cells, when on 10 kPa substrates. The results show that organization of the actin cytoskeleton does affect cell efficiency. While this further confirms the hypothesis, we decided to not expand this experiment for different stiffness substrates. Since inhibiting ROCK affects many major pathways, the effects may be too upstream to determine that efficiency is decreased solely due to a reduction in actin polarization.

The future directions for this study are to use treatments which can enhance or restrict certain integrin binding, to affect the efficiency of cells on different stiffness substrates. A FAK inhibitor can be used to inhibit certain integrin binding, in an attempt to decrease the efficiency of cells. Cells can also be treated with manganese, in the form of manganese chloride, to increase integrin binding and to promote focal adhesion formation. Manganese would be used in an attempt to increase the efficiency of cells. Finally, it would be useful to stain cells for actin and focal adhesions, to visualize their structure due to substrate stiffness. Results from these proposed experiments are hypothesized to show that cells on stiffer substrates have more efficient organization of stress fibers which allow them to transmit traction stress more efficiently.

2.4 Methods

2.4.1 Cell Culture and Reagents

Human umbilical vein endothelial cells (HUVECs) were cultured using endothelial growth medium 2 (EGM-2), by Lonza. HUVECs were maintained until their fifth passage. HUVECs were transfected to express the PercevalHR and pHRed probes, to allow measurement of cellular ATP/ADP ratios [68, 71]. To inhibit Rho-associated protein kinase (ROCK), HUVECs were treated with 10 μM Y27632 in DMSO over a period of 3 hours.

2.4.2 Polyacrylamide Gel Preparation

Polyacrylamide (PA) gels were created using methods previously described [72, 73]. The ratio of acrylamide to bis-acrylamide was altered to tune gel stiffness from 0.5 kPa to 20 kPa. PA gels were coated with 0.1 mg/mL rat tail type I collagen.

2.4.3 Traction Force Microscopy

Traction force microscopy was performed using methods previously described [57, 73]. PA gels were created at different stiffnesses containing 0.5 μm fluorescent beads. HUVECs were

cultured on PA gels for 24 hours in 2 mL EGM-2. After 24 hours, a LSM800 confocal microscope was used to take images of HUVECs and fluorescent images of the beads, at the surface of the PA gel. HUVECs were then removed from the gel using trypsin. More fluorescent images of the beads were then acquired. Cell areas and bead displacements between the two sets of fluorescent images were found using LIBTRC [74].

2.4.4 Fluorescent Microscopy

PercevalHR, pHRed, TMRM, and 2-NBDG uptake were measured on a LSM800 confocal microscope using a 20x/0.8 objective. PercevalHR was excited using a 488/405 nm laser and emission was collected through a 450-550 nm bandpass filter. pHRed was excited using a 561/488 nm laser and emission was collected through a 576 nm long-pass filter. 2-NBDG and TMRM were performed using methods previously described [75]. 2-NBDG was excited using a 488 nm laser and emission was collected through a 490-650 nm bandpass filter. TMRM was excited using a 561-nm laser and emission was collected through a 562-700 nm bandpass filter. All images were quantified using ImageJ.

2.4.5 ATPlite

HUVECs were seeded on top of PA gels of stiffnesses 0.5-20 kPa in a 6-well plate at a density of 200,000 cells per well. Cells were cultured in 2 mL EGM-2. 24 hours after seeding, EGM-2 was removed and 1.5 mL of ATPlite's mammalian cell lysis solution was introduced per well. The plates were placed in an orbital shaker and cells were allowed to lyse for 5 minutes. Lyse solution, containing the lysed HUVECs, was then moved from wells of the 6-well plate to wells of a 96-well plate. 50 µL of lyse solution was moved into each well of the 96-well plate. 50 µL of ATPlite's substrate solution was then pipetted into each well of the 96-well plate. The 96-well plate was dark adapted for 10 minutes. Luminescence was measured using the Cytation 5 plate reader. We then ensured that a similar number of cells were lysed when using different stiffness PA gels. To do this, HUVECs were cultured on PA gels of stiffnesses 0.5-20 kPa in a separate 6-well plate. After culturing in EGM-2 for 24 hours, cells were removed with the use of trypsin and counted using a hemocytometer.

2.4.6 Statistical Analysis

Statistical analyses were performed using GraphPad Prism 9. For finding comparisons between two groups, an unpaired Student's t-test was used. For finding comparisons between multiple groups, a one-way analysis of variance (ANOVA) was used, with multiple comparisons of means between groups. The figures used show all data points found, the mean, and standard error.

REFERENCES

[1] National Cancer Institute (2020). Cancer Statistics. Available from https://www.cancer.gov/about-cancer/understanding/statistics.

[2] Hanahan, D. (2022). Hallmarks of cancer: new dimensions. *Cancer discovery*, *12*(1), 31-46.

[3] American Cancer Society (2019). Cancer Treatment & Survivorship Facts & Figures 2019-2021. American Cancer Society. Available from https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/cancer-treatment-and-survivorship-facts-and-figures/cancer-treatment-and-survivorship-facts-and-figures-2019-2021.pdf.

[4] Hapach, L. A., Carey, S. P., Schwager, S. C., Taufalele, P. V., Wang, W., Mosier, J. A., ... & Reinhart-King, C. A. (2021). Phenotypic heterogeneity and metastasis of breast cancer cells. *Cancer research*, *81*(13), 3649-3663.

[5] Peltomäki, P. (2012). Mutations and epimutations in the origin of cancer. *Experimental cell research*, *318*(4), 299-310.

[6] Sheikh, A., Hussain, S. A., Ghori, Q., Naeem, N., Fazil, A., Giri, S., ... & Al Tamimi, D. M. (2015). The spectrum of genetic mutations in breast cancer. *Asian Pacific Journal of Cancer Prevention*, *16*(6), 2177-2185.

[7] Ståhl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., ... & Frisén, J. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, *353*(6294), 78-82.

[8] Burgess, D. J. (2019). Spatial transcriptomics coming of age. *Nature Reviews Genetics*, *20*(6), 317-317.

[9] Rao, A., Barkley, D., França, G. S., & Yanai, I. (2021). Exploring tissue architecture using spatial transcriptomics. *Nature*, *596*(7871), 211-220.

[10] Patrick, E., Taga, M., Ergun, A., Ng, B., Casazza, W., Cimpean, M., ... & Mostafavi, S. (2020). Deconvolving the contributions of cell-type heterogeneity on cortical gene expression. *PLOS Computational Biology*, *16*(8), e1008120.

[11] Li, J., Chen, S., Pan, X., Yuan, Y., & Shen, H. B. (2021). CCST: Cell clustering for spatial transcriptomics data with graph neural network.

[12] Zhao, E., Stone, M. R., Ren, X., Guenthoer, J., Smythe, K. S., Pulliam, T., ... & Gottardo, R. (2021). Spatial transcriptomics at subspot resolution with BayesSpace. *Nature Biotechnology*, *39*(11), 1375-1384.

[13] Fu, H., Xu, H., Chong, K., Li, M., Ang, K. S., Lee, H. K., ... & Chen, J. (2021). Unsupervised spatially embedded deep representation of spatial transcriptomics. *Biorxiv*.

[14] Hu, J., Li, X., Coleman, K., Schroeder, A., Ma, N., Irwin, D. J., ... & Li, M. (2021). SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature methods*, *18*(11), 1342-1351.

[15] Dong, K., & Zhang, S. (2022). Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature communications*, *13*(1), 1-12.

[16] Elyanow, R., Zeira, R., Land, M., & Raphael, B. J. (2021). STARCH: Copy number and clone inference from spatial transcriptomics data. *Physical Biology*, *18*(3), 035001.

[17] Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic acids research*, *47*(D1), D419-D426.

[18] Monjo, T., Koido, M., Nagasawa, S., Suzuki, Y., & Kamatani, Y. (2022). Efficient prediction of a spatial transcriptomics profile better characterizes breast cancer tissue sections without costly experimentation. *Scientific reports*, *12*(1), 1-12.

[19] Wang, R., Peng, G., Tam, P. P., & Jing, N. (2022). Integration of computational analysis and spatial transcriptomics in single-cell study. *Genomics, Proteomics & Bioinformatics*.

[20] Pomaznoy, M., Sethi, A., Greenbaum, J., & Peters, B. (2019). Identifying inaccuracies in gene expression estimates from unstranded RNA-seq data. *Scientific reports*, *9*(1), 1-10.

[21] Perez-Carrasco, R., Beentjes, C., & Grima, R. (2020). Effects of cell cycle variability on lineage and population measurements of messenger RNA abundance. *Journal of the Royal Society Interface*, *17*(168), 20200360.

[22] Cardoso, J. R., Pereira, L. M., Iversen, M. D., & Ramos, A. L. (2014). What is gold standard and what is ground truth?. *Dental press journal of orthodontics*, *19*, 27-30.

[23] Chen, Y., Zhou, S., Li, M., Zhao, F., & Qi, J. (2022). STEEL enables high-resolution delineation of spatiotemporal transcriptomic data.

[24] United States Census Bureau (2021). Gini Index. Available from https://www.census.gov/topics/income-poverty/income-inequality/about/metrics/gini-index.html.

[25] Sitthiyot, T., & Holasut, K. (2020). A simple method for measuring inequality. *Palgrave Communications*, *6*(1), 1-9.

[26] Sjöstedt, E., Zhong, W., Fagerberg, L., Karlsson, M., Mitsios, N., Adori, C., ... & Mulder, J. (2020). An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science*, *367*(6482), eaay5947.

[27] Boyle, E. I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J. M., & Sherlock, G. (2004). GO:: TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, *20*(18), 3710-3715.

[28] The Gene Ontology Consortium. The Gene Ontology Resource. Available from http://geneontology.org.

[29] Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic acids research*, *47*(W1), W191-W198.

[30] Obinata, D., Takada, S., Takayama, K. I., Urano, T., Ito, A., Ashikari, D., ... & Inoue, S. (2016). Abhydrolase domain containing 2, an androgen target gene, promotes prostate cancer cell proliferation and migration. *European Journal of Cancer*, *57*, 39-49.

[31] Allgöwer, C., Kretz, A. L., von Karstedt, S., Wittau, M., Henne-Bruns, D., & Lemke, J. (2020). Friend or foe: S100 proteins in cancer. *Cancers*, *12*(8), 2037.

[32] Zhu, Y., Qian, Y., Li, Z., Li, Y., & Li, B. (2021). Neoantigen-reactive T cell: An emerging role in adoptive cellular immunotherapy. *MedComm*, *2*(2), 207-220.

[33] Paul Boyce (2021). What is the Gini Coefficient. BoyceWire. Available from https://boycewire.com/what-is-the-gini-coefficient/.

[34] Wang, X., Allen, W. E., Wright, M. A., Sylwestrak, E. L., Samusik, N., Vesuna, S., ... & Deisseroth, K. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*, *361*(6400), eaat5691.

[35] Moffitt, J. R., Bambah-Mukku, D., Eichhorn, S. W., Vaughn, E., Shekhar, K., Perez, J. D., ... & Zhuang, X. (2018). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science*, *362*(6416), eaau5324.

[36] Gay, L., Baker, A. M., & Graham, T. A. (2016). Tumour cell heterogeneity. *F1000Research*, *5*.

[37] Marusyk, A., & Polyak, K. (2010). Tumor heterogeneity: causes and consequences. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, *1805*(1), 105-117.

[38] Zanotelli, M. R., & Reinhart-King, C. A. (2018). Mechanical forces in tumor angiogenesis. *Biomechanics in Oncology*, 91-112.

[39] MedlinePlus (2020). TP53 gene. National Library of Medicine (US). Available from: https://medlineplus.gov/genetics/gene/tp53/.

[40] Chial, H. (2008). Proto-Oncogenes to Oncogenes to Cancer. Nature Education 1 (1): 33 What Drives Cancer Cells to Grow and Divide Uncontrollably Turning into Cancer? Studies of Proto-Oncogenes Reveal Some Clues about How Normal Cellular Processes Mutate and Go Awry.

[41] Zanotelli, M. R., Zhang, J., & Reinhart-King, C. A. (2021). Mechanoresponsive metabolism in cancer cell migration and metastasis. *Cell Metabolism*, *33*(7), 1307-1321.

[42] Belhabib, I., Zaghdoudi, S., Lac, C., Bousquet, C., & Jean, C. (2021). Extracellular matrices and cancer-associated fibroblasts: targets for cancer diagnosis and therapy?. *Cancers*, *13*(14), 3466.

[43] Taufalele, P. V., VanderBurgh, J. A., Muñoz, A., Zanotelli, M. R., & Reinhart-King, C. A. (2019). Fiber alignment drives changes in architectural and mechanical features in collagen matrices. *Plos one*, *14*(5), e0216537.

[44] Bordeleau, F., Mason, B. N., Lollis, E. M., Mazzola, M., Zanotelli, M. R., Somasegar, S., ... & Reinhart-King, C. A. (2017). Matrix stiffening promotes a tumor vasculature phenotype. *Proceedings of the National Academy of Sciences*, *114*(3), 492-497.

[45] VanderBurgh, J. A., Potharazu, A. V., Schwager, S. C., & Reinhart-King, C. A. (2020). A discrete interface in matrix stiffness creates an oscillatory pattern of endothelial monolayer disruption. *Journal of cell science*, *133*(18), jcs244533.

[46] Reinhart-King, C. A., Dembo, M., & Hammer, D. A. (2003). Endothelial cell traction forces on RGD-derivatized polyacrylamide substrata. *Langmuir*, *19*(5), 1573-1579.

[47] Reinhart-King, C. A., Dembo, M., & Hammer, D. A. (2005). The dynamics and mechanics of endothelial cell spreading. *Biophysical journal*, *89*(1), 676-689.

[48] Kraning-Rush, C. M., Carey, S. P., Califano, J. P., & Reinhart-King, C. A. (2012). Quantifying traction stresses in adherent cells. *Methods in Cell Biology*, *110*, 139-178.

[49] Califano, J. P., & Reinhart-King, C. A. (2010). Substrate stiffness and cell area predict cellular traction stresses in single cells and cells in contact. *Cellular and molecular bioengineering*, *3*(1), 68-75.

[50] Kniazeva, E., & Putnam, A. J. (2009). Endothelial cell traction and ECM density influence both capillary morphogenesis and maintenance in 3-D. *American Journal of Physiology-Cell Physiology*, *297*(1), C179-C187.

[51] Zebda, N., Dubrovskyi, O., & Birukov, K. G. (2012). Focal adhesion kinase regulation of mechanotransduction and its impact on endothelial cell functions. *Microvascular research*, *83*(1), 71-81.

[52] Park, J. S., Burckhardt, C. J., Lazcano, R., Solis, L. M., Isogai, T., Li, L., ... & Danuser, G. (2020). Mechanical regulation of glycolysis via cytoskeleton architecture. *Nature*, *578*(7796), 621-626.

[53] Ge, H., Tian, M., Pei, Q., Tan, F., & Pei, H. (2021). Extracellular matrix stiffness: new areas affecting cell metabolism. *Frontiers in Oncology*, *11*, 631991.

[54] Fraisl, P., Mazzone, M., Schmidt, T., & Carmeliet, P. (2009). Regulation of angiogenesis by oxygen and metabolism. *Developmental cell*, *16*(2), 167-179.

[55] Califano, J. P., & Reinhart-King, C. A. (2010). Exogenous and endogenous force regulation of endothelial cell behavior. *Journal of biomechanics*, *43*(1), 79-86.

[56] Dembo, M., Oliver, T., Ishihara, A., & Jacobson, K. (1996). Imaging the traction stresses exerted by locomoting cells with the elastic substratum method. *Biophysical journal*, *70*(4), 2008-2022.

[57] Kraning-Rush, C. M., Califano, J. P., & Reinhart-King, C. A. (2012). Cellular traction stresses increase with increasing metastatic potential. *PloS one*, *7*(2), e32572.

[58] Mosier, J. A., Wu, Y., & Reinhart-King, C. A. (2021). Recent advances in understanding the role of metabolic heterogeneities in cell migration. *Faculty Reviews*, *10*.

[59] Zanotelli, M. R., Rahman-Zaman, A., VanderBurgh, J. A., Taufalele, P. V., Jain, A., Erickson, D., ... & Reinhart-King, C. A. (2019). Energetic costs regulated by cell mechanics and confinement are predictive of migration path during decision-making. *Nature communications*, *10*(1), 1-12.

[60] Sarangi, B. R., Gupta, M., Doss, B. L., Tissot, N., Lam, F., Mège, R. M., ... & Ladoux, B. (2017). Coordination between intra-and extracellular forces regulates focal adhesion dynamics. *Nano letters*, *17*(1), 399-406.

[61] Doss, B. L., Pan, M., Gupta, M., Grenci, G., Mège, R. M., Lim, C. T., ... & Ladoux, B. (2020). Cell response to substrate rigidity is regulated by active and passive cytoskeletal stress. *Proceedings of the National Academy of Sciences*, *117*(23), 12817-12825.

[62] Yeh, Y. C., Ling, J. Y., Chen, W. C., Lin, H. H., & Tang, M. J. (2017). Mechanotransduction of matrix stiffness in regulation of focal adhesion size and number: reciprocal regulation of caveolin-1 and β1 integrin. *Scientific reports*, *7*(1), 1-14.

[63] Pasqualini, F. S., Agarwal, A., O'Connor, B. B., Liu, Q., Sheehy, S. P., & Parker, K. K. (2018). Traction force microscopy of engineered cardiac tissues. *PloS one*, *13*(3), e0194706.

[64] Li, D., & Wang, Y. L. (2020). Mechanobiology, tissue development, and tissue engineering. In *Principles of tissue engineering* (pp. 237-256). Academic Press.

[65] Hoon, J. L., Tan, M. H., & Koh, C. G. (2016). The regulation of cellular responses to mechanical cues by Rho GTPases. *Cells*, 5(2), 17.

[66] Dunn, J., & Grider, M. H. (2021). Physiology, adenosine triphosphate. In *StatPearls [Internet]*. StatPearls Publishing.

[67] Ladoux, B., Mège, R. M., & Trepat, X. (2016). Front–rear polarization by mechanical cues: From single cells to tissues. *Trends in cell biology*, *26*(6), 420-433.

[68] Tantama, M., Martínez-François, J. R., Mongeon, R., & Yellen, G. (2013). Imaging energy status in live cells with a fluorescent biosensor of the intracellular ATP-to-ADP ratio. *Nature communications*, *4*(1), 1-11.

[69] Yang, J. S., Hsu, J. W., Park, S. Y., Li, J., Oldham, W. M., Beznoussenko, G. V., ... & Hsu, V. W. (2018). GAPDH inhibits intracellular pathways during starvation for cellular energy homeostasis. *Nature*, *561*(7722), 263-267.

[70] Messi, Z., Bornert, A., Raynaud, F., & Verkhovsky, A. B. (2020). Traction forces control cell-edge dynamics and mediate distance sensitivity during cell polarization. *Current Biology*, *30*(9), 1762-1769.

[71] Tantama, M., Hung, Y. P., & Yellen, G. (2011). Imaging intracellular pH in live cells with a genetically encoded red fluorescent protein sensor. *Journal of the American Chemical Society*, *133*(26), 10034-10037.

[72] Califano, J. P., & Reinhart-King, C. A. (2008). A balance of substrate mechanics and matrix chemistry regulates endothelial cell network assembly. *Cellular and Molecular Bioengineering*, *1*(2), 122-132.

[73] Schwager, S. C., Bordeleau, F., Zhang, J., Antonyak, M. A., Cerione, R. A., & Reinhart-King, C. A. (2019). Matrix stiffness regulates microvesicle-induced fibroblast activation. *American Journal of Physiology-Cell Physiology*, *317*(1), C82-C92.

[74] Dembo, M., & Wang, Y. L. (1999). Stresses at the cell-to-substrate interface during locomotion of fibroblasts. *Biophysical journal*, *76*(4), 2307-2316.

[75] Schwager, S. C., Mosier, J. A., Padmanabhan, R. S., White, A., Xing, Q., Hapach, L. A., ... & Reinhart-King, C. A. (2022). Link between glucose metabolism and epithelial-to-mesenchymal transition drives triple-negative breast cancer migratory heterogeneity. *Iscience*, 25(10), 105190.

[76] The Human Protein Atlas. Protein Atlas. Available from https://www.proteinatlas.org.

APPENDIX

Protocol for Polyacrylamide Gel Polymerization Inside of V28 Seahorse Plate

Purpose: To allow the metabolic measurements of cells on different stiffness substrates. The changes in the protocol from the usual PA gel protocol are due to the oxygen on the surface of plastic well plates. Oxygen inhibits PA gel polymerization. Only stiffer stiffness PA gels have been observed to work with this protocol, due to their shorter polymerization time, allowing less oxygen to enter the solution.

1. Remove N6 from fridge and allow to come to room temperature before opening. Do not open if cold. (Take out fridge at the start and leave for at least 30 min before opening.)
2. Combine the following components in a 50 ml centrifuge tube to prepare the gel mixture needed. Gel compliance is based on the ratio of acrylamide:bis-acrylamide.

E (Pascals)	Acrylamide %	**Vol. acrylamide (ml)**	Bis-acrylamide %	**Vol. bis-acrylamide (ml)**	**HEPES pH6 (ml)**	**TEMED (µl)**	**MilliQ water (ml)**
10000	7.5	**3.75**	0.35	**3.5**	**2.6**	**10**	**8.64**
15000	12	**6**	0.13	**1.3**	**2.6**	**10**	**8.59**
20000	12	**6**	0.19	**1.9**	**2.6**	**10**	**7.99**
30000	12	**6**	0.28	**2.8**	**2.6**	**10**	**7.09**

3. pH solution to 6.0 by adding ~35 µl of 2M HCl. First add 30 µl and check on pH meter, then slow add HCl until pH is 6.00-6.05. Vortex solution to mix.
4. Remove 845 µl of gel mixture and place in a 5 ml plastic culture tube.
5. Degas solution for 30 minutes by placing the tube in a vacuum flask and seal with a rubber stopper. Tap flask gently on table periodically to let bubbles rise.
6. While the solution is degassing, place the plastic V28 Seahorse plate inside the plasma cleaner. Using a timer, allow the plasma flow for at least 10 minutes. This step will allow the solution to spread out more evenly along the wells of the Seahorse plate.
7. Weigh out 5.6 mg of N6 linker in a 1.7 ml Eppendorf tube. You need 5.6 mg N6 per tube of gel mixture.
8. Make a 10% ammonium persulfate (APS) in MilliQ water solution (0.05 g APS into 500 µl MilliQ water). You only need a small amount to make gels, but you need extra to polymerize any leftover solution before discarding.
9. Make a 5% Sodium Sulfite (Na_2SO_4) in MilliQ water solution (0.01 g Sodium Sulfite into 200 µl MilliQ water). Sodium Sulfite is an oxygen scavenger and is what allows the PA gel solution to polymerize in plastic wells, which contain a lot of oxygen on their surface. The goal is to make your final gel solution contain ~30 mM of Sodium Sulfite.

10. Add 70 µl of 200 proof ethanol to N6. Mechanically mix (swirling) with pipette tip until well mixed. Turn micropipette up to ~90 µl slowly suck up N6/ethanol and add it to the 845 µl of acrylamide mixture.
11. Polymerize gels. Lay out all solutions and micropipettors with tips prior to beginning polymerization (For 24-well Seahorse plate: P20 = 15 ul, P200= 65-85 ul, P200 = 40 ul, P1000= 500 ul to mix).
    a. Add 65-80 µl of the Sodium Sulfite solution into 5 ml tube with acrylamide/N6 solutions. Polymerization will begin to initiate after this – work quickly after this point.
    b. Add 15 µl of the APS solution into 5 ml tube with acrylamide/N6 solutions.
    c. Use P1000 to mix gently, pipetting up and down, being careful not to introduce bubbles.
    d. Add 40 uL gel solution to the V28 Seahorse plate.
    e. Let gels polymerize for 30 minutes to 1 hour (the softer the gel, the longer the polymerization time).
12. While the gels are polymerizing, prepare a 0.1 mg/ml collagen in 50 mM HEPES (pH 8). Keep solution on ice.
13. As soon as gels polymerize, place collagen onto the gels. Use 50-100 uL per gel for the wells of the Seahorse plate.
14. Once collagen is on all gels, move the Seahorse plate to the refrigerator. Incubate at 4°C for at least 2 hours.
15. Prepare a 1:1000 solution of ethanolamine in 50mM HEPES (pH8). (1 uL of ethanolamine for each 1 mL of HEPES buffer).
16. After incubating the Seahorse plate at 4°C for 2 hours, pipette the collagen off the top of the gels. Place 50-100 uL of the 1:1000 ethanolamine solution onto the gels.
17. Incubate gels at room temperature for 30 minutes.
18. Rinse gels with PBS – add liquid to side of well, not directly onto gel.
19. Place gels in PBS + 2% P/S and store at 4C for up to 2 weeks.
20. UV gels for 1 hr prior to use.