

DISSERTATION

DEVELOPMENT OF SINGLE CELL SHAPE MEASURES AND QUANTIFICATION OF SHAPE CHANGES  
WITH CANCER PROGRESSION

Submitted by:

Elaheh Alizadeh

Department of Chemical and Biological Engineering

In partial fulfilment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2018

Doctoral Committee:

Advisor: Ashok Prasad

Jennifer DeLuca

Brian Munsky

Christopher D. Snow

Copyright by Elaheh Alizadeh 2018

All Rights Reserved

## ABSTRACT

### DEVELOPMENT OF SINGLE CELL SHAPE MEASURES AND QUANTIFICATION OF SHAPE CHANGES WITH CANCER PROGRESSION

In spite of significant recent progress in cancer diagnostics and treatment, it is still the second leading cause of death in the United States. Some of the complexity of cancer arises from its heterogeneity. Cancer tumors in each patient are different than other patients. Even different tumors from one patient could differ from each other. Such a high diversity of tumors makes it challenging to correctly characterize cancer and come up with the best treatment plan for each patient. In order to do that, a complex combination of clinical and histopathological data need to be collected. This dissertation provides the evidence that the shape of the cells can be used in conjunction with other methods for a more reliable cancer characterization.

In this study, experimental studies, numerical representation of the cell shape, big data analysis methods, and machine learning techniques are combined to provide a tool to better characterize cancer cells using their shape information. It provides evidence that cell shape encodes information about the cell phenotype, and demonstrates that the former can be used to predict the latter. This dissertation proposes detailed quantitative methods for quantifying the shape and structure of a cell and its nucleus. These features are classified into three main categories of textural, spreading and irregularity measures, which are then sub-categorized into nine different shape categories. Textural measures are used to quantify changes in actin organization for the cells perturbed with cytoskeletal drugs. Using the spreading and irregularity measures, it is shown that the changes in actin structure lead to significant changes in irregularity of the boundary of a cell and spreading of the cell and nuclei.

Using these methods, the shape of retina, breast, and osteosarcoma cancer cells are quantified and it is shown that the majority of cells have similar changes in their shape once they become cancerous. Then, a neural network is trained on the shape of the cells which leads to an excellent prediction of class of cancer cells. This study shows that even though cancer cells have different characteristics, they can be categorized into clinically relevant subgroups using their shape information alone.

## ACKNOWLEDGMENTS

I have been lucky to be trained by several amazing teachers and supervisors during my education. I am using this opportunity to express my gratitude to everyone who supported me throughout my educational journey!

Firstly, I am so grateful to my adviser Dr. Ashok Prasad, for his continuous support and valuable knowledge and time he shared with me. I would like to thank you for helping me to grow as a research scientist. Working under your supervision and guidance allowed me to grow greatly both scientifically and professionally.

I firmly believe that my research would have not been as successful without help of Jordan Castle, Wenlong XU, Katherine Schuamberg, Jackie Foss, Analia Quirk, Cameron Taylor, and Samanthe Lyons. Thank you all for your help. I also want to thank my committee Dr. Jennifer DeLuca, Dr. Brian Munsky and Dr. Christopher Snow for your time and guidance.

I am deeply thankful for my parents who supported me from the moment they taught me how to hold a pen till this moment. Thank you to my mom, dad and my brother for believing in me and helping me to stay strong.

Finally, last but by no means least, I would like to thank my husband, Soran Shadman, for his endless patience, love, and encouragement. Many thanks to you for supporting me spiritually and scientifically throughout my PhD study.

## DEDICATION

This work is dedicated to my uncle, Majid Azizzadeh Niari, whom I lost to cancer when he was 29. He was diagnosed only one week after he got married and fought it for nine months. He was not the only one who I lost the day he passed away. After his tragic death nobody was the same.

I deeply wish no one is made to watch their loved ones to suffer....

## TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iii
DEDICATION.....	iv
Chapter 1:           Introduction.....	1
1.1 The relevance of cell shape.....	1
1.2 Cell shape encodes and affects cell state.....	2
1.3 The measurement of cell shape.....	9
1.4 Mechanism of action and the control of shape.....	11
1.5 Dissertation Outline.....	14
Bibliography.....	19
Chapter 2:           Development of shape quantification methods and testing their performance on the cells perturbed with different cytoskeletal drugs.....	25
2.1 Introduction.....	25
2.2 Experimental Methods.....	27
2.3 Imaging methods.....	27
2.4 Data analysis methods.....	28
2.4.1 Quantifying changes in multidimensional shape space.....	28
2.4.2 Pearson correlation.....	28
2.5 Results.....	29
2.5.1 Drugs which directly perturb microtubules, myosin-II or actin leave a unique signature on structure and shape of a cell.....	29
2.5.2 Developing Shape quantifiers:.....	29

2.6 Conclusion:	36
Bibliography	40
Chapter 3: Changes in cell shape are correlated with metastatic potential in murine and human osteosarcomas	44
3.1 Summary	44
3.2 Introduction	44
3.3 Results	47
3.3.1 Pairwise Comparisons: The four paired cell lines demonstrated two distinct trends of cell shape changes	47
3.3.2 Highly metastatic cells differ in roundness, elongation and variability of perimeter	48
3.3.3 Highly Metastatic Cell Lines Show Shape Differences in the Nucleus	49
3.3.4 Multivariate Techniques show overlapping but distinct cell populations	50
3.3.5 Identification of Cells by Machine Learning	51
3.4 Discussion	51
3.5 Materials and Methods	54
3.5.1 Cell lines and cell culture	54
3.5.2 Immunofluorescence microscopy	55
3.5.3 Preparation of surfaces	55
3.5.4 Contact angle measurements	55
3.5.5 Cell volume measurement using Scepter Cell Counter:	56
3.5.6 Image Processing	56
3.5.7 Data Analysis	57
Bibliography	69
Chapter 4: Measuring systematic changes in invasive cancer cell shape using Zernike moments	75

4.1 Summary.....	75
4.2 Introduction.....	75
4.3 Experimental.....	79
4.3.1 Cell Lines and Cell Culture: .....	79
4.3.2 Preparation of surfaces .....	79
4.3.3 Immunofluorescence microscopy .....	79
4.3.4 Zernike moments .....	80
4.3.5 Zernike Moments Calculation .....	81
4.3.6 Principal Component Analysis (PCA).....	82
4.3.7 Neural Network Classifier.....	83
4.3.8 Weight Analysis in neural networks.....	84
4.4 Results and discussion .....	85
4.4.1 Paired cell lines are distinguishable and cluster similarly. ....	85
4.4.2 Cell shapes of more invasive cancer cells are distinguishable from their less invasive parental line. ....	85
4.4.3 Three pairs of cell lines showed the same kind of shape differences in multidimensional Zernike moment space .....	86
4.4.4 Cell shapes can be classified accurately for type 1 shape changes collectively, but not for mixed populations.....	87
4.4.5 Cell shape changes on substrates are similar for hydrophilic to hydrophobic substrates, and are more significant for type 1 invasive cancer cells.....	88
4.4.6 Even order Zernike moments contribute more to the classification. ....	89
4.5 Conclusion .....	89
Bibliography .....	100

Chapter 5:	Quantifying single cell shape changes with cancer progression.....	104
5.1	Introduction.....	104
5.2	Methods .....	106
5.2.1	Cell culture: .....	106
5.2.2	Image Processing.....	107
5.2.3	Shape quantification: .....	107
5.2.4	Statistical testing of differences.....	108
5.2.5	PCA analysis .....	108
5.2.6	Machine Learning Method: .....	109
5.3	Results and discussion .....	110
5.3.1	Analyzing the variability of cell shape quantifiers for different cell lines.....	110
5.3.2	Analyzing shape changes between cell lines within each cell type .....	112
5.3.3	Cancer progression is accompanied with similar shape changes across most cancer types.....	113
5.4	Conclusion .....	114
Bibliography	.....	124
Chapter 6:	Conclusion.....	126
Appendix I:	supplementary information for chapter 2.....	129
Appendix II:	supplementary information for chapter 3 .....	145
Appendix III:	supplementary information for chapter 4.....	158
Appendix IV:	supplementary information for chapter 5.....	165

## Chapter 1: Introduction<sup>1</sup>

### 1.1 The relevance of cell shape

All of us, scientists and laypersons alike, are fascinated with biological form! The intricate patterns of a leaf vein, the geometric arrangement of sepals and petals in a flower, the bizarre universe of bacterial shapes are examples of a bewildering variety of shapes and forms that are encountered in living organisms. It has always been natural to ask whether these shapes performed a function and how specifically did they arise. Just over a century ago, D’Arcy Wentworth Thompson published his famous book, *On Growth and Form*, which argued that physical and mathematical laws are needed to explain biological forms. D’Arcy Thompson was also one of the first scientists to wonder whether there is a physical and mechanical basis for the shape of cells. We now realize that cell shape is formed out of the subtle interplay of genetics with physics, but we are still searching for fundamental principles that govern that interaction. However we have accumulated a lot of information about how form and function are related for single cells, allowing us to begin to use shape and form both as readouts of phenotype as well as aspects that could be used for engineering specific cellular responses. This review will provide an overview of the recent progress we have made in interpreting and controlling mammalian cell shape, specifically adherent cells on surfaces. We should note here that cell shapes in bacteria are no less bewildering and fascinating than those of mammalian cells, but they will not be our concern here.

Cells of specialized tissues look quite different, and clearly that difference is related to function. The shapes of kidney or liver cells for example as opposed to the shapes of neurons are a good example. However when cultured in identical conditions a family of adherent cells will adopt a number of seemingly random morphologies. Over the last few years we have begun interrogating the seeming randomness of these morphologies and interpreting the information they carry. We present below a brief review of the current state of the field of interpreting, understanding and controlling cell shape.

---

<sup>1</sup> A review paper based on this chapter, co-authored with my advisor Ashok Prasad, has been submitted for publication, and is under review at the time of submission of this thesis.

## 1.2 Cell shape encodes and affects cell state

In modern cell biology perhaps the first theoretical argument for the determination of cell shape was made by Donald Ingber and his co-workers, as a natural extension of the tensegrity hypothesis. Inspired by the visionary architect, Buckminster Fuller, the basic tensegrity hypothesis applied to mammalian cells held that the shape of the cell was determined by a mechanical balance arising from contractile actin forces acting on a rigid tensed scaffolding formed by the microtubules[1]. In a review paper from 1989 (printed in an edited volume that summarizes the current understanding of cell shape in 1989[2]), Ingber and Folkman put forward indirect evidence that cell shape could regulate function, such as the relation of spreading with anchorage dependent growth (discussed below) and modulation of shape during angiogenesis[3].

How can shape be explained by tensegrity? Using physical models constructed with elastic cords and poles, Ingber and colleagues argue that an unsupported tensegrity structure is spherical, like cells in suspension, because that is the minimum stress structure. When such a structure is attached to a substrate that resists shear, spontaneous flattening takes place. If the structure is attached to a malleable substrate however, it rounds up, and if a “nucleus” is included within the model, the nucleus also deforms when the structure is attached to a substrate. They suggest that the proteins involved in these tensile cytoskeletal structures may show significantly altered thermodynamic properties as compared with proteins not experiencing significant stresses. In fact we now know that there are many proteins that are mechanosensitive, specifically proteins that are stretch sensitive, that are implicated in the conversion of mechanical information into cellular outcome in cells [4],[5]. It was also pointed out that mechanical forces could be transmitted directly to the nucleus since intermediate filaments provided a direct connection from the cell membrane to the nuclear envelope. Thus many of the themes that would emerge in later work were anticipated early on, largely arising from the natural implications of the recognition that cell shape determination must be, at least in part, based on physical force balances.

Changes in cell shape had been noticed early for the case of cells growing in contact, and were used to explain contact inhibition of growth[6], since as cells became more crowded, they became more compact in shape, which limited the uptake of mitosis promoting growth factors. Shape was also invoked to explain anchorage-dependence of untransformed mammalian cell growth on the basis of the observation that when adhesion was limited by coating the substrate with different concentrations of poly(2-hydroxyethyl methacrylate) (or poly(HEMA)), the more spherical cells showed lower rates of DNA synthesis, as measured by 3H-thymidine incorporation assay[7]. However it was

pointed out that growth appears to be affected not by cell shape but by total spread area since cells of different shapes but with the same spread area showed similar rates of DNA synthesis [3]. Later studies noted that the nucleus was also markedly flattened and spread in proliferating lens epithelial cells, while non-proliferating spherical cells had a more rounded nucleus[8], suggesting presciently that nuclear shape may play a role in the relation of shape with proliferation. Other explanations for cell shape focused on balances of forces exerted and experienced by the cell. Thus the basis for the fried egg shape of many adherent cells was studied using a scanning acoustic microscope in Ref [9] and a force based explanation of the cross-sectional outline sought in the balance between actomyosin contractility at the periphery and hydrostatic pressure on the cell interior due to the cortical actin later on the top.

Techniques of surface lithography began to be adopted in biology in the 90's and the use of micro-contact printing of adhesive proteins on micron sized islands allowed for direct manipulation of cell spread area, and later even shape[10]. Ingber's group took a leading role in this and showed that the extent of cell spreading directly correlated with the extent of proliferation when cells were cultured on adhesive islands of various sizes, from those that restricted cell spreading to those that freely permitted it. In the early 2000s a seminal paper from Christopher Chen's group demonstrated the importance of cell spreading for differentiating mesenchymal stem cells (MSCs)[10]. The hypothesis tested was a direct consequence of the earlier work on cell density and cell shape. The authors noted that mesenchymal stem cells were reported to differentiate into adipocytes, osteoblasts and chondrocytes only when plated at the appropriate densities [11]. They hypothesized that the effect of density was via its effect on cell shape and set out to test this hypothesis directly by restricting the shape of cells using a micro-patterning technique. They reported that human MSC (hMSC) cells in osteoblast and adipocyte differentiation media differentiated into osteoblasts when plated at low density and into adipocytes when plated at high density after 4 weeks of culture. Significantly, cells retained a memory of the earlier culture condition when re-suspended and re-plated for at least 1 week. They then micro-contact printed fibronectin on PDMS substrates to generate isolated fibronectin "islands" of either 1024, 2025 or 10,000  $\mu\text{m}^2$  in area, surrounded by non-adhesive regions. The hMSC cells were then cultured on these different surfaces in mixed media. They found that adipogenesis occurred only on small islands, osteogenesis on large islands and intermediate sized islands supported both lineages. By using a Rho-associated protein kinase<sup>2</sup>

---

<sup>2</sup>Replace by: ROCK is a kinase that is a downstream effector of the Rho GTPases, which are central controllers of cytoskeletal properties.

(ROCK) inhibitor they showed that adipogenesis is controlled by cytoskeletal tension, and RhoA<sup>3</sup> regulates the commitment switch between osteogenic and adipogenic fates. Thus this paper established a link between cell shape, mechanical tension and lineage commitment. This was one of the early papers that provided direct evidence for a role for mechanotransduction in lineage commitment, and helped inspire a slew of later work that cultured MSCs on patterned substrates of different kinds and studied their impact on lineage commitment. In particular, MSCs were found to be sensitive to the shape of the patterns they were cultured on, thus again underlining the role of cell shape in changing cell phenotype[12]. Significantly however, while the earlier work could only address the spread size of cells, Ref. [12] from the group of Milan Mrksich, showed that apart from size, shape was also a factor in determining function. This paper showed that MSC differentiation was sensitive to the curvature of the perimeter of the adhesive islands they were cultured on, as well as the aspect ratio. Islands with shapes with higher local curvature generally promoted a slightly larger differentiation along the osteogenic pathway. Similarly, islands with a higher aspect ratio also were found to be osteogenesis promoting. The difference was attributed to differences arising from changes in the tension properties of the actin cytoskeleton, with high tension actin structures promoting osteogenesis. Microarray analysis showed that the high tension phenotype was associated with a higher expression of MAP kinase pathways and noncanonical members of the Wnt pathway, including RhoA and ROCK. Cells on rounded flower-like shapes that tended to prefer the adipogenic pathway showed elevated expression of Wnt inhibitory molecules.

A few years previously, in a seminal paper, Denis Discher's group had shown that MSCs were sensitive to the mechanical properties of the substrate they were cultured on [13], which sparked off many research programs trying to control MSC differentiation using substrate properties. While there is a significant body of work generated on the sensitivity of stem cells to mechanical cues [14], much of this literature does not consider cell shape as a causal factor in the lineage commitment process. Nevertheless cells cultured on substrates of different stiffness and on patterned substrates of different properties do adopt shapes that are often quite different from cells cultured on flat surfaces. Thus cell shape may also be a factor determining the cellular response to substrate properties. Similarly, cell shape may possibly be a major factor explaining MSC differentiation into osteoblasts on certain patterned surfaces with random disorder [15].

---

<sup>3</sup> Small GTPase protein associated with stress fibers formation and actomyosin contractility.

While the literature in the early 2000s had firmly established the importance of cell spreading, there were still relatively little work on trying to understand cell shape. However, that required the establishment of high-throughput methods of image analysis, improved microscopy techniques and modern techniques of analysis of large data sets. One of the early noteworthy papers in this regard was from Nobert Perrimon's group [16], which linked cell shape to signaling proteins and gene expression, using high-throughput methods. They performed a genetic screen of 249 gene-overexpression treatment conditions using the *Drosophila* BG-2 cell line as their model system. They imaged over 12,000 cells and carried out quantitative morphological analysis, and then used neural networks to classify the morphological signatures of different treatment conditions into different clusters. More interestingly, they could classify morphological signatures according to the genes that were perturbed, thus demonstrating a direct connection between specific proteins, their associated pathways and cell morphology. For example, cells over-expressing RhoGEF3 as well as p190RhoGAP, SCAR, slingshot and a few other proteins all clustered together and showed a very round morphology with few protrusions of any kind, suggesting that these proteins, or others downstream of them, suppressed cell protrusions. Large flattened cells with extensive lamellipodia resulted from the over-expression of SIF, RacV12 and RacF28L, which leads to over activation of the Rac pathway. In another interesting related paper [17], the authors asked whether morphological signatures are sufficient to make inferences about signaling pathways. They chose a family of 13 RhoGAPs, and performed all single knockdowns and almost all double knockdowns of these proteins using RNAi. They then compared the morphological features of the knockdown cells with the morphological features of the same cells with RhoGTPase overexpression, using a statistical algorithm to infer synergistic interactions as well as non-interactions between the RhoGAPs. Their method performed remarkably well, underscoring the conclusion that the relation between phenotype and morphology is robust enough to make inferences about signaling pathways, at least for major cytoskeleton associated proteins. This work contributed to building a framework for recognizing cellular phenotype using "high-content" images of cells, i.e. images of cells with the cytoarchitecture labeled and visible [18].

Cell shape and the cytoarchitecture are of course closely related, since it is the cytoskeleton that controls shape. However, in practice, high-content images may include DNA stains and staining of organelles, which of course contain significant phenotypic information too. For example, Refs [19],[20] used morphology, staining intensity and spatial distribution of the cellular nuclei, microtubules and Golgi compartments to screen biologically active

molecules. However since this information is different from that contained in cell shape and the cytoskeletal architecture, and a recent review exists [21], we do not discuss this branch of research further.

There have been a number of empirical correlative studies that have used morphological characteristics to predict some functional attribute of the cells being studied, especially in the context of cancer. Here we highlight some of the noteworthy ones that specifically focused on studying morphological measures.

One of the significant changes in morphology at the single cell level is associated with the Epithelial to Mesenchymal transition (EMT). Ren et al [22] studied measures of shape that best characterize A549 cells after undergoing transforming growth factor- $\beta$ 1 (TGF- $\beta$ 1) induced EMT. In their list of 29 morphological features they find roundness and the radius ratio as the best able to distinguish control cells from the cells that underwent EMT. Metastasis is of course closely related with EMT, and a number of studies have asked whether metastasis is predictable using morphology. In our group we studied spread cell shape of eight osteosarcoma cell lines, four of which were highly metastatic and four were of low metastatic capability [23],[24]. Using geometric measures of cell area, elongation and irregularity, as well as a representation of cell shape using Zernike moments, we showed that metastatic osteosarcoma cell lines appear to be characterized by two types of shape changes, that we called Type I and Type II. In Type I changes, metastatic cells are smaller, more elongated and more irregular while Type II metastatic cells are characterized by a large increase in cell size, accompanied by significant rounding of the spread cell shape. Similar trends were observed by using Zernike moments. We found that a neural network could predict the metastatic capacity of cell lines using morphological markers with almost 99% accuracy for a related pair of cell lines, and even more remarkably, with about 90% accuracy for all the cell lines from Type I [23].

The Bakal group has produced a number of papers of note that we discuss briefly below. Continuing with the theme of morphological changes in cancer, Sailem et al [25] developed a quantitative and predictive relationship between cancer gene expression, cell shape and tumor clinical phenotypes, using 18 breast cancer cell lines. They discovered that there were 504 genes in their dataset that had a significant correlation with the morphological measures that they used. Using Gene Set Enrichment Analysis (GSEA) many of these genes were found to be connected with genes that are known to regulate cell morphogenesis, migration, differentiation and adhesion processes. Of genes known to be involved in EMT, only one, SMAD3, was found to correlate with breast cancer cell shape, specifically its aspect ratio. They used databases of known protein interactions to build a shape-gene interaction network, which incorporates the genes that are directly linked to specific shape parameters. Finally, they also looked for subsets of

genes, described as morphological metagenes, whose expression level correlated with a specific shape feature by using multiple regression analysis. Analysis of the diagnostic importance of the metagenes using breast cancer patient data revealed that the metagenes did correlate with tumor grade[25]. NFkB was found to be related with morphology too, and in another beautiful paper [26], the authors found that NFkB activation leads to distinguishable morphological signatures that can be used for prediction. In particular they find that the nuclear/cytoplasmic NFkB ratio is positively correlated with ruffiness, which measures variations in membrane intensity, and negatively with cell and nuclear area, and that other parameters have a non-linear and complex relationship with NFkB ratio.

One of the powerful tools used by the Bakal group is the use of large RNAi screens to examine the effects of knockdowns of individual genes on morphology, allowing them to ask questions about the correlation between gene expression, morphology, and metastasis that were hitherto difficult. One of their interesting observations is that instead of leading to novel cell shapes, RNAi knockdowns lead to enrichment of certain morphologies, suggesting that stochastic gene expression in control cells leads to a diverse population of cell shapes. In a paper published in 2017 [27] they use the RNAi screens to discover Rho GTPase regulators that lead to significant changes in the frequency of their observed 5 discrete cell shape categories. As an example they show that depletion of a RhoGEF called ECT2 increases the population of large spread cells many-fold for LM2 and MDA-MB-231 breast cancer cell lines. They can also tune the YAP/TAZ nuclear localization (specifically the nuclear/cytoplasmic YAP ratio) by depleting different genes. YAP (Yes-associated protein) and its homolog TAZ, are Hippo pathway genes that play a significant role in transducing mechanical signals to the nucleus [27], [28]. Nuclear localization of these transcription factors is a marker for activation of this pathway, and in another paper [28] they model the relation between the YAP ratio and cell shape using image-based analysis and regression models. To identify the proteins that directly regulate the YAP ratio, they perturb the localization of YAP using drugs or siRNA, and record the YAP ratio and cell shape. Using cell shape and their regression models, they then predict the YAP ratio for cells with specific genes knocked-down by siRNA, and specifically look for cases where the prediction does not match the observed YAP ratio. These cases would then correspond to either upregulation or downregulation of YAP localization. They find that LATS1 and LATS2 directly inhibit YAP localization, and Cdc42, Rac1,  $\beta$ -PIX, PAK2, and PAK4 are direct activators of YAP localization. In MCF10A human mammary gland cells the YAP ratio is correlated positively with cell area, percent protrusion, and protrusion extent and negatively correlated with Nucleus to cell area, neighbor fraction, and local cell

density. However, in four variants of MDA-MB-231 human breast tumor cells, the YAP ratio is uncoupled from cell-ECM contact and area but not from cell density [27], [28].

Cancer cells are known to show two forms of motility, amoeboid and mesenchymal, and both forms are associated with characteristic shapes. It had been shown previously that melanoma cells appear to spontaneously shift between the amoeboid and mesenchymal shapes when plated in 3D collagen matrices [29]. A deeper analysis of these conversions in PTEN-null WM266.4 melanoma cells led to the identification of different routes in shape space that cells use when they are transitioning between shapes, as well as specific Rho GTPase regulators associated with these routes.

Another comprehensive empirical study [30] concentrated on pancreatic ductal adenocarcinoma, a devastating human cancer with a very poor prognosis. Using 11 PDAC cell lines, as well as 10 breast cancer cell lines, they developed a high-throughput imaging and computer vision and analysis system and collected morphological data for thousands of cell lines. However in contrast with other work, this paper did not find a morphological signature of metastasis, except for a lower degree of heterogeneity, suggesting that PDAC cells may not show many shape changes.

Apart from metastasis, one could ask whether chemo-resistance information is encoded in shape. In two publications, Paqualato and co-authors [31],[32], study the relation between cell shape of a chemo-resistant and a chemo-sensitive colon cancer cell line. They find significant differences between the shapes of these two cell lines. One possible explanation of these differences could be in different shape characteristics of cancer stem cells, which are usually believed to be more chemo-resistant.

Apart from cancer, prediction of the lineage commitment of MSCs has also been shown to be reflected in changes in morphology. Matsuoka et al. use cell shape to non-invasively predict osteogenic differentiation potential of hMSCs, based on phase contrast images, quantified by nine morphology parameters and bone formation markers to build a machine-learning model that was capable of predicting osteogenic differentiation of hMSCs from new patients, with excellent accuracy [33]. The role of shape may be important in maintaining the stem cell niche, possibly through the effect of topography on shape, since MSCs cultured on small adhesive islands were shown to maintain the multipotent stem cell phenotype more reliably than those on smooth surfaces [34].

These image-based methods to discover patterns in biological data are not yet standardized enough for wide adoption. Part of the reason is that more data and more mechanistic details are required before specific morphological

changes are identified with specific stimuli or phenotypes. However the plethora of papers also use a plethora of shape parameters and data analysis methods that can be bewildering at first sight. A recent review of data analysis methods gives some hope of standardization of methods [35]. Here we briefly review some of the shape parameters that have been adopted in others, and our, work.

### **1.3 The measurement of cell shape**

Cell shape can be visualized using both phase-contrast [36], and fluorescence, though phase contrast is typically not efficient in picking up the cell perimeter accurately, and fluorescence microscopy using labelled actin provides better detail of the outline of the cell and the actin structures inside the cell. The cell outline is quantified by using a segmentation algorithm that identifies the cell boundary, and actin or other cytoskeletal structures can be captured using intensity-based grey-scale images (Figure 1-1).

The two dimensional outline and the grey-scale image now need to be converted into shape quantifiers. There are two broad classes of measures of shape quantifiers. The first class consists of measures of some user-defined aspect of shape or texture. We will refer to these measures as geometric or textural quantifiers and they include standard measures such as the cell area, perimeter, aspect ratio, circularity, fractal dimension of the perimeter, roughness of the texture, polarization of the texture, etc. We discuss some of these measures in more detail below. The second class of shape quantifiers is based on a basis-function, or similar, representation of the shape. In general, the two dimensional outline of the cell can be thought of as a function in two dimensions and expanded in some orthonormal series expansion such as Fourier series. The coefficients of the expansion then are the numbers representing the shape of the cell. The advantage of this method is that if the series representation is sufficiently compact, the shape of the cell is completely represented in a vector of just a few numbers. However, the disadvantage is that we often cannot interpret what changes in those coefficients are telling us about shape changes. Also, we do not know a priori whether the characteristics of shape being picked up by these abstract measures are the right measure for distinguishing between biological shape changes.

The two most common basis function expansions used have been Fourier series and Zernike polynomials based expansions. Fourier series are used to decompose a 1-dimensional (1D) signal, and thus the cell perimeter first needs to be rendered into two 1D vectors. Two alternative ways of doing this are by either separately decomposing the X and Y components of the pixels in the boundary in Cartesian coordinates, or their  $\rho$  and  $\theta$  components in the Polar coordinates into linear combinations of sin and cos basis functions with different frequency. The vector of

Fourier coefficients in this expansion can be used as a shape descriptor for that cell. On the other hand, Zernike polynomials are used to decompose a 2-dimensional (2D) image into linear combination of Zernike polynomials, i.e. each polynomial is a 2D image-based basis function. The coefficient of basis function in this expansion, called Zernike Moments, can be used as descriptor of the cell image. The Zernike polynomials are orthonormal polynomials on the unit disk, and have the advantage of being naturally rotation-invariant [37]. However they are hard to interpret in an intuitive way.

In general, different quantifiers have been used in the literature to ask different questions. It is possible however to say something about the properties that good shape or morphology quantifiers should possess. Pincus and Theriot [36] introduce three criteria for good shape quantifiers. The first criterion is fidelity, i.e. discarding minimum possible true information, and not adding noise to the data. This can be tested by reconstructing the shape from the quantifiers. The second criterion is that the quantifiers should capture variations that are biologically important, and discard unimportant variations. Note however that not all changes are necessarily going to be reflected in morphological changes. The third and last criterion is that each quantifier should be individually meaningful and interpretable. However it is not clear whether this is a useful criterion, since it rules out basis function expansions such as Zernike moments that work quite well. Pincus and Theriot [36] go on to compare different shape quantifiers of shapes of *Caulobacter Crescentus* (CC), Madin Darby canine kidney (MDCK), and Nicaraguan cichlid epithelial keratocytes (NE) cells. They compare Fourier basis functions, Zernike moments as well as PCA and ICA representations of the underlying shape data, as well as of the Fourier and Zernike reconstructions. Their overall conclusion is that the PCA representation of shape appears to perform the best. However they were strongly limited by the choice of only a 16 component feature set. Within such a small representation, methods such as Zernike moments will not work well; we found that about 30-50 Zernike moments yield both fidelity and discrimination power [23]. However we have also found that radial-based Fourier representations provide the most compact high-fidelity reconstructions of cell shape as shown in Figure 1-2.

While basis function representations provide complete information, the geometric measures, while suffering from a certain ad hoc character, do have the property of being information-rich, that is they can encapsulate many features of shape into one measure. As such there is considerable value in arriving at a standard list of image features that should be used by the community, even while evaluating newer descriptors. The Bakal group has used a variety of geometric measures, and has argued for the recognition of a limited number of shape types. The six shape types

that they specifically identify are ellipse, small round, large round, star, spindle and tear drop. However in our experience two-dimensional mammalian cell data does not always show the presence of specific separable shape classes, and the shape of cells does appear to belong to a continuum. It is interesting that melanoma cells were found to possess identifiable shape classes in 3d culture but not in 2d culture [29], suggesting that perhaps shape classes may exist in the native environment of the cell.

Our work has also shown the importance of textural measures that pick up aspects of the actin distribution, especially its linear or circular polarization and heterogeneities in its density. As an example, we found it necessary to distinguish between the cells with homogenous actin distribution versus the cells in which actin is organized concentrically around the nucleus. The shape feature we developed to distinguish between these cases is based on dividing the gray scale image of a cell into ten concentric bands[38]. The bands are determined in the following way. All the lines that connect center of mass of a cell to the cell boundary is divided into 10 equally spaced lines. Band  $n$  is where all the  $n^{\text{th}}$  lines reside. Then the average actin intensity in each band is calculated. We also quantify the homogeneity of the actin distribution by measuring variations in actin fluorescence intensity. It is worth mentioning that although these textural measures were applied only to actin structure, they should be relevant for other cytoskeletal polymers too. Detailed discussion of the scores of different measures that are typically used are reported in the next chapter.

#### **1.4 Mechanism of action and the control of shape**

The basic tool for controlling cell shape is by the use of patterned substrates. Some types of patterned substrates may control cell shape directly, such as fibronectin islands of different shapes, which cells must adopt in order to adhere successfully. Other types of patterned substrates provide contact guidance cues, such as grooves for example, that adherent cells will usually align along. A third type of patterned substrates control the adhesion of cells to surfaces by controlling the size and the distribution of focal adhesions that cells can form. These patterned substrates affect cell shape but may also affect mechanotransductive signaling by modulation of the focal adhesion properties.

We still lack complete understanding of the effect of cell shape on gene expression and cell phenotype. The hypothesis to which we subscribe is that cells adopt shapes that correspond to the constraints of the substrates and their own mechanical properties. When cell spreading is constrained in a particular way, it leads to changes in signaling from focal adhesions as well as mechanical forces acting directly on the nucleus, transmitted through the cytoskeleton.

These mechanical forces could result in mechanotransductive signaling on the nuclear envelope as well as changes in gene expression patterns due to changes in chromosome mingling.

Signaling through focal adhesions, involving Src kinases and the interplay between Rho and Rac activation, and its role in cell spreading have been extensively reviewed in the past[39]–[41]. These have also formed the basis of biophysical studies of cell spreading (reviewed in [42]) and mathematical models of cell shape based on the interplay of Rho and Rac signaling [43], [44]. However the link between the organization and contractile forces exerted by the cytoskeleton and the change of state of the cell is still being uncovered. A major role here appears to be played by Yap/Taz signaling, that relays matrix and shape dependent signals through Rho and actin contractility to the cell nucleus [5], [45], [46]. Even more recent work, discussed below, has led to a new hypothesis that links cell shape to state via changes in chromosome organization.

Work from the Shivashankar laboratory has established a possible functional link between shape and phenotype by showing that these changes appear to work through changes in nuclear morphology [47]. Adherent cells were observed to have a perinuclear actin cap just below the apical surface that exerted compressive forces on the nucleus, but the structure of this actin cap was modulated by cell geometry. In particular cells cultured on square islands showed a mesh-like actin structure while cells cultured on an elongated islands displayed distinct parallel fibers that led to a flattened and elongated nuclear shape [48]. In order to study the effects of cell shape on gene expression, they cultured NIH 3T3 cells on fibronectin islands of triangular, spherical, square and rectangular shapes, and carried out transcriptomic analysis using microarrays [49]. They found significant differences on gene expression between cells cultured on different shapes of the same area as well as different areas of the same shape. They found a strong correlation between the cell shape and the nuclear height and the projected nuclear area. Cell size and geometry was found to regulate both nuclear volume and histone acetylation. While actin contractility can act through direct mechanotransductive signaling, they postulated an interesting and significant hypothesis that changes in nuclear volume and geometry affect gene expression by changing the intermingling of chromosome territories[50]. This hypothesis is built upon recent work on chromosome territories that has shown that the spatial organization of chromosomes within the eukaryotic nucleus is an important part of nuclear architecture with functional implications for cellular phenotype (reviewed in Refs. [51], [52]). In a set of papers in 2017, significant correlations were seen between gene expression and the organization of chromosome territories in undifferentiated and differentiating stem cells [53], [54], and cell geometry was shown to lead to repositioning of the chromosome territories, which correlated

with changes in gene expression [55]. While the principles of the role of cell geometry on chromosome intermingling are still being worked out, there are a promising and exciting hypothesis that would help lay the ground for the rational engineering of cell shape to influence phenotype.

Patterned substrates have been reviewed extensively [56]–[58], including notably their possible role in cell based therapies [59] and we will not review the literature again, but highlight some of the work relevant for controlling cell shape. Perhaps the best-known phenomenon is that of contact guidance, wherein cells grow along grooves present on the substrate. Contact guidance has been deployed extensively in attempts to control axonal growth from neurons [60], which is important, for example, in neural prostheses. It has been shown in multiple studies that neurons use contact guidance cues from the underlying topography to determine polarization and axon growth. More remarkably, neuronal markers were found to be upregulated when MSCs as well as ESCs were grown on nano-grooved surfaces [61], [62].

Nanopatterned substrates that modulate the size and the shape of adhesive spots exercise control over cell shape and function by constraining focal adhesions. A detailed study of this effect, reported in Slater et al. [63], showed that small spots of under-100 nm in diameter changed the distribution of focal adhesions and led to enhanced migration, and a very dynamic morphology as compared with larger spots (~200nm and 400nm) and smooth controls. Nanopatterned surfaces have been widely used to control osteogenic differentiation of mesenchymal stem cells, reviewed in [57] and [64].

An exciting advance in exerting fine control over cell shape by patterning surfaces is based on a method of patterning substrates using computer-guided laser ablation rather than lithography masks, which allows much greater control over the position, size and geometry of adhesive spots than the traditional technique. This technique can be used for biomimetic patterning by using fluorescently labelled focal adhesions or the actual morphology of spread cells to make either biomimetic adhesive spots, or biomimetic adhesive islands, which recapitulate the typical shape of any particular cell or cell type [65]. This technique was used to recapitulate the patterns corresponding to a typical adipocyte, and the differentiation of human MSCs cultured on these patterns was compared to those cultured on square and circular patterns as well as smooth controls [66] It was found that when treated with mixed media that promotes both osteogenesis and adipogenesis, the pattern of focal adhesions that were derived from adipocytes showed a much higher percentage of adipogenesis and no osteogenesis, compared with smooth controls. Interestingly the square and circular patterns too showed an intermediate level of adipogenesis. These results in outcome were reflected in the

nuclear to cytoplasmic ratio of the adipogenic regulator, PPAR $\gamma$ . Cell-type mimetic patterned substrates therefore can be used for much subtler control of differentiation outcome than simple geometric patterns.

As our review of the literature has indicated, cell shape is closely linked with cytoskeletal properties and the cellular response to the substrate, both of which reflect cell state. Furthermore, the causal links run both ways and changes in cell shape lead to changes in cell state in ways that appear predictable. Changes in cell shape too appear informative about the changes in state, and may become tools in the future, both for understanding changes in cells upon treatments as well as for guiding therapeutics in diseases like cancer. However we still lack causal understanding of how cell shape is determined, even on simpler two-dimensional substrates, though many of the links have become clearer in the last few years. In summary, an explosion of work in recent years has shown us that cell shape should be treated as a complex but predictable emergent property of cell state, whose understanding and control can have profound implications for deeper biological understanding as well as for biomedical engineering.

### **1.5 Dissertation Outline**

In this research, the feasibility of cancer phenotype identification using the shape of cancer cells is studied. Figure 1-3 demonstrates the foundation and aim of this dissertation. As presented in this figure, gene expression changes cytoskeletal properties which play a role in determining shape of a cell. In 2 dimensional cultures, substrate properties also change the cell shapes. It is shown that the role of cytoskeleton of the cell, substrate properties, and gene expression of a cell in final shape of the cell can be used to learn and predict the behavior of the cancer cells.

Due to the lack of appropriate and sufficient measures to quantify the shape changes for the cells under study, different quantification methods are developed in this dissertation to numerically represent the shape of the cells. These methods are detailed in Chapter 2. In brief, shape quantification methods are divided into three major group of textural, spreading, and irregularity measures which quantify gray scale image of the actin structure in a cell, the binary image of cell or nuclei, and radial information of the pixels in the boundary of a cell, respectively. Then, their performance on capturing shape changes on the cells perturbed by 6 different cytoskeletal drugs is explored.

In Chapters 3 and 4, the spreading measures are used to quantify shape changes of paired osteosarcoma cancer cell lines; each containing one high invasive line derived from low invasive line with selection for metastasis. Statistical analysis such as T-test and principal component analysis are used to visualize the shape distribution of

various cancer types and calculate the shape differences among them. Then, the shape information is used to predict the invasive potential of the cancer cells using neural network classifiers.

Chapter 5 uses the methods developed in the previous chapters and builds upon their results to characterize the shape changes in a larger number of cells with cancer progression. This chapter demonstrates that the shape changes of majority of cancer cells which are derived from different tissues (e.g. breast, retina or bone) follow a similar pattern with cancer progression.

This dissertation concludes with Chapter 6 in which the main findings are summarized, and future directions are proposed.

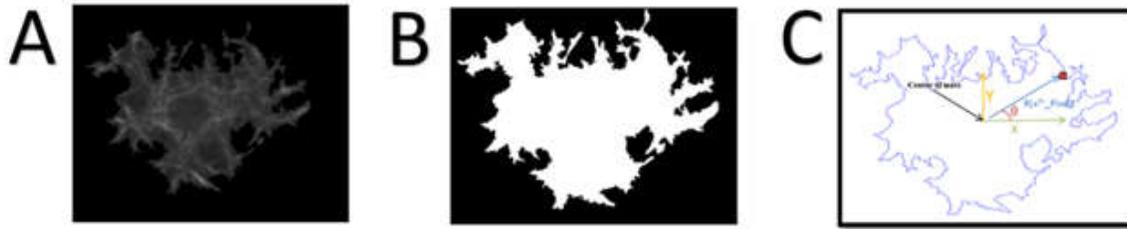


Figure 1-1 Representation of cell shape A. Gray scale images using labeled actin. Intensity of each pixel is recorded and represented as a greyscale intensity plot. B. 2D outline: Position of each pixel in the boundary is recorded in polar or Cartesian coordinates. C. Binary image: All the pixels in the image can take only two intensities. One intensity is for inside the cell and other intensity is for outside the cell.

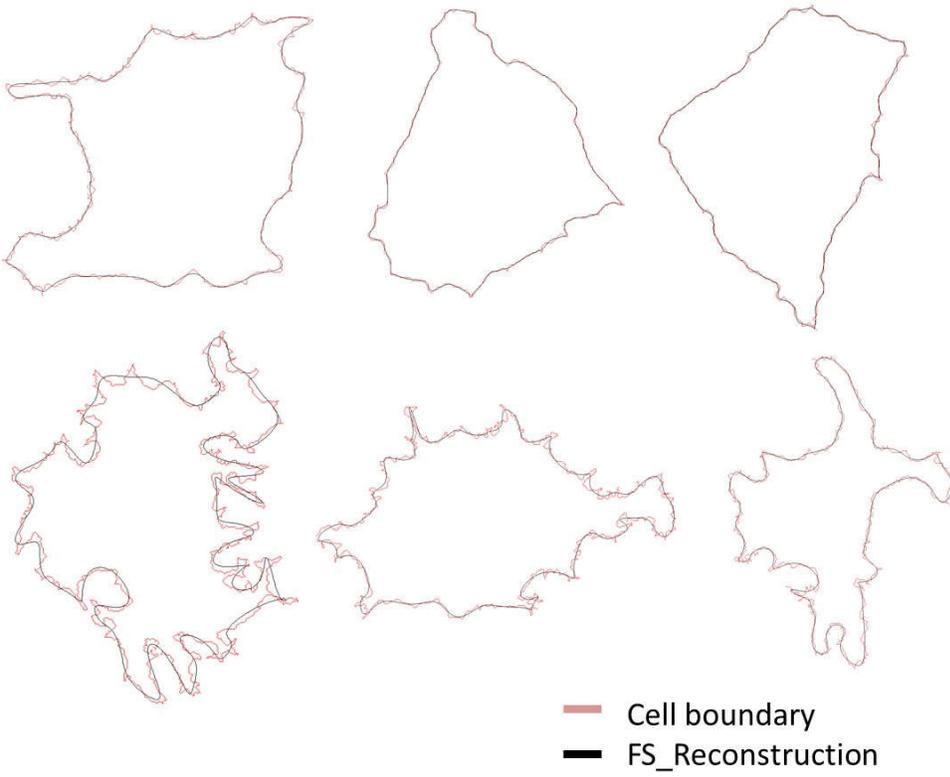


Figure 1-2 Reconstruction of the cells using Fourier decomposition. The red line is the actual boundary of the cell and black line is the reconstruction of the cell using the first 35 terms in the Fourier series expansion.

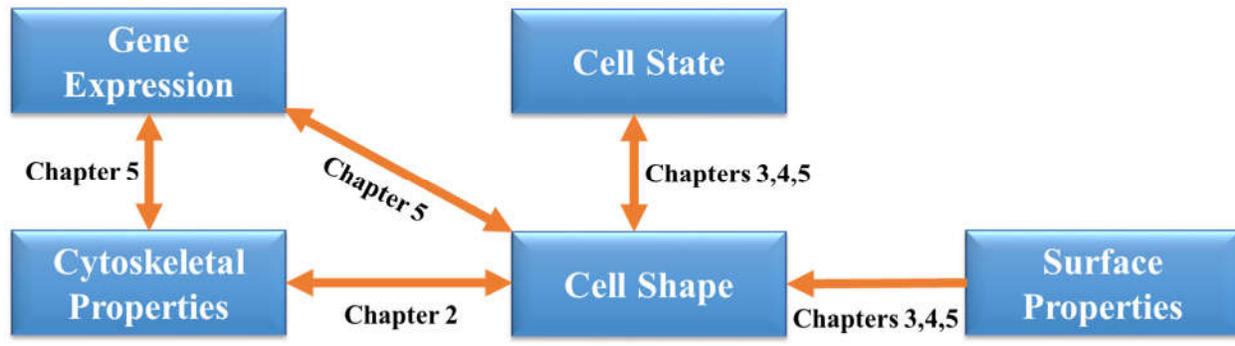


Figure 1-3. The foundation and aim of dissertation. In this dissertation, role of cytoskeleton, gene expression and substrate properties on cell shape is explored and it is used to predict cell behavior with application in cancer diagnostics.

## Bibliography

- [1] D. E. Ingber, "Cellular tensegrity: defining new rules of biological design that govern the cytoskeleton," *J Cell Sci*, vol. 104 ( Pt 3, pp. 613–627, 1993.
- [2] W. D. Stein and F. C. N.-Q. . C. 1989 Bronner, *Cell shape : determinants, regulation, and regulatory role*. San Diego: Academic Press, 1989.
- [3] D. E. Ingber and J. Folkman, "Tension and Compression as Basic Determinants of Cell Form and Function: Utilization of a Cellular Tensegrity Mechanism," in *Cell shape : determinants, regulation, and regulatory role*, W. D. Stein and F. Bronner, Eds. San Diego: Academic Press, pp. 3–31.
- [4] N. Wang, J. D. Tytell, and D. E. Ingber, "Mechanotransduction at a distance: mechanically coupling the extracellular matrix with the nucleus," *Nat Rev Mol Cell Biol*, vol. 10, no. 1, pp. 75–82, 2009.
- [5] S. Piccolo, S. Dupont, and M. Cordenonsi, "The biology of YAP/TAZ: hippo signaling and beyond," *Physiol Rev*, vol. 94, no. 4, pp. 1287–1312, 2014.
- [6] L. N. Castor, "Flattening, movement and control of division of epithelial-like cells," *J Cell Physiol*, vol. 75, no. 1, pp. 57–64, 1970.
- [7] J. Folkman and A. Moscona, "Role of cell shape in growth control," *Nature*, vol. 273, no. 5661, pp. 345–349, 1978.
- [8] M. Iwig and D. Glaesser, "On the role of microfilaments in cell-shape-mediated growth control of lens epithelial cells," *Cell Tissue Kinet*, vol. 18, no. 2, pp. 169–182, 1985.
- [9] J. Bereiter-Hahn, I. Karl, H. Luers, and M. Voth, "Mechanical basis of cell shape: investigations with the scanning acoustic microscope," *Biochem Cell Biol*, vol. 73, no. 7–8, pp. 337–348, 1995.
- [10] R. McBeath, D. M. Pirone, C. M. Nelson, K. Bhadriraju, and C. S. Chen, "Cell shape, cytoskeletal tension, and RhoA regulate stem cell lineage commitment," *Dev. Cell*, vol. 6, pp. 483–495, 2004.
- [11] M. F. Pittenger *et al.*, "Multilineage potential of adult human mesenchymal stem cells," *Science (80-. )*, vol. 284, no. 5411, pp. 143–147, 1999.
- [12] K. A. Kilian, B. Bugarija, B. T. Lahn, and M. Mrksich, "Geometric cues for directing the differentiation of

- mesenchymal stem cells,” *Proc Natl Acad Sci U S A*, vol. 107, no. 11, pp. 4872–4877, 2010.
- [13] A. J. Engler, S. Sen, H. L. Sweeney, and D. E. Discher, “Matrix elasticity directs stem cell lineage specification,” *Cell*, vol. 126, no. 4, pp. 677–689, 2006.
- [14] K. H. Vining and D. J. Mooney, “Mechanical forces direct stem cell behaviour in development and regeneration,” *Nat Rev Mol Cell Biol*, vol. 18, no. 12, pp. 728–742, 2017.
- [15] M. J. Dalby *et al.*, “The control of human mesenchymal cell differentiation using nanoscale symmetry and disorder,” *Nat. Mater.*, vol. 6, no. 12, pp. 997–1003, Dec. 2007.
- [16] C. Bakal, J. Aach, G. Church, and N. Perrimon, “Quantitative Morphological Signatures Define Local Signaling Networks Regulating Cell Morphology,” *Science (80-. )*, vol. 316, no. 5832, 2007.
- [17] O. Nir, C. Bakal, N. Perrimon, and B. Berger, “Inference of RhoGAP/GTPase regulation using single-cell morphological data from a combinatorial RNAi screen,” *Genome Res*, vol. 20, no. 3, pp. 372–380, 2010.
- [18] J. Wang, X. Zhou, P. L. Bradley, S. F. Chang, N. Perrimon, and S. T. Wong, “Cellular phenotype recognition for high-content RNA interference genome-wide screening,” *J Biomol Screen*, vol. 13, no. 1, pp. 29–39, 2008.
- [19] M. Tanaka *et al.*, “An unbiased cell morphology-based screen for new, biologically active small molecules,” *PLoS Biol*, vol. 3, no. 5, p. e128, 2005.
- [20] C. L. Adams *et al.*, “Compound classification using image-based cellular phenotypes,” *Methods Enzym.*, vol. 414, pp. 440–468, 2006.
- [21] M. Boutros, F. Heigwer, and C. Laufer, “Microscopy-Based High-Content Screening,” *Cell*, vol. 163, no. 6, pp. 1314–1325, 2015.
- [22] Z.-X. Ren, H.-B. Yu, J.-S. Li, J.-L. Shen, and W.-S. Du, “Suitable parameter choice on quantitative morphology of A549 cell in epithelial - mesenchymal transition.,” *Biosci. Rep.*, pp. 1–7, 2015.
- [23] E. Alizadeh, S. M. Lyons, J. M. Castle, and A. Prasad, “Measuring systematic changes in invasive cancer cell shape using Zernike moments,” *Integr Biol*, 2016.
- [24] S. M. Lyons *et al.*, “Changes in cell shape are correlated with metastatic potential in murine and human osteosarcomas,” *Biol. Open*, vol. 5, no. 3, pp. 289–299, 2016.
- [25] H. Z. Sailem and C. Bakal, “Identification of clinically predictive metagenes that encode components of a

- network coupling cell shape to transcription by image-omics.” *Genome Res.*, vol. 27, no. 2, pp. 196–207, Feb. 2017.
- [26] J. E. Sero, H. Z. Sailem, R. C. Ardy, H. Almuttaqi, T. Zhang, and C. Bakal, “Cell shape and the microenvironment regulate nuclear translocation of NF- $\kappa$ B in breast epithelial and tumor cells.” *Mol. Syst. Biol.*, vol. 11, no. 3, p. 790, Mar. 2015.
- [27] P. Pascual-Vargas, S. Cooper, J. Sero, V. Bousgouni, M. Arias-Garcia, and C. Bakal, “RNAi screens for Rho GTPase regulators of cell shape and YAP/TAZ localisation in triple negative breast cancer,” *Sci. Data*, vol. 4, p. 170018, 2017.
- [28] J. E. Sero and C. Bakal, “Multiparametric Analysis of Cell Shape Demonstrates that  $\beta$ -PIX Directly Couples YAP Activation to Extracellular Matrix Adhesion,” *Cell Syst.*, vol. 4, no. 1, p. 84–96.e6, 2017.
- [29] Z. Yin *et al.*, “A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes,” *Nat Cell Biol.*, vol. 15, no. 7, pp. 860–871, 2013.
- [30] P.-H. Wu *et al.*, “Evolution of cellular morpho-phenotypes in cancer metastasis,” *Sci. Rep.*, vol. 5, no. 1, p. 18437, 2016.
- [31] A. Pasqualato *et al.*, “Shape in migration: Quantitative image analysis of migrating chemoresistant HCT-8 colon cancer cells,” *Cell Adhes. Migr.*, vol. 7, no. 5, pp. 450–459, 2013.
- [32] A. Pasqualato *et al.*, “Quantitative shape analysis of chemoresistant colon cancer cells: Correlation between morphotype and phenotype,” *Exp. Cell Res.*, vol. 318, no. 7, pp. 835–846, 2012.
- [33] F. Matsuoka *et al.*, “Morphology-Based Prediction of Osteogenic Differentiation Potential of Human Mesenchymal Stem Cells,” *PLoS One*, vol. 8, no. 2, 2013.
- [34] D. Zhang and K. A. Kilian, “The effect of mesenchymal stem cell shape on the maintenance of multipotency,” *Biomaterials*, vol. 34, no. 16, pp. 3962–3969, 2013.
- [35] J. C. Caicedo *et al.*, “Data-analysis strategies for image-based cell profiling,” *Nat. Methods*, vol. 14, no. 9, pp. 849–863, 2017.
- [36] Z. PINCUS and J. A. THERIOT, “Comparison of quantitative methods for cell-shape analysis,” *J. Microsc.*, vol. 227, no. 2, pp. 140–156, Aug. 2007.

- [37] E. C. Kintner, "On the Mathematical Properties of the Zernike Polynomials," *Opt. Acta Int. J. Opt.*, vol. 23, no. 8, pp. 679–680, 1976.
- [38] J. H. Slater *et al.*, "Recapitulation and Modulation of the Cellular Architecture of a User-Chosen Cell of Interest Using Cell-Derived, Biomimetic Patterning," *ACS Nano*, vol. 9, no. 6, pp. 6128–6138, Jun. 2015.
- [39] B. Geiger, J. P. Spatz, and A. D. Bershadsky, "Environmental sensing through focal adhesions," *Nat Rev Mol Cell Biol*, vol. 10, no. 1, pp. 21–33, 2009.
- [40] C. D. Lawson and K. Burridge, "The on-off relationship of Rho and Rac during integrin-mediated adhesion and cell migration," *Small GTPases*, vol. 5, p. e27958, 2014.
- [41] S. Huvneers and E. H. Danen, "Adhesion signaling - crosstalk between integrins, Src and Rho," *J Cell Sci*, vol. 122, no. Pt 8, pp. 1059–1069, 2009.
- [42] U. S. Schwarz and S. A. Safran, "Physics of adherent cells," *Rev. Mod. Phys.*, vol. 85, no. 3, pp. 1327–1381, 2013.
- [43] W. R. Holmes, J. Park, A. Levchenko, and L. Edelstein-Keshet, "A mathematical model coupling polarity signaling to cell adhesion explains diverse cell migration patterns," *PLoS Comput Biol*, vol. 13, no. 5, p. e1005524, 2017.
- [44] W. R. Holmes and L. Edelstein-Keshet, "Analysis of a minimal Rho-GTPase circuit regulating cell shape," *Phys Biol*, vol. 13, no. 4, p. 46001, 2016.
- [45] A. Totaro *et al.*, "YAP/TAZ link cell mechanics to Notch signalling to control epidermal stem cell fate," *Nat. Commun.*, vol. 8, p. 15206, 2017.
- [46] S. Dupont *et al.*, "Role of YAP/TAZ in mechanotransduction," *Nature*, vol. 474, no. 7350, p. 179, 2011.
- [47] N. M. Ramdas, Q. Li, and G. V. Shivashankar, "Regulation of nuclear morphology by actomyosin components and cell geometry," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2015–Novem, pp. 342–345, 2015.
- [48] Q. Li, A. Kumar, E. Makhija, and G. V. Shivashankar, "The regulation of dynamic mechanical coupling between actin cytoskeleton and nucleus by matrix geometry," *Biomaterials*, vol. 35, no. 3, pp. 961–969, 2014.
- [49] N. Jain, K. V. Iyer, A. Kumar, and G. V. Shivashankar, "Cell geometric constraints induce modular gene-

- expression patterns via redistribution of HDAC3 regulated by actomyosin contractility,” *Proc. Natl. Acad. Sci.*, vol. 110, no. 28, pp. 11349–11354, 2013.
- [50] C. Uhler and G. V. Shivashankar, “Geometric control and modeling of genome reprogramming,” *Bioarchitecture*, vol. 6, no. 4, pp. 76–84, 2016.
- [51] T. Cremer and M. Cremer, “Chromosome territories,” *Cold Spring Harb Perspect Biol*, vol. 2, no. 3, p. a003889, 2010.
- [52] C. Lanctot, T. Cheutin, M. Cremer, G. Cavalli, and T. Cremer, “Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions,” *Nat Rev Genet*, vol. 8, no. 2, pp. 104–115, 2007.
- [53] S. Maharana, K. V Iyer, N. Jain, M. Nagarajan, Y. Wang, and G. V Shivashankar, “Chromosome intermingling—the physical basis of chromosome organization in differentiated cells,” *Nucleic Acids Res*, vol. 44, no. 11, pp. 5148–5160, 2016.
- [54] Y. Wang *et al.*, “Coupling between chromosome intermingling and gene regulation during cellular differentiation,” *Methods*, vol. 123, pp. 66–75, 2017.
- [55] Y. Wang, M. Nagarajan, C. Uhler, and G. V Shivashankar, “Orientation and repositioning of chromosomes correlate with cell geometry-dependent gene expression,” *Mol Biol Cell*, vol. 28, no. 14, pp. 1997–2009, 2017.
- [56] M. J. P. Biggs, R. G. Richards, and M. J. Dalby, “Nanotopographical modification: a regulator of cellular function through focal adhesions,” *Nanomedicine*, vol. 6, no. 5, pp. 619–633, 2010.
- [57] M. Laura E, M. Rebecca J, B. Manus J. P, K. Fahsai, O. Richard O. C, and D. Matthew J, “Nanotopographical Control of Stem Cell Differentiation,” *J. Tissue Eng.*, vol. 2010, p. 120623, 2010.
- [58] M. Nikkhah, F. Edalat, S. Manoucheri, and A. Khademhosseini, “Engineering microscale topographies to control the cell-substrate interface,” *Biomaterials*, vol. 33, no. 21, pp. 5230–5246, 2012.
- [59] A. A. Abdeen and K. Saha, “Manufacturing Cell Therapies Using Engineered Biomaterials,” *Trends Biotechnol*, vol. 35, no. 10, pp. 971–982, 2017.
- [60] J. M. Corey and E. L. Feldman, “Substrate patterning: an emerging technology for the study of neuronal behavior,” *Exp Neurol*, vol. 184 Suppl, pp. S89-96, 2003.
- [61] E. K. F. Yim, S. W. Pang, and K. W. Leong, “Synthetic nanostructures inducing differentiation of human

- mesenchymal stem cells into neuronal lineage,” *Exp Cell Res*, vol. 313, no. 9, pp. 1820–1829, 2007.
- [62] M. R. Lee *et al.*, “Direct differentiation of human embryonic stem cells into selective neurons on nanoscale ridge/groove pattern arrays,” *Biomaterials*, vol. 31, no. 15, pp. 4360–4366, 2010.
- [63] J. H. Slater *et al.*, “Modulation of endothelial cell migration via manipulation of adhesion site growth using nanopatterned surfaces,” *ACS Appl. Mater. Interfaces*, vol. 7, no. 7, pp. 4390–4400, Feb. 2015.
- [64] S. Dobbenga, L. E. Fratila-Apachitei, and A. A. Zadpoor, “Nanopattern-induced osteogenic differentiation of stem cells - A systematic review,” *Acta Biomater*, vol. 46, pp. 3–14, 2016.
- [65] J. H. Slater *et al.*, “Recapitulation and Modulation of the Cellular Architecture of a User-Chosen Cell of Interest Using Cell-Derived, Biomimetic Patterning,” *ACS Nano*, vol. 9, no. 6, pp. 6128–6138, 2015.
- [66] A. Shukla, J. H. Slater, J. C. Culver, M. E. Dickinson, and J. L. West, “Biomimetic Surface Patterning Promotes Mesenchymal Stem Cell Differentiation,” *ACS Appl Mater Interfaces*, vol. 8, no. 34, pp. 21883–21892, 2016.

## **Chapter 2: Development of shape quantification methods and testing their performance on the cells perturbed with different cytoskeletal drugs<sup>4</sup>**

### **2.1 Introduction**

A plethora of studies have shown that cell shape is intimately related to cell phenotype. Consideration of the biophysical basis for cell shape leads naturally to the hypothesis that cell shape reflects cell state. The final shape of a cell is determined by the balance between the internal and external forces exerted on the cell boundary. The cell exerts forces and responds to external forces, from the extracellular matrix (ECM) or from neighboring cells, with the help of molecular motors and the cellular cytoskeleton [1],[2]. Thus the cytoskeleton is the ultimate determinant of cell shape. The cellular cytoskeleton is a complex network, which consists of three major kinds of filaments, actin, microtubules and intermediate filaments that form a cross-linked dynamic meshwork in the cytoplasm of a cell. Assembly of these cytoskeletal filaments in the cell provides shape and structure to the cell [3],[4]. Forming the cytoskeleton structure is similar to building a structure using a LEGO toy. However the difference between them is that in the cytoskeleton, the protein subunits making up the filaments self-assemble according to different signals that they receive from cell's microenvironment or from the cell itself. The dynamic control of the cytoskeleton structure, based on control of polymerization and depolymerization, is the mechanism by which the cell can control its mechanical properties. The polymerization dynamics of actin and contractile forces generated by myosin motors also generates the necessary forces for the cell to migrate.

Contractile forces generated by myosin motors within cytoskeletal networks, membrane extension caused by actin polymerization, changes in osmotic pressure by opening of water or ion channels are examples of internal forces that play a role in shape of a cell. External forces leading to shape changes are applied through neighboring cells or extra-cellular matrix (ECM)[5]. Cytoskeleton filaments can generate and also resist mechanical stresses and cell deformation. But they can also eventually reorganize and change their structure, thereby sometimes relaxing external stresses. Different mechanical properties of the cell cytoskeleton and ECM will lead to different shapes for the cell.

---

<sup>4</sup> This chapter will be submitted to a journal to be published as a research paper. Therefore, some content, in particular the background information, partly overlaps with other chapters which is regrettable but unavoidable.

Therefore, cell morphology and cytoskeleton structure of a cell are linked and may provide information about its cellular and environmental conditions.

Researchers have used cell shape as a predictor of cell phenotype and behavior. As an example, we used Zernike moments and geometric parameters as a measure of cell shape to distinguish between high metastatic and low metastatic osteosarcoma cancer cell lines with 99% accuracy. Sailem et al have used cell shape as predictor of tumor grade [6], while Sero et al[7] predict changes in the nuclear/cytoplasmic ratio of NF $\kappa$ B, a transcription factor involved in cancer progression, from changes in shape of the cells. Similarly Sero et al[8] found that in MCF10A breast epithelial cells, and four variants of MDA-MB-231 breast cancer cell lines, the nuclear/cytoplasmic ratio of YAP (Yes-associated protein), a mechanosensor involved in cell growth, proliferation, differentiating and tumorigenesis, is correlated with cell shape. Pasqualato et al[9] ,[10] reported significant shape differences between chemoresistant and chemosensitive human colon cancer cell lines. Keren et al use phase contrast microscopy of keratocytes and found that cells with large aspect ratio move faster[11]. Cooper et al[12] look at the relation between cell shape and two forms of motility, amoeboid and mesenchymal, in PTEN-null WM266.4 melanoma cells. They found that cells with amoeboid motility displayed what they described as elliptical, small-round or large-round morphology, while cells with mesenchymal motility showed star, spindle or tear drop morphology. They observe changes in frequency distributions of the six shapes by depletion of Rho-family GTPases which are regulators of cell shape. Ren et al[13] showed that shape parameters can distinguish between A549 cells during Epithelial to Mesenchymal transition (EMT). Matsuoka et al use cell shape to non-invasively predict osteogenic differentiation potential of Human Mesenchymal Stem Cells (hMSCs) which is validated using commonly used osteogenic markers[14].

The importance of cell morphology as a tool for understanding and predicting cell behavior, and development of advanced microscopes and high throughput screening techniques raises the need to appropriately quantify shape of a cell. Depending on the cell type and questions being asked different quantifiers can be used, and we refer to the space of all quantifiers as shape space. In the shape space each cell will be represented by a point. If the shape in the population is random, or if inappropriate shape measures chosen, the points will be distributed randomly in the shape space. Therefore, some criteria should be considered when choosing proper shape quantifiers in order to learn from numerical representations of shape. Pincus et al[15] introduce three criteria for good shape quantifiers. The first criterion is fidelity. A measure with high fidelity is one which discards at most a small amount of true information and adds at most a small amount of noise to the data. The second criterion is capturing biologically important variations,

and discarding unimportant variations within populations of interest. Thus a shape measure should be able to distinguish between cell types with different treatment conditions. The last criterion is that each quantifier be individually meaningful and interpretable. For the first criterion, the higher the fidelity the better the quantifier is. However, the last two criteria can be prioritized differently depending on the problem and the cells under study.

In all the examples provided above, different measures are used by different groups, but in all of the cases they only use information about cell boundary. Here, we show that in addition to information about the cell boundary, the structure of the actin cytoskeleton also carries important information, and taking it into account can provide more insight to the cell's behavior.

In order to quantify and characterize the role of the cytoskeleton on cell shape, we use pharmacological modulators of the cytoskeleton, and quantify the consequent changes in cell shape and actin structure. In the process, we develop new measures to quantify both the cell boundary as well as the actin distribution within the cell. Then we use these measures to quantify changes in actin structure using texture information of actin stained images. We use the actin structure measures to verify that the drugs affect actin distribution. Then the changes in shape of the cell and nuclei by perturbing cytoskeleton is quantified and analyzed. We show that even though cytoskeleton structure is related to final shape of the cell, their relation is complex and nonlinear.

## **2.2 Experimental Methods**

For this study, DUNN and DLM8 osteosarcoma cancer cells are used. DLM8 line is derived from DUNN cell line with selection for metastasis. Therefore, DLM8 is closely related to DUNN except for degree of its invasiveness. Cells are treated with Cytoskeletal drugs with description and conditions listed in the Table 2-1. After 3 hours of treatment with these drug cells are washed and then are fixed and fluorescently stained for nuclei and actin. All the drugs are dissolved in Dimethyl sulfoxide (DMSO) and to drop its effect on cell shape and actin structure control study is treated with DMSO with the same molarity as other drugs. Representative image of each cell line treated with these drugs are shown in Figure 2-1.

## **2.3 Imaging methods**

To quantify cell shape, the cell first needs to be imaged using a microscope. Phase contrast microscopy is sometimes used to record the shape of cells. With a phase contrast microscopy, perimeter of the cells will be distorted and background noise is usually high[16]. Fluorescence microscopy provides better detail of the outline of the cell

and the structure inside the cell. In order to use this technique, biological molecules of choice should be fluorescently labeled. We use fluorescent microscopy to image the cell and nuclei of the cells in this study.

## 2.4 Data analysis methods

### 2.4.1 Quantifying changes in multidimensional shape space

Each cell can be represented as a point in multidimensional shape space. For ease of analysis this vector can be projected down to a lower-dimensional space of the first few principal components (PCs). Differences between different conditions or different cell lines can be represented by p-value between their shape distributions for each PC space. While doing multiple comparisons, we often pick the most informative principal component for our analysis. This is done in the following manner. We pick the PC whose worst case, i.e. largest p-value among all comparisons, is better (i.e. smaller p-value) than that of any other PC, so that it is the best single measure for distinguishing between all the comparisons. In other words, for each principal component maximum p-value between all comparisons is calculated as Equation 2.1.

$$\text{MaxP}_{(PC1)} = \text{Max} \{ \text{p-value}_{(PC1, 1)}, \dots, \text{p-value}_{(PC1, n)} \} \quad \text{Equation 2.1}$$

Where  $\text{p-value}_{(PC1, n)}$  is the p-value for  $n^{\text{th}}$  comparison in the first principal component. Then from the first four PCs, the one that has the smallest  $\text{MaxP}$ , maximum p-value between different comparisons, is chosen. In each shape category, we choose the best principal component by this criteria, and will describe them as the Primary Principal Components (PPC) of each shape quantification measure or category, for each analysis.

### 2.4.2 Pearson correlation

Pearson correlation coefficient is calculated between features as shown in Equation 2.2. Here  $r$  is Pearson correlation coefficient,  $x_i$  is the feature  $x$  for  $i^{\text{th}}$  sample,  $y_i$  is the feature  $y$  for  $i^{\text{th}}$  sample,  $\mu_x$  is the average value of the feature  $x$  for all of the samples, and  $\mu_y$  is the average value of the feature  $y$  for all of the samples. This coefficient is calculated for all 14 cell lines and drug combinations (2 cell lines x 7 drugs) and the averaged coefficient is recorded. The results are shown in heat map plots in Appendix I: Figure S2.1. In these plots the diagonal elements are correlation between features with themselves, so they are all 1. Also, Pearson correlation coefficient is symmetric so each heat map plot will be symmetric as well.

$$r = \frac{\sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^n (x_i - \mu_x)^2} \sqrt{\sum_{i=1}^n (y_i - \mu_y)^2}} \quad \text{Equation 2.2}$$

## 2.5 Results

### 2.5.1 Drugs which directly perturb microtubules, myosin-II or actin leave a unique signature on structure and shape of a cell.

We found that the most dramatic and unique effects on cell shape arise from drugs that directly perturb microtubules and actin filaments. As shown in Figure 2-1, control cells acquire a polygonal shape on surfaces. Depolymerizing actin using Cytochalasin D leads to rounded cells where the actin has re-organized into alternating concentric rings of high and low density. Within each ring the actin is structured in very unique radial streaks! Increasing the depolymerization rate of microtubules using Nocodazole leads to a small rounded morphology. Inhibiting myosin II activity and decreasing cell's contractility using Blebbistatin leads to increase in irregularities in the cell boundary. These changes in shape and structure of a cell are similar for both DUNN and DLM8 cell lines. Changes in structure and morphology of the cells using other drugs are subtle and not easily identifiable by eye. Below we explore the performance of different shape quantifiers for analyzing the shape changes under these treatment conditions.

### 2.5.2 Developing Shape quantifiers:

The actin structure is obtained from pixel intensities of labelled actin, which we can represent by a greyscale image (Figure 1-1A). These greyscale images of cells, or images of stained nuclei, are converted into a binary image (Figure 1-1B) and also the coordinates of the edges are recorded (Figure 1-1C), yielding the cell boundary. To extract features of the morphology from this numerical data, we introduce three classes of shape measures, described as textural, spreading, and irregularity classes. Textural measures, include subcategories called band based measurements, gray scale fractal dimension, and gray scale measures which are calculated based on the textural representation of actin, Figure 1-1A. Spreading measures, which include measures based on a Zernike moment representation of shape as well as subcategories involving basic geometric parameters such as area and perimeter, are mostly extracted from the binary representation of the cell (Figure 1-1B), its convex hull, or a similar image of the nuclei. Irregularity measures include parameters such as the waviness and roughness, which use information regarding the pixels at the boundary of the cell, Figure 1-1C. Each of these shape categories have been described below.

#### 2.5.2.1 Textural measures

To identify changes in actin distribution of the cell textural measures are used. In these measures only gray scale representation of the cells, shown in Figure 1-1A, are used. With these measures we can quantify actin

distribution in each cell and verify if the drug changes actin distribution significantly. Below, three types of textural measures that are used in this paper are introduced.

#### 2.5.2.1.1 Band based measurements

This parameter is sensitive to changes in the radial symmetry of the distribution of actin (or any other structure under analysis). We divide the image into 10 equally spaced concentric regions ( $\Delta r$ ) around the center of mass of the image, called bands, as shown in Appendix I: Figure S2.2. Five quantifiers are used to measure the differences in actin distribution between these bands. First, average intensity for each band is calculated. Then indices of the bands which have the lowest or the highest average intensity and their value are recorded. The last measurement is what is called above average adjusted intensity of the bands which is formulated in Appendix I: Table S2-1 along with other band based measures[17].

In all the cells treated with Cytochalasin D, actin has very unique symmetrical distribution. In these cells, dense foci are formed around the nucleus. The bands located in the central region of the cells are void of actin. At the outer bands, short linear actin structures aligned along the radius are observed.

#### 2.5.2.1.2 Gray scale fractal dimension

Fractal dimension (FD) is a measure of roughness of objects that has been applied to the characterization of texture in engineered and natural images. For planar images the fractal dimension lies between 2 and 3. There are many methods to calculate this measure, and we chose the box counting method which is as follows. First the binary image is covered with evenly spaced grid with side length of  $\epsilon$ . Then, the number of boxes which cover the fractal image are counted. This process is repeated by decreasing side length and FD is calculated based on Equation 2.3

$$FD = \lim_{\epsilon \rightarrow 0} \frac{\log N(\epsilon)}{\log(\frac{1}{\epsilon})} \quad \text{Equation 2.3}$$

Where  $\epsilon$  is size of each box in the grid and the variable  $N(\epsilon)$  is the number of the boxes which contain the fractal.

We calculate FD based on a binarization of the greyscale image using edge detection methods to identify actin voids. This is done using four different edge detection methods in Matlab. To provide an example, binarized images of a cell from DUNN cell line treated with Cytochalasin D using these four edge detection methods are shown in Appendix A: Figure S2.3. We found that different edge detection methods pick up different aspects of the actin structure, and so for every cell their gray scale image is binarized using these four methods. Then, their FD is calculated.

### 2.5.2.1.3 Other gray scale measures

Haralik et, al [18] introduced a procedure for quantifying the texture of satellite images based on the spatial relation between the gray tone of neighboring pixels in an image. They used these measures for classifying the satellite images. In this method the gray level co-occurrence matrix (GLCM, sometimes called Spatial-Dependence Matrix) is calculated for gray scale images. For an image that has intensities of 1, 2, ..., g, the co-occurrence matrix is a  $g \times g$  matrix such that its  $ij^{\text{th}}$  element is the number of the times that a pixel in the image has intensity equal to “i” and the pixel at a pre-defined distance  $\bar{d}$  (which we choose to be 1 pixel) of it has intensity of j. An example of a 4pixels x 4pixels image with 5 gray tone levels is shown in Appendix A: Figure S2.4.

In a GLCM matrix when there are very few dominant transitions in gray tone of an image in neighboring pixels, the matrix will have small value for all of the entries. Each diagonal element of a GLCM matrix represents the number of the times that gray tone does not change in the neighboring pixel. Large diagonal elements imply that the image is homogenous. Large numbers in the far upper right and far lower left of the matrix implies large transitions in intensity and high contrast in an image. After calculating the GCLM matrix, shown for the example image in Appendix I: Figure S2.4C, 23 different measures are calculated to quantify the texture in an image. The list of measures are tabulated in Appendix I: Table S2-2[19], [18], [20], [21]. Contrast is one of the parameters. For interpretation of this measure, it is useful to keep in mind that the contrast in an image is proportional to the changes in gray tone,  $n=|i-j|$ , so the far upper right and far lower left which have bigger value of n will have bigger contribution to the contrast parameter. For homogenous images, the diagonal entries will be large, and will have a bigger contribution to the homogeneity measure.

In this study we calculate the GLCM matrix for the vector  $\bar{d}$  equal to 1 pixel in magnitude and with directions of  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $125^\circ$  and then the average value for each parameter is reported. This process leads to rotation invariant measures.

### 2.5.2.1.4 Cytoskeletal drugs changes actin organization significantly which are quantifiable by textural measures

Table 2-2 demonstrates quantified changes in all the measures for all the drugs. As demonstrated in this table the changes in textural measures for all the drugs are significant. This means that all the drugs that perturbed actin either directly, or indirectly through perturbing other cytoskeletal components, lead to significant changes in actin

cytoskeleton organization. In each drug condition, changes in textural features for different cells are similar for both DUNN and DLM8 cell lines except for PP2, where the direction of shape changes are the opposite of each other.

### **2.5.2.2 Spreading measures**

In this class of shape quantifiers, the binary image of cell and nuclei is used. These measures include geometric parameters for nuclei and cell, convex hull measures and Zernike moment for cells.

#### **2.5.2.2.1 Geometric measures**

Parameters like area, perimeter, major axis of fitted ellipse, minor axis of fitted ellipse, their ratio are examples of geometric measures. All geometric measures used in this study to quantify cell and nuclei's shapes are listed in Appendix I: Table S2-4 and S2-5.

#### **2.5.2.2.2 Convex Hull measurements**

A convex polygon is a closed polygon such that the connecting line between any pair of points, lies completely inside the polygon. Also, one can walk between any two points inside the polygon in a straight line without leaving the polygon. Appendix I: Table S2-5A and B shows an example of a convex and non-convex polygon. The convex hull of a 2D shape is the smallest convex polygon that encloses the whole shape[22]. An example of the convex hull of a 2D cell shape is shown in Appendix I: Figure S2.5C. Convex hull geometric measures used in this study are listed in Appendix I: Table S2-6.

#### **2.5.2.2.3 Zernike moments:**

To calculate Zernike moments, the image of the cell is projected to Zernike Polynomial basis function and the coefficients are called Zernike moments. Zernike moments are complex numbers whose magnitude is rotation invariant, and can be made displacement invariant by moving the cell to the center of the imaging frame such that the center of mass of the image falls on the geometric center of the frame. The procedure will be discussed extensively in chapter 4[23].

#### **2.5.2.2.4 Actin reorganization changes the spreading measures for both the cell and the nucleus**

To explore changes in 2D shape of a cell and nuclei by changing actin distribution we compare the PPC of the geometric measures of the cell and nuclei. As shown in Table 2-2 perturbing actin significantly changes cell geometric measures other than DLM8 cell line treated with Jasplakinolide. Interestingly changes in actin structure not

only changes cell geometric measure, but it also changes nuclei geometric measures for all conditions other than both cell lines treated with Blebbistatin and Dunn cell line treated with Jasplakinolide and PP2 drugs.

Zernike moments and Convex Hull parameters are similar in some respects. Although their quantification method is very different, as shown in reconstruction of the image of cells using Zernike moments (Appendix I: Figure S2.6), both measures ignore the irregularities and fine fluctuations in the boundary. As demonstrated in Table 2-2 changes in the PPC for both measures are not distinguishable for DUNN cell lines treated with PP2, and both cell lines treated with FAKI 14. Moreover, hull geometric measure does not change significantly for DLM8 cells treated with Cytochalasin-D and DUNN cells treated with Jasplakinolide. In addition, Zernike moments does not change significantly for the DLM8 cells treated with Blebbistatin and DUNN cells treated with Cytochalasin-D. All other drugs lead to distinguishable changes in convex hull and Zernike moment measures.

Since many cell shape parameters are expected to be correlated with each other, we performed a correlation analysis of all the measures within each shape quantification category, which are shown in Appendix I: Figure S2.7. This results indicates that Zernike moments, cell geometric measures, size measures in nuclei, and convex hull geometric measures are highly correlated with each other.

### **2.5.2.3 Irregularity of boundary measures:**

In addition to cell spreading, irregularities in the boundary also carries information about state of the cell. For example an irregular border could arise due to a large number of filopodia in the cell, which is a signature of a highly dynamic cytoskeleton. A highly contractile cell may retract from focal adhesions at the boundary creating many membrane protrusions that increase variability. To quantify irregularity of the cell boundary, the 2D boundary of the cell is used as shown in Figure 1-1. We use the pixel positions that mark the boundary of the cell to calculate waviness, which estimates the periodic variation in the boundary, and roughness, which measures the non-periodic variation in the boundary, as discussed below.

#### **2.5.2.3.1 Waviness measures**

Using a Fourier series, a signal can be expanded in terms of linear combinations of orthogonal basis functions of sines and cosines with increasing frequencies. In the Fourier series expansion, it is assumed that the input signal is periodic. The outline of a cell is a closed curve, hence it is counted as periodic signal. The boundary of the cell could be represented in Cartesian coordinates,  $x$  and  $y$ , or polar coordinates,  $\rho$  and  $\theta$ . Regardless of coordinate selection, it

will lead to two independent signals which can be separately written as linear combination of cos and sin basis functions as equation 2.4.

$$f(x) = C_{x,0} + \sum_{n=1}^{nPixel} A_{x,n} \cos(wnx) + B_{y,n} \sin(wnx) \quad \text{Equation 2.4.}$$

Where nPixel is the number of the pixels in the boundary of a cell and  $\frac{w}{2\pi}$  is fundamental frequency which is equal to  $\frac{1}{2\pi \cdot nPixel}$ . The variable n is an integer number and n\*w is the nth frequency in the decomposition. Variables  $A_n$  and  $B_n$  are amplitudes of n<sup>th</sup> frequency and  $f(x)$  is the input signal.

If we use all the frequencies, we can reconstruct the cell. But the number of the descriptors will be too high. There is a tradeoff between the number of frequencies (shape descriptors) used and the accuracy of the reconstruction. We are interested in quantifying shape without dealing with high number of parameters. Since the amplitude decreases with increasing frequency, we can filter higher frequencies and just use lower frequencies to decrease the number of descriptors used for the analysis. Here, with qualitative analysis of reconstructed cells with different frequency we decided to use only the first 35 frequencies as descriptor of the cell. Reconstruction of shape of a few cells using these 35 frequencies is excellent, as shown in Figure 1-2.

However, a remaining issue for the Fourier coefficients is that they are not rotationally invariant. We can construct a rotationally invariant measure by using Equation 2.5, which removes the phase difference in the Fourier expansion.

$$C_n = \sqrt{A_n^2 + B_n^2} \quad \text{Equation 2.5.}$$

This also reduces the number of parameters to half. Even though reconstruction of cell shape will not be possible with the rotationally invariant measures, our results show that they are nevertheless useful parameters for distinguishing between cells with different degree of variations in the radius. We will refer to these parameters as Waviness parameters. Interestingly, for Fourier parameters, all the features show low correlation coefficient with each other (~0.35), suggesting that each coefficient carries mostly independent information. This observation was expected since we use orthogonal basis functions.

### 2.5.2.3.2 Roughness

As explained earlier, in the Fourier series decomposition of shape, higher frequencies have small amplitude and we neglect them. However, we can still obtain information about small amplitude variations from those high

frequency terms. Following the method of [24] to account for high frequency measures, we reconstruct the cell with the 35 Fourier components and subtract it from the original signal. Now we can calculate the roughness measures from the remainder of the signal. These parameters are also highly correlated with each other and are listed in Appendix I: Table S2-7.

#### **2.5.2.3.3 Cytoskeletal reorganization leads to changes in irregularities of a cell's boundary**

As shown in Table 2-2, waviness measures for all the drug conditions, except for DUNN cell lines treated with Jasplakinolide or PP2 and both cell lines treated with FAKI 14, changes with respect to control study significantly. In addition, the roughness measures change significantly for both cell lines treated with Blebbistatin, Cytochalasin-D, FAKI 14, and Nocodazole. In both cell lines, Jasplakinolide does not change the roughness measure significantly, which is also the case for the DUNN cell lines treated with PP2.

#### **2.5.2.4 Different categories of shape quantifiers represent non-redundant shape information**

While we have shown that each of our major categories and subcategories serve as good shape measures, in that they can be used to look for interpretable shape changes in different drug conditions, it is not clear whether we need all of them to represent shape. In order to estimate the degree of redundant information carried by the different shape categories, we calculated the Pearson correlation coefficient between all the features from different shape categories (Data not shown). In general, features from two different shape categories are weakly correlated ( $<0.4$ ) except for Convex Hull and Cell Geometric features which are highly correlated with each other. There are a few other specific exceptions for which the features are also highly correlated. They are as follows. Zernike moments with  $n < 16$  and  $m = 0$  are highly correlated with Area. The correlation coefficient between Cell Area and Zernike Moment (0\_0) is 1. It decreases with increasing order till it is almost zero for Zernike Moment (22\_0), then it becomes negative and increases in magnitude (Appendix I: Figure S2.7 ). The coefficient C\_0 from Fourier features has a correlation of 1 with the Mean Cell Radius. It is also highly correlated with Cell Area (0.93) and Convex Hull area. Since the Fourier decomposition is based on a radial representation of the cell shape, C\_0 is a measure of the average cell radius, and should be expected to show these high correlations. The Fractal Dimension measures also have high correlation with Cell Geometric, Nuclei Geometric, Gray Scale measures, and Zernike moments with  $m = 0$  and  $n < 17$ . However in spite of these specific cases, the relative high number of weak correlation between different shape categories, suggests that these shape categories contain non-redundant information about cell shape. Thus, quantitative shape analysis should

ideally be carried out with representations from all of these shape categories in order for most efficient discrimination between different experimental conditions.

## **2.6 Conclusion:**

In this paper we introduce and provide tools to quantify cell shape and cytoskeletal structure based on two dimensional images of cell morphology and actin structure. This work is among the few that quantifies structure of the actin. Using statistically significant differences between experimental conditions in principal component space as our measure, we explore the capacity of these quantitative measures in capturing biologically important information. We perturb the cytoskeleton of the cells with different drugs and explore their effect on cell shape and its structure. We first use textural measures to verify that actin structure of the cell changes in the cells treated by cytoskeletal drugs and then we explore changes in the cell's and nuclei's 2D shape and irregularities of the cell boundary accompanied by changes in actin structure. Here we showed that most of the drugs used in this study directly or indirectly lead to significant changes in actin structure. Then we explored effect of changes in structure of the cell on irregularities of cell boundary, nuclei spreading, and cell spreading. In most of the cases changes in actin structure are accompanied with significant changes in irregularities of cell boundary, and cell and nuclei spreading. The results show that textural measures and spreading measures are related but their relation is not linear and in general the two classes of measures carry non-redundant information. It is worth mentioning that although we implemented texture measures to quantify actin structure, it can be used to quantify other cell structures as well. These shape quantification methods presented in this paper will prove useful for computer aided diagnostics in diseases such as cancer that are associated with cytoskeletal perturbations, assessment of qualitative cellular changes in different experimental conditions, and for mechanistic understanding of the determination of cell shape.

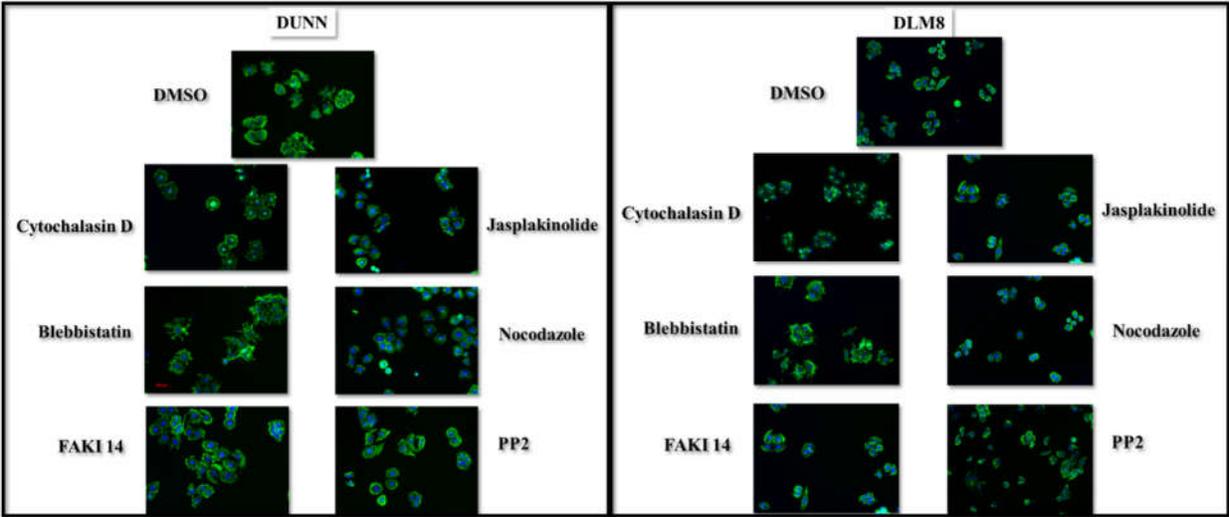


Figure 2-1: Representative images of DUNN and DLM8 osteosarcoma cancer cell lines with different drug treatments. Blue color represents nuclei and green color represents actin cytoskeleton.

Table 2-1. Description of the drugs used to perturb cytoskeleton.

<b>Drug</b>	<b>Blebbistatin</b>	<b>Cytochalasin D</b>	<b>Jasplakinolide</b>	<b>Faki 14</b>	<b>Nocodazole</b>	<b>PP2</b>
<b>Target</b>	Inhibits non-muscle myosin II activity	Inhibits actin polymerization	1. Promotes actin polymerization 2. stabilizes actin filaments	Inhibits focal adhesion kinase	Depolymerizing microtubules	Inhibits Src kinase
<b>Application in cancer therapy</b>	Inhibits cancer invasion[25]	Used as chemotherapeutic agent[[26], [27].	Reduces lung metastases of systemic Lewis lung carcinoma[28].	Showed anti-metastatic, anti-neoplastic and anti-angiogenic properties [29]–[32].	Has chemotherapeutic properties by arresting cells in mitosis[33].	Prevention of cancer metastasis[34].
<b>Effect</b>	1.Reduces F-actin contractility 2.Increase in cell spreading and 3.Decreases cell motility[7]	Decreases cell contractility	Stabilizes F-actin[35]	Reduces cell motility and focal adhesion turnover[36].	Cell shapes which lack tails and long protrusions and decreased cell asymmetry[37].	1. Induces strong cell-cell contact. 2.Enhances E-cadherin/catenin expression which are strongly associated with the actin cytoskeleton[34].
<b>Concentration (uM)</b>	10	1	0.01	1	5	1

Table 2-2. Changes in primary principal component of all measures for different drugs according to T-test analysis. In general all of the drugs lead to significant changes in actin texture of the cell and cell geometric parameters. Changes in all other parameters depends on the cell line and the drug treatment.

Comparisons: Control vs	Cell Line	Band Based	Fractal Dimension	Gray Scale	Cell Geometric	Nuclei	Hull Geometric	Zernike Moment	Waviness	Roughness
Blebbistatin	DUNN	1.00E-07	2.60E-14	0.008	1.60E-78	0.006	3.10E-14	3.00E-21	2.00E-35	2.70E-45
	DLM8	3.00E-08	6.30E-45	6.00E-09	6.00E-25	0.004	7.90E-08	0.0018	2.00E-12	5.30E-16
Cytochalasin D	DUNN	3.00E-50	2.10E-31	2.00E-23	3.50E-21	2.00E-22	2.50E-06	0.0092	5.00E-10	3.30E-27
	DLM8	4.00E-31	2.10E-13	1.00E-23	4.50E-15	1.00E-18	0.05285	1.00E-05	3.00E-07	8.20E-07
Faki 14	DUNN	2.00E-09	5.16E-31	1.00E-23	5.20E-09	2.00E-04	0.04184	0.1194	0.1398	2.20E-12
	DLM8	4.00E-04	6.60E-22	2.00E-21	1.50E-06	2.00E-06	0.07864	0.0997	0.3125	7.80E-07
Jasplakinolide	DUNN	2.00E-04	2.10E-13	9.00E-05	1.30E-08	0.001	0.02999	4.00E-05	0.0036	0.051
	DLM8	0.003	2.80E-09	2.00E-04	0.00207	5.00E-04	1.20E-05	0.0005	0.0001	0.02117
Nocodazole	DUNN	6.00E-13	5.60E-31	3.00E-25	1.30E-28	1.00E-04	4.60E-10	9.00E-07	1.00E-12	3.30E-16
	DLM8	0.001	9.80E-14	4.00E-07	1.20E-14	6.00E-06	3.00E-09	6.00E-14	3.00E-21	5.30E-13
PP2	DUNN	2.00E-21	1.40E-11	2.00E-10	2.50E-06	0.003	0.03602	0.0243	0.0089	0.00387
	DLM8	4.00E-07	0.00035	3.00E-09	0.00043	8.00E-10	2.70E-09	4.00E-13	4.00E-11	0.00078

## Bibliography

- [1] D. A. Fletcher and R. D. Mullins, “Cell mechanics and the cytoskeleton,” *Nature*, vol. 463, no. 7280, pp. 485–92, Jan. 2010.
- [2] T. D. Pollard and J. A. Cooper, “Actin, a Central Player in Cell Shape and Movement,” *Science (80-. )*, vol. 326, no. 5957, 2009.
- [3] D. A. Fletcher and R. D. Mullins, “Cell mechanics and the cytoskeleton,” *Nature*, vol. 463, no. 7280, pp. 485–492, 2010.
- [4] D. E. Ingber, Z. Sun, and H. Betensky, “E , Cytoskeletal ~ Ec ~ an ~ Cs ~ Control in Angiogene,” *Science (80-. )*, vol. 7, no. 12, 1994.
- [5] E. Paluch and C.-P. Heisenberg, “Biology and Physics of Cell Shape Changes in Development,” *Curr. Biol.*, vol. 19, no. 17, pp. R790–R799, Sep. 2009.
- [6] H. Z. Sailem and C. Bakal, “Identification of clinically predictive metagenes that encode components of a network coupling cell shape to transcription by image-omics,” *Genome Res.*, vol. 27, no. 2, pp. 196–207, Feb. 2017.
- [7] J. E. Sero, H. Z. Sailem, R. C. Ardy, H. Almuttaqi, T. Zhang, and C. Bakal, “Cell shape and the microenvironment regulate nuclear translocation of NF- $\kappa$ B in breast epithelial and tumor cells,” *Mol. Syst. Biol.*, vol. 11, no. 3, p. 790, Mar. 2015.
- [8] J. E. Sero and C. Bakal, “Multiparametric Analysis of Cell Shape Demonstrates that  $\beta$ -PIX Directly Couples YAP Activation to Extracellular Matrix Adhesion,” *Cell Syst.*, vol. 4, no. 1, p. 84–96.e6, Jan. 2017.
- [9] A. Pasqualato *et al.*, “Shape in migration: Quantitative image analysis of migrating chemoresistant HCT-8 colon cancer cells,” *Cell Adhes. Migr.*, vol. 7, no. 5, pp. 450–459, 2013.
- [10] A. Pasqualato *et al.*, “Quantitative shape analysis of chemoresistant colon cancer cells: Correlation between morphotype and phenotype,” *Exp. Cell Res.*, vol. 318, no. 7, pp. 835–846, 2012.
- [11] K. Keren *et al.*, “Mechanism of shape determination in motile cells,” *Nature*, vol. 453, no. 7194, pp. 475–480, 2008.

- [12] S. Cooper, A. Sadok, V. Bousgouni, and C. Bakal, "Apolar and polar transitions drive the conversion between amoeboid and mesenchymal shapes in melanoma cells.," *Mol. Biol. Cell*, vol. 26, no. 22, pp. 4163–70, Nov. 2015.
- [13] Z.-X. Ren, H.-B. Yu, J.-S. Li, J.-L. Shen, and W.-S. Du, "Suitable parameter choice on quantitative morphology of A549 cell in epithelial - mesenchymal transition.," *Biosci. Rep.*, pp. 1–7, 2015.
- [14] F. Matsuoka *et al.*, "Morphology-Based Prediction of Osteogenic Differentiation Potential of Human Mesenchymal Stem Cells," *PLoS One*, vol. 8, no. 2, 2013.
- [15] Z. PINCUS and J. A. THERIOT, "Comparison of quantitative methods for cell-shape analysis," *J. Microsc.*, vol. 227, no. 2, pp. 140–156, Aug. 2007.
- [16] S. M. Wilson and A. Bacic, "Preparation of plant cells for transmission electron microscopy to optimize immunogold labeling of carbohydrate and protein epitopes," *Nat. Protoc.*, vol. 7, p. 1716, Aug. 2012.
- [17] J. H. Slater *et al.*, "Recapitulation and Modulation of the Cellular Architecture of a User-Chosen Cell of Interest Using Cell-Derived, Biomimetic Patterning," *ACS Nano*, vol. 9, no. 6, pp. 6128–6138, 2015.
- [18] K. S. Robert M. HARALICK and I. DINSTEN, "Textural Features for Image classification," *IEEE Trans. Syst. Man. Cybern.*, vol. SMC-3, no. 6, pp. 610–621, 1973.
- [19] "Properties of gray-level co-occurrence matrix." [Online]. Available: [https://www.mathworks.com/help/images/ref/graycoprops.html?searchHighlight=glcm&s\\_tid=doc\\_srchtile](https://www.mathworks.com/help/images/ref/graycoprops.html?searchHighlight=glcm&s_tid=doc_srchtile).
- [20] L. Soh, C. Tsatsoulis, and S. Member, "Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices," *Ieee Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 780–795, 1999.
- [21] Avinash Uppuluri, "GCLM feature calculation." [Online]. Available: [https://www.mathworks.com/matlabcentral/fileexchange/22354-glcm-features4-m--vectorized-version-of-glcm-features1-m--with-code-changes-?s\\_tid=prof\\_contriblnk](https://www.mathworks.com/matlabcentral/fileexchange/22354-glcm-features4-m--vectorized-version-of-glcm-features1-m--with-code-changes-?s_tid=prof_contriblnk).
- [22] B. Gärtner, M. Hoffmann, and E. Welzl, "Computational Geometry." [Online]. Available: <https://www.ti.inf.ethz.ch/ew/Lehre/CG13/index.html>.
- [23] E. Alizadeh, S. M. Lyons, J. M. Castle, and A. Prasad, "Measuring systematic changes in invasive cancer cell

- shape using Zernike moments,” *Integr. Biol.*, vol. 8, no. 11, pp. 1183–1193, 2016.
- [24] J. De and Villanueva, “DESIGN AND IMPLEMENTATION OF A CALCULATION PROGRAM IN MATLAB PLATFORM FOR EVALUATION OF ROUGHNESS,” 2015.
- [25] T. Nelson, M. Roddy, D. M. Appledorn, and V. Groppi, “The myosin II inhibitor blebbistatin more potently inhibits 3-D cell invasion than 2-D cell migration in a human metastatic tumor cell,” *Cancer Res.*, vol. 71, no. 18 Supplement, pp. C1–C1, 2011.
- [26] M. Trendowski, J. N. Zoino, T. D. Christen, C. Acquafondata, and T. P. Fondy, “Preparation, In Vivo Administration, Dose-Limiting Toxicities, and Antineoplastic Activity of Cytochalasin B,” *Transl. Oncol.*, vol. 8, no. 4, pp. 308–317, Aug. 2015.
- [27] M. Trendowski, J. M. Mitchell, C. M. Corsette, C. Acquafondata, and T. P. Fondy, “Chemotherapy with cytochalasin congeners in vitro and in vivo against murine models,” *Invest. New Drugs*, vol. 33, no. 2, pp. 290–299, Apr. 2015.
- [28] H. Takeuchi, G. Ara, E. A. Sausville, and B. Teicher, “Jasplakinolide: interaction with radiation and hyperthermia in human prostate carcinoma and Lewis lung carcinoma,” *Cancer Chemother. Pharmacol.*, vol. 42, no. 6, pp. 491–496, 1998.
- [29] C. G. Kang, H. J. Han, H.-J. Lee, S.-H. Kim, and E.-O. Lee, “Rho-associated kinase signaling is required for osteopontin-induced cell invasion through inactivating cofilin in human non-small cell lung cancer cell lines,” *Bioorg. Med. Chem. Lett.*, vol. 25, no. 9, pp. 1956–60, May 2015.
- [30] V. Gabarra-Niecko, M. D. Schaller, and J. M. Dunty, “FAK regulates biological processes important for the pathogenesis of cancer,” *Cancer Metastasis Rev.*, vol. 22, no. 4, pp. 359–74, Dec. 2003.
- [31] G. W. McLean, E. Avizienyte, and M. C. Frame, “Focal adhesion kinase as a potential target in oncology,” *Expert Opin. Pharmacother.*, vol. 4, no. 2, pp. 227–34, Feb. 2003.
- [32] T. P. Hecker and C. L. Gladson, “Focal adhesion kinase in cancer,” *Front. Biosci.*, vol. 8, pp. s705-14, May 2003.
- [33] K. Endo, M. Mizuguchi, A. Harata, G. Itoh, and K. Tanaka, “Nocodazole induces mitotic cell death with apoptotic-like features in *Saccharomyces cerevisiae*,” *FEBS Lett.*, vol. 584, no. 11, pp. 2387–2392, 2010.

- [34] J. Nam, Y. Ino, M. Sakamoto, and S. Hirohashi, "Src family kinase inhibitor PP2 restores the E-cadherin/catenin cell adhesion system in human cancer cells and reduces cancer metastasis.," *Clin. Cancer Res.*, vol. 8, no. 7, pp. 2430–6, 2002.
- [35] S. C. Posey and B. E. Bierer, "Actin stabilization by jasplakinolide enhances apoptosis induced by cytokine deprivation.," *J. Biol. Chem.*, vol. 274, no. 7, pp. 4259–4265, Feb. 1999.
- [36] D. D. Schlaepfer, C. R. Hauck, and D. J. Sieg, "Signaling through focal adhesion kinase.," *Prog. Biophys. Mol. Biol.*, vol. 71, no. 3–4, pp. 435–78, 1999.
- [37] V. I. Rodionov, F. K. Gyoeva, E. Tanaka, A. D. Bershadsky, J. M. Vasiliev, and V. I. Gelfand, "Microtubule-dependent Control of Cell Shape and Pseudopodial Activity Is Inhibited by the Antibody to Kinesin Motor Domain," vol. 123, no. 6, pp. 1811–1820, 1993.

## **Chapter 3: Changes in cell shape are correlated with metastatic potential in murine and human osteosarcomas<sup>5</sup>**

### **3.1 Summary**

Metastatic cancer cells for many cancers are known to have altered cytoskeletal properties, in particular to be more deformable and contractile. Consequently shape characteristics of more metastatic cancer cells may be expected to have diverged from those of their parental cells. To examine this hypothesis we study shape characteristics of paired osteosarcoma cell lines, each consisting of a less metastatic parental line and a more metastatic line, derived from the former by *in vivo* selection. Two-dimensional images of four pairs of lines were processed. Statistical analysis of morphometric characteristics shows that shape characteristics of the metastatic cell line are partly overlapping and partly diverged from the parental line. Significantly the shape changes fall into two categories, with three paired cell lines displaying a more mesenchymal-like morphology, while the fourth displaying a change towards a more rounded morphology. A neural network algorithm could distinguish between samples of the less metastatic cells from the more metastatic cells with near perfect accuracy. Thus subtle changes in shape carry information about the genetic changes that lead to invasiveness and metastasis of osteosarcoma cancer cells.

### **3.2 Introduction**

Despite significant advances in treatment of cancer, it remains the leading cause of death in both men and women under 80 years of age in the US [1], with metastasis as the cause of 90% of human deaths from cancer [1, 2]. Understanding and prevention of cancer invasion and metastasis is key in reducing cancer mortality [2]. Multiple studies have pointed out that the acquisition of invasiveness appears to require changes in mechanical properties of cancer cells, which may be linked to the functional properties that are necessary for metastasis [3, 4]. To form successful metastases, tumor cells must navigate a complex, multi-stage process including: detachment from primary tumor, migration to vascular supply, intravasation, survival and transit in blood or lymphatic vessels, extravasation, and successful growth and adhesion in a new site [5]. Metastatic cells have been found to be softer or more deformable

---

<sup>5</sup> This work has been published in Biology Open and is reproduced here under the Creative Commons License. Lyons, Samanthe M., et al. "Changes in cell shape are correlated with metastatic potential in murine and human osteosarcomas." *Biology open* 5.3 (2016): 289-299. I am co-first author on the publication with Samanthe M Lyons, and my contribution to this paper is participation in the experiments and the data analysis using machine learning methods. The work is being presented here in its entirety to maintain the intellectual coherence of the project.

than non-metastatic cells in analysis with atomic force microscopy [6-9] and optical lasers [10-12]. In addition to cell deformability, multiple studies have shown that molecules responsible for cell-ECM and cell-cell adhesion interactions, including cadherins and integrins, are down-regulated or altered in cancer cells [13-17]. Cancer cell deformability is linked with invasiveness and can be an indicator of metastatic potential [3, 18-20]. However softness is just one aspect of cellular biomechanics. Cells are active objects and can exert contractile forces on the extracellular matrix; there are some reports that more invasive cells are more contractile [21]. Understanding and identifying altered biomechanical properties of aggressive cancer cells can provide crucial information for assessing the invasiveness of cancer cells.

Direct assays of mechanical changes require fairly complex and expensive instrumentation. However, one can hypothesize that changes in biomechanical properties, including changes in cytoskeletal properties and expression of adhesion proteins, would translate into changes in cell shape. It has been shown previously that changes in gene expression of genes with cytoskeletal function leads to shape changes that can be detected using morphometric characteristics [22]. Cytoskeletal gene expression changes that are signatures of metastatic capacity therefore may be detectable by morphometric analysis. Ability to detect such changes would be of great use in assessing cancer clinically.

One of the gold standards of predicting clinical outcome of cancer is tumor grading which includes assessment of cellular morphology from tumor tissue samples. Tumor grading schemes focus on overt changes in cellular morphology such as mitotic index, degree of nuclear pleomorphism and degree of tumor necrosis [23, 24]. What is not known is whether morphometric parameters of the two-dimensional shape of the cell are sensitive to the changes in cellular properties that accompany the acquisition of invasiveness.

Our paper is based on the hypothesis that subtle changes in cellular properties should manifest themselves in small but detectable morphological changes because of the importance of cytoskeletal changes for acquisition of invasive capacity. The biomechanical changes that accompany the emergence of aggressive tumor cells should be detectable by assaying the changes in shape of these tumor cells. Moreover, the observation of specific changes in shape may be linked with specific genetic changes in cancers.

Anecdotal evidence for the change in cell shape has been well documented. For example, the epithelial to mesenchymal transition (EMT), associated with development of the invasive phenotype in carcinomas [25], is often

accompanied by acquisition of a mesenchymal-like elongated spindle morphology [26, 27]. Studies have found that tumors which have formed metastases at the time of diagnosis have significantly higher grades, and thus grossly altered morphology, than non-metastatic osteosarcomas [28]. An understanding of the relationship between cell shape and the invasiveness of the cancer would lead to a deeper understanding of the relationship between carcinogenic transformation and shape regulation and may allow for a more accurate assessment of cancer outcome from cancer biopsies. Assays that can reliably estimate the percentage of potentially invasive cells in a heterogeneous sample of primary tumor cells may be of great value for guiding therapeutics.

We utilized osteosarcoma cell lines due to the high metastatic rate of the cancer. Osteosarcoma (OSA) is the most common primary bone tumor of dogs [29] and humans [30]. OSA has a high rate of metastasis and routinely forms metastases to the lung, often before the primary tumor is diagnosed and more than 80% of human OSA patients have metastases at the time of diagnosis [31-34], most with pulmonary metastases [35, 36].

Comparing the morphology of cells that were closely related (except for their degree of invasiveness) was important to minimize variables that would make the sample less homologous. We therefore sought paired osteosarcoma lines, where a more aggressive cell line was developed from a less aggressive ancestor without the use of exogenous transforming agents, as these agents may alter naturally occurring genetic changes leading to metastatic properties of osteosarcoma [37]. Without exogenous agents, we can attribute changes in cell morphology more directly to the difference in metastatic potential, as the *in vivo* development of the highly metastatic line more accurately represents the natural process of formation of metastases.

The morphology-related genetic changes that accompany transformation include both changes in cytoskeletal properties as well as changes in adhesive properties [38]. We decided to use surfaces of different hydrophobicity in our experiments to explore this possibility as more hydrophobic surfaces are less amenable to protein deposition [39] and thereby are less favorable to cell adhesion than hydrophilic surfaces. We prepared three different glass surfaces of varying hydrophobicity (Appendix II: Figure S3.1). These are Glass Detergent washed and Air dried (GDA, contact angle 27.6°), Glass Acid etched and Air dried (GAA, contact angle too small to measure), and Siliconized Ethanol Treated (SET, contact angle 99°).

We cultured four paired osteosarcoma cell lines with low and high metastatic potential: DUNN and DLM8; K12 and K7M2; MG63 and MG63.2; and SaOS2 and SAOS-LM7 on these three surfaces for 48 hours, and then fixed,

stained and imaged the cells. For simplicity we refer to each pair by the first letter of the parental line, i.e. we refer to the pairs as the D, K, M and S pairs of lines. We stained the cells for actin, the plasma membrane and nucleus. We developed a high-throughput, quantitative image analysis algorithm that chose individual cells not in contact with others, segmented, optimized and thresholded the images to obtain accurate representations of two-dimensional shape and then processed the images to extract 29 morphometric measurements: 21 cellular and 8 nuclear (Appendix II: Table S3-2). Representative images of the eight different cell lines are shown in Figure 3-1. Since here we are specifically looking for interpretable geometric differences, we did not consider other morphological representations such as shape representations in basis function expansions [40]. We then subjected the data to statistical analysis to understand the differences between the high metastatic and low metastatic cell lines, using pairwise comparisons as well as by the multivariate Principal Component Analysis (PCA) and Nonmetric Multidimensional Scaling (NMDS). We developed a neural network machine-learning algorithm to try to distinguish between cells from the high metastatic and low metastatic cell lines.

### 3.3 Results

#### 3.3.1 Pairwise Comparisons: The four paired cell lines demonstrated two distinct trends of cell shape changes

The 29 morphometric parameters were classified into five categories of cell shape: (i) projected cell size, (ii) cell roundness vs elongation, (iii) shape variability, (iv) nuclear size, and (v) nuclear shape. We identified a subset of the 29 parameters that were most often statistically significant across the various cell lines by performing pairwise t-tests between different morphometric measurements of the low metastatic line (low-met) and high metastatic line (high-met) within a paired cell line. In order to adjust for multiple testing, we performed t-tests on all 29 parameters using the Holm-Bonferroni correction [41], and identified the parameters that remain significantly different (Appendix II: Table S3-2 and S3-3). The data showed that metastatic cell lines show distinct differences in shape compared to their non-metastatic counterparts. We discovered that three of the four paired cell lines, i.e. the D, K, and S lines displayed a similar pattern of shape changes, while the fourth line, i.e. the M-line, displayed a different pattern (Figure 3-2 and Appendix II: Table S3-2). This suggests that morphological changes due to acquisition of metastatic potential fall into two distinct classes. For simplicity we denote these two patterns as type-1 and type-2. When pooled into two classes, the type-1 cells showed significant differences in 28 out of the 29 parameters we tested, while the type-2 cells showed significant differences in 26 parameters (Appendix II: Table S3-3). Highly metastatic cancer cells differ in cell volume and projected cell area.

A striking gross morphological difference between the high-met and low-met line of each pair is a systematic difference in two-dimensional projected area. Less metastatic type-1 cell lines have a significantly larger projected cell area (Figure 3-2A), and on average, the type-1 high metastatic lines are 30.7% smaller in area. The type-2 M lines showed the opposite trend, with highly metastatic cells being significantly more spread out, more than double the size of the parental line on some surfaces (Appendix II: Table S3-2). To determine whether cell volume corresponded with cell area, we measured cell volume using a handheld Scepter<sup>®</sup> counter (Methods and Materials). We found that for most pairs the cell volume and the cell area followed the same trend, i.e. the high-met line was smaller in both area and volume for two lines of type I (K and S), while the high-met M line was larger in volume and area than its less metastatic pair. The D line showed an opposite trend with a large volume but smaller area for the high-met line. However the percentage difference in mean volume is much smaller than the percentage change in mean area, suggesting a difference in the spreading behavior of the cells for both type-1 and type-2 cells. This trend of a smaller projected high-met line for type-1 cells and a larger projected high-met line for type-2 cells is consistent across all 12 measures of two-dimensional cell size (Appendix II: Table S3-2). Within the type 1 cells, the largest difference was shown by the metastatic K12 cells that were over 50% smaller on GDA while the smallest difference was shown by the LM7 cells which were about 23% smaller (Appendix III: Supplementary Table 5.2).

The change in size is also anisotropic, as the type-1 high-met lines have a minor axis that is 22% smaller, while the major axis is only 10.5% smaller, thus the minor axis percent difference is about twice that of the major axis (Figure 3-2B&C), indicating elongation of the high-met type-1 cells relative to the low-met cells. The smaller size and elongated shape of high-met lines in type-1 osteosarcoma pairs are consistent across all three type-1 lines. The type-2 M cell lines showed the opposite trend for the major and minor axes with the high-met line having a larger minor axis by 67% and larger major axis by 48%, respectively.

### **3.3.2 Highly metastatic cells differ in roundness, elongation and variability of perimeter**

The second category of cell shape assesses how round versus elongated the cell is, and is best represented by the aspect ratio. As suggested by the major and minor axis differences discussed above, the type-1 highly metastatic cell lines had a significantly larger aspect ratio than the low-met lines, indicative of cell elongation (Figure 3-2D). On average, the type-I highly metastatic cells were about 19% more elongated than the low metastatic cells. The maximum change here were the LM7 cells on SET with a 60% increase in the aspect ratio, while the smallest were the DLM8 cells with just about 13% increase on SET surfaces.

Variability of cell shape was characterized by the cell shape parameters of solidity and the coefficient of variation of radii drawn either to the cell perimeter from the center of mass of the convex hull (CV Rad Hull) or from the center of mass of the bounding circle to points on the convex hull (CV Rad Circle). The highly metastatic type-1 cell lines had more variability in radii drawn to both the perimeter (Appendix II: Table S3-2 and Figure 3-2E) and the convex hull. Another interesting measure is the circularity of the perimeter, which measures the deviation of the average shape from that of a circle. Circularity of the cell perimeter is significantly different between the high-met and low-met type-1 lines, by a little over 37% on average (Appendix II: Table S3-2).

The type-2 M lines showed the opposite trend to the ones listed above. The low metastatic cells had an aspect ratio which was about 22% larger on the GAA surface, and about 15% larger overall. The type-2 low-met line also displayed greater variability in shape than the high-met line with the CV of the perimeter radius larger by about 19% on average and by about 25% on the GAA surface. Similarly the CV of the Hull radius was larger by almost 28% on average for the low-met line. The circularity of the low-met type-2 line was also larger than its high-met partner, in contrast to the behavior shown by the type-1 lines.

### **3.3.3 Highly Metastatic Cell Lines Show Shape Differences in the Nucleus**

Interestingly, the shape parameters of the nucleus also showed statistically significant differences between the high and low metastatic lines (Figure 3-3). Nuclear size was larger for the low metastatic cells for all cell lines, including both type 1 and type 2. However while the larger nuclear size for low-met cells was statistically significant for the type-1 cells on GAA and SET surfaces, as well as for all surfaces, it was significant for the type-2 line only on the GAA surface. In line with the difference in nuclear area, the major and minor axes were larger for the low metastatic cells for all four pairs of cell lines (Figure 3-3 B&C). However, the nuclei aspect ratio showed mixed results, with the high metastatic lines demonstrating a larger aspect ratio for the D and M lines, while the low metastatic lines demonstrated a larger aspect ratio for the K and S lines (Figure 3-3D).

This analysis also underscored the fact that every cell line contained a heterogeneous collection of cell shapes. The distributions of each parameter overlapped, which was not surprising given the fact that we chose the paired lines on the grounds that they were close to each other in genetic space. In the light of these results, we asked whether we could still see these differences using a multivariate measure by utilizing all the descriptors together.

### 3.3.4 Multivariate Techniques show overlapping but distinct cell populations

We performed a principal component analysis (PCA) of the multivariate data, comparing each paired line separately (Figure 3-4 and Appendix II: Figure S3.3). The PCA showed that the geometric characteristics of each cell type were overlapping but clustered distinctly within the space formed by the first three principal components. The overlap between the characteristics of the paired cell lines indicates that the high-met line is still not too dissimilar from the low-met line. However the genetic changes that accompany the acquisition of invasive characteristics have also resulted in the cell shape parameters drifting away from that of the original cell. The maximum overlap of the first three principal components can be seen in the SAOS-LM7 and Saos2 pair (Figure 3-4D). The type-1 cells collectively show distinct clustering of the low-met and high-met populations (Figure 3-4E), which is lost when we club the type-1 and type-2 cells together (Figure 3-4F).

To test whether we could obtain a better separation using a nonlinear technique, we supplemented PCA by non-metric multidimensional scaling (NMDS) [42]. NMDS is an ordination technique that seeks to find the “best” coordinates for representing multivariate data in a lower k-dimensional ordination space. It does so by assessing and optimizing the agreement between ranked distance between data vectors in the original higher-dimensional space and the corresponding distance between them in k-space. Departure from this agreement is formally measured as “stress”. Other groups have used multidimensional scaling (MDS) to visually separate subpopulations of mesenchymal cells, further using this analysis to predict the fate of differentiating stem cells [43]. We used permutational multivariate analysis of variance to obtain the  $R^2$  values, where in the NMDS context  $R^2$  is a measure of the proportion of the distance variation of the data that is explained by cell line, i.e. from the high-met or low-met comparison within each paired line. Figure 3-5 shows the NMDS results for the best-performing surfaces, and shows that the geometric characteristics overlap between paired lines but nevertheless cluster distinctly. The  $R^2$  values are tabulated in Table 3-1 along with their p-values. The maximum proportion of the distance variation that can be attributed to cell line is 0.16 for the D-lines on the GAA surface, 0.2 for the K-lines and 0.06 for the S-lines (both on the GDA surface), and 0.24 for the M-lines on the SET surface (0.22 on GDA). All the  $R^2$  values are statistically significant and indicate that cell shape parameters of the high-met line, despite significant overlap, have diverged from those of the low-met line. Other surfaces show varying levels of overlap but in general support this conclusion (Appendix II: Figure S4.3). Interestingly data points corresponding to the high-met line for both type-1 and type-2 cells occupy a greater area in 2-space, suggesting that the high-met lines are characterized by greater heterogeneity of the shape parameters.

### **3.3.5 Identification of Cells by Machine Learning**

We asked whether these subtle but significant differences in cell shape are sufficient to construct a classification algorithm that could correctly classify the low-met and high-met cells. We wrote a neural network machine-learning algorithm to classify a cell into either the low-met or the high-met class, based on its geometric parameters alone, as described in the Methods section. Following standard practice we divided our data into three mutually exclusive subsets for training, optimizing and validating the neural network respectively. The trained algorithm was then tested blind on the third subset, the validation set, which was not used for any parameter adjustment.

The accuracy of classification of the algorithm was found to lie between 60% and 92%, (Table 3-2) suggesting as high as about 40% and as low as 8% overlap of parameters. The latter figure is much lower than expected from the preceding analysis, probably due to the efficiency of the neural network in picking up subtle differences. Single cells from all the four lines can be classified with at least 80% accuracy on at least one surface.

Next we asked whether the classification algorithm is capable of accurately classifying random samples of cells from the high-met and the low-met lines. This process can be construed as a simulation of what would happen in a clinical setting: the heterogeneous cancer cell population taken from a tumor biopsy or aspirate would be assayed using morphometric characteristics. The decision algorithm used was that if the majority of cells in the sample are of type A, the sample is of type A, and with this simple rule the algorithm achieves near perfect classification of samples into the correct cell type (Table 3-3). For every line there is at least one surface where samples can be classified with greater than 95% accuracy. Even for the S-line, where NMDS revealed only a 6% maximum difference between the cell line parameters, the neural network achieves a maximum classification accuracy of 99%.

Note that the algorithm performs relatively poorly when used to classify samples from all low-met lines against all high-met lines as compared with when it is used on type-1 cells and type-2 cells separately. Thus, shape changes in the three paired lines in type-1 appear similar enough that despite originating from different species and different cell lines, they can be accurately classified into high-met and low-met cells with reasonable accuracy.

## **3.4 Discussion**

We have shown that highly metastatic osteosarcoma cell lines derived from less metastatic parental cells show differences in shape that can be broadly classified into two types. In type-1, displayed by 3 out of 4 paired cell

lines, the highly metastatic cells are smaller in two-dimensional area, more elongated, and the radius of the cell perimeter from the center of mass is more variable. In type-2 cells, displayed by one cell line pair, the cells become larger, more rounded, with a less variable perimeter. In both cases, the distribution of the geometric characteristics that we measured was more diverse for the high-met cell line. There was a significant overlap between the parameters of the low-met and the high-met line. However use of multivariate data analysis techniques such as PCA and NMDS indicated that despite the overlap, the data points of the two cell lines clustered slightly differently from each other. The differences were sufficient to enable a trained neural network to correctly classify an individual cell as belonging to the high-met or low-met line with over 80% accuracy on at least one surface, and to almost perfectly classify samples of cells from either line.

Our data suggest that genomic changes leading to acquisition of invasive properties also give rise to detectable shape changes, and hence shape changes carry information about the state of the cell. While this study was restricted to these four pairs of osteosarcoma cell lines, we suspect that the broad conclusions are more general. Genetic changes that drive the acquisition of invasive properties may affect cell shape in various ways. For example, there could be down-regulation of adhesive proteins, a softening of the cell due to down-regulation of keratin and up-regulation of vimentin and changes in cellular contractility due to Rho-ROCK signaling. Each of these is likely to have a different set of effects on cell shape, and requires further investigation. Identifying and understanding the full typology of shape changes could have a major impact on our knowledge of metastasis and its relation with the cellular cytoskeleton. It may be eventually possible to read out genetic changes corresponding to specific changes in cell shape [22]. Determining the causal links between genetic changes and the shape of the cell are outside the scope of the present paper (and are future goals), but we provide evidence that these links exist since functional changes in invasive properties correlates with changes in cell shape.

Our discovery that shape changes fall into two types or classes is also potentially significant. Our hypothesis arising from this work is that these two classes correspond to the two modes of cell migration, i.e. mesenchymal and amoeboid [44]. Mesenchymal motion consists of cell polarization, extension, substrate binding followed by actin-based contraction and release of focal adhesions at the trailing edge. This kind of migration is dependent upon adhesion receptors, as well as on the expression of enzymes that degrade the extracellular matrix such as MMPs [44]. However when enzyme activity of MMPs is blocked, cells are found to move in an amoeboid manner, wherein the cell squeezes itself into the empty spaces in the extracellular matrix. The two modes of motion are associated with different

morphologies, with the mesenchymal mode corresponding to an elongated morphology and the amoeboid mode corresponding to a rounded morphology [45]. Recent studies have shown that the amoeboid mode of migration is associated with Rho signaling through ROCK and requires the protein ezrin, which links the cell membrane and the cellular cytoskeleton [46]. Thus downregulation of MMPs and upregulation of ezrin appears associated with amoeboid motility. The highly metastatic MG63.2 line was found to be characterized by downregulation of MMPs and upregulation of ezrin [37], suggesting that its preferred mode of motility could be amoeboid and providing an explanation for the rounded morphology it possesses as compared with the parent MG63 line. This suggests that cancer cells may acquire intrinsic preference for one mode of motility over the other as they acquire invasive characteristics, even if they are capable of switching modes of motility [44, 47]. Thus type-1 cells could have an intrinsic preference for mesenchymal motility while type-2 cells could have an intrinsic preference for amoeboid motility.

Our work suggests that it is possible to develop a consistent reproducible framework for computational morphometrics of cell shape. Further work is required to validate and refine the framework through use of other cell lines, including primary tumor lines, other cancer types and species. A reproducible quantitative framework is important for improving the subjectivity of traditional morphological analysis performed by trained histopathologists. While there is a strong correlation between tumor grade and metastatic outcome, there is not yet an ability to predict metastatic potential, based on tumor grade, in individual cases [28]. One study found low reliability of the grading of chondrosarcomas, despite the fact that grading scores guide therapeutic decision-making [48]. A summary of numerous studies on the reliability and reproducibility of urologic, prostate or renal cell cancer grading found low agreement and reproducibility [49]. Our work provides some evidence that computational image processing based morphometry to assess tumor grade may help overcome some of these challenges.

A small number of recent publications have highlighted the functional importance of cell shape by using high throughput image analysis to characterize the relation between cellular morphology and cellular properties. Treiser et al. [43] used quantitative morphometric descriptors along with MDS to predict differentiation of mesenchymal stem cells along bone or fat lineages at an early time point. They showed that subtle genetic differences between cells proceeding down the two lineages could be inferred from looking at small changes in cellular morphometrics. Yin et al. [50] utilized high throughput imaging and computational methods to classify *Drosophila* haemocyte cells into 5 discrete shapes based upon quantitative shape and morphology metrics, and argued that transitions between these shapes are switch-like. They utilized RNAi to identify genes which play a large role in regulating cell shape, including

demonstrating that the loss of PTEN induces elongation of cells. They did not however look for systematic differences between closely related cancer cell lines. While we have not tried to ascertain whether specific types of shapes are present in our data, the message of this paper is that differences in quantitative shape parameters, even within the same type, should carry useful information about the internal state of the cell. The overlap between the multidimensional shape parameters in principal component space or in NMDS space indicates that each pair of the cell lines we study has not diverged significantly in shape characteristics. However both these studies support our contention in this paper that the understanding of cell shape can give significant insight into cell properties and function.

Studies of metastasis in cancer cells have focused mainly on changes at the level of gene, protein and microRNA expression, and to a smaller extent, at the level of cellular mechanics. In contrast our work demonstrates that these changes do lead to reproducible changes in shape. More work needs to be done to construct a more comprehensive typology of shape changes in cancer, especially in other cancer types, and to achieve a mechanistic understanding of how changes in gene and protein expression result in changes in cell shape.

### **3.5 Materials and Methods**

#### **3.5.1 Cell lines and cell culture**

We utilized four paired cell lines; two of murine origin: DUNN and DLM8, K12 and K7M2, and two of human origin: MG63 and MG63.2, SaOS2 and SAOS-LM7. All metastatic lines (DLM8, K7M2, SAOS-LM7, and MG63.2) have significantly higher rates of pulmonary metastasis reported in the literature with a 200-fold increase in MG63.2, 100% efficacy of DLM8 relative to no pulmonary metastases in DUNN, 100% efficacy of SAOS-LM7, and a 90% efficacy of K7M2 relative to 33% of K12. Additionally, MG63.2, DLM8, SAOS-LM7 and K7M32 cells were reported to show greater migration and invasion than their low-metastatic counterparts: MG63, DUNN, Saos2 and K12 [37, 51-53]. MG63.2 is reported to have weaker heterotypic adhesion than MG63, while K7M2 have higher initial rates of adhesion but no difference in ultimate adhesion [37, 53].

DUNN, DLM8, K12, and K7M2 cell lines were a gift from Dr. D. Thamm (Colorado State University), MG63 and MG63.2 cell lines were a gift from Dr. D. Duval (Colorado State University), and Saos2 and SAOS-LM7 a gift from Dr. E. S. Kleinerman (MD Anderson Cancer Center). All cell lines were maintained under typical culture conditions at 37°C and 5% carbon dioxide concentration in Dulbecco's Modified Eagle Medium (DMEM) (Sigma). DMEM was supplemented with 10% fetal bovine serum (Atlas Biologicals), 20mM Hepes (Sigma), and 100 Units/ml

penicillin with 100 µg/ml streptomycin (Fisher Scientific-Hyclone). Cell lines were not independently authenticated or tested for contamination by us.

### **3.5.2 Immunofluorescence microscopy**

Cells were cultured on indicated substrate for 48 hours. Cells were stained with Wheat Germ Agglutinin, Alexa Fluor 594 Conjugate (Molecular Probes). Cells were fixed in 4% paraformaldehyde then stained with AlexaFluor 488 Phalloidin and DAPI (Molecular Probes). Cells were imaged under a 20X objective on a Zeiss Axioplan 2 fluorescence microscope (Zeiss, Thornwood, NY, USA) using filter sets: DAPI BP 445/50 blue filter, HQ Texas Red BP 560/40, and Green BP 474/28.

### **3.5.3 Preparation of surfaces**

Three different surfaces were prepared for this work from either a #1.5 22mm x 22mm glass coverslip (Richard Allen Scientific) or a #2 22mm x22mm siliconized glass coverslip (Hampton Research). The formulated surfaces follow: Glass Detergent washed and Air dried (GDA), Glass Acid etched and Air dried (GAA), and Siliconized Ethanol Treated (SET). GAA and GDA surfaces were initially prepared by sonication for 30 min. in a mild detergent solution. Following sonication, the coverslips were sequentially rinsed with milliQ (MQ) water, isopropyl alcohol (IPA), and a second rinse with MQ water prior to any further downstream processing. In the case of the GDA surface, no further downstream processing was required and the surfaces were blown dry with sterile 0.2 µm filtered nitrogen with an air gun from an in house boil off nitrogen source. GAA surface was subjected to a downstream 1M hydrochloric acid etching at 60°C for 12-16 hours. After the etching period, the coverslips went through the same rinse process described above (MQ to IPA back to MQ) before being blown dry in the same manner as the GDA surface. The SET surface was subjected to a rinse in 100% ethyl alcohol and then they were blown dry with the nitrogen gun to ensure removal of any residual liquid and debris. Prior to use in cell culture, all surfaces were exposed to UV sterilization to minimize potential contamination risks.

### **3.5.4 Contact angle measurements**

Contact angles for different substrates were measured using sessile drop method by Rame Hart Goniometer (Model # 100\_25\_M). 3 microliter of milliQ water were placed on XYZ plane using needle. Image were captured and analyzed using Rame Hart DROP Image Advanced software. Contact angle were measured for 3 different spot on one slide and this was repeated 3 times on different slides to see variability of slide's contact angle. Representative images are shown in Appendix II: Figure S3.1.

### **3.5.5 Cell volume measurement using Scepter Cell Counter:**

Volume measurements were made by the Scepter™ Handheld Automated Cell Counter, Millipore, with a 60  $\mu\text{m}$  sensor. First, cells were plated in a culture dish. Once they were ready to be split, they were trypsinized and re-suspended in 1-X PBS. After checking that cell density was in the operating range (10,000–500,000 cells/mL) the Scepter sensor was submerged in the cell suspension. The upper and lower gates were adjusted to remove debris information, and cell volume information recorded.

The distributions of cell volumes was well approximated by a log-normal distribution. Thus we log-transformed and calculated the mean and the standard error of the mean of the resulting normal distribution. The mean cell volume for the cell lines were 1.01, 1.12, 1.60, 1.41, 1.49, 1.43, 1.66, 1.80 pico-liter for DUNN, DLM8, K12, K7M2, SAOS2, LM7, MG63 and MG63.2, respectively. Thus the percentage differences between the volume of the high-met line from the low-met line are -10.9% (D lines), +11.9% (K lines), +4% (S lines) and -8.4% (M lines). We performed t-tests against the null hypothesis that the volume data for both pairs of each paired line came from a distribution with the same mean. The mean volumes for the two partners in a paired line were statistically different from each other, with p-values much smaller than 0.05 in each case.

### **3.5.6 Image Processing**

Images of isolated cells that were not in contact with other cells were chosen. To ensure adequate statistical power we picked a sample of 100 such cell images (so that the power of the test for comparing means would be 80% at 1% significance level for a half-standard deviation effect size). Images were collected blind in the sense that the students doing the imaging were not previously aware of any differentiating characteristics discovered. The images were collected at one time for each cell line on each surface. The image processing involved three distinct steps; enhancement, conversion into binary format, and automated cropping of each cell for measurement of shape metrics. Three channels were captured as described above. Prior to processing the images were converted into 16 bit TIFF images. The exported TIFF images were loaded into MATLAB where the actin, membrane, and nuclei channels were enhanced separately by contrast stretching. The actin and membrane images were combined into a single TIFF images to get full characterization of the shape. Finally, erosion with a three pixel mask was applied to sharpen edge boundaries. The enhanced TIFF images of the combined membrane-actin image and separate nuclei image were exported into ImageJ analysis for manual conversion into binary formatted images by thresholding (a representative example is shown in Appendix II: Figure S3.2). Once the images had been converted into binary format, they were

again loaded into MATLAB for shape analysis. Segmentation was achieved through use of the built-in MATLAB function toolbox so each cell could be individually cropped and reconstructed in a new image in which shape measurements (listed in Appendix II: Table S3-1) could be made on both the cell and corresponding nuclei and stored for statistical analysis. Minimum Bounding Circle was found using MATLAB function `minboundcircle`, open source code developed by John D'Errico [54]. Scripts for image processing will be made available upon request.

### **3.5.7 Data Analysis**

#### **3.5.7.1 T-Test**

Individual cell metrics were compared as discussed in the results section utilizing the built in MATLAB `ttest2` function which returns a test for the null hypothesis that the data come from independent random samples with normal distributions and equal means without assuming equal variance. This is a two-tailed test. The null hypothesis is initially rejected at a 5% significance level. All 29 parameters are then retested with the significance level determined by the Holm-Bonferroni correction for multiple tests.

#### **3.5.7.2 Principal Component Analysis (PCA):**

PCA is a method to project each sample in specific dimension to a space with equal or smaller dimension. This process is done in such a way that the first principal component has the maximum variance, second principal component has the next maximum variance, and this rule continues for subsequent components. The principal component vectors also form an orthogonal basis. We used singular value decomposition (SVD) to perform PCA on the data. First, the data was standardized so that mean of new data is zero and standard deviation is 1. Then, SVD of the data was computed and the principal components extracted from the right singular vectors of the data. Each data point was then projected into the space formed by the first three principal components, and was plotted for visualization. The variance captured by the first three principal components lie in the range 44%-47% of the total variance for every comparison made.

#### **3.5.7.3 Nonmetric Multidimensional Scaling (NMDS)**

We performed separate analyses for each of the three surfaces (GAA, GDA, and SET) and each paired cell line (D, K, S and M). First, each of the 29 cell morphology variables were relativized by dividing each value by the maximum value. Statistical software R (version 3.1.2) and package `vegan` were used to perform all statistical analyses.

The Bray-Curtis dissimilarity index was used to perform NMDS. Based on observed stress, convergence behavior, Shepard plots, and parsimony,  $k$  was chosen to be 3.

The ordination pattern was scaled as follows before plotting. First, centering was done to move the origin to the average of the axes. Second, principal components were used to rotate the configuration so that the variance of points was maximized for the first dimension, with the second dimension explaining the maximum variance of points unexplained by the first. We then displayed the ordination pattern in 2-space. For the factor “Metastatic capacity” (with levels Low and High) and the factor “cell line pair” (with levels D, K, M, and S), we generated two separate color-coded plots with 95% confidence ellipses and labeled locations of the level centroids. For GAA, GDA, and SET, observed stresses were approximately 0.07, 0.07, and 0.08.

Permutational multivariate analysis of variance using Bray-Curtis distance between cells (PERMANOVA) was used to obtain  $R^2$  values for “metastatic capacity”. Specifically,  $R^2$  is a measure of the proportion of the data (distance) variation explained by “metastatic capacity”.

#### **3.5.7.4 Machine Learning**

A multilayer perceptron (MLP) neural network with one hidden layer, adapted with permission from a version used by Dr. Charles Anderson for teaching [55], was used to classify data, and is available from the corresponding author upon request. A back-propagation learning algorithm, which uses a Scaled Conjugate Gradient, SCG, was used to design the MLP. The activation function used in this model should have some characteristics including the following. 1-It should be computationally simple. 2-For small weights it should be close to linear and as the weights magnitude increases the function should become non-linear. 3- Its derivative should be computationally simple. 4-The magnitude of the derivative should decrease as the weight’s magnitude grows. 5-The maximum value of the magnitude of the derivatives should be limited. Two common choices for activation functions are sigmoid or tanh function and we used the latter in our model. The SCG was adapted from Nabney’s netlab library [56, 57]. Each data set was partitioned into test, training and validation data at 50%, 25%, and 25% of data respectively unless noted otherwise. The test and training data sets were used to find the best attribute combinations, number of hidden units and weight parameter values in the non-linear logistic regression model. Initial parameters are chosen randomly. Training data was used to fit parameters by maximizing a likelihood function which is shown in equation 3.1.

$$L(\mathbf{w}) = \prod_{n=1}^N \prod_{k=1}^K p(C = k | \mathbf{x}_n)^{t_{n,k}}$$

$$w \text{ ere } t_{n,k} = \begin{cases} 1 & \text{w ere } n = k \\ 0 & \text{ot er} \end{cases}$$

Equation 3.1

$$, p(C = k | \mathbf{x}_n) = \frac{e^{y_{n,k}}}{\sum_{m=1}^K e^{y_{m,k}}}$$

$$, \text{ and } y_{n,k} = w_{0,k} + \sum_{m=1}^{nH} w_{m,k} \text{ Tan } (\sum_{j=0}^D v_{j,m} x_{n,j})$$

Where N is the number of sample, K is the number of classes. Function  $p(C = k | \mathbf{x}_n)$  is probability of class to be k given the sample  $\mathbf{x}_n$ . The variable  $y_{n,k}$  is predicted target value for sample n. The variables w and v are the weights in the model. The variable nH is the number of hidden units and D is the number of features for each sample. Finally, the variable  $x_{n,j}$  is  $j^{\text{th}}$  feature for  $n^{\text{th}}$  sample. To add bias parameter, D+1 features are used and  $x_{n,0}=1$  for all the samples.

Testing data was then used to calculate the percentage of cells classified correctly (test percent). To optimize the model, training and test data were repartitioned and an average test percent was calculated for different attribute combinations and function structure; we selected the optimal attribute combination and function structure based on the maximum average test percent. After the function structure was chosen the test and training data sets were combined for one last round of optimization of the weights. The optimized model was then used to predict the class that each individual cell in the validation data belongs to with no further adjustment of parameters.

To test the accuracy on random samples of cells from each population, after identifying the function structure with the training and test data, we took 100 random paired samples of 10 cells each from the validation data set. The percentage of cells in each sample predicted to be class 1 are recorded ( $P$ ). Thus the percentage of cells predicted to be class 2 = 1 -  $P$ . A decision threshold was determined utilizing the false negative rate (FNR) and true positive rate (TPR). When  $P$  was bigger than the decision threshold, the sample was classified as class 1, and when it was smaller than the decision threshold as class 2. As detailed in Appendix II: Table S3-4, the threshold was optimal at 0.6. From the total 100 pairs, the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) were calculated. Using this information, accuracy, false negative rate (FNR) and true negative rate (TPR) were calculated as defined below:

$$\text{Accuracy} = \frac{TP + TN}{TN + FN + TP + FP}$$

$$FNR = \frac{FN}{FN + TP}$$

$$TPR = \frac{TP}{FN + TP}$$

#### Acknowledgements

We gratefully acknowledge Dr. Jennifer DeLuca and Dr. Diego Krapf for helping us start these experiments, and Dr. Chuck Anderson for permission to use and adapt his neural network algorithm. We thank Dr. Dawn Duval and Dr. E. S. Kleinerman for the gift of cell lines. This work was supported through an NSF CAREER grant (PHY-1151454 to A.P.).

#### Author Contributions

A.P., S.M.L. and D.T. designed experiments, S.M.L., E.A., J.C., B.S. performed experiments, S.M.L., E.A., J.M. and A.P. analyzed data, J.M. wrote image processing code, K.S. and E.A. performed the machine-learning algorithm based analysis including code writing and results analysis, P.T. performed the NMDS analysis and A.P. and SML wrote the paper. All authors participated in editing, and read and approved the final manuscript.

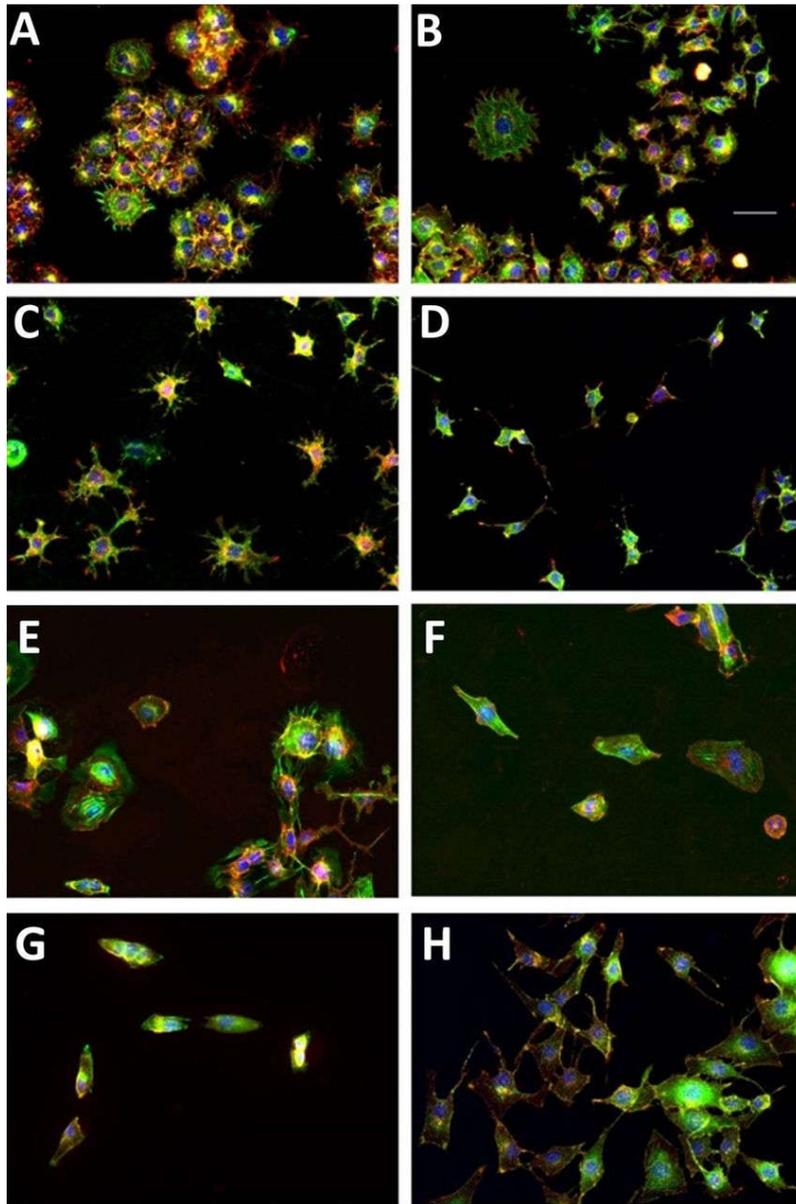


Figure 3-1 Representative images of the four cell lines using fluorescence microscopy. Each set of two panels represent the low metastatic (left) and the high metastatic (right) partner of a paired cell line. The cells nuclei (blue), the actin cytoskeleton (green) and the lipid membrane (red) of fixed cells are stained and are pseudo-colored as indicated for contrast. Note that the yellow color indicates the overlap of the red (membrane) and green (actin) channels. The cell lines are: (A) Dunn, (B) DLM8, (C) K12, (D) K7M2, (E) Saos2, (F) SAOS-LM7, (G) MG63 and (H) MG63.2. In all panels, scale bar is 50  $\mu$ m.

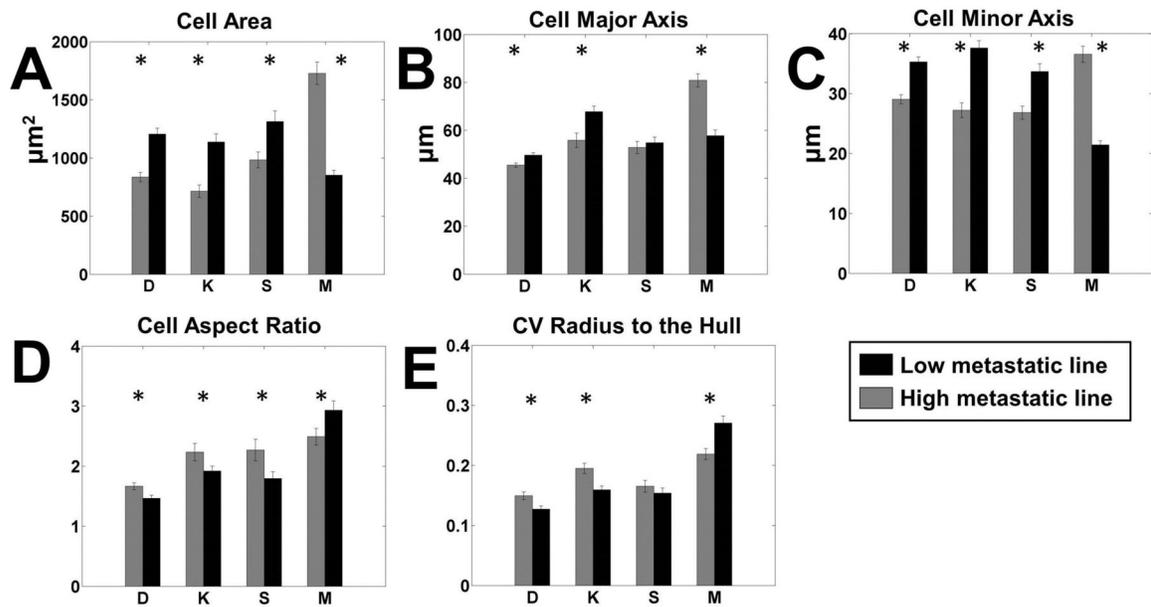


Figure 3-2 Pairwise comparison of most significant cell shape parameters. Each panel shows the comparison between high metastatic (grey) and low metastatic (black) cell lines for a single significant parameter on all surfaces. The paired lines are indicated by letters as follows. D: DUNN and DLM8; K: K12 and K7M2; S: Saos2 and SAOS-LM7; M: MG63 and MG63.2. (A) Cell area, (B) cell major axis, (C) cell minor axis, (D) cell aspect ratio and (E) coefficient of variation (CV) of the radius from the center of mass to the hull. n=100 for each cell line on each surface. \*P<0.05 by two-tailed t-test satisfying the Holm–Bonferroni criteria for all variables.

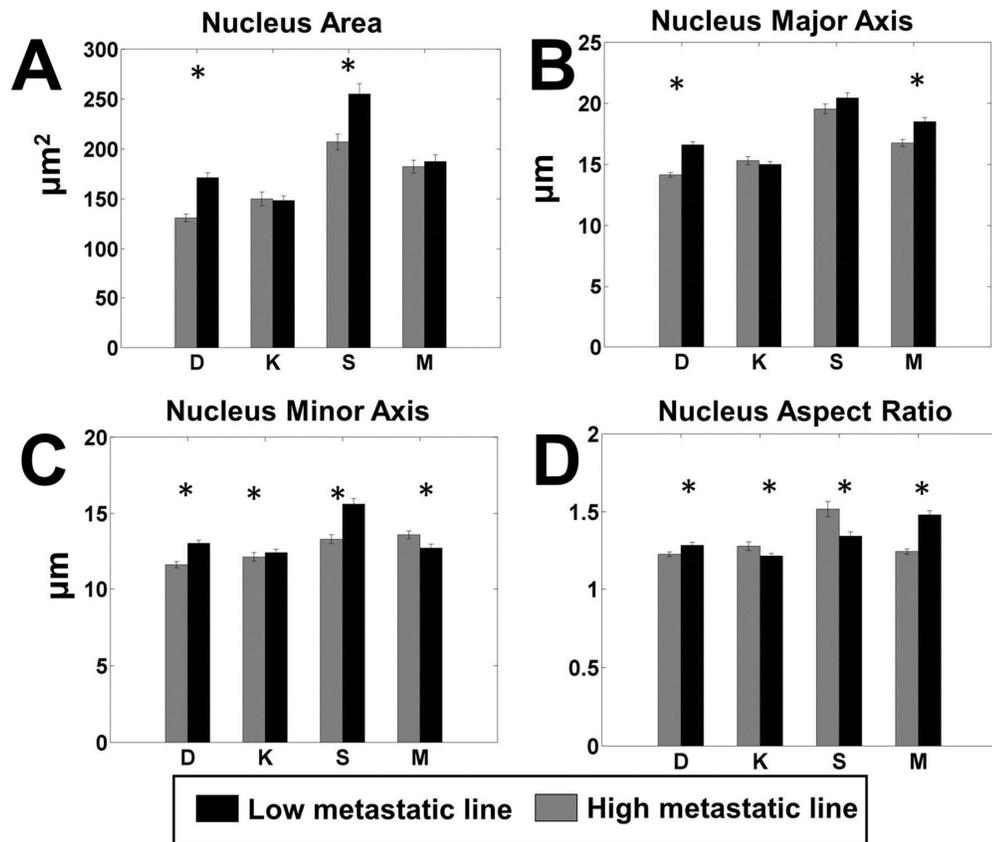


Figure 3-3 Pairwise comparison of the most significant parameters of nucleus shape. Each panel shows the comparison between high metastatic (grey) and low metastatic (black) cell lines for a single significant parameter. The paired lines are indicated by letters as follows. D: DUNN and DLM8; K: K12 and K7M2; S: Saos2 and SAOS-LM7; M: MG63 and MG63.2. (A) Nuclear area, (B) major axis of the nucleus, (C) minor axis of the nucleus and (D) aspect ratio of the nucleus. n=100 for each cell line on each surface. \*P<0.05 by two-tailed t-test satisfying the Holm–Bonferroni criteria for all variables.

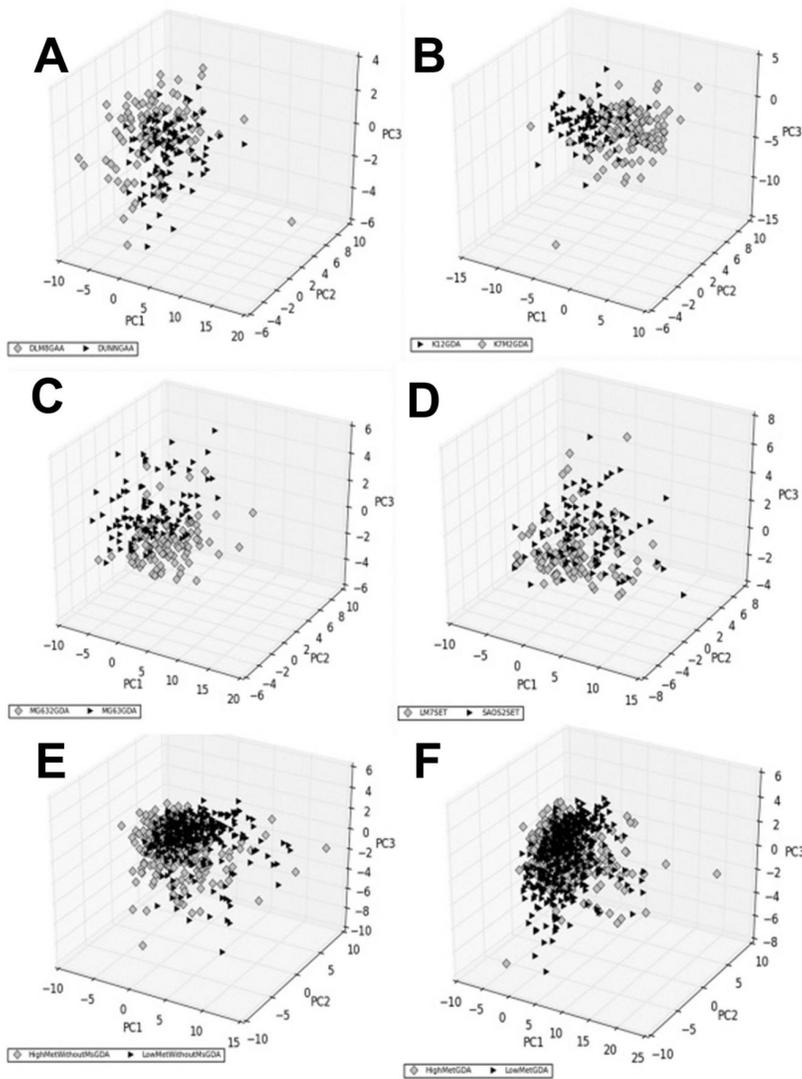


Figure 3-4 Principal components of shape characteristics. The shape characteristics data for each cell is projected onto the first three principal components of the combined data of each comparison. In this figure, comparisons for each paired cell lines that performed best are shown, as determined by visual inspection and global comparisons. The grey diamonds represent the high metastatic cell line(s) while the black triangles represent the low metastatic line(s). Each panel represents one comparison as follows: (A) DUNN vs DLM8 on GAA, (B) K12 vs K7M2 on GDA; (C) MG63 vs MG63.2 on GDA; (D) Saos2 vs SAOS-LM7 on SET; (E) all type-1 low metastatic lines versus high metastatic lines on GDA and (F) all low metastatic versus high metastatic (i.e. both types combined) on GDA. n=100 for each cell line on each surface.

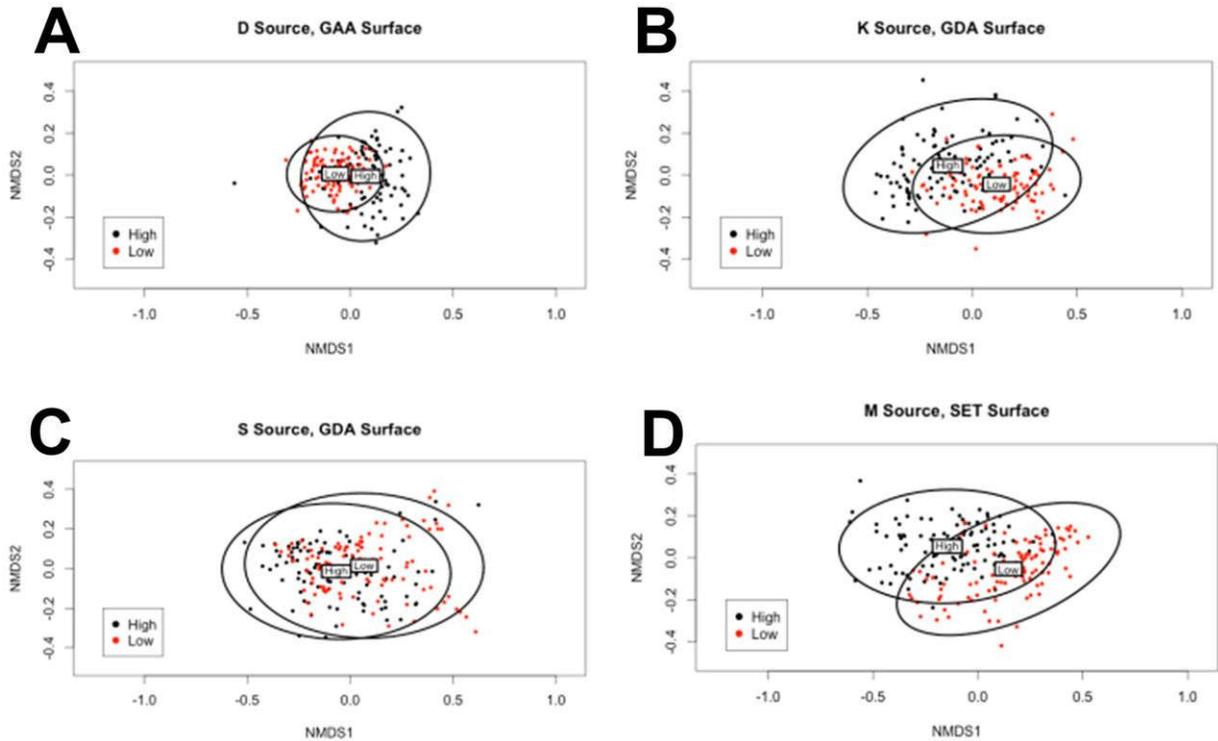


Figure 3-5 Nonmetric multidimensional scaling. Each panel represents an ordination pattern formed by comparison of geometric characteristics of a low metastatic and a high metastatic cell line on the surface that showed the highest R2 value for the pair. Each point represents the shape parameters of a single cell, plotted in black if high metastatic and red if low metastatic. The ellipses represent 95% confidence intervals with the labels 'High' and 'Low' marking the centroid positions of the corresponding cell line. The comparisons are as follows: (A) DUNN (low) and DLM8 (high) on GAA; (B) K12 (low) and K7M2 (high) on GDA; (C) Saos2 (low) and SAOS-LM7 (high) on GDA and (D) MG63 (low) and MG63.2 (high) on SET. n=100 for each cell line on each surface.

Table 3-1 Nonmetric multidimensional scaling statistics. Comparisons between a paired cell line on each surface. The ‘Stress’ is a measure of the departure of the ranked distances of the cells in the low-dimensional NMDS space from that in the original high dimensional space. The low numbers in the table indicate that NMDS was able to preserve the ranked differences. The  $R^2$  (for NMDS) is an average measure of the proportion of the total distance between cells that can be explained by the membership in the two lines, i.e. high-met and low-met. Both the  $R^2$  values and their P-values were calculated using permutation multivariate analysis of variance. The surface abbreviations are as follows: GAA, glass acid etched; GDA, glass detergent washed; SET, siliconized glass, ethanol treated. n=100 for each cell line on each surface.

Cell line	Surface	Stress	$R^2$	<i>P</i> -value
D	GAA	0.07	0.16	0.001
	GDA	0.07	0.05	0.001
	SET	0.07	0.06	0.001
K	GAA	0.06	0.06	0.001
	GDA	0.06	0.20	0.001
	SET	0.06	0.04	0.001
S	GAA	0.06	0.04	0.001
	GDA	0.06	0.06	0.001
	SET	0.06	0.05	0.001
M	GAA	0.07	0.11	0.001
	GDA	0.06	0.22	0.001
	SET	0.05	0.24	0.001

Table 3-2 Proportion of individual cells correctly identified by the neural network algorithm. The numbers represent the proportion of the sum of true positives and true negatives to all cases (see Materials and Methods section). Each row is a specific indicated comparison while the columns represent the surface on which the cells were cultured, with the last column representing the results of data from all surfaces combined. The surface abbreviations are as follows: GAA, glass acid etched; GDA, glass detergent washed; SET, siliconized glass, ethanol treated.

	GAA	GDA	SET	ALL
DUNN vs DLM8	0.9	0.6	0.76	0.62
K12 vs K7M2	0.74	0.82	0.6	0.72
Saos2 vs SAOS-LM7	0.76	0.78	0.84	0.74
MG63 vs MG63.2	0.84	0.94	0.88	0.92
All low-met vs high met	0.61	0.64	0.67	0.59
Type-1 low-met vs high-met	0.65	0.64	0.68	0.67

Table 3-3 Accuracy in sample identification of the neural network. The numbers represent the proportion of random samples (with sample size 10) from the high metastatic and low metastatic cell lines that were correctly identified by the neural network algorithm. The accuracy is the proportion of the sum of true positives and true negatives to all cases (see Materials and Methods section), hence the maximum possible accuracy is 1. Each row is a different comparison as specified, and the columns represent the three surfaces separately and combined (last column). The surface abbreviations are as follows: GAA, glass acid etched; GDA, glass detergent washed; SET, siliconized glass, ethanol treated.

Comparison	GAA	GDA	SET	ALL
DUNN vs DLM8	1	0.69	0.91	0.83
K12 vs K7M2	0.97	1	0.96	0.73
Saos2 vs SAOS-LM7	0.99	0.99	0.99	0.95
MG63 vs MG63.2	1	1	1	1
All low-met vs all high-met	0.67	0.73	0.88	0.7
All type-1 low-met vs high met	0.89	0.81	0.9	0.9

## Bibliography

- [1] Siegel, R., et al., *Cancer statistics, 2014*. CA Cancer J Clin, 2014. **64**(1): p. 9-29.
- [2] Gupta, G.P. and J. Massague, *Cancer metastasis: building a framework*. Cell, 2006. **127**(4): p. 679-95.
- [3] Suresh, S., *Biomechanics and biophysics of cancer cells*. Acta Biomater, 2007. **3**(4): p. 413-38.
- [4] Makale, M., *Cellular mechanobiology and cancer metastasis*. Birth Defects Res C Embryo Today, 2007. **81**(4): p. 329-43.
- [5] Chambers, A.F., A.C. Groom, and I.C. MacDonald, *Dissemination and growth of cancer cells in metastatic sites*. Nat Rev Cancer, 2002. **2**(8): p. 563-72.
- [6] Xu, W., et al., *Cell stiffness is a biomarker of the metastatic potential of ovarian cancer cells*. PLoS One, 2012. **7**(10): p. e46609.
- [7] Li, Q.S., et al., *AFM indentation study of breast cancer cells*. Biochem Biophys Res Commun, 2008. **374**(4): p. 609-13.
- [8] Cross, S.E., et al., *AFM-based analysis of human metastatic cancer cells*. Nanotechnology, 2008. **19**(38): p. 384003.
- [9] Cross, S.E., et al., *Nanomechanical analysis of cells from cancer patients*. Nat Nanotechnol, 2007. **2**(12): p. 780-3.
- [10] Guck, J., et al., *Optical deformability as an inherent cell marker for testing malignant transformation and metastatic competence*. Biophys J, 2005. **88**(5): p. 3689-98.
- [11] Guck, J., et al., *The optical stretcher: a novel laser tool to micromanipulate cells*. Biophys J, 2001. **81**(2): p. 767-84.

- [12] Runge, J., et al., *Evaluation of single-cell biomechanics as potential marker for oral squamous cell carcinomas: a pilot study*. Oral Dis, 2014. **20**(3): p. e120-7.
- [13] Berx, G. and F. van Roy, *Involvement of members of the cadherin superfamily in cancer*. Cold Spring Harb Perspect Biol, 2009. **1**(6): p. a003129.
- [14] Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation*. Cell, 2011. **144**(5): p. 646-74.
- [15] Cavallaro, U. and G. Christofori, *Cell adhesion and signalling by cadherins and Ig-CAMs in cancer*. Nat Rev Cancer, 2004. **4**(2): p. 118-32.
- [16] Chen, J.C., Y.C. Fong, and C.H. Tang, *Novel strategies for the treatment of chondrosarcomas: targeting integrins*. Biomed Res Int, 2013. **2013**: p. 396839.
- [17] Rathinam, R. and S.K. Alahari, *Important role of integrins in the cancer biology*. Cancer Metastasis Rev, 2010. **29**(1): p. 223-37.
- [18] Tullberg, K.F. and M.M. Burger, *Selection of B16 melanoma cells with increased metastatic potential and low intercellular cohesion using Nuclepore filters*. Invasion Metastasis, 1985. **5**(1): p. 1-15.
- [19] Paszek, M.J., et al., *Tensional homeostasis and the malignant phenotype*. Cancer Cell, 2005. **8**(3): p. 241-54.
- [20] Kenny, P.A., et al., *The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression*. Mol Oncol, 2007. **1**(1): p. 84-96.
- [21] McGrail, D.J., et al., *Actomyosin tension as a determinant of metastatic cancer mechanical tropism*. Phys Biol, 2015. **12**(2): p. 026001.
- [22] Bakal, C., et al., *Quantitative morphological signatures define local signaling networks regulating cell morphology*. Science, 2007. **316**(5832): p. 1753-6.

- [23] Straw, R.C., et al., *Canine mandibular osteosarcoma: 51 cases (1980-1992)*. J Am Anim Hosp Assoc, 1996. **32**(3): p. 257-62.
- [24] Kirpensteijn, J., et al., *Prognostic significance of a new histologic grading system for canine osteosarcoma*. Vet Pathol, 2002. **39**(2): p. 240-6.
- [25] Kalluri, R. and R.A. Weinberg, *The basics of epithelial-mesenchymal transition*. J Clin Invest, 2009. **119**(6): p. 1420-8.
- [26] Odenwald, M.A., J.R. Prosperi, and K.H. Goss, *APC/beta-catenin-rich complexes at membrane protrusions regulate mammary tumor cell migration and mesenchymal morphology*. BMC Cancer, 2013. **13**: p. 12.
- [27] Cowden Dahl, K.D., et al., *The epidermal growth factor receptor responsive miR-125a represses mesenchymal morphology in ovarian cancer cells*. Neoplasia, 2009. **11**(11): p. 1208-15.
- [28] Loukopoulos, P. and W.F. Robinson, *Clinicopathological relevance of tumour grading in canine osteosarcoma*. J Comp Pathol, 2007. **136**(1): p. 65-73.
- [29] Morello, E., M. Martano, and P. Buracco, *Biology, diagnosis and treatment of canine appendicular osteosarcoma: similarities and differences with human osteosarcoma*. Vet J, 2011. **189**(3): p. 268-77.
- [30] Ottaviani, G. and N. Jaffe, *The epidemiology of osteosarcoma*. Cancer Treat Res, 2009. **152**: p. 3-13.
- [31] Link, M.P., et al., *The effect of adjuvant chemotherapy on relapse-free survival in patients with osteosarcoma of the extremity*. N Engl J Med, 1986. **314**(25): p. 1600-6.
- [32] Ward, W.G., et al., *Pulmonary metastases of stage IIB extremity osteosarcoma and subsequent pulmonary metastases*. J Clin Oncol, 1994. **12**(9): p. 1849-58.
- [33] Kaste, S.C., et al., *Metastases detected at the time of diagnosis of primary pediatric extremity osteosarcoma at diagnosis: imaging features*. Cancer, 1999. **86**(8): p. 1602-8.

- [34] Yonemoto, T., et al., *Prognosis of osteosarcoma with pulmonary metastases at initial presentation is not dismal*. Clin Orthop Relat Res, 1998(349): p. 194-9.
- [35] Jaffe, N., et al., *Single and multiple metachronous osteosarcoma tumors after therapy*. Cancer, 2003. **98**(11): p. 2457-66.
- [36] Kager, L., et al., *Primary metastatic osteosarcoma: presentation and outcome of patients treated on neoadjuvant Cooperative Osteosarcoma Study Group protocols*. J Clin Oncol, 2003. **21**(10): p. 2011-8.
- [37] Su, Y., et al., *Establishment and characterization of a new highly metastatic human osteosarcoma cell line*. Clin Exp Metastasis, 2009. **26**(7): p. 599-610.
- [38] Cavallaro, U. and G. Christofori, *Cell adhesion in tumor invasion and metastasis: loss of the glue is not enough*. Biochim Biophys Acta, 2001. **1552**(1): p. 39-45.
- [39] Grinnell, F. and M.K. Feld, *Fibronectin adsorption on hydrophilic and hydrophobic surfaces detected by antibody binding and analyzed during cell adhesion in serum-containing medium*. J Biol Chem, 1982. **257**(9): p. 4888-93.
- [40] Pincus, Z. and J.A. Theriot, *Comparison of quantitative methods for cell-shape analysis*. J Microsc, 2007. **227**(Pt 2): p. 140-56.
- [41] Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 289-300.
- [42] Anderson, M.J., *A new method for non-parametric multivariate analysis of variance*. Austral Ecology, 2001. **26**: p. 32-46.
- [43] Treiser, M.D., et al., *Cytoskeleton-based forecasting of stem cell lineage fates*. Proc Natl Acad Sci U S A, 2010. **107**(2): p. 610-5.

- [44] Wolf, K., et al., *Compensation mechanism in tumor cell migration: mesenchymal-amoeboid transition after blocking of pericellular proteolysis*. J Cell Biol, 2003. **160**(2): p. 267-77.
- [45] Sanz-Moreno, V. and C.J. Marshall, *The plasticity of cytoskeletal dynamics underlying neoplastic cell migration*. Curr Opin Cell Biol, 2010. **22**(5): p. 690-6.
- [46] Sahai, E. and C.J. Marshall, *Differing modes of tumour cell invasion have distinct requirements for Rho/ROCK signalling and extracellular proteolysis*. Nat Cell Biol, 2003. **5**(8): p. 711-9.
- [47] Liu, Y.J., et al., *Confinement and low adhesion induce fast amoeboid migration of slow mesenchymal cells*. Cell, 2015. **160**(4): p. 659-72.
- [48] Eefting, D., et al., *Assessment of interobserver variability and histologic parameters to improve reliability in classification and grading of central cartilaginous tumors*. Am J Surg Pathol, 2009. **33**(1): p. 50-7.
- [49] Engers, R., *Reproducibility and reliability of tumor grading in urological neoplasms*. World J Urol, 2007. **25**(6): p. 595-605.
- [50] Yin, Z., et al., *A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes*. Nat Cell Biol, 2013. **15**(7): p. 860-71.
- [51] Asai, T., et al., *Establishment and characterization of a murine osteosarcoma cell line (LM8) with high metastatic potential to the lung*. Int J Cancer, 1998. **76**(3): p. 418-22.
- [52] Jia, S.F., L.L. Worth, and E.S. Kleinerman, *A nude mouse model of human osteosarcoma lung metastases for evaluating new therapeutic strategies*. Clin Exp Metastasis, 1999. **17**(6): p. 501-6.
- [53] Khanna, C., et al., *An orthotopic model of murine osteosarcoma with clonally related variants differing in pulmonary metastatic potential*. Clin Exp Metastasis, 2000. **18**(3): p. 261-71.
- [54] D'Errico, J. *minbouncircle.m*. 2015 [cited 2015 July 3]; Available from: <http://www.mathworks.com/matlabcentral/fileexchange/347>

- [55] 67-a-suite-of-minimal-bounding-objects/content/MinBoundSuite/minboundcircle.m.
- [56] Anderson, C. *CS545: Machine Learning*. 2015 [cited 2015 June 29, 2015]; Available from:  
<http://www.cs.colostate.edu/~anderson/cs545/index.html/doku.php>.
- [57] Møller, M.F., *A scaled conjugate gradient algorithm for fast supervised learning*. *Neural Networks*, 1993.  
**6**(4): p. 525-533.
- [58] Nabney, I.T., *Netlab: Algorithms for Pattern Recognition*. 2002, London: Springer.

## Chapter 4: Measuring systematic changes in invasive cancer cell shape using Zernike moments<sup>6</sup>

### 4.1 Summary

We study the shape characteristics of osteosarcoma cancer cell lines on surfaces of differing hydrophobicity using Zernike moments to represent cell shape. We compare the shape characteristics of four invasive cell lines with a corresponding less-invasive parental line on three substrates. Cell shapes of each pair of cell lines are quite close and display overlapping characteristics. To quantitatively study shape changes in high-dimensional parameter space, we project down to principal component space and define a vector that summarizes average shape differences. Using this vector, we find that three of the four pairs of cell lines show similar changes in shape, while the fourth pair shows a very different pattern of changes. We find that shape differences are sufficient to enable a neural network to classify cells accurately as belonging to the highly invasive or the less invasive phenotype. The patterns of shape changes were also reproducible for repetitions of the experiment. We also find that shape changes on different substrates show similarities between the eight cells studied, but the differences were typically not enough to permit classification. Our paper strongly suggests that shape may provide a means to read out the phenotypic state of some cell types, and shape analysis can be usefully performed using a Zernike moment representation.

### 4.2 Introduction

Cells that are from different tissues and are of different types take different shapes which are closely related to their function. For example red blood cells, whose main function is delivering oxygen to the body tissue, have a biconcave shape, which may maximize oxygen absorption through increasing the surface area and also eases movement through capillaries[1][2]. Nerve cells have long and thin extensions (called axons) to send electrical signals and other extensions (called dendrites) that receive the signals to and from other neurons in the brain and the rest of the body. This morphology allows each cell to connect to thousands of other cells. Epithelial cells in a tissue are

---

<sup>6</sup> This work has been published in Integrative Biology and is reproduced here under the Creative Commons License. Alizadeh, Elaheh, et al. "Measuring systematic changes in invasive cancer cell shape using Zernike moments", *Integr. Biol.* 8.11 (2016): 1183-1193. I am first author on this publication. The work is being presented here in its entirety to maintain the intellectual coherence of the project.

densely packed in polygonal shapes to seal certain regions[1]. The list of examples could be multiplied, but these are some pieces of evidence that overall cell shape is strongly influenced by cell function.

However even for cells belonging to one tissue or one type, experiments suggest that by changing the phenotypic state of a cell, the cell shape changes as well. By phenotypic state of the cell we mean properties such as whether the cell is migratory or not, actively dividing or quiescent etc. For example when cancer cells undergo Epithelial to Mesenchymal transitions, their migration and invasion ability increases. When this transition is chemically induced in cancer cell lines or epithelial cells, their shape changes from epithelial morphology to spindle-shaped morphology[3],[4]. Similarly by inducing apoptosis, HT-1080 cells showed smaller and rounder morphology with smooth surface in comparison with the controls[5]. In fact, this rounded morphology is often used as a marker to identify apoptosis in a cell population[6]. Mesenchymal stem cells that undergo adipocytic differentiation look rounded, while those that differentiate into osteoblasts look more spread out. Again this could be function related since a round, spherical shape allows for maximal lipid storage in adipose tissue, while cell spreading facilitates osteoblast matrix deposition during bone remodeling[7].

Even more interesting is the experimental evidence that by forcing a cell to take a specific shape we can influence the cell to adopt a particular phenotypic state. Thus cell shape also can influence phenotype. By forcing mesenchymal stem cells to adopt a flatter shape, they differentiated into osteogenic cells. However dictating a rounded shape triggered adipogenesis[7]. Other experiments have also shown that differentiation of mesenchymal stem cells can be directed by altering their shape[8]. In addition, cell spreading can modulate cell cycle progression. Human and bovine capillary endothelial cells cultured on substrates with extracellular matrix (ECM) coated islands to restrict cell shape switched from growth to apoptosis by this process[9]. Thus, cell shape not just reflects cell function, but may affect it because changing cell shape changes its behavior at least for some cell types.

The shape of a cell in two-dimensional culture is generally different from its shape in vivo. However even though two-dimensional culture is not physiological, shape on two-dimensional surfaces may carry information about cell phenotype that can be decoded if we knew enough about the shape-state relationship. Furthermore, it is not a priori apparent that shape under physiological conditions would be more informative. For example suspended cells often take a spherical shape because of the surface tension arising from plasma membrane and actin cortex, which tries to make the cell surface area the smallest possible, similar to a soap bubble. Arguably very little information can be extracted from a smooth sphere! However cell spreading on an external substrate depends on active adhesions

between the substrate and adhesive proteins, and depends upon factors like integrin expression, cellular contractility and cellular stiffness[1]. Hence, potentially there is more information buried in the shape of the spread cell than its 3D shape in a suspension.

Early stages of cell spreading on the substrate appear to be determined by passive viscoelastic processes like what happens in a bag of fluid[1],[10]. However later stages of adhesion are mostly determined by the properties of the cellular cytoskeleton. Since the cytoskeleton is composed of crosslinked polymer networks, it provides the cell with elastic restoring forces that resist shear deformations[1],[11]. These polymer networks, especially the actin network, are also dynamic, creating pushing forces through rapid polymerization at the boundaries of the cell, especially at the ends of filopodia and lamellipodia. Molecular motors slide actin filament relative to each other and generate pulling forces in the cell, that determine its contractility[1]. The actin polymer network may reorganize under these contractile forces, giving rise to structures such as stress fibers[12]. Pushing forces generated by polymerization, pulling forces generated by molecular motors and adhesion by FAs, along with fluidity of a cell will be the main determinants of cell shape[1]. In the adhering process, volume of a cell stays almost constant but its surface area can increase by up to 50%[11].

To summarize therefore, the evidence suggests that the phenotypic state of the cell affects the shape that the cell adopts on a surface, through its effect on the cytoskeletal properties of the cell. This is our main hypothesis. To test this hypothesis we will begin by examining whether changes in the phenotypic state of the cell are accompanied by reproducible changes in shape. Cancer cells are a good platform to begin with, since it is known that more invasive cancer cells have altered mechanical properties when compared with less-invasive cells.

There is significant evidence that invasive cancer cells have altered mechanical properties. They are softer[13],[14] and deform easily in response to applied stresses[15]–[17]. Their cytoskeletal fluidity changes[17]. In addition, more invasive cells are more contractile[18]. If these mechanical changes are reflected in cell shape we may be able to assess invasiveness of the cell by evaluating its shape. In a previous publication[19] we had studied paired osteosarcoma cell lines where one cell of the pair is derived from the other by selection for invasiveness. By comparing a handful of geometric parameters of the cell shape, such as area, aspect ratio, circularity and roundness, we were able to show that there were systematic shape changes in the four pairs of cell lines. However the choice of shape parameters was somewhat arbitrary since it is not clear what aspects of cell shape actually carry the most information. In order to quantify cell shape in a more rigorous way, we decided to move to a complete basis function representation

of cell shape that preserves all the information about the shape of the cell, based on Zernike moments. These were first proposed for diffraction theory by Zernike[20] and then have been widely used by the image processing community. They have also been used for character and human face recognition[21], and for classifying breast cancer tumors as benign or malignant based on mammography images of breasts[22], for automated organ recognition methods[23] and recognition of subcellular structures[24]. Zernike moments have also been previously favorably reviewed as a set of useful quantitative shape descriptors for single cells[25], but to our knowledge have not yet been applied practically towards single cell shape analysis. Zernike moments are calculated using the orthogonal Zernike polynomial basis set, and the orthogonality property implies that there is no redundancy between different moments[26] Another important property of Zernike moments is that their magnitude is rotation invariant. Finally, we can reconstruct a cell image by these moments. These characteristics of Zernike moments make them suitable to be used as shape descriptors[27].

A high dimensional set of descriptors like Zernike moments raises the problem of how to reliably compare changes in cell shape under different conditions in such high dimensional parameter space. A pairwise comparison of moments is not meaningful since each moment represents a more abstract feature of the shape (Appendix III: Figure S4.1). To solve this problem we developed a novel method we call PC vector, which looks at directions of movement in Principal Component space. Shape changes which are identical will have PC vectors that are parallel and of the same size. Using this method we analyze shape changes between less invasive and more invasive osteosarcoma cell lines, cells on different substrates and different types of osteosarcoma cells.

An alternative method of working with high dimensional data that has the added advantage of being intrinsically non-linear is machine learning using neural networks. Neural networks can be trained to distinguish one cell type from another, and the trained neural network thus becomes a classifier for that cell type. In most of the comparisons in this paper, unless otherwise stated, we trained the neural network on 50% of the data, optimized some of its parameters on 25%, and performed validation on 25% of the data that was not used for the training. The percentage accuracy of the neural network is a measure of the difference in shape characteristics of the cell lines being compared, but the false positive and false negative rates also carry information because they represent the overlap between the shapes in the cell types.

## 4.3 Experimental

### 4.3.1 Cell Lines and Cell Culture:

We utilized four paired osteosarcoma cell lines; two of them were derived from the mouse, i.e. DUNN and DLM8, K12 and K7M2, and two of them were human: MG63 and MG63.2, SAOS2 and SAOS-LM7. Each paired cell line includes one low invasive cell line and a derived high invasive cell line. High invasive cell lines are (DLM8, K7M2, SAOS-LM7, and MG63.2) and their corresponding low invasive cell lines are DUNN, K12, SAOS2, MG56, respectively [28]–[31]. We will refer to each paired cell line by first letter of their low invasive cell line's name, i.e. as D, K, S, and M pairs, followed by either L or H for low metastatic and high metastatic respectively. Thus DUNN cell lines will be referred to as DL. Appendix III: Table S4-1 lists the cell lines and the corresponding abbreviations used to refer to them. We use only the first letter when we refer to the pairs, i.e. the D pair comprises the DL and the DH lines. Representative images are shown in Ref[32].

DUNN, DLM8, K12, and K7M2 cell lines were a gift from Dr. D. Thamm (Colorado State University), MG63 and MG63.2 cell lines were a gift from Dr. D. Duval (Colorado State University), and SAOS2 and SAOS-LM7 a gift from Dr. E. S. Kleinerman (MD Anderson Cancer Center). All cell lines were maintained under typical culture conditions at 37°C and 5% carbon dioxide concentration in Dulbecco's Modified Eagle Medium (DMEM) (Sigma). DMEM was supplemented with 10% fetal bovine serum (Atlas Biologicals), and 100 Units/ml penicillin with 100 µg/ml streptomycin (Fisher Scientific-Hyclone).

### 4.3.2 Preparation of surfaces

For this study, we used three different substrates using the protocol described previously[32]. We utilized either a 1.5 22mm x 22mm glass coverslip (Richard Allen Scientific) or a #2 22mm x22mm siliconized glass coverslip (Hampton Research). Surfaces are named “Glass Acid etched and Air dried” (GAA), “Glass Detergent washed and Air dried” (GDA), “Siliconized glass Ethanol Treated” (SET). For GAA, GDA, and SET substrate contact angles were measured to be almost 0°, 27.6°, and 99°, respectively[32]. To minimize contamination, surfaces were exposed to UV light before seeding the cells.

### 4.3.3 Immunofluorescence microscopy

After culturing for 48 hours, we stained with Wheat Germ Agglutinin, Alexa Fluor 594 Conjugate (Molecular Probes), which is a membrane stain, and after fixing cells were stained with AlexaFluor 488 Phalloidin and DAPI (Molecular Probes) to stain for actin filaments and the nucleus as previously described[32]. Nuclei staining were used

to avoid cells that were undergoing division and had multiple nuclei in the dataset. Membrane and Actin staining was used to capture two dimensional spreading of cells. Imaging was carried out using a 20X objective on a Zeiss Axioplan 2 fluorescence microscope (Zeiss, Thornwood, NY, USA). The following filter sets were used: DAPI BP 445/50 blue filter, HQ Texas Red BP 560/40, and Green BP 474/28. Image processing was carried out as previously described[32].

#### 4.3.4 Zernike moments

Zernike polynomials are one of an infinite number of complete sets of polynomials in two variables,  $\rho$  and  $\theta$  which form an orthogonal basis set on the unit disk[26]. Early orders of Zernike polynomials have been plotted in cc. Each Zernike polynomial consists of a real valued radial polynomial and complex azimuthal function shown in Equation 4.1

$$V_{n,m}(\rho, \theta) = R_{n,m}(\rho)e^{jm\theta} . \quad \text{Equation 4.1}$$

Radial polynomial is in the form of Equation 4.2:

$$R_{n,m} = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s!((n+|m|)/2-s)!((n-|m|)/2-s)!} \rho^{n-2s} . \quad \text{Equation 4.2}$$

Where n represents order of the Zernike polynomials which are non-negative integers and m represents repetitions of Zernike polynomials which satisfies the constraint of  $(n-|m|) = \text{even}$  and  $|m| \leq n$ . The Zernike moments are orthogonal and they satisfy the following orthogonality condition:

$$\int_0^{2\pi} \int_0^1 V_{n,m}(\rho, \theta) V_{p,q}(\rho, \theta) \rho d\rho d\theta = \begin{cases} \frac{\pi}{n+1} & n = p, m = q \\ 0 & \text{ot erwise} \end{cases} . \quad \text{Equation 4.3}$$

By using the image function,  $f(\rho, \theta)$  and this orthogonality relation we can project the image function to each basis set and find the coefficient corresponding to that basis, the Zernike moments. Zernike moments of order n and repetition m for image function of  $f(\rho, \theta)$  can be computed using Equation 4.4.

$$Z_{n,m} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 f(\rho, \theta) V_{n,m}(\rho, \theta) \rho d\rho d\theta . \quad \text{Equation 4.4}$$

Since we are dealing with digital images, we replaced integral by summation and normalized coordinate of the image into (0, 1) using mapping transform. In computation of Zernike moments, only information inside the unit circle is used and Zernike moment computed in this case does not have any information about outside the unit circle. Thus, the following equation is used to calculate discrete form of the Zernike moment of a digital image with size N x N [22]:

$$\begin{aligned}
Z_{n,m} &= \frac{n+1}{\lambda_n} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) V_{n,m}(x,y) \\
&= \frac{n+1}{\lambda_n} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) R_{n,m}(\rho_{xy}) e^{-jm\theta_{xy}} .
\end{aligned}$$

Equation 4.5

Here  $0 \leq \rho_{xy} \leq 1$  and  $\lambda_n$  is normalization factor. Normalization factor for discrete images is number of pixels located inside the unit circle. This factor is  $\pi$  for continuous problems as introduced in Equation 4.4. Transformed radius,  $\rho_{xy}$ , and phase,  $\theta_{xy}$  at pixel of (x,y) are computed by:

$$\rho_{xy} = \frac{\sqrt{(2x - N + 1)^2 + (N - 1 - 2y)^2}}{N} .$$

Equation 4.6

$$\theta_{xy} = \tan^{-1} \left( \frac{N - 1 - 2y}{2x - N + 1} \right)$$

The reason for using the magnitude of Zernike moments instead of the complex moments is that the magnitude of Zernike moments are invariant under rotation. Thus there is no difference between a cell and its rotated image since it can be achieved by capturing the image with specific orientation of slides. In order to test this property we rotated a set of cell shapes by 120 degrees and projected its shape to the first two principal components (Appendix III: Figure S4.2). We also tested this data using t-tests and machine learning. In the t test analysis, just 3 out of 256 moments were statistically different (without controlling for multiple testing). In PCA, as demonstrated in Appendix III: Figure S4.2, each data point for each class was slightly shifted for another class and there is no sign of clustering for two classes. In Machine learning algorithm, no matter which threshold number we used all of the samples were predicted as one class. Hence, we just use magnitude of Zernike moments as our cell shape descriptors.

#### 4.3.5 Zernike Moments Calculation

Since Zernike moments depend on translation of the object, after images are processed, the centroid of the mass is translated to center of the image. In order to eliminate the parts of the image that does not have useful information, images are cropped so that the final size of images are 563 pixels x 563 pixels for every cell. Then, Zernike moments extracted as described above. We used moments of order 0 to 30 with all of their non-negative repetitions. This number was chosen on the basis of analysis of reconstructed images using Zernike moments.

Reconstruction of one rounded cell and one elongated cell using different orders of Zernike moments is demonstrated in Appendix I: Figure S2.1. From this figure it can be seen that the more moments we use, the more information we extract from cell shape and subsequently better image we get. However, computing Zernike moments is time consuming and higher order moments are susceptible to noise<sup>32</sup>. Based on the quality of the reconstructed images, we decided to use the magnitude of Zernike moments up to order 30. It can be shown from definition of Zernike moments that:

$$|Z(n)(m)| = |Z(n)(-m)|. \quad \text{Equation 4.7}$$

So, we only use magnitude of Zernike moments for non-negative repetitions, giving us a total of 256 Zernike moments as shape descriptors.

We used publically available code originally written by Tahmasbi et al[22], with some adjustments, to calculate magnitude of Zernike moments of each cell up to order 30 with non-negative positive repetitions. First, radial polynomial is calculated for a basis function using Equation 4.2. Then, basis function is computed according to Equation 4.1 and 4.2. Finally, Zernike moment is computed by projection of the image function to Zernike basis function based on Equation 4.5. This step is done for all the orders and non-negative repetitions and is repeated for all of the images of interest. The shape of the cell is thus represented by the set of the Zernike moments up to order 30, and all statistical operations described in the paper are carried out on this representation of cell shape.

#### 4.3.6 Principal Component Analysis (PCA)

PCA is a statistical procedure that finds new variables, or principal components, that are linear combinations of original variables and chosen so that the first principal component is the direction in which the data has the largest variation, the second principal component is an orthogonal direction with the next maximum variance, and so on. To calculate principal components of the Zernike moment data we use Singular Value decomposition (SVD). SVD is factorization of a matrix to three matrices in form of:

$$X = U \Sigma V^*. \quad \text{Equation 4.8}$$

Where  $U$  is called left-singular vectors of  $X$ ,  $V^*$  is called right singular vector of  $X$ , and  $\Sigma$  is diagonal matrix. Diagonal arrays of  $\Sigma$  matrix are called singular values of  $X$  and are variance of each principal component. In this Matrix variances are sorted in descending order based on value of eigenvalues. Columns of Right singular vector of  $X$  are principal components of our data and will be our new space. Each squared eigenvalue is variance of information

captured by its corresponding principal component[33]. For enhancing clustering in principal component space, we random sample the Zernike moment representation of 10 cells each, a 100 times for each class, and calculate the mean of each Zernike moment for all the cells in a sample. The data were then centered and the principal components of the data were computed using SVD. Finally, the data set was projected to the space of the first two principal components (accounting for about 60% of the total variance) to visualize clustering of two categories. For quantifying shape changes we carried out PCA without sampling on the raw data for all the cells in the comparison together, and then calculated the center of mass of the cluster of points for each class. The vector from the center of mass of one class to that of another we call the PC vector, and it represents the shape differences between the two classes. The direction of the vector in PC space represents the direction of changes while the size of the vector represents the magnitude of changes. Two vectors that are parallel represent shape changes in exactly the same direction relative to the original starting point. The magnitude of the vectors represents the magnitude of the shape changes. Thus by calculating the magnitude and the angle of the vector that connects center of mean of one class to another class we can quantitatively describe the changes in two clusters.

#### **4.3.7 Neural Network Classifier**

We developed a multilayer perceptron (MLP) neural network with one hidden layer, written in Python to classify our data. The learning algorithm used is back-propagation. As described in chapter 3, the activation function used in this study is  $\tanh(x)$ . To find the best attribute combinations, the best number of hidden units and weight parameter values we randomly partition the Zernike moment cell shape data into train, test and validation data sets with 50%, 25%, and 25% of the data respectively. The algorithm is initialized with random parameter values and the training data was used to fit parameters by maximizing a likelihood function. After finding the network we need to see what percentage of the cells are classified correctly (true test percent). We calculate the true test percent on test data set, which is not used in training the network. To find the best network, train and test data were repartitioned and average of true test percent was calculated. We can alter network setting by changing attribute combinations and model's topology. Average of true test percent was calculated and stored for different network settings. Optimized model was chosen to be the one corresponding to the structure of the maximum true test percent calculated before. Finally, the optimized model was used to predict the class of each individual cell in the validation data, which has not been used in training and optimizing the network.

We applied the algorithm to samples of cells from the validation set. We took 100 random samples of 10 cells each from the validation set for each category of cells. The percentage of cells in each sample predicted to be class 1 are recorded as P (S1). Since the exact category of the cell sample is known, false negative rate (FNR) and true positive rate (TPR) was calculated. We defined a decision threshold. When P (S1) was bigger than the decision threshold, the sample was classified as class 1, and when it was smaller than the decision threshold it was assigned to class 2. A good decision threshold number is the one that gives low FNR and high TPR. The optimal decision threshold was at 0.6[32]. From the total 100 pairs, the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) were calculated. Using this information, accuracy was calculated as defined below:

$$Accuracy = \frac{TP+TN}{TN+FN+TP+FP}$$

$$FNR = \frac{FN}{FN+TP}$$

$$TPR = \frac{TP}{FN+TP}$$

Equation 4.9

We identify good prediction as predictions that has TPR of 0.8 or higher and FNR of 0.2 is less.

We also carried out classification tests where we trained the data on one cell type and classified it on another cell type (Table 4-1 and Appendix III: Table S4-2). Here we train the neural network on 60% of the data, test on 40% of data, validate on 100% of the other paired cell line, which we call the blind set, since the neural network has not seen any of the cell shape data from this cell line during training or testing. When we train on one paired cell line and test on another, we use shape data from 60 low metastatic cells and 60 high metastatic cells for training, 40 low and 40 high for testing from one paired line. We then validate on the shape data of 200 cells of the blind set. When we train on two paired cell lines, we use exactly the same proportions, but the numbers double. The analysis of prediction accuracy is exactly as described above.

#### 4.3.8 Weight Analysis in neural networks

Machine learning algorithm uses Equation 4.10 to classify different classes, where  $X_j$  is jth attribute, i.e., the jth Zernike moment.  $V_{ij}$  is the weight used for jth Zernike moment in the ith hidden layer.  $V_{0i}$  is a constant. Each hidden layer has a weight given by  $W_i$ , which determines the importance of that layer in the classification. To find the most important moments contributing to the machine learning classification we sorted the weight matrix in descending

order based on  $W_i$  value, then we sorted the weights of the Zernike moments for each hidden layer based on the value of  $V_{ij}$ . We also separately sorted the moments based on the p-value of a simple t-test for significant difference, and we compared the two.

$$Y = \sum_{i=1}^{n^H} W_i \tan \left( \sum_{j=1}^{n^A} V_{ij} X_j + V_{0i} \right). \quad \text{Equation 4.10}$$

## 4.4 Results and discussion

### 4.4.1 Paired cell lines are distinguishable and cluster similarly.

The first question we asked was whether cell shapes were actually meaningful, in the sense that cell lines adopted a specific shape that was different from other cell lines, and closely related cell lines clustered together. In order to answer this question we decided to use a neural network binary classifier algorithm. For all such classification questions we used the algorithm in the same way. We first define two categories and train the neural network to distinguish between them based on the Zernike moments of cell shape. We then use the results of the neural network on the test set to estimate whether the network was successful in this classification.

For the first set of comparisons we tested whether the neural network can correctly classify different type of osteosarcoma cancers that are derived from the same species on the same substrates. We did this test separately for Mouse and Human cell lines, on the GDA substrate. For the mouse cell lines we compared the D lines with the K lines and for the human cell lines we compared the S lines with the M lines. Each category contained shape characteristics of at least 200 cells. We had high accuracy for both classifications, which was 100% for D cell lines versus K cell lines and 97% for M lines versus S lines on GDA substrate (Table 4-2).

We checked to see whether these results are reflected in the PCA data, and as shown in Figure 4-1, the D and the K lines cluster quite distinctly while the S and the M lines have a definite overlap but are distinguishable. Thus the shapes of the two cell lines in each pair of cell lines are related and can be distinguished from the other pair of cell lines in the same species.

### 4.4.2 Cell shapes of more invasive cancer cells are distinguishable from their less invasive parental line.

We then ran the neural network for classification of less invasive and more invasive cells. We have 4 paired cell lines and we have attempted this classification for all three substrates individually, and also all of the substrates combined as one group, which makes it 16 comparisons. **Table 4-3** demonstrates the accuracy of each prediction. It

can be seen that the prediction accuracy is almost 100% for at least one of the hydrophilic surfaces for all four cell lines. There was a small decline in accuracy for the SET substrate, especially for the K lines, but the classification for the other three lines was over 90% accurate. We achieved almost perfect prediction for D and S cell line on GAA substrate and for K and M cell lines on GDA substrate.

There were some substrates that showed a significant decline in accuracy of classification. This was true for the K cell lines on GAA and SET substrate and similarly to the S line when they are on the GDA substrate.

#### **4.4.3 Three pairs of cell lines showed the same kind of shape differences in multidimensional Zernike moment space**

The next question was whether shape changes seen in each pair of cell lines could be compared and classified. Cell shape space is high dimensional, but the changes in shape are likely to live on some lower dimensional manifold of the parameter space. Moving to principal component space is a good way of finding a representation of shape changes in a lower dimensional space. Since principal components are deterministic, we can visualize and compare changes in shape between different cell types by looking at the direction and magnitude of shape changes, as detailed in the Experimental section.

The results of the principal component analysis on the GAA, GDA, and SET substrates are shown in Figure 4-2A, Figure 4-2B, and Figure 4-2C, respectively. On the GAA substrate we can see that shape changes are mostly in the direction of PC1. PC1 decreases for D, K, S and increases for M lines. PC2 decreases for K and M lines and it increases for S and D cell line.

On GDA substrate, Figure 4-2B, the D, K, S pairs of lines show exactly the same behavior for both PC1 and PC2, which both decrease. The M line shows the opposite behavior. Similarly on the SET substrate the D, K, S lines show a decrease in PC1 and an increase in PC2, while the M cell line again shows an increase in PC1 and a decrease in PC2. Interestingly the SET substrate shows a relatively larger increase in PC2 compared with the other two substrates.

If we restrict our attention to PC1 only, then for all three substrates the change in PC1 is quite similar, with the D, K, S cell lines showing a decrease in PC1 and the M line showing an increase in PC1. Except for one case (K on GAA) the change in PC2 for the D, K, S lines are also similar and opposite to the change displayed by the M line.

To summarize, therefore, the shape changes observed are of two different types, and three pairs of cell lines, the D, K, S lines, show shape changes of one kind while the M lines show the opposite shape changes.

To test whether the shape changes seen are reproducible, we realized that we had data for the D lines on the GDA substrate from two separate experimental batches, separated by about a year. We used the PC vector technique using both Zernike moments as well as a set of 29 geometric parameters used previously[32] on the experimental data from these two batches. As can be seen in Appendix III: S4.3, despite a significant batch effect, the change in shape displayed by the PC vector is identical for the geometric parameters and very similar, and in fact identical for PC1, for the Zernike moments.

#### **4.4.4 Cell shapes can be classified accurately for type 1 shape changes collectively, but not for mixed populations.**

To test whether the two types of shape changes seen affected the accuracy of the classification, we decided to make our categories more general by choosing paired cell lines that are derived from the same species and have been cultured on the same substrate. We assigned high invasive cell lines to one category and low invasive ones to the other category, which makes it 6 comparisons (Table 4-4) in all. We found that the neural network could distinguish between high and low invasive Mouse cell lines on GDA with 94% accuracy and on SET with 87% accuracy. This comparison did not perform well on GAA. Classification for human cell lines failed on all the substrates. In light of the previous result, this is expected since the direction of shape changes for the two types are completely different, hence our algorithm cannot distinguish between the two categories.

This can also be seen when we examine the data in PC space. As shown in Figure 4-3A, the murine DL and KL cell lines are separated from their paired high invasive lines, DH and KH, with small area of overlap. In contrast, for human cell line (Figure 4-3B) we see a significantly bigger area of overlap.

Since cell shapes can be accurately classified for the type 1 shape changes, we asked whether a neural network classifier on one type 1 paired cell line would be able to accurately predict the high and low metastatic cell lines of another paired cell line. Using the data decomposition described in the Experimental section, we carried out 36 such comparisons (Appendix III: Table S4-2). Since neural networks are notoriously famous for having the capacity to over fit the data, it is not clear a priori whether they would do a good job of prediction on a cell type whose characteristics have not been used at all for training. Despite this, we found that while the majority of the 36 classification tasks do

not meet our criteria for good performance, in about one-third of the cases the classification almost meets or exceeds the criteria (Appendix III: Table S4-2). The best performing cases have been highlighted in Table 4-1, and 5 out of the 9 cases involve training on the SL and SH cell lines, suggesting that the shape changes seen in these cell lines are representative of the entire population.

#### **4.4.5 Cell shape changes on substrates are similar for hydrophilic to hydrophobic substrates, and are more significant for type 1 invasive cancer cells.**

We then asked whether there were similarities between shape changes for the same cell lines when cultured on different substrates. The average direction of the shape changes as shown by the PC vector for seven of the eight cell lines is similar for the change between the GAA surfaces to the SET surface (Figure 4-4A). All seven of these cell lines showed a negative change in PC1. The SH cell line showed opposite pattern, i.e. a positive change in PC1. The shape changes between the GDA surface and the SET surfaces are less significant and are negligible for 2 of the 8 cell lines. Of the remaining 6 cell lines, 5 show a similar, i.e. negative change in PC1. These results are shown in Figure 4-4B. The shape changes between GAA and GDA were small (Figure 4-4C). The GAA substrate is glass with contact angle of around 0 degrees, GDA substrate is glass with contact angle of around 27 degrees, and SET substrate is siliconized glass with contact angle of around 95 degrees. Hence, sorting two substrates based on their degree of the contact angle differences, will lead to GAA-SET having the maximum difference followed by GDA-SET. The observation of similar direction of changes for PC1 for the transition from a hydrophilic (GAA) to a hydrophobic (SET) surface supports the hypothesis that cell shape is strongly affected by substrate properties.

We then asked whether the cell shape changes on substrates were sufficient for a classifier to classify them correctly. We have tried all possible combinations, i.e. all the cell lines on GAA versus GDA, GDA versus SET and GAA versus SET. The accuracy results for this group of classifications are shown in **Table 4-5**. In general, when comparing cell lines on GAA versus GDA, and GDA versus SET we do not get a good accuracy, but we get good accuracy for 4 out of 8 comparisons when we compare cells that are grown on GAA substrates versus SET substrates. Again this is in line with the differences in hydrophobicity mentioned above.

It is interesting to note that in general the neural network does a good job of classification when the magnitude of the PC vector between the two classes is larger and does not do well when the magnitude is small. Additionally, as can be seen in **Table 4-5**, we have at least one good prediction for high invasive cell lines from type 1 (i.e. 3 good predictions out of 3 cell lines) in comparison with low invasive type 1 cell lines (i.e. 1 good predictions out of 3),

which suggests that low invasive cell line shapes are more robust in response to increasing hydrophobicity than high invasive cell lines belonging to type 1. The M cells were least affected by the changes in hydrophobicity of the substrate.

The neural net results are also in accordance with the visual appearance of the data clusters in two dimensional PC space. By plotting the data for each cell line on each substrate in principal component space and looking at clusters we can see that there is considerable overlap between the distributions (Appendix III: Figure S4.4).

#### **4.4.6 Even order Zernike moments contribute more to the classification.**

We did a simple t-test to see how many of the moments are statistically different in each classification and we compared those moments with the weight matrix of the neural network classification. We did not find global statistically different moments for all classifications but we did find that there are common moments for most of the classification, which are typically moments of low order with repetition of zero, such as 0\_0, 2\_0, 4\_0, 6\_0, and 8\_0.

We compared the first five important moments based on p-value with a 4 by 4 matrix of important moments based on the largest weights as discussed in the Experimental section. We found that there is a correlation between our t-test results and our machine learning result, i.e. the larger the number of statistically different moments, the better our classification was. However the converse is not true, i.e. having a large number of statistically different moments does not guarantee a good classification. For each classification we noticed that some of the important moments based on the t-test also had among the largest weights for the neural net. In most of the cases, when machine learning picks up large weights for attributes with small p-value we get a good classification.

For identifying moments that contribute the most to the PC analysis, we looked at attributes that had largest coefficients in first two principal components. Here too in most of the cases moments with repetitions of zero and of early orders, such as 0\_0, 2\_0, 4\_0, 6\_0, and 8\_0, are the most important attributes. These moments are also symmetrical. Thus these moments are sensitive to aspects of cell shape that distinguish less invasive cells from more invasive cells in osteosarcomas.

### **4.5 Conclusion**

We studied the shape characteristics of eight cell lines cultured on three kinds of substrates using a multidimensional shape representation provided by Zernike moments. Our data strongly suggests that the shape features of different cell types on a substrate are distinguishable and closely related cells are similar in their shape

characteristics. Our data shows that significant shape changes are associated with acquisition of invasive characteristics, despite heterogeneous populations. However these shape changes fell into two types, based mostly on the direction of change of the first principal component of the Zernike moments. The two types corresponded with the two types of shape changes that we found in an earlier study using a small set of descriptive measures such as area and elongation[32]. We also found that while shape changes on different substrates were small, the shape changes on going from the hydrophilic GAA substrate to the hydrophobic SET substrate was similar in PC space for the majority of the cell lines. The data therefore suggests that cell shape reflects cell state acting through the biophysical parameters of the cytoskeleton and the physical and chemical properties of the substrate.

While these results still need to be tested against other adherent cells from different types of cancers, it does appear as if cell shape can in principle be treated as a novel source of information about cellular phenotype, which may be very useful in biology and in medicine. Previous work has also pointed out that cell shape may give useful information about phenotype, and in fact cells may display “switch-like” transitions between various shape-states[34]. Cellular morphology have been shown to be linked to signaling and genetic networks[35], though the mechanistic relationships remain to be understood. Finally there is potentially more information in shape parameters and their dynamics of live cells as opposed to fixed cells[36]. Using basis function expansions like Zernike moments are likely to be useful for the dynamic analysis of shape changes too. Cell shape may be emerging as a new window for non-invasively understanding important phenotypic changes in cells. However we still need to have a better understanding of the causal relationships between cell shape and phenotype. Machine learning and other statistical classification and analysis methods may have much to offer here especially since fluorescent imaging of multiple biomarkers can be linked with shape changes in a high throughput way[37].

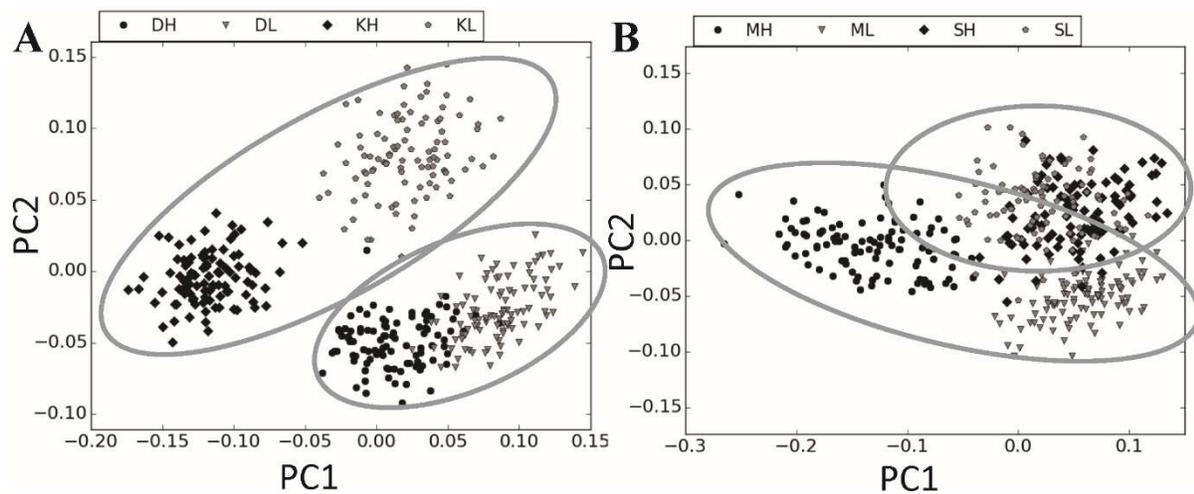


Figure 4-1 Paired cell lines in principal component space. The shape parameters are projected to the space formed by the first two principal components. Every data point represents a cell. Each cell line is represented by 100 cells, and thus a paired line is represented by 200 data points. A) Mouse D and K paired cell lines. D paired cell lines cluster different than K paired cell lines in this space. B) Human S and M paired cell lines. There is a small overlap between S lines and M lines.

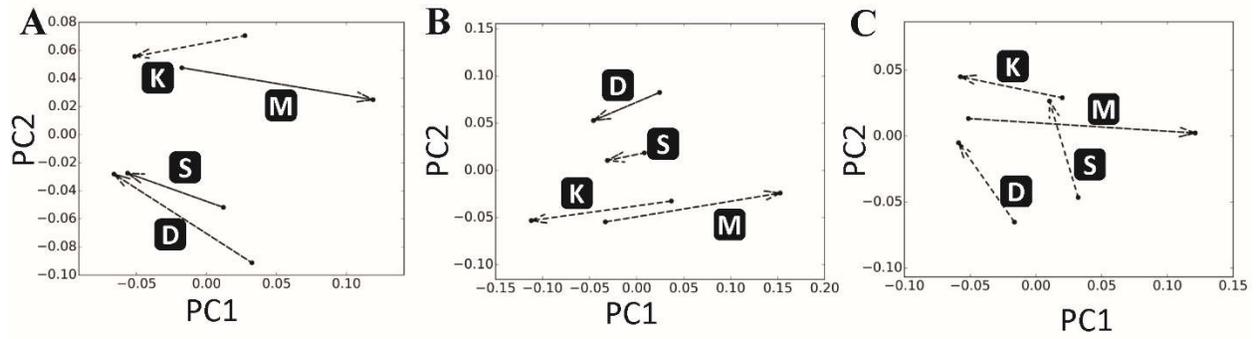


Figure 4-2 Shape differences between high and low invasive cell lines as displayed by PC vectors for each paired cell line. Every vector represents the shape differences of a single paired cell line on that specific surface. The base of the vector is positioned at the mean of the low invasive cell line and tip of the vector is the mean of the high invasive cell line. A) PC vector on GAA substrate, PC1 and PC2 captures 59.1% of the variation in the data. B) PC vector on GDA substrate, PC1 and PC2 captures 60.1% of the variation in the data. C) PC vector on SET substrate, PC1 and PC2 captures 61.2% of the variation in the data.

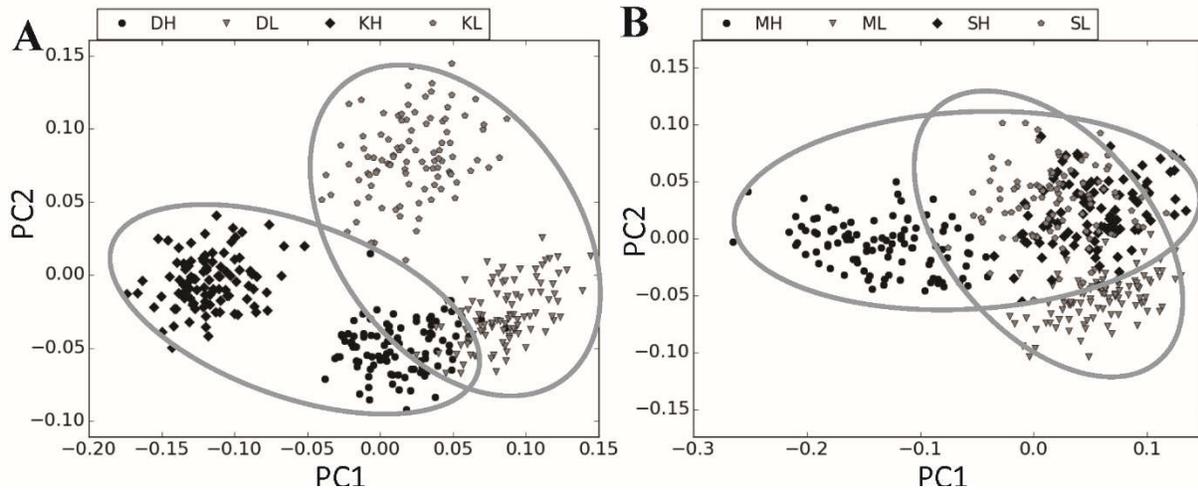


Figure 4-3 High invasive cell lines versus low invasive lines in principal component space on GDA substrate. Each point represents a cell and shape parameters of all the cells originating from a single species are plotted together. This is the same data as in Figure 4.1 but the ellipses here are drawn around cells of similar metastatic potential, i.e. low or high. A) Mouse lines. High invasive cell lines are separated from low invasive lines with small area of overlap. B) Human lines. There is significant overlap between high invasive cell lines and low invasive lines.

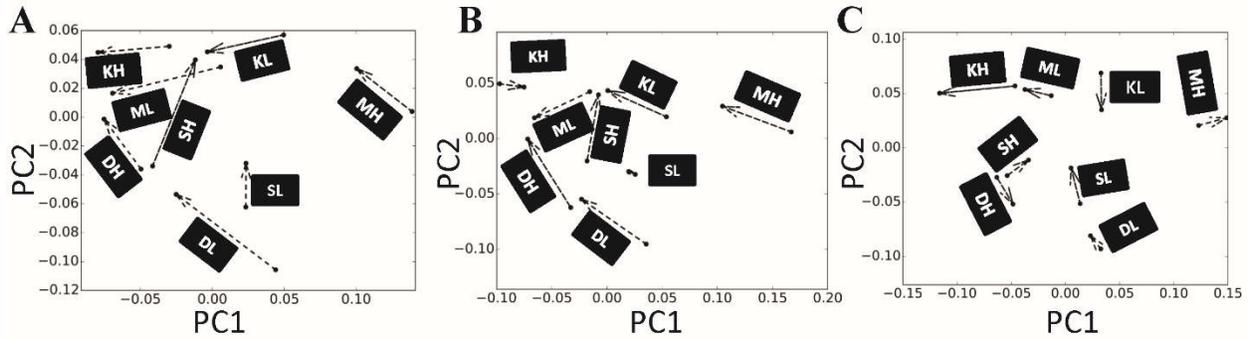


Figure 4-4 Changes in shape of the cells due to changing hydrophobicity of the substrate. As previously each vector represents the changes in shape, with the base of the vector at the mean shape for the hydrophilic substrate and tip of the vector at the mean of shape for hydrophobic substrate in the principal component space. A) GAA versus SET substrate. PC1 and PC2 captures 60 % of the variation in the data. PC1 changes for all the lines other than S line are in the same direction. B) GDA versus SET substrate. PC1 and PC2 capture 60.4% of the variation in the data. PC1 changes are in the same direction for all the cell lines other than high invasive line of S and K lines. C) GAA versus GDA substrate. PC1 and PC2 capture 59.5% of the variation in the data. PC1 changes for all the cells other than high invasive line of S lines are in the same direction.

Table 4-1 Accuracy of machine learning when the neural network is trained and validated on different cell lines. The first column shows the cell lines used for training and testing, while the second column shows the cell lines used for validation. The 9 predictions shown here are the best in a set of 36 classification tasks we conducted, whose results are shown in Table S4-2.

Training and Test cell lines	Validation cell lines	Substrate	Accuracy
SL versus SH	KL and KH	GDA	0.895
SL and DL versus SH and DH	KL and KH	GDA	0.945
SL and SH versus KL, KH	DL and DH	GAA	0.92
DL versus DH	KL and KH	GDA	0.85
DL versus DH	KL and KH	SET	0.785
DL versus DH	SL and SH	GAA	0.81
KL and DL versus KH and DH	SL and SH	GAA	0.775
SL versus SH	DL and DH	SET	0.81
SL versus SH	KL and KH	All	0.83

Table 4-2. The accuracy of distinguishing between paired lines from the same species. Machine learning algorithm can predict the type of mouse cell lines better than the human cell lines.

Classification Task	GDA
DL and DH versus KL and KH	1
ML and MH versus SL and SH	0.97

Table 4-3. The accuracy of machine learning algorithm in predicting high invasive cell lines from their parental low invasive lines. DL and DH and SL and SH on GAA, KL and KH and ML and MH on GDA are distinguishable with almost perfect accuracy.

Classification Task	GAA	GDA	SET	All
DL versus DH	0.995	0.91	0.9	0.96
KL versus KH	0.8	1	0.85	0.93
SL versus SH	0.995	0.65	0.99	0.66
ML versus MH	0.97	0.99	0.95	0.97

Table 4-4 Accuracy of machine learning in distinguishing between cells that are derived from the same species but have different degree of invasiveness. For mouse lines, on GDA and SET surfaces there is a good prediction. However, it cannot distinguish between cells that are from human.

Classification Task	GAA	GDA	SET
DL and KL versus DH and KH (Mouse)	0.8	0.94	0.875
ML and SL versus MH and SH (Human)	0.77	0.68	0.545

Table 4-5 Accuracy of machine learning in distinguishing between the same cells that are grown on different substrates. For most of the comparisons shape changes were not sufficient for our classifier to classify them accurately by our criteria.

Cell Line Used for Classification	GAA versus GDA	GAA versus SET	GDA versus SET
DL	0.73	0.89	0.965
DH	0.495	0.915	0.865
KL	0.675	0.745	0.61
KH	0.53	0.8(2)	0.49
SL	0.75	0.415	0.575
SH	0.73	0.94	0.78
ML	0.62	0.765	0.585
MH	0.285	0.72	0.715

#### Acknowledgements

We gratefully acknowledge support from an NSF CAREER grant (PHY-1151454) to A.P.

## Bibliography

- [1] U. S. Schwarz and S. A. Safran, “Physics of adherent cells,” *Rev. Mod. Phys.*, vol. 85, no. 3, pp. 1327–1381, 2013.
- [2] CK-12, “Cell size and shape - Advanced.” .
- [3] S. Ding *et al.*, “Induction of an EMT-like transformation and MET in vitro,” *J. Transl. Med.*, vol. 11, p. 164, Jul. 2013.
- [4] Y. Tang, G. Herr, W. Johnson, E. Resnik, and J. Aho, “Induction and Analysis of Epithelial to Mesenchymal Transition,” *J. Vis. Exp.*, no. 78, p. 50478, 2013.
- [5] E. J. Mahdi, A. M. Alshahrani, A. A. Abdulsatar, and J. G. Mahdi, “Morphological evaluation of apoptosis induced by salicylates in HT-1080 human fibrosarcoma cells,” *J. Microsc. Ultrastruct.*, vol. 2, no. 1, pp. 20–27, 2014.
- [6] S. Elmore, “Apoptosis: A Review of Programmed Cell Death,” *Toxicol. Pathol.*, vol. 35, no. 4, pp. 495–516, 2007.
- [7] R. McBeath, D. M. Pirone, C. M. Nelson, K. Bhadriraju, and C. S. Chen, “Cell shape, cytoskeletal tension, and RhoA regulate stem cell lineage commitment,” *Dev Cell*, vol. 6, no. 4, pp. 483–495, 2004.
- [8] K. A. Kilian, B. Bugarija, B. T. Lahn, and M. Mrksich, “Geometric cues for directing the differentiation of mesenchymal stem cells,” *Proc Natl Acad Sci U S A*, vol. 107, no. 11, pp. 4872–4877, 2010.
- [9] C. S. Chen, M. Mrksich, S. Huang, G. M. Whitesides, and D. E. Ingber, “Geometric Control of Cell Life and Death,” *Science (80-. )*, vol. 276, no. 5317, 1997.
- [10] D. Cuvelier *et al.*, “The Universal Dynamics of Cell Spreading,” *Curr. Biol.*, vol. 17, no. 8, pp. 694–699, 2007.
- [11] N. C. Gauthier, M. A. Fardin, P. Roca-Cusachs, and M. P. Sheetz, “Temporary increase in plasma membrane tension coordinates the activation of exocytosis and contraction during cell spreading,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 35, pp. 14467–14472, 2011.
- [12] S. Pellegrin and H. Mellor, “Actin stress fibres,” *J. Cell Sci.*, vol. 120, no. 20, pp. 3491–3499, 2007.

- [13] J. Rother, H. Noding, I. Mey, and A. Janshoff, “Atomic force microscopy-based microrheology reveals significant differences in the viscoelastic response between malign and benign cell lines,” *Open Biol.*, vol. 4, no. 5, 2014.
- [14] E. C. and Y.-S. J. and J. T. and R. W. and J. R. and J. K. G. Sarah, “AFM-based analysis of human metastatic cancer cells,” *Nanotechnology*, vol. 19, no. 38, p. 384003, 2008.
- [15] J. Guck *et al.*, “Optical Deformability as an Inherent Cell Marker for Testing Malignant Transformation and Metastatic Competence,” *Biophys. J.*, vol. 88, no. 5, pp. 3689–3698, 2005.
- [16] W. Xu, R. Mezencev, B. Kim, L. Wang, J. McDonald, and T. Sulchek, “Cell Stiffness Is a Biomarker of the Metastatic Potential of Ovarian Cancer Cells,” *PLoS One*, vol. 7, no. 10, p. e46609, 2012.
- [17] M. F. Coughlin *et al.*, “Cytoskeletal stiffness, friction, and fluidity of cancer cell lines with different metastatic potential,” *Clin. Exp. Metastasis*, vol. 30, no. 3, pp. 237–250, 2013.
- [18] J. M. and Q. M. N. K. and J. A. I. and M. R. D. Daniel, “Actomyosin tension as a determinant of metastatic cancer mechanical tropism,” *Phys. Biol.*, vol. 12, no. 2, p. 26001, 2015.
- [19] S. M. Lyons *et al.*, “Changes in cell shape are correlated with metastatic potential in murine and human osteosarcomas,” *Biol Open*, vol. 5, no. 3, pp. 289–299, 2016.
- [20] F. Zernike, *Diffraction theory of the cut procedure and its improved form, the phase contrast method.* .
- [21] C. Singh, E. Walia, and N. Mittal, “Rotation invariant complex Zernike moments features and their applications to human face and character recognition,” *IET Comput. Vis.*, vol. 5, no. 5, pp. 255–265, 2011.
- [22] A. Tahmasbi, F. Saki, and S. B. Shokouhi, “Classification of benign and malignant masses based on Zernike moments,” *Comput. Biol. Med.*, vol. 41, no. 8, pp. 726–735, 2011.
- [23] D. Broggio *et al.*, “Comparison of organs’ shapes with geometric and Zernike 3D moments,” *Comput. Methods Programs Biomed.*, vol. 111, no. 3, pp. 740–754, 2013.
- [24] M. V Boland and R. F. Murphy, “A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells,” *Bioinformatics*, vol. 17, no. 12, pp. 1213–1223, Dec. 2001.
- [25] Z. PINCUS and J. A. THERIOT, “Comparison of quantitative methods for cell-shape analysis,” *J. Microsc.*,

- vol. 227, no. 2, pp. 140–156, Aug. 2007.
- [26] S.-K. Hwang and W.-Y. Kim, “A novel approach to the fast computation of Zernike moments,” *Pattern Recognit.*, vol. 39, no. 11, pp. 2065–2076, 2006.
- [27] H. Mahi, H. Isabaten, and C. Serief, “Zernike Moments and SVM for Shape Classification in Very High Resolution Satellite Images,” *Int. Arab J. Inf. Technol.*, vol. 11, no. 1, pp. 43–51, 2014.
- [28] T. Asai *et al.*, “Establishment and characterization of a murine osteosarcoma cell line (LM8) with high metastatic potential to the lung,” *Int. J. Cancer*, vol. 76, no. 3, pp. 418–422, 1998.
- [29] C. Khanna, J. Prehn, C. Yeung, J. Caylor, M. Tsokos, and L. Helman, “An orthotopic model of murine osteosarcoma with clonally related variants differing in pulmonary metastatic potential,” *Clin. Exp. Metastasis*, vol. 18, no. 3, pp. 261–271, 2000.
- [30] S.-F. Jia, L. L. Worth, and E. S. Kleinerman, “A nude mouse model of human osteosarcoma lung metastases for evaluating new therapeutic strategies,” *Clin. Exp. Metastasis*, vol. 17, no. 6, pp. 501–506, 1999.
- [31] Y. Su *et al.*, “Establishment and characterization of a new highly metastatic human osteosarcoma cell line,” *Clin. Exp. Metastasis*, vol. 26, no. 7, pp. 599–610, 2009.
- [32] S. M. Lyons *et al.*, “Changes in cell shape are correlated with metastatic potential in murine and human osteosarcomas,” *Biol. Open*, 2016.
- [33] D. P. and M. G. Berrar, W. Dubitzky, Eds., *A Practical Approach to Microarray Data Analysis*. Springer, 2009.
- [34] Z. Yin *et al.*, “A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes,” *Nat. Cell Biol.*, vol. 15, no. 7, pp. 860–71, Jul. 2013.
- [35] C. Bakal, J. Aach, G. Church, and N. Perrimon, “Quantitative Morphological Signatures Define Local Signaling Networks Regulating Cell Morphology,” *Science (80-. )*, vol. 316, no. 5832, 2007.
- [36] S. Gordonov, M. K. Hwang, A. Wells, F. B. Gertler, D. A. Lauffenburger, and M. Bathe, “Time series modeling of live-cell shape dynamics for image-based phenotypic profiling,” *Integr. Biol.*, vol. 8, no. 1, pp. 73–90, 2016.
- [37] B. Nicola, L. S. Yang, L. H. Kuan, and L. Lit-Hsin, “Large-scale image-based screening and profiling of

cellular phenotypes,” *Cytom. Part A*, vol. 91, no. 2, pp. 115–125, Jul. 2016.

## Chapter 5: Quantifying single cell shape changes with cancer progression

### 5.1 Introduction

Although methods of cancer diagnostics and treatment have improved over the past few decades, it still takes lives of many people and is ranked as the second leading cause of death in the US [1]. In contrast to non-metastasizing cancer tumors that do not propagate to other parts of the body, metastasizing tumors grow uncontrollably and start traveling to other parts of body and colonize other organs. This process is responsible for 90% of the deaths from cancer tumors [2], [3]. Cancer statistics show that 5-year cancer survival rate is higher for patients who have been diagnosed in early stages of their cancer disease [4]. Accurate diagnosis leads to an efficient treatment plan and can help eliminate the harsh therapy unless it is necessary. Therefore early and accurate cancer diagnosis is a key factor in the success of patient's treatment plan. To diagnose a cancer, a complex combination of clinical and histopathological data is utilized, but cancer diagnosis could be improved by incorporation of other information. This paper is based on the hypothesis that it is possible to gain additional useful information about cancer cells based on a quantitative examination of their shape.

The shape of a cell is dynamic and at any given time it is the outcome of the surrounding mechanical and biochemical cues, internal forces and cell's signaling state[4]. Numerous reports have shown that altering the phenotypic state of cells leads to changes in shape of the cells or vice versa[5]-[6]-[7]-[8]. These experiments demonstrate that there is a feedback loop between the shape of the cell and its phenotypic state and therefore we can learn about the latter by using the information from the former.

Our central hypothesis is that cell shape and the invasive propensity of the cancer cell are linked together; implying that useful information regarding the invasive phenotype of the cell is buried in its shape. By examining the relation between cancer cell shape and its invasiveness we can learn more about cancer cells and improve the diagnosis of the metastatic potential of cancer.

The existence of the correlation between the shape of the cell and cancer phenotype has been shown previously. Sailem et al have used cell shape to predict tumor grade [9]. Sero et al have used cell shape to predict the changes in NF $\kappa$ B ratio which is a transcription factor involved in cancer progression [10]. They found that activation of Yes-associated protein (YAP) which participates in tumorigenesis is correlated with cell shape for MCF10A

cells[11]. Pasqualato et al found that cell shape is significantly different for Human colon cancer cell line HCT-8 and 5-FU chemoresistant line HCT-8FUres[12].

In our previous publications we were able to use cell shape and a neural network to build a model to distinguish between low invasive osteosarcoma cancer lines and their paired high invasive cancer line with 99%-100% accuracy [13], [14]. We then trained the neural network on shape information of one paired cell line and tested the model on another paired cell line achieving an accuracy of 92% [15]. Thus, even though the paired cell line of the training set were derived from different species and had different initial and final cell shape, when they underwent metastasis their shape changes followed a similar pattern. The neural network was able to pick up those patterns in one pair and correctly classified low invasive and high invasive line of the other pair.

In this study, we seek to find out whether acquisition of metastatic potential is accompanied with specific cell shape changes across different types of cancers. We use three different cancer cell types which are breast cells, Osteosarcoma cancer cells, and human retina cells. Names of the cell lines used in this study along with their acronyms used throughout the paper are listed in Table 5-1.

The cell lines used as breast cells are MCF10A, MCF7 and MDA-MB-231. MCF10A is a non-tumorigenic human mammary gland epithelial cell line. This cell line, which is immortalized, has been used in vitro studies as a model for normal breast cells[16]. MCF7 cell line is a breast cancer line which exhibits a low aggressive behavior. MDA-MB-231 is a triple negative breast cancer cell line which is derived from a metastatic site and is highly invasive and is used as a model for metastasis. MDA-MB-231 cells show a high degree of heterogeneity in their shape and gene expression which lead to poor prognosis[17]. Representative images of breast cells are shown in Appendix IV: Figure S5.1A.

The next cell type used in this study is derived from human Retinal pigment epithelium cells, ARPE-19. In this set of cell lines, the parental cell line is transfected step by step to transform to three cancer cell lines. First, they are engineered to express hTERT for immortalization. Then, they are transfected by p53dd. This process inactivates the tumor suppressor. Afterwards, they are co-transfected with Cyclin, which increases proliferation signaling, and CDK4 R24C, which prevents the kinase from interacting with p16 and p15 CDK inhibitors. Finally, the cells are transfected with Rasv12 or AKTmyr, or MekDD oncogenes. Representative images of retina cells are shown in Appendix IV: Figure S5.2.

The last cell type used in this study is osteosarcoma cancer cells. Four paired osteosarcoma cancer cell lines are used to study their shape changes. In each paired cell line, there is one high metastatic cell line which is derived from low metastatic line with selection for metastasis. Cell lines used here are DUNN and DLM8, K12 and K7M2, SAOS2 and LM7, MG63 and MG63.2. First two pairs are derived from mouse and the last two pairs are derived from human. For this study, we have used the data from our previous publications [14], [18].

In this study, by using cell lines that have the same parental line, we can study changes in the shape of the cells caused by cancer progression. In order to determine whether shape of the cell is related to its invasiveness, we first need to quantify it. As detailed in Chapter 2, we have developed 256 shape quantifiers with 9 different shape categories which each quantify some aspect of shape and structure of the cell and its nuclei. To test if shape changes between different classes of cells are statistically different, we use statistical tests such as T-test and principal component analysis. To test if shape differences between two different classes of the cells are sufficient to distinguish between them, we use neural network classifier.

## 5.2 Methods

### 5.2.1 Cell culture:

MCF10A, a human normal epithelial cell line, was obtained from the American Type Culture Collection (ATCC, Manassas, VA). MCF7 and MDA-MB-231 cells were gifts from Dr. Brian McNaughton at Colorado State University (CSU). DUNN, DLM8, K12, and L7M2 cell lines were gifts from Dr. D. Thamm (CSU). MG63 and MG63.2 cell lines were gifted from Dr. D. Duval (CSU). SAOS2 and LM7 cell lines were donated from Dr. E. S. Kleinerman (MD Anderson Cancer Center). Retinal laboratory-transformed cell lines were developed in Jennifer DeLuca's laboratory at CSU and were gifted to our lab along with normal APRE19 cell line.

Retinal cell lines were cultured in DMEM/F12 (Hyclone) supplemented with 10% Equafetal Fetal Bovine Serum (Atlas Biologicals) and 1% antibiotics (Hyclone). As recommended by ATCC, MCF10A breast cancer cells were cultured in MEGM kit (Lonza/Clonetics Corporation) without the GA-1000 (gentamycin-amphotericin B mix) provided with the kit and with 100 ng/ml cholera toxin. All other cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Sigma) supplemented with 10% Equafetal Fetal Bovine Serum (Atlas Biologicals), 100 Units/ml penicillin and 100 µg/ml streptomycin (Fisher Scientific-Hyclone). MCF7 and MDA\_MB\_231 cells were supplemented with 20mM Hepes (Sigma).

All cell lines were maintained under standard culture conditions at 37 °C and 5% carbon dioxide concentration. They were seeded either on glass detergent washed and air dried (GDA) substrate using the protocol described in Chapter 3 or on GDA substrates coated with Fibronectin (FN) with density of 2 $\mu$ g/cm<sup>2</sup>. Breast cells were seeded on FN substrate (they did not attach to GDA substrate) in duplicates. Retinal cells were seeded on GDA substrates, in triplicates, and FN substrates, in duplicates. Osteosarcoma cell lines were cultured on GDA substrates. All cells were seeded with 7.5 E4 cells/ml density in 6 well plates on substrate of choice. Retinal and breast cells were fixed after 22 hours and Osteosarcoma cells were fixed after 48 hours. After fixing, cells were fluorescently labeled with DAPI (Bio legend) for nuclei and with Acti-stain 488 phalloidin (Cytoskeleton, Inc.) for actin.

### **5.2.2 Image Processing**

In order to strike a balance between the throughput and the accuracy of the image processing, the image processing is designed to be fully supervised by the operator to reduce the number of artifacts and also automated as much as possible to speed-up the process. The image processing code is stream-lined into a consecutive step-by-step workflow which consists of four steps as follows. 1) A graphic user interface (GUI) enables the binary thresholding of the actin channel and nucleus channel. While a thresholding value is suggested automatically by Otsu's method, users can easily adjust the thresholding value using a slide bar in the GUI by visually checking the original image and the thresholded image displayed side-by-side in the GUI. 2) Cell declustering in the thresholded images is done using an optimized template in the open-source software CellProfiler [19]. 3) The outputs from the CellProfiler are then visually examined by the operator and corrections can be made if necessary using the modules functionalized into this step. To facilitate the following analyses, each cell is centered and saved separately into one 1024x1024 image. Except for CellProfiler, all the other image processing codes were programmed in-house using Matlab (Mathworks) and are available upon requests. Also available are the detailed protocol on how to use the image processing codes and the CellProfiler template.

### **5.2.3 Shape quantification:**

We collected 100 samples per cell line per replicate. The morphological measurements are carried out as described in Chapter 2. We quantified shape of each cell using a large number of measures that can be classified into three broad categories and 9 sub-categories. The three categories are Texture, Spreading and Irregularities. The Textural measures use the gray scale image of labelled actin, and they include the subcategories described as Band based, Grayscale fractal dimension, and Gray scale measures. The Spreading category includes subcategories

involving geometric properties of the nucleus and the cell, which are based on the binary image of nuclei and cells. It also includes Zernike moment and Convex hull geometric parameters of binary image of the cells. Using information about radius of each pixel in the boundary, the different measures in the Irregularity category, including measures of Waviness and Roughness, were calculated. List of measures in each 9 shape categories are detailed in Chapter 2.

#### **5.2.4 Statistical testing of differences**

To find the best parameters that describe the shape of the cell for each cell line, we used T-test analysis. The criteria for “desired feature” were that the selected feature should be similar in replicates and be different for different cell lines. These criteria were based on p-value using the T-test between two cell populations under comparison. We tested the similarity criterion on comparison of two replicates in one cell line. Dissimilarity criterion was tested on different cell lines when all the replicates were combined. The higher the p-value is, the more similar two classes are. So good measures are features with high p-value in similarity tests and low p-value in dissimilarity tests for most of the comparisons.

This analysis was conducted for each cancer cell type separately, and the results are summarized in heat map plots shown in Figure 5-1. In this figure each shape category and each cancer cell type is plotted in separate panel. Each column represents one shape parameter. The first heat map row for retina cell type and breast cell type are the results for similarity comparisons, and each row in this panel is for one cell line when one replicate is compared with another replicate. The second heat map row for these cell types represents the dissimilarity results and each row is comparison of combined replicates from one cell line against combined replicates of other cell line. Since there are no replicates for Osteosarcoma cell lines data is not available for similarity test of this cell type. Therefore, only dissimilarity criterion was tested on this cell type.

#### **5.2.5 PCA analysis**

Since we deal with high dimensional data due to large number of parameters used to quantify cell shape, Principal Component Analysis (PCA) was used to reduce the dimension and have better visualization of shape of different cell lines/types. In principal component space, first principal component (PC1) captures the highest variation in the data and the second principal component (PC2) captures the next highest variation in the data. In our data sets, typically the first 4 principal components captured 99% of the variation. So, we only worked with first 4 principal component for each shape category. Each shape category captured information about different aspect of cell shape and structure. Therefore, we performed PCA on each 9 shape categories separately. PCA is not necessarily a

classification method. As described in Chapter 2, in each shape category we used T-test to choose principal component which lead to better separation for all of the classes under comparison. In other words, we picked the PC whose worst case, i.e. the largest p-value among all comparisons, was better (i.e. smaller) than that of any other PC, so that it was the best single measure for distinguishing between all the comparisons for that shape category. After choosing the best PC for each shape category, we scatter plotted the data. In the scatter plots each point in every class is average of 10 random samples from one class in principal component space. However, the selection of PC is performed by using information of single cells.

### 5.2.6 Machine Learning Method:

We used multilayer perceptron (MLP) neural network with one hidden layer to distinguish between two cell classes using their shape information. As described in Chapter 3,  $\tanh(x)$  is used as activation function. Cells of each class were divided into 5-folds partitions. Three partitions were used to train the model. One partition was used as validation data set to optimize the model. One other partition was used to test the model. The test set was used in neither training nor optimizing the model. This procedure was repeated 20 times until all of the partitions were used once in the test set and once in the validation set. The average accuracy was reported afterwards. As introduced in Chapters 3 and 4, we have used majority rule when reporting the outcome of the prediction. In majority rule, instead of predicting single cell's class we predicted the classes of sample of 10 cells, all collected randomly from one class. If six or more of the cells were predicted to be class 1, the whole sample was assigned to be from class 1. Otherwise it is assigned to be class 2. Since we know which class the sample belonged to, we calculated the true positive rate (TPR) and the true negative rate (TNR) of the classification. In our analysis, sometimes the classes did not have equal number of samples. So, we reported the accuracy of the prediction of the model to be:

$$Accuracy = \frac{TPR \cdot nS_P + TNR \cdot nS_N}{nS_P + nS_N}, \quad \text{Equation 5.1}$$

Where  $nS_P$  is number of samples in the positive class and  $nS_N$  is number of samples in the negative class.

It should be emphasized that the number of samples collected from each class for majority rule is equal to number of samples in each class. We used the accuracy calculated from the majority rule prediction to optimize the model. To optimize the model structure we had to find the optimum number of hidden units and the best attribute combinations. This was conducted by removing the lowest weights in the model as described in Chapters 3 and 4.

## 5.3 Results and discussion

### 5.3.1 Analyzing the variability of cell shape quantifiers for different cell lines

Our own [14], [15] and others previous work has shown that there is significant heterogeneity in shape, even for a single cell line, that reflects the heterogeneity of the cell state. Moreover, the heterogeneity can be expected to be different for different cell lines. Thus, one important question is whether the shape characteristics of a cell line are nevertheless identifiable despite this heterogeneity. This question can be posed as a question about the distribution of cell shape quantifiers of a particular cell line in shape space, what we call the morph-space of the cell line. If this distribution is reasonably compact, at least for a set of quantifiers, the morph-space of the cell line is reproducible and the cell line is easily distinguishable by those shape quantifiers.

#### 5.3.1.1 Different cell lines have different degrees of shape heterogeneity.

As explained in the method section, shape differences between two classes can be quantified using T-test analysis for each shape feature and visualized in a heat map plot. To visualize the shape similarities within each cell line and between their different replicates, the T-test analysis was performed and the summary of the results are shown in Appendix IV: Figure S5.3. In this figure the lighter the color is, the more similar the replicates of each cell lines are. As shown in this figure, different cell lines have different degrees of shape differences between their replicates. For example, the replicates of normal retinal cell lines are more similar than the replicates of AKT<sup>myr</sup> cell line on Fibronectin substrates, Appendix IV: Figure S5.3A. Moreover, the shape differences between replicates depend on substrate properties. As an example, the replicates of normal retina cells on fibronectin substrates, Appendix IV: Figure S5.3A, have more repeatable shape features than the replicates of the same cell line on GDA substrates, Appendix IV: Figure S5.3B. In addition, the shape differences between replicates of each cell line depends on the shape feature under comparison. For example, the Roughness parameter within replicates of AKT<sup>myr</sup> cell line on Fibronectin substrate is a repeatable parameter,  $p\text{-value} < 0.01$ , while Gray Scale FD parameters is not,  $p\text{-value} > 0.05$ .

#### 5.3.1.2 Several shape measures serve as good shape parameters for distinguishing between cell lines across the most comparisons.

We first asked whether some shape measures were good measures for distinguishing between different cell lines globally, i.e. for most cases. The condition used for a good measure is as following: a good shape feature is the one which p-value between technical replicates of the same cell line is significantly higher than p-value of the comparisons between different cell lines. Figure 5-1 shows all the comparisons that we carried out to identify optimal

shape parameters, organized by cell lines, and classes of shape parameters. It shows that some shape measures from each category are optimal in all the comparisons, though the specific number is cell-line and comparison specific. Since there is no way of specifying beforehand which shape categories will prove to be most useful for a cell type, this result underscores the importance of using a large number of shape categories in the analysis.

Figure 5-1 also shows that basis function methods, i.e. Fourier and Zernike expansions, perform well in distinguishing between samples. Both types of expansions pass the similarity test quite well for retina cell lines (Figure 5-1 B), and display differences in quite a few coefficients for the dissimilarity test for retina and osteosarcoma cells. Overall, waviness parameters do not meet the criteria for good measure in breast cell lines since p-values within replicates are in the same range as p-values between different cell lines. However, Zernike moments perform excellently in both similarity and dissimilarity test for breast cells.

It is worth mentioning that Waviness and Zernike measures in some cases complement each other i.e. where Waviness measures are not distinguishing parameters, Zernike moments can be used as good distinguishing parameters or vice versa. As an example, low metastatic and high metastatic line of K paired line, K12 and K7M2, are distinguishable with Waviness measures. However, they are not distinguishable with Zernike moments, Figure 5-1 C. On the other hand, M (MG63 and MG63.2) and S (SAOS2 and LM7) osteosarcoma cell lines are very similar in Waviness measures but distinguishable with Zernike moments, Figure 5-1 C. This suggests that these two types of shape measures capture different aspects of cell shape.

Cell, nuclei, and convex hull geometric parameters are also excellent in both similarity and dissimilarity tests. These include measures connected with spread size, elongation and irregularity of the boundary of a cell. However, the fine features of the irregularity of the boundary of the cell that are measured in the roughness parameters show strong cell line specificity. In particular, breast tissue cells (normal or cancerous) require a more selective group of parameters for distinguishing them, while osteosarcoma cells (Figure 5-1C) appear to be more easily distinguishable.

Each comparison between cells picks up specific shape features that are important for that comparison and not necessarily good for many others. Overall, gray scale measures are the best parameters which meet both similarity and dissimilarity criteria for almost all the cases.

### **5.3.1.3 Fibronectin substrates alter the shape and size of cells and their morph-space.**

To test whether cell shape on plain glass substrates differs from cell shape on fibronectin substrates, we compared the cell shapes of a subset of cell lines between these substrates. As shown in Appendix IV: Figure S5.3a-e, in general replicates of normal retina, RAS<sup>v12</sup>, and MEK<sup>DD</sup> cell lines cultured on GDA surfaces share less similarities with each other than their replicates on fibronectin coated surfaces. So, their morph-space on GDA is bigger than FN substrates. For AKT<sup>myr</sup> cell line opposite trend was observed.

### **5.3.2 Analyzing shape changes between cell lines within each cell type**

#### **5.3.2.1 Low metastatic and high metastatic breast cancer cells share similar shape features with normal breast cells but not with each other.**

According to Figure 5-2, distribution of high metastatic breast cell line is always apart from the distribution of the low metastatic cell line for all the shape categories. However, as demonstrated in this Figure, the parameters of the normal breast cell line overlaps with both the low metastatic cell line and the high metastatic cell line, depending on the shape category. As an example, normal cells overlap with high metastatic lines in Gray scale FD, Cell geometric, Hull geometric, ZM, and Roughness measures. Moreover, normal cells also overlap with low metastatic cells in all Textural measures, Spreading measures, and Roughness parameters. Therefore, cancer progression leads to the changes in some aspects of shape but not all. This figure suggests that it should be possible to distinguish between the low metastatic and high metastatic breast cancer cell lines by shape alone. Our neural network results also show that the classification accuracy for these two cell lines leads to 98% accuracy of prediction (Table 5-2)

#### **5.3.2.2 Transformation of normal retina cells to cancerous lines leads to significant shape changes, but with different characteristics on GDA or FN substrates.**

Figure 5-3 demonstrates shape distribution of retinal cells on GDA substrates. On GDA substrate normal cells are apart from cancer cells for all the shape categories other than ZM and Hull geometric measures. This implies that cancerous transformation leads to significant shape changes between cancer cells and normal retina cells regardless of their type on ordinary glass substrates. However, there is a major overlap between distribution of various cancer cells on GDA substrate in all shape categories other than cell and nuclei measures. Therefore, GDA substrate divides the retinal cells into cancerous and non-cancerous cells. On the other hand on FN substrates, Figure 5-4, in general, different cancer cells are more separated from each other in comparison with when they are on GDA substrates. Hence, in general FN is a better substrate in terms of separating retinal cancer cell lines from each other as

well as unifying the replicates of each cell line, and GDA is a better substrate in terms of separating retinal cancer lines from normal line. As shown in Table 5-2, our neural network algorithm can distinguish between normal and each cancer cell line with excellent accuracy ( $>0.98\%$ ) for both substrates.

### **5.3.2.3 Osteosarcoma cell lines show partially overlapping shape features but are easily distinguished on average.**

As discussed in Chapter 2 and 3 using a small number of shape features, 3 out of 4 paired cell lines share similar patterns in their shape changes, which we described as type 1 cell lines. However, the fourth paired line, MG63 and MG63.2, which we called type 2 line, showed exactly opposite trend. In Figure 5-5, low metastatic cells of all type 1 paired cell lines are combined and plotted against their high metastatic cell lines. As shown in this figure, the high metastatic cells are more homogenous in cell shape than the low metastatic line and the high metastatic and low metastatic cell lines are separated from each other. In order to understand the primary contributor to the separation, in Figure 5-6 the PC which has the best separation between all pairs is selected, and its scatter plot is graphed. As shown in this figure, the D paired line is more homogenous than other paired lines for all the shape categories especially in gray scale measure. All the type 1 paired lines show similarities in shape with significant overlap, but the type 2 cell line is separated from the type 1 cell lines in all the figures. In this figure, plots that represent distributions of spreading measures for these cell lines are good examples of how combinations of the right measures can lead to good classification of the cells under study. As shown in Cell geometric and nuclei geometric PPC plot, if each data point is projected into x axis (cell geometric-PC3), four paired lines will overlap significantly. The same is true when they are projected into y-axis (nuclei geometric\_PC1). However, the combination of cell geometric and nuclei geometric leads to an improved classification of all four paired lines. As tabulated in Table 5-2, our neural network algorithm was able to predict the class of the cells in each paired cell lines with high accuracy ( $>96\%$ ).

### **5.3.3 Cancer progression is accompanied with similar shape changes across most cancer types.**

In our previous study based on Zernike moments of cell shape [14], we showed that changes in the first principal component (PC1) of Zernike moments for type 1 paired lines undergoing metastasis were similar and opposite to type 2 paired line, MG63 and MG63.2. We would like to test whether the cell lines of other types which become cancerous or undergo metastasis experience similar shape changes using our expanded and optimized list of features. For this purpose, we calculated the changes in the mean of first principal component for each pair of lines undergoing metastasis and repeated this for all shape categories. In each pair there was one normal/low metastatic line

and one cancerous/high metastatic line. The list of the comparisons are tabulated in Appendix IV: Table S5-1. If the changes in PC1 are significant ( $p\text{-value} < 0.001$ ), it is noted whether it increases or decreases. In PCA, weight of different feature in each principal component can be negative or positive. Therefore, an increase in the principal component does not imply an increase in a feature. However, in each plot changes in PC1 of each pair can be compared with other pairs, i.e. if PC1 increases for both it means their shape changes similarly. If they are in opposite direction, it means their shape changes are not similar. Here, the M lines were used as our reference. We are interested to know the direction of changes in PC1 of the other pairs when the changes in PC1 for M cell line is statistically different ( $p\text{-value} < 0.001$ ). The results for this analysis is listed in Appendix IV: Table S5-1. The changes in PC1 for M cell line was not significant for Gray scale, band based, and nuclei measures. Of all the 6 remaining shape categories other than roughness, significant shape changes for all the comparisons were exactly opposite of shape changes seen for M cell line. In roughness measures, the shape changes from normal retina cell line to AKT retina cancer line was the only comparison which followed similar shape changes as the M line. These mean that the shape changes accompanied with cancer progression in most of these cell lines follows similar pattern regardless of type of the measures used to quantify the shape. Note that we have not seen other shape changes that appear to resemble type 2 shape changes, but there are reports in the literature that appear to descriptively match the shape behavior we see in the M-lines[20] .

We asked whether shape differences between cell lines are sufficient so that the neural network classifier can be used to distinguish between low metastatic and high metastatic cell lines while all cell types are combined. For this purpose, all the shape features from 9 shape categories were combined and used as the input of the classifier. Results for classifications are tabulated in Table 5-2, and they show that the neural network classifier can distinguish between low metastatic and high metastatic or normal and cancer cells with a good accuracy. As discussed earlier, when the classes to be classified are just from one cell type, excellent classification accuracy is observed (Accuracy > 96%). However, the accuracy of prediction decreases to 74% when the cells to be classified belong to different cell types. This number goes up to 82% when only type 1 cells are used and 87% when only the normal and cancerous cells are compared. Comparing only the cancer cells and those that undergo metastasis, results in an accuracy of prediction of 80%.

## 5.4 Conclusion

In previous publications we had shown that cell shape could be used to distinguish osteosarcoma cells of low metastatic capacity from those with high metastatic capacity. However, these results were based on a small number

of geometric features and Zernike moments of shape of the cells. We wanted to know whether our results generalize to other cancers, and whether other features of shape, especially textural features that describe the distribution of actin, can also help discriminate between cells of different metastatic potential. We were also interested in understanding the cell shape changes during cancerous transformation. An *in vitro* system of carcinogenesis based on a specific cocktail of genetic perturbations developed by the Deluca laboratory in CSU provided the opportunity to test the effect of specific oncogenic insults on cell shape. Thus in this study we used a larger array of shape features, incorporating several new features that we developed that are discussed in greater detail in Chapter 2, and we used these measures to analyze changes in cell shape in a larger array of cell lines. Our data strongly suggests that cell shape can be used as a marker of invasive potential for many cancers.

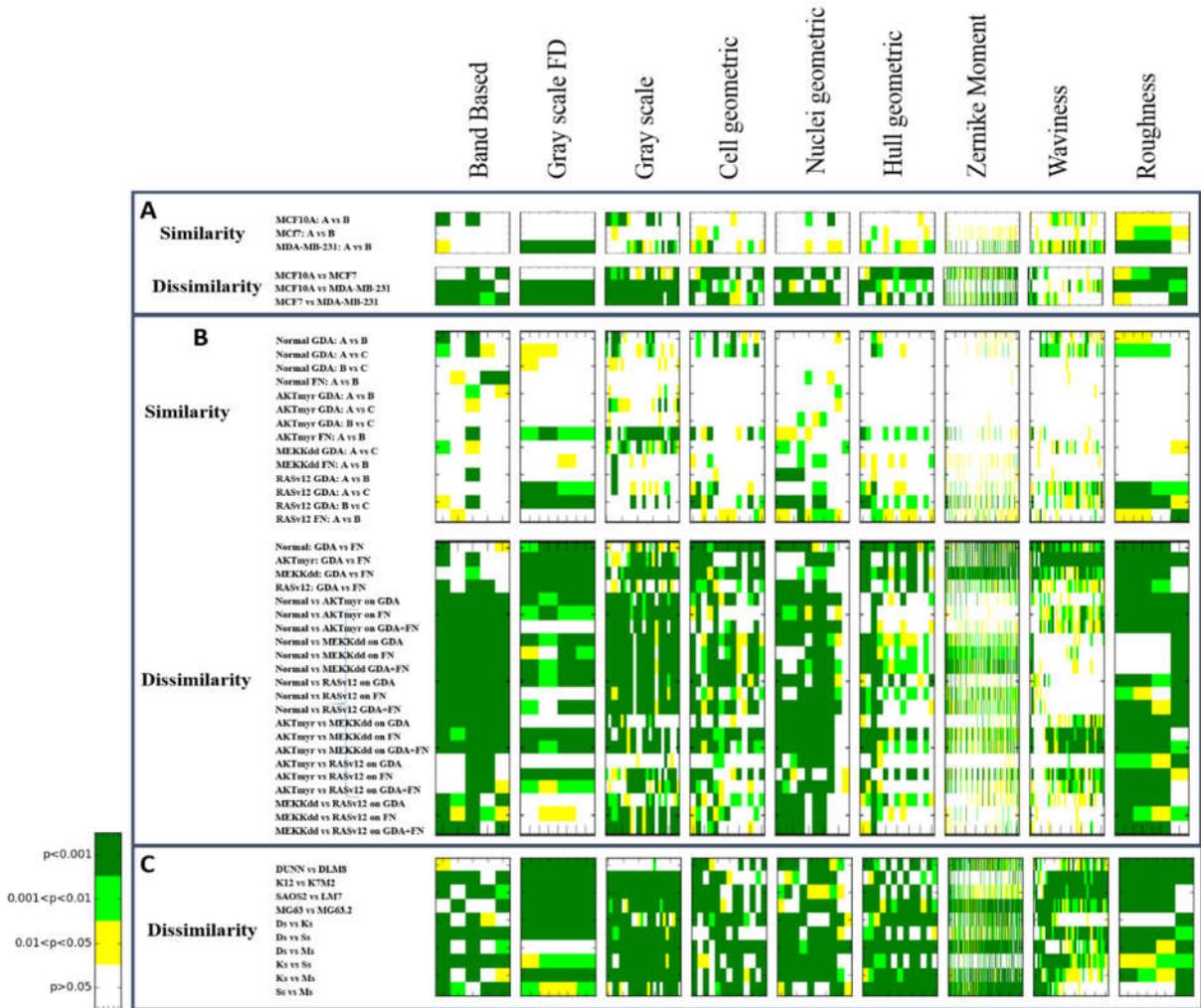


Figure 5-1 T-tests results. Heat map plot of p-values for similarity tests, when replicates of the same cell line are compared, and dissimilarity tests, when replicates of cell lines are combined and compared against other lines. Each column in this plot represent one shape measure and each shape category is plotted separately. The lighter the color is the more similar the samples are for that shape measure. As shown in this figure, similarity tests take brighter color than dissimilarity tests.

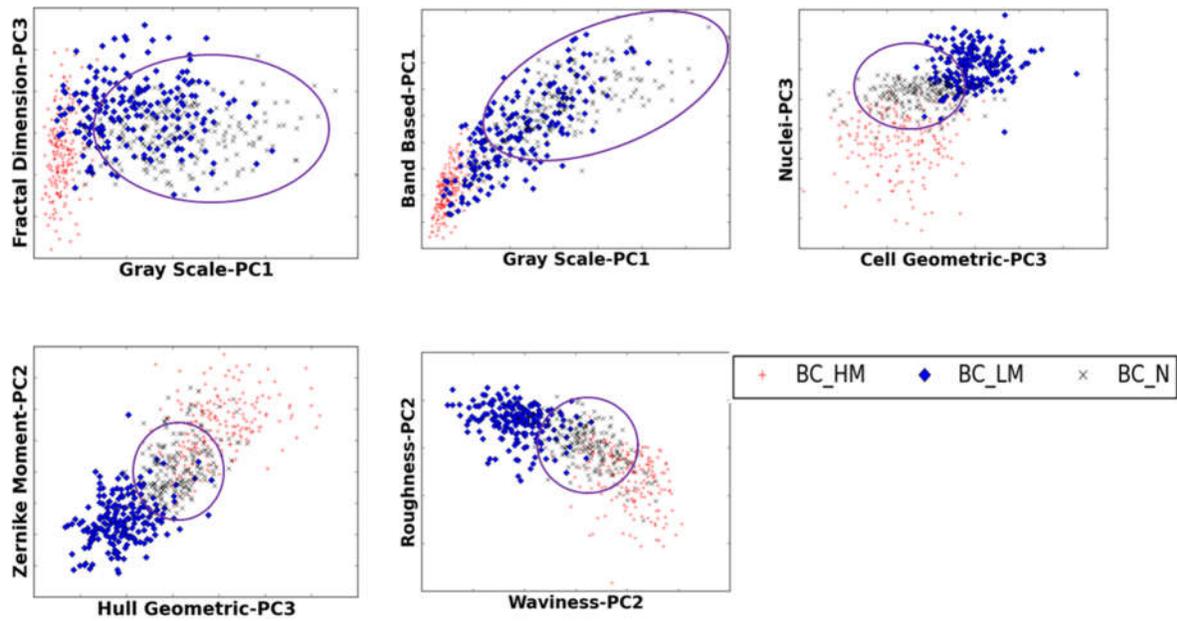


Figure 5-2 Scatter plot in PPC for all the breast cell lines. Depending on the shape category normal cells sometimes overlap with low metastatic cells and sometimes with high metastatic cells. The two hardly overlap with each other. The purple ellipse is drawn to enclose normal cells to guide the eye.

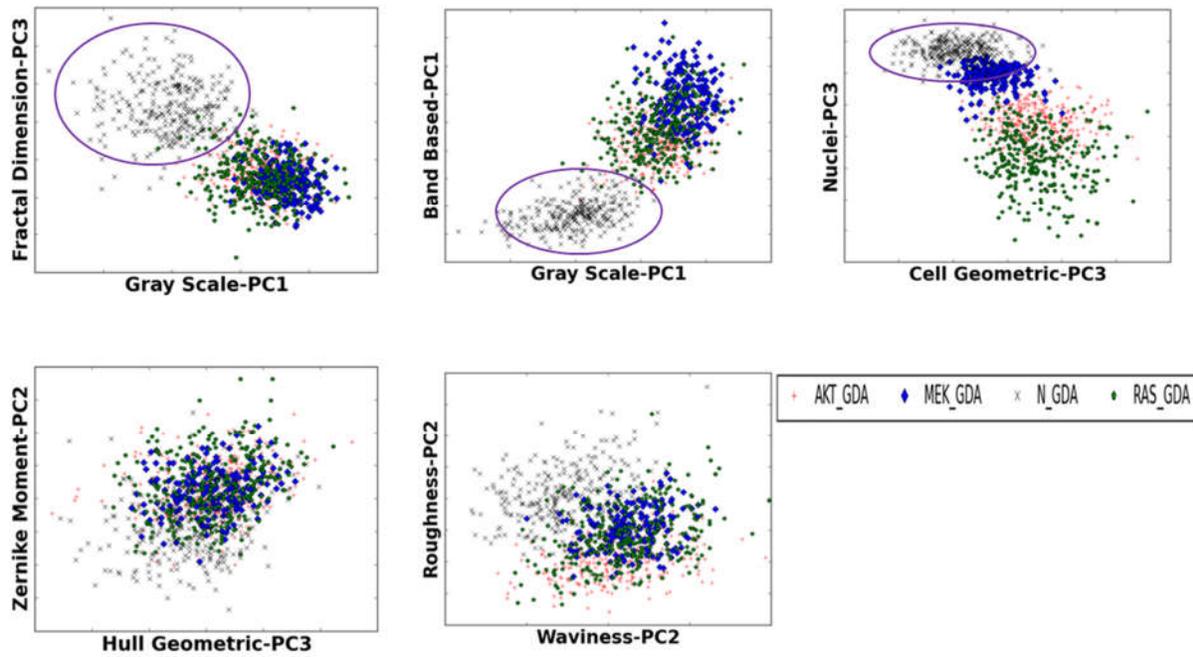


Figure 5-3 Scatter plot in PPC for all Retinal cells on GDA substrate. Cancerous cell lines overlap with each other on GDA substrate but they are far away from normal cells. The purple ellipse is drawn to enclose normal cells to guide the eye.

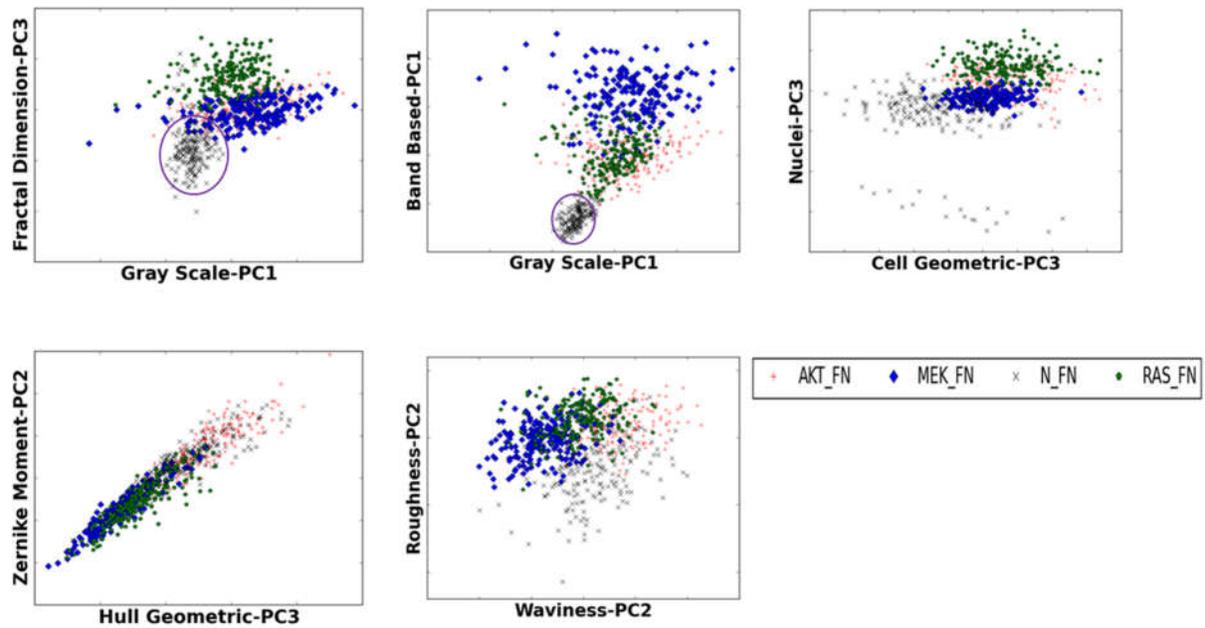


Figure 5-4 Scatter plot PPC for all Retinal cells on FN substrate. On FN substrate cancerous cells overlap less with each other. Normal cells still can be distinguished from cancerous cells. The purple ellipse is drawn to enclose normal cells to guide the eye.

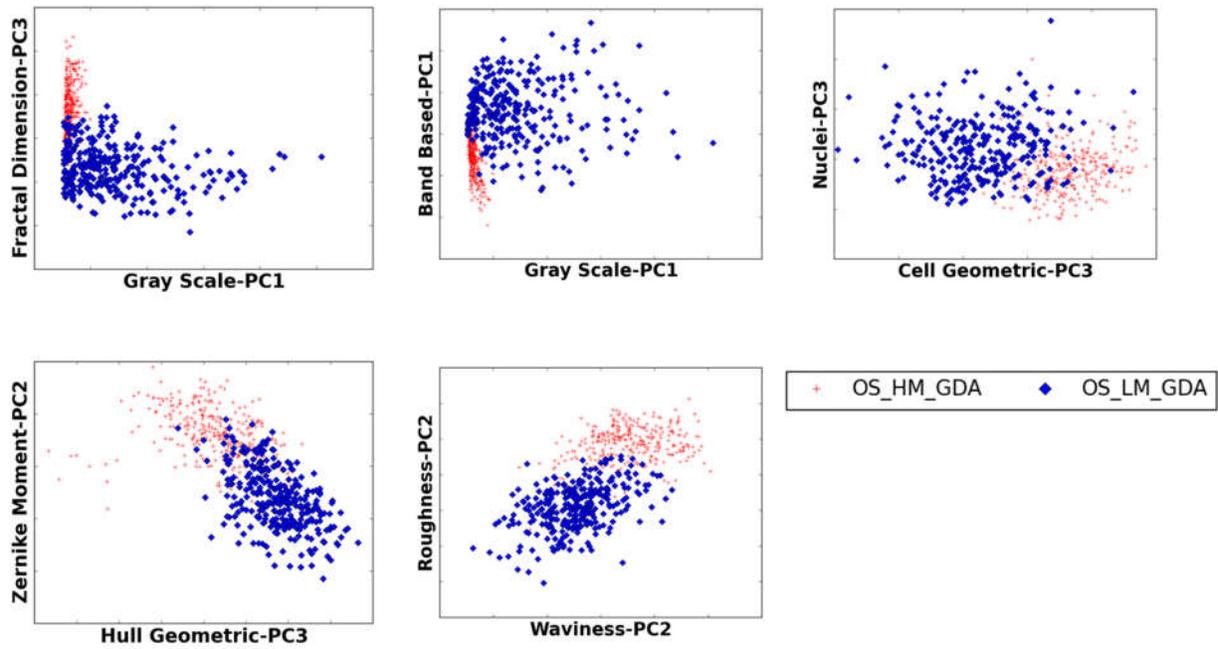


Figure 5-5 Scatter plot in PPC for type 1 osteosarcoma cells. Low metastatic cells and high metastatic cells cluster separately. High metastatic cell lines also appear to have more uniform shape distribution than low metastatic cell lines.

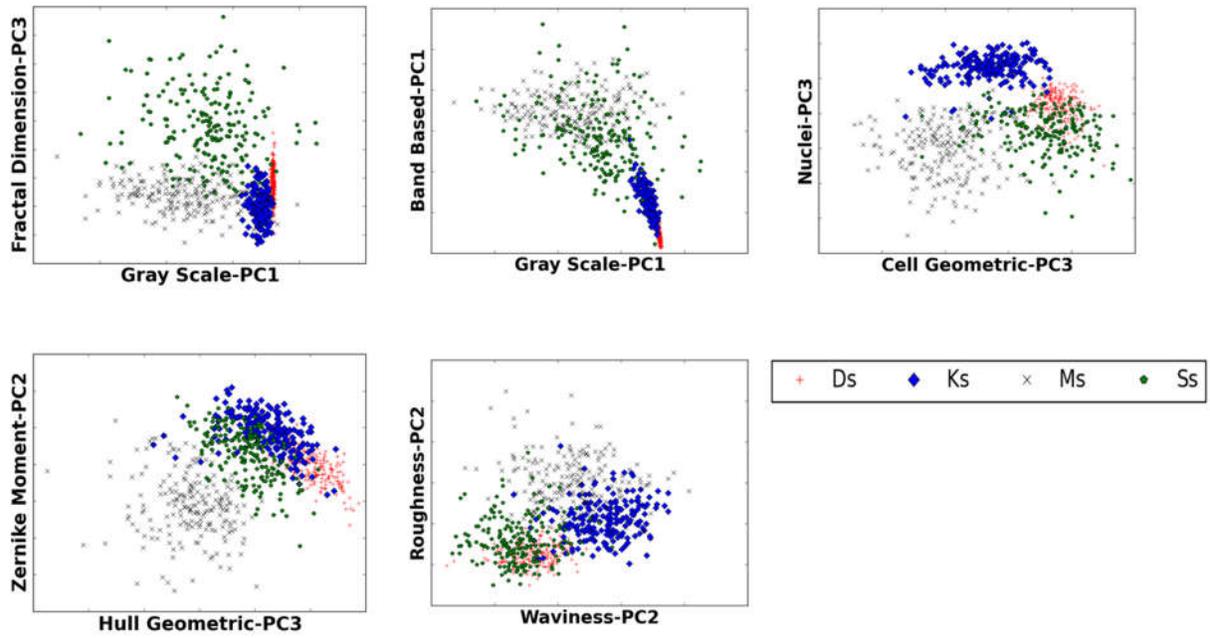


Figure 5-6 Scatter plot in PPC for Osteosarcomacells. Shape parameters of both the high and low metastatic lines in each paired line are plotted together. It can be seen that paired lines cluster together, and that spreading measures can separate paired lines with good accuracy.

Table 5-1 Cell lines and acronyms used in chapter 5.

Cell Types		Species	Cell lines	Acronyms	Substrates (Replicates)	Substrates (Replicates)
Breast Cells		Human	MCF10A	BC_N*	FN** (2)	
		Human	MCF7	BC_LM***	FN (2)	
		Human	MDA-MB-231	BC_HM†	FN (2)	
Osteosarcoma Cancer cells	Ds	Mouse	DUNN	OS_DL	GDA††(3)	
			DLM8	OS_DH	GDA(3)	
	Ks	Mouse	K12	OS_KL	GDA(3)	
			K7M2	OS_KH	GDA(3)	
	Ms	Human	MG63	OS_ML	GDA(3)	
			MG63.2	OS_MH	GDA(3)	
	Ss	Human	SAOS2	OS_SL	GDA(3)	
			LM7	OS_SH	GDA(3)	
Retina Cells			APRE19	RC_N	FN (2)	GDA*(3)
			AKTmyr	RC_AKT	FN (2)	GDA(3)
			MEKDD	RC_MEK	FN (2)	GDA(2)
			Rasv12	RC_RAS	FN (2)	GDA(3)

\*N: Normal cells ; \*\*FN: Fibronectin coated glass ;

\*\*\*LM: Low Metastatic line; †HM: High Metastatic line;

††GDA: Glass Detergent washed Air dried.

Table 5-2. The accuracy of Machine learning classification. Type A cells are the normal cells which become cancerous and type B cells are low metastatic cells which become high metastatic.

<b>Comparison</b>	<b>Prediction</b>
Breast: Low metastatic vs high metastatic	0.98125
Retina: Normal versus AKTmyr on GDA substrate	0.9975
Retina: Normal versus MEKKDD on GDA substrate	0.9885
Retina: Normal versus RASv12 on GDA substrate	0.997917
Osteosarcoma: DUNN vs DLM8	0.96
Osteosarcoma: K12 vs K7M2	1
Osteosarcoma: SAOS2 vs LM7	0.995
Osteosarcoma: MG63 vs MG63.2	0.99125
Retina: Normal versus AKTmyr on FN substrate	0.984375
Retina: Normal versus MEKKDD on FN substrate	0.99375
Retina: Normal versus RASv12 on FN substrate	0.98375
All cell types on GDA substrate	0.749348
All type 1 cells on GDA	0.8275
All type A on GDA	0.875833
All type B on GDA	0.873125

## Bibliography

- [1] “Leading Cause of death.” [Online]. Available: <http://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm>.
- [2] G. P. Gupta and J. Massagué, “Cancer Metastasis: Building a Framework,” *Cell*, vol. 127, no. 4, pp. 679–695, Nov. 2006.
- [3] R. Siegel, J. Ma, Z. Zou, and A. Jemal, “Cancer statistics, 2014,” *CA. Cancer J. Clin.*, vol. 64, no. 1, pp. 9–29, Jan. 2014.
- [4] Z. Yin, H. Sailem, J. Sero, R. Ardy, S. T. C. Wong, and C. Bakal, “How cells explore shape space: A quantitative statistical perspective of cellular morphogenesis,” *BioEssays*, vol. 36, no. 12, pp. 1195–1203, Dec. 2014.
- [5] J. Y. Lim and H. J. Donahue, “Cell Sensing and Response to Micro- and Nanostructured Surfaces Produced by Chemical and Topographic Patterning,” *Tissue Eng.*, vol. 13, no. 8, pp. 1879–1891, Aug. 2007.
- [6] K. A. Kilian, B. Bugarija, B. T. Lahn, and M. Mrksich, “Geometric cues for directing the differentiation of mesenchymal stem cells,” *Proc. Natl. Acad. Sci.*, vol. 107, no. 11, pp. 4872–4877, 2010.
- [7] R. McBeath, D. M. Pirone, C. M. Nelson, K. Bhadriraju, and C. S. Chen, “Cell shape, cytoskeletal tension, and RhoA regulate stem cell lineage commitment,” *Dev. Cell*, vol. 6, pp. 483–495, 2004.
- [8] M. Thery, “Micropatterning as a tool to decipher cell morphogenesis and functions,” *J. Cell Sci.*, vol. 123, no. 24, pp. 4201–4213, Dec. 2010.
- [9] H. Z. Sailem and C. Bakal, “Identification of clinically predictive metagenes that encode components of a network coupling cell shape to transcription by image-omics,” *Genome Res.*, vol. 27, no. 2, pp. 196–207, Feb. 2017.
- [10] J. E. Sero, H. Z. Sailem, R. C. Ardy, H. Almuttaqi, T. Zhang, and C. Bakal, “Cell shape and the microenvironment regulate nuclear translocation of NF- $\kappa$ B in breast epithelial and tumor cells,” *Mol. Syst. Biol.*, vol. 11, no. 3, p. 790, Mar. 2015.
- [11] J. E. Sero and C. Bakal, “Multiparametric Analysis of Cell Shape Demonstrates that  $\beta$ -PIX Directly Couples YAP Activation to Extracellular Matrix Adhesion,” *Cell Syst.*, vol. 4, no. 1, p. 84–96.e6, Jan. 2017.

- [12] A. Pasqualato *et al.*, “Quantitative shape analysis of chemoresistant colon cancer cells: Correlation between morphotype and phenotype,” *Exp. Cell Res.*, vol. 318, no. 7, pp. 835–846, 2012.
- [13] S. M. Lyons *et al.*, “Changes in cell shape are correlated with metastatic potential in murine and human osteosarcomas,” *Biol. Open*, 2016.
- [14] E. Alizadeh, S. M. Lyons, J. M. Castle, and A. Prasad, “Measuring systematic changes in invasive cancer cell shape using Zernike moments,” *Integr. Biol.*, vol. 8, no. 11, pp. 1183–1193, 2016.
- [15] E. Alizadeh, M. Lyons, M. Castle, and A. Prasad, “Integrative Biology Measuring systematic changes in invasive cancer cell shape using Zernike moments †,” *Integr. Biol.*, vol. 8, pp. 1183–1193, 2016.
- [16] F. D’Anselmi *et al.*, “Metabolism and cell shape in cancer: a fractal analysis,” *Int. J. Biochem. Cell Biol.*, vol. 43, no. 7, pp. 1052–1058, 2011.
- [17] P.-H. Wu *et al.*, “Evolution of cellular morpho-phenotypes in cancer metastasis,” *Sci. Rep.*, vol. 5, no. 1, p. 18437, 2016.
- [18] S. M. Lyons *et al.*, “Changes in cell shape are correlated with metastatic potential in murine and human osteosarcomas,” *Biol. Open*, vol. 5, no. 3, pp. 289–299, 2016.
- [19] L. Kametsky *et al.*, “Improved structure, function and compatibility for cellprofiler: Modular high-throughput image analysis software,” *Bioinformatics*, vol. 27, no. 8, pp. 1179–1180, 2011.
- [20] G. Cantelli *et al.*, “TGF- $\beta$ -Induced Transcription Sustains Amoeboid Melanoma Migration and Dissemination,” *Curr. Biol.*, vol. 25, no. 22, pp. 2899–2914, 2015.

## Chapter 6: Conclusion

The emergence of super resolution microscopy and high throughput imaging techniques has provided the tools to collect a large number of images of cells. Combination of image processing methods with imaging techniques provides clear images of the structure and boundary of the cells. Interpreting the images demand methods to numerically represent them. This dissertation provides experimentally validated methods to numerically represent the shape and structure of cells and uses the developed methods to predict the invasiveness of cancer cells, demonstrating that quantitative analysis of cell shape could have useful applications in cancer diagnostics.

Three major shape measures of textural, spreading and irregularity have been introduced in Chapter 2. Then, these measures have been sub-categorized into nine different shape categories. Each shape category captures different aspect of shape of a cell. Using Pearson correlation coefficient, it was shown that the shape categories are generally weakly correlated with each other and therefore, they do not contain redundant information. Moreover, it was shown through-out the experimental results that the best shape category for each problem is different and therefore there is a need to use all shape categories. Different methods such as T test, PCA and PPA were utilized to find the best shape category for each problem separately.

In this study these measures were used to quantify actin structure and shape of the cells. Even though other researchers have only used the qualitative changes in actin structure, this work has shown that the textural measure can also quantify subtle changes in the actin structure that are not easily identifiable by eye. It should be noted that the textural measures can also be used for quantification of any other structures such as membrane and microtubules of a cell. They also can be used to quantify the image of slice of a tumor which could be helpful tool in cancer diagnostic rather than having pathologist to examine them by eye which requires a lot of training and does not lead to consistent diagnostics.

In Chapter 2, it was shown that when the cells are perturbed with cytoskeletal drugs, their actin structure changes significantly. Then it was confirmed that the cell spreading and irregularity measures also changes for some drugs. Therefore, actin organization is correlated with irregularity and spreading measures; however this correlation is not linear. Some deep learning models could be used to find this complex nonlinear relation between actin organization and cell spreading and irregularity. Textural measures introduced in Chapter 2 can also be used in image

of slices of tumors rather than single cells. The advantage of using these measures is that the results are more consistent which has less human error and inconsistency and it is not time consuming and does not require expensive training.

In Chapter 3, some of the spreading measures introduced in Chapter 2 were used to quantify the changes in shape of four paired osteosarcoma cell lines, which undergo metastasis. These paired lines fall into two groups. Three out of four paired lines, which we classified as type 1, become smaller, less elongated and more irregular when they become metastatic and the last paired line, classified as type 2, shows the opposite changes in shape. The same trend was also observed in Chapter 4, where Zernike moments were used and projected into PCA space. It is worth mentioning that even though Zernike moments are categorized as spreading measures, their quantification method is dramatically different with other spreading measures used in Chapter 3; yet they lead to consistent results.

Using supervised classification method, neural network classifier, good accuracy of prediction between low metastatic and high metastatic lines of each osteosarcoma paired line was achieved. It should be noted that in supervised learning methods some known samples should be provided to the machine to be trained on. This becomes challenging when predicting class of unknown samples. Similar changes between low metastatic and high metastatic cells in type 1 osteosarcoma cell lines showed that the paired lines undergo similar shape changes even though their shape are not similar. Interestingly, with training the neural network classifier on one pair of type 1 cells, it was able to predict class of low metastatic and high metastatic cells of other pairs from type 1. This shows some promising results and raises the hope to be able to characterize unknown cancer cell samples using cell shape information of known samples.

In Chapter 5, cells from retinal epithelial cells, which were transfected with three different oncogenes, along with normal and cancerous breast cell lines, and paired osteosarcoma cancer cell lines were used to study changes in cell shape with cancer progression within this broader range of cell lines. Interestingly, all the shape changes seen in the cell lines used fell into the type 1 category except for one pair of osteosarcoma lines. However, their similarities were not enough to train the neural network on a pair from one cell type, as an example breast line, and predict class of other cell type, as an example osteosarcoma pair. Deep learning methods which uses more layers and is more complex might be able to better predict the class of cells across different types of cancer. This still needs a lot more experiment and validations, but there might be some signature of primary site in the cell shape as well. In other words, cancer cells which start from the same primary tissue might undergo similar shape changes but different from the cells that are from a different tissue.

For this work, isolated cells were used, but this also could be used for the cells that are completely surrounded by other cells. We believe that the shape of isolated cells could be different than cells in clusters. It is recommended that neighboring factor, perimeter of a cell which is in contact with other cells over its total perimeter, be used as another feature of a cell.

Cancer is the second leading cause of deaths in the United States. Ninety percent of the deaths among cancer patients are from cancer metastasis. Patients who have been diagnosed in early stage of their cancer have higher 5-year survival rate. Therefore, improving diagnostic methods is a vital need. In cancer diagnostics a complex combination of clinical and histopathological data are used. This includes qualitative examination of tissue. This work provides the evidence that cell shape on two-dimensional substrates, can be potentially used as additional information to better diagnose cancer; along with providing the tools to quantitatively represent the cell shape.

Appendix I: supplementary information for chapter 2

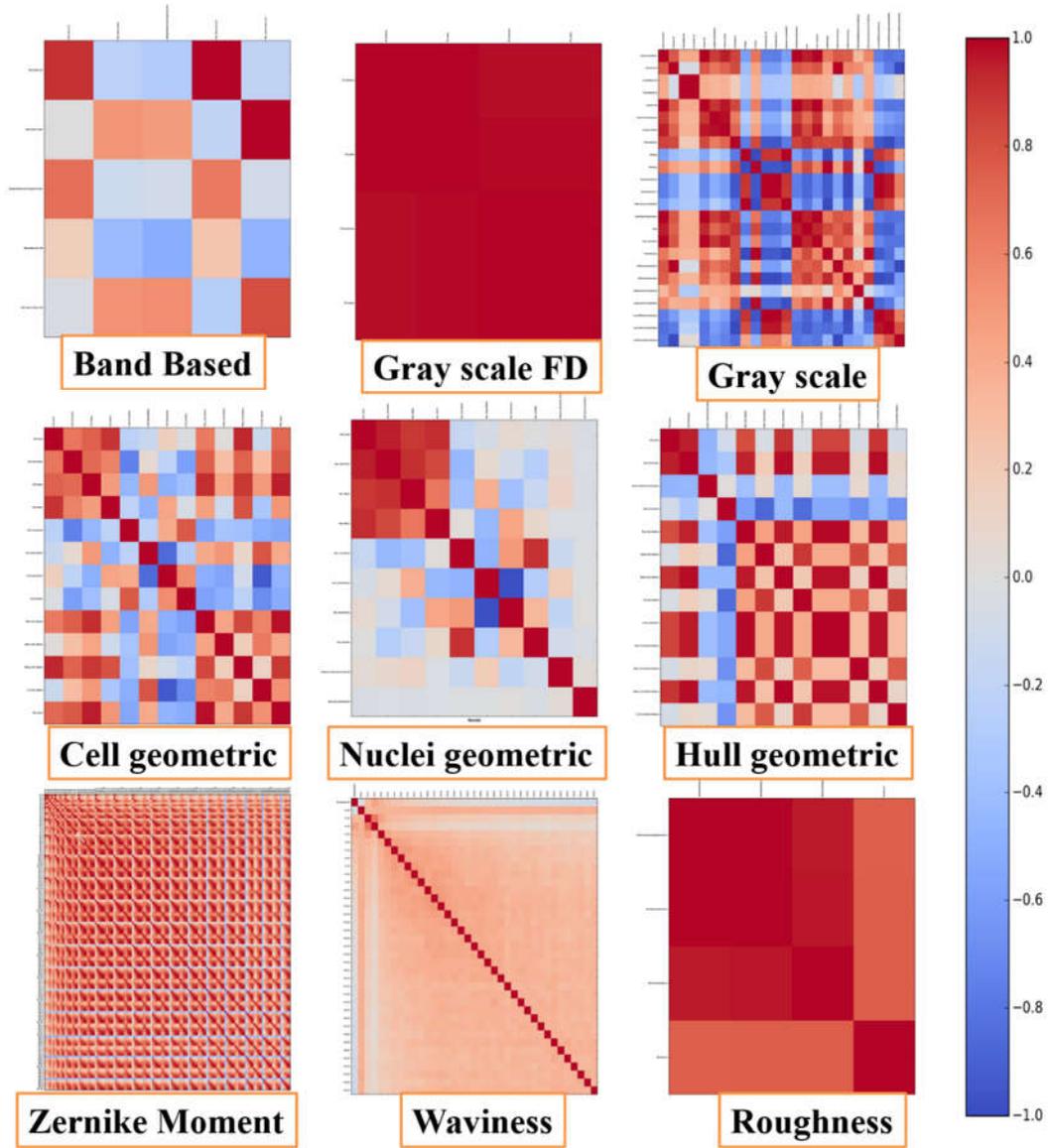


Figure S2.1 Pearson correlation coefficient within each shape category. Within each shape category features are highly correlated other than band based, waviness (Fourier), and some of the nuclei measures.

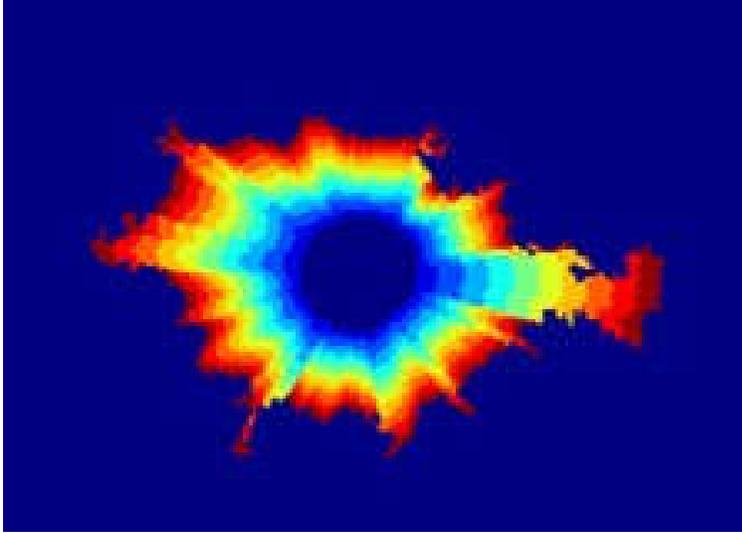


Figure S2.2 Band based division of textural image of a cell. Cell is divided into 10 equally spaced radial bands. All the measures calculated based on this representation is listed in Table S2-1.

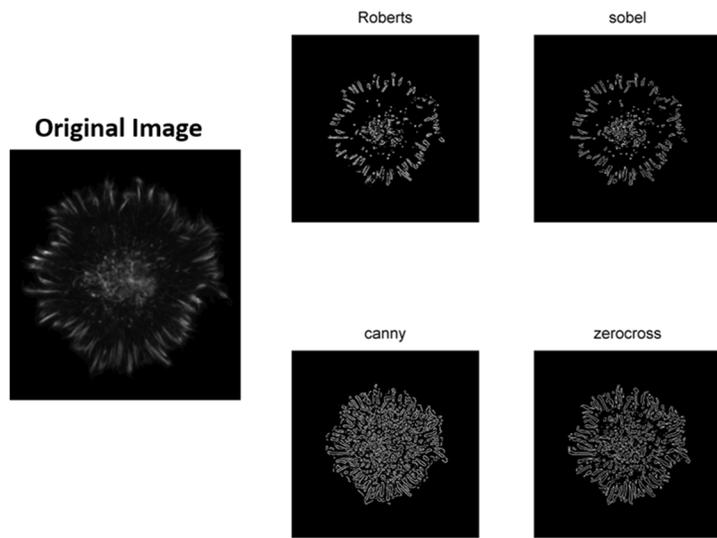


Figure S2.3. Matlab's Edging methods used to binerize gray scale image of actin. Four different methods of Roberts, Sobel, Canny, and zerocross are used in this dissertation.

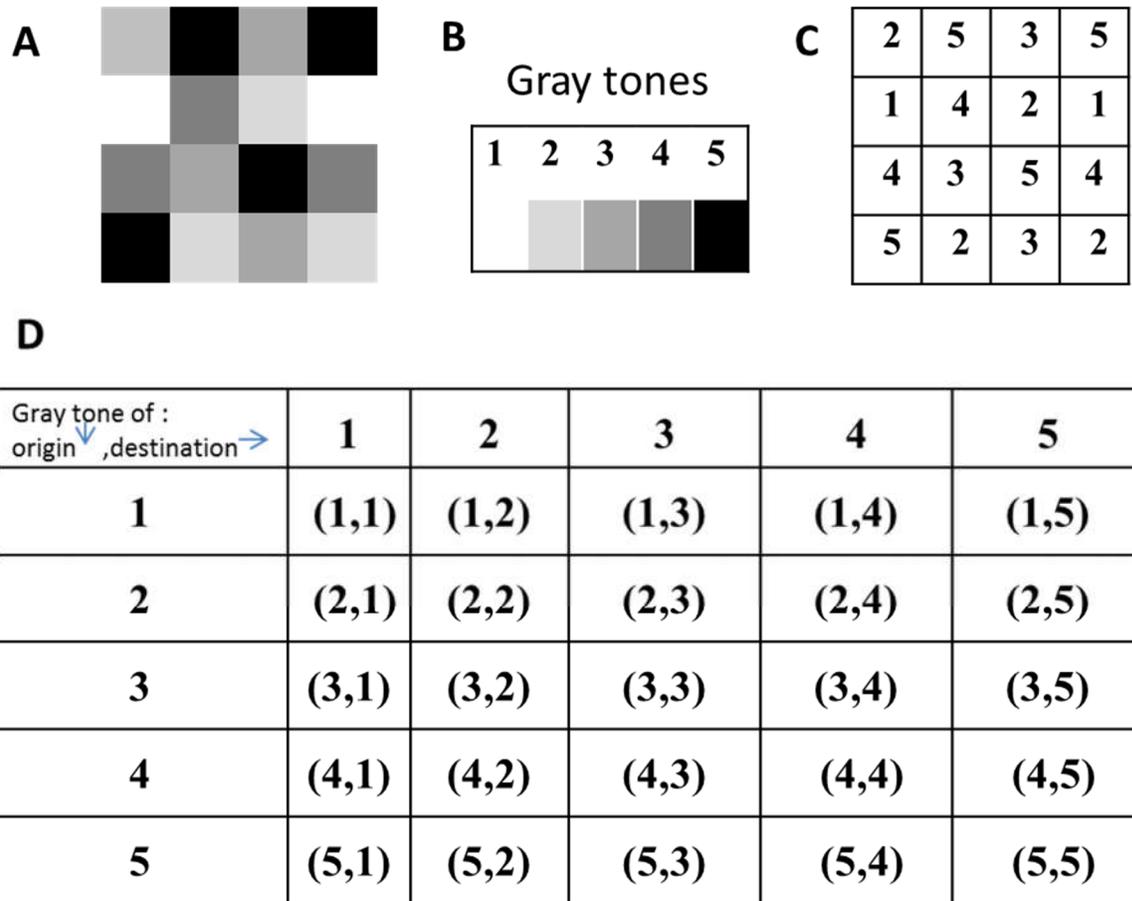


Figure S2.4 Calculation of GCLM matrix. A) A cartoon of a 4pixels x 4pixels image. B) The image has 5 gray tones as shown. C) Numerical representation of image based on gray tone is shown. D) GCLM Matrix is 5 x 5 element and its  $ij^{\text{th}}$  element is number of the times which pixel with intensity of  $j$  is in distance  $\mathbf{d}$  from the pixels with intensity of  $i$ .  $\mathbf{d}$  is an arbitrary vector but should be defined in advance. In this dissertation  $\mathbf{d}$  is defined to have length of one and it is defined in 4 directions of  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ . Gray scale measures are calculated based on four directions and their average is reported.

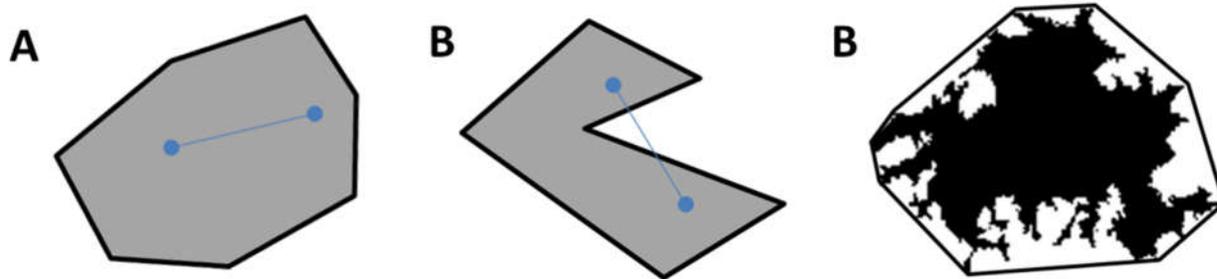


Figure S2.5 Convex Hull. A cartoon of a A) convex and B) non-convex polygon. A convex polygon is a polygon which any line connecting two points inside the polygon falls inside the polygon. C. Convex hull surrounding a cell.

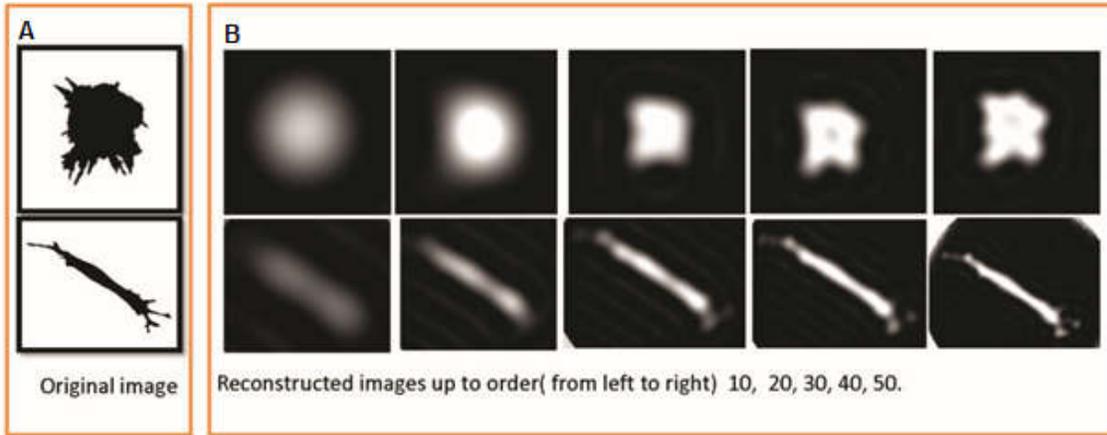


Figure S2.6 Reconstruction of cell images for two different shaped cells. A. Original images. B. Reconstructed images using Zernike moments up to order (from left to the right) 10, 20, 30, 40, 50.

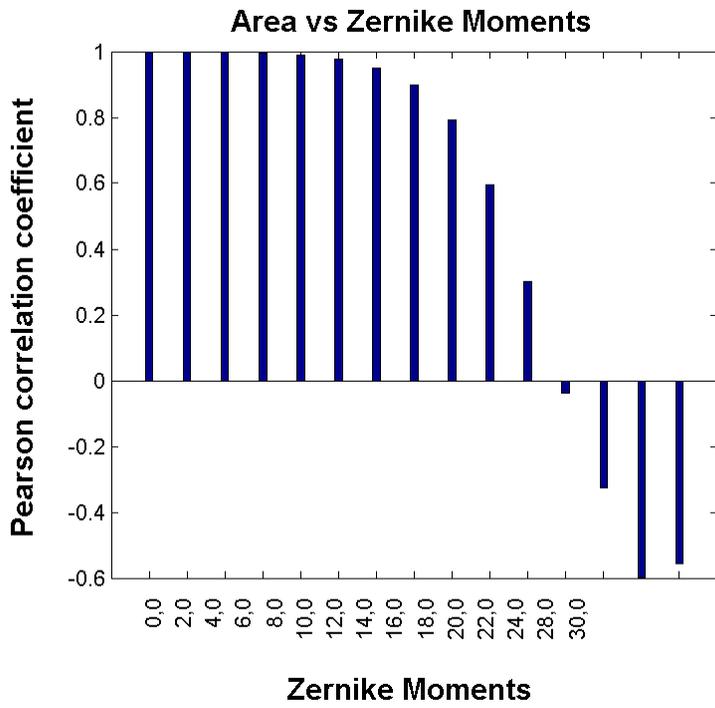


Figure S2.7 Pearson correlation coefficient between Zernike Moments with  $m=0$  and Cell Area. Correlation is maximum and positive with Zernike moment of  $0_0$ , it decreases with increasing the order and it is almost 0 for Zernike moment of  $22_0$  and it is negative for higher orders and increases in magnitude.

Table S2-1: Band based measurements used in the dissertation[39].

Parameter		Formula
Max band intensity	Intensity of the band with maximum average intensity	$\text{Max} (\{I_{B_1}, \dots, I_{B_{10}}\})$
Max band index	Location of the band with maximum average intensity	The n such that $I_n = \text{Max} (\{I_{B_1}, \dots, I_{B_{10}}\})$
Minimum band intensity	Intensity of the band with minimum average intensity	$\text{Min} (\{I_{B_1}, \dots, I_{B_{10}}\})$
Index of minimum band intensity	Location of the band with minimum average intensity	The m such that $I_m = \text{Min} (\{I_{B_1}, \dots, I_{B_{10}}\})$
Above average adjusted intensity	Adjusted, weighted sum of average actin intensities greater than the mean intensity.	$\frac{\sum_{k=1}^{10} s(I_{B_k}) I_{B_k}}{\text{Max} (\{I_{B_1}, \dots, I_{B_{10}}\}) \sum_{k=1}^{10} s(I_{B_k})}$ <p>Where <math>S(I_{B_k}) =</math></p> $\begin{cases} 1 & \text{if } I_{B_k} \geq \text{mean}(\{I_{B_1}, \dots, I_{B_{10}}\}) \\ 0 & \text{ot otherwise} \end{cases}$

Table S2-2 Gray scale measures used in this dissertation [19], [18], [20], [21] . See TableS2-3 for the parameters used in this table.

Parameter	Formula
Autocorrelation	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ij p(i, j)$
Contrast_M[	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g}  i - j ^2 p(i, j)$
Correlation_M	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j}$
Correlation_P	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (ij)p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
Cluster SD	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^2 p(i, j)$
Cluster Prominence	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^4 p(i, j)$
Cluster Shade	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^3 p(i, j)$
Dissimilarity	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \{ i - j  p(i, j)\}$
Energy	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j)^2$
Entropy	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \log(p(i, j))$
Homogeneity_M	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i, j)}{1 +  i - j }$

Homogeneity_P	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{1 + (i-j)^2}$
Maximum probability	$MAX_{i,j} p(i,j)$
Sum squares variance	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 p(i,j)$
Sum_average	$\sum_{i=2}^{2N_g} i p_{x+y}(i)$
Sum variance	$\sum_{i=2}^{2N_g} (i - \text{SumEntropy})^2 p_{x+y}(i)$
Sum Entropy	$\sum_{i=2}^{2N_g} p_{x+y}(i) \log(p_{x+y}(i))$
Difference Variance	$\sum_{i=0}^{N_g-1} i^2 p_{x-y}(i)$
Difference Entropy	$\sum_{i=0}^{N_g-1} p_{x-y}(i) \log(p_{x-y}(i))$
Information Measure of Correlation1	$\frac{\text{Entropy } H_{XY1}}{\max\{H_X, H_Y\}}$
Information measure Correlation 2	$\sqrt{(1 - e^{-2(H_{XY2} - H_{XY})})}$
Inverse Difference Normalized(INN) (Based on Matlab code)	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{1 + \frac{ i-j }{N_g}}$
Inverse Difference Moment Normalized Based on Matlab code	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{1 + \frac{(i-j)^2}{N_g^2}}$

Table S2-3. Parameters used in the definitions of gray scale measures in the Table S2-2.

Parameter	Formula
$P(i,j)$	$ij^{\text{th}}$ entry of co-occurrence probability matrix
$N_g$	Highest intensity of gray level
$\mu$	Mean value of $p(i,j)$
$p_x(i)$	$\sum_{j=1}^{N_g} p(i,j)$
$p_y(i)$	$\sum_{i=1}^{N_g} p(i,j)$
$\mu_x$	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} i p(i,j)$
$\mu_y$	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} j p(i,j)$
$\sigma_x^2$	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu_x)^2 p(i,j)$
$\sigma_y^2$	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu_y)^2 p(i,j)$
$p_{x+y}(k)$	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)$ <p>Where <math>i+j=k</math>, <math>k=2,\dots,2N_g</math></p>
$p_{x-y}(k)$	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)$ <p>Where <math>i-j=k</math>, <math>k=0,\dots,N_g-1</math></p>
HX	$\sum_{i=1}^{N_g} p_x(i) \log(p_x(i))$

HY	$\sum_{i=1}^{N_g} p_y(i) \log(p_y(i))$
HXY1	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \log(p_x(i)p_y(j))$
HXY2	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_x(i)p_y(j) \log(p_x(i)p_y(j))$

Table S2-4. Geometric parameters.

Parameter	Description	Unit
Cell area	Area of the cell	Pixel <sup>2</sup>
Cell perimeter	Perimeter of the cell	Pixel
Cell major axis	The major axis of an ellipse drawn around the cell	Pixel
Cell minor axis	The minor axis of an ellipse drawn around the cell	Pixel
Cell circularity (Smoothness)	$\frac{4\pi \text{ Area}}{\text{Perimeter}^2}$	Unit-less
Cell aspect Ratio	$\frac{\text{Cell Major axis}}{\text{Cell Minor axis}}$	Unit-less
Cell roundness	$\frac{4 \text{ Area}}{\pi \text{ Cell major axis}^2}$	Unit-less
Cell solidity	$\frac{\text{Convex hull area}}{\text{Cell area}}$	Unit-less
Max cell Radius	Maximum radius of the cell	Pixel
Cell radius ratio	$\frac{\text{Max radius of the cell}}{\text{Min radius of the cell}}$	Unit-less
Mean Cell Radius	The mean of all radii drawn from the hull's centroid to an exterior point	Pixel
CV Cell Radius	The relative variation of radii drawn from the cell's center to an exterior point. Given by the Standard of Deviation of all Radii divided by the mean of all radii	Unit-less
Max Span	The maximum distance from one point on the convex hull to another	Pixel

Table S2-5. Convex hull geometric parameters.

<b>Parameter</b>	<b>Description</b>	<b>Unit</b>
Hull area	Area of the convex hull	Pixel <sup>2</sup>
Hull perimeter	Perimeter of the convex hull.	Pixel
Ratio of perimeter of Hull to cell	Ratio of hull perimeter to cell perimeter.	Unit-less
Hull circularity	$\frac{4\pi \text{ Hull area}}{\text{Hull perimeter}^2}$	Unit-less
Max hull Radius	Maximum distance from centroid of Hull to an exterior point on the hull.	Pixel
Hull radius ratio	Maximum/minimum radius of hull.	Unit-less
Mean hull radius	The mean of all radii drawn from the hull's centroid to an exterior point.	Pixel
CV hull radius	$\frac{STD \text{ hull radius}}{Mean \text{ hull radius}}$	Unit-less
Bounding circle diameter	The diameter of the bounding circle drawn around the cell.	Pixel
Max circle to hull radius	The maximum distance from the center of the bounding circle to an edge of the convex hull.	Pixel
Circle to hull radius ratio	$\frac{\text{Max circle to hull radius}}{\text{Min circle to hull radius}}$	Unit-less
Mean circle to hull radius	The mean of all radii drawn from the circle's center to the hull	Pixel
CV_Circle2Hull_Radius	$\frac{STD(\text{ all radii drawn from the circle's center to the hull } )}{\text{Mean circle to hull radius}}$	Unit-less

Table S2-6. Nuclei geometric parameters.

<b>Parameter</b>	<b>Description</b>	<b>Unit</b>
Nuc area	Area of the nucleus	Pixel <sup>2</sup>
Nuc perimeter	Perimeter of the nucleus	Pixel
Nuc major	Major axis of an ellipse drawn around the nucleus	Pixel
Nuc minor	Minor axis of an ellipse drawn around the nucleus	Pixel
Nuc circularity	$\frac{4\pi \text{ Nuc area}}{\text{Nuc perimeter}^2}$	Unit-less
Nuc aspect ratio	$\frac{\text{Nuc major}}{\text{Nuc minor}}$	Unit-less
Nuc roundness	$\frac{4 \text{ Nuc area}}{\pi \text{ Nuc major}^2}$	Unit-less
Nuc solidity	$\frac{\text{Area of convex hull of nuclei}}{\text{Nuc area}}$	Unit-less
Distance nuc to cell centroid	Distance between the centroids of cell and nucleus	Pixel
Nuc to cell orientation	Angle between the distance nuc to cell centroid and cell major axis	Degree

Table S2-7. Roughness measurements.

Parameter	Formula	Units
Arithmetic average of absolute values	$R_a = \frac{1}{n} \sum_{i=1}^n  y_i $	Pixel
Root Mean Squared	$R_a = \sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2}$	Pixel
Ten-Point mean roughness	$R_{ZUS} = \frac{(Y_{p1} + Y_{p2} + Y_{p3} + Y_{p4} + Y_{p5}) + (Y_{v1} + Y_{v2} + Y_{v3} + Y_{v4} + Y_{v5})}{5}$	Pixel
Maximum peak to valley Roughness	$R_y = R_p + R_v$	Pixel
Maximum height of the profile	$R_{yt} = R_p - R_v$	Pixel
Skewness	$R_{sk} = \frac{1}{nR_q^3} \sum_{i=1}^n y_i^3$	Pixel <sup>3</sup>
Kurtosis	$R_{ku} = \frac{1}{nR_q^4} \sum_{i=1}^n y_i^4$	Pixel <sup>4</sup>

## Appendix II: supplementary information for chapter 3

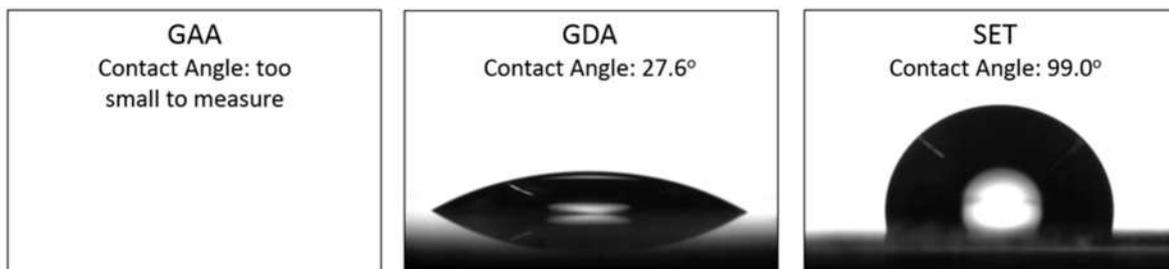


Figure S3.1 Contact angle measurements of the surfaces. Side view of a drop on GAA, GDA, and SET substrates. Hydrophobicity increases from left to right. Contact angle was too small to be measured for GAA substrates. It was 27.6° on average for GDA substrates, and 99.0° on average for SET.

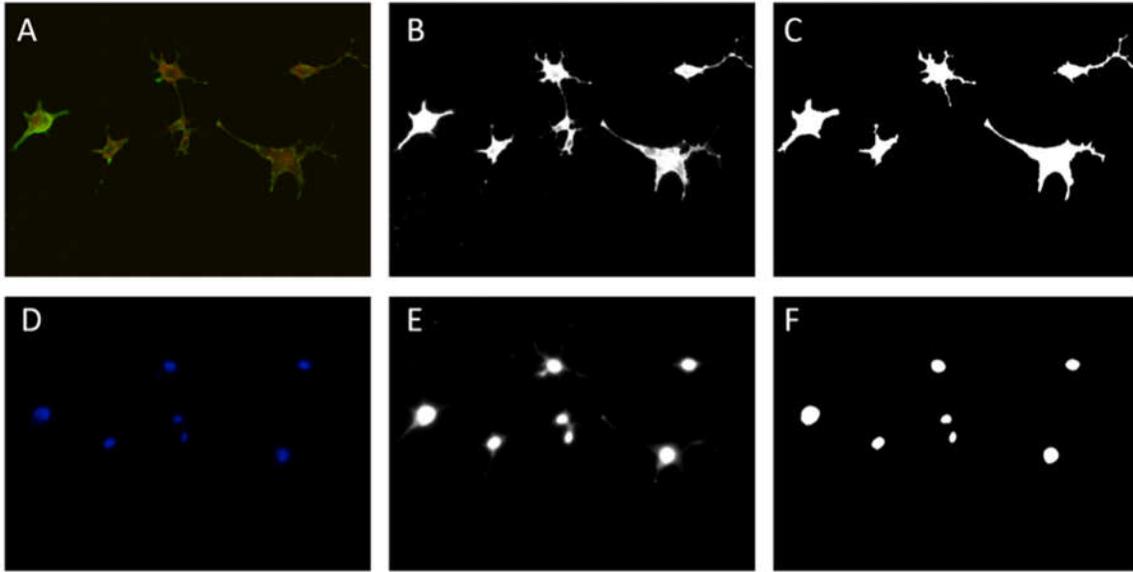


Figure S3.2 Image processing example for shape metrics. A) Membrane and actin images are enhanced and combined as cell image. D) Enhanced nuclei image. B) Sharpened edge boundaries for cell image. E) Sharpened edge boundaries for nuclei image. C) Binary image of cell. F) Binary image of nuclei.

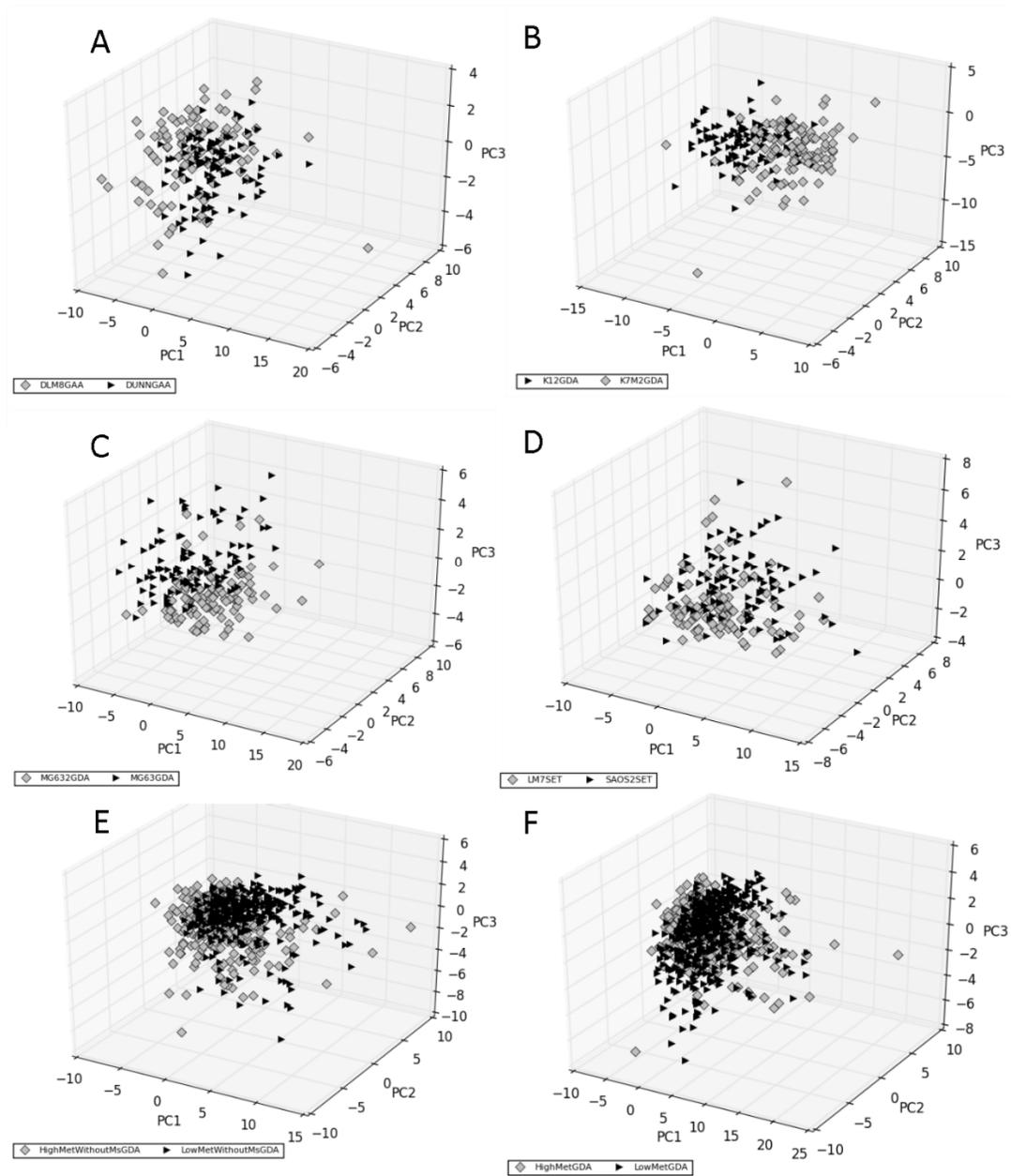


Figure S3.3 Principal component analysis. The principal component based comparisons that were performed, apart from those reported in the main paper figures. Each panel represents shape data projected on the first three Principal Components of all the morphometric characteristics of the cells. Each panel represents one comparison indicated in the legend. The gray diamond signs represent the high-met lines and the black triangle signs represent the low met lines.

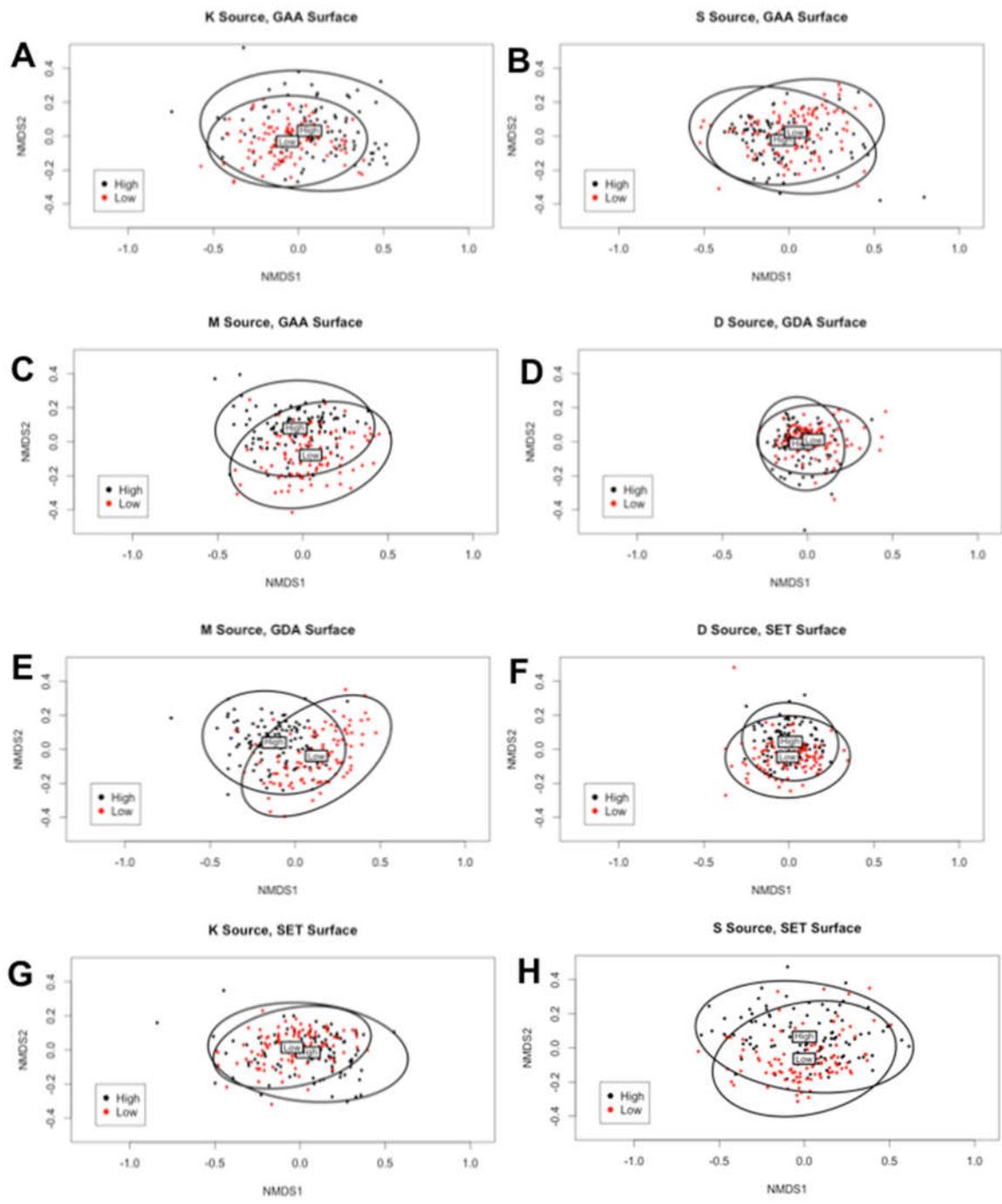


Figure S3.4 Nonmetric multidimensional scaling (NMDS) analysis. Each panel reports the NMDS based ordination plot of all comparisons we made except for those shown in the main paper figures. Each panel represents a specific comparison, with the legend indicating the pair of cell lines considered and the surface on which they were cultured. The labels “High” and “Low” refer to the data centroids of the high metastatic and low metastatic cell line of the pair. The ellipses represent 95% confidence intervals.

Table S3-1 Morphometric parameters. Description and units of the morphometric measures used for this analysis.

<b>Parameter</b>	<b>Description</b>	<b>Unit</b>
Cell Area	Area in pixels of the cell	pixels <sup>2</sup>
Cell Perimeter	Perimeter in pixels of the cell	pixels
Cell Major	The major axis of an ellipse drawn around the cell	pixels
Cell Minor	The minor axis of an ellipse drawn around the cell	pixels
Circularity of Cell	$4\pi(\text{Area})/\text{Perimeter}^2$	unitless (0-1)
Aspect Ratio of cell	Major axis of ellipse/minor axis of ellipse	unitless
Roundness of cell	$4(\text{Area})/(\pi(\text{Elliptical Major Axis})^2)$	unitless (0-1)
Solidity Cell	Area of convex hull of cell / area of the cell	unitless (0-1)
Max span across the hull	The maximum distance from one point on the convex hull to another	pixels
Area of the hull	Area of the convex hull (smallest convex polygon that encloses the cell)	pixels <sup>2</sup>
Perimeter of the hull	Perimeter of the convex hull	pixels
Circularity of the hull	$4\pi(\text{area of hull}/\text{hull perimeter}^2)$	unit less (0-1)
Max radius of the hull	Maximum Distance from centroid of Hull to an exterior point on the hull	pixels
Max/Min Radius of hull	Maximum radius of hull / minimum radius of hull (see max radius definition above)	unitless
CV Rad Hull	The relative variation of radii drawn from the hull's center to an exterior point. Given by the Standard of Deviation of all Radii divided into the mean of all radii	unitless
Mean Radius of hull	The mean of all radii drawn from the hull's centroid to an exterior point	pixels
Diameter of bounding circle	The diameter of the bounding circle drawn around the cell	pixels
Max radius from circle to hull	The maximum distance from the center of the circle to an edge of the convex hull	pixels
Max/Min Radius from Circle	Maximum radius from the center of the circle to an edge of the convex hull / minimum radius from the center of the circle to an edge of the convex hull	unit-less
CV Radius from circle to hull	The relative variation of radii drawn from the circle's center to the hull. Given By the Standard of Deviation of all Radii divided into the mean of all radii	unit-less
Mean radius from circle to hull	The mean of all radii drawn from the circle's center to the hull	pixels

Nucleus area	area of the nucleus in pixels <sup>2</sup>	pixels <sup>2</sup>
Nucleus perimeter	perimeter of the nucleus in pixels	pixels
Nucleus major	Major axis of an ellipse drawn around the nucleus	pixels
Nucleus minor	Minor axis of an ellipse drawn around the nucleus	pixels
Nucleus circularity	$4\pi (\text{Area})/\text{Perimeter}^2$	unitless (0-1)
Nucleus aspect ratio	Major axis of ellipse/minor axis of ellipse	unitless
Nucleus roundness	$4(\text{Area})/(\pi (\text{Elliptical Major Axis})^2)$	unitless (0-1)
Nucleus solidity	Area of convex hull of cell / area of the cell	unit less (0-1)

TableS3-2 Complete t-test results. The t-test results for paired comparisons between the high met and low met cell line for 29 morphometric characteristics for each paired line on all surfaces. The abbreviations used are as follows: GDA: Glass Detergent washed; GAA: Glass Acid etched; SET: Siliconized glass, Ethanol Treated. L: low metastatic cell line; H: high metastatic cell line. The cell lines are referred to as follows: (i) DL: DUNN; (ii) DH: DLM8; (iii) KL: K12; (iv) KH: K7M2; (v) SL: SAOS2; (vi) SH: LM7-Saos; (vii) ML: MG63; (viii) MH: MG63.2

CELL LINES	DL vs DH on GDA			KL vs KH on GDA		
	Significant (1) or not (0)	p-value	Cell Line with higher mean	Significant (1) or not (0)	p-value	Cell Line with higher mean
Area	1	1.70E-07	L	1	9.81E-23	L
Perimeter	1	0.001571742	L	1	1.33E-23	L
Major axis	1	1.00376E-04	L	1	5.13E-09	L
Minor axis	1	7.65E-05	L	1	3.13E-23	L
Circularity	0	0.983677269	H	1	1.35E-13	H
Aspect Ratio	0	0.434960439	H	1	4.90416E-04	H
Roundness	0	0.054961952	L	0	0.734672912	L
Solidity	1	0.019938295	L	1	1.13E-05	H
Max Span Hull	1	9.54706E-04	L	1	7.66E-10	L
Area Hull	1	3.76E-05	L	1	8.95E-18	L
Perimeter Hull	1	1.12529E-04	L	1	3.56E-15	L
Circularity Hull	0	0.111165111	L	1	6.27E-08	L
Max Rad Hull	1	0.00304524	L	1	8.59E-09	L
Max/Min Rad Hull	0	0.451931518	H	1	7.15E-08	H
CV Rad Hull	0	0.345243155	H	1	4.08E-08	H
Mean Rad Hull	1	9.79594E-04	L	1	4.61E-12	L
Diameter Bounding Circle	1	0.001415995	L	1	4.44E-10	L
Max Rad Circle	1	0.001427271	L	1	4.42E-10	L
Max/Min Circle	0	0.311128332	H	1	0.00196701	H
CV Rad Circle	0	0.696827332	H	1	0.001900904	H
Mean Rad Circle	1	0.001463127	L	1	4.57E-11	L
Nucleus Area	1	6.38E-07	L	1	0.040689462	H
Nucleus Perimeter	1	2.33E-07	L	1	0.013025999	H
Nucleus Major	1	2.76E-12	L	1	0.016036836	H
Nucleus Angle	0	0.799829655	L	0	0.750885339	H
Nucleus Circularity	1	0.018320778	H	1	0.004260063	L
Nucleus Aspect Ratio	1	5.11E-07	L	0	0.202738897	H
Nucleus Roundness	1	6.62E-07	H	0	0.077355812	L
Nucleus Solidity	0	0.223859205	H	0	0.227590653	L
	<b>SL v SH on GDA</b>			<b>ML vs MH on GDA</b>		
	Significant (1) or not (0)	p-value	Cell Line with higher mean	Significant (1) or not (0)	p-value	Cell Line with higher mean
Area	1	0.024815142	L	1	3.70E-15	H
Perimeter	1	8.07E-10	L	1	1.68E-20	H
Major axis	1	0.022839291	L	1	1.69E-10	H





Nucleus Area	1	0.006289331	L	1	0.006707588	L
Nucleus Perimeter	1	0.001425287	L	1	5.23337E-04	L
Nucleus Major	0	0.847030057	0	1	7.52E-09	L
Nucleus Angle	1	6.92E-06	L	0	0.674316166	H
Nucleus Circularity	1	0.002641323	H	1	5.12E-06	H
Nucleus Aspect Ratio	1	3.23E-07	H	1	8.87E-13	L
Nucleus Roundness	1	1.01E-07	L	1	1.98E-13	H
Nucleus Solidity	1	4.08E-06	H	0	0.960896383	H
	<b>DL vs DH on SET</b>			<b>KL vs KH on SET</b>		
	Significant (1) or not (0)	p-value	Cell Line with higher mean	Significant (1) or not (0)	p-value	Cell Line with higher mean
Area	1	7.07E-08	L	1	0.001359293	L
Perimeter	1	1.06E-02	L	1.00E-00	1.92E-04	L
Major axis	0	0.31028643	L	1	0.013883123	L
Minor axis	1	1.18E-05	L	1	0.003635536	L
Circularity	0	0.084699895	L	1	0.00244293	H
Aspect Ratio	1	0.031597212	H	0	0.795298041	L
Roundness	1	1.07E-08	L	0	0.373415036	H
Solidity	1	4.36E-06	L	0	0.16293527	H
Max Span Hull	0	0.537074144	L	1	0.025277059	L
Area Hull	1	0.004863484	L	1	0.021865475	L
Perimeter Hull	0	0.100952481	L	1	0.008627989	L
Circularity Hull	1	2.86E-06	L	0	0.098235037	L
Max Rad Hull	0	0.76158252	L	0	0.062899103	L
Max/Min Rad Hull	1	0.017263712	H	0	0.343331551	H
CV Rad Hull	1	0.001714675	H	0	0.433784196	H
Mean Rad Hull	0	0.425063751	L	0	0.050041448	L
Diameter Bounding Circle	0	0.541368617	L	1	0.024810174	L
Max Rad Circle	0	0.53999839	L	1	0.024854432	L
Max/Min Circle	0	0.243584829	H	0	0.829351169	H
CV Rad Circle	1	0.029595895	H	0	0.590918861	L
Mean Rad Circle	0	0.447429427	L	0	0.0588553	L
Nucleus Area	1	2.20E-07	L	0	0.399109003	H
Nucleus Perimeter	1	1.60E-08	L	0	0.278611101	H
Nucleus Major	1	6.76E-09	L	0	0.136841267	H
Nucleus Angle	0	0.216745907	L	0	0.927478174	H
Nucleus Circularity	1	3.67561E-04	H	0	0.273993333	L
Nucleus Aspect Ratio	0	0.223888859	L	0	0.092011377	H
Nucleus Roundness	0	0.2625561	H	0	0.082731096	L
Nucleus Solidity	0	0.074281351	H	0	0.124702902	L

	SL v SH on SET			ML vs MH on SET		
	Significant (1) or not (0)	p-value	Cell Line with higher mean	Significant (1) or not (0)	p-value	Cell Line with higher mean
Area	1	1.64641E-04	L	1	1.49E-17	H
Perimeter	1	1.15E-06	L	1.00E-00	3.21E-21	H
Major axis	0	0.217680048	0	1	7.31E-14	H
Minor axis	1	1.84E-10	L	1	3.38E-20	H
Circularity	1	3.24E-05	H	1	1.94E-15	L
Aspect Ratio	1	2.12E-06	H	0	0.875889932	L
Roundness	1	1.26E-05	L	1	9.62E-06	L
Solidity	1	0.00673473	H	1	3.05E-24	L
Max Span Hull	0	0.976305895	0	1	1.29E-15	H
Area Hull	1	7.10E-05	L	1	8.41E-19	H
Perimeter Hull	0	0.168120354	0	1	1.29E-20	H
Circularity Hull	1	6.66E-05	L	0	0.111580393	L
Max Rad Hull	0	0.741891225	0	1	1.86E-15	H
Max/Min Rad Hull	1	4.48139E-04	H	1	0.048981042	L
CV Rad Hull	1	0.002198444	H	1	0.014724811	L
Mean Rad Hull	0	0.747565824	0	1	2.13E-19	H
Diameter Bounding Circle	0	0.927051607	0	1	6.85E-16	H
Max Rad Circle	0	0.925962896	0	1	6.85E-16	H
Max/Min Circle	1	0.041942415	H	1	0.0064255	L
CV Rad Circle	1	0.006119484	H	1	4.95E-06	L
Mean Rad Circle	0	0.715080747	0	1	1.84E-19	H
Nucleus Area	1	1.91E-07	L	0	0.889749117	H
Nucleus Perimeter	1	7.18E-07	L	0	0.20859413	L
Nucleus Major	1	0.005380744	L	1	0.004868009	L
Nucleus Angle	1	3.16E-10	L	0	0.825562251	H
Nucleus Circularity	1	0.002951905	H	1	6.79E-14	H
Nucleus Aspect Ratio	1	8.20E-06	H	1	2.97E-08	L
Nucleus Roundness	1	2.01E-05	L	1	1.65E-08	H
Nucleus Solidity	1	1.64E-05	H	1	4.05E-09	H

Table S3-3 True positive and true negative rates of different thresholds for sample classification. Here we show the true positive and true negative rate data against different thresholds used to classify a sample of cells as belonging to the high-met or low met lines. Thresholds tested were [0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9]. In each case, if the proportion of cells in the sample of class high-met were greater than the threshold, the sample was classified as belonging to that class. Good predictions are considered to be those where both true positives and true negatives were equal to or greater than 0.8 (i.e. 80%). It is clear that a threshold of 0.6 has the best performance. The abbreviations used to describe cell lines and the surfaces are as follows: GDA: Glass Detergent washed; GAA: Glass Acid etched; SET: Siliconized glass, Ethanol Treated. L: low metastatic cell line; H: high metastatic cell line. The cell lines are referred to as follows: (i) DL: DUNN; (ii) DH: DLM8; (iii) KL: K12; (iv) KH: K7M2; (v) SL: SAOS2; (vi) SH: LM7-Saos; (vii) ML: MG63; (viii) MH: MG63.2. TPR stands for true positive rate and TNR stands for true negative rate.

Comparison	Classes	Threshold	0.3	0.4	0.5	0.6	0.7	0.8	0.9
D Lines GAA									
	DH	TPR	1	1	0.99	0.93	0.79	0.41	0.12
	DL	TNR	1	1	1	1	1	1	1
D Lines GDA									
	DH	TPR	1	1	0.99	0.84	0.49	0.24	0.06
	DL	TNR	0.06	0.22	0.51	0.79	0.97	1	1
D Lines SET									
	DH	TPR	1	1	1	0.93	0.67	0.25	0.09
	DL	TNR	0.72	0.96	1	1	1	1	1
D Lines All surfaces									
	DH	TPR	1	1	1	0.94	0.87	0.6	0.21
	DL	TNR	0.14	0.38	0.64	0.84	0.92	1	1
K Lines GAA									
	KH	TPR	1	1	1	0.99	0.98	0.82	0.39
	KL	TNR	0.38	0.73	0.97	1	1	1	1
K Lines GDA									
	KH	TPR	1	1	1	1	1	0.98	0.68
	KL	TNR	0.38	0.73	0.95	0.99	1	1	1
K Lines SET									
	KH	TPR	1	1	0.97	0.85	0.57	0.32	0.05
	KL	TNR	0	0.03	0.16	0.47	0.82	0.95	1
K Lines All surfaces									
	KH	TPR	1	1	0.97	0.82	0.51	0.27	0.08
	KL	TNR	0.25	0.52	0.82	0.92	0.99	1	1
M Lines GAA									
	MH	TPR	1	1	1	1	0.94	0.71	0.32
	ML	TNR	0.75	0.97	1	1	1	1	1
M Lines GDA									
	MH	TPR	1	1	1	1	1	0.99	0.7
	ML	TNR	0.97	1	1	1	1	1	1
M Lines SET									
	MH	TPR	1	1	1	1	1	0.96	0.75
	ML	TNR	0.93	1	1	1	1	1	1
M Lines All surfaces									

	MH	TPR	1	1	1	0.97	0.88	0.7	0.41
	ML	TNR	0.83	0.94	0.97	0.99	1	1	1
S Lines GAA									
	SH	TPR	1	1	1	1	1	0.95	0.6
	SL	TNR	0.04	0.25	0.53	0.8	0.98	1	1
S Lines GDA									
	SH	TPR	1	1	1	0.96	0.82	0.51	0.2
	SL	TNR	0.27	0.62	0.89	0.99	1	1	1
S Lines SET									
	SH	TPR	1	1	1	0.96	0.81	0.5	0.14
	SL	TNR	0.11	0.34	0.62	0.89	1	1	1
S Lines All surfaces									
	SH	TPR	1	1	0.98	0.93	0.77	0.45	0.22
	SL	TNR	0.12	0.33	0.56	0.83	0.96	1	1
All H vs L GAA									
	H	TPR	1	0.99	0.98	0.9	0.65	0.38	0.1
	L	TNR	0.04	0.18	0.39	0.66	0.86	0.94	0.97
All H vs L GDA									
	H	TPR	1	1	0.99	0.96	0.77	0.48	0.27
	L	TNR	0	0.05	0.17	0.37	0.7	0.92	0.99
All H vs L SET									
	H	TPR	1	0.99	0.92	0.75	0.53	0.31	0.12
	L	TNR	0.1	0.35	0.64	0.84	0.97	1	1
All H vs L All surfaces									
	H	TPR	1	1	0.98	0.92	0.73	0.42	0.16
	L	TNR	0.11	0.33	0.48	0.68	0.93	1	1
Type-1 H vs L GAA									
	H	TPR	1	1	0.98	0.93	0.78	0.45	0.23
	L	TNR	0.07	0.21	0.49	0.73	0.9	0.98	1
Type-1 H vs L GDA									
	H	TPR	1	1	1	0.86	0.7	0.46	0.22
	L	TNR	0.24	0.49	0.75	0.95	0.98	1	1
Type-1 H vs L SET									
	H	TPR	1	1	1	0.95	0.88	0.56	0.29
	L	TNR	0.28	0.57	0.78	0.97	1	1	1
Type-1 H vs L All surfaces									
	H	TPR	1	0.98	0.94	0.88	0.69	0.38	0.17
	L	TNR	0.32	0.54	0.72	0.89	0.99	1	1

Appendix III: supplementary information for chapter 4

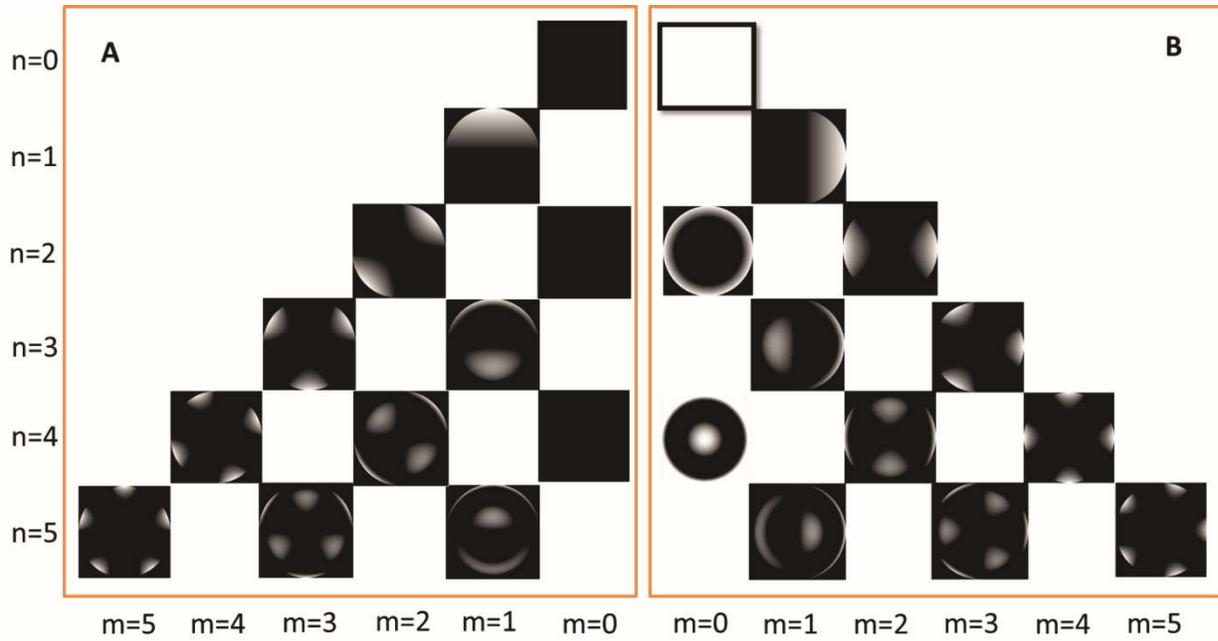


Figure S4.1 Zernike polynomials representation. In this figure, Zernike polynomials are plotted from  $n=0$  to  $n=5$ . A) Imaginary part of Zernike polynomials. B) Real part of Zernike polynomials. Zernike polynomials with even order and zero repetition are symmetrical under rotation.

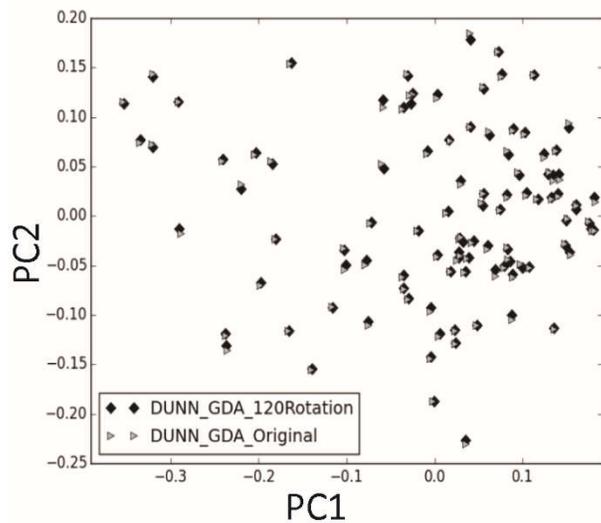


Figure S4.2 Rotation invariant Zernike moments. Zernike moments of DUNN cell line on GDA substrate and its Zernike moments of 120 degrees rotated images in principal component space shows that they are almost equal. This confirms that, Zernike moments used in this study are rotation invariant.

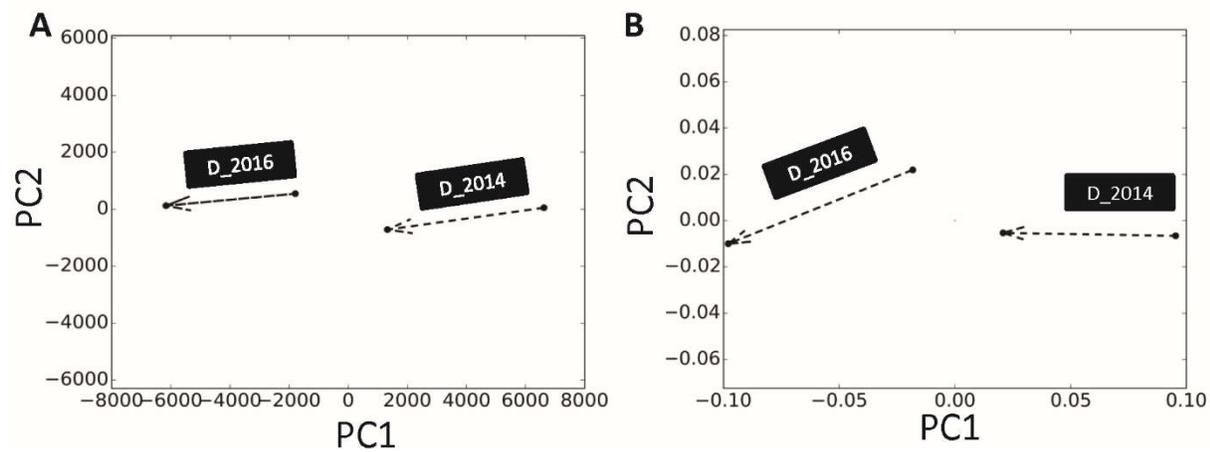


Figure S4.3 Reproducibility of shape changes. Data from two separate experimental batches year in principal component space (A) Geometric parameters: PC1 and PC2 changes are similar for experiments done in 2014 and 2016 (B) Zernike moments: PC1 changes is very similar for the experiments done in 2014 and 2016.

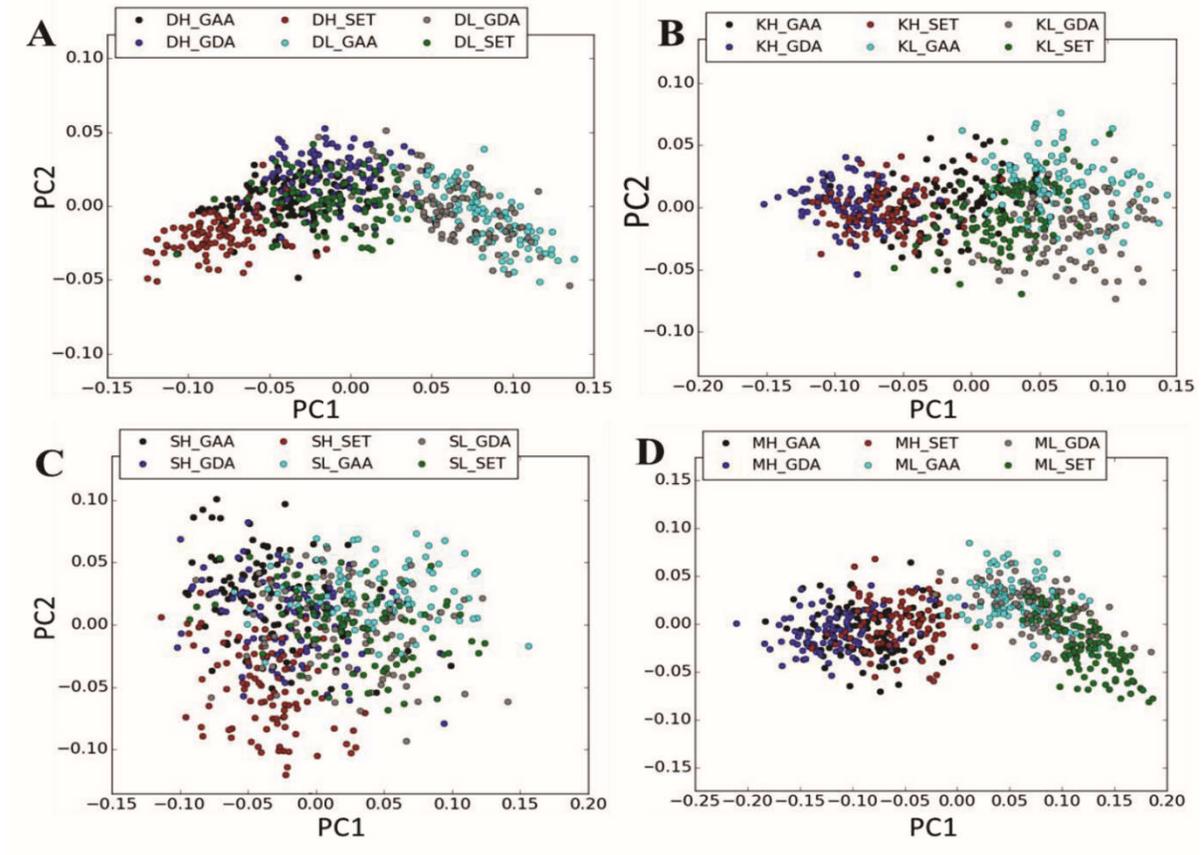


Figure S4.4 Clustering of each paired line on different substrates in principal component space. A) D paired cell lines. B) K paired cell lines. C) S paired cell lines. D) M paired cell lines.

Table S4-1 The cell lines and the abbreviations used to describe them in this paper. The abbreviation is based on the first letter of each pair of cell lines, with the second letter denoting whether it is classified as having low or high metastatic potential. Cell lines 1-4 are murine while cell lines 5-8 are human.

S. No.	Original Cell Line Name	Metastatic Classification	Nomenclature in this paper
1.	DUNN	Low	DL
2.	DLM8	High	DH
3.	K12	Low	KL
4.	K7M2	High	KH
5.	SAOS2	Low	SL
6.	SAOS-LM7	High	SH
7.	MG63	Low	ML
8.	MG63.2	High	MH

Table S4-2 Prediction accuracy of the neural network when it is trained on one set of cell lines and another set of cell lines are used for validation. The best performing cases are highlighted here and reported in Table 4 of the main text.

Training and Test cell lines	Validation cell lines	Substrate	Accuracy
DL versus DH	KL versus KH	All	0.69
DL versus DH	KL versus KH	GAA	0.52
DL versus DH	KL versus KH	GDA	0.85
DL versus DH	KL versus KH	SET	0.785
DL versus DH	SL versus SH	All	0.68
DL versus DH	SL versus SH	GAA	0.81
DL versus DH	SL versus SH	GDA	0.625
DL versus DH	SL versus SH	SET	0.69
KL versus KH	DL versus DH	All	0.67
KL versus KH	DL versus DH	GAA	0.565
KL versus KH	DL versus DH	GDA	0.525
KL versus KH	DL versus DH	SET	0.615
KL versus KH	SL versus SH	All	0.63
KL versus KH	SL versus SH	GAA	0.52
KL versus KH	SL versus SH	GDA	0.66
KL versus KH	SL versus SH	SET	0.405
SL versus SH	DL versus DH	All	0.725
SL versus SH	DL versus DH	GAA	0.635
SL versus SH	DL versus DH	GDA	0.665
SL versus SH	DL versus DH	SET	0.81
SL versus SH	KL versus KH	All	0.83
SL versus SH	KL versus KH	GAA	0.57
SL versus SH	KL versus KH	GDA	0.895
SL versus SH	KL versus KH	SET	0.64
SL and DL versus SH and DH	KL versus KH	All	0.725
SL and DL versus SH and DH	KL versus KH	GAA	0.58
SL and DL versus SH and DH	KL versus KH	GDA	0.945
SL and DL versus SH and DH	KL versus KH	SET	0.495
SL and KL versus KL and KH	DL versus DH	All	0.675
SL and KL versus KL and KH	DL versus DH	GAA	0.925
SL and KL versus KL and KH	DL versus DH	GDA	0.54
SL and KL versus KL and KH	DL versus DH	SET	0.205
KL and DL versus KH and DH	SL versus SH	All	0.66
KL and DL versus KH and DH	SL versus SH	GAA	0.775
KL and DL versus KH and DH	SL versus SH	GDA	0.59
KL and DL versus KH and DH	SL versus SH	SET	0.66

Table S4-3 Sorted weights for PCA. For each principal component Zernike moments are sorted based on their weights. First column is sorted Zernike moments for PC1 and third column is sorted Zernike moments for PC2. It can be seen that in PC1 early even orders with m=0 and for PC2 later even orders with m=2 have the largest weights.

PC1		PC2	
Attributes	Weights	Attributes	Weights
4_0	0.401406	10_2	0.110873
6_0	0.382081	12_2	0.109819
2_0	0.309363	14_2	0.099882
8_0	0.278669	8_2	0.097855
10_2	0.160788	18_4	0.08918
12_2	0.155303	15_1	0.088992
8_2	0.138404	16_2	0.088989
10_0	0.131511	16_4	0.086615
14_2	0.126728	17_1	0.085775
0_0	0.116983	20_4	0.085324
6_2	0.09455	13_1	0.082725
16_4	0.09018	18_2	0.08064
16_2	0.08985	19_1	0.080287
18_4	0.086394	22_4	0.079989
14_4	0.086118	20_2	0.078865

**Appendix IV: supplementary information for chapter 5**

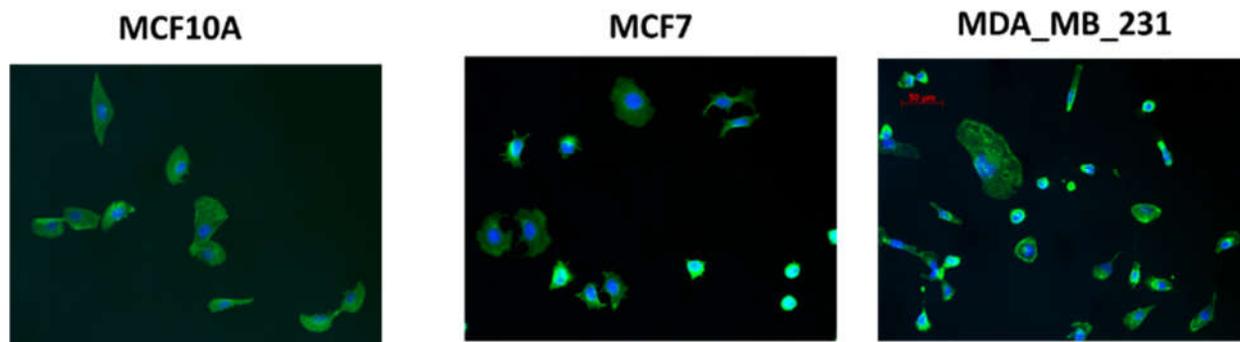


Figure S5-1 Representative images of breast cells. The cells are grown on fibronectin coated glasses. The scale bar is 50um.

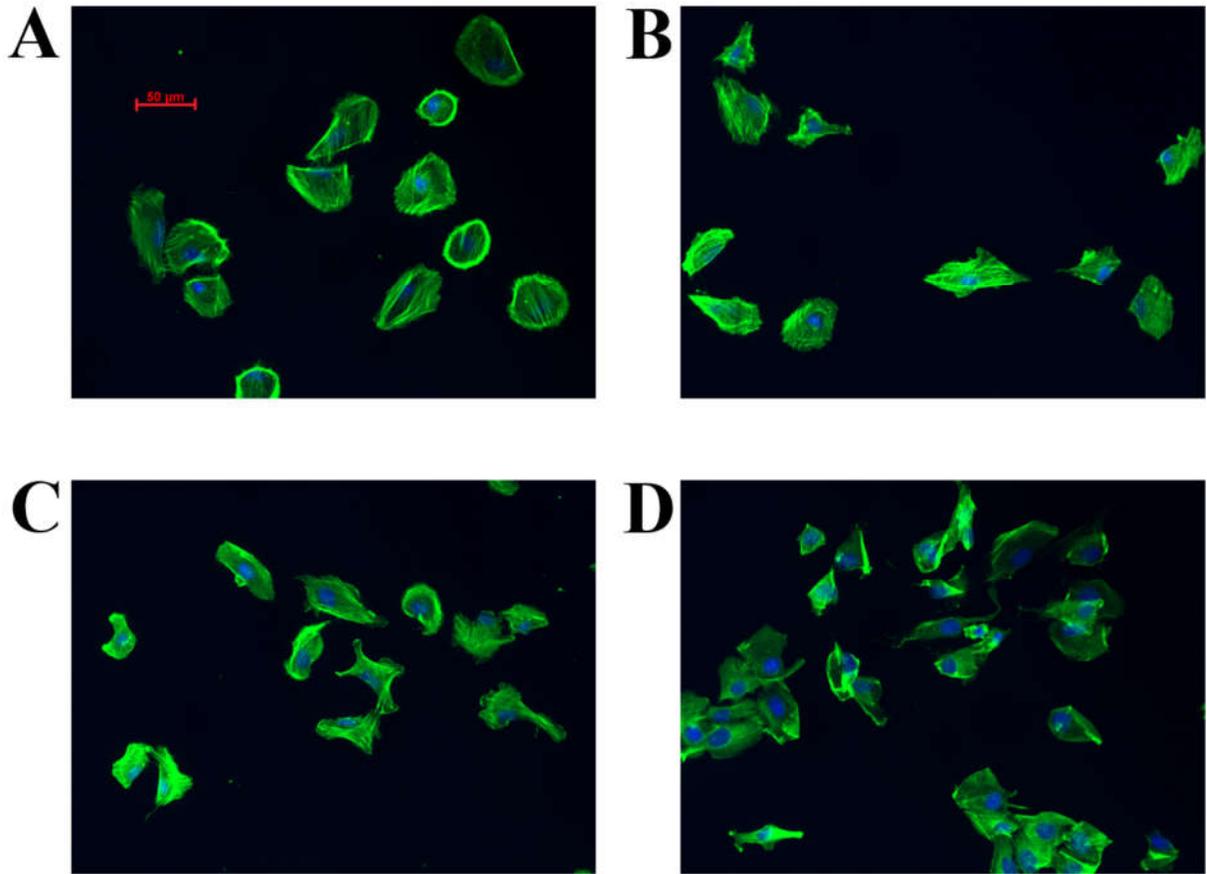


Figure S5-2: Representative images of Retinal cells. A) Normal epithelial ARPE19 cells. B) Cancerous AKT<sup>myr</sup> cells. C) Cancerous MEKK<sup>DD</sup> cells. D) Cancerous RAS<sup>v12</sup> cells. The representative images belong to the cells grown on GDA substrate. The scale bar is 50um.

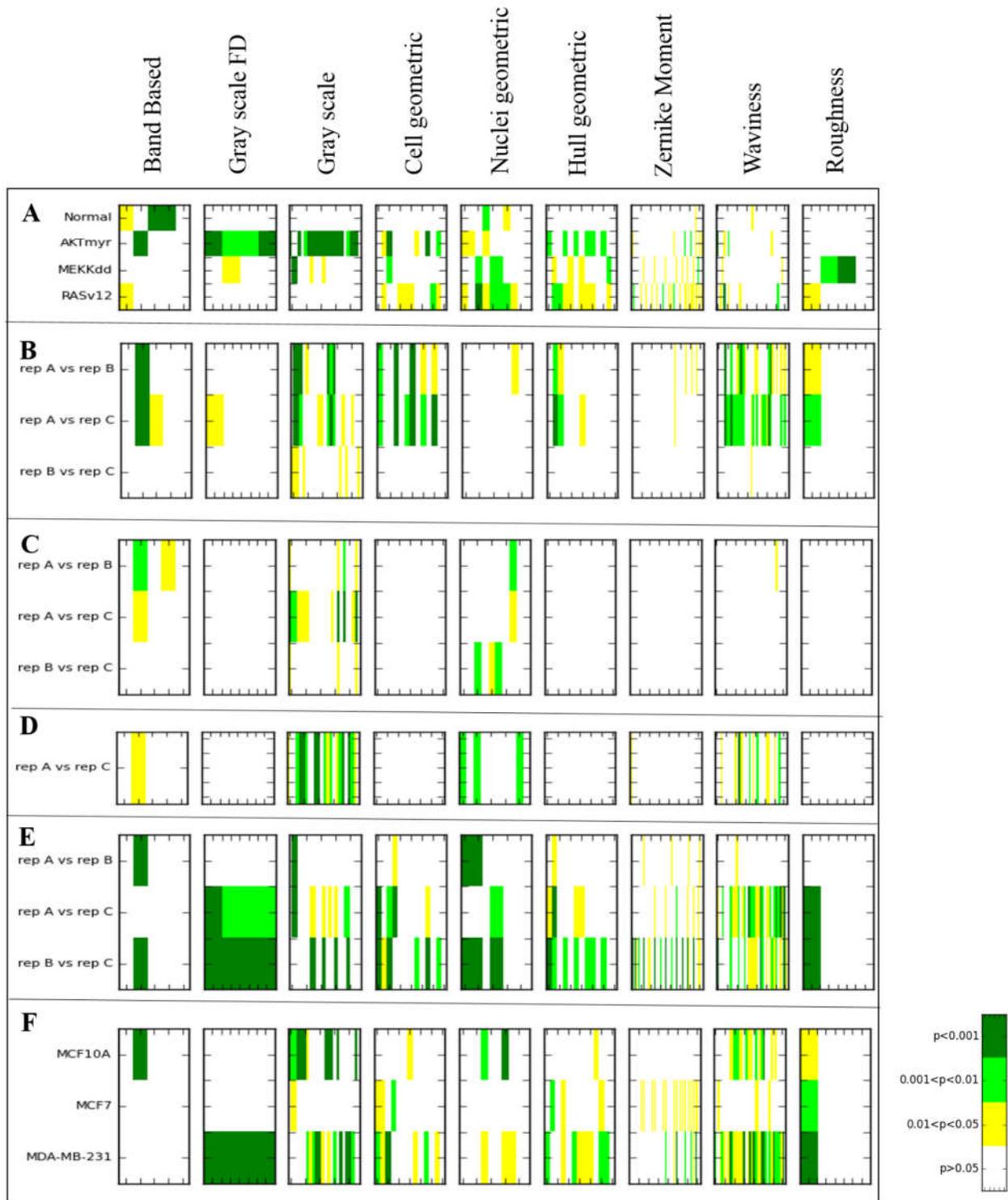


Figure S5-3 Similarity tests. Heat map plot of p-value of comparison between replicates of each cell line is plotted. The lighter the color is the larger p-values are. So, the lighter colors means cells that are being compared are more similar. A) Retina cells on FN substrate. B) Normal retinal cells on GDA substrate. C) Retinal AKT<sup>myr</sup> cells on GDA substrate. D) Retinal MEK<sup>Kdd</sup> cells on GDA substrate. Retinal RAS<sup>V12</sup> cells on GDA substrate. Breast cells on FN substrates.

Table S5-1 The direction of changes in PC1 with cancer progression for different categories. In shape categories which changes in PC1 for M cell line with cancer progression is significant (p-value<0.001), the direction of changes in other cell lines with cancer progression is exactly opposite of M cell line. Results for remaining shape categories of gray scale, band based, and nuclei measures are not shown here since changes in PC1 for M cell line was not significant for these shape categories.

Gray scale FD			Cell geometric			Hull geometric		
Osteosarcoma cells	DUNN vs DLM8	↓	Osteosarcoma cells	DUNN vs DLM8	↓	Osteosarcoma cells	DUNN vs DLM8	↓
	K12 vs K7M2	↓		K12 vs K7M2	↓		K12 vs K7M2	↓
	SAOS2 vs LM7	↓		SAOS2 vs LM7	↓		SAOS2 vs LM7	↓
	MG63 vs MG63.2	↑		MG63 vs MG63.2	↑		MG63 vs MG63.2	↑
Breast cells	MCF7 vs MDA_MB_231	↓	Breast cells	MCF7 vs MDA_MB_231	↓	Breast cells	MCF7 vs MDA_MB_231	↓
Retina cells	Normal retina vs AKTmyr	↓	Retina cells	Normal retina vs AKTmyr	↓	Retina cells	Normal retina vs AKTmyr	↓
	Normal vs MEKKdd	↓		Normal vs MEKKdd	↓		Normal vs MEKKdd	↓
	Normal vs RASv12	↓		Normal vs RASv12	↓		Normal vs RASv12	↓
Zernike Moment			Waviness			Roughness		
Osteosarcoma cells	DUNN vs DLM8	↓	Osteosarcoma cells	DUNN vs DLM8	↑	Osteosarcoma cells	DUNN vs DLM8	↓
	K12 vs K7M2	↓		K12 vs K7M2	↑		K12 vs K7M2	↓
	SAOS2 vs LM7	↓		SAOS2 vs LM7	↑		SAOS2 vs LM7	↓
	MG63 vs MG63.2	↑		MG63 vs MG63.2	↓		MG63 vs MG63.2	↑
Breast cells	MCF7 vs MDA_MB_231	↓	Breast cells	MCF7 vs MDA_MB_231	↓	Breast cells	MCF7 vs MDA_MB_231	↓
Retina cells	Normal retina vs AKTmyr	↓	Retina cells	Normal retina vs AKTmyr	↓	Retina cells	Normal retina vs AKTmyr	↑
	Normal vs MEKKdd	↓		Normal vs MEKKdd	↓		Normal vs MEKKdd	↓
	Normal vs RASv12	↓		Normal vs RASv12	↓		Normal vs RASv12	↓

-  Significant increase in PPC
-  Significant decrease in PPC
-  Not significant( p-value> 0.001)