Computers in Biology and Medicine 126 (2020) 104044

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: http://www.elsevier.com/locate/compbiomed

# Cellular morphological features are predictive markers of cancer cell state

Elaheh Alizadeh [a], Jordan Castle [b], Analia Quirk [a,c], Cameron D.L. Taylor [a,c], Wenlong Xu [a], Ashok Prasad [a,c,*]

[a] Department of Chemical and Biological Engineering, USA
[b] Department of Biology, USA
[c] School of Biomedical Engineering, Colorado State University, Fort Collins, CO, 80523, USA

ARTICLE INFO

*Keywords:*
Cell morphology
Metastasis
Cell shape
Machine learning
Cytoskeletal structure
Actin

ABSTRACT

Even genetically identical cells have heterogeneous properties because of stochasticity in gene or protein expression. Single cell techniques that assay heterogeneous properties would be valuable for basic science and diseases like cancer, where accurate estimates of tumor properties is critical for accurate diagnosis and grading. Cell morphology is an emergent outcome of many cellular processes, potentially carrying information about cell properties at the single cell level. Here we study whether morphological parameters are sufficient for classification of single cells, using a set of 15 cell lines, representing three processes: (i) the transformation of normal cells using specific genetic mutations; (ii) metastasis in breast cancer and (iii) metastasis in osteosarcomas. Cellular morphology is defined as quantitative measures of the shape of the cell and the structure of the actin. We use a toolbox that calculates quantitative morphological parameters of cell images and apply it to hundreds of images of cells belonging to different cell lines. Using a combination of dimensional reduction and machine learning, we test whether these different processes have specific morphological signatures and whether single cells can be classified based on morphology alone. Using morphological parameters we could accurately classify cells as belonging to the correct class with high accuracy. Morphological signatures could distinguish between cells that were different only because of a different mutation on a parental line. Furthermore, both oncogenesis and metastasis appear to be characterized by stereotypical morphology changes. Thus, cellular morphology is a signature of cell phenotype, or state, at the single cell level.

## 1. Introduction

Although methods of cancer diagnostics and treatment have improved over the past few decades, it is the second leading cause of death globally [1,2]. Around 90% of the deaths from cancer are due to metastasis [3], and the 5-year cancer survival rate is higher for patients who have been diagnosed in early stages of their cancer disease [4], but early diagnosis is also required to be accurate, otherwise it can increase patient morbidity [5]. Overdiagnosis is one of the pitfalls of early detection, and has been estimated to range between 10% and 22% in three follow-up studies of randomized controlled trials [6]. There is a great need therefore for accurate prediction of the properties of the tumor, which can help optimize therapy. The morphology of cells from samples extracted via fine-needle biopsy is already examined by experienced cytologists for tumor grading, but tumor grading is characterized by low reproducibility and low accuracy [7–10]. Other assays such as gene expression-based methods have had limited success, mostly for breast cancers [11]. One major reason for the lack of methods to predict metastasis is the heterogeneity of the tumor, which may be composed of cells that may differ drastically in their properties, without displaying major mutational differences. These include cancer cells that have acquired varying degrees of stemness, including cancer stem cells (CSCs) as well as well-differentiated cells that still resemble the tissue of origin (12). An important example is the Epithelial-to-Mesenchymal Transition (EMT), believed to underlie cancer metastasis in many cancers, which can be induced in cells without new gene mutations [13]. Drug-resistant cancer cells have also been shown to be a stochastically generated sub-population that have transitioned to a drug-resistant state [14]. Thus, high-throughput methods that could classify single cells according to their state or phenotype would be a significant advance in our ability to predict tumor metastasis, and possibly other signatures of disease.

This paper is based on the hypothesis, recently reviewed in Ref. [15,

* Corresponding author. Department of Chemical and Biological Engineering, USA.
*E-mail address:* ashok.prasad@colostate.edu (A. Prasad).

https://doi.org/10.1016/j.compbiomed.2020.104044
Received 13 July 2020; Received in revised form 4 October 2020; Accepted 6 October 2020
Available online 8 October 2020

16], that quantitative measures of cell shape and actin organization (which we call **cell morphometrics**) can distinguish between cells in different states, and may be capable of recognizing cancer cells that have acquired metastatic characteristics, de-differentiated into CSCs and other subpopulations in a tumor. Cell morphometrics as defined here include the shape of the cell and nucleus, and the structure and organization of the actin cytoskeleton. The hypothesis above is based on a body of previous work, beginning from early experimental studies that observed phenotypic changes in cells, such as differentiation along a particular lineage, arising from imposed shape alterations [17–22]. Studies in Drosophila [23] and mammalian cells [4,24,25] have confirmed that cell shape can provide state information, and may be sensitive to perturbations of specific proteins [24,26–28]. Shape characteristics of breast cancer cell lines have been linked with NFkB signaling [27], changes in the YAP/TAZ nuclear to cytoplasmic ratio (24) and even chemoresistance in colon cancer [29,30]. In pancreatic cancer a significant reduction in shape heterogeneity was observed to be a marker for metastasis [31]. In previous work we have shown that some measures of two-dimensional spread cell outline, as well its representation using Zernike moments, can be used to distinguish between osteosarcoma cell lines with differing metastatic potential [32,33]. We also found that shape changes between less and more metastatic osteosarcoma cell lines fell into two broad groups, Type 1 and Type 2. Type 1 shape changes appear to be similar to stereotypical Epithelial to Mesenchymal Transition (EMT) shape changes, where the cells become more elongated and spindly, while Type 2 changes made cells larger, more rounded and more amoeboid, resembling some invasive melanoma [34] and breast cancer cells (i.e. BT549) [35]. However, shape parameters were overlapping and the ability to classify single cells based on shape alone was poor. In order to improve shape classification at the single cell level, we developed a number of new shape parameters including an efficient Fourier series representation of shape and a number of quantities that measure various aspects of the texture of the actin cytoskeleton, incorporated in the TISMorph toolbox [36].

In this paper we apply the shape parameters in this toolbox to 15 cell lines, organized in 3 sets, to address a number of questions pertaining to morphological changes in cancer and metastasis. The three sets of cell lines represent metastatic transformation in human and murine osteosarcoma, in human breast cancer and the oncogenesis process driven by mutations in major human oncogenes respectively. With the help of statistical analysis and machine learning applied to the quantitative morphological parameters of these cells, we ask whether (i) shape changes in highly metastatic breast cancer as compared with less metastatic cancer are similar to shape changes seen in invasive osteosarcomas; (ii) cells undergoing oncogenic transformation can be distinguished from the parental line, and from each other, through morphology alone; (iii) shape changes between "normal" cells and cancer cells are similar across cancer types, and (iv) can the specially crafted shape parameters that we developed contribute to improved classification at the single cell level for distinguishing between all of these classes of cell types.

We find that crafted morphology metrics appear to be a sensitive readout of cell state, even at the single cell level, and cell types can be distinguished at the single cell level with high accuracy. This is even true when a parental cell line is transformed by specific oncogenes; the morphological metrics can still distinguish the transformed cells from each other. Furthermore, there seem to be many similar changes in morphology that accompany the process of oncogenic transformation, with the breast cancer cells studied also displaying type 1 morphology changes.

The three sets of cell lines are as follows. One set of eight are osteosarcomas, composed of 4 subsets of paired lines, two murine and two human, each with a low metastatic parental line and a derived high metastatic daughter line [32,33]. Three of the 7 remaining cell lines are MCF10A, MCF7 and MDA-MB-231, which constitute a very common studied set for breast cancer. MCF10A is a non-tumorigenic human mammary gland epithelial cell line, which has been widely used as a model for normal breast cells [37]. MCF7 is the most widely studied breast cancer line, which has been shown to have relatively low metastatic potential [38]. 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 [39, 40]. The final four cell lines are the Retinal pigment epithelium cells, ARPE-19, and three derived lines, in which the parental cell line is engineered into three cancer cell lines by a stepwise process culminating in transfection with either $Ras^{v12}$ or $AKT^{myr}$, or $Mek^{DD}$ oncogenes. Representative images of the osteosarcoma cell lines can be found in Ref. [33], and the others are shown in Fig A.1 and Fig A.2.

## 2. Methods

### 2.1. Cell lines and culture

The 15 cell lines used in this study are listed in Table 1. MCF10A, a human normal breast 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). The osteosarcoma lines, DUNN, DLM8, K12, and K7M2 cell lines were gifts from Dr. D. Thamm (CSU). MG63 and MG63.2 cell lines were gifted from Dr. D. Duval (CSU). SAOS2 and SAOS-LM7 cell lines were donated from Dr. E. S. Kleinerman (MD Anderson Cancer Center). The Retinal Pigment Epithelium (ARPE) laboratory-transformed cell lines were developed in Jennifer DeLuca's laboratory at CSU and were gifted to our lab along with the normal APRE19 cell line.

Retinal cell lines were cultured in Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (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) component 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 20 mM 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 previously [33] or on GDA substrates coated with Fibronectin (FN) with density of 2 μg/cm$^2$. 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 h and Osteosarcoma cells were fixed after 48 h. After fixing, cells were fluorescently labelled with DAPI (Bio legend) for nuclei and with Acti-stain 488 phalloidin (Cytoskeleton, Inc.) for actin, and then imaged using fluorescent microscope. (Fig. 1A–C).

### 2.2. Image processing

Images are processed using the steps demonstrated in Fig. 1D–E. The segmentation of actin and nuclei is done in CellProfiler software [41] (Fig. 1D). Then using script developed in house using Matlab and ImageJ, cells are visually inspected and corrected as follows. When two cells are identified as one object by CellProfiler, a line separating these two cells are drawn manually and one identified object is split into two separate objects. In addition when one cell is identified as two objects by the CellProfiler software these two objects are merged manually by

**Table 1**
Names of the cell lines and their acronyms used in the text.

| Cell Types | | Species | Cell lines | Acronyms | Substrates (Replicates) | Substrates (Replicates) |
|---|---|---|---|---|---|---|
| Breast Cells | | Human | MCF10A | BC_N[a] | FN[b] (2) | |
| | | Human | MCF7 | BC_LM[c] | FN (2) | |
| | | Human | MDA-MB-231 | BC_HM[d] | FN (2) | |
| Osteosarcoma Cancer cellls | Ds | Mouse | DUNN | OS_DL | GDA[e](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) | |
| | | | SAOS-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) |

[a] N: Normal cells.
[b] FN: Fibronectin coated glass.
[c] LM: Low Metastatic line.
[d] HM: High Metastatic line.
[e] GDA: Glass Detergent washed Air dried.

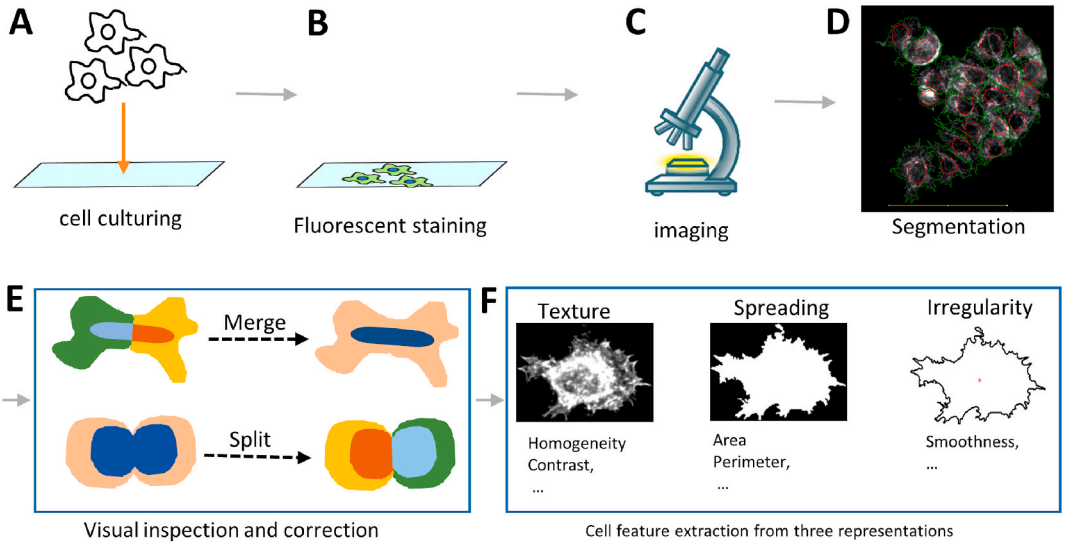


**Fig. 1. Data collection procedure.** A) Cells are cultured on glass substrate with and without fibronectin coating. B) Cells are stained for actin and nuclei. C) Then they are imaged using fluorescent microscopy. D) Cells are segmented using CellProfiler software (scale bar 100 μm). E) An operator visually inspects the segmentation and wrong segmentations are corrected. E) Cells are isolated and morphological features are extracted from three different representations of texture, spreading and irregularity.

clicking on them and using merge option. These steps are explained in detail in our previous publication [36], and the morphology is quantified as briefly described below. Codes are available online (details in Ref. [36]).

### 2.3. Morphology quantification

We collected between 100 and 300 individual cell images, per cell line, which were used to calculate the cellular morphometrics. These morphological quantifiers can be classified into three broad categories, spreading, waviness and textural measures. Each of these categories are features that are extracted from different representation of cell (Fig. 1F). Spreading measures are those calculated from a two-dimensional binary representation (Fig. 1F-Spreading) of the spread cell and include geometric characteristics of the cell and the nucleus, such as area, aspect ratio, solidity and circularity. They also include a two-dimensional representation using Zernike moments, which are an expansion in Zernike polynomials. Waviness measures are calculated from the coordinates of the cell boundary (Fig. 1F-Irregularity) and include a Fourier series expansion of the coordinates as well as measures of roughness of the boundary. Textural measures use the gray scale image of labelled actin (Fig. 1F-Texture), and they measure various characteristics of actin structure of the cell, such as its extent of linear or circular organization, its planar fractal dimension etc. A list of all the measures (over 256 in total), with formulae and descriptions, can be found in Ref. [36]. These 256 shape parameters can be classified into 9 shape categories: (i) Cell geometric (ii) Nuclei geometric (iii) Hull geometric (iv) Grayscale texture (v) Band based (vi) Gray scale fractal dimension (FD) (vii) Zernike Moments (viii) Waviness (Fourier) parameters and (ix) Roughness. Details of these shape measures can be found in Refs. [36]. In this paper we club together categories (i), (ii) and (iii) as "Geometric" measures, and (i)-(vi) and (ix) as "crafted" measures, since they are not complete mathematical descriptions of morphology but have been human-crafted to measure some aspect of it. Zernike and Fourier descriptors are also described as basis function-based measures.

### 2.4. 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 our data sets the first 4 principal components captured 99% of the total variation in the data, hence we restricted ourselves to the first 4 principal components for each shape category. Since each shape category captures information about a different aspect of cell shape and structure, we performed PCA on each of the 9 shape categories separately. In order to visualize if different cell classes cluster separately we picked the principal component (among the first 4) which leads to better separation for all of the classes under comparison, i.e. the PC whose worst case (based on the p-value of a *t*-test) was better than that of any other PC, so that it was the best single measure for distinguishing between all the comparisons for that shape category. We call this PC the "Primary Principal Component" or PPC. For visualization purposes we plotted the data shown in the figures as a scatter plot where each point in every class is an average of 10 random samples from that class in principal component space.

### 2.5. Machine learning methods for classification

In order to distinguish between cell classes, we used two commonly employed machine learning methods, Artificial Neural Networks (ANN) and Support Vector Machines (SVM). Machine learning analysis was carried out using the Neural Network Toolbox and the Statistics and Machine Learning Toolbox of the software suite, Matlab®. The ANNs used were a standard two-layer feedforward neural network with one hidden layer, usually consisting of 20 nodes, and two output nodes for binary classification. The network used a sigmoid transfer function in the hidden layer, and a softmax transfer function in the output layer. The training method used was the scaled conjugate gradient method with default parameters (minimum performance gradient $1 \times 10^{-6}$; weight for second derivative approximation $5 \times 10^{-5}$; lambda $= 5 \times 10^{-7}$), and the performance was assessed by using cross-entropy. Cross-entropy for a given target value, t, is defined as -tLn(p) where p is the output of the corresponding output neuron. Training continues till the mean cross-entropy is minimized. We employed 10-fold cross validation for the neural network analysis presented here. Basically, data was divided into a test set (10%) and a training set (80%) and a validation set (10%). Model fitting and optimization was performed on the training and validation sets, while the test set was used for pure prediction. The process was repeated 10-times. The results reported are averages of the results for the pure prediction on the test set.

Linear SVMs work by finding the best hyperplane that separates the datapoints of the two classes. We did not need to use non-linear SVMs since the performance of the linear models was quite good. However, the best linear models were found by optimizing the hyper-parameters of the model to minimize misclassification on the test set using 5-fold cross-validation, using an inbuilt Matlab procedure (in the function *fitcsvm*). In order to prevent "bleed-through" of the data from the test set into the training set, we carried out a hierarchical or nested cross-validation procedure in the following way. We divided the entire data set into four blocks. We removed each block by turns, and used the remaining 75% of the data for training and optimization of hyperparameters to minimize misclassification, using 5-fold cross-validation. The final optimal model was then used for a pure prediction on the 25% of the data that was heldout, and the classification accuracy calculated. The procedure was repeated for each of the 4 holdout blocks of the data, and the results reported are averages over the 4 blocks. For both ANNs and SVMs classification accuracy reported here is the percentage of true predictions out of the total test set. Confusion matrices are reported in the Supplementary information.

It should be noted that most of the results presented here are based on the set of features described as "crafted features" above, which comprise a vector of 96 numbers, that are not all independent. Our smallest dataset is the osteosarcoma dataset where we only collected 100 cell images per cell line. However even for these comparisons, despite the number of features, our algorithms performed well without any modification or optimization. We think that this is because morphological metrics for less and more metastatic osteosarcomas are generally well separated even in linear PCA analysis (except for DUNN and DLM8 cell lines, where the accuracy is correspondingly low). The more difficult comparisons reported here are between sets of at least 300 cells per class, where the 'curse of dimensionality' would be less operative.

In earlier publications [32,33] predictive accuracy was calculated based on the performance of the algorithm on predicting the class of a small sample of 10 cells from each class, using a simple majority rule, i.e. 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. This is similar to the cumulative probability that the number of successes (i.e. cells of the right class) in 10 Bernoulli trials is greater than 5 when sampling from a population with a given proportion, r, of cells of the right class, and can be calculated in Matlab as 1-binocdf (5, 10, r). This is the value reported as "sample accuracy" in the tables and is an estimate of the classification accuracy on a small sample of cells such as that would be obtained by a biopsy. In other words, if we had taken a sample of 10 cells from the metastatic class, and the accuracy of the ANN/SVM to distinguish that class from the less metastatic class is r, the "sample accuracy" number calculates the probability (as a percentage) that more than 5 cells in that sample are classified as metastatic.

Data and sample codes are available in the Supporting Information of this paper.

## 3. Results

### 3.1. Most shape measures are useful parameters for classifying cell lines

We first asked whether shape measures were good measures for distinguishing between different cell lines globally. We used a relatively broad criterion here for a good measure, i.e. a good shape feature is the one which p-value between technical replicates of the same cell line is not significant at the 5% level, but the p-value of the comparisons between different cell lines is, for at least some cell lines. Fig. 2 shows all the comparisons that we carried out to identify good shape parameters, organized by cell lines, and classes of shape parameters. It shows that some shape measures from each category are good 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.

We found that human-crafted morphological parameters mostly met our criteria for good shape parameters. Cell, nuclei, and convex hull geometric parameters pass both similarity and dissimilarity tests. These include measures connected with spread size, elongation and irregularity of the boundary of a cell. Gray scale measures are the best parameters which meet both similarity and dissimilarity criteria for almost all the cases. Roughness measures are good only for specific cell lines. In particular, breast tissue cells (normal or cancerous) cannot be easily distinguished by these measures, while osteosarcoma cells appear to be more easily distinguishable.

Fig. 2 also shows that basis function methods, i.e. Fourier and Zernike expansions, do not perform as well as the constructed features of the other parameters. However, both types of expansions pass the similarity test quite well for retina cell lines, and display differences in quite a few coefficients for the dissimilarity test for retina and osteosarcoma cells. Overall, the Fourier parameters do not meet the criteria for good measures 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 better for both similarity and dissimilarity test for

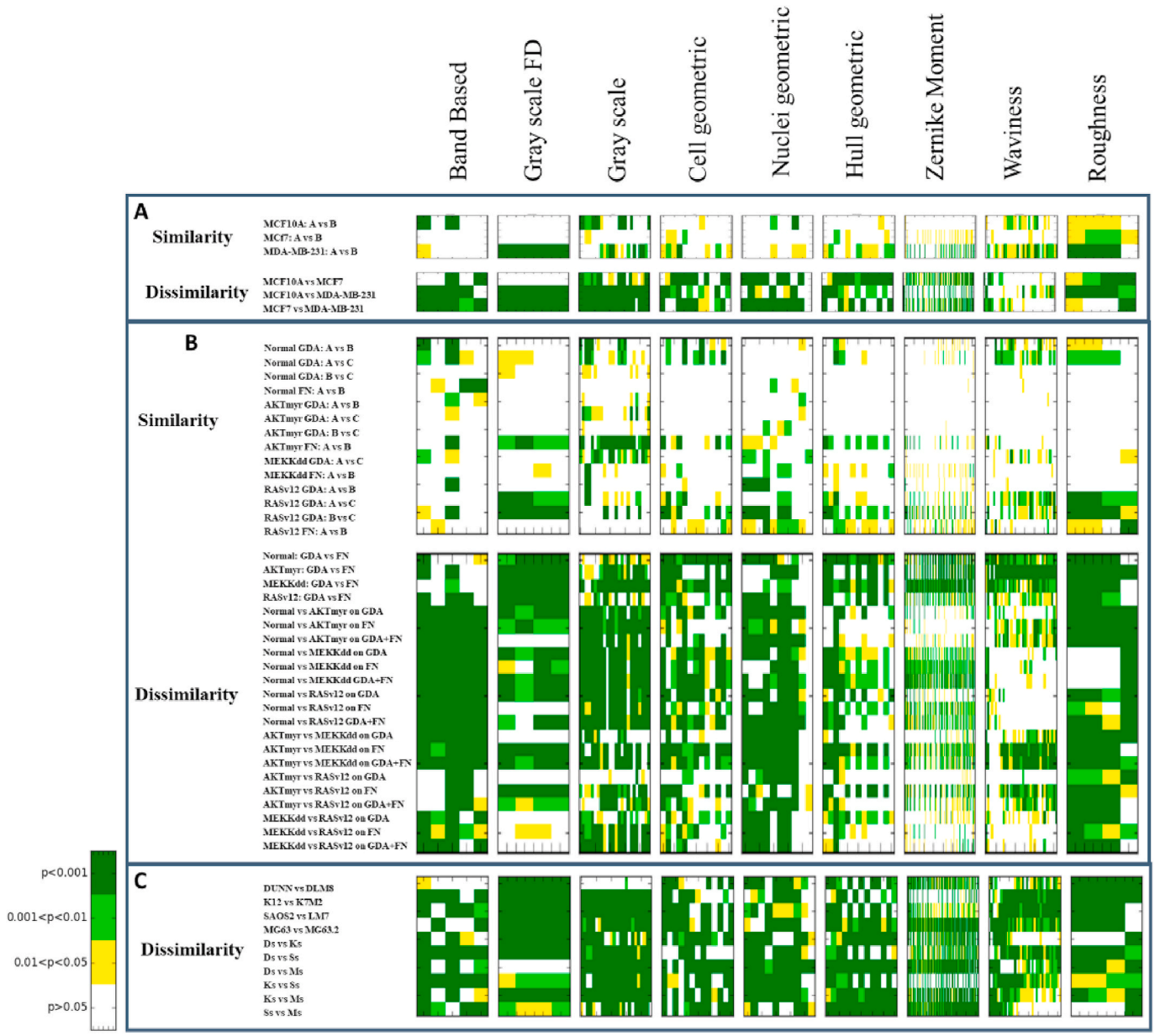


**Fig. 2. T-tests results to find "good" parameters.** Heat map plot of p-values for similarly tests, when replicates of the same cell line are compared, and dissimilarity tests, when replicates of cell lines are combined and compared against other lines. Each column in this plot represent one shape measure and each shape category is plotted separately. The lighter the color is the more similar the samples are for that shape measure. A good measure, as defined here, is mostly dark green in dissimilarity and white in similarity tests.

breast cells.

### *3.2. Low metastatic and high metastatic breast cancer cell morphologies partially overlap with normal breast epithelium cells but have distinguishing features*

In order to understand the major changes in morphological parameters, we classified our shape parameters into 9 subcategories, reduced the dimensionality of the morphospace by calculating the principal components of each of the 9 categories separately, and then picked one principal component (the Primary Principal Component or PPC) from each that was best for classification (see Methods). This method projects down the high-dimensional morpho-space into a 9-dimensional space, where each axis represents a shape category. Though still high dimensional, this method allows us to gain some insight into morphological differences, as can be seen from two-dimensional projections of the data in Fig. 3. This figure shows that shape changes between normal tissue, low metastatic cancer and high metastatic cancer is not a linear progression from normal to extreme, but in fact normal cells show significant overlap with both the types of cancer cells. Interestingly, in their geometric characteristics normal cells lie somewhere in between the two classes of cancer cells. However, the actin cytoskeleton does display progressive changes in structure, from normal to high metastatic, as suggested by the Grayscale and Band based measures (2nd panel).

If cell morphologies overlap significantly, how useful are morphological metrics for distinguishing between single cells? One test for determining whether cell morphologies are different at the single cell level is the accuracy of classification using standard classifiers such as ANNs or SVMs. In accordance with standard practice, these classifiers are trained on a subset of the data, and their performance assessed by a pure prediction on the remaining data. This process is then repeated on different divisions of the data. A high accuracy in the pure prediction

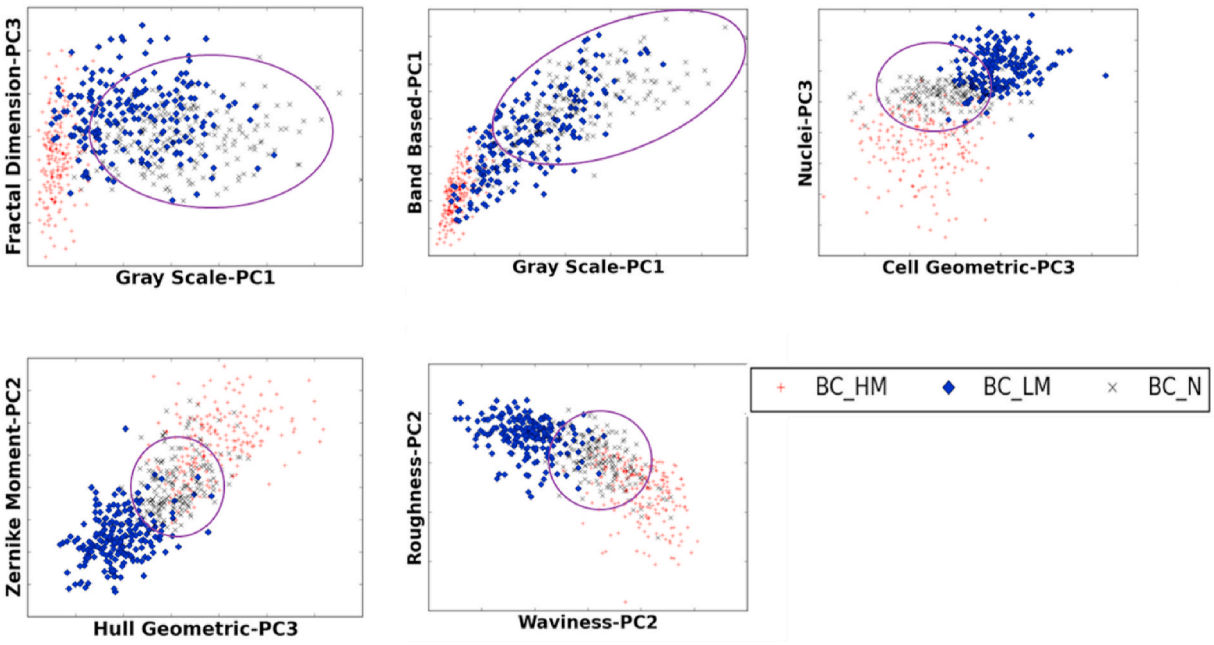


**Fig. 3. Scatter plot using the PPC for all the breast cell lines.** BC_HM (high metastatic): MDA-MB-231; BC_LM (low metastatic): MCF7; BC_N (normal): MCF10A. Depending on the shape category normal cells sometimes overlap with low metastatic cells and sometimes with high metastatic cells. The high and low metastatic cell lines overlap significantly less with each other. The purple ellipse is drawn by hand to enclose normal cells to guide the eye.

thus implies that the trained model is not just a good fit but it has predictive accuracy. In line with results from other cell lines, discussed below, we found that the crafted measures, were significantly more information-rich than the basis function-based quantities (see Methods for definitions of both). Thus, crafted measures could correctly identify 88% of a dataset consisting of both low metastatic and high metastatic breast cancer cells (Table 2, Table A1), while Zernike moments and Waviness measures could correctly identify 74% and 65.5% respectively. An accuracy of 88% at the single cell level translates to 100% accuracy in distinguishing small samples of cells from different classes from each other. In the remainder of the paper we have used these crafted measures for classifying cells using ANNs or SVMs, unless otherwise stated.

### 3.3. *Transformation of normal retina cells to cancerous lines lead to characteristic shape changes*

Stepwise cellular transformation into cancerous cells is a model system to study oncogenesis, and can help in leading to an understanding of key pathways that are perturbed in cancer. Using the transformed ARPE cells based on mutations of the Akt, Mek and Ras pathways allows us to ask whether oncogenic transformation leads to unambiguous shape changes, and whether there are specific morphological signatures of the genetic mutations in each case. We find that in PPC space we can distinguish between normal cells and transformed cells quite easily by eye (Fig. 4). However, the morphometrics of the transformed cells are highly overlapping, possibly because Ras and Mek are both part of the Ras-Erk MAPK pathway, which has strong crosstalk with the Akt-PI3K pathway. Significantly, the ANN and the SVM analysis shows that normal cells can be distinguished from transformed cells at the single cell level with an accuracy of between 90% and 94% for ANNs and similar numbers for SVMs on GDA substrate (Table 2, Table A1).

Fig. 4 also shows that in some projections the morphometrics of the three transformed lines are only partially overlapping, leading to the expectation that they are actually more separable in high-dimensional morphometric space. Both the ANNs and the SVMs bear this out. Using ANNs, $Akt^{myr}$ and $Mek^{DD}$ cells can be correctly classified at the single cell level with an accuracy of 87.5%, $Akt^{myr}$ and $Ras^{v12}$ with an accuracy of 85% and $Ras^{v12}$ and $Mek^{DD}$ with an accuracy of 81% (the corresponding SVM figures are 84.8%, 85.83% and 81.4%). This strongly suggests that cell morphology is a highly sensitive readout of cell state, and mutations that are quite close to each other and have

**Table 2**
**Classification accuracy (in percentages) using SVMs or ANNs.** The Best Sample Accuracy is a rounded-off estimate of the probability that a sample of 10 cells from either of the two categories would have more than 5 cells correctly classified as belonging to that category. The categories are explained in the Methods section.

| Classification Experiment | SVM Accuracy | Neural Network Accuracy | Best Sample Accuracy |
|---|---|---|---|
| Breast Cancer: Low metastatic vs high metastatic | 88.25 | 88.25 | 100 |
| Breast Cancer: Normal vs Low Metastatic | 90.5 | 93.25 | 100 |
| Breast Cancer: Normal vs High Metastatic | 88.0 | 88.75 | 100 |
| RPE: Control versus $AKT^{myr}$ | 90.33 | 92.83 | 100 |
| RPE: Control versus $MEK^{DD}$ | 91.75 | 90.4 | 100 |
| RPE: Control versus $RAS^{v12}$ | 93.83 | 94.83 | 100 |
| RPE: Control versus $AKT^{myr}$ on FN substrate | 84 | 86.75 | 99 |
| RPE: Control versus $MEK^{DD}$ on FN substrate | 85.5 | 86.5 | 100 |
| RPE: Control versus $RAS^{v12}$ on FN substrate | 91.5 | 88.25 | 100 |
| Osteosarcoma: MG63 vs MG63.2 | 89.0 | 92.46 | 100 |
| Osteosarcoma: DUNN vs DLM8 | 77.5 | 80.0 | 97 |
| Osteosarcoma: K12 vs K7M2 | 97.5 | 98.0 | 100 |
| Osteosarcoma: SAOS2 vs SAOS-LM7 | 93.5 | 93.0 | 100 |
| All low metastatic vs High Metastatic | 69.5 | 81.67 | 97 |
| All type 1 low vs high metastatic | 75.2 | 79.5 | 92 |
| All Normal vs low metastatic | 88.67 | 92.57 | 100 |
| All Normal vs type 1 high metastatic | 91.4 | 94.4 | 100 |

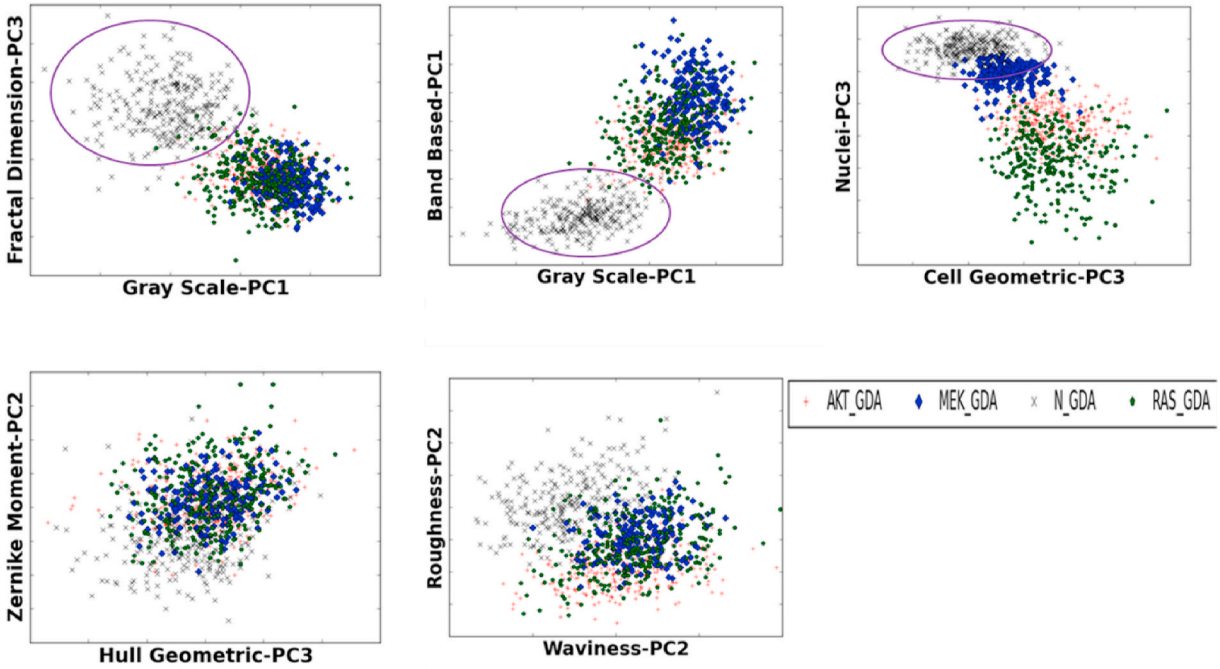


**Fig. 4. Scatter plot using the PPC for ARPE and transformed cells on glass substrate.** GDA: Detergent washed glass. AKT: Aktmyr; MEK: MEKDD, Ras: Rasv12 cells. Transformed cell lines overlap with each other but cluster separately from normal cells in the first three projections. The purple ellipse is drawn to enclose normal cells to guide the eye.

overlapping downstream consequences may nevertheless be distinguishable on the basis of morphology features alone.

### 3.4. Cells are accurately distinguishable on fibronectin substrates despite changes in parameters

Cell morphology is sensitive to the properties of the substrate that they are cultured on. In practical applications many cancer cells adhere poorly to glass, and require substrates coated with fibronectin or other such adhesion promoting proteins. We therefore tested whether the fibronectin coating changes our conclusions for the ARPE transformed cells, discussed above. As shown in Fig. A.3, the qualitative features of the shape parameters look similar to those in Fig. 4, i.e. there is visible separation between the normal cells and the transformed cells, but the transformed cells are in general highly overlapping. Each transformed cell line is distinguishable at the single cell level from the controls with an accuracy of 86.75%, 86.5% and 88% for $Akt^{myr}$, $Mek^{DD}$ and $Ras^{v12}$ respectively, using ANNs, which is marginally smaller than the same cell lines on uncoated glass substrates. Interestingly the morphology of the $Akt^{myr}$ cell line is closer to the other two on fibronectin, since the classification accuracy of ANNs drops to 84.75% and 76.25% when classifying sets with $Akt^{myr}$ and $Mek^{DD}$ or $Akt^{myr}$ and $Ras^{v12}$ respectively. The accuracy of the $Ras^{v12}$ and $Mek^{DD}$ classification only marginally decreases to 80%. It is interesting to speculate whether signaling through fibronectin brings $Akt^{myr}$ cells closer to the phenotype of the other two transformed cell lines.

### 3.5. Osteosarcoma cell lines of different metastatic potential are easily distinguished by morphology

We had previously shown [32,33] that paired osteosarcoma cell lines were easily distinguishable by morphology, based on a smaller set of shape features. We had also found that 3 out of 4 paired cell lines shared very similar patterns in their shape changes, but the fourth paired line, MG63 and MG63.2, displayed a very different trend. We describe these shape changes as Type 1 and Type 2 respectively. The larger set of shape features in the toolbox used in this work strengthens these conclusions. In Fig. 5, low metastatic cells of the 6 Type 1 cell lines are combined and plotted together. It can be seen that though there is some overlap, the two classes are separable by eye. It is interesting that in some of the projections shown in Fig. 5, the high metastatic cells are more homogenous than the low metastatic cells. However we did not find a consistent decrease in CV of the Geometric features at least between the low metastatic and high metastatic line for osteosarcomas, unlike Ref. [31] found for pancreatic cancer. For each pair of cell lines, the machine learning analysis is quite successful in separating cells based on morphological features. The most similar pair is the DUNN and DLM8 pairs, where cells were correctly classified with about 80% accuracy, while it was 98% for the other murine pair, K12 vs K7M2 and around 93% for both of the human pairs. The classification accuracy using ANNs between all high and low metastatic cell lines was only about 81.7%, and between all Type 1 high and low metastatic osteosarcomas, it was 79.5% (but with a 40-node network).

### 3.6. Cancer progression is accompanied by similar shape changes for cancer types studied

Next, we asked how similar morphology changes across all cell lines were, between normal and transformed cells as well as low metastatic and high metastatic cells. For this purpose, we calculated the changes in the mean of first principal component for each comparable pair of lines comprising of either a low and high metastatic set or a normal and transformed set, separately for all 9 shape categories. The list of the comparisons are shown pictorially in Fig. 6. This figure is color-coded to identify significant changes in PC1 (p-value$<$0.001 in order to account for multiple comparisons), and whether the mean value increases or decreases across the comparison. Note that an increase in the principal component does not imply an increase in a feature. However, changes in PC1 between each pair can be compared with other pairs, i.e. if PC1 increases for both it implies that their shape changes similarly. If they are in opposite direction, it means their shape changes are not similar.

The most striking feature of Fig. 6 is the comparison of the shape changes between the MG63 and MG63.2 lines with all of the others. The changes in PC1 for this comparison was not significant for Gray scale,

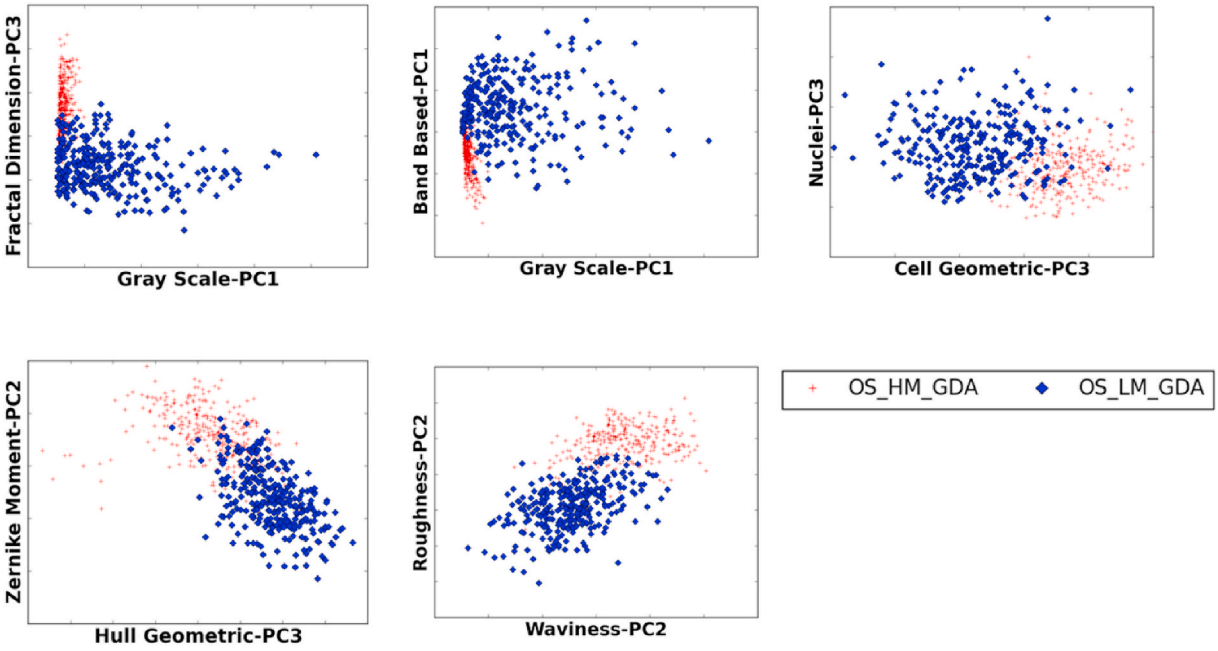


**Fig. 5. Scatter plot using the PPC for Type 1 osteosarcoma cells.** GDA: Detergent washed glass; OS_HM: High metastatic and OS_LM: Low metastatic osteosarcomas. Low metastatic and high metastatic cells cluster separately. In some projections high metastatic morphologies are more homogenous than those of low metastatic cell lines.

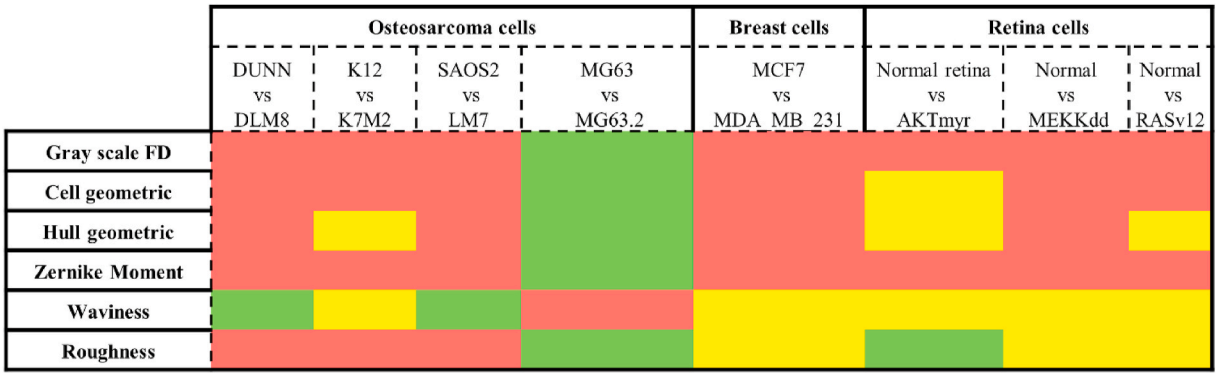


**Fig. 6. The direction of changes in PC1 with cancer progression for different categories.** Only those categories for which changes in PC1 for MG63 vs MG63.2 cell lines were significant (p-value<0.001) are shown. Yellow boxes imply changes in PC1 were not significant (p-value>0.001) for that comparison. Green boxes signify an increase in PC1 and red boxes signify a decrease in PC1. The figure suggests that morphology differences across most comparisons are similar to each other and different from the MG63/MG63.2 differences.

band based, and nuclei measures. Of all the 6 remaining shape categories other than Roughness, significant shape changes for all the other comparisons were exactly opposite of shape changes seen for M cell line. For the Roughness measures, the shape changes between ARPE cell line to $AKT^{myr}$ cancer line was the only comparison which followed similar shape changes as the MG63/MG63.2 lines.

These results suggest that shape changes accompanied with cancer progression from normal to transformed, and from low metastatic cancers to high metastatic cancers in these 15 cell lines follows a similar pattern, regardless of type of the measures used to quantify the shape. This pattern is very different from that exhibited by the MG63/MG63.2 cells, which we call Type 2 changes. 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 characteristics observed [34].

Finally, we asked whether shape differences between cell lines are sufficiently similar for the ANN or SVM classifiers to classify cells accurately when the cell lines are combined together in broad classes. In Table 2 we combined different classes of cells together, and we find that the ANN can classify single cells belonging to all cancers into the low metastatic and high metastatic category with an accuracy of 81.67%. Here we have assumed that the transformed ARPE cells belong to the low metastatic category. Since the morphology differences in these cells resemble those that we call Type 1 differences, we clubbed all Type 1 cells together and found that classification accuracy was lower at about 79.5%. However, the differences between normal cells and cancer cells were much more marked. The ANN could accurately classify cells into either the normal and low metastatic class or the normal and high metastatic class with an accuracy of 92.57% and 94.4% respectively. For these comparisons a linear SVM did not uniformly perform as well as an ANN. It should be noted that for small samples of cells, a 75% accuracy at the level of a single cell, translates into an accuracy of almost 100% in correctly identifying a sample of 30 cells from either of the two populations.

## 4. Conclusion

This work is based on a comparison of morphological parameters of 15 cell lines with different properties. Together they represent a number of major biological transformations. One set of osteosarcoma lines are divided into 4 paired sets, consisting of a low metastatic parental line and a high metastatic derived line. A second set is a commonly used triplet of cell lines used to study breast cancer, and its metastasis. A third set of lines represent laboratory guided oncogenic transformations and are a model system to study key oncogenes. These cell lines thus afforded us an opportunity to ask a number of questions about the relation between cell morphology and cell state [15].

Cell morphology as defined here included parameters of the two-dimensional outline of cells on a surface, as well as the measures of actin structure extracted from labelled actin in fixed cells [36]. We use a large number of shape parameters, including what we call "crafted measures" that are quantitative measures of some qualitative property of morphology, as well as more mathematically complete basis function expansions. While we can reconstruct the outline of the cell from the latter with excellent accuracy, we find that the crafted measures are actually more information rich, and are able to correctly identify the class membership of most single cells. In particular we find that at the single cell level we can classify high metastatic cells of either osteosarcomas or breast cancers from their low metastatic counterparts with high accuracy. Moreover, even when clubbed together, these classes are sufficiently dissimilar that we can identify them reasonably well. This raises the questions of whether there are a limited number of stereotypical shape changes accompanying metastatic transformation. Our previous work [33], as well as the current work, has suggested that metastatic transformation may fall into only two classes of morphology changes. However, many more cancer cell types need to be studied before we can be sure of this.

Our analysis of oncogenic transformation shows that there are clear morphological signatures of cancerous transformation. More excitingly, the three oncogenes studied appear to be distinguishable from each other by morphology alone. This is somewhat surprising since all three oncogenes operate in closely related pathways. In many projections from high-dimensional morpho-space the parameters of these cell lines were strongly overlapping. Nevertheless, in the higher-dimensional morpho-space, SVMs and ANNs could correctly identify most cells by morphology alone. These results indicate that major mutational signatures of patient cancers may be detectable from advanced versions of these methods applied to biopsy images.

These results also underscore the fact that it is impossible to get the full measure of a high dimensional data set by projections down to two dimensions and we require an intrinsically high dimension analysis method for that. We used machine learning methods that have become very standard by now: Linear Support Vector Machines and Artificial Neural Nets with one hidden layer. This is well behind the current state of the art in machine learning. However, our intent was not to optimize machine learning classification methods, but to gain insight with limited data. That is the main reason for our choice of methods. There is significant scope for more advanced methods along with much larger datasets, especially for applications in diagnostics. However, we will not be able to fully interpret and make use of the machine learning analysis without an advance in mechanistic insight. While there have been important breakthroughs in cancer diagnostics using conventional methods such as genetic testing (for breast cancer), no single method is likely to be a panacea, and the key is going to be integration of multiple pieces of information. Optimizing the information that can be extracted from already existing methods of diagnosis could therefore be a useful tool to improve accuracy in the prediction of the tumor properties.

## 5. Discussion

This work is part of the growing body of literature that utilizes high dimensional morphological parameters to make inferences about cell state, which has been recently reviewed [15,16]. Early papers showed linkages between cell shape and signaling, and identified the role of Rho GTPase proteins in cell morphology in Drosophila cell lines [23,25,28]. More recently, a few papers have applied morphological analysis to classify metastatic cells in specific cancers. Ren et al. [42] used 29 morphological parameters to classify shape changes in A549 lung adenocarcinoma cell lines after stimulation by TGBβ, which promotes metastasis transformation through the Epithelial to Mesenchymal transformation (EMT) [12]. Laboratory induced EMT is known to induce dramatic morphological changes, also studied in transformed MCF10-A cells using high content imaging by Leggett et al. [43]. In both cases only a very small set of parameters were needed since induced-EMT leads to very dramatic morphological changes that are easily distinguishable by eye. In contrast we study a larger set of subtler and more complex changes, whose morphological signatures are not easily distinguishable by eye, using a larger set of parameters. Wu et al. [31] do a large study of cancer metastasis of about 39,000 cells coming from 13 patient derived tumor samples of pancreatic ductal adenocarcinoma (PDAC). They used a low magnification technique and calculated shape descriptors based on a description of the boundary coordinates of the cell and nucleus that were then projected to PCA space. They did not find a specific morphological signature of metastatic cells but found that they displayed greater heterogeneity. They did not study textural parameters that we did in this study. A very recent study by the same group studied a large number of MDA-MB-231 breast cancer cells [44], using a different but more comprehensive set of morphological parameters of the cell and nucleus boundary, as well as parameters related to nucleus-cell positioning. They showed that even a population of cells coming from one cell line showed significant heterogeneity, and, in agreement with our results, single cell morphology could be used to identify metastasis. While many of the parameters they use overlap with ours in this paper, they did not use textural parameters that we found to be very useful in characterizing cells and did not study oncogenic transformation, or sarcoma cells. They also used only a single cell line. However, their results are very significant because they showed that the morphological variation within a single cell line may have functional consequences.

All of these studies reinforce our central message that significantly more information could be extracted from cell morphology than is currently normal practice. We should point out here similar developments have taken place in other fields of biomedical imaging such as radiomic analysis [45] where multi-dimensional morphological metrics of MRI or other biomedical images have been successfully used along with big data techniques to make predictive models of cancer properties and patient outcome [46,47]. Cell morphology reflects cell state because of the close link of the cellular cytoskeleton and of cellular adhesion with cell physiology. Cell shape is dynamic, but it is a consequence of the balance of forces generated by the cytoskeleton, the mechanical properties of the cell, the properties of the extracellular matrix and that of the adhesion molecules. For non-motile cells therefore, it reflects a long-lived cytoskeletal state, which in turn reflects an internal physiological state. While the mapping between cell state and morphology is likely to be many-to-one, our data, and that of others, show that there is sufficient granularity in that mapping for that knowledge to be extremely useful to us.

A deeper understanding of cell morphology requires studies, both modeling and experiments, that are more mechanistic and that are performed at the single cell level, that are well beyond the scope of this one. Our results, and those of others, suggest that some aspects of morphological heterogeneity within a population of cells reflect phenotypic differences in the cells, and that cellular morphology may give us unprecedented insight, when fully understood, into cellular heterogeneity. Unlike other single cell techniques, this has the virtue of being relatively easier (just requiring imaging) and in principle can be carried out dynamically. In fact, it is very possible that using only live-cell phase contrast imaging, along with cellular morphometrics coupled

with data analysis and algorithms, we will be able to obtain a dynamic noninvasive real-time window into cell state.

## Author contributions

A.P. conceived of the idea and A.P. and E.A. planned the work for this publication; E.A., J.C., A.Q. and C. D.L.T. carried out the experiments and the image processing; E.A. and W.X. wrote the imaging processing code and the code to calculate the morphometrics; A.P. and E.A. analyzed the data and wrote the paper. All authors read and approved the final version.

## Declaration of competing interest

None Declared.

## Acknowledgements

This work was supported by a National Science Foundation (USA) CAREER grant PHY- 1151454 to A.P.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2020.104044.

## References

[1] CDC - national center for health statistics https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm December 15, 2018 2018.
[2] Network GBoDC, Global Burden of Disease Study, Institute for Health Metrics and Evaluation (IHME), Seattle, 2017.
[3] R. Siegel, J. Ma, Z. Zou, A. Jemal, Cancer statistics, Ca - Cancer J. Clin. 64 (1) (2014) 9–29, 2014.
[4] Z. Yin, H. Sailem, J. Sero, R. Ardy, S.T. Wong, C. Bakal, How cells explore shape space: a quantitative statistical perspective of cellular morphogenesis, Bioessays 36 (12) (2014) 1195–1203.
[5] J.D. Schiffman, P.G. Fisher, P. Gibbs, Early detection of cancer: past, present, and future, Am Soc Clin Oncol Educ Book (2015) 57–65.
[6] J.L. Carter, R.J. Coletti, R.P. Harris, Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods, BMJ 350 (2015) g7773.
[7] I. Damjanov, F. Fan, Cancer Grading Manual, Springer, 2007.
[8] R. Engers, Reproducibility and reliability of tumor grading in urological neoplasms, World J. Urol. 25 (6) (2007) 595–605.
[9] D. Eefting, Y.M. Schrage, M.J. Geirnaerdt, S. Le Cessie, A.H. Taminiau, J.V. Bovee, et al., Assessment of interobserver variability and histologic parameters to improve reliability in classification and grading of central cartilaginous tumors, Am. J. Surg. Pathol. 33 (1) (2009) 50–57.
[10] K. Tanaka, T. Hasegawa, T. Nojima, Y. Oda, J. Mizusawa, H. Fukuda, et al., Prospective evaluation of Ki-67 system in histological grading of soft tissue sarcomas in the Japan Clinical Oncology Group Study JCOG0304, World J. Surg. Oncol. 14 (2016) 110.
[11] M.J. van de Vijver, Y.D. He, L.J. van't Veer, H. Dai, A.A. Hart, D.W. Voskuil, et al., A gene-expression signature as a predictor of survival in breast cancer, N. Engl. J. Med. 347 (25) (2002) 1999–2009.
[12] C.L. Chaffer, R.A. Weinberg, A perspective on cancer cell metastasis, Science 331 (6024) (2011) 1559–1564.
[13] R. Kalluri, R.A. Weinberg, The basics of epithelial-mesenchymal transition, J. Clin. Invest. 119 (6) (2009) 1420–1428.
[14] S.M. Shaffer, M.C. Dunagin, S.R. Torborg, E.A. Torre, B. Emert, C. Krepler, et al., Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance, Nature 546 (7658) (2017) 431–435.
[15] A. Prasad, E. Alizadeh, Cell form and function: interpreting and controlling the shape of adherent cells, Trends Biotechnol. 37 (4) (2019) 347–357, https://doi.org/10.1016/j.tibtech.2018.09.007.
[16] R.A. Marklein, J. Lam, M. Guvendiren, K.E. Sung, S.R. Bauer, Functionally-relevant morphological profiling: a tool to assess cellular heterogeneity, Trends Biotechnol. 36 (1) (2018) 105–118.
[17] R. McBeath, D.M. Pirone, C.M. Nelson, K. Bhadriraju, C.S. Chen, Cell shape, cytoskeletal tension, and RhoA regulate stem cell lineage commitment, Dev. Cell 6 (4) (2004) 483–495.
[18] C. Chen, J. Alonso, E. Ostuni, G. Whitesides, D. Ingber, Cell shape provides global control of focal adhesion assembly, Biochem. Biophys. Res. Commun. 307 (2) (2003) 355–361.
[19] L.E. Dike, C.S. Chen, M. Mrksich, J. Tien, G.M. Whitesides, D.E. Ingber, Geometric control of switching between growth, apoptosis, and differentiation during angiogenesis using micropatterned substrates, In Vitro Cell. Dev. Biol. Anim. 35 (8) (1999) 441–448.
[20] C.S. Chen, M. Mrksich, S. Huang, G.M. Whitesides, D.E. Ingber, Geometric control of cell life and death, Science 276 (5317) (1997) 1425–1428.
[21] D. Zhang, K.A. Kilian, The effect of mesenchymal stem cell shape on the maintenance of multipotency, Biomaterials 34 (16) (2013) 3962–3969.
[22] K.A. Kilian, B. Bugarija, B.T. Lahn, M. Mrksich, Geometric cues for directing the differentiation of mesenchymal stem cells, Proc. Natl. Acad. Sci. U.S.A. 107 (11) (2010) 4872–4877.
[23] Z. Yin, A. Sadok, H. Sailem, A. McCarthy, X. Xia, F. Li, et al., A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes, Nat. Cell Biol. 15 (7) (2013) 860–871.
[24] P. Pascual-Vargas, S. Cooper, J. Sero, V. Bousgouni, M. Arias-Garcia, C. Bakal, RNAi screens for Rho GTPase regulators of cell shape and YAP/TAZ localisation in triple negative breast cancer, Sci Data 4 (2017) 170018.
[25] C. Bakal, J. Aach, G. Church, N. Perrimon, Quantitative morphological signatures define local signaling networks regulating cell morphology, Science 316 (5832) (2007) 1753–1756.
[26] J.E. Sero, C. Bakal, Multiparametric analysis of cell shape demonstrates that β-PIX directly couples YAP activation to extracellular matrix adhesion, Cell Systems 4 (2017) 84–96.e6.
[27] H.Z. Sailem, C. Bakal, Identification of clinically predictive metagenes that encode components of a network coupling cell shape to transcription by image-omics, Genome Res. 27 (2017) 196–207.
[28] O. Nir, C. Bakal, N. Perrimon, B. Berger, Inference of RhoGAP/GTPase regulation using single-cell morphological data from a combinatorial RNAi screen, Genome Res. 20 (3) (2010) 372–380.
[29] A. Pasqualato, V. Lei, A. Cucina, S. Dinicola, F. D'Anselmi, S. Proietti, et al., Shape in migration: quantitative image analysis of migrating chemoresistant HCT-8 colon cancer cells, Cell Adhes. Migrat. 7 (5) (2013) 450–459.
[30] A. Pasqualato, A. Palombo, A. Cucina, M.A. Mariggio, L. Galli, D. Passaro, et al., Quantitative shape analysis of chemoresistant colon cancer cells: correlation between morphotype and phenotype, Exp. Cell Res. 318 (7) (2012) 835–846.
[31] P.-H. Wu, J.M. Phillip, S.B. Khatau, W.-C. Chen, J. Stirman, S. Rosseel, et al., Evolution of cellular morpho-phenotypes in cancer metastasis, Sci. Rep. 5 (2016) 18437.
[32] E. Alizadeh, S.M. Lyons, J.M. Castle, A. Prasad, Measuring systematic changes in invasive cancer cell shape using Zernike moments, Integr Biol (Camb). (2016).
[33] S.M. Lyons, E. Alizadeh, J. Mannheimer, K. Schuamberg, J. Castle, B. Schroder, et al., Changes in cell shape are correlated with metastatic potential in murine and human osteosarcomas, Biol Open 5 (3) (2016) 289–299.
[34] J. Ahn, V. Sanz-Moreno, C.J. Marshall, The metastasis gene NEDD9 product acts through integrin beta 3 and Src to promote mesenchymal motility and inhibit amoeboid motility, J. Cell Sci. 125 (Pt 7) (2012) 1814–1826.
[35] R.M. Neve, K. Chin, J. Fridlyand, J. Yeh, F.L. Baehner, T. Fevr, et al., A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes, Canc. Cell 10 (6) (2006) 515–527.
[36] E. Alizadeh, W. Xu, J. Castle, J. Foss, A. Prasad, TISMorph: a tool to quantify texture, irregularity and spreading of single cells, PloS One 14 (6) (2019), e0217346.
[37] H.D. Soule, T.M. Maloney, S.R. Wolman, W.D. Peterson Jr., R. Brenz, C. M. McGrath, et al., Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10, Canc. Res. 50 (18) (1990) 6075–6086.
[38] S. Comsa, A.M. Cimpean, M. Raica, The story of MCF-7 breast cancer cell line: 40 years of experience in research, Anticancer Res. 35 (6) (2015) 3147–3154.
[39] A.J. Minn, G.P. Gupta, P.M. Siegel, P.D. Bos, W. Shu, D.D. Giri, et al., Genes that mediate breast cancer metastasis to lung, Nature 436 (7050) (2005) 518–524.
[40] Y. Kang, P.M. Siegel, W. Shu, M. Drobnjak, S.M. Kakonen, C. Cordon-Cardo, et al., A multigenic program mediating breast cancer metastasis to bone, Canc. Cell 3 (6) (2003) 537–549.
[41] C. McQuin, A. Goodman, V. Chernyshev, L. Kamentsky, B.A. Cimini, K.W. Karhohs, et al., CellProfiler 3.0: next-generation image processing for biology, PLoS Biol. 16 (7) (2018), e2005970.
[42] Z.X. Ren, H.B. Yu, J.S. Li, J.L. Shen, W.S. Du, Suitable parameter choice on quantitative morphology of A549 cell in epithelial-mesenchymal transition, Biosci. Rep. 35 (3) (2015).
[43] S.E. Leggett, J.Y. Sim, J.E. Rubins, Z.J. Neronha, E.K. Williams, I.Y. Wong, Morphological single cell profiling of the epithelial-mesenchymal transition, Integr Biol (Camb). 8 (11) (2016) 1133–1144.
[44] P.-H. Wu, D.M. Gilkes, J.M. Phillip, A. Narkar, T.W.-T. Cheng, J. Marchand, et al., Single-cell morphology encodes metastatic potential, Science Advances 6 (4) (2020), eaaw6938.
[45] S. Rizzo, F. Botta, S. Raimondi, D. Origgi, C. Fanciullo, A.G. Morganti, et al., Radiomics: the facts and the challenges of image analysis, Eur Radiol Exp 2 (1) (2018) 36.
[46] A. Chaddad, C. Desrosiers, B. Abdulkarim, T. Niazi, Predicting the gene status and survival outcome of lower grade glioma patients with multimodal MRI features, IEEE Access 7 (2019) 75976–75984.
[47] A. Chaddad, P. Daniel, S. Sabri, C. Desrosiers, B. Abdulkarim, Integration of radiomic and multi-omic analyses predicts survival of newly diagnosed IDH1 wild-type glioblastoma, Cancers 11 (8) (2019).