A robust method to quantify cell morphological heterogeneity

Jude M. Phillip[1, 5,*], Kyu-Sang Han[1,*], Wei-Chiang Chen[1], Denis Wirtz[1-4, +], Pei-Hsun Wu[1, +]

[1] Department of Chemical and Biomolecular Engineering, Johns Hopkins Physical Sciences Oncology Center, Johns Hopkins Institute for Nanobiotechnology (INBT), Johns Hopkins University, Baltimore, Maryland 21218, USA; [2] Department of Pathology, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA; [3] Department of Oncology, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA; [4] Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA; [5] Department of Medicine, Division of Hematology and Oncology, Weill Cornell Medicine, New York, New York, 10065, USA;

[*] These authors contributed equally to this work;

[+]Correspondence should be addressed to: P-H. W pwu@jhu.edu and D.W. wirtz@jhu.edu

Keywords: cell morphology; single-cell phenotyping; shape modes; cellular heterogeneity; entropy

Abstract

Cell morphology encodes essential information on many underlying biological processes. It is commonly employed by clinicians and researchers in the study, diagnosis, prognosis, and treatment of human diseases. Quantification of cell morphology has seen tremendous advances in recent years. However, effectively defining morphological shapes and evaluating the extent of morphological heterogeneity within a cell population remain challenges. Here we present a protocol and software for the analysis of cell morphology data using the VAMPIRE algorithm. This algorithm enables cell profiling through the classification of cells into shape modes based on equidistant points along cell and nuclear contours. Examining the distributions of cell morphologies across automatically identified shape modes provides an effective visualization scheme that relates cell shapes to cellular subtypes, which are defined by endogenous and exogenous cellular conditions. In addition, these shape mode distributions offer a direct and quantitative way to measure the extent of morphological heterogeneity within cell populations. This protocol is highly automated and fast, taking less than 10 min to analyze 10,000 cells.

Introduction

Cell morphology is commonly employed by clinicians and researchers in the study, diagnosis, prognosis, and treatment of human diseases. Fundamentally, cellular morphology represent the ensemble imprints of highly interactive molecular networks, including metabolic, proteomic, epigenomic and genomic components[1–6]. The coordinated orchestration of these inter-dependent cellular programs are critical to properly govern cellular behavior[4] and ultimately determine cell responses to perturbations and stressors, such as microenvironmental cues, biomechanical stimuli[7,8], and pharmacological treatments[9–11]. Advances in high-content imaging[6,12,13], image processing[12,14,15], and machine learning[16–18] have greatly improved the throughput and accuracy of cell morphological measurements and have bolstered its utility in digital pathology[19–22], biomarker identification[1,23], and phenotypic screens[10,24–26].

Cell morphology is traditionally quantified using a handful of geometric parameters[12,27], delineating the size (e.g. area, perimeter) and shape (e.g. shape factor, aspect ratio) of cells and their corresponding nuclei. These measures are often complemented by measurements of protein expression based on fluorescent labels and their intensity patterns and localization within the cell. Measuring cell and nuclear sizes can be readily achieved using open-source segmentation algorithms, such as CellProfiler[28] and ImageJ/Fiji[29,30].

However, defining and quantifying cellular shapes are more complicated. Classically, shape descriptors, such as shape factor ($4\pi A/P^2$, where A is the surface area of the morphology and P its perimeter), aspect ratio (long axis length/short axis length) and eccentricity (see **glossary**), all measure the deviation of a cell's shape from a circle. While these geometric parameters are geared towards biological simplicity and provide the ability to quickly and directly detect differences among tested cell conditions, these parameters tend to insufficiently capture the complexities of cell shapes[1].

To illustrate this issue, we described the morphologies of mouse embryonic fibroblasts (MEFs) using conventional shape features, including shape factor, solidity, and aspect ratio. Taking a subset of cells which have highly similar values of these parameters, we still observed a high degree of morphological variability among individual cells, suggesting that conventional cell morphology parameters may be insufficient to capture

cellular differences (**Figure 1**). Furthermore, mesenchymal cells on flat substrates, or cells embedded inside more physiological 3D collagen gels, which often feature extensive dendritic protrusions and nuclear blebs[34,36–38] are similarly difficult to distinguish using these traditional parameters.

A popular approach to address this shortcoming consists in defining additional geometric descriptors of cell morphology, some of which are based on the curvature and roughness of the cell and nuclear contours[12,27]. This has led to an expansion of morphological descriptors, with the premise that these additional descriptors would help to better define and differentiate cellular subtypes. While increasing the number of shape descriptors allow users to capture more complex cell morphologies, effective visualization of differences in cell morphology, and assigning biological meaning for these additional morphology descriptors are challenging.

To address this challenge, we recently developed a cell morphology analysis software to provide improved visualization and quantitative analysis of complex morphological changes in cells. The software, which we named Visually Aided Morpho-Phenotyping Image Recognition (VAMPIRE), is highly automated and allows to process large datasets rapidly.

Development of the protocol

VAMPIRE analysis was initially developed to better interpret morphological data that we acquired for a set of 11 pancreatic cancer cell lines using a custom high-throughput microscopy imaging system[1]. Our goal was to identify a potential morphological signature of metastasis in pancreatic ductal adenocarcinoma (PDAC). Among the samples used, five were collected from patient-derived primary tumors, four were obtained from liver metastatic (LM; four lines), and two were non-neoplastic pancreatic epithelial cell lines. For direct visual assessment of cell and nuclear shapes, we randomly selected subsets of individual cells traces after alignment and found no overt morphological differences between primary tumor cells and liver-metastasis cells, due partly to the irregularity of cell shapes.

To measure cell shapes, we examined commonly used morphological features, such as spreading area, shape factor, and aspect ratio. These features could not reflect the observed extent of cell shape variations, since even a small subset of cells displaying an

extremely narrow range of values of these conventional shape descriptors appeared radically different from each other.

To address this problem, we established and validated VAMPIRE analysis, which provides morphological information beyond classically defined geometric parameters[1,23]. VAMPIRE analysis is a visual aid that compares cell morphologies by first identifying representative shape modes (see **Glossary**) among all cell shapes present within a cell population, then determines the abundance of cells classified within each shape mode. VAMPIRE comprises four essential computational steps (**Figure 2**): I) the determination and registration of the coordinates of equally-spaced points along nuclear and cellular contours to define morphological descriptors; II) the reduction in the number of morphological descriptors using principal component analysis (PCA); III) the identification of shape modes through unsupervised clustering analysis, and IV) the analysis of shape mode distributions of all tested cell samples.

To represent highly complex shapes of cells and nuclei, a sufficient number of equally spaced points along their boundaries or contours (typically 50, **Figure 2A**) are used to define high-dimensional "features". These coordinates of contour points are normalized to unify cell sizes and aligned to eliminate any effects of rotational variations among cells (**Figure 2B**). After the cell contours are normalized and registered, eigen shape vectors from the PCA (see **glossary**) that comprise 95% of the observed variance are computed and used as reduced features for cell and/or nuclei shapes (**Figure 2C**). Next, representative shape modes are empirically identified from the populations of cells by applying K-means clustering algorithm[39] to the reduced shape features determined from PCA (**Figure 2D-E**), then each cell and/or nucleus is binned into a shape mode. The centroid locations of each cluster in the PCA-reduced features from K-means clustering analysis can be directly used to reconstruct and visualize the morphology of each shape mode. The shape modes are then used to classify the shapes of individual cells and occurrences of cells in different shape modes per condition provides both quantitative and visual insights of cell morphology. Among several classification methods such as DBscan, OPTICS, Meanshift, and K-means, we chose to use K-means clustering algorithm for its fast calculation and simplicity in setting the parameters.

In the study of pancreatic cancer cells (see above), VAMPIRE analysis showed that metastasized cells present significantly lower heterogeneity than primary tumor cells. We also found a similar morphological signature of metastasis for a cohort of 10 breast cancer cell lines[1]. We further deciphered the relative contributions to heterogeneity from cell cycle, cell-cell contacts, cell stochasticity and heritable morphological variations. In a separate study, the utility of VAMPIRE analysis was further demonstrated by investigating the morphological signature of healthy aging using skin dermal fibroblast cells[22]. We found that cellular age could be used to stratify individuals on the basis of cell morphology using a cohort of 32 samples of primary dermal fibroblasts collected from individuals between 2 and 96 years of age.

Overview of the protocol

In this protocol, for the accessibility and easiness to perform a VAMPIRE shape analysis, we established a python-based graphic user interface (GUI), VAMPIRE GUI. We note that VAMPIRE GUI does not provide a segmentation tool; it analyzes cell and/or nuclear shapes that are already detected and segmented. The segmentation can be performed using software such as ImageJ/Fiji [29,30] and CellProfiler[28], with easy integration of the segmentation results into VAMPIRE GUI.

To help the user explore the software and all its functionalities, we provide two small image datasets in the **Supplementary Data** under"example fluorescence images"**.** See the directory of supplementary data in S1 section of "README.pdf" in the **Supplementary Data** to locate any provided example images and workflow in this **Overview of the protocol** section. The example outputs of VAMPIRE analysis using provided image datasets are also included in "README.pdf" and **Supplementary Data** under "Example output". Before applying VAMPIRE analysis to their own data, users should first run VAMPIRE analysis using provided image datasets following the detailed procedure provided in the **Procedure** section. We also illustrate the power of VAMPIRE analysis by analyzing the morphology of mouse embryonic fibroblasts (MEFs) confined to adhesive micropatterns (akin to spatial restriction of cells in tissue) in the presence and absence of nuclear protein Lamin A/C, and for dermal fibroblasts derived from healthy individuals of increasing age (see **Anticipated results**).

The overall protocol is composed of four main parts: segmentation of images, formatting segmentation data, generating a VAMPIRE model, and applying the VAMPIRE model. The procedure starts with segmentation of fluorescent images of cells to generate greyscale images of segmented cells (**Step 1**). These segmented images can be obtained using standard segmentation tools, including but not limited to CellProfiler and ImageJ. In the example workflow/data, we use CellProfiler to segment the fluorescent images of cells.

To import segmented cells to VAMPIRE, the segmented images need to be organized in a specified format for use in the VAMPIRE GUI (**Step 2**). The segmented images must be either 8-bit or 16-bit binary images and image pixels with non-zero integer value representing the detected cells area. This required format is standard in most segmentation software. Once segmented images are imported into the VAMPIRE GUI, it reads the images to obtain coordinates of contour points, and a subset of other morphological parameters.

Once the dataset to be analyzed by VAMPIRE is segmented and properly organized, the user decides the set of images to be used to train a VAMPIRE model by specifying them in a comma separated values (CSV) file (**Step 3**). We refer these specified images as "training set" hereafter. An example CSV file of this list, "segmented image sets to build model.csv", can be found in **Supplementary Data**. The VAMPIRE model built based on the training set will be saved in a designated local folder. (**Step 4-10**). The model can then be applied to a new set of segmented images by specifying them in a new CSV file (**Step 11-13**). We note that VAMPIRE analysis can also be done by training the model and applying the same model to the training set.

The output of the VAMPIRE model includes a plot showing the frequency distribution of each shape mode per condition, the CSV files that contain the shape mode information for individual cells. (**Step 13**). Specifically, data for each cell include: the "xy" coordinates of cell centroids within the image, the spread area, circularity, aspect ratio, and assigned shape mode index (IDX), as well as the goodness of the shape mode classification for each cell that we refer to as "contour fit" (see **Glossary**). The datasheet can be directly linked to the features generated by CellProfiler, which makes VAMPIRE and CellProfiler analyses complementary. Example datasheets showing the results of the analysis using both platforms are provided in **Supplementary Data**, labeled "CellProfiler datasheet c1.csv" and "VAMPIRE datasheet c1.csv".

For all the procedures, approximate timing is indicated. This time corresponds to the time it takes for an experienced user. More time may be required when using VAMPIRE for the first time.

Applications of VAMPIRE

We have previously demonstrated the utility of VAMPIRE with two studies, (a) the morphological changes displayed by human pancreatic cancer cells as they spread from the primary tumor to the liver[1], and (b) the morphological changes that dermal fibroblasts derived from individuals undergo during healthy aging[23].

In general, VAMPIRE can be applied to any set of segmented images of cells or nuclei to detect and analyze changes in their morphology across multiple conditions in cell-culture model systems. For instance, VAMPIRE can be applied to the study of cell morphology changes in response to changes in cell cycle state and genetic and epigenetic status[1–3,7,23,36–38]. VAMPIRE analysis is similarly applicable to other types of perturbations, including mechanical perturbations[40,41]. VAMPIRE analysis is also suitable for drug screening [9,10,13]. Changes in cell morphology are often used in high-throughput biochemical and discovery screens[42]. The large volume of data that is typically generated in such screens makes it difficult to visually inspect cell responses. Here, the application of VAMPIRE provides users with the ability to rapidly classify phenotypically distinct cellular conditions in large amount of data and identify drug-induced changes in the frequency distributions of shape modes.

VAMPIRE analysis can be applied to the cellular images derived beyond standard 2D cell culture models. We have recently demonstrated the utility of VAMPIRE analysis for cells embedded in 3D collagen matrices[1]. In that study, we obtained the 2D contours of cells from the z-projected images of these cells. VAMPIRE analysis showed that shape modes for cells in 3D cultures were distinctly more protrusive than the same cells in more traditional 2D cultures[1]. VAMPIRE analysis can also be applied to study changes in nuclear shapes in tissue sections. A growing number of studies have shown that nuclear shape can encode prognostic information for patients in different types of cancers[43,44]. Segmented nuclei of cells in tissue sections can be imported directly into the VAMPIRE worflow to assess changes in nuclear morphology associated with tumor progression and patients outcomes[1].

Limitations of VAMPIRE

An assumption of VAMPIRE analysis is that the shapes of segmented cells and nuclei faithfully represent the original cell and nuclear shapes. The accuracy of this segmentation, for instance using CellProfiler, relies on the user properly optimizing the image processing pipeline, choosing appropriate noise-reduction filters, and using suitable thresholding parameters. If the segmentation is not accurate, the shape modes generated using VAMPIRE will not be representative of the actual shape of cells and nuclei. To address this potential issue, the user should evaluate the accuracy of segmentation before running VAMPIRE. This can be done via visual inspection by overlaying segmented cell contours onto the original image to gauge deviations. If the deviation between the segmented contours and the original images is substantial, the results from VAMPIRE analysis will not be reliable.

A challenge for any cell-morphological tool is the analysis and classification of highly complex cell shapes, such as cells with highly protrusive morphologies. Although VAMPIRE can compute a vast number of features from the coordinates of points along the shape outlines to examine the complexity of cell shapes, the use of a reduced number of coordinates along with the PCA can still lead to detected shape modes of limited spatial resolution. In this case, a morphological analysis that provides for the direct quantification of cell protrusions[7] or takes better account of cell protrusions[27] is more suitable. The distance between the computationally assigned shape modes and actual cell shapes provides a direct way to examine the fitness of the VAMPIRE model applied to any given image dataset. If this distance is large, the VAMPIRE model has failed. To further examine the fitness of VAMPIRE analysis, we have included a measure called "distance from cluster center" (see **Glossary**), which delineates the goodness of the match between shape modes and individual cell contours as part of the output data provided to users. Note that the fitness score depends on the parameters used in VAMPIRE analysis and can be improved by increasing the number of shape modes or eliminating "outlier" cells, as described in details below in the **Experimental design** section.

Comparison with other methods

Two commonly used tools for cell shape analysis are CellProfiler's measurement modules and MorpholibJ[45], a plugin for ImageJ[29] These tools segment images and extract an extensive

list of features, such as shape factor, eccentricity and Zernike number. As mentioned previously, too many shape descriptors can limit the ease of biological interpretation and visualization of morphological data. Here, VAMPIRE presents three major advantages compared to conventional approaches to assess and visualize cell morphology [12,27,30]: 1) VAMPIRE provides an *in silico* visualization scheme that makes it easy to compare and learn biologically meaningful information from complex morphology data[1,23]. 2) VAMPIRE bypasses the feature-extraction step to avoid using overly abstract morphological parameters. 3) VAMPIRE provides a direct means to compute the degree of cellular heterogeneity in cell populations across tested conditions using the Shannon entropy[46], based on the abundance of cells within each of the identified shape mode:

$$S = -\sum p_i ln(p_i)$$

Here, S is the Shannon entropy and p_i is the occurrence of each shape mode.

Experimental design - Selection of parameters for VAMPIRE analysis

Within the VAMPIRE interface, a key input parameter for establishing the model is the number of shape modes. We encourage the user to tune this parameter to obtain optimal results. Here, we briefly present the underlying basis for the selection of the number of shape modes. During the dimensional reduction steps, we implement K-means clustering to relate each individual cell to the centroid of each cluster, where the distance from the cluster centroid is defined as the contour fit (see **Glossary**). This K-means clustering stratifies cells on the principle of minimizing a parameter known as the inertia. This inertia is calculated as the sum of the squared distance between the cluster centroid and each data point within the cluster (**Figure 4A**). Inertia can be thought of as the metric that defines how internally coherent clusters are, with the optimal inertia value being zero.

Fundamentally, increasing the number of clusters reduce the inertia and improve cluster coherence. To illustrate the effect of the number of clusters on the inertia, we plotted the number of clusters as a function of the inertia for cells cultured on adhesive micropatterns (**Figure 4B**). We observed an elbow-shaped decay function, at which point there was only a minimal benefit to increasing the number of clusters

Control experiments

Examining cells of pre-defined shapes is the most straightforward way to validate VAMPIRE analysis. Using adhesive micropatterning techniques, the user evaluates the morphologies of cells confined to pre-defined adhesive shapes (see **Anticipated results**). As a result, cells cultured on circular and triangular adhesive micropatterns should exhibit shape modes that are predominantly circular and triangular, respectively.

Materials

Equipment

- A computer with at least 8GB of RAM running Microsoft Windows 10 (64 bit)

Software

- VAMPIRE executable software
- CSV editor (e.g. Microsoft Excel, Numbers)
- Choice of a standard segmentation tool:
    - CellProfiler 3.1.9 software (https://cellprofiler.org/releases/)
    - ImageJ/FIJI (https://imagej.net/Fiji/Downloads)
    - MATLAB (https://www.mathworks.com/downloads)

Procedure

Segment images of cells or nuclei ● TIMING 10-60 min

1| Segment the fluorescence images to identify the boundaries of cells or nuclei. The VAMPIRE GUI does not segment cells. User should accomplish this task with software, including, but not limited to, ImageJ, MATLAB, and CellProfiler. More information on how to use these segmentation tools can be found on their official websites

- ImageJ: https://imagej.net/Segmentation;
- MATLAB: https://www.mathworks.com/help/images/detecting-a-cell-using-image-segmentation.html;
- CellProfiler: https://cellprofiler.org/tutorials.

To better demonstrate the VAMPIRE analysis procedure, we provide sample images of fluorescently tagged cells in the **Supplementary Data** under "Example images" folder and its corresponding results through VAMPIRE analysis procedure. Two sample sets, MEF_LMNA-- and MEF_wildtype, are provided, and correspond to mouse embryonic fibroblast cells with and without Lamin A knockout, respectively. Throughout this **Procedure** section, refer to the directory of supplementary files in "README.pdf" to locate example data and results. We have provided segmented example images using CellProfiler, as well as a sample CellProfiler segmentation pipeline in **Supplementary files**. Note that the example workflow is designed using CellProfiler version 3.1.9, and it may not work with more recent versions of CellProfiler.

? TROUBLESHOOTING

2| Convert the segmented image data to the required format that is compatible with VAMPIRE analysis, if needed. Invidiual segmented images require either 8-bit or 16-bit grayscale tiff images. Area within each cell must have the same value, and each cell must have unique and non-zero value (**Figure 5A**). Segmented images for the same condition should be placed in the same folder. Segmented images with multiple fluorescent channels may also be saved in the same folder. In this case, the segmented images must have filenames that distinguish objects by channel (i.e. xy001c1.tif and xy001c2.tif). A sample format of segmented images is provided in the **Supplementary Data** for reference.

? TROUBLESHOOTING

Build shape-analysis VAMPIRE model ● TIMING 3-10 min

3| Generate a CSV file to specify which segmented image sets to use to construct a VAMPIRE model. In this CSV file, the first row contains column labels. Each column specifies information about the specific segmented images. From the second row, each column should be filled with information of a specific segmented image set with the following order:

i. "set ID" : row index number.

ii. "condition name" : description of a image set.

iii. "set location" : the location/path of the folder containing segmented images

iv. "tag" : A string of text. Only segmented images in the set location with filenames containing the tag will be identified and analyzed. For example, if

"tag" is set as "c1", for an image set location containing segmented images from multiple channels (i.e. xy001c1.tif, xy001c2.tif, xy002c1.tif, xy002c2.tif) only image filenames containing "c1" (i.e. xy001c1.tif and xy002c1.tif) will be analyzed. If the field is empty, all segmented images in the set location will be analyzed.

v. "note" : any information about the image set for your own record. This is not used in the VAMPIRE analysis.

An example CSV file named "Segmented image sets to build model.csv" can be found in "README.pdf". Users can download and direct modify the example CSV files using Excel or other CSV editors. To use the example segmented images provided in the **Supplementary Data** for the following analysis, the user may need to update the set location column in the example CSV file with the actual location of the example segmented images.

4| Launch VAMPIRE Graphic User Interface (GUI) by opening the VAMPIRE.exe file from **Supplementary Data**.

5| Locate the CSV file generated in **step 3** to build VAMPIRE model in the "Build Model" section of the VAMPIRE GUI. Click "Load CSV". This will open a popup window for the user to select the CSV file.

6| Specify the number of coordinates to extract from the cell contours in Build Model section of VAMPIRE GUI under "number of coordinates" box. The default value is fifty. A higher number of coordinates will better represent the object boundary at the expense of analysis speed. A lower number of coordinates may not capture the details of the object boundary and the result of analysis may under-represent the actual cell morphology.

7| Determine the number of shape modes in the "Build Model" section of the VAMPIRE GUI under "number of shape modes" box. The default value is ten. To optimize this number, refer to the ***Selection of parameters for VAMPIRE analysis*** section in the **Introduction**.

8| Specify where to output the model in the "Build Model" section of VAMPIRE GUI under the "Model output folder" box.

9| Name the model in the "Build Model" section of VAMPIRE GUI under the "Model name" box. This name will be used to generate a pickle file which contains model parameters.

10| Click "Build Model" in VAMPIRE GUI to generate a VAMPIRE model based on the specified parameter values provided in **steps 6 and 7**. Once the model is generated, it will be saved to the output folder specified in **step 8**. Within this new folder, VAMPIRE model data will be saved into a subfolder "[model name]" that contains:

- A VAMPIRE model file which is named as "[model name].pickle".
- A subfolder named "[model name] figures" that contains:
  - The overlay of randomly selected 20 raw shapes falling into each shape mode named "registered objects.png".
  - The dendrogram showing the level of correlation between shape modes named "shape mode dendrogram.png".

Example output files of this step are provided in the **Supplementary Data**, under "Example output". These files are generated from the example segmented images provided in **step 2,** using the default values of parameters from **step 6 and 7**.

? TROUBLESHOOTING

Analyze cell shapes with VAMPIRE model ● TIMING 1-10 min

11| Repeat **step 3** to specify the sets of segmented images to apply the VAMPIRE model on. If you need to prepare new sets of segmented images, repeat step 1 and 2 to do so. The format of the CSV file remains the same. Once the user generates the CSV file, go back to the VAMPIRE GUI. In the "Apply Model" section of the VAMPIRE GUI, click "load CSV". This will open a popup window for the user to select the CSV file.

12| Specify the previously built model to analyze the segmented images. Click "load model" button to choose pickle file generated in step 10. Refer to "README.pdf" locate the pickle file.

13| Perform the VAMPIRE analysis on the specified images by clicking apply "Apply Model" in VAMPIRE GUI. When this process is finished, a new folder will be created named "Result based on [model name]" inside the VAMPIRE model folder. This new folder contains a collection of distributions showing the fractional abundance for cells within each shape mode, with the percent of cells within each shape mode denoted on the top of the bars. See S4 section of "README.pdf" in the **Supplementary Data** to see example outputs. Each

distribution is saved with the naming convention: "Shape mode distribution_[*condition*].png". Clicking the "Apply Model" button also generates a VAMPIRE datasheet CSV file in each segmented image set folder. Each datasheet CSV contains:

- Filename : name of the segmented image file that contains the object
- ImageID : ID number of the segmented image file
- ObjectID : ID number of the object within the segmented image file
- X and Y: location of the object's center of mass within the segmented image
- Area: area of the object
- Perimeter: length of object's circumference
- Lengths of major axis and minor axis
- Circularity: shape factor calculated by $\frac{4\pi A}{P^2}$. Its value varies from 0 to 1. The circularity of a perfect circle is 1.
- Aspect ratio: it is calculated by major axis length divided by minor axis length.
- Shape mode ID number: number that represents the shape mode where each cell belongs to.
- Distance from cluster center: distance between the cluster center and this object

Example output files of this step is provided in the **Supplementary Data**, under "Example output". These files are generated using the VAMPIRE model provided in **Supplementary Data** under the same folder "Example output". See the directory of Supplementary Data in S1 section of "README.pdf" to locate the output files.

Troubleshooting

Troubleshooting guidance can be found in Table 2.

TABLE 2| Troubleshooting table.

Step	Problem	Possible reason	Solution
1	Cannot run pipeline: the pipeline did not identify any image sets	User did not load any images in the "Images" module.	Drag and drop images into the "Images" module of CellProfiler.

1	Subfolder under CellProfiler output folder is named “None”	The metadata extraction rule is incorrect	Modify the extraction rule under “Metadata” module in CellProfiler.
10	IndexError: arrays used as indices must be an integer	Segmented images does not contain any cell or nucleus	Check if segmented images have correct format as specified in step 2, and have at least one cell or nucleus.
10	RuntimeWarning: Mean of empty slice	The number of objects is less than the number of clusters	Provide images with a greater number of cells than the number of clusters.
10	Permission denied	CSV file is open while the analysis is running	Close all CSV files open and repeat step 10.

Timing

The timing information below is estimated based on the analysis of 10,000 cells using an i7-8700k Intel CPU with 5.0 GHz clock speed on Windows 10 pro OS.

Step 1-2, Segment cells or nuclei, 5~10 mins
Step 3, Create a list of images to build the shape-analysis model, 1-3 mins
Steps 4-10, Build shape-analysis model in VAMPIRE, 1-5 mins
Steps 11-13, Application of the model to analyze shapes across conditions, 1-5 mins
Total, steps 1-13, complete VAMPIRE analysis, 8-23 mins

BOX 1 | GLOSSARY

Eigen shapes vectors—Mathematical descriptors used to describe cell shapes based on the principal component analysis (PCA) of cellular shape features. Once determined, a linear

combination of eigen shapes reconstructs the original shape of each cell.

Shape modes—mathematical descriptors of cell and nuclear shapes based on clustering analysis of user-specified eigen shape vectors. Once these shape modes are identified, the abundance of cells within each mode is assessed and the entropy to determine the extent of heterogeneity can be computed.

Shannon entropy—mathematical description used to quantify the degree of diversity within a population of cells based on the number of shape modes and the abundance of cells within each shape mode. It is given by the general equation:

$$S = -\sum p_i ln(p_i)$$

S is the Shannon entropy and p_i is the occurrence of cells in each shape mode.

Cellular heterogeneity—property that describes the extent of cell-to-cell variations within a cell population.

Eccentricity—a measure of how similar a cell shape is to a circle or an ellipse, calculated as the ratio of the distance between the geometric centroid and the focus point by the distance between the geometric centroid and the vertex.

Distance from cluster center— euclidean distance between the 50 equidistant points along the contour of an individual cell and those comprising the shape mode.

PCA— Abbrevation for principle component analysis. PCA is a mathematical technique for reducing the dimensionality of large datasets, increasing interpretability but at the same time minimizing information loss by finding new uncorrelated variables, principal components, from possibly correlated variables.

Anticipated results

To demonstrate the utility of VAMPIRE, we examined the shapes of mouse embryonic fibroblasts (MEFs) in response to different surface topographies. These cells are either wild type (MEF LMNA +/+) or deficient in lamin A/C (MEF LMNA-/-). Cells were seeded onto three different glass surfaces: 1. circular, 2. triangular shape fibronectin coated islands, surrounded by poly-ethylene glycol (PEG) passivated regions, and 3. surface without any island nor PEG. Cells were incubated overnight, then fixed and stained with DAPI and Alexa

Fluor 488 Phalloidin, highlighting the nuclear DNA and F-actin fibers respectively. Cells and their corresponding nuclei were segmented using CellProfiler, then the contours were analyzed using VAMPIRE with 10 shape modes and 50 contour points (**Figure 6A**).

We quantified the shape mode distribution for each of the all probed conditions and examined whether cells on patterns exhibited associations with particular shape modes that resembled circles and triangles (**Figure 6B**). As expected, results showed that both LMNA +/+ and LMNA -/- cells seeded on un-patterned surfaces exhibited a more mixed shape profile i.e., similar aboundance in all identified cellular shape modes, as opposed to the cells seed on the pattened substrates. Cells seeded on circular patterns exhibited an enrichment in the circular shape mode (mode 4) with an average abundance of 55% and 52% of the total cell populations respectively, compared to 8.1% and 21% of those seeded on unpatterned substrate. Cells seeded on triangular patterns the two triangular shape modes, the "sharp" (mode 1) and "blunted" vertex (mode 2) triangles, while decreasing the cell aboundance in irregular cell shapes (mode 6-9) (**Figure 6B**). Interestingly, LMNA -/- cells seeded on triangular patterns were classified as "blunt" (mode 2) three times more with abundance of 34% than "shape" (mode 1) with abundance of 12%. We did not observe such difference between two shape mode abundance in LMNA +/+ cells. This bias suggests that the deficiency in lamin A/C limits the ability of these cells to form acute angle vertices, potentially through defective nucleo-cytoskeletal connections[41,47]. Our results reveal that cells can respond morphologically differently to the same shape constrains and VAMPIRE analysis can visualize and quantify the subtle differences.

We computed the Shannon entropy for the cell populations and observed no significant differences between LMNA +/+ and LMNA -/- within the same micropattern (**Figure 6B**). However, looking across conditions, we observe a significant decrease in the population heterogeneity for both LMNA +/+ and LMNA -/- seeded on circular patterns, relative to cells seeded on unpatterned surfaces and triangular patterns. The aspect ratio of LMNA +/+ cells increased from 1.66 (no pattern) to 2.20 (triangle pattern), suggesting a more elongated shape for these cells; however, their shape factors increased from 0.34 (no pattern) to 0.51 (triangle pattern), suggesting the cell shape was rounder. These seemingly contradictory results measured by shape factor and aspect ratio suggests that VAMPIRE

analysis can provide direct visual insight to better monitor the transition of cell morphology than classical morphology parmaters.

We also examined the association between cellular morphology and chronological ages of dermal fibroblasts with a VAMPIRE analysis model from a panel of seven healthy individuals[23]. Previously, we demonstrated that cell and nuclear morphology of dermal fibroblasts encode key information about the biological age for healthy individuals[23]. Using ten shape modes, the VAMPIRE analysis shows a decrease in the frequency of cells having rounded morphologies shape modes, and an increase in cells having non-rounded morphologies with increasing age. This is measured by a negative age-correlations for shape modes 1 and 2 with rounded shape, and positive age-correlations for non-rounded shape modes 3, 4 and 7 (**Figure 7A**). The correlation is calculated using Pearson's correlation coefficient. We also note that computing standard shape parameters, including shape factor and aspect ratio, yielded very similar values for the cells in different shape modes, (SF: 0.77-0.83, and AR: 1.51-1.64), even for shape modes having opposite trends in age correlations (R: -0.6 and +0.6) —i.e. shape modes 1 and 3. Furthermore, circular shape modes 1 and 2 have very similar shape parameters (SF and AR) to ellipsoidal shape modes 9 and 10 (**Figure 7B**). VAMPIRE analysis, however, is capable of visually identifying cell morphological changes that otherwise would have been unnoticed.

Boasting the utility beyond cultured cells, we have successfully implemented VAMPIRE analysis to tissues. Here we analyzed the human epidermis and reticular dermis using hematoxylin and eosin (H&E) stained tissue section (**Figure 8A**). Note that we segmented nuclei in the sections using a custom image analysis algorithm. To compare the morphology of cells in epidermis and reticular dermis region, we built VAMPIRE model using nuclei segmented from the scanned image of a H&E stained skin tissue biopsy from a 79 years old donor. We observed that shape modes 1 through 3 were more elongated (i.e. less circular) than modes 4 through 10 (**Figure 8B**). As expected, VAMPIRE analysis shows that nearly 50% of dermal cells are classified as modes 1 through 3, as compared to only 6.4% for epidermal cells (**Figure 8B**).

DATA AVAILABILITY The datasets generated during and/or analyzed during the current study are available from the corresponding author(s) upon reasonable request; an example dataset is provided in Supplementary Data.

CODE AVAILABILITY

The source code is available on Github : https://github.com/kukionfr/VAMPIRE_open . The code can be accessed and used by readers without restriction.

REFERENCES

1. Wu, P. H. *et al.* Evolution of cellular morpho-phenotypes in cancer metastasis. *Sci. Rep.* **5**, (2015).
2. Chen, W.-C. *et al.* Functional interplay between the cell cycle and cell phenotypes. *Integr. Biol.* **5**, 523 (2013).
3. Chambliss, A. B., Wu, P. H., Chen, W. C., Sun, S. X. & Wirtz, D. Simultaneously defining cell phenotypes, cell cycle, and chromatin modifications at single-cell resolution. *FASEB J.* **27**, 2667–2676 (2013).
4. Bakal, C., Aach, J., Church, G. & Perrimon, N. Quantitative morphological signatures define local signaling networks regulating cell morphology. *Science (80-.).* **316**, 1753–1756 (2007).
5. Rohban, M. H. *et al.* Systematic morphological profiling of human gene and allele function via cell painting. *Elife* **6**, (2017).
6. Wu, P.-H. *et al.* Single-cell morphology encodes metastatic potential. *Sci. Adv.* **6**, eaaw6938 (2020).
7. Guo, Q. *et al.* Modulation of keratocyte phenotype by collagen fibril nanoarchitecture in membranes for corneal repair. *Biomaterials* **34**, 9365–9372 (2013).
8. Sero, J. E. *et al.* Cell shape and the microenvironment regulate nuclear translocation of NF-κB in breast epithelial and tumor cells. *Mol. Syst. Biol.* **11**, 790 (2015).
9. Simm, J. *et al.* Repurposed High-Throughput Images Enable Biological Activity Prediction For Drug Discovery. *doi.org* 108399 (2017). doi:10.1101/108399
10. Bray, M.-A. *et al.* A dataset of images and morphological profiles of 30 000 small-

molecule treatments using the Cell Painting assay. *Gigascience* **6**, 1–5 (2017).

11. Wawer, M. J. *et al.* Toward performance-diverse small-molecule libraries for cell-based phenotypic screening using multiplexed high-dimensional profiling. *Proc. Natl. Acad. Sci.* **111**, 10911–10916 (2014).
12. Bray, M. A. *et al.* Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* **11**, 1757–1774 (2016).
13. Beghin, A. *et al.* Localization-based super-resolution imaging meets high-content screening. *Nat. Methods* **14**, 1184–1190 (2017).
14. Carpenter, A. E. Extracting rich information from images. *Methods Mol. Biol.* **486**, 193–211 (2009).
15. Meijering, E., Carpenter, A. E., Peng, H., Hamprecht, F. A. & Olivo-Marin, J. C. Imagining the future of bioimage analysis. *Nat. Biotechnol.* **34**, 1250–1255 (2016).
16. Piccinini, F. *et al.* Advanced Cell Classifier: User-Friendly Machine-Learning-Based Software for Discovering Phenotypes in High-Content Imaging Data. *Cell Syst.* **4**, 651-655.e5 (2017).
17. Danuser, G. Computer vision in cell biology. *Cell* **147**, 973–978 (2011).
18. Falk, T. *et al.* U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67–70 (2019).
19. Gabril, M. Y. & Yousef, G. M. Informatics for practicing anatomical pathologists: Marking a new era in pathology practice. *Modern Pathology* **23**, 349–358 (2010).
20. Fuchs, T. J. & Buhmann, J. M. Computational pathology: Challenges and promises for tissue analysis. *Computerized Medical Imaging and Graphics* **35**, 515–530 (2011).
21. Sarnecki, J. S. *et al.* A robust nonlinear tissue-component discrimination method for computational pathology. *Lab. Investig.* **96**, 450–458 (2016).
22. Beck, A. H. *et al.* Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival. *Sci. Transl. Med.* **3**, 108ra113-108ra113 (2011).
23. Phillip, J. M. *et al.* Biophysical and biomolecular determination of cellular age in humans. *Nat. Biomed. Eng.* **1**, 0093 (2017).
24. Pegoraro, G. & Misteli, T. High-Throughput Imaging for the Discovery of Cellular Mechanisms of Disease. *Trends in Genetics* **33**, 604–615 (2017).
25. Lang, P., Yeow, K., Nichols, A. & Scheer, A. Cellular imaging in drug discovery. *Nature*

Reviews Drug Discovery **5**, 343–356 (2006).

26. Loo, L. H., Wu, L. F. & Altschuler, S. J. Image-based multivariate profiling of drug responses from single cells. *Nat. Methods* **4**, 445–453 (2007).
27. Sailem, H. Z., Sero, J. E. & Bakal, C. Visualizing cellular imaging data using PhenoPlot. *Nat Commun* **6**, 1–6 (2015).
28. McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* (2018). doi:10.1371/journal.pbio.2005970
29. Schindelin, J. *et al.* Fiji: An open-source platform for biological-image analysis. *Nature Methods* (2012). doi:10.1038/nmeth.2019
30. Hodneland, E., Kögel, T., Frei, D. M., Gerdes, H. H. & Lundervold, A. CellSegm - a MATLAB toolbox for high-throughput 3D cell segmentation. *Source Code Biol. Med.* **8**, (2013).
31. Jayatilaka, H. *et al.* EB1 and cytoplasmic dynein mediate protrusion dynamics for efficient 3-dimensional cell migration. *FASEB J.* (2018). doi:10.1096/fj.201700444RR
32. Giri, A. *et al.* The Arp2/3 complex mediates multigeneration dendritic protrusions for efficient 3-dimensional cancer cell migration. *FASEB J.* (2013). doi:10.1096/fj.12-224352
33. Fraley, S. I. *et al.* A distinctive role for focal adhesion proteins in three-dimensional cell motility. *Nat. Cell Biol.* **12**, 598–604 (2010).
34. Jayatilaka, H. *et al.* Synergistic IL-6 and IL-8 paracrine signalling pathway infers a strategy to inhibit tumour cell migration. *Nat. Commun.* **8**, (2017).
35. Jayatilaka, H. *et al.* Tumor cell density regulates matrix metalloproteinases for enhanced migration. *Oncotarget* **9**, 32556–32569 (2018).
36. Phillip, J. M., Aifuwa, I., Walston, J. & Wirtz, D. The Mechanobiology of Aging. *Annu. Rev. Biomed. Eng.* **17**, 113–141 (2015).
37. Kim, D.-H. *et al.* Volume regulation and shape bifurcation in the cell nucleus. *J. Cell Sci.* **129**, 457–457 (2016).
38. Yu, Y. *et al.* Inhibition of Spleen Tyrosine Kinase Potentiates Paclitaxel-Induced Cytotoxicity in Ovarian Cancer Cells by Stabilizing Microtubules. *Cancer Cell* **28**, 82–96 (2015).
39. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*

(2011).

40. Kim, D. H. & Wirtz, D. Focal adhesion size uniquely predicts cell migration. *FASEB J.* **27**, 1351–1361 (2013).
41. Kim, J.-K. *et al.* Nuclear lamin A/C harnesses the perinuclear apical actin cables to protect nuclear morphology. *Nat. Commun.* **8**, 2123 (2017).
42. Zheng, W., Thorne, N. & McKew, J. C. Phenotypic screens as a renewed approach for drug discovery. *Drug Discovery Today* (2013). doi:10.1016/j.drudis.2013.07.001
43. Kashyap, A., Jain, M., Shukla, S. & Andley, M. Role of nuclear morphometry in breast cancer and its correlation with cytomorphological grading of breast cancer: A study of 64 cases. *J. Cytol.* (2018). doi:10.4103/JOC.JOC_237_16
44. Seethala, R. R. *et al.* Noninvasive follicular thyroid neoplasm with papillary-like nuclear features: A review for pathologists. *Modern Pathology* (2018). doi:10.1038/modpathol.2017.130
45. Legland, D., Arganda-Carreras, I. & Andrey, P. MorphoLibJ: Integrated library and plugins for mathematical morphology with ImageJ. *Bioinformatics* (2016). doi:10.1093/bioinformatics/btw413
46. Altschuler, S. J. & Wu, L. F. Cellular Heterogeneity: Do Differences Make a Difference? *Cell* **141**, 559–563 (2010).
47. Kim, D. H. & Wirtz, D. Cytoskeletal tension induces the polarized architecture of the nucleus. *Biomaterials* **48**, 161–172 (2015).

Acknowledgement

This work was supported in part by the National Institutes of Health Grants U54CA143868 (DW), R01CA174388 (DW), and U01AG060903 (DW, JMP, PHW)

Author information

Jude M. Phillip

Present address: Department of Medicine, Division of Hematology and Oncology, Weill Cornell Medicine, New York, New York, 10065, USA

Kyu-Sang Han
Present address: Department of Chemical and Biomolecular Engineering, Johns Hopkins Physical Sciences Oncology Center, Johns Hopkins Institute for Nanobiotechnology (INBT), Johns Hopkins University, Baltimore, Maryland 21218, USA

Wei-Chiang Chen
Present address: Department of Chemical and Biomolecular Engineering, Johns Hopkins Physical Sciences Oncology Center, Johns Hopkins Institute for Nanobiotechnology (INBT), Johns Hopkins University, Baltimore, Maryland 21218, USA

Denis Wirtz
Present address: Department of Chemical and Biomolecular Engineering, Johns Hopkins Physical Sciences Oncology Center, Johns Hopkins Institute for Nanobiotechnology (INBT), Johns Hopkins University, Baltimore, Maryland 21218, USA

Pei-Hsun Wu
Present address: Department of Chemical and Biomolecular Engineering, Johns Hopkins Physical Sciences Oncology Center, Johns Hopkins Institute for Nanobiotechnology (INBT), Johns Hopkins University, Baltimore, Maryland 21218, USA

These authors contributed equally: Jude M. Phillip, Kyu-Sang Han

Affiliations
Department of Chemical and Biomolecular Engineering, Johns Hopkins Physical Sciences Oncology Center, Johns Hopkins Institute for Nanobiotechnology (INBT), Johns Hopkins University, Baltimore, Maryland 21218, USA
Jude M. Phillip, Kyu-Sang Han, Wei-Chiang Chen, Denis Wirtz, Pei-Hsun Wu

Department of Pathology, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA

Denis Wirtz

Department of Oncology, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA
Denis Wirtz

Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA
Denis Wirtz

Department of Medicine, Division of Hematology and Oncology, Weill Cornell Medicine, New York, New York, 10065, USA
Jude M. Phillip

Contributions

JMP and PHW designed and conducted experiments; PHW, JMP, DW and WC conceived analysis and workflow of VAMPIRE; PHW developed the original VAMPIRE software; KSH converted the VAMPIRE software from MATLAB to Python; KSH developed the graphical user interface of VAMPIRE; KSH and JMP analyzed and plotted data; PHW and DW supervised the study; JMP, DW, KSH and PHW wrote and edited the protocol; DW, JMP and PHW secured funding.

Corresponding authors

Correspondence to Pei-Hsun Wu and Denis Wirtz

Ethics declarations

Competing interests

The authors declare no competing interests.

FIGURE CAPTIONS

Figure 1. Cells confined to narrow ranges of traditional morphological parameters still exhibit highly variable shapes. Scatter plot showing the distributions of aspect ratio, shape factor, and solidity of 37,750 mouse embryonic fibroblast cells (MEF) grown on collagen-coated substrates. The subset of 10 cells highlighted in red display substantial morphological heterogeneity, despite highly similar aspect ratio, circularity and solidity.

Figure 2. Overview of VAMPIRE analysis, from the extraction of contour coordinates to the automatic generation of shape modes A. The contour of a cell described by 50 equidistant points along its contour. **B.** Unaligned (left) shapes of a set of cells are pooled, normalized by size, and aligned (right). **C**. Eigen-shape vectors (i.e., principal components or PCs) are obtained from a principal component analysis (PCA) of the contour coordinates of aligned cells. **D**. Reconstructed cell shape from a reduced number of eigen-shape vectors. The reduced number of eigen-shape vectors was dafulted at the number of vectors that comprise 95% of the shape variations among all assessed cells. **E**. Representative cellular shape modes are obtained by applying a k-means clustering method to a set of cell morphology data described by the reduced number of eigen-shape vectors.

Figure 3. Overview of VAMPIRE implementation with the VAMPIRE GUI. A. The VAMPIRE Graphic User Interface (GUI). **B.** Flow diagram illustrating key steps in the implementation of VAMPIRE analysis with VAMPIRE GUI. Image of cells are first segmented into binary image that highlight cellular region and/or nuclear region. In VAMPIRE GUI top sections allows user to specify analysis parameters and (multi-)set of segmented images to be used to create a VAMPIRE analysis model (highlighted in red). Once the VAMPIRE analysis model is established, the user can specify (multi-)set of segmented images to be analyzed with the previous established model (highlighted in blue).

Figure 4. **Determinants of cluster coherence in the shape mode distributions. A.** Schematic illustrating the concept of inertia in K-means clustering. The inertia is measured by total distances of all data points to the centroids of their corresponding subtype. The lower the inertia value indicates better clustering result with more coherence. **B.** The inertia in principal decays with increasing number of clusters. The corresponding cluster number at the elbow point where the inertia decay rate starts to drop is the suggested cluster number to use in VAMPIRE for k-means clustering. The inertia is calculated as the distance from the cluster centroid to each of 17,093 mouse embryonic fibroblast cell and nucleus. VAMPIRE models were built using number of clusters = 2, 5, 10, 15, 20, 25, 30, and 35 with repeats of five (N=5). Because of the random initialization of K-means clustering, the cluster results (centroid and distance) change between repeats. The errorbars represent the standard deviation of inertia between repeats.

Figure 5. **VAMPIRE analysis of example *LMNA*+/+ and *LMNA*-/- mouse embryonic fibroblasts. A.** Images of phalloidin-stained (top) wild-type (*LMNA*+/+, left) and lamin-deficient (*LMNA*-/-, right) mouse embryonic fibroblasts are provided Supplementary Data so that users can test the protocol and adapt it for their own application. Images of segmented cells (bottom) are also provided in the Supplementary Data. Segmentation is obtained through CellProfiler and the pipeline file used in CellProfiler is also provided so that users can adapt and test it. **B**. Bar graphs show the distribution of cell shape modes from the VAMPIRE analysis with the example images of MEFs. Numbers above the bars represent the abundances [%] of cells in each shape mode.

Figure 6. VAMPIRE analysis of mouse embryonic fibroblasts seeded on adhesive micro-patterned surfaces. A. Fluorescence microscopy images of wild-type (*LMNA*+/+) and lamin-deficient (*LMNA*-/-) mouse embryonic fibroblasts cultured on circular (top row) and triangular (middle row) adhesive fibronectin-coated micropatterns. Control cells (bottom row) are placed on the fibronectin-coated glass. Cells were fixed and stained for F-actin (red) and nuclear DNA (blue). Segmented fluorescence images (right). On the left are the raw images of cells and their nuclei with the segmented contours highlighted in yellow; on the left are the same cells color-coded according to the shape mode to which they belong.

Inserts are magnified views of cells. The identified shape modes are located on the right of the panel. **B**. Shape mode distributions for MEF++ and MEF-- cultured on circular or triangular micropatterns (top and middle rows) and unpatterned surfaces (bottom row). The table on the right displays the values for traditional morphological parameters, including average area, shape factor, and aspect ratio of cells, as well as the number of cells analyzed (N), lamin A/C status and the Shannon entropy of the cells. The number of cells or nuclei analyzed for each condition are N= 3888, 3453, 1956, 2874, 3538, 1384 from the top row to bottom row. The measurements for each condition were calculated as the mean of all objects detected in two distinct samples. These results show that traditional morphological parameters cannot discriminate between the nuclear morphological responses of MEF++ and MEF-- on different adhesive micropatterns (right table). In contrast, the differential morphological response of these cells is readily revealed when measured via shape mode distributions (left color-coded table).

Figure 7. VAMPIRE analysis of human dermal fibroblasts from donors of different ages. A. Distributions of nuclear shape modes for dermal fibroblasts from age 3 to 96. Each row is a distribution for each donor with sample number of nuclei of N=643, 420, 407, 531, 373, 575, 637 respectively from age 3 to 96. Cells from younger donors populate the rounder shape mode (mode 1 and 2), while cells from older donors have nuclei classified that populate the non-rounded shape modes (mode 3, 4, and 7). **B.** Pearson's correlation (R), shape factor (SF), and aspect ratio (AR) of each nuclear shape mode. R is correlation between age and abundance of nuclei in a specific shape mode. SF and AR are calculated as the mean of all nuclei classified in each shape mode across all ages. SF and AR of each shape mode are calculated as the mean of all nuclei identified as that shape mode across all ages.

Figure 8. Analysis of nuclear shape in H&E stained tissue sections with VAMPIRE. A. Images of a skin tissue section stained with hematoxylin and eosin (H&E) and obtained from the cancer genome atlas (TCGA case ID: EE-A20I). Nuclei in the epidermis and the reticular dermis regions were segmented and analyzed with VAMPIRE. **B.** Bar graphs show the distribution of nuclei shape modes, comparing epidermal cells (N=1579) and dermal cells

(N=498) using VAMPIRE analysis. Numbers above the bars represent the abundances [%] of nuclei in each shape mode.