nature communications

Article

https://doi.org/10.1038/s41467-025-58082-0

Automated prediction of fibroblast phenotypes using mathematical descriptors of cellular features

Received: 24 May 2024

Accepted: 5 March 2025

Published online: 22 March 2025

Check for updates

Alex Khang[1,2], Abigail Barmore[1,2], Georgios Tseropoulos[1,2], Kaustav Bera[1,2], Dilara Batan[2,3] & Kristi S. Anseth[1,2]

Fibrosis is caused by pathological activation of resident fibroblasts to myofibroblasts that leads to aberrant tissue stiffening and diminished function of affected organs with limited pharmacological interventions. Despite the prevalence of myofibroblasts in fibrotic tissue, existing methods to grade fibroblast phenotypes are typically subjective and qualitative, yet important for screening of new therapeutics. Here, we develop mathematical descriptors of cell morphology and intracellular structures to identify quantitative and interpretable cell features that capture the fibroblast-to-myofibroblast phenotypic transition in immunostained images. We train and validate models on features extracted from over 3000 primary heart valve interstitial cells and test their predictive performance on cells treated with the small molecule drugs 5-azacytidine and bisperoxovanadium (HOpic), which inhibited and promoted myofibroblast activation, respectively. Collectively, this work introduces an analytical framework that unveils key features associated with distinct fibroblast phenotypes via quantitative image analysis and is broadly applicable for high-throughput screening assays of candidate treatments for fibrotic diseases.

Fibrosis can occur in virtually any organ including the heart, kidney, liver, skin, and lungs and is implicated in nearly 45% of deaths in the modern world[1]. Moreover, fibrosis and cancer associated fibroblasts are known to play a role in tumor metastasis[2]. Fibrosis occurs when normal wound healing processes are dysregulated, leading to excessive deposition of structural extracellular matrix (ECM) components, such as collagen, and results in aberrant stiffening of tissues and organs. If left unresolved, the fibrotic process causes abnormal remodeling of the ECM, organ malfunction or even failure, and in severe cases, death[3]. To date, only two drugs, OFEV® (nintedanib) and Esbriet® (pirfenidone), are approved to treat fibrosis but are marketed only for the treatment of idiopathic pulmonary fibrosis (IPF)[4].

While the role of the fibroblast to myofibroblast transition is well recognized in fibrosis and disease progression, the complex nature of the disease, especially its initiation and progression as in the case of IPF[5], has led to a scarcity of pharmaceutical therapies. Typically, fibroblasts reside within stromal tissue and are responsible for the secretion and remodeling of ECM components, thereby playing a critical role in many wound healing processes. In response to insult, quiescent fibroblasts undergo a phenotypic activation to myofibroblasts characterized by increased proliferation, ECM synthesis, and contractility mediated by expression of alpha-smooth muscle actin (αSMA) stress fibers. When periods of repeated or severe insult occur, the tissue-resident myofibroblast phenotype can persist, which results in excessive matrix deposition and pathological tissue/organ stiffening. Subsequently, excessive tissue/organ stiffening further encourages myofibroblast differentiation and initiates a positive feedback loop. Due to their pivotal role in fibrosis, fibroblasts and their activation to myofibroblasts have been extensively studied in virtually every organ system[6–9].

[1]Department of Chemical and Biological Engineering, University of Colorado Boulder, Boulder, CO, USA. [2]The BioFrontiers Institute, University of Colorado Boulder, Boulder, CO, USA. [3]Department of Biochemistry, University of Colorado Boulder, Boulder, CO, USA. e-mail: kristi.anseth@colorado.edu

A hallmark and widely used indicator of myofibroblast activation is the organization of αSMA into discrete stress fibers, which is not found in quiescent fibroblasts[9–12]. Traditionally, researchers and pathologists have used immunostaining of tissue sections or fibroblast cultures to visualize αSMA; however, grading the severity and discreteness of the fibers has remained largely subjective. Prior published methods have used absolute or normalized fluorescence intensity of αSMA+ cells or tissues, but these measurements can vary with sources of antibodies, staining protocols, and different microscopes and imaging parameters (e.g., laser power and gain). Moreover, significant information, such as cell morphology, can go unused in fibroblast phenotyping. Some studies that have considered cell morphological measurements used qualitative assessments to classify cells[13]. Despite the widespread occurrence of fibrotic diseases and the desire to discover drugs that can target pathological myofibroblasts, a universally accepted method for quantifying fibroblast phenotypes is currently lacking and further complicates cross-study comparisons.

Towards automated and objective methods to characterize fibroblast phenotypes, researchers have begun to use neural networks to differentiate between quiescent and activated fibroblasts based on manually curated training data[14]. Although effective, these models require substantial training datasets, and the predictions are typically based on abstract features with limited physical and/or physiological interpretations. However, we recognize that promising approaches have been developed toward interpreting deep-learned features[15,16]. Furthermore, neural networks often exhibit high complexity, which might be unnecessary for the categorical classification of fibroblasts. As an alternative approach, others have used cell and nuclear morphologies to classify fibroblasts, which provides an avenue to better understand how myofibroblasts are different and not simply if they are different[17–19]. However, to date, no approach exists to combine morphological features and αSMA stress fiber architecture to grade fibroblast phenotypes.

In the present study, we developed a quantitative image analysis approach that includes both cell morphological features and expression of discrete αSMA stress fibers to demarcate quiescent and activated fibroblast phenotypes and to visualize their shapes across the activation spectrum. Our efforts focus on shape and the presence of αSMA that is specifically organized into discrete stress fibers, as it is a widely used characteristic to delineate fibroblasts from myofibroblasts[6,13,14,17,20,21]. As a result, data sets of fluorescence images are routinely collected to characterize fibroblast and myofibroblast populations and are readily available. Motivated by the prevalence of valve and cardiac fibrosis in the general population[22,23], we use valve interstitial cells (VICs)[24–29] isolated from porcine aortic valves and cardiac fibroblasts (CFs)[30,31] isolated from adult rat ventricles as model fibroblast cells. VICs and CFs were induced into an activated state using both mechanical (soft vs stiff microenvironment) and biochemical (TGF-β) stimuli, respectively, to test and validate our analysis approach on several scenarios that promote myofibroblast activation. First, we train and validate models on the VIC dataset and then use the model to grade test data sets comprised of VICs seeded on stiff substrates and treated with small molecule drugs 5-azacytidine and bisperoxovanadium (HOpic) (bpV(HOpic)) that inhibit and promote myofibroblast activation, respectively. As a final example, we extend the VIC-trained models to grade the CF dataset and compare the grading to models trained directly on the CF dataset (Appendix A in Supplemental Information). In our image analysis methodology, we use Fourier series to mathematically describe segmented fibroblast shapes and use the variable coherence to quantify the anisotropy of αSMA stress fibers in a manner independent from absolute fluorescence intensity. This information is then combined and used in exploratory data analysis to discover, and importantly visualize, the greatest variation between fibroblasts and myofibroblasts. Physical interpretation of these differences allows for the identification of morphological features (e.g., cell area, aspect ratio) that vary significantly between quiescent and activated fibroblasts. Finally, we test the ability of minimal key features to predict fibroblast phenotype and compare the automated algorithm to manual grading done via an independent visual assessment. Overall, we report a method using mathematical descriptors of cellular features to discover key measures that can predict fibroblast phenotypes with high levels of accuracy using a straightforward, but robust, method. In addition, the method is compatible with any arbitrary cellular shape, regardless of complexity, and performs cellular phenotyping based on physically interpretable features. Looking forward, the developed methodology can minimize human bias and serve as a high throughput method for grading fibroblast phenotypes during drug screening and complement pre-existing high-throughput assays of candidate treatments for fibrotic diseases[32–34]. Moreover, we anticipate that the established methodology may be extended to analyze fibroblast phenotypes in situ and augment histological assessment of diseased tissues. In a broader context, the current method presents the potential of a gold standard across research groups to ensure consistent fibroblast grading that would largely bolster the field of fibrosis research and expedite drug discovery.

Results

Mathematical descriptors of cell and nuclear shapes

Cell and nuclear shapes are known to be influenced by changes in gene and protein expression, and in this regard, can be viewed as proxies of information regarding phenotypic state[18,19,35–37], rendering methods to quantify cell morphologies useful towards understanding cell function. Here, we extended Fourier analysis[38–40] to create mathematical descriptions of cellular (Fig. 1a) and nuclear (Fig. 1b) shapes and produce faithful reconstructions of their boundaries (Fig. 1c) from an optimized number of coefficients. Fourier analysis enables the reconstruction of arbitrary 2D shapes with varying degrees of precision, determined by the user-defined number of Fourier harmonics (n) (Fig. 1d). Two parametric functions $x(\theta)$ and $y(\theta)$ are employed to relate the arc length around a given cell/nuclei boundary (θ) to the corresponding x- and y- Cartesian coordinates. Here, the arc length of the shape is normalized such that the total length is 2π. Essentially, this results in a one-to-one mapping of the original shape to a unit circle, ensuring that every cartesian x,y point corresponds to exactly one arc length θ in the range $[0, 2\pi]$. The equations $x(\theta)$ and $y(\theta)$ take the form

$$x(\theta) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos n\theta + b_n \sin n\theta \tag{1}$$

$$y(\theta) = \frac{c_0}{2} + \sum_{n=1}^{\infty} c_n \cos n\theta + d_n \sin n\theta \tag{2}$$

where a_O, a_n, b_n, c_O, c_n, and d_n are Fourier coefficients determined by solving overdetermined systems of linear equations. Ultimately, this process yields Fourier coefficients that contain information regarding cellular morphology. However, it is important to note that the physical meaning of the coefficients is not obvious and instead offers a unique opportunity to employ a data-driven approach to analyze cell shapes. Here, the Fourier series representations of VIC body and nuclear geometries reduce the total number of variables being analyzed by representing many boundary points with appreciably fewer coefficients (e.g., > 1000 boundary points are reproduced accurately with 50 Fourier coefficients). Furthermore, the Fourier series produces a smoothed representation of the target shape when n is optimized and is key to recovering naturally smooth cell body and nuclear boundaries that can be distorted when represented by discrete pixels during the imaging and digitization process. The value of n was determined

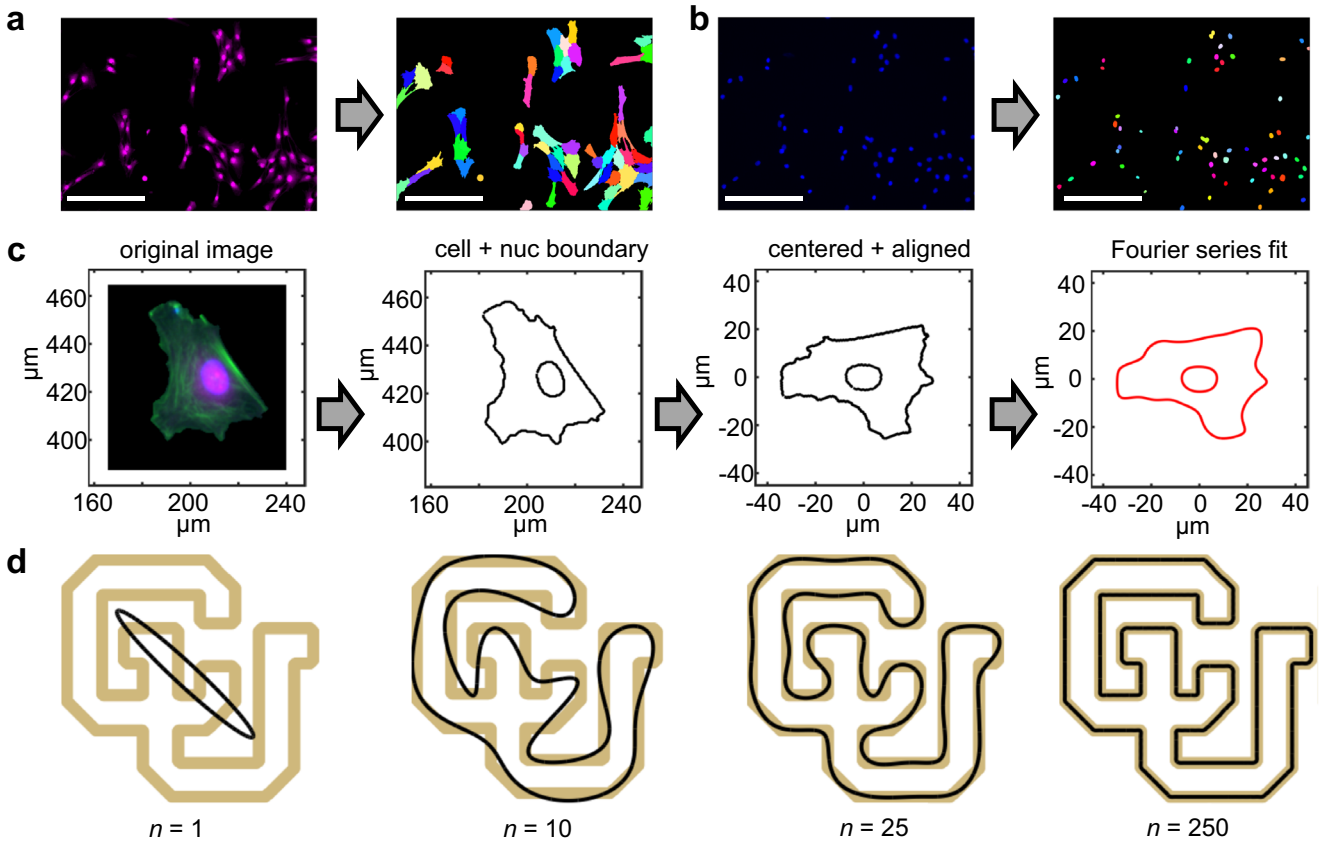

Fig. 1 | Segmentation and mathematical representation of VIC body and nuclear morphologies. a The VIC cytoplasm is imaged and segmented using a marker-based watershed algorithm with the (**b**) segmented nuclear geometries serving as markers. **c** VIC morphological analysis is achieved through segmenting the cell and nuclear boundaries, centering, alignment of the VIC along its longest axis, and fitting a Fourier series expansion to the VIC body and nuclear shapes (body: $n = 15$ harmonics, nuclei: $n = 5$ harmonics). **d** Fourier series expansion used to represent the geometry of a closed shape. Note that as the harmonic number n increases, the Fourier series expansion (black outline) converges to the target shape (gold outline). Stained images: green - αSMA, magenta – cytoplasm, blue – nuclei. Scale bars: 200 μm.

empirically through plotting the discrepancy between the Fourier representation and target shape (residual sum of squares) with respect to increasing n and identifying where further increases in n no longer result in meaningful improvements to the Fourier representation. This process was completed for 100 randomly selected cells (Supplemental Fig. 1). We determined that $n = 15$ (62 coefficients total) and $n = 5$ (22 coefficients total) were sufficient to faithfully reconstruct cellular bodies and nuclei, respectively. We also computed the Bayesian Information Criterion[41] and Akaike Information Criterion[42] which further justified the choice of n for cell bodies and nuclei (see Supplemental Fig. 1).

Quantification of αSMA stress fiber structures

After achieving mathematical descriptors of VIC cell body and nuclear shape, we developed complementary analytical methods to characterize αSMA stress fibers. Specifically, we computed coherence (c), which reflects the local degree of anisotropy[43–45] and is a sensitive metric of αSMA stress fiber organization because of their fibrous nature. To compute c, we first computed the x- and y-gradients (I_x and I_y, respectively) of αSMA fluorescent images on a pixel-by-pixel basis using the Sobel-Feldman operator[46]. Next, we assembled the structure tensor S

$$S = \begin{bmatrix} W * I_x^2 & W * (I_x I_y) \\ W * (I_x I_y) & W * I_y^2 \end{bmatrix} \quad (3)$$

where W is a Gaussian smoothing kernel with a standard deviation of 4 and the symbol * denotes convolution. Eigendecomposition of S was performed to obtain eigenvalues (λ_1, λ_2) and eigenvectors ($\boldsymbol{e}_1$, $\boldsymbol{e}_2$). If $\lambda_1 = \lambda_2$, then the gradient has no preferred direction which can occur in the case of rotational symmetry or homogeneity (Fig. 2a, left-most panel). If $\lambda_1 > \lambda_2$, then λ_1 represents the greatest gradient (i.e., change in intensity) squared at a given pixel that is oriented along $\boldsymbol{e}_1$ whereas λ_2 represents the gradient squared along the orthogonal direction $\boldsymbol{e}_2$ (Fig. 2a, middle three panels). In the case that $\lambda_1 > 0$ and $\lambda_2 = 0$, the structure in the image is completely aligned and is locally one-dimensional (Fig. 2a, right-most panel). The relative difference in magnitude between λ_1 and λ_2 reflects the local degree of anisotropy and is quantified by c

$$c = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \quad (4)$$

whose value ranges from 0 to 1. We demonstrate that c is resistant to changes in absolute fluorescence intensity, distance from the focal plane, and pixel resolution (Supplemental Fig. 2), highlighting its versatility in various imaging scenarios. Moreover, c is a continuous variable whose limits represent completely diffuse αSMA staining ($c = 0$) and discrete αSMA stress fibers ($c = 1$) (Fig. 2a). In this analysis, c provides a quantifiable measure of the amount of αSMA organized in stress fibers (Fig. 2b), thereby providing a continuous measure of the quiescent to activated VIC phenotypic spectrum. The computed values of c were large in regions with discrete αSMA stress fibers and low in

https://doi.org/10.1038/s41467-025-58082-0

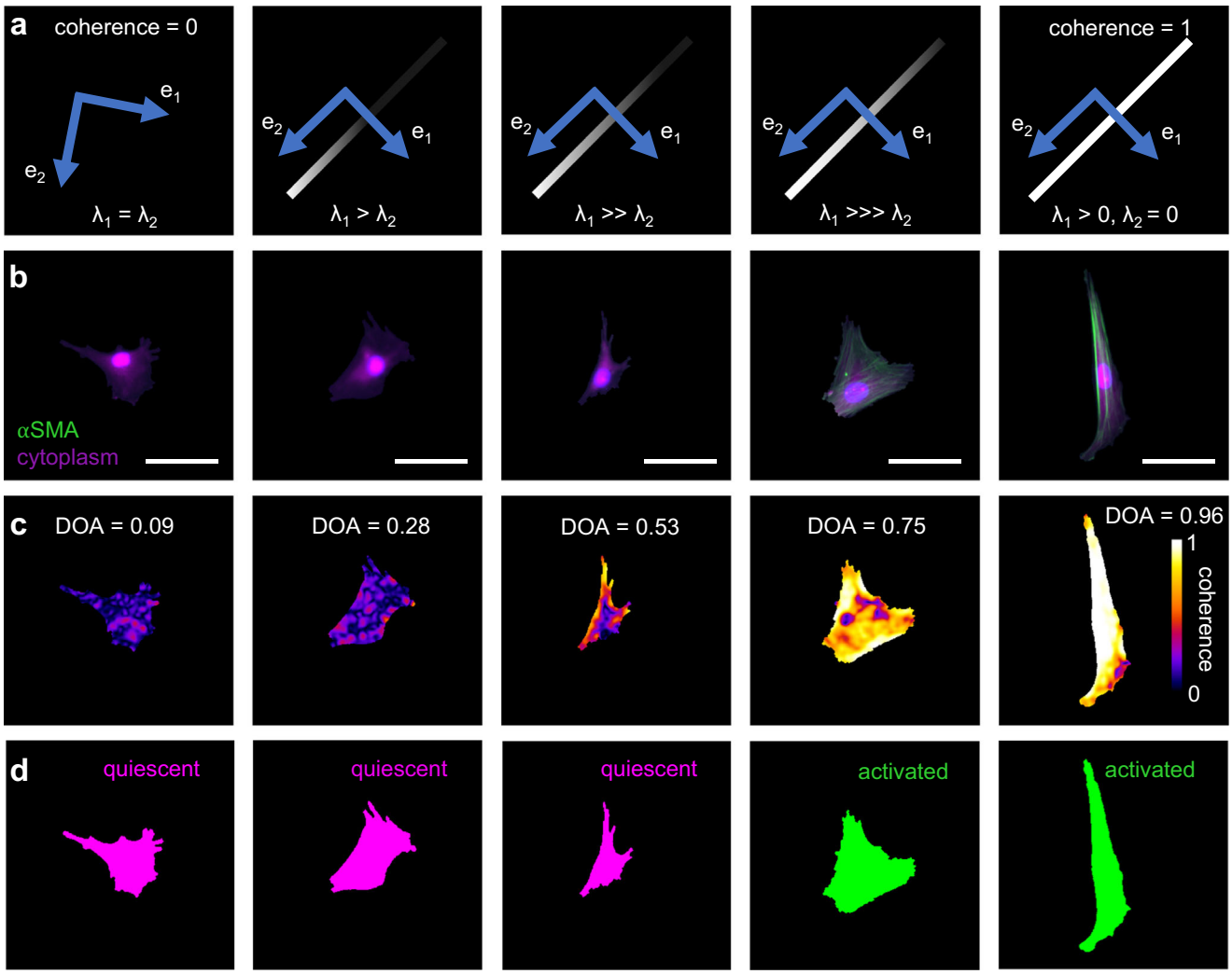

Fig. 2 | Quantification of αSMA stress fiber architecture. a Schematic demonstrating the computation of coherence. Coherence, at a point, increases as the greatest intensity gradient (λ_1) becomes greater than the orthogonal intensity gradient (λ_2). For a fiber, λ_1 is oriented along the cross-fiber direction ($\mathbf{e}_1$) whereas λ_2 is oriented along the preferred fiber direction ($\mathbf{e}_2$). In the case of homogenous intensity values (left most panel), the directions for $\mathbf{e}_1$ and $\mathbf{e}_2$ are random but orthogonal. **b** Representative immunostaining images depicting VICs with increasing levels of αSMA expression and anisotropy from left to right. **c** Degree of anisotropy (DOA) is computed for the VICs shown in panel (**b**) by taking the median of the coherence values at pixel locations comprising the VIC body. **d** Manual grading of the VICs shown in panel (**b**) as quiescent or activated. DOA analysis was performed for every cell used in this study (> 3000). Stained images: green - αSMA, magenta – cytoplasm. Scale bars: 50 μm.

regions with no αSMA stress fibers (Fig. 2b, c). Since *c* is computed on a per-pixel basis, we sought a single number to represent the overall degree of anisotropy (DOA) of a given cell. To this end, the median *c* value was computed for each VIC to represent the DOA. We then compared a currently prevalent manual grading system of activated and quiescent VICs with our calculated DOA. High values of DOA correspond with activated VICs, while quiescent VICs have low values of DOA (Fig. 2c, d), thus providing confidence that DOA aligns with traditional grading of VIC phenotypes while offering a refined and quantitative approach. Collectively, this quantitative method characterizes αSMA stress fiber expression and offers an alternative to manual grading of VIC phenotypes into binary classifications.

Extraction of morphological features that correlate with VIC activation

Prior methods to quantify and describe cellular morphologies have relied on domain knowledge to select and manipulate raw data into intuitive and meaningful features such as cell area, perimeter, solidity, circularity, aspect ratio, and curvature[35]. In the method presented, experimenter subjectivity is minimized during the process of morphological feature extraction and its subsequent correlation to αSMA anisotropy which reflects the activated phenotype. The methodology was first trained and validated on a dataset generated by seeding VICs on soft (E= ~5 kPa) (Fig. 3a) and stiff (E= ~13 kPa) (Fig. 3b) hydrogel matrices that mimic healthy and fibrotic valve tissue matrix and are known to induce quiescent and activated VIC phenotypes, respectively[47–49]. A single metric, the degree of anisotropy (DOA), was used to quantify the amount of αSMA organized into discrete stress fibers present within the VICs which reflects their level of myofibroblast activation (Fig. 2b, c). Next, feature vectors were generated for every cell, which were comprised of the Fourier coefficients describing the cell body and nuclei and the associated DOA value (Supplemental Table 1). The feature vectors were standardized and concatenated into a single array, and then principal component analysis (PCA) was performed. PCA resulted in principal component (PC) axes in which the VICs were the most variant. Visualization of the eigenshapes, which show the greatest shape variation along orthogonal PC axes, revealed

https://doi.org/10.1038/s41467-025-58082-0

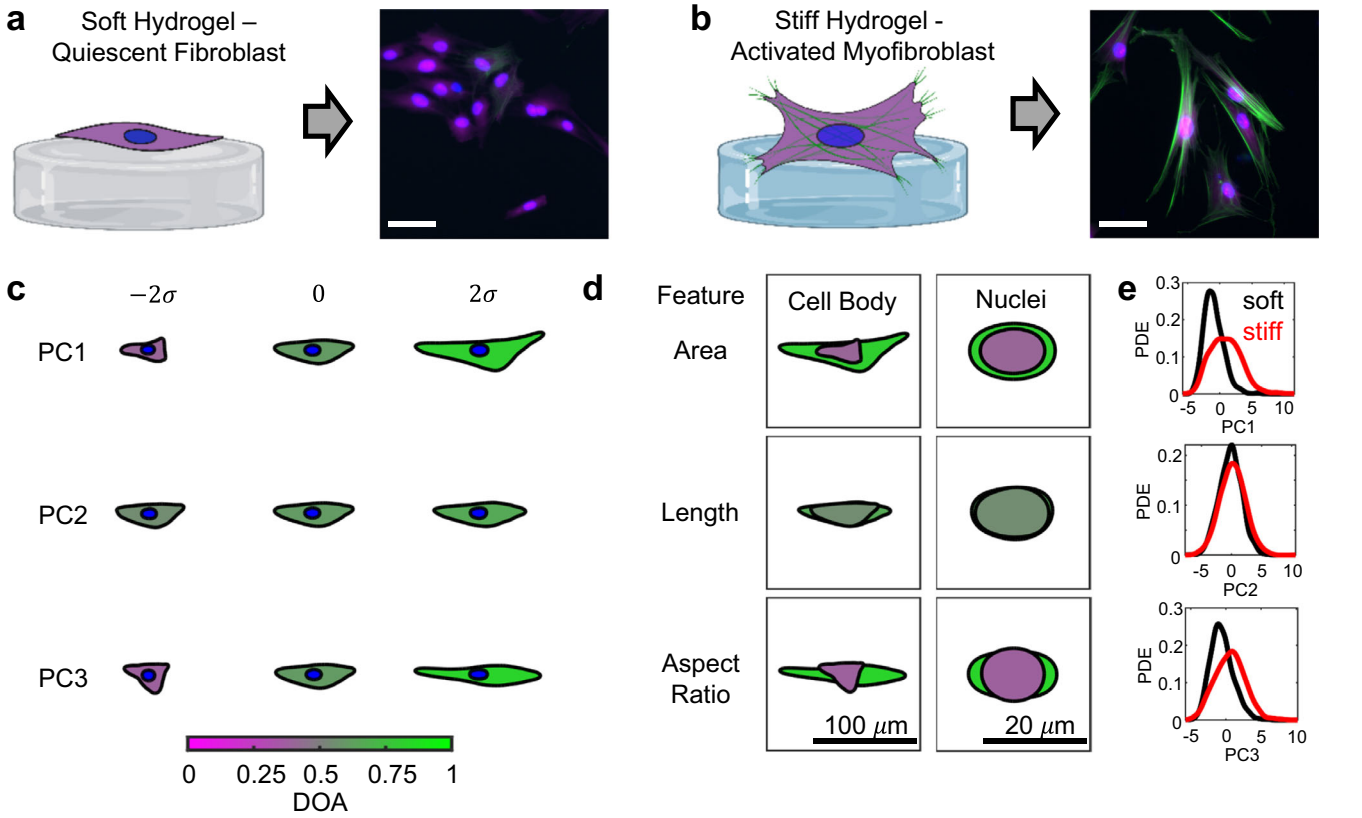

Fig. 3 | An approach combining morphological and αSMA expression data to reveal relevant morphological features to demarcate quiescent and activated VICs seeded on engineered hydrogel materials with tunable mechanical properties. a VICs seeded on top of mechanically soft hydrogels exhibit a quiescent phenotype characterized by low expression of αSMA stress fibers. **b** VICs seeded on top of mechanically stiff hydrogels transition to an activated myofibroblast phenotype characterized by high expression levels of αSMA. **c** VIC body and nuclear eigenshapes computed along principal components (PCs) 1-3 and colored with the corresponding DOA values. **d** VIC body and nuclear eigenshapes at the extremes (-2σ and 2σ) are superimposed and colored with the corresponding DOA values for PCs 1-3 accompanied with a physical interpretation of the primary morphological feature that varies along each PC. **e**, Probability density estimates (PDEs) depicting the PC score distributions for VICs seeded on soft (black) and stiff (red) hydrogels for PCs 1–3. Stained images: green - αSMA, magenta – cytoplasm, blue-nuclei. Scale bars: 50 μm. Schematics created in BioRender. Khang, A. (2025) https://BioRender.com/r87h898. Source data are provided as a Source Data file.

morphological features that co-varied with DOA (Fig. 3c). This process allowed for physical interpretation of morphological features that are most indicative of the activated phenotype. For example, we interpreted the morphological features of PCs 1, 2, and 3 as area, length, and aspect ratio, respectively (Fig. 3d). For clarity, the Fourier series expansion is not needed to quantify these morphological metrics. The Fourier series coefficients encode information regarding fibroblast morphology and are used in exploratory data analysis to visualize and interpret morphological changes that occur along the activation spectrum. We computed the probability density estimates (PDE) for VICs seeded on soft and stiff hydrogels for every PC and found that on stiff, activation-promoting hydrogels, VICs showed larger PC scores along PCs 1 and 3 compared to VICs seeded on soft, quiescence-promoting hydrogels (Fig. 3e). No appreciable difference was observed between the soft and stiff condition for PC 2. To assess the practical significance of the observed differences, the effect size for each PC axis was represented using Cohen's d[50,51]. A Cohen's d value of 0.94 was observed between the soft and stiff condition along PC1 (large to very large effect size) and 0.61 along PC3 (medium to large effect size, Supplemental Table 2). Although seeding VICs on soft and stiff hydrogels resulted in statistically significant differences in PC scores along some of the other PC axes ($p < 0.001$, Supplemental Table 2), the effect sizes were small (Cohen's $d < 0.5$, lower than medium effect size) indicating that the differences, while statistically significant, may not be of practical importance and were therefore not chosen for further analysis. Taken together, these results suggest cell and nuclear area and aspect ratio are morphological features that are correlated with VIC activation whereas cell and nuclear length are not (Fig. 3e). Furthermore, there was no significant correlation observed between cell area and aspect ratio ($r = 0.16$) nor between nuclear area and aspect ratio ($r = -0.10$), suggesting that they are orthogonal morphological features that should be considered separately. We observed experimentally imaged cellular morphologies (Supplemental Fig. 3) that aligned closely with the eigenshapes generated for PCs 1-3 (Fig. 3c). This finding provides confidence that the current approach produces realistic eigenshapes and highlights its relevance toward experimentally observed cellular geometries. In addition, the identified cells underline the innate heterogeneity of VICs observed in vitro.

Selection and validation of features that can grade VIC phenotypes

PCA and generation of eigenshapes along the PC axes (Fig. 3) identified that cell and nuclear area and aspect ratio covaried with DOA and were potential morphological features that could be used to grade quiescent and activated VICs (Fig. 4a). For each identified feature and DOA variable, logistic regression models were developed to grade VIC phenotypes based on morphology and αSMA anisotropy, respectively (Fig. 4b). In addition, we trained a multivariable

https://doi.org/10.1038/s41467-025-58082-0

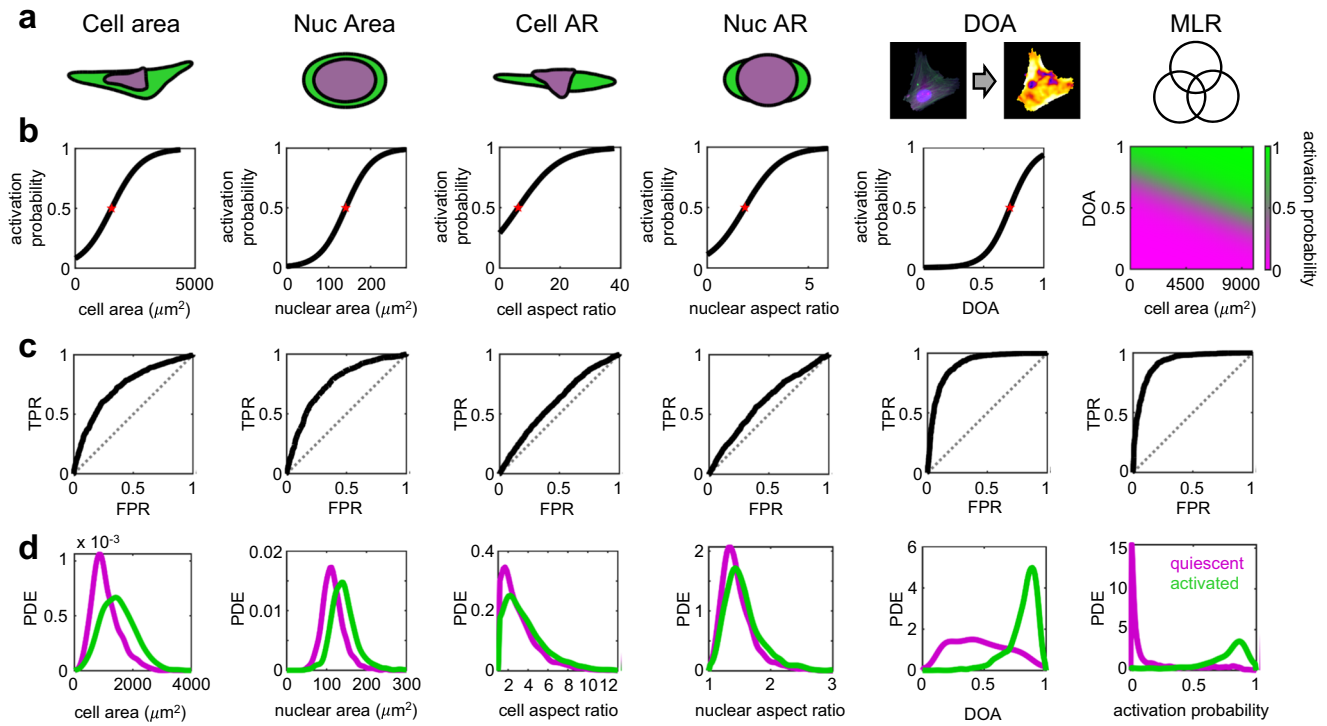

Fig. 4 | Efficacy of selected features toward demarcating quiescent and activated VICs induced by seeding on soft and stiff hydrogels. a Visualization of selected features that most likely vary between quiescent and activated VICs. The unity symbol denotes a multivariable logistic regression (MLR) model which considers all selected features. **b** Logistic regression curves depicting the activation probability of VICs versus the selected features. The red stars denote the quantitative threshold value where the activation probability of VICs ≥ 0.5. **c** Receiver operating characteristic curves generated by plotting the true positive rate (TPR) against the false positive rate (FPR) at ever-increasing threshold values of the selected features. The ROC curve for a random classifier is shown by the gray-dotted line. **d** Probability density estimates (PDEs) depicting the distributions of the selected features for quiescent (magenta) and activated (green) VICs. From left to right, the columns represent results for cell area, nuclear area, cell aspect ratio, nuclear aspect ratio, DOA, and the MLR model. Source data are provided as a Source Data file.

logistic regression (MLR) model that combined the effects of each identified feature and DOA. For visualization of the MLR model, we varied the parameters cell area and DOA, which differed greatly between quiescent and activated VICs (Supplemental Table 3), while keeping the remainder of the features constant and equal to the mean value observed for the VICs analyzed in the present study. Logistic regression models were employed due to their simplicity and interpretability. Specifically, the coefficients of the trained models inform on the appropriate threshold value to grade VICs for each identified feature (Supplemental Table 4). Moreover, the coefficients of the MLR model inform on the relative contribution or importance of each feature in a combined approach to grade VICs (Supplemental Table 5). Grading was accomplished by setting a threshold value (red star in Fig. 4b, Supplemental Table 4) for each feature that was defined by an activation probability ≥ 50%. To assess the performance of individual binary classifiers at varying threshold values, receiver operating characteristic (ROC) curves were generated for the features and the MLR model (Fig. 4c). ROC curves allowed for visual inspection of the area underneath the curve (AUC) which serves as a single quantitative descriptor of the predictive capability of a given variable. The AUC for cell area (0.73), nuclear area (0.76), DOA (0.90), and the MLR model (0.92) were high, thus identifying these as suitable variables for grading VIC phenotypes. In contrast, the AUC for cell (0.60) and nuclear (0.60) aspect ratio were relatively lower (Supplemental Table 6) and thus less predictive of VIC phenotypes. Of further note, activated VICs had significantly larger cell and nuclear areas and higher values of DOA (Fig. 4d), which all showed Cohen's d values greater than 0.8 (large to very large effect sizes, Supplemental Table 3). The largest effect size was observed in the MLR model (Cohen's d = 2.22, huge effect size). No appreciable differences were observed between activated and quiescent VICs in terms of cell and nuclear aspect ratios (Fig. 4d) and their effect sizes were considered small to medium with a Cohen's d less than 0.5 (Supplemental Table 3). Overall, the training of logistic regression models allows for the grading of VICs as quiescent or activated based on morphological characteristics, αSMA anisotropy, or both. Extended analysis with ROC curves assesses the overall performance of each model and showed that some features (cell area, nuclear area, DOA, and the MLR model) are better predictors of VIC phenotype than others (cell and nuclear aspect ratio). However, all features performed better than a random classifier which is denoted by the gray-dotted lines in Fig. 4c.

To validate the performance of the logistic regression models generated for each feature (Fig. 4b), we compared their predictions with manual phenotypic grading of fluorescent imaging datasets comprised of VICs stained for their cytoplasm, nuclei (DAPI), and αSMA stress fibers (Fig. 5a). A k-fold cross validation (k = 10) was used, in which the data was split into 10 random groups with 9 groups used to train the logistic regression models and the 10th serving as the validation data set (Fig. 5b). This process was iterated 10 times so that each group served as the validation data set exactly once. The percentage of cell phenotypes that were correctly predicted in the validation data set by the logistic regression models was recorded for every iteration (Fig. 5c). The validation showed that cell and nuclear area are good predictors of VIC phenotype with an accuracy of 69 ± 3% and 70 ± 3%, respectively; however, the cell and nuclear aspect ratios were less predictive with an accuracy of 60 ± 3% and 61 ± 3%, respectively. The DOA variable had the highest

https://doi.org/10.1038/s41467-025-58082-0

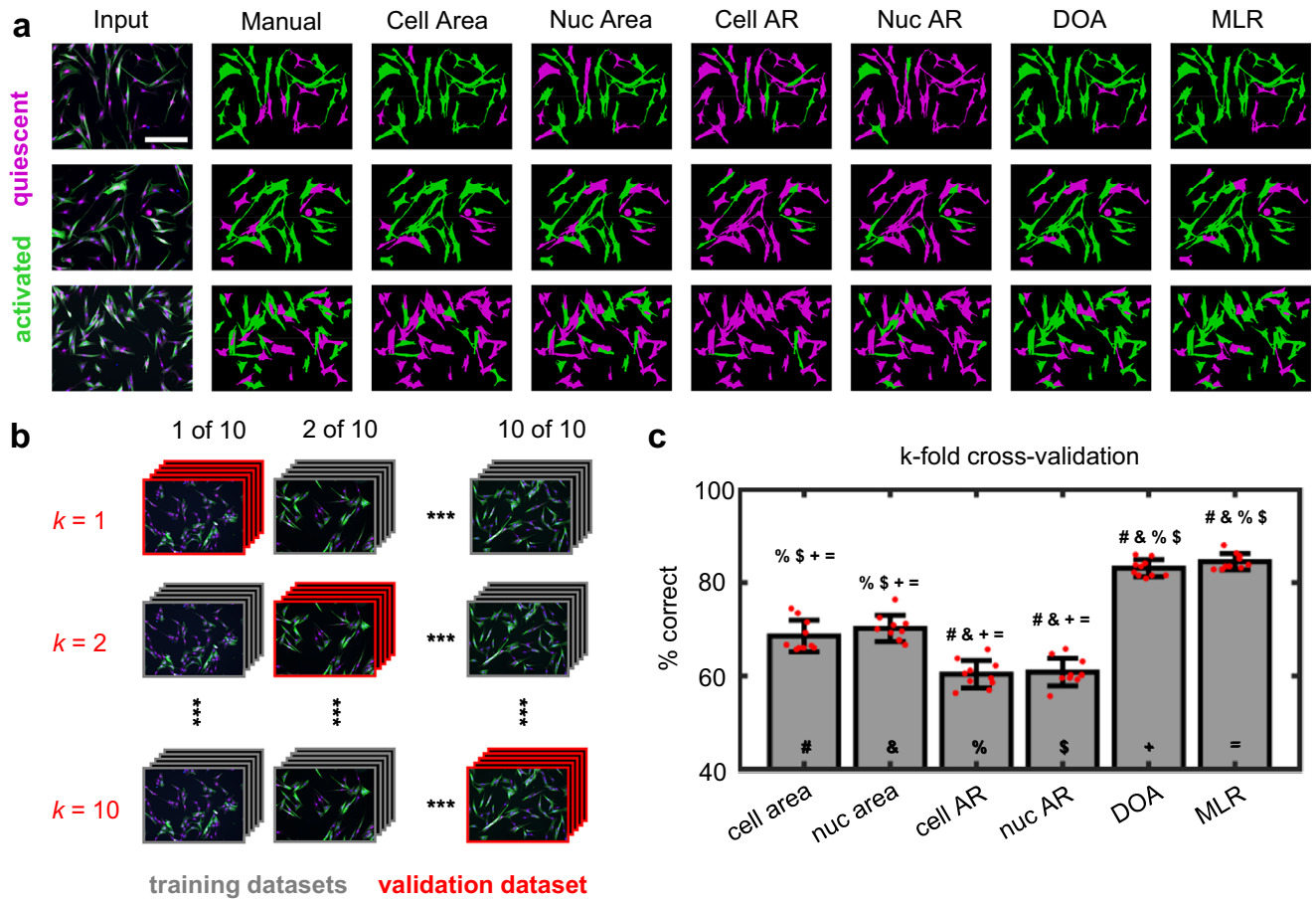

Fig. 5 | Validation of selected features toward grading activated and quiescent VICs. a Three representative examples of the grading process depicting the input image, manual grading, grading done by quantitative thresholding of selected features (cell area, nuclear area, cell aspect ratio (cell AR), nuclear aspect ratio (nuc AR), and DOA) and grading done via a multivariable logistic regression (MLR) model that combines information from all the selected features. Quantitative thresholds were determined via logistic regression. The grading process was performed for every image used in this study (> 200 images). **b** To assess the predictive capabilities of the selected features, *k*-fold cross-validation was performed in which the dataset was randomly partitioned into *k* equal sized subsets. Of the *k* subsets, *k-1* subsets were used to train logistic regression models and the remaining subset was used as the validation subset. The cross-validation process was repeated *k* times, with each of the *k* subsets serving as the validation subset exactly one time. For this analysis, $k = 10$. **c** Bar plots showing compiled *k*-fold cross-validation results depicting the average percentage of VICs graded correctly for each selected feature and the MLR model when compared to manual grading ($k = 10$). The error bars denote standard deviation. Each feature is represented with a symbol. The symbols on top of each bar plot denote the features that are significantly different ($p < 0.05$) from the current feature as determined by a repeated measures one-way ANOVA with Bonferroni correction applied for multiple comparisons ($p = 2.12e\text{-}31$, $F(5) = 243.75$, $\eta^2 = 0.96$). Exact *p*-values for multiple comparisons are provided in Appendix I. Stained images: green - αSMA, magenta – cytoplasm, blue-nuclei. Scale bars: 200 μm. Source data are provided as a Source Data file.

predictive capability among the single variables tested with an accuracy of 83 ± 2%. Beyond assessing the predictive capabilities of individual features, we also assessed the combination of all features on predicting VIC phenotypes using a MLR model (Supplemental Table 5). Overall, the MLR model showed the best predictive capabilities with an accuracy of 84 ± 2%; however, combinations of features did not result in any significant improvements in the predictions compared to the DOA variable alone.

Performance of models on test dataset

The logistic regression models, which were trained on VICs seeded on soft and stiff hydrogels, were employed for drug screening purposes and used to grade VICs from a test dataset where VICs were seeded on top of stiff, activation-promoting hydrogels and treated with the small molecule drug 5-azacytidine (AZA)[52] (Fig. 6a). AZA has been reported to be protective against fibrosis by increasing the expression of phosphatase and tension homolog (PTEN)[53–57]. The logistic regression models for cell area, nuclear area, DOA, and the MLR model showed excellent performance in predicting the phenotype of VICs in the AZA dataset with accuracies of 74 ± 4%, 72 ± 3%, 74 ± 6%, and 83 ± 4%, respectively (Fig. 6b, c). The logistic regression models for cell and nuclear aspect ratio showed reduced performance with an accuracy of 61 ± 1% and 55 ± 3% in the AZA dataset, respectively. Based on manual grading, 68 ± 11% of VICs seeded on stiff hydrogels were activated, 13 ± 8% of VICs seeded on soft hydrogels were activated, and 27 ± 5% of VICs seeded on stiff hydrogels with AZA treatment were activated (Fig. 6d). Of note, the predictions made by the MLR model on the soft and stiff training dataset as well as on the AZA test dataset largely resembled manual grading (stiff - 76 ± 6% activated; soft - 9 ± 10% activated; AZA - 29 ± 5% activated) and were statistically indistinguishable as determined by paired t-tests ($p > 0.05$), highlighting the ability of the current approach to recapitulate traditional methods while employing quantitative and data-driven methodologies. These results reflect that the AZA treatment group more closely resembles VICs on soft hydrogels than VICs on stiff hydrogels in terms of the percentage of activated cells observed, suggesting that AZA may be useful toward protecting against fibrosis by inhibiting myofibroblast activation.

As an additional drug screening example, we used the logistic regression models to grade VICs from a test dataset where VICs were

https://doi.org/10.1038/s41467-025-58082-0

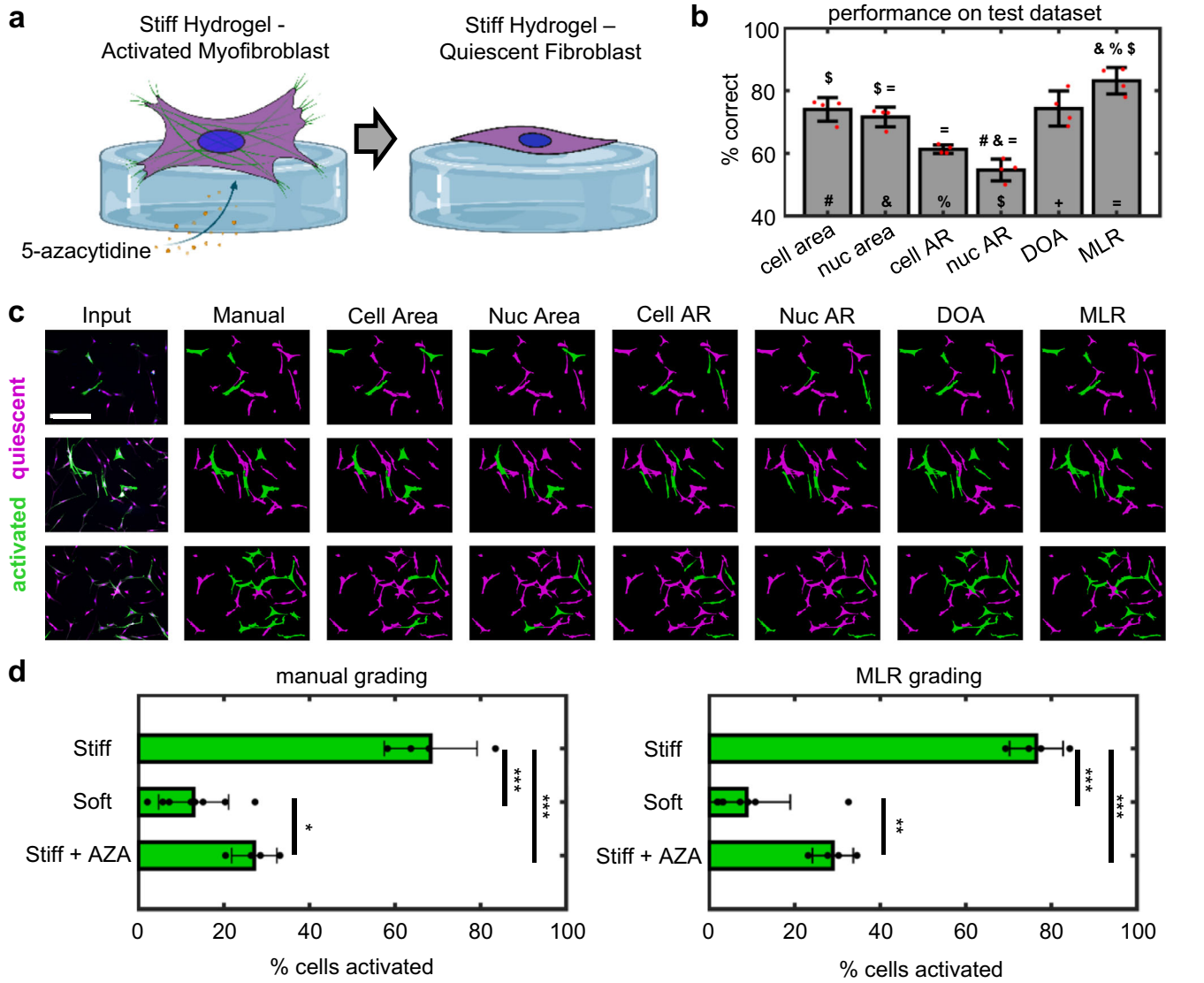

Fig. 6 | Model performance on drug screening test dataset: 5-azacytidine. **a** VICs were seeded on top of stiff, activation-promoting hydrogel surfaces and treated with the small molecule drug 5-azacytidine (AZA), which inhibits myofibroblast activation. **b** The average percentage of VICs in the test datasets graded correctly for each selected feature and the MLR model as compared to manual grading ($n = 4$ replicates from cells pooled from 10 pig hearts). The error bars denote standard deviation. Each feature is represented with a symbol. The symbols on top of each bar plot denote the features that are significantly different ($p < 0.05$) from the current feature as determined by a repeated measures one-way ANOVA with Bonferroni correction applied for multiple comparisons ($p = 1.88e\text{-}9$, $F(5) = 61.36$, $\eta^2 = 0.95$). Exact p-values for multiple comparisons are provided in Appendix I. **c** Three representative examples of the grading process depicting the input image, manual grading, grading done by quantitative thresholding of cell area, nuclear area, cell aspect ratio, nuclear aspect ratio, DOA and the MLR model. **d** The percentage of activated VICs when seeded on stiff hydrogels ($n = 4$), soft hydrogels ($n = 8$), and stiff hydrogels treated with AZA ($n = 4$) as determined by manual grading and the MLR model. Statistically significant differences ($p < 0.05$) were observed among all groups in manual grading ($p = 2.91e\text{-}7$, $F(2) = 59.32.13$, $\eta^2 = 0.90$) and grading done via the MLR model ($p = 2.93e\text{-}8$, $F(2) = 87.24$, $\eta^2 = 0.93$) as assessed by a one-way ANOVA with Bonferroni correction applied for multiple comparisons. No statistically significant differences were found between manual grading and grading performed by the MLR model for the soft ($p = 0.08$, $t(7) = 2.08$, Cohen's D [95% CI] = 0.45 [−0.54 1.44]), stiff ($p = 0.06$, $t(3) = -2.87$, Cohen's D [95% CI] = 0.92 [-0.54, 2.38]), and AZA conditions ($p = 0.56$, $t(3) = -0.65$, Cohen's D [95% CI] = 0.36 [−1.03 1.76]) as determined by two-sided, paired t-tests. The bars represent the mean and the error bars denote standard deviation. Stained images: green - αSMA, magenta – cytoplasm, blue-nuclei. Scale bars: 200 μm. Schematics created in BioRender. Khang, A. (2025) https://BioRender.com/t10b968. Source data are provided as a Source Data file.

seeded on top of soft, quiescence-promoting hydrogels and treated with the small molecule drug (bpV(HOpic))[58] (Fig. 7a). The compound bpV(HOpic) has been reported to inhibit the expression of PTEN and we hypothesized that it would promote myofibroblast activation even under quiescence-promoting conditions. The logistic regression models for DOA and the MLR model showed excellent performance in predicting the phenotype of VICs in the bpV(HOpic) dataset with accuracies of 78 ± 2% and 81 ± 3%, respectively (Fig. 7b, c). The logistic regression models for cell area, nuclear area, cell aspect ratio, and nuclear aspect ratio showed reduced performance with accuracies of 68 ± 2%, 69 ± 5%, 64 ± 7%, and 63 ± 7%, respectively. Based on manual grading, 34 ± 6% of VICs seeded on soft hydrogels with bpV(HOpic) treatment were activated (Fig. 7d). Of note, the predictions made by the MLR model on the bpV(HOpic) test dataset resembled manual grading (27 ± 10% activated) and were statistically indistinguishable as determined by a paired t-test ($p > 0.05$), further highlighting the ability of the current approach to recapitulate traditional grading methods. These results reflect that

https://doi.org/10.1038/s41467-025-58082-0

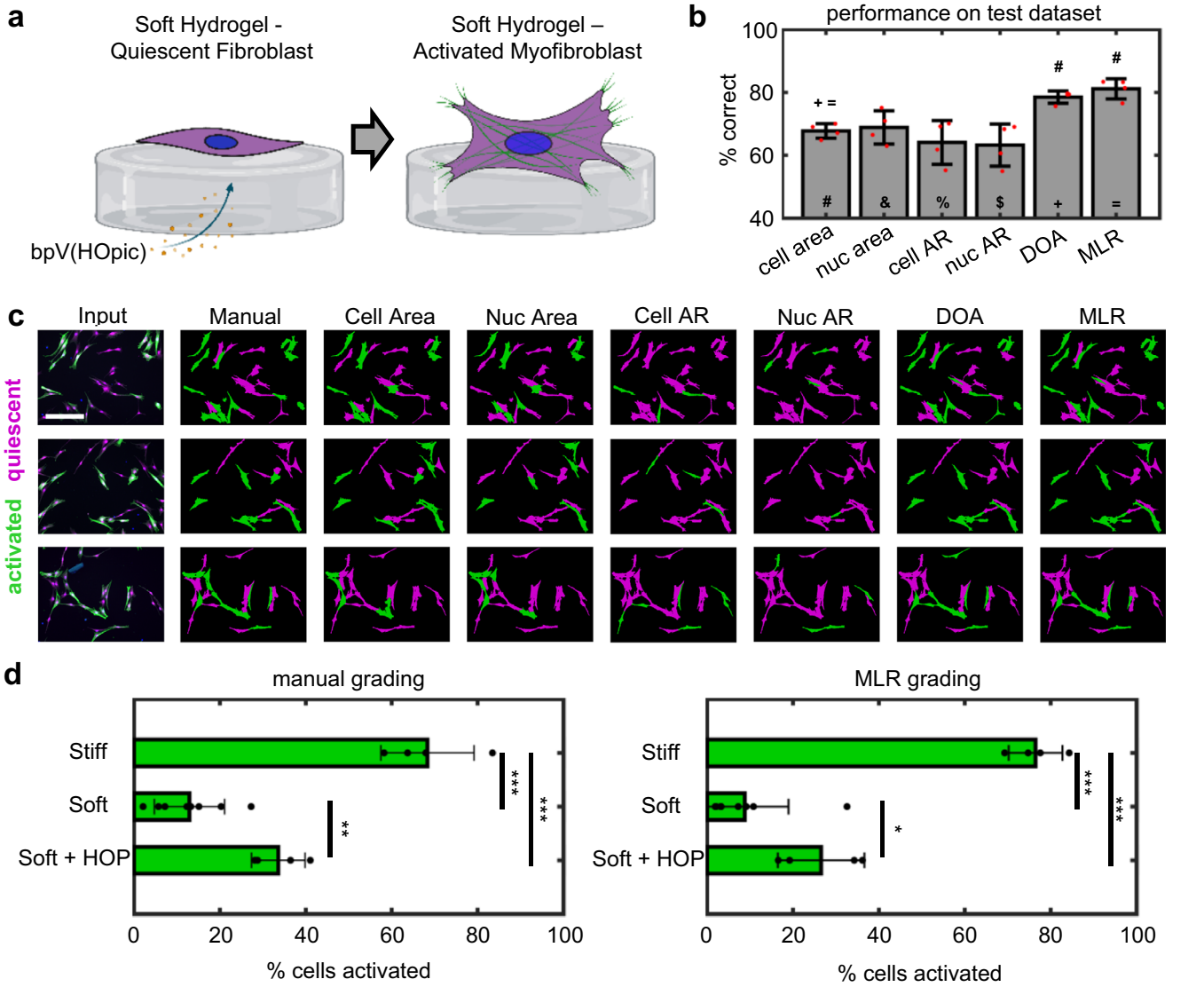

Fig. 7 | Model performance on drug screening test dataset: bpV(HOpic). a VICs were seeded on top of soft, quiescence-promoting hydrogel surfaces and treated with the small molecule drug bpV(HOpic), which promotes myofibroblast activation. **b** The average percentage of VICs in the test datasets graded correctly for each selected feature and the MLR model as compared to manual grading ($n = 4$ replicates from cells pooled from 10 pig hearts). The error bars denote standard deviation. Each feature is represented with a symbol. The symbols on top of each bar plot denote the features that are significantly different ($p < 0.05$) from the current feature as determined by a repeated measures one-way ANOVA with Bonferroni correction applied for multiple comparisons ($p = 1.66e\text{-}7$, $F(5) = 32.18$, $\eta^2 = 0.91$). Exact p-values for multiple comparisons are provided in Appendix I. **c** Three representative examples of the grading process depicting the input image, manual grading, grading done by quantitative thresholding of cell area, nuclear area, cell aspect ratio, nuclear aspect ratio, DOA and the MLR model. **d**, The percentage of activated VICs when seeded on stiff hydrogels ($n = 4$), soft hydrogels ($n = 8$), and soft hydrogels treated with bpV(HOpic) (HOP, $n = 4$) as determined by manual grading and the MLR model. Statistically significant differences ($p < 0.05$) were observed among all groups in manual grading ($p = 3.7e\text{-}7$, $F(2) = 56.94$, $\eta^2 = 0.90$) and grading done via the MLR model ($p = 1.14e\text{-}7$, $F(2) = 69.57$, $\eta^2 = 0.91$) as assessed by a one-way ANOVA with Bonferroni correction applied for multiple comparisons. No statistically significant differences were found between manual grading and grading performed by the MLR model for the soft ($p = 0.08$, $t(7) = 2.08$, Cohen's D [95% CI] = 0.45 [−0.54 1.44]), stiff ($p = 0.06$, $t(3) = -2.87$, Cohen's D [95% CI] = 0.92 [-0.54, 2.38]), and HOP conditions ($p = 0.07$, $t(3) = 2.80$, Cohen's D [95% CI] = 0.84 [−0.61 2.29]) as determined by two-sided, paired t-tests. The bars represent the mean and the error bars denote standard deviation. Stained images: green - αSMA, magenta – cytoplasm, blue-nuclei. Scale bars: 200 μm. Schematics created in BioRender. Khang, A. (2025) https://BioRender.com/z65q983. Source data are provided as a Source Data file.

VICs on soft hydrogels treated with bpV(HOpic) contained more activated VICs than when seeded on soft hydrogels alone, suggesting that bpV(HOpic) promotes myofibroblast activation. Taken together with our AZA results, both drug screens support that PTEN promotes VIC quiescence and is a druggable target with promise toward mitigating fibrosis. In addition to confirming the ability of the models to make predictions on unseen test datasets, these results showcase the utility of the present approach in drug-screening applications toward predicting the activation landscape of VICs in response to candidate anti-fibrotic drugs.

Benchmarking against pre-existing explicit shape analysis methods, a deep-learning method, and a high-dimensional cell phenotyping method

The Fourier shape reconstruction method used herein holds several advantages over previous methods of explicit shape analysis. To demonstrate these advantages, we benchmark the shape reconstruction method we used against previously published approaches. We benchmark against a variant form of Fourier series shape analysis[59], which converts shape boundary points to polar coordinates (r, θ), and parametrizes every radial value r with respect to the angle around the shape

https://doi.org/10.1038/s41467-025-58082-0

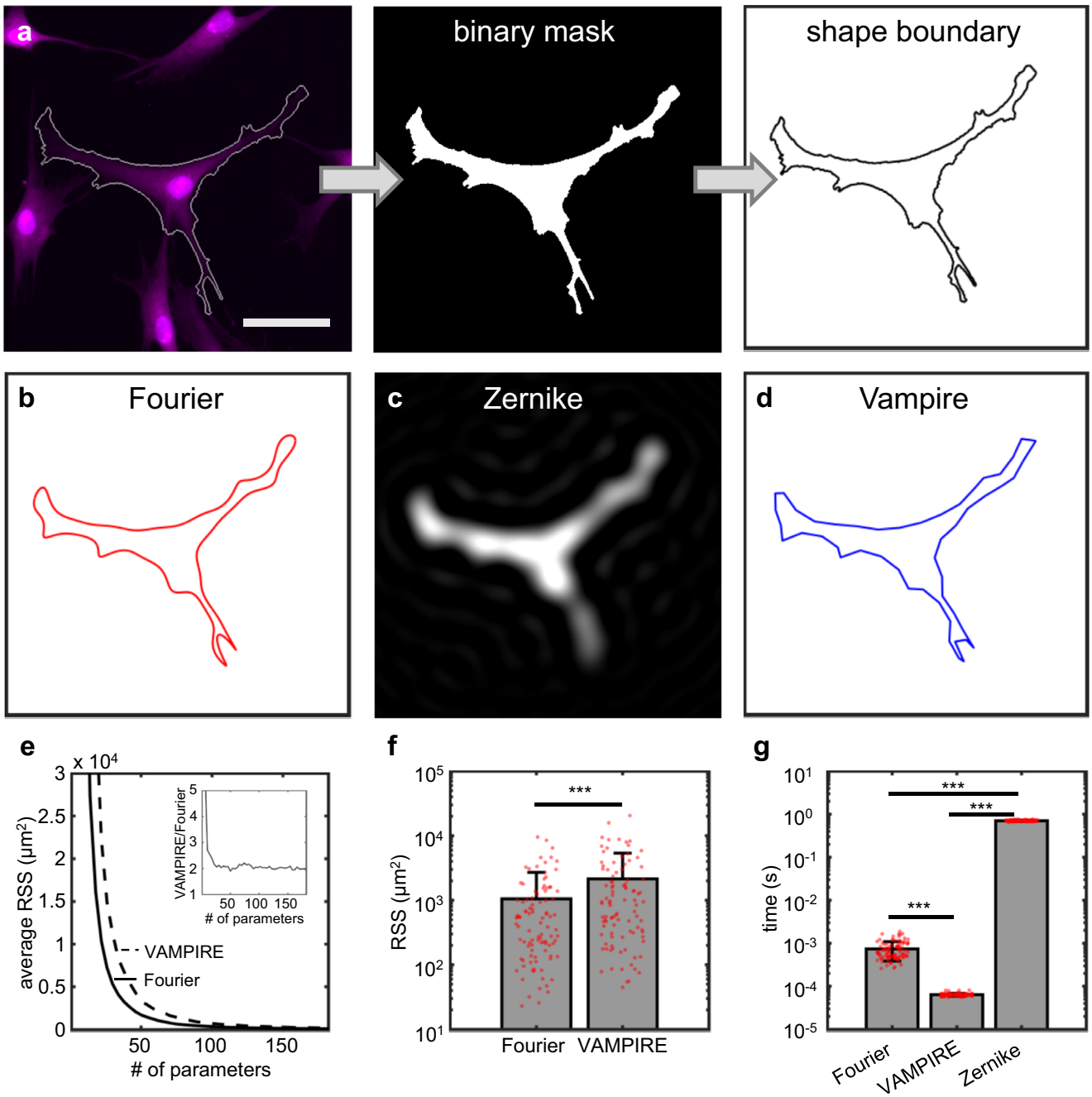

Fig. 8 | Benchmarking against alternative shape analysis methods. **a** Segmentation of a VIC from immunofluorescent images to achieve a binary mask and boundary representation. **b** Fourier series representation of the shape in (**a**) using 62 total model parameters (15 Fourier harmonics). **c**, Zernike polynomial representation of the shape in panel (**a**) using N = 45 degrees. **d** VAMPIRE representation of the shape in (**a**) using 62 total model parameters (31 equidistant *x*-*y* boundary points). **e** Average residual sum of squares (RSS) between the original and reconstructed shape plotted against the total number of parameters for Fourier series (solid) and VAMPIRE (dotted) representations of the same 100 cells assessed in Supplemental Fig. 1. The insert depicts the ratio of the average VAMPIRE RSS to the average Fourier series RSS with increasing number of parameters. Note that the Fourier series method consistently results in approximately half the error of VAMPIRE. **f** The average RSS for Fourier series and VAMPIRE representations using 62 total model parameters for n = 100 cells selected at random. The error bars denote standard deviation. The RSS for Fourier series representations were statistically less than that of VAMPIRE representations as assessed via a paired t-test (p = 9.07e-9, t(99) = −6.28, Cohen's D [95% CI] = 0.42 [0.14 0.70]). **g** The average time required to parameterize cell shapes into Fourier series, VAMPIRE, and Zernike polynomial representations for n = 100 cells selected at random. The error bars denote standard deviation. Statistically significant differences ($p < 0.05$) in time required to parameterize cell shapes were observed among all techniques as assessed by a repeated measures one-way ANOVA with Bonferroni correction applied for multiple comparisons (p = 2.11e-283, F(2) = 7.09e4, η^2 = 1.0). Stained images: magenta – cytoplasm and nuclei. Scale bars: 50 μm. Source data are provided as a Source Data file.

θ (Appendix D in Supplemental Information). In brief, we report that this variant of Fourier series analysis is suitable for simple contours and cannot adequately reproduce every VIC shape analyzed in this study. In contrast, the Fourier series shape analysis employed in the current study allows for the reconstruction of any closed contour, regardless of complexity. This is due to the differences in parameterization. Specifically, in our approach, the x and y portions of every boundary coordinate is parameterized independently as a function of the arc length

around the shape which ensures a one-to-one mapping. Other approaches that directly convert cartesian coordinates to polar coordinates do not ensure a one-to-one mapping, especially in the case of complex cell geometries, and introduces contradictions when solving for Fourier coefficients that profoundly limit shape reconstruction quality (Supplemental Fig. 9).

Next, we benchmarked our methodology against Zernike polynomial[60] and VAMPIRE[18] reconstructions (Fig. 8 and Appendix D in Supplemental Information). Reconstructions for a representative VIC shape (Fig. 8a) were developed using Fourier series (15 harmonics totaling 62 parameters, Fig. 8b), Zernike polynomials (45 moments totaling 1217 parameters, Fig. 8c), and VAMPIRE (31 points totaling 62 parameters, Fig. 8d) for comparison. In general, Zernike polynomial reconstructions perform well for small images (Supplemental Fig. 14) but result in blurry approximations of VIC shapes (Figs. 8c and Supplemental Fig. 15), even when a large number of parameters are used for shape reconstruction (Supplemental Figs. 16 and 17). VAMPIRE represents shapes using equidistant points along the boundary of the contour (Fig. 8d). We compute the error in shape representation between Fourier series analysis and VAMPIRE and find that Fourier series representations consistently resulted in approximately half the error ($1053 \pm 1650\ \mu m^2$ vs $2140 \pm 3257\ \mu m^2$ using 62 total parameters), regardless of the number of parameters used to generate the shape (Fig. 8e, f). Comparing the quality of reconstructions with Zernike polynomials is limited as they produce images rather than discrete boundaries. However, we compared the computation time required to parameterize shapes into Fourier series, VAMPIRE, and Zernike polynomial representations using the same number of parameters (62 total) and observed that Zernike polynomials require significantly more computation time per cell (0.70 ± 0.03 s) than Fourier series ($7.4e\text{-}4 \pm 3.5e\text{-}4$ s) and VAMPIRE ($6.4e\text{-}5 \pm 4.8e\text{-}6$ s) ($p \leq 0.001$ Fig. 8g). This difference is of practical importance when analyzing a large number of cellular shapes as total computation time increases considerably. We also observed that Fourier series parameterization requires more time than VAMPIRE parameterization, however the difference is of little practical significance considering both methods can perform parameterization in milliseconds or less. We also analyzed the training datasets comprised of VICs seeded on soft and stiff hydrogel substrates using Zernike polynomials and VAMPIRE to compare the eigenshapes generated with these methods with those produced using Fourier series (Appendix D in Supplemental Information, Supplemental Fig. 11–13 & 18–20). From visual inspection, we observed that Zernike polynomial eigenshapes roughly approximate those produced by Fourier series representations (Supplemental Figs. 18–20). Thus, we conclude that Zernike polynomials may produce similar eigenshapes as Fourier series representations but are limited by the need for a large number of parameters and significantly longer computational time that still ultimately result in blurry representations. This suggests that our method is more deterministic in eigenshape reconstruction and more computationally efficient, both of which are significant advantages. In addition, we observed that combining VAMPIRE representations with DOA for PCA enabled the generation of eigenshapes that closely resembled those generated using Fourier series. However, VAMPIRE eigenshapes did not display appreciable changes in DOA along the PC axes. This is in direct contrast to results obtained using Fourier series (Fig. 3c) and Zernike polynomials (Supplemental Fig. 18–20). This finding suggests that VAMPIRE representations result in greater variation in z-scores among shape representation parameters which dominates over the variation arising from differences in αSMA stress fiber expression. This suggests that VAMPIRE may not be the most suitable tool to grade fibroblast phenotypes by combining morphological and αSMA expression information.

Deep learning approaches have been employed to phenotype several cell types, including cancer cells[16], microglia[15], and fibroblast cells[14]. As phenotypic markers and signatures vary between cell types, it is difficult to design a holistic model that can phenotype them all. In addition, it is difficult to achieve an 'apples to apples' comparison of the performance of models that have been designed for different cell types. With this in mind, we benchmark our approach against a deep learning model that was designed to classify quiescent and activated cardiac fibroblasts (CFs) based on αSMA stress fiber expression[14] (Appendix E in Supplemental Information). This comparison serves as external validation and enables us to extend our methodology to analyze an externally sourced dataset that was also independently graded[61]. In addition, we showcase the versatility of our approach on a different fibroblast type and on data gathered at another institution using different microscopy hardware. The dataset is comprised of fluorescent images of CFs seeded on hydrogel matrices with varying moduli to achieve a spectrum of quiescent to activated phenotypes (Supplemental Fig. 21a). The authors stained for nuclei (blue), f-actin (red), and αSMA (green). The f-actin and nuclei channels were used to segment the nuclei and cell body using a water-shed algorithm performed in the software CellProfiler[62]. The same analysis that was done for our VIC dataset such as Fourier series shape reconstruction, DOA computation, PCA, eigenshape generation, logistic regression model training, and *k*-fold cross-validation was performed on this CF dataset (see methods). In summary, we observed that an MLR model developed using our approach resulted in increased accuracy over the deep learning model ($77 \pm 2\%$ vs 67%). Thus, we conclude that our method has the potential to outperform a deep learning approach to quantify quiescent and activated fibroblasts, while doing so in a simple and interpretable manner.

Lastly, we benchmark our approach against a high-dimensional cell phenotyping software, CellProfiler[62] (Appendix F in Supplemental Information). CellProfiler is a powerful tool toward quantifying differences in cell state in high-throughput studies and has been combined with specific fluorescent dyes to augment morphological profiling of cells such as U2OS, A549, MCF-7, 3T3, and HeLA in a process termed Cell Painting[62–65]. CellProfiler has also augmented the study of cellular responses to chemical and genetic pertubations[64,66]. In brief, CellProfiler is capable of extracting hundreds to thousands of features which quantify the morphology and fluorescence of cells within immunostaining images. We inputted the cell body and nuclei segmentations from our VIC training dataset into CellProfiler and obtained morphological and fluorescence descriptors of fluorescence images comprised of the VIC cytoplasm (CellMask), nuclei (DAPI), and αSMA using the modules "MeasureObjectSizeShape", "MeasureObjectIntensity", and "MeasureTexture". This resulted in 302 features for both the cell body and nuclei of VICs (604 features total per cell). Next, we standardized and performed principal component analysis on these features which revealed that the greatest variation in the dataset along PC axis 1 were attributable to differences in αSMA staining (Supplemental Table 13). This observation is in line with the variation in DOA we observed along PC1 (Fig. 3c). For the remainder of the analysis, we focused on benchmarking the top 5 variables affecting PC1 as determined from CellProfiler against our measure, DOA, as they all focus on quantifying αSMA stress fiber expression. The logistic regression models for var1-5 showed fair performance in predicting the phenotype of VICs in the test dataset with accuracies of $71 \pm 6\%$, $70 \pm 6\%$, $68 \pm 6\%$, $68 \pm 5\%$, and $70 \pm 6\%$, respectively, with DOA resulting in the best accuracy ($74 \pm 6\%$, Supplemental Figs. 26b and c). These results suggest that DOA may outperform metrics from CellProfiler in quantifying αSMA expression. This is likely because DOA better quantifies αSMA organized into stress fibers than variables such as max intensity, standard deviation, and entropy that were identified by CellProfiler (Supplemental Table 13). Broadly, this suggests that our methodology is a more direct and accurate approach to quantitatively grade fibroblast phenotypes and requires substantially fewer total dimensions to do so (85 vs 604).

Discussion

Traditional methods of grading myofibroblast activation have relied on manual and visual inspection for the presence of discrete αSMA stress fibers, with morphological characteristics being of secondary importance if not ignored altogether. Moreover, previous approaches have relied on subjective grading based on domain knowledge and qualitative assessment. Other studies that quantify fibroblast phenotypes commonly rely on metrices computed using absolute αSMA fluorescence intensity, which largely depends on sample preparation methods, hardware used (e.g., microscope), and imaging parameters. In the present study, we develop a method that can leverage and combine morphological and αSMA stress fiber information derived solely from fluorescent images to demarcate quiescent fibroblasts from activated myofibroblasts. The use of the DOA variable allows for a quantitative method to characterize discrete αSMA stress fiber expression, and therefore phenotypic activation, through measuring their anisotropy. The use of Fourier series to obtain coefficients to represent cell and nuclear morphological data in an abstract manner allows for data-driven approaches to discover morphological features that are correlated with myofibroblast activation. Of note, the current method is intended for use on images that are routinely taken to assess fibroblast phenotypes and requires no additional experimental effort.

DOA performed the best among the individual variables used to predict VIC phenotypes (Fig. 5c). This finding is expected as DOA reflects the presence of discrete αSMA stress fibers, which is the main characteristic that is used to manually grade fibroblast activation. However, the present standard of practice relies on manual and visual assessment to detect αSMA fibers, which is not practical for high-throughput studies that can generate thousands of images of fibroblasts[32–34]. Although automated approaches exist to measure αSMA intensity, they do not measure the organization of αSMA into stress fibers, which is required to differentiate myofibroblasts from quiescent fibroblasts[9,11,12]. An advantage of the current approach is the utility of the DOA variable to quantify the organization of αSMA into stress fibers, which may reduce labor and eliminate individual bias in fibroblast phenotyping. DOA performed equally as well as the MLR model in *k*-fold cross validation (accuracy of 83 ± 2% vs 84 ± 2%, Fig. 5). However, when tested on unseen datasets, the multivariable logistic regression model performs considerably better than DOA alone (74 ± 6% vs 83 ± 4% in Fig. 6b and 78 ± 2% vs 81 ± 3% in Fig. 7b). Therefore, we conclude that the MLR model outperforms DOA on unseen datasets through combining both αSMA stress fiber information and cell morphology. In addition, the current method was able to reveal morphological features that are indicative of VIC phenotypic activation as evidenced by both cell and nuclear area proving comparable in accuracy to DOA when grading the AZA test dataset (Fig. 6b). In summary, the MLR model resulted in the best predictions on the test datasets (Fig. 6, 7), underscoring the utility of a quantitative grading scheme that combines both morphological and αSMA stress fiber information. Looking forward, we envision that the MLR model may be used to test the effectiveness of identified drugs by comparison to negative and positive controls (e.g., VICs seeded on soft and stiff hydrogels, respectively).

Traditionally, grading of fibroblast phenotypes is largely binary with quiescent and activated being the two main categories. In the context of the current study, individual features and a combined approach are used to perform binary grading to compare their predictive capabilities against the current gold standard: manual binary grading. However, fibroblast phenotypes likely reside on a spectrum. In appreciation of this fact, the developed methodology can also predict the activation probability of fibroblasts on a continuous scale (Supplemental Fig. 4). Looking forward, this capability could lead to the characterization of sub-phenotypes that lie between quiescence and phenotypic activation.

We recognize that a large body of pre-existing work has focused on phenotyping cells using methods such as explicit shape analysis[18,59,60], deep learning[14–16,67,68], and high-dimensional approaches[62–66]. An advantage of our approach lies in the use of Fourier series shape analysis which can mathematically describe any 2D contour, regardless of complexity (Appendix D in Supplemental Information). In addition, our method results in more accurate shape reconstructions and require fewer model parameters to do so compared to previous approaches[18,60] (Appendix D in Supplemental Information). Moreover, we demonstrate that our approach can perform equally as well and potentially outperform deep learning models and high dimensional phenotyping methods (Appendices E&F in Supplemental Information). In addition, our method makes accurate fibroblast phenotyping decisions that are based on an improved metric of αSMA stress fiber expression and physically interpretable morphological features.

Fibroblast cells naturally reside within 3D microenvironments and the current methods can be extended to analyze 3D imaging data. DOA can be easily modified to detect out of plane αSMA stress fiber anisotropy while spherical harmonics (the 3D extension of Fourier series) can be used to describe 3D cellular shapes[29,40]. Moreover, the current methods are not limited to analyzing VICs and can be used to study other fibroblast cell types, as we have demonstrated with cardiac fibroblasts (Appendix C in Supplemental Information). In addition, the current methods can be extended to segment and analyze bright field images, presenting the opportunity to quantify fibroblast phenotypes in real time and longitudinally to assess dynamic phenotypic shifts. Lastly, the presented methodology has potential to augment and improve research in the field of fibrosis through the analysis of cellular and nuclear morphologies derived from histology of healthy and diseased patient samples. This analysis could include parameters which describe nuclear activity such as the chromatin condensation parameter[49,69]. In summary, our work establishes a method to learn how fibroblast morphologies change with respect to phenotypic activation and leverages this knowledge to quantitatively grade fibroblast phenotypes. Looking forward, the developed methodology can complement high-throughput screening of small molecule drug libraries by efficiently pinpointing drugs that induce a quiescent fibroblast phenotype and provide rational for further testing of candidate anti-fibrotic agents.

Methods

The relevant animal protocols used in this study were approved by the Institutional Animal Care and Use Committee at the University of Colorado Boulder (Protocol #2351). The data analyzed in this study are available on a public repository (https://scholar.colorado.edu/concern/datasets/mw22v725h).

Seeding aortic valve interstitial cells on soft and stiff hydrogels

Primary porcine aortic valve interstitial cells (VICs) were isolated from dissected porcine aortic valve leaflets[70]. Porcine hearts were obtained and pooled from 10 pigs from Animal Biotech Industries and washed in Earle's balanced salt solution (Sigma) supplemented with 1 μg/ml amphotericin B (Gibco), 50 μg/ml streptomycin (ThermoFisher), and 50 U/ml of penicillin (ThermoFisher). Next, the leaflets were digested within a collagenase solution comprised of 250 U/ml of type II collagenase (Worthington) dissolved within Earle's balanced salt solution for 30 min. This primary digestion removed the valvular endothelial cells which were discarded. The leaflets were then transferred to a fresh collagenous solution and were incubated with gentle shaking for 60 min. Next, the solution containing the leaflets was vortexed for 5 min to help dislodge the VICs and was then poured through a 100 μm cell strainer to collect the VICs. The isolated VICs were then pelleted and seeded within expansion media comprised of M199 media (ThermoFisher) supplemented with 10% Fetal Bovine Serum (FBS)

(ThermoFisher), 1 μg/ml amphotericin B, 50 μg/ml streptomycin, and 50 U/ml of penicillin. The VICs were cultured until ~80% confluency and detached using trypsin (ThermoFisher) for passaging, freezing, and experimentation.

For our study, we pooled cells from 10 pig hearts to minimize individual biological variability and focus on differences driven by experimental conditions. Given the small size of valve tissues, isolating cells from a single heart yields a limited number of cells, which may not be sufficient for experimentation. Pooling allowed us to maximize cell yield while ensuring that observed effects were not biased by inter-animal variation. By standardizing cell contributions from each heart and processing all samples under identical conditions, we reduced the risk of batch effects and maintained experimental consistency. Additionally, pooling enhanced reproducibility by averaging out potential outlier effects from individual pigs, making our results more representative of the broader population. This strategy enabled us to investigate condition-dependent cellular responses while maintaining statistical rigor using pooled biological replicates. To clarify, the reported n values refer to replicates from our pooled cell source derived from 10 pig hearts.

VICs were seeded on soft (E = ~2.8 kPa, $n = 8$) and stiff (E = ~13.4 kPa, $n = 4$) hydrogels as previously described[71] (Fig. 3a, b). In the soft condition, VICs tend to remain quiescent and proliferate less than VICs seeded on stiff, activation-promoting substrates. To ensure equivalent numbers of VICs from both soft and stiff conditions in the training dataset, we performed 4 additional replicates of the soft condition. In brief, poly(ethylene glycol) photodegradable diacrylate (PEGdiPDA; MW ~ 3,400 Da) was co-polymerized with poly(ethylene glycol) monoacrylate (PEGA; MW ~ 400 Da) and acrylamide diethylene glycol – diethylene glycol – glycine – arginine – glycine – aspartic acid – serine – glycine (Ac-OOGRGDSG) in PBS through a redox-initiated free radical polymerization. Hydrogel formulations were prepared with the following concentrations: 7.0 wt% PEGdiPDA, 6.8 wt% PEGA, 5 mM Ac-OOGRGDSG, 0.2 M ammonium persulfate (APS), and 0.1 M tetramethylethylenediamine (TEMED). ~100 μm thick hydrogels were formed on acrylated glass coverslips after the addition of APS and TEMED. The hydrogels were allowed to cure for 6 min before being washed and stored in PBS prior to cell seeding. This formulation resulted in stiff hydrogels. To photodegrade and produce soft hydrogels, a separate cohort of stiff hydrogels were irradiated with 365 nm UV light at 10 mW/cm^2 intensity for 6 min prior to cell seeding. VICs were seeded at a density of 15,000 cells/cm^2 on 12 mm soft and stiff hydrogels within media comprised of M199 media supplemented with 1% FBS, penicillin-streptomycin, and amphotericin B for 72 h before fixing and staining. To generate drug-screening test data sets, a separate cohort of VICs seeded on stiff hydrogels were treated with 10 μM of the small molecule drug 5-azacytidine for 72 h before fixing and staining ($n = 4$). Furthermore, an additional cohort of VICs seeded on soft hydrogels were treated with 100 nM of the small molecule drug bpV(HOpic) for 72 h before fixing and staining ($n = 4$).

Immunostaining and imaging

Samples were fixed in 4% paraformaldehyde (PFA, Electron Microscopy Sciences) for 20 min at room temperature before washing twice with PBS for 5 min each. Next, the samples were permeabilized with 0.1% TritonX-100 (Fisher Scientific) in PBS for 20 min at room temperature. Then, non-specific binding was blocked for 1 h using 5% bovine serum albumin (BSA, Sigma-Aldrich) dissolved in PBS before adding mouse anti-αSMA primary antibody (Cat. No. ab7817, Abcam) diluted in 5% BSA at a ratio of 1:200. The samples were then incubated overnight at 4°C. The next day, the primary antibody was removed, and the samples were washed twice with PBST (0.5% Tween-20 in PBS) and once in PBS. Then, a secondary antibody solution containing 1:200 goat-anti mouse Alexa 488 (Cat. No. A-11001, ThermoFisher), 1:1000 DAPI, and 1:5000 HCS Cell Mask Orange (Life Technologies) was added for 1 h and then rinsed 2x with PBST and once with PBS before imaging. Imaging was performed using an Operetta high-content confocal microscope (PerkinElmer) with a 20x air objective resulting in 1360 × 1024 × 10 images at a resolution of 0.49 μm/pixel in the x-y plane and 1 μm through the z-plane. The z-plane that was most in focus for each field of view was used for analysis.

Segmentation, centering, and alignment

The boundaries of cellular bodies were segmented using a marker-based watershed algorithm (Fig. 1a) that used segmented nuclei as markers (Fig. 1b). The segmentation was performed in the Harmony High-Content Imaging and Analysis Software (PerkinElmer). We briefly note that marker-based watershed segmentation can be achieved in a myriad of programs including Matlab, Python, ImageJ, and CellProfiler[62] and is not a point of novelty in the current study. Segmentation quality was manually checked and cells that were segmented poorly (e.g., erroneous cell boundaries resulting in incomplete segmentation or inclusion of neighboring cells) were discarded from further analysis (Supplemental Fig. 5). Moreover, dividing cells evidenced by nuclei splitting, were also discarded. In addition, only cellular bodies that were completely in the field of view (i.e., not touching the image boundary) were included in the analysis. Despite the growing recognition for the need to evaluate the effects of cellular segmentation error on downstream analysis[72,73], there are currently no standard methods to detect and correct erroneous cellular segmentations in an unsupervised manner. Therefore, qualitative/subjective validation of segmentation is used herein and by others[18]. As the goal of the current manuscript is to develop quantitative approaches to perform phenotyping of fibroblasts, we do not focus on developing methods to improve image segmentation or to automate the detection and removal of poor segmentations from analysis. In the future, automated methods to detect and remove/improve poor segmentations may be developed to augment the current method presented herein. The pixels comprising the segmented cell boundaries and the segmented nuclei boundaries were converted to cartesian points by multiplying the pixel location by the image resolution (0.49 μm/pixel). To ensure that morphological analysis was invariant to cell location and pose, the cell bodies and nuclei were centered and aligned so that their centroids coincided with the origin and their longest axes were parallel to the x-axis (Fig. 1c). The longest axis was defined as the direction which accounted for the greatest variation in cell/nuclei boundary points as determined by conducting principal component analysis (PCA) on the cartesian points of the cell/nuclei boundary and then aligning the cell/nuclei along the first principal component axis. "Flipping" of the shapes along the vertical and/or horizontal axis was also permitted to ensure that the greatest portion of the cell/nuclei area was in the first quadrant. This type of shape-preserving transformation was performed to ensure that all large cellular structures (such as protrusions) lie in approximately the same location across all cells to facilitate consistent analysis.

Fourier series expansion

We employed Fourier-based methods to obtain mathematical descriptions of cellular bodies and nuclei to quantify their morphologies[40]. Two parametric functions x(θ) and y(θ) are employed to relate the arc length around a given cell/nuclei boundary (θ) to the corresponding x- and y- Cartesian coordinates. Here, the arc length of the shape is normalized such that the total length is 2π. Essentially, this results in a one-to-one mapping of the original shape to a unit circle, ensuring that every cartesian x,y point corresponds to exactly one arc length θ in the range $[0, 2\pi]$ (Eqs. 1 and 2). In Eqs. 1 and 2, the variables a_0, a_n, b_n, c_0, c_n, and d_n are Fourier coefficients determined by solving overdetermined systems of linear equations. Equations 1 and 2 can be

represented in matrix form as

$$\begin{bmatrix} 1/2 & \cos(1*0) & \sin(1*0) & \cdots & \cos(n*0) & \sin(n*0) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1/2 & \cos(1*2\pi) & \sin(1*2\pi) & \cdots & \cos(n*2\pi) & \sin(n*2\pi) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ b_1 \\ \vdots \\ a_n \\ b_n \end{bmatrix} = \begin{bmatrix} x(0) \\ \vdots \\ x(2\pi) \end{bmatrix}, \tag{5}$$

$$\begin{bmatrix} 1/2 & \cos(1*0) & \sin(1*0) & \cdots & \cos(n*0) & \sin(n*0) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1/2 & \cos(1*2\pi) & \sin(1*2\pi) & \cdots & \cos(n*2\pi) & \sin(n*2\pi) \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ d_1 \\ \vdots \\ c_n \\ d_n \end{bmatrix} = \begin{bmatrix} y(0) \\ \vdots \\ y(2\pi) \end{bmatrix}. \tag{6}$$

Here, n is the number of user-determined Fourier harmonics used to reproduce the shape. As n increases, the Fourier description converges to the target shape (Fig. 1d). Representing the boundaries of cell bodies and nuclei using Fourier series resulted in Fourier coefficients that contained information on cell and nuclei morphology (e.g., both size and shape) that can be used to analyze cellular bodies and nuclei in a data-driven manner.

Computation of coherence and degree of anisotropy (DOA)

Coherence was computed to achieve a quantitative descriptor of αSMA stress fiber anisotropy or discreteness within imaged cells using previously described methods[43] (Fig. 2a). We first computed the x- and y-gradients (I_x and I_y, respectively) of αSMA images on a pixel-by-pixel basis using the Sobel-Feldman operator[46] which consists of two 3 × 3 kernels

$$K_x = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} \tag{7}$$

$$K_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}. \tag{8}$$

The convolution of K_x and K_y with an image I results in the gradients

$$I_x = K_x * I \tag{9}$$

$$I_y = K_y * I. \tag{10}$$

Here, the symbol $*$ denotes the 2D convolution operation. Next, we assembled the structure tensor (Eq. 3). In Eq. 3, W is a Gaussian smoothing kernel with a standard deviation of 4 pixels and here again the symbol $*$ denotes convolution. A standard deviation of 4 was chosen as the diameter of αSMA stress fibers in our fluorescence images spanned ~4 pixels. It is important to note that the standard deviation value will vary depending upon imaging parameters and the size (in pixels) of the biological structure of interest. Eigendecomposition of S was performed to obtain eigenvalues (λ_1, λ_2) and eigenvectors ($\boldsymbol{e}_1$, $\boldsymbol{e}_2$). If $\lambda_1 = \lambda_2$, then the gradient has no preferred direction which can occur in the case of rotational symmetry or homogeneity (Fig. 2a, left-most panel). If $\lambda_1 > \lambda_2$, then λ_1 represents the greatest gradient (i.e., change in intensity) squared at a given pixel which is oriented along $\boldsymbol{e}_1$ whereas λ_2 represents the gradient squared along the orthogonal direction $\boldsymbol{e}_2$ (Fig. 2a, middle three panels). In the case that $\lambda_1 > 0$ and $\lambda_2 = 0$, the structure in the image is completely aligned and is locally one-dimensional (Fig. 2a, right-most panel). The relative difference in magnitude between λ_1 and λ_2 reflects the local degree of anisotropy. This is quantified by coherence (c), as shown in Eq. 4, whose value ranges from 0 to 1. The median coherence value of the pixels comprising the cell body was computed to achieve a single quantitative metric of the degree of anisotropy (DOA) of αSMA stress fibers (Fig. 2c). Traditionally, manual inspection of the presence of discrete αSMA stress fibers governed the grading of quiescent and activated fibroblasts (Fig. 2d). In addition to being computed programmatically[74], DOA can also be computed using existing tools in ImageJ (Appendix H in Supplemental Information).

Concatenation and standardization of variables, principal component analysis, and generation of eigenshapes

Feature vectors were established for each cell by combining the Fourier coefficients for the cell body and nucleus with the corresponding DOA value (Supplemental Table 1). The feature vectors were concatenated into an array such that every row contained information for one cell and every column represented a unique variable (either a Fourier coefficient or DOA). The variables were standardized as follows

$$Z = \frac{x - \mu}{\sigma} \tag{11}$$

where Z represents the standard score, x is the observed value of the variable, μ is the mean of the variable, and σ is the standard deviation of the variable. Standardization of variables was performed to ensure that subsequent PCA was not affected by the magnitude of the variables. PCA was then performed to investigate the greatest sources of variation among the feature vectors which resulted in eigenvalues and eigenvectors (U). Next, principal component (PC) scores were computed using the following

$$V = XU \tag{12}$$

where V contains the resulting PC scores and X is the original data matrix containing the standardized feature vectors for every cell. Eigenshapes were generated along the principal axes by first taking the average of the PC scores across all cells analyzed. Then, the average PC score for the principal axis of interest was altered systematically from -2*STD to +2*STD to produce eigenshapes using

$$X' = \bar{V}U^{-1} \tag{13}$$

where $\bar{V}$ is the altered average PC scores and X' is the recovered standardized feature vector. To generate the eigenshape, X' is first converted from Z-scores back to non-standardized variables to recover Fourier coefficients and DOA. The Fourier coefficients are used to generate the x- and y-cartesian coordinates of the eigenshape using Eqs. 1 and 2, respectively (Fig. 3c). The eigenshape is colored based on its DOA value. Superimposing the -2*STD and +2*STD eigenshapes along PC axes 1-3 allowed for physical interpretation of each PC axis and for the identification of features (Fig. 3d).

Logistic regression models of features

Logistic regression models were trained for each feature (Fig. 4a) using a manually graded VIC data set in which VICs were demarcated as quiescent (activation probability = 0) or activated (activation probability = 1). The training process resulted in models that predicted VIC phenotype based on individual features (Fig. 4b). The training process

involved fitting the ground truth data set to the logistic function

$$p(x) = \frac{1}{1 + e^{-(x-\mu)/s}} \tag{14}$$

where $p(x)$ is the probability of VICs being activated, x is a value of the feature of interest, μ is a location parameter such that $p(\mu) = 0.5$, and s is a scaling parameter. In practice, the ground truth data set had activation probabilities of 0 or 1 whereas the fitted logistic regression model can produce continuous activation probabilities from 0 to 1. The coefficients for the logistic regression models were determined by minimizing the overall negative log-likelihood. To combine the effects of multiple features (e.g., cell area, DOA, cell aspect ratio) into a single model for predicting VIC activation, we trained a multi-variable logistic regression model

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m)}} \tag{15}$$

where p is activation probability, β_{0-m} are fitted parameters, and x_{1-m} are values of features. The coefficients were determined in similar fashion to those in Eq. 14 by minimizing the overall negative log-likelihood.

Receiver operating characteristic curves

Receiver operating characteristic (ROC) curves[75] were generated to visualize the predictive capability of each feature and the MLR model (Fig. 4c). This was accomplished by setting an ever-increasing threshold value for a specific feature of interest and computing the true positive rate (TPR) and false positive rate (FPR) at every threshold value. For the MLR model, the probability at which a fibroblast is deemed activated was varied. The TPR and FPR were computed by comparing the grading of VIC phenotype by the current threshold value to the ground truth data set generated through manual grading. Next, the TPR was plotted against the FPR at every threshold value to achieve a visual representation of the feature's predictive capability. In general, variables with high predictive capabilities will produce an ROC curve that approaches the top left corner of the plot whereas worse predictors will produce an ROC curve closer to the diagonal axis of the plot with the absolute worst predictors (i.e., random classifiers) forming a line of identity. The area under the curve (AUC) is computed to serve as a single quantitative descriptor of the predictive capability of each feature and the MLR model. AUC ranges from 0.5 (random classifier) to 1.0 (perfect classifier).

k-fold cross validation

We performed k-fold cross validation to validate the ability of identified features to grade fibroblast phenotypes (Fig. 5b). Here, we chose $k = 10$. The entire data set was randomly separated into 10 equal groups where 9 of the groups were used to train the logistic regression models and the remaining group was used for validation. The percent of VICs in the validation data set that were predicted correctly as compared to manual grading was computed and served as an indicator of model accuracy (Fig. 5c). This process was iterated 10 times so that each group served as the validation group exactly once.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The imaging data used in this study are available in the University of Colorado Boulder's database, CU Scholar, under accession code mw22v725h [https://scholar.colorado.edu/concern/datasets/mw22v725h][76] Source data are provided with this paper.

Code availability

The code used to develop the model, perform the analyzes and generate results in this study is publicly available and has been deposited in Github at https://github.com/alexkhang18/Fourier-Series-Shape-Analysis-Fibroblasts, under MIT license. The specific version of the code associated with this publication is archived in Zenodo and is accessible via https://doi.org/10.5281/zenodo.14798858[74].

References

1. Henderson, N. C., Rieder, F. & Wynn, T. A. Fibrosis: from mechanisms to medicines. *Nature* **587**, 555–566 (2020).
2. Cox, T. R. & Erler, J. T. Fibrosis and cancer: partners in crime or opposing forces? *Trends Cancer* **2**, 279–282 (2016).
3. Wynn, T. A. & Ramalingam, T. R. Mechanisms of fibrosis: therapeutic translation for fibrotic disease. *Nat. Med.* **18**, 1028–1040 (2012).
4. Rogliani, P., Calzetta, L., Cavalli, F., Matera, M. G. & Cazzola, M. Pirfenidone, nintedanib and N-acetylcysteine for the treatment of idiopathic pulmonary fibrosis: a systematic review and meta-analysis. *Pulm. Pharmacol. Ther.* **40**, 95–103 (2016).
5. Glass, D. S. et al. Idiopathic pulmonary fibrosis: current and future treatment. *Clin. Respir. J.* **16**, 84–96 (2022).
6. Hinz, B. Myofibroblasts. *Exp. Eye Res.* **142**, 56–70 (2016).
7. Hinz, B. & Lagares, D. Evasion of apoptosis by myofibroblasts: a hallmark of fibrotic diseases. *Nat. Rev. Rheumatol.* **16**, 11–31 (2020).
8. Schuster, R., Rockel, J. S., Kapoor, M. & Hinz, B. The inflammatory speech of fibroblasts. *Immunol. Rev.* **302**, 126–146 (2021).
9. Hinz, B. et al. The myofibroblast one function, multiple origins. *Am. J. Pathol.* **170**, 1807–1816 (2007).
10. Younesi, F. S., Miller, A. E., Barker, T. H., Rossi, F. M. V. & Hinz, B. Fibroblast and myofibroblast activation in normal tissue repair and fibrosis. *Nat. Rev. Mol. Cell Biol.* **25**, 617–638 (2024).
11. Younesi, F. S., Son, D. O., Firmino, J. & Hinz, B. Myofibroblasts, methods and protocols. *Methods Mol. Biol.* **2299**, 17–47 (2021).
12. Hinz, B. Masters and servants of the force: the role of matrix adhesions in myofibroblast force perception and transmission. *Eur. J. Cell Biol.* **85**, 175–181 (2006).
13. Ali, M. S. et al. Correlation between valvular interstitial cell morphology and phenotypes: a novel way to detect activation. *Tissue Cell* **54**, 38–46 (2018).
14. Hillsley, A., Santos, J. E. & Rosales, A. M. A deep learning approach to identify and segment alpha-smooth muscle actin stress fiber positive cells. *Sci. Rep.* **11**, 21855 (2021).
15. Wu, Z. et al. DynaMorph: self-supervised learning of morphodynamic states of live cells. *Mol. Biol. Cell* **33**, ar59 (2022).
16. Zaritsky, A. et al. Interpretable deep learning uncovers cellular properties in label-free live cell images that are predictive of highly metastatic melanoma. *Cell Syst.* **12**, 733–747.e6 (2021).
17. Hillsley, A. et al. A strategy to quantify myofibroblast activation on a continuous spectrum. *Sci. Rep.* **12**, 12239 (2022).
18. Phillip, J. M., Han, K.-S., Chen, W.-C., Wirtz, D. & Wu, P.-H. A robust unsupervised machine-learning method to quantify the morphological heterogeneity of cells and nuclei. *Nat. Protoc.* **16**, 754–774 (2021).
19. Phillip, J. M. et al. Biophysical and biomolecular determination of cellular age in humans. *Nat. Biomed. Eng.* **1**, 0093 (2017).
20. Klingberg, F., Hinz, B. & White, E. S. The myofibroblast matrix: implications for tissue repair and fibrosis. *J. Pathol.* **229**, 298–309 (2013).
21. Baum, J. & Duffy, H. S. Fibroblasts and myofibroblasts: what are we talking about? *J. Cardiovasc. Pharm.* **57**, 376–379 (2011).
22. Frangogiannis, N. G. Cardiac fibrosis. *Cardiovasc. Res.* **117**, 1450–1488 (2020).
23. Goody, P. R. et al. Aortic Valve Stenosis. *Arter., Thromb., Vasc. Biol.* **40**, 885–900 (2020).

24. Khang, A., Meyer, K. & Sacks, M. An inverse modeling approach to estimate 3D aortic valve interstitial cell stress fiber force levels. *J. Biomech. Eng.* **145**, 121005 (2023).
25. Khang, A., Steinman, J., Tuscher, R., Feng, X. & Sacks, M. S. Estimation of aortic valve interstitial cell-induced 3D remodeling of poly(ethylene glycol) hydrogel environments using an inverse finite element approach. *Acta Biomater.* **160**, 123–133 (2023).
26. Tuscher, R. et al. Functional differences in human aortic valve interstitial cells from patients with varying calcific aortic valve disease. *Front. Physiol.* **14**, 1168691 (2023).
27. Khang, A., Lejeune, E., Abbaspour, A., Howsmon, D. P. & Sacks, M. S. On the three-dimensional correlation between myofibroblast shape and contraction. *J. Biomech. Eng.* **143**, 094503 (2021).
28. Khang, A. et al. Quantifying heart valve interstitial cell contractile state using highly tunable poly(ethylene glycol) hydrogels. *Acta Biomater.* **96**, 354–367 (2019).
29. Khang, A., Nguyen, Q., Feng, X., Howsmon, D. P. & Sacks, M. S. Three-dimensional analysis of hydrogel-imbedded aortic valve interstitial cell shape and its relation to contractile behavior. *Acta Biomater.* **163**, 194–209 (2023).
30. Peter, A. K. et al. Cardiac fibroblasts mediate a sexually dimorphic fibrotic response to β-adrenergic stimulation. *J. Am. Hear. Assoc.: Cardiovasc. Cerebrovasc. Dis.* **10**, e018876 (2021).
31. Ceccato, T. L. et al. Defining the cardiac fibroblast secretome in a fibrotic microenvironment. *J. Am. Hear. Assoc.: Cardiovasc. Cerebrovasc. Dis.* **9**, e017025 (2020).
32. Good, R. B. et al. A high content, phenotypic 'scar-in-a-jar' assay for rapid quantification of collagen fibrillogenesis using disease-derived pulmonary fibroblasts. *BMC Biomed. Eng.* **1**, 14 (2019).
33. Marwick, J. A. et al. Application of a high-content screening assay utilizing primary human lung fibroblasts to identify antifibrotic drugs for rapid repurposing in COVID-19 patients. *SLAS Discov.* **26**, 1091–1106 (2021).
34. Gerckens, M. et al. Phenotypic drug screening in a human fibrosis model identified a novel class of antifibrotic therapeutics. *Sci. Adv.* **7**, eabb3673 (2021).
35. Haupt, A. & Minc, N. How cells sense their own shape – mechanisms to probe cell geometry and their implications in cellular organization and function. *J. Cell Sci.* **131**, jcs214015 (2018).
36. Wu, P.-H. et al. Evolution of cellular morpho-phenotypes in cancer metastasis. *Sci. Rep.* **5**, 18437 (2015).
37. Wu, P.-H. et al. Single-cell morphology encodes metastatic potential. *Sci. Adv.* **6**, eaaw6938 (2020).
38. Rohlf, F. J. & Archie, J. W. A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Culicidae). *Syst. Zoöl* **33**, 302 (1984).
39. Rohlf, F. J. Relationships among eigenshape analysis, Fourier analysis, and analysis of coordinates. *Math. Geol.* **18**, 845–854 (1986).
40. Shen, L., Farid, H. & McPeek, M. A. Modeling three-dimensional morphological structures using spherical harmonics. *Evolution* **63**, 1003–1016 (2009).
41. Schwarz, G. et al. Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464 (1978).
42. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723 (1974).
43. Püspöki, Z., Storath, M., Sage, D. & Unser, M. Transforms and operators for directional bioimage analysis: a survey. *Adv. Anat., Embryol., cell Biol.* **219**, 69–93 (2016).
44. Fonck, E. et al. Effect of aging on elastin functionality in human cerebral arteries. *Stroke* **40**, 2552–2556 (2009).
45. Rezakhaniha, R. et al. Experimental investigation of collagen waviness and orientation in the arterial adventitia using confocal laser scanning microscopy. *Biomech. Model. Mechanobiol.* **11**, 461–473 (2012).
46. Sobel, I. & Feldman, G. An isotropic 3x3 image gradient operator. https://doi.org/10.13140/rg.2.1.1912.4965 (1968).
47. Wang, H., Tibbitt, M. W., Langer, S. J., Leinwand, L. A. & Anseth, K. S. Hydrogels preserve native phenotypes of valvular fibroblasts through an elasticity-regulated PI3K/AKT pathway. *Proc. Natl Acad. Sci.* **110**, 19336–19341 (2013).
48. Walker, C. J. et al. Extracellular matrix stiffness controls cardiac valve myofibroblast activation through epigenetic remodeling. *Bioeng. Transl. Med.* **7**, e10394 (2022).
49. Walker, C. J. et al. Nuclear mechanosensing drives chromatin remodelling in persistently activated fibroblasts. *Nat. Biomed. Eng.* **5**, 1485–1499 (2021).
50. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* (Academic Press, 1977).
51. Sawilowsky, S. S. New effect size rules of thumb. *J. Mod. Appl. Stat. Methods* **8**, 597–599 (2009).
52. Strand, K. A. et al. High throughput screen identifies the DNMT1 (DNA Methyltransferase-1) inhibitor, 5-Azacytidine, as a potent inducer of PTEN (Phosphatase and Tensin Homolog). *Arter. Thromb., Vasc. Biol.* **40**, 1854–1869 (2020).
53. Parapuram, S. K. et al. Loss of PTEN expression by dermal fibroblasts causes skin fibrosis. *J. Investig. Dermatol.* **131**, 1996–2003 (2011).
54. Parapuram, S. K. et al. Loss of PTEN expression by mouse fibroblasts results in lung fibrosis through a CCN2-dependent mechanism. *Matrix Biol.* **43**, 35–41 (2015).
55. Lu, S. et al. PTEN (Phosphatase and Tensin Homolog) protects against Ang II (Angiotensin II)-induced pathological vascular fibrosis and remodeling—brief report. *Arter. Thromb., Vasc. Biol.* **40**, 394–403 (2020).
56. Moulton, K. S. et al. PTEN deficiency promotes pathological vascular remodeling of human coronary arteries. *JCI Insight* **3**, e97228 (2018).
57. Jones, C. E. et al. Pten regulates collagen fibrillogenesis by fibroblasts through SPARC. *PLoS ONE* **16**, e0245653 (2021).
58. Schmid, A. C., Byrne, R. D., Vilar, R. & Woscholski, R. Bisperoxovanadium compounds are potent PTEN inhibitors. *FEBS Lett.* **566**, 35–38 (2004).
59. Fregin, B. et al. High-throughput single-cell rheology in complex samples by dynamic real-time deformability cytometry. *Nat. Commun.* **10**, 415 (2019).
60. Khotanzad, A. & Hong, Y. H. Invariant image recognition by Zernike moments. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**, 489–497 (1990).
61. Hillsley, A. et al. Data for: a deep learning approach to identify and segment α-smooth muscle actin stress fiber positive cells. *Texas Data Repository*. https://doi.org/10.18738/t8/lrwtyj (2021).
62. Stirling, D. R. et al. CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinform* **22**, 433 (2021).
63. 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).
64. Chandrasekaran, S. N. et al. Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. *Nat. Methods* **21**, 1114–1121 (2024).
65. Cimini, B. A. et al. Optimizing the Cell Painting assay for image-based profiling. *Nat. Protoc.* **18**, 1981–2013 (2023).
66. Gustafsdottir, S. M. et al. Multiplex cytological profiling assay to measure diverse cellular states. *PLoS ONE* **8**, e80999 (2013).
67. Kobayashi, H., Cheveralls, K. C., Leonetti, M. D. & Royer, L. A. Self-supervised deep learning encodes high-resolution features of protein subcellular localization. *Nat. Methods* **19**, 995–1003 (2022).
68. Kräter, M. et al. AIDeveloper: deep learning image classification in life science and beyond. *Adv. Sci.* **8**, 2003743 (2021).

69. Irianto, J., Lee, D. A. & Knight, M. M. Quantification of chromatin condensation level by image processing. *Méd. Eng. Phys.* **36**, 412–417 (2014).
70. Johnson, C. M., Hanson, M. N. & Helgeson, S. C. Porcine cardiac valvular subendothelial cells in culture: cell isolation and growth characteristics. *J. Mol. Cell. Cardiol.* **19**, 1185–1193 (1987).
71. Kloxin, A. M., Kasko, A. M., Salinas, C. N. & Anseth, K. S. Photodegradable hydrogels for dynamic tuning of physical and chemical properties. *Science* **324**, 59–63 (2009).
72. Chen, H. & Murphy, R. F. Evaluation of cell segmentation methods without reference segmentations. *Mol. Biol. Cell* **34**, ar50 (2023).
73. Wiggins, L., O'Toole, P. J., Brackenbury, W. J. & Wilson, J. Exploring the impact of variability in cell segmentation and tracking approaches. *Microsc. Res. Tech.* https://doi.org/10.1002/jemt.24715 (2024).
74. Khang, A. et al. alexkhang18/Fourier-Series-Shape-Analysis-Fibroblasts: Version 1. *Github* https://doi.org/10.5281/zenodo.14798859 (2025).
75. Park, S. H., Goo, J. M. & Jo, C.-H. Receiver Operating Characteristic (ROC) curve: practical review for radiologists. *Korean J. Radiol.* **5**, 11–18 (2004).
76. Batan, D. et al. Immunostained fluorescent images of valve interstitial cells seeded on soft and stiff hydrogels [Data set]. *University of Colorado Boulder* https://doi.org/10.25810/degy-jc81 (2025).

Acknowledgements

We thank Dr. Tova L. Ceccato for the rat cardiac fibroblast imaging dataset. We acknowledge funding from the National Institutes of Health T32 HL007822 and F32 HL176073 awarded to A.K., Helen Hay Whitney Foundation award number F1339 to K.B., American Heart Association 20PRE35200068 awarded to D.B. and National Institutes of Health R01 HL142935 and R01 HL132353 awarded to KSA.

Author contributions

A.K., D.B., and K.S.A. designed and performed all the experiments. A.B. and A.K. performed image analysis, mathematical analysis, and data analysis. A.K., G.T., K.B., and K.S.A. contributed to writing and editing the paper.

Competing interests

The authors report no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-58082-0.

Correspondence and requests for materials should be addressed to Kristi S. Anseth.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.