Supplementary Information

1. Support vector machine

The support vector machines are supervised learning models to classify data represented by vectors in multi-dimensional spaces with boundaries defined by associated algorithms. In detail, the training set of the classification problem is in the form: $\{\boldsymbol{x}_i, y_i\}, i = 1,2,...,n$, where $\boldsymbol{x}_i \in \Re^d$ are vectors representing the data, and $y_i \in \{+1,-1\}$ are class labels for two classes.

In this study, we used support vector machines with linear kernel, with which the classifier boundaries are hyperplanes in the form $\boldsymbol{w} \cdot \boldsymbol{x} + b = 0$ where $\boldsymbol{w}$ is the normal vector of the hyperplane and b is the term determining offset. For separable cases the decision function for classification is then defined as $f(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{x} + b$ so that $f(\boldsymbol{x}_i) \geq 1$ if $y_i = 1$ or $f(\boldsymbol{x}_i) \leq -1$ if $y_i = -1$. Thus the margin between two classes is $\frac{2}{\|\boldsymbol{w}\|}$. SVMs optimize the classification by maximizing the margin and the problem can be formulated as:

$$\underset{(\boldsymbol{w},b)}{\arg\min}\left(\frac{1}{2}\|\boldsymbol{w}\|^2\right) \tag{S1}$$

subject to $y_i(\boldsymbol{w} \cdot \boldsymbol{x} + b) \geq 1$, (for any $i = 1,2,...,n$)

For non-separable cases, compromise of the decision functions and mislabeling are allowed, which are described by the “slack variables”: $\xi_i \geq 0, i = 1,2,...,n$. Thus SVMs optimize the classification by both maximization of the margin and minimization of the penalty caused by the slack variables. The problem is formulated as:

$$\underset{(\boldsymbol{w},b,\xi)}{\arg\min}\left(\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n}\xi_i\right) \tag{S2}$$

subject to $y_i(\boldsymbol{w} \cdot \boldsymbol{x} + b) \geq 1 - \xi_i, \xi_i \geq 0$

Introducing Lagrange multipliers, we can solve the problem by optimizing the Lagrangian:

$$\arg \min_{(\boldsymbol{w},b,\xi)} \max_{(\boldsymbol{\alpha},\boldsymbol{\beta})} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\left[y_i(\boldsymbol{w} \cdot \boldsymbol{x} + b) - 1 + \xi_i\right] - \sum_{i=1}^{n}\beta_i\xi_i \tag{S3}$$

Therefore the normal vector of the hyperplane is $\boldsymbol{w} = \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i$. For data points correctly classified and out of the margin, $\alpha_i = 0, \xi_i = 0$. For data points on the margin boundary, $0 < \alpha_i < C, \xi_i = 0$ and $f(\boldsymbol{x}_i) = \pm 1$. For data points within the margin or misclassified, $\alpha_i = C, \xi_i > 0$. Any $\boldsymbol{x}_i$ with $\alpha_i > 0$ are named support vectors which contribute to define the normal vector of the classifier hyperplane. Therefore, support vectors are those data points on the margin boundary or with compromised decision function, suggesting they are not well classified by the trained classifier.

2. Selection of "median cells"

The "median cells" are single cells that we selected to represent the population morphological difference of two classes in each pairwise SVM comparison. Subsequent to each loop of the SVM training on supercells at supercell size of 5 with the selected features, the trained classifier was employed to classify the original single cells. Single cells with decision functions higher than 1 or lower than -1 were recorded as well classified single cells respectively for two classes. We then selected those single cells which are always well classified by the classifier of each training loop.

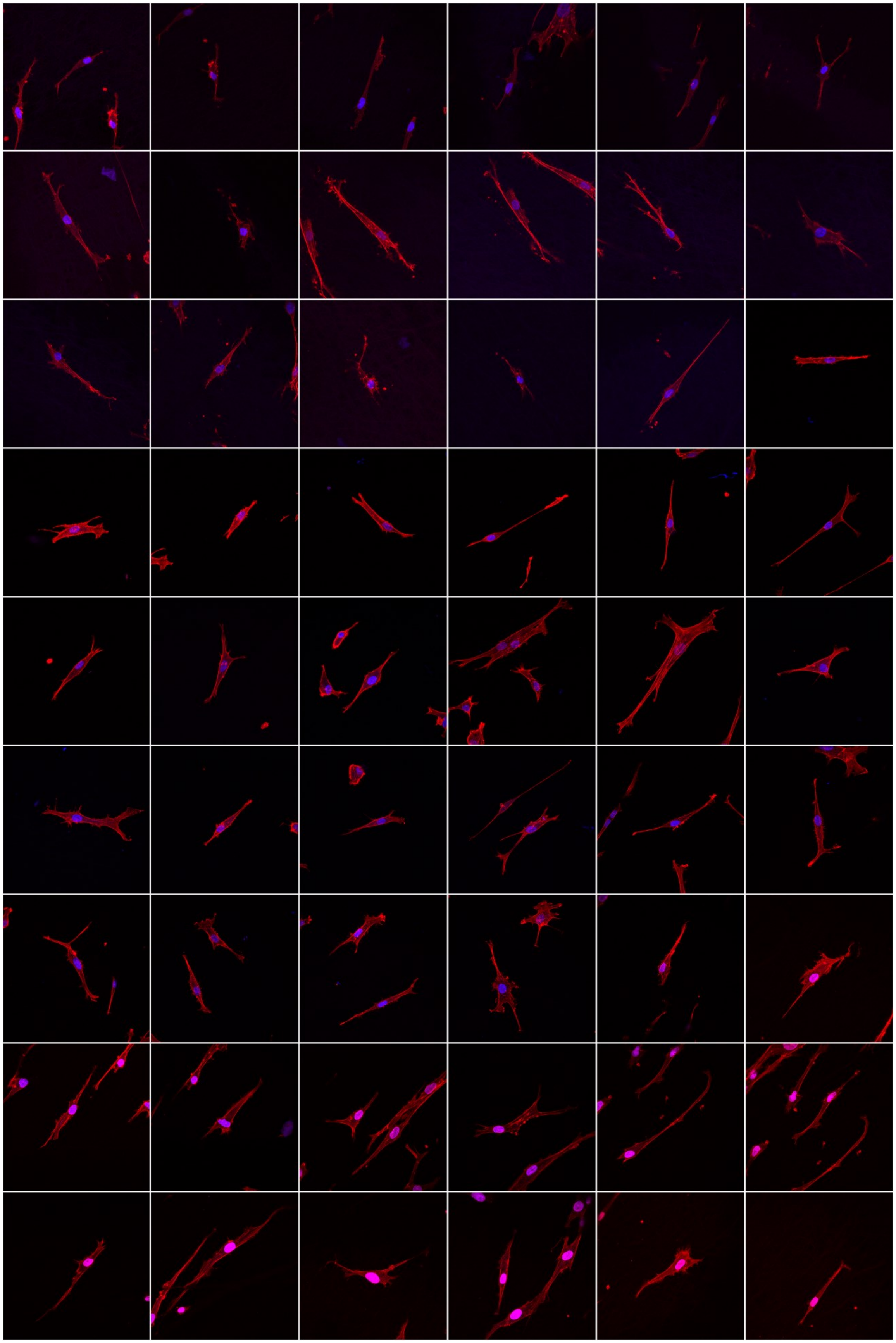
In the next step, for each class we calculated the total Euclidean distance from each selected single cell to all the other selected single cells in metric space established with only the selected features. The list of the selected single cells of each class was then sorted according to the total Euclidean distances. We then extracted the binary segmentations of those single cells with the lowest total Euclidean distances to represent the morphology of each class in Fig. 5.

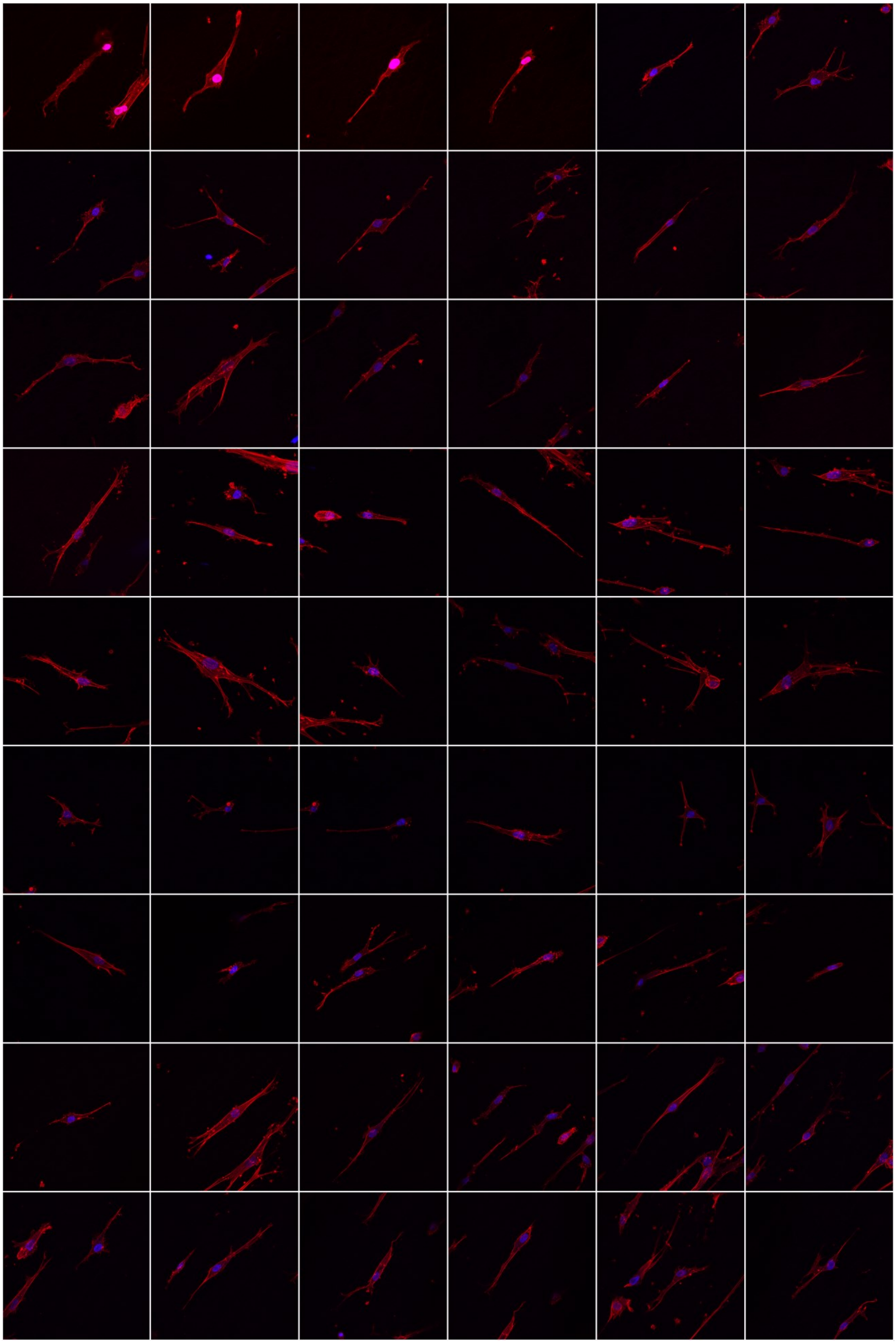
Supplementary Table 1

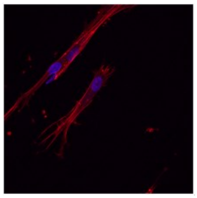
Explanations of some shape metrics of the total 22 shape metrics

Metric Category	Metrics	Explanations
Cell Size	1. Area	Area of the 2-D projection of the cell
	2. Perimeter	Perimeter of the 2-D projection of the celll
	3. Mean Boundary Distance	Average distance from the geometrical center of the cell outline to each representative point on the cell boundary
	4. Max Boundary Distance	Maximum distance from the geometrical center of the cell outline to each representative point on the cell boundary
	5. Major Axis Length	Length of the major axis of the ellipse that has the same normalized second central moments as the 2-D projection of the cell
	6. Minor Axis Length	Length of the minor axis of the ellipse that has the same normalized second central moments as the 2-D projection of the cell
	7. Max Major Branch Length	For each major branch, assuming the branch origin is the intersection point of the cell equivalent circle and the extracted branch skeleton. The farthest distance from the intersection point of the cell equivalent circle and the extracted branch skeleton to the major branch boundary points is taken as the length of the major branch. PCA analysis is applied to boundary points of the major branch (the part out of the equivalent circle). Width of each branch is defined as the average of double of the projection of the branch boundary points to the direction of the secondary principal component.
	8. Mean Major Branch Length	
	9. Max Major Branch Width	
	10. Mean Major Branch Width	
Shape Pattern	11. Circularity	Ratio of the perimeter of the cell to the perimeter of the cell equivalent circle. The equivalent circle is defined as the circle having the same area and geometrical center with the cell. Circularity is normalized so that circularity of a circle is 1
	12. Solidity	Area of the cell over the convex hull encompassing the cell. The convex hull of a shape is the shortest boundary that can be drawn around that shape with no negative curvature segments (i.e. invaginations). A smaller solidity therefore indicates more or larger invaginations.
	13. Aspect Ratio	Ratio of major axis length over minor axis length of the ellipse that has the same normalized second central moments as the 2-D projection of the cell

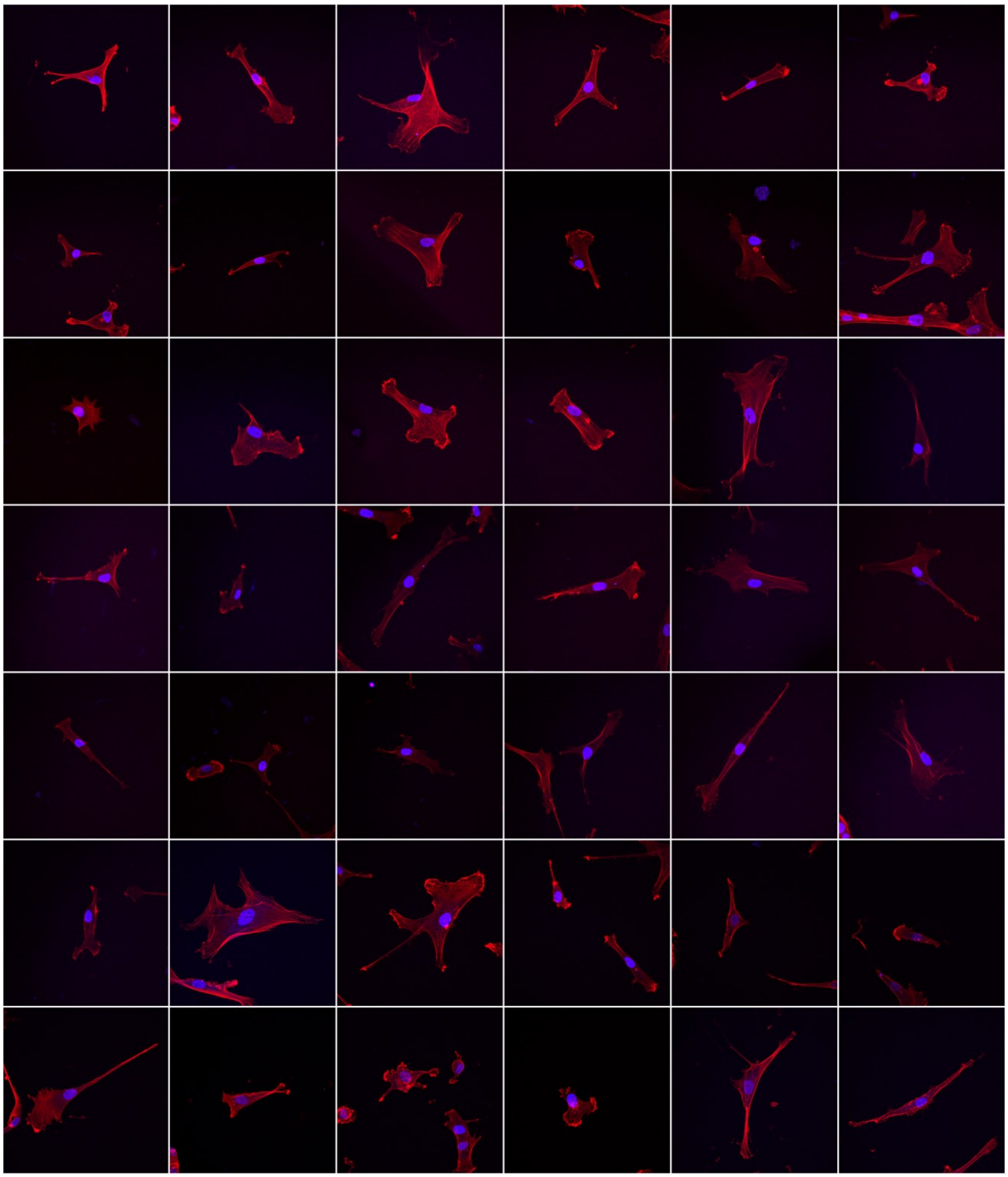
	14. Num. of Major Branches	In the skeleton structure of the cell, major branches are identified as the branches extending out the central part defined as the part of cell inside the equivalent circle (the circle which is centered at the geometrical centroid of the cell and has the same area with the cell)
	15. Max Branch Length/ Major Axis Length	
	16. Mean Branch Length/ Equivalent Radius	Equivalent radius is defined as the radius of the circle which is centered at the geometrical centroid of the cell and has the same area with the cell
Local Curvature Information	17. Mean Negative Curvature	The curvature at a certain point is defined as the reciprocal of the radius of the circular arc which fits that point and the two neighborhood points 2 points away from it. Mean negative curvature is the sum of negative curvatures over the perimeter of the cell outline
	18. Invagination	Number of concave regions with continuous negative curvature on the boundary, normalized by the perimeter
	19. Tortuosity	An integral of square of derivative of curvature, divided by the length of a curve
	20. Num. of Tiny Protrusions	Small protrusions are identified as humps of positive curvature on the cell boundary which contains at least 5 boundary points
	21. Num. of Protrusions per Branch	Average number of protrusions per branch in a cell
	22. Average Boundary Interval Between Protrusions	Perimeter over the number of the protrusions

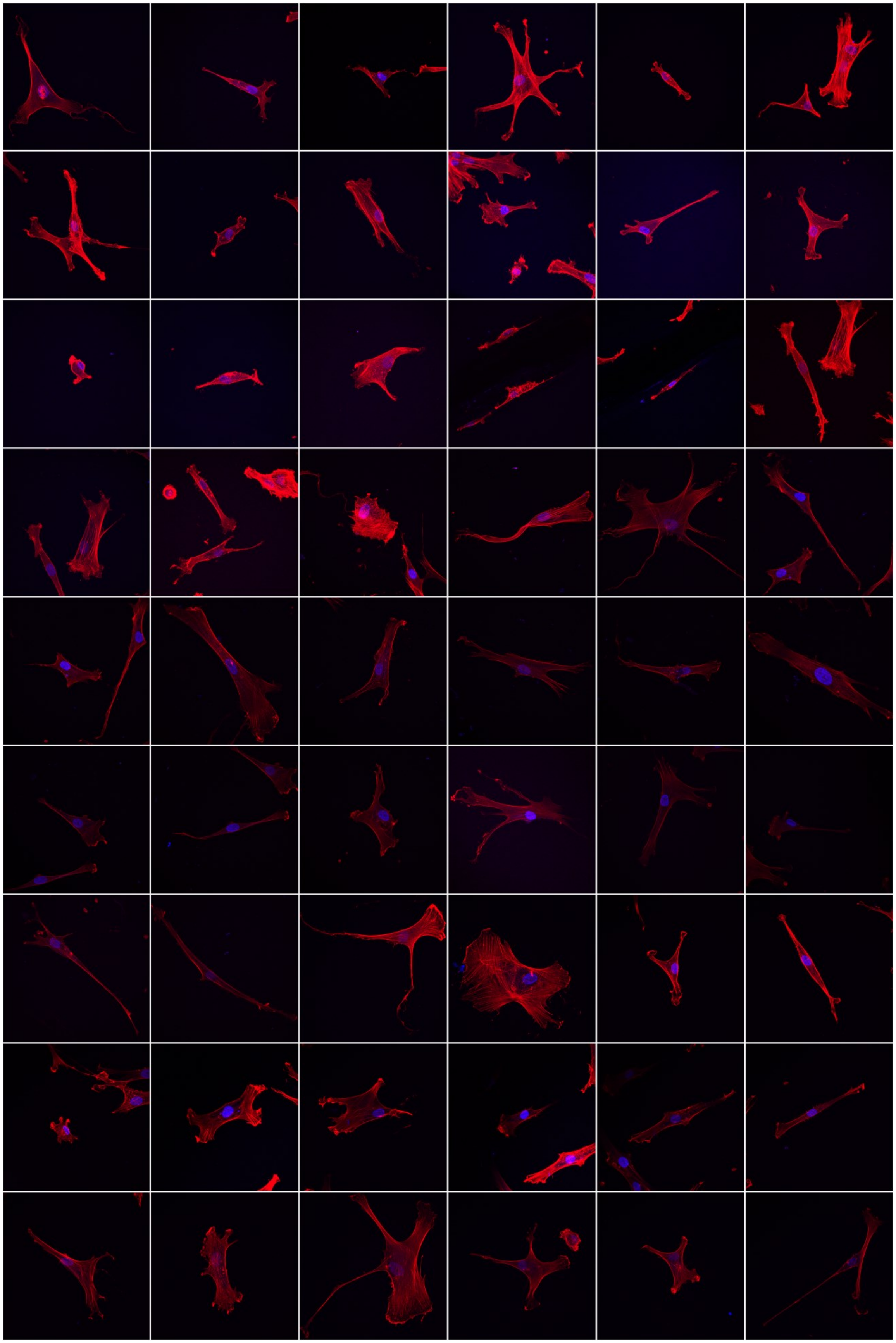


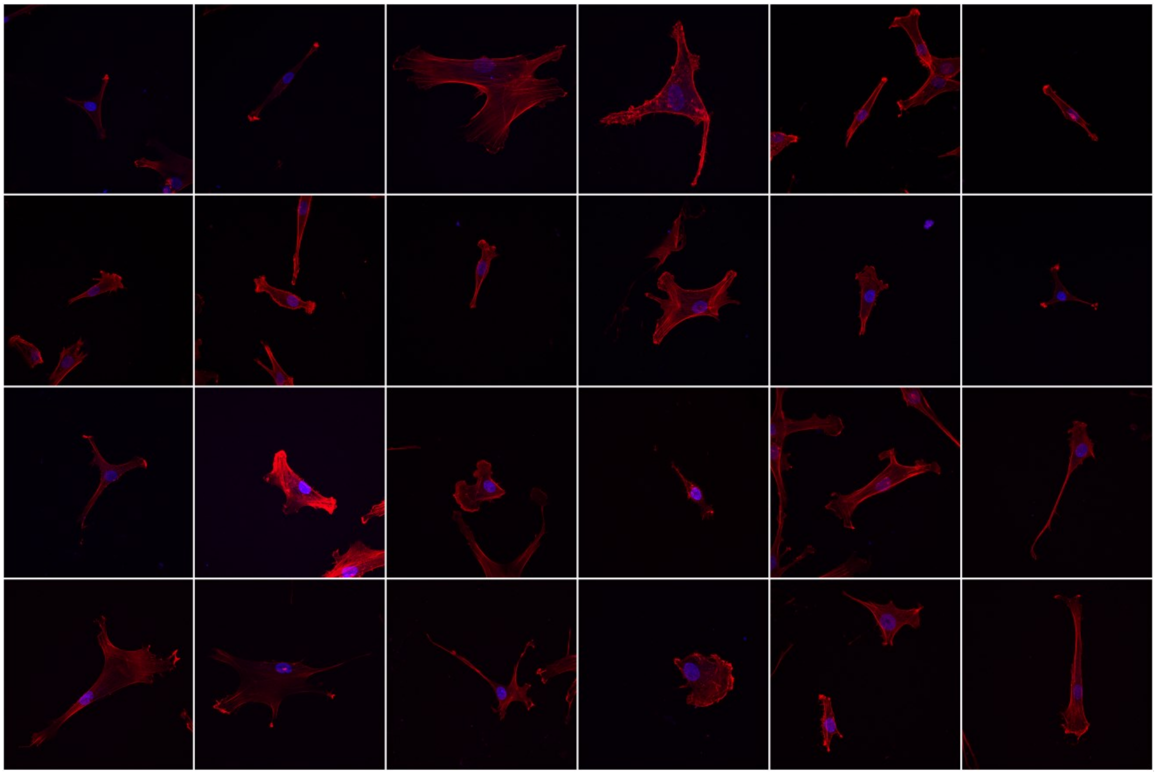




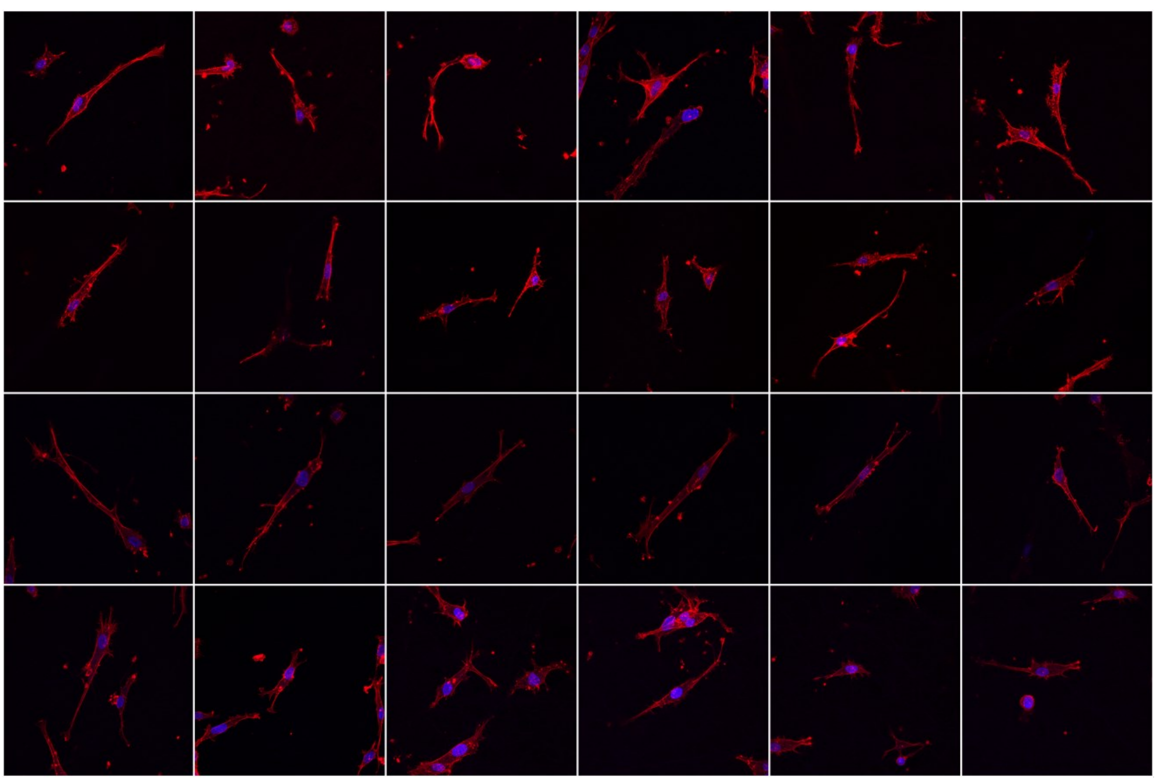
Maximum intensity projections of 3-D z-stack images in xy plane of hBMSCs of FS

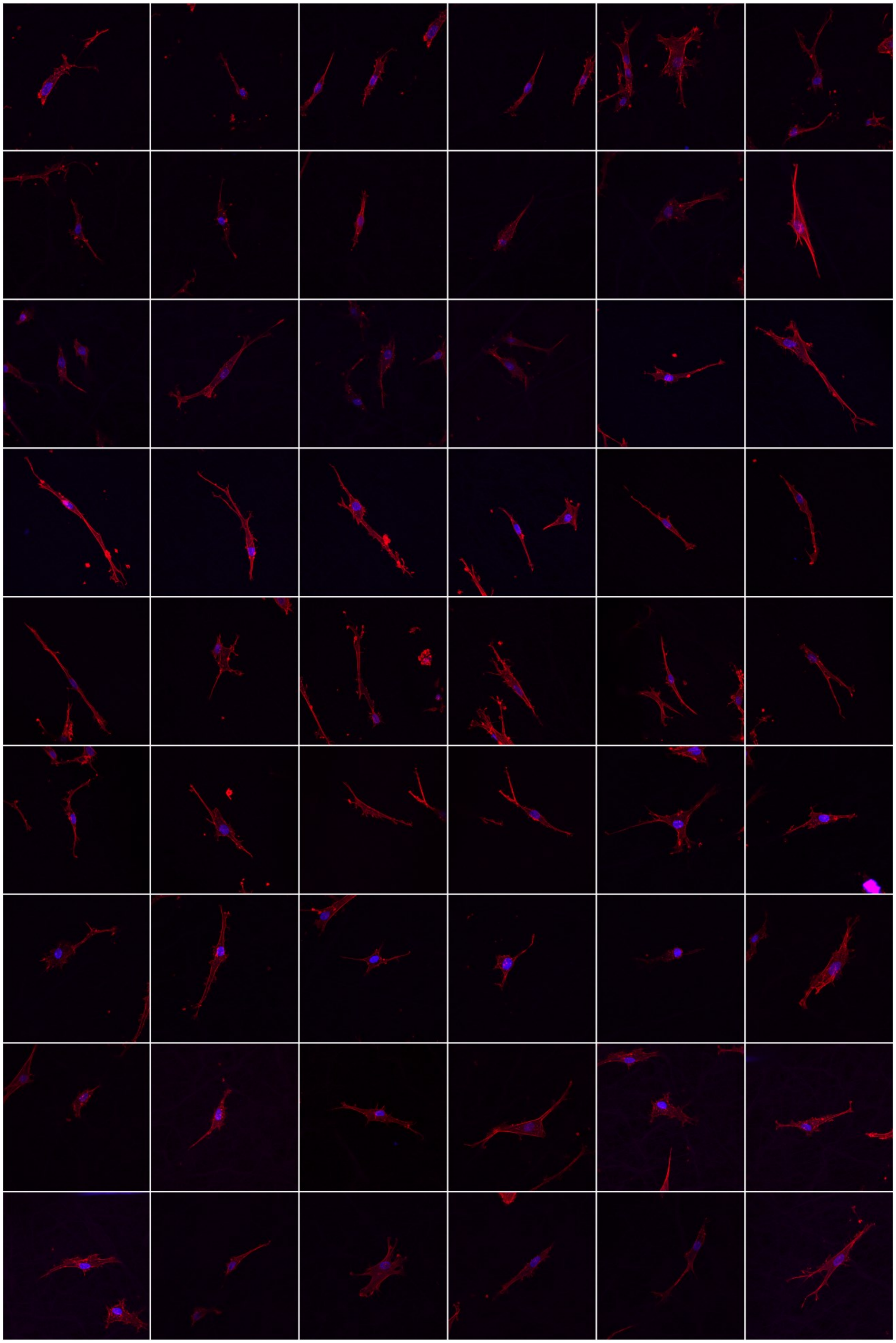


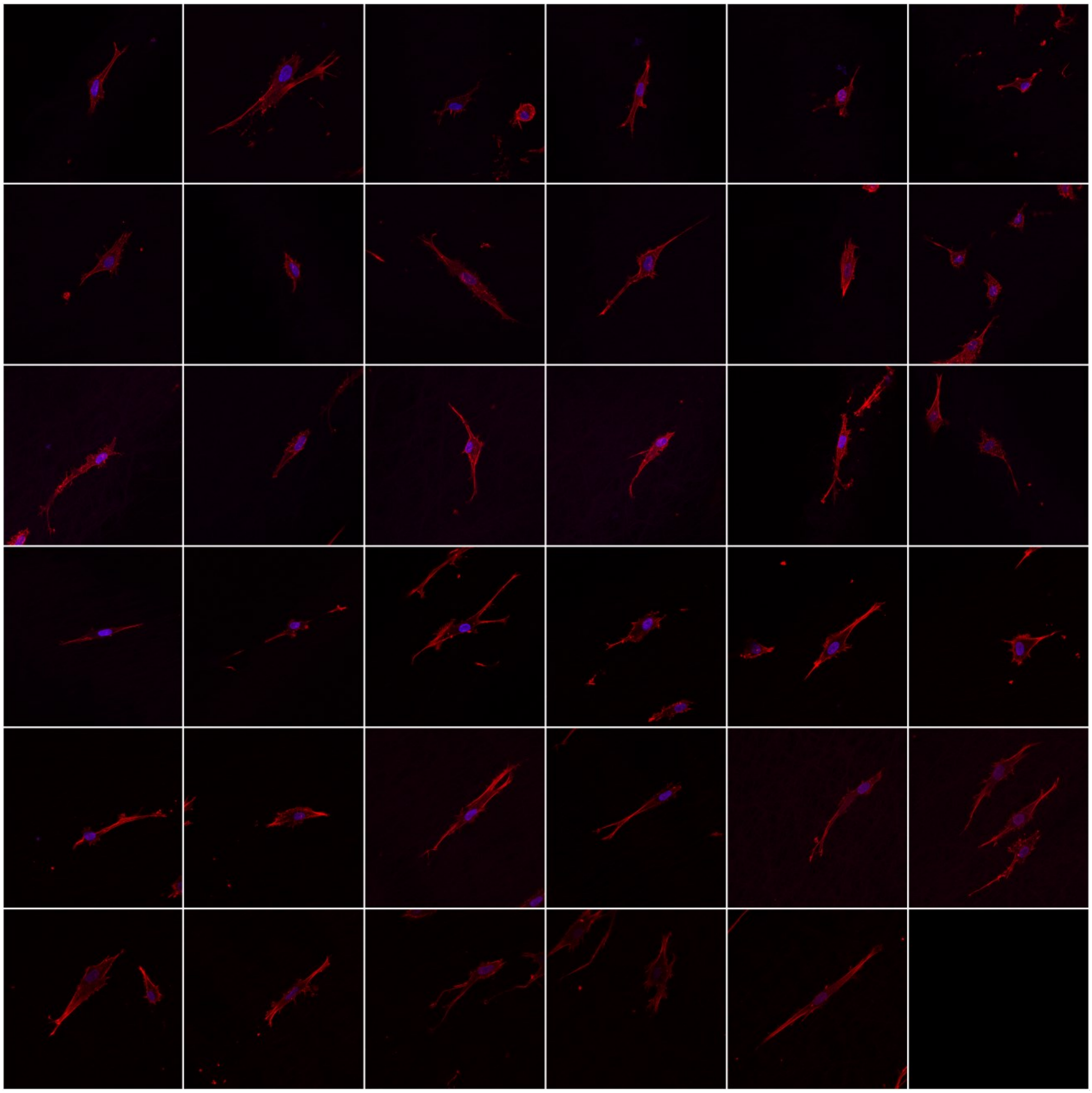




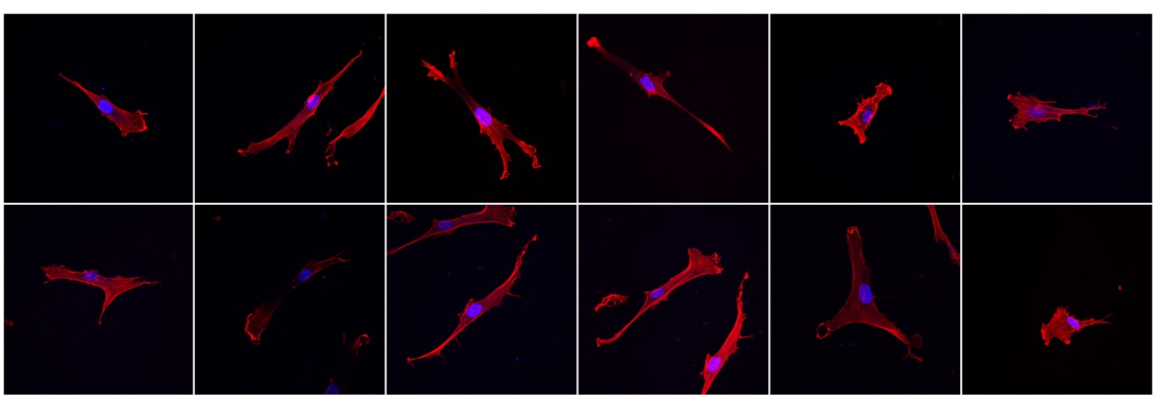
Maximum intensity projections of 3-D z-stack images in xy plane of hBMSCs of SC

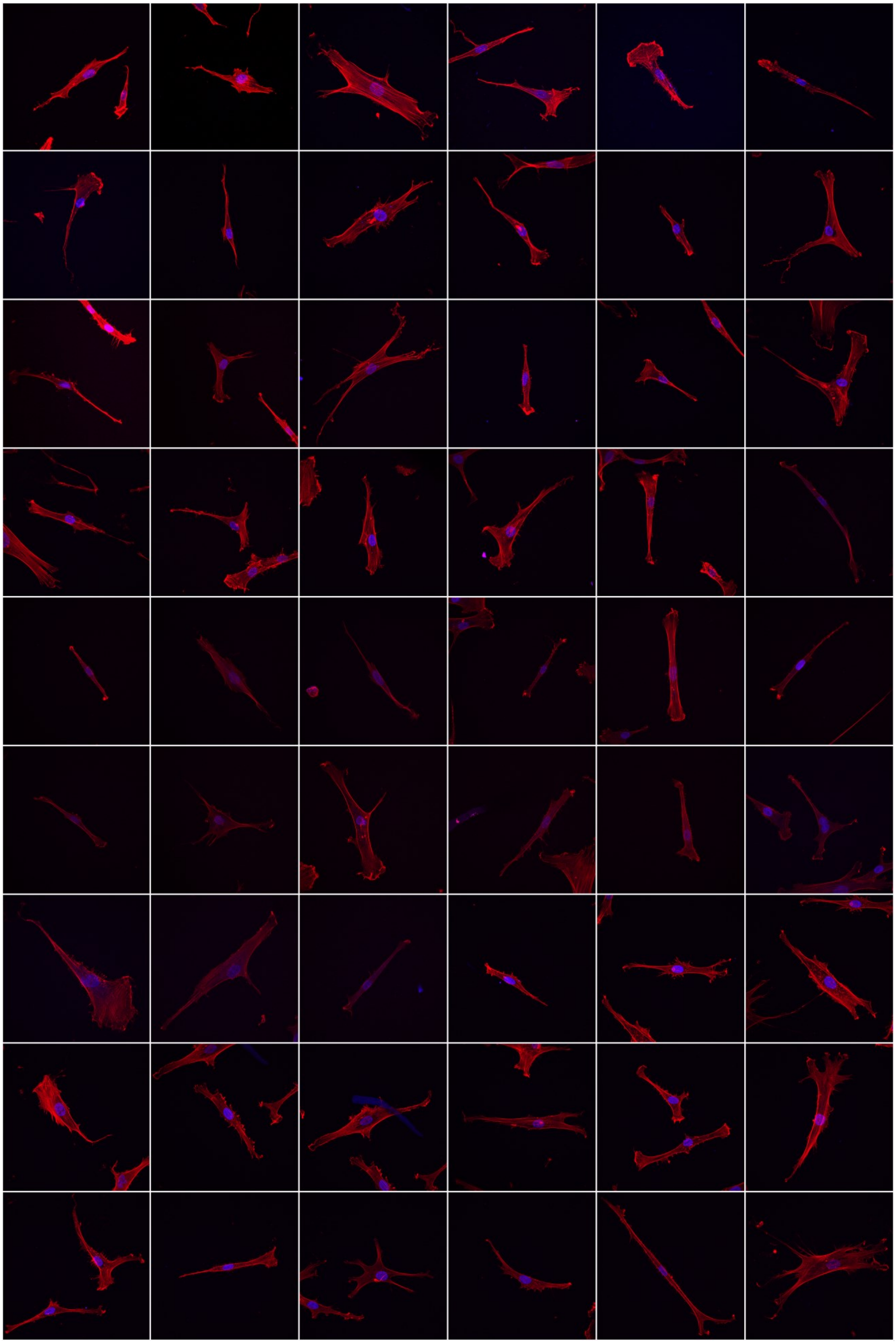


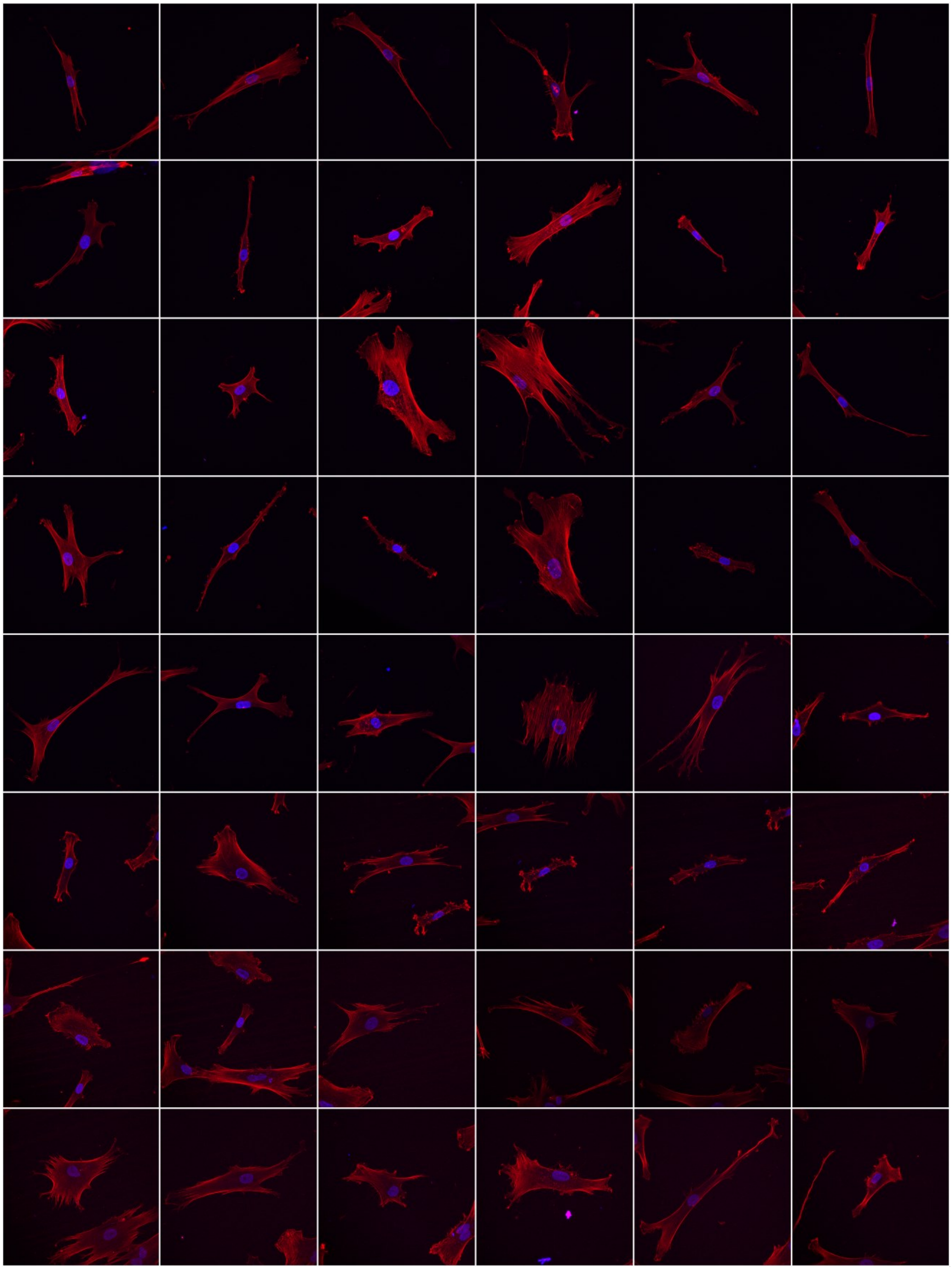




Maximum intensity projections of 3-D z-stack images in xy plane of hBMSCs of FS+OS



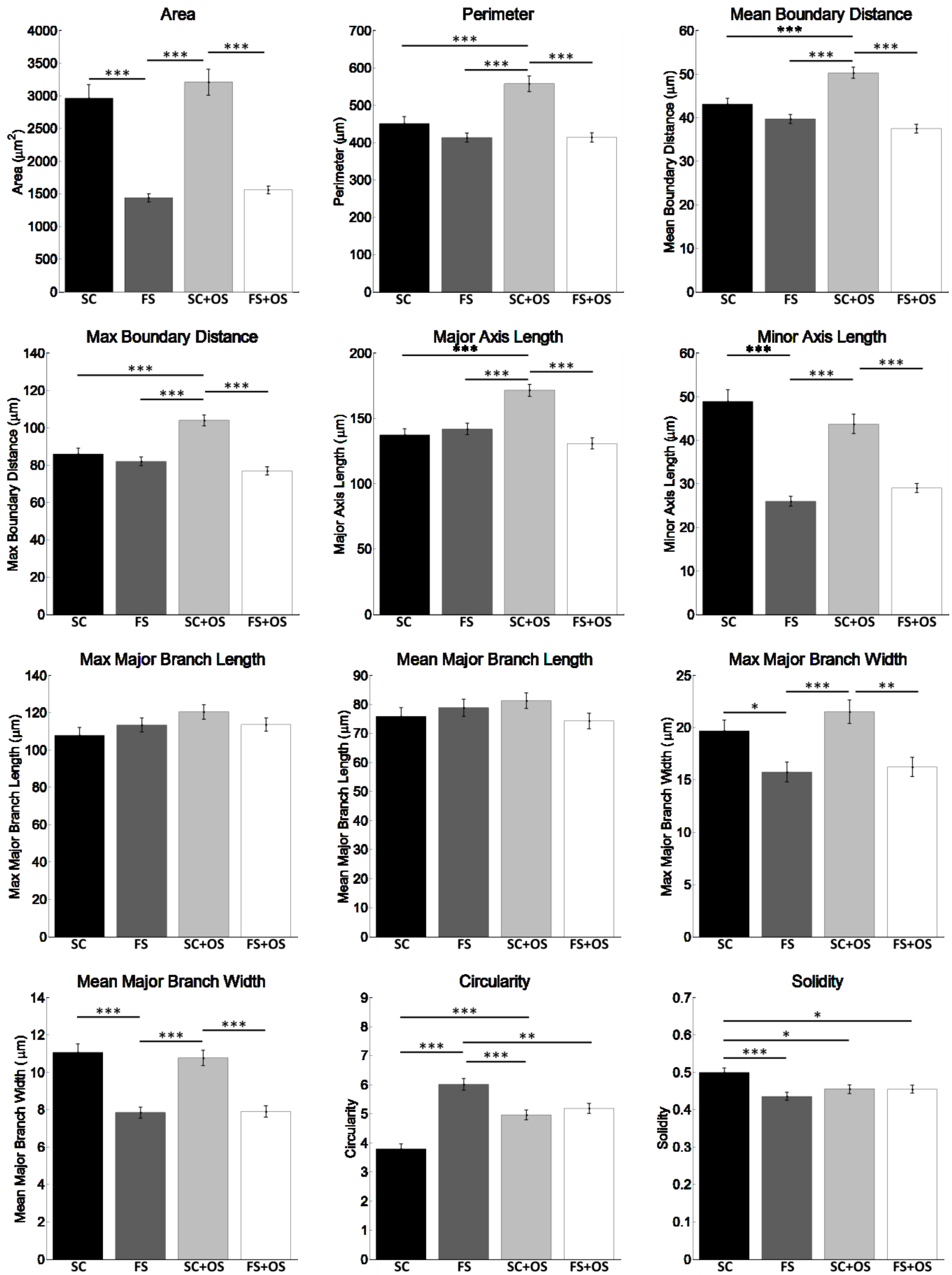


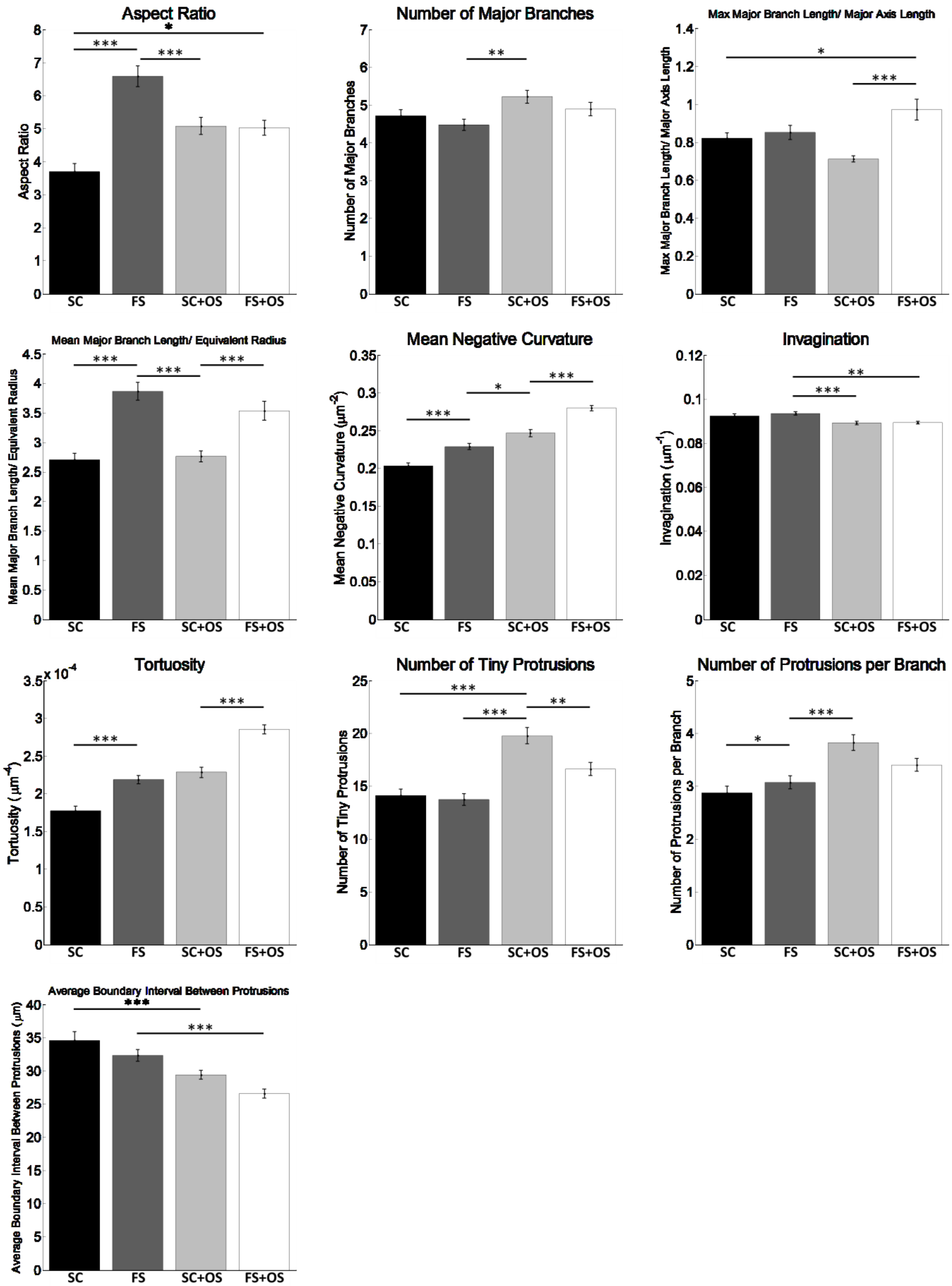


Maximum intensity projections of 3-D z-stack images in xy plane of hBMSCs of SC+OS

Supplementary Fig 1. Maximum intensity projections of 3-D z-stack images of actin (red) and nucleus (blue).

a

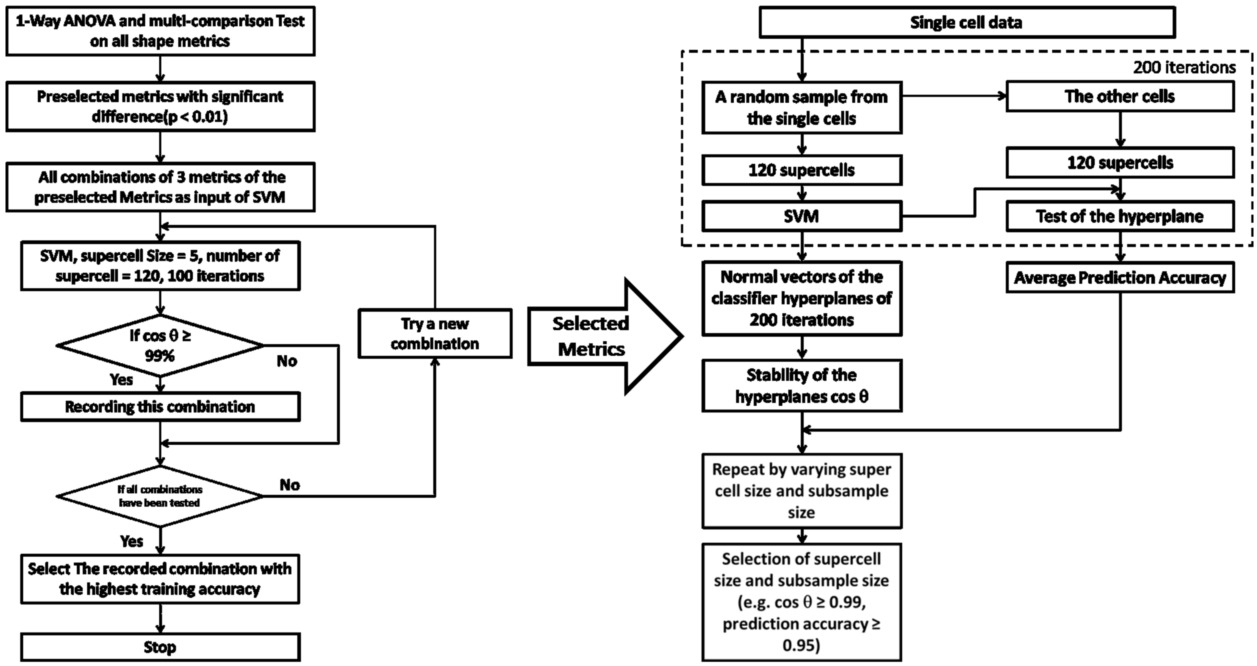

Area
Area (μm²)
Perimeter
Perimeter (μm)
Mean Boundary Distance
Mean Boundary Distance (μm)
Max Boundary Distance
Max Boundary Distance (μm)
Major Axis Length
Major Axis Length (μm)
Minor Axis Length
Minor Axis Length (μm)
Max Major Branch Length
Max Major Branch Length (μm)
Mean Major Branch Length
Mean Major Branch Length (μm)
Max Major Branch Width
Max Major Branch Width (μm)
Mean Major Branch Width
Mean Major Branch Width (μm)
Circularity
Solidity
SC
FS
SC+OS
FS+OS


Aspect Ratio
Aspect Ratio
Number of Major Branches
Number of Major Branches
Max Major Branch Length/ Major Axis Length
Max Major Branch Length/ Major Axis Length
Mean Major Branch Length/ Equivalent Radius
Mean Major Branch Length/ Equivalent Radius
Mean Negative Curvature
Mean Negative Curvature (μm-2)
Invagination
Invagination (μm-1)
Tortuosity
Tortuosity (μm-4)
x 10-4
Number of Tiny Protrusions
Number of Tiny Protrusions
Number of Protrusions per Branch
Number of Protrusions per Branch
Average Boundary Interval Between Protrusions
Average Boundary Interval Between Protrusions (μm)
SC
FS
SC+OS
FS+OS

b

Metric Category	Metrics	FS vs. SC	FS vs. FS+OS	FS vs. SC+OS	SC vs. FS+OS	SC vs SC+OS	FS+OS vs SC+OS
Cell Size	1. Area	***	ns	***	***	ns	***
	2. Perimeter	ns	ns	***	ns	***	***
	3. Mean Boundary Distance	ns	ns	***	**	***	***
	4. Max Boundary Distance	ns	ns	***	ns	***	***
	5. Major Axis Length	***	**	***	***	***	ns
	6. Minor Axis Length	***	***	*	***	***	***
	7. Max Major Branch Length	ns	**	***	*	*	ns
	8. Mean Major Branch Length	***	***	ns	***	***	***
	9. Max Major Branch Width	ns	ns	***	ns	***	***
	10. Mean Major Branch Width	***	ns	***	***	ns	***
Shape Pattern	11. Circularity	***	ns	ns	*	*	ns
	12. Solidity	***	***	***	**	**	ns
	13. Aspect Ratio	ns	**	***	*	***	**
	14. Num. of Major Branches	ns	ns	**	ns	ns	ns
	15. Max Branch Length/ Major Axis Length	ns	ns	ns	ns	ns	ns
	16. Mean Branch Length/ Equivalent Radius	ns	ns	ns	ns	ns	ns
Local Curvature Information	17. Mean Negative Curvature	*	ns	***	ns	ns	**
	18. Invagination	***	ns	***	***	ns	***
	19. Tortuosity	ns	ns	***	*	***	ns
	20. Num. of Tiny Protrusions	ns	***	ns	***	***	ns
	21. Num. of Protrusions per Branch	ns	ns	ns	*	ns	***
	22. Average Boundary Interval Between Protrusions	***	ns	***	***	ns	***

Supplementary Fig 2. (a) Shape metrics quantifying cell shapes of hBMSCs in FS, SC, FS +OS and SC+OS. All error bars represent standard error of the mean. (b) Summary of the statistical analysis of each cell shape metrics with 1-way ANOVA and Tukey Multicomparison test (ns: $p > 0.05$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$).



Supplementary Fig 3. Flow chart of the training and feature selection procedure followed by the flow chart of subsampling validation for the selected metrics.