Biomaterials 104 (2016) 104–118

Contents lists available at ScienceDirect

Biomaterials

journal homepage: www.elsevier.com/locate/biomaterials

Machine learning based methodology to identify cell shape phenotypes associated with microenvironmental cues

Desu Chen [a], Sumona Sarkar [c], Julián Candia [b, d, e], Stephen J. Florczyk [c], Subhadip Bodhak [c], Meghan K. Driscoll [b], Carl G. Simon Jr. [c], Joy P. Dunkers [c], Wolfgang Losert [b, *]

[a] Biophysics Program, University of Maryland, College Park, MD, United States
[b] Department of Physics, University of Maryland, College Park, MD, United States
[c] Biosystems & Biomaterials Division, National Institute of Standards & Technology, Gaithersburg, MD, United States
[d] School of Medicine, University of Maryland, Baltimore, MD, United States
[e] Center for Human Immunology, National Institutes of Health, Bethesda, MD, United States

ARTICLE INFO

Article history:
Received 27 April 2016
Received in revised form
17 June 2016
Accepted 19 June 2016
Available online 25 June 2016

Keywords:
Cell morphology
Machine learning
Supercell
Fibrous substrates
Stem cell

ABSTRACT

Cell morphology has been identified as a potential indicator of stem cell response to biomaterials. However, determination of cell shape phenotype in biomaterials is complicated by heterogeneous cell populations, microenvironment heterogeneity, and multi-parametric definitions of cell morphology. To associate cell morphology with cell-material interactions, we developed a shape phenotyping framework based on support vector machines. A feature selection procedure was implemented to select the most significant combination of cell shape metrics to build classifiers with both accuracy and stability to identify and predict microenvironment-driven morphological differences in heterogeneous cell populations. The analysis was conducted at a multi-cell level, where a "supercell" method used average shape measurements of small groups of single cells to account for heterogeneous populations and microenvironment. A subsampling validation algorithm revealed the range of supercell sizes and sample sizes needed for classifier stability and generalization capability. As an example, the responses of human bone marrow stromal cells (hBMSCs) to fibrous vs flat microenvironments were compared on day 1. Our analysis showed that 57 cells (grouped into supercells of size 4) are the minimum needed for phenotyping. The analysis identified that a combination of minor axis length, solidity, and mean negative curvature were the strongest early shape-based indicator of hBMSCs response to fibrous microenvironment.

1. Introduction

The morphology of a cell is influenced by a combination of many intracellular mechanical processes, interactions with other cells and the surrounding extracellular matrix [1–7]. Thus, cell morphology reflects the integrative effect of many distinct processes and signaling pathways across different scales [4,5] and may be a valuable descriptor of cell behaviors in differentiation [8–14], function or dysfunction [15], migration [16–18] and cancer progression [19]. For example, a recent study by Marklein et al. [8] demonstrates over 90% accuracy in the prediction of day 35 mineralization of human bone-marrow derived mesenchymal stem cells (hMSCs) cultures of varying donors and passages based on day 3 cell morphology. In another recent study by Unadkat et al. [10], cell morphology was also investigated as an indicator of cell genotypic and phenotypic responses. Beyond being a possible indicator, some studies have shown that either affecting cell morphology with surface topographical cues [20–23] or directly manipulating cell morphology through geometric constraints of cell adhesive regions can elicit genotypic or phenotypic alterations [5–7,24]. Therefore, cell morphology may contribute as a descriptor, indicator or intermediate factor in characterizing cell-material interactions. High-throughput single-cell bioimaging has enabled the quantification of heterogeneous cell population with many cell shape features that are increasingly difficult to interpret. In addition, the complex biomaterial microenvironment can also

* Corresponding author.
E-mail address: wlosert@umd.edu (W. Losert).

http://dx.doi.org/10.1016/j.biomaterials.2016.06.040
0142-9612/

contribute to the heterogeneity of cell shape response. Innovative analytical tools must be developed to identify and combine key cell shape features correlated with biological outcome while accounting for both multi-parametric complexity and biological heterogeneity.

Multi-parametric single-cell data are widely used in biomaterials studies with technologies such as bioimaging, single-cell PCR and flow cytometry. In order to associate multi-parametric single cell data with cell-material interactions, appropriate computational and statistical tools are required to quantify the informative content of data and describe differences between cell populations. Common statistical methods typically used are Student's *t*-test and ANOVA analyses. These approaches describe differences of the multi-parametric data by comparing the values of each single metric across different cell populations with a statistical hypothesis test which outputs a p-value [25,26]. This has proven valuable to determine individual metrics that may be important in characterizing cell-material interactions. However, if we intend to describe the cell phenotypes for cell populations with more comprehensive representations by combining multiple metrics, these approaches are limited as they omit correlations between metrics in describing cell population differences.

Representations of multi-parametric data can be obtained by other statistical methods, such as principal component analysis (PCA) and singular value decomposition (SVD) [8,27,28]. More recent methods (for instance, self-organizing maps [29] and multidimensional scaling [9]) can achieve reduced multidimensional representations of cell morphology. However, these methods bring other limitations. In particular, they are not designed to separate different classes optimally and, the achieved dimensional reduction introduces more abstract descriptions of the system in terms of linear or non-linear combinations of metrics, bringing difficulties to determine relevant features in defining the cell phenotypes. To address these limitations, we have developed an approach to overcome several of these limitations by generating multi-dimensional linear classifiers that allow simple interpretation for classification and phenotyping in reduced metric space.

In this study, we investigated the morphology of human bone marrow stromal cells (hBMSCs) in fibrous substrates compared to that of cells on flat films (Fig. 1a) in presence or absence of osteogenic differentiation media. Fibrous materials are widely used in both research and clinical applications of tissue engineering and regeneration medicine. Previous studies had demonstrated that hBMSCs cultures on fibrous substrates developed osteogenic differentiation after 50 days of culturing without any osteogenic supplement [21]. Morphological response of hBMSCs in fibrous substrates is being investigated as a possible mechanism for osteogenic differentiation observed in this microenvironment [21,30–33]. This hypothesis is supported by several studies describing mechanistic associations between hMSCs shape and subsequent differentiation [5–7]. However, only a few individual cell shape features have been investigated for their association with differentiation, and cell morphologies vary greatly across a fibrous substrate. To address this limitation, we have developed an analysis framework for multi-parametric single-cell data based on support vector machines (SVMs) [34–36] to quantify shape differences of hBMSCs populations and associate them with different microenvironments (Fig. 1b). SVM classifiers are designed to find the optimal classification boundary that separates data points in the multidimensional shape metric space. We investigated a wide range of shape metrics to quantify global and local shape features, including for example cell size and aspect ratio, cytoskeletal branching, and local boundary curvature. Moreover, the resulting SVM classifiers provided a selection of reduced shape metrics to quantify hBMSCs shape phenotypes in specific microenvironments.

The heterogeneous cell population and the heterogeneous microenvironment may cause variability in cell morphology, where difference between shapes of single cells within the same culture environment are observed. Within the SVM scenario, variability in cell morphology can lead to highly overlapping cell populations and, thus poorly performing classifiers on the single-cell level. In order to address single-cell heterogeneity from different sources, a method of averaging shape metrics over a small subset of randomly selected cells known as “supercell” averaging [36,37] was implemented to improve the training and prediction accuracies of the SVM classifiers. Instead of solely focusing on phenotypes on single-cell level, the SVM/supercell paradigm allowed consideration of cell shape phenotypes associated to small groups of cells, i.e. “supercells”. The random sampling used to generate supercells can introduce uncertainty in the SVM classifier. The tradeoff between prediction accuracy, supercell averaging and uncertainty in the classifier were quantitatively determined in this study. Furthermore, by introducing a subsampling validation procedure, we studied the sample size as another important limiting factor in the construction of single-cell or supercell phenotypes and its effects on classifier prediction accuracy. By combining multiple metrics and learning at small cell group levels, the SVM/supercell paradigm quantitatively identified changes in population behavior of cell morphology for four different conditions. Building on this approach, a systematic analysis of the cell morphological response to the physical and biochemical properties of their surrounding microenvironment could be carried out.

2. Materials and methods

Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

2.1. Preparation of substrates

To fabricate poly(ε-caprolactone) (PCL) fibrous substrates (FS), PCL solution (0.15 g/mL in 5: 1 vol ratio of chloroform: methanol) was dispensed by a 3 mL syringe and pump (0.5 mL/h) through a 21 gauge 1″ shaft, flat tip, dispensing needle over an array of 0.95 cm^2 tissue-culture polystyrene (TCPS) disks arranged on a grounded aluminum foil over a 6 h period. The distance between the syringe needle and the target TCPS disk array was 20.4 cm. The needle was connected to a positive lead of 13.5 kV. To help the adhesion of PCL fibers over the target TCPS disk array, the disks were sprayed with 70% by mass ethanol solution every 10 min to enhance fiber deposition to the disks. The diameter of fabricated PCL fibers was (589 ± 116) nm (n = 151) as determined by scanning electron microscopy imaging (2.00 kV, 5000×).

PCL spin-coated films (SC) were prepared by spin-coating PCL solution (0.7 mL, 0.1 g/mL in glacial acetic acid) on tissue-culture polystyrene dishes at 1100 rpm for 30 s. Films were air dried at room temperature overnight and heated above 60 °C for 4 to 5 times in order to achieve a reproducible cobblestone pattern in the films [38]. Films were punched into disks of 0.95 cm^2. The surface roughness of the SC is (92.76 ± 10.69) nm determined by atomic force microscopy.

FS and SC disks were placed in 48-well tissue-culture polystyrene plates. Plates were sterilized by ethylene oxide for 12 h and then purged under vacuum for 2 days. Before cell seeding, each well was fully wetted with basal cell culture media (α-minimum