# Image Informatics for Studying Signal Transduction in Cells Interacting with 3D Matrices

Dimitrios S. Tzeranis*[a], Jin Guo[b], Chengpin Chen[b], Ioannis V. Yannas[a,c], Xunbin Wei[b], Peter T.C. So[a,c]

[a]Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; [b]Institutes of Biomedical Sciences, Fudan University, 138 Yi Xue Yuan Road, Shanghai 200032, China; [c]Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

## ABSTRACT

Cells sense and respond to chemical stimuli on their environment via signal transduction pathways, complex networks of proteins whose interactions transmit chemical information. This work describes an implementation of image informatics, imaging-based methodologies for studying signal transduction networks. The methodology developed focuses on studying signal transduction networks in cells that interact with 3D matrices. It utilizes shRNA-based knock down of network components, 3D high-content imaging of cells inside the matrix by spectral multi-photon microscopy, and single-cell quantification using features that describe both cell morphology and cell-matrix adhesion pattern. The methodology is applied in a pilot study of TGFβ signaling via the SMAD pathway in fibroblasts cultured inside porous collagen-GAG scaffolds, biomaterials similar to the ones used clinically to induce skin regeneration. Preliminary results suggest that knocking down all rSMAD components affects fibroblast response to TGFβ1 and TGFβ3 isoforms in different ways, and suggest a potential role for SMAD1 and SMAD5 in regulating TGFβ isoform response. These preliminary results need to be verified with proteomic results that can provide solid evidence about the particular role of individual components of the SMAD pathway.

**Keywords:** Image informatics, cell-matrix interactions, spectral multi-photon imaging, TGFβ signaling, SMAD pathway, myofibroblast differentiation, wound contraction

## 1. INTRODUCTION

Cells sense and respond to their environment via signal transduction pathways, cascades of proteins and small molecules that transmit information (in the form of molecule concentration and modification) inside the cell [1-3]. Existing knowledge about signal transduction pathways is based on many years of experiments using established biochemical and genetic methods. However as signal transduction studies adopt system-wide approaches and reach larger scales, and require information-rich datasets, there is increasing need for novel high-throughput genomic and proteomic [4-7], or imaging-based methods.

One imaging-based method for studying cell signaling is image informatics. Image informatics refers to methodologies that combine cell perturbation (e.g. via genetic manipulation), large scale imaging, image processing and statistics to study complex signal transduction pathways. So far, image informatics has been applied in studies of cells cultured on standard culture dishes [8-12]. Unfortunately, the 2D surface of a culture dish presents to cells an environment that is very different compared to the microenvironments felt inside tissues. Since the insoluble microenvironment of cells affects critically their phenotypes [13], it is important to develop imaging-based studies that can study cells inside more physiologically relevant 3D matrix models.

This study presents an implementation of image informatics that can be used to study signal transduction pathways in cells that interact with a 3D matrix. A key feature compared to previous image informatics approaches is that the output of the cell-matrix system is quantified not just based on morphometric features of the cells, but also on morphometric features that describe the cell-matrix adhesion pattern. This study provides preliminary data from a pilot study of TGFβ isoform signaling via the SMAD pathway, a mechanism that has been implicated in regulating wound contraction and myofibroblast differentiation, both of which are critical phenotypes of cancer pathology and wound healing [14-16].

# 2. MATERIALS AND METHODS

## 2.1 The Concept of Image Informatics

Image informatics refers to single-cell imaging-based approaches that infer information about the components of a signal transduction pathway. The pathway of interest is modeled as a graph that consists of interacting proteins (nodes) that interact with each other in some particular topology (edges). The system is triggered by extracellular stimuli (e.g. growth factors, cytokines, hormones), and transmits signals through the nodes (usually via phosphorylation post-translational modification) towards either the cell nucleus where it guides gene expression or other cytosolic processes (e.g. cytoskeletal remodeling, apoptosis) [1]. In some cases the topology between the pathway proteins is only party known. In other cases, there is a good a-priori knowledge about the elementary interactions between the proteins of the pathway (based on the vast literature of biochemistry, protein binding assays, and genetics), however the relative importance of the nodes and edges can be cell-specific or it may be altered due to pathology or drug effects (e.g. a healthy cell and a cancer cell may use different edges of the graph to transmit signals induced by the same stimuli). In image informatics, the pathway of interest is modeled as a system whose outputs are single-cell assays (features) that are quantified based on images. Image informatics infers the roles of individual pathway components by examining how the input-output response of the system is altered when one or more components of the system are knocked down.
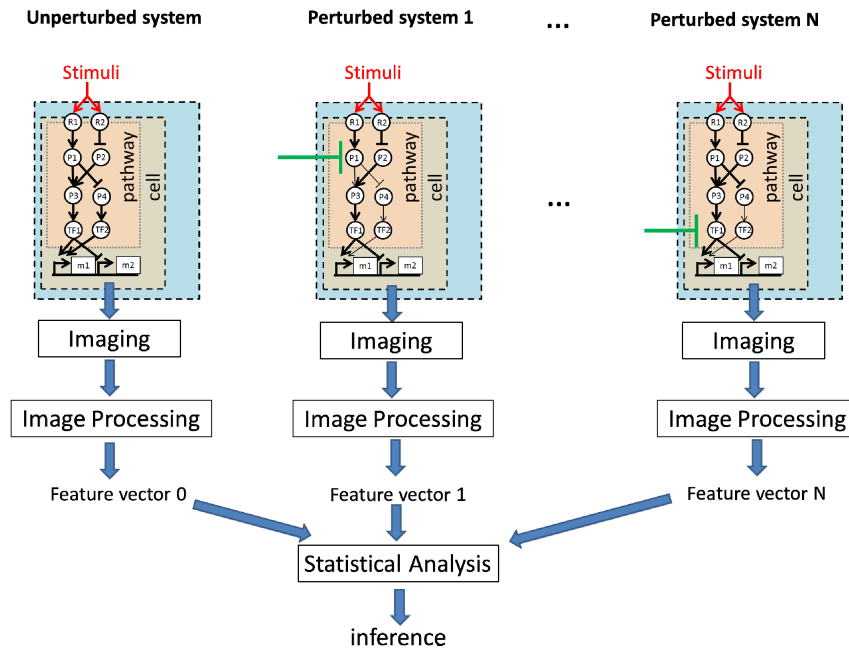


Figure 1. Overview of the image informatics approach. The role of individual components of a signal transduction pathway is inferred by quantifying how the response (evaluated by optical-based assays called features) of the system to particular stimuli changes when one or more components of the pathway are knocked down.

## 2.2 Implementing Image Informatics for Signal Transduction Studies in Cells Interacting with 3D matrices

The proposed image informatics pathway consists of five steps, see Figure 2.
- *System perturbation*: knock down (or least significantly down-regulate) one or more components of the pathway of interest in a cell line of interest. This can be achieved by RNAi or by using small molecules (e.g. drugs, kinase inhibitors). In this study, stable knock-down of one SMAD pathway component in primary human fibroblasts is implemented by lentiviral-based stable expression of shRNA that targets the target protein mRNA. Four rSMAD isoforms (SMAD1,2,3,5) were targeted using commercial shRNA vectors (Sigma Aldrich).
- *System stimulation*: Wild type (WT) cells or cells whose component X is knocked down ($X^-$) are seeded inside the 3D matrix of interest. After cell adhesion on the matrix, cells are stimulated by molecules known to excite the pathway of interest. In this study, fibroblasts (WT, or shRNA-treated) are stimulated by TGFβ1 or TGFβ3, growth factors known to excite the SMAD pathway [17]. The porous collagen-GAG scaffolds used are similar to

biomaterial grafts used clinically to induce skin regeneration [14]. After stimulation, cells are incubated for appropriate duration (in this study 5 days) before quantification.

- *Quantification*: The cell-matrix system is imaged by a 3D imaging microscopy modality in order to quantify the system response using single-cell imaging-based assays. In this study, fibroblasts were fixed, immuno-labeled with mouse anti-α smooth muscle actin (αSMA) antibody detected by an Alexa Fluor488-conjugated anti-mouse antibody, and then counter-stained using the CMTMR cell tracker and the Hoechst33342 nucleic stain. 3D images were acquired by a spectral multi-photon microscope equipped with a 16-channel multi-anode PMT and single-photon counting capability [18]. Alternatively, some cell-matrix samples can be lysed and analyzed by standard proteomic assays.

- *Image Processing*: The purpose of image processing is to identify single cells in the imaging data, and then calculate for each cell a vector of N features. The system response of each treatment condition (TC; corresponds to applying a particular stimulation to a particular cell type) is then described by a feature matrix, whose M columns contain the N features of the M single cells identified from the images. This step is quite complex, instrument-dependent, and computational intensive. In this study, spectral multi-photon data are initially processed using a probabilistic processing pipeline that was developed for processing low-signal high-content spectral images (Tzeranis et al., in preparation). The pipeline utilizes spectral unmixing and Bayesian classification to segment the image pixels into four classes based on the known types of objects known to be present in each class: the "cytoplasm" class (contains CMTMR, Alexa fluor 488), the "nucleus" class (contains CMTMR, Hoechst 33342), the "scaffold" class (contains fluorescence and second harmonic emission by collagen), and the "void" class (empty space between cells and scaffold). After image segmentation, single cells are identified using published seeding and growing algorithms [19,20]. Due to the complexity of this step, the validity of the identified cell objects need to be manually verified by a human user. Finally, for each cell, a set of 45 morphometric features are calculated. The features describe the shape of the cell (based on the CMTMR emission), the shape of the nucleus (based on Hoechst33342 emission), the shape of cell-matrix adhesion pattern (this is approximated based on "cytoplasm" pixels that neighbor "scaffold" pixels), and the expression of contractile phenotypes by cells (based on the Alexa Fluor 488 emission).

- *Statistical Processing*: Statistical processing consists of two steps. Low-level statistical analysis quantifies the degree of "difference" (using some sort of "distance" metric) between the feature arrays of different TC. High-level statistical analysis combines the "differences" between the responses of different TC with a-priori knowledge on the biology of the signal transduction pathway of interest in order to infer conclusions about the role of pathway components. This study implements only low-level statistical processing and follows the procedure proposed by [12]: the feature matrix of each TC is initially normalized and its dimension is reduced by principal component analysis. The "difference" between the responses of two TC is calculated using the Mahalanobis distance metric.
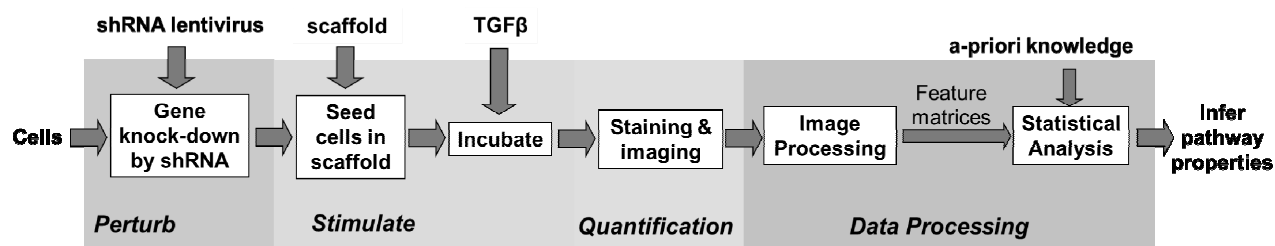


Figure 2. Schematic of the steps of the proposed image informatics methodology as applied in a pilot study of TGFβ isoform signaling via the SMAD pathway in primary fibroblasts inside porous collagen scaffolds.

## 3. RESULTS

Figure 3 shows representative images of the outcome of the various steps of the image processing pipeline that converts raw spectral multi-photon microscopy data into single cell objects that are ready for quantification. Figure 3a shows an intensity image obtained by summing the photon counts of the 16 channels of the spectral image. The staining protocol has been optimized so that the mean fluorescent emission from "cytoplasm" pixels and "nucleus" are at most one order of magnitude larger compared to the intrinsic fluorescent emission from scaffold strut pixels. Figure 3b shows the outcome of spectral unmixing. Results show that Hoechst33342 emission is localized in cell nuclei, CMTMR emission in the whole volume of cells (cytoplasm and nucleus), collagen emission is present in the scaffold, and Alexa Fluor 488 emission (label for expression of contractile αSMA fibers) is indeed present in the cytoplasm of some cells. Figure 3c shows the outcome of the image segmentation algorithm. Images of fibroblasts inside collagen scaffolds reveal that in

general fibroblasts and myofibroblasts (cells that express significant αSMA) interact extensively with their surrounding matrix and much less with each other. Although a few clusters of cells of moderate size are present (for example the one located in the bottom left part of Figure 3c), neither in WT cells nor in any one of the SMADx⁻ cells there was observed a tendency for cells to form cell niches. Figure 3d highlights the selected single-cell objects that will be quantified. Usually a small percent of the cells (less than 30%) present in each field satisfies the necessary conditions and is selected for further processing. The remaining cells are not selected either because they lack a nucleus (it is located in a plane not present in the data) or because they are part of a big cell cluster that makes cell seeding/splitting calculations hard.
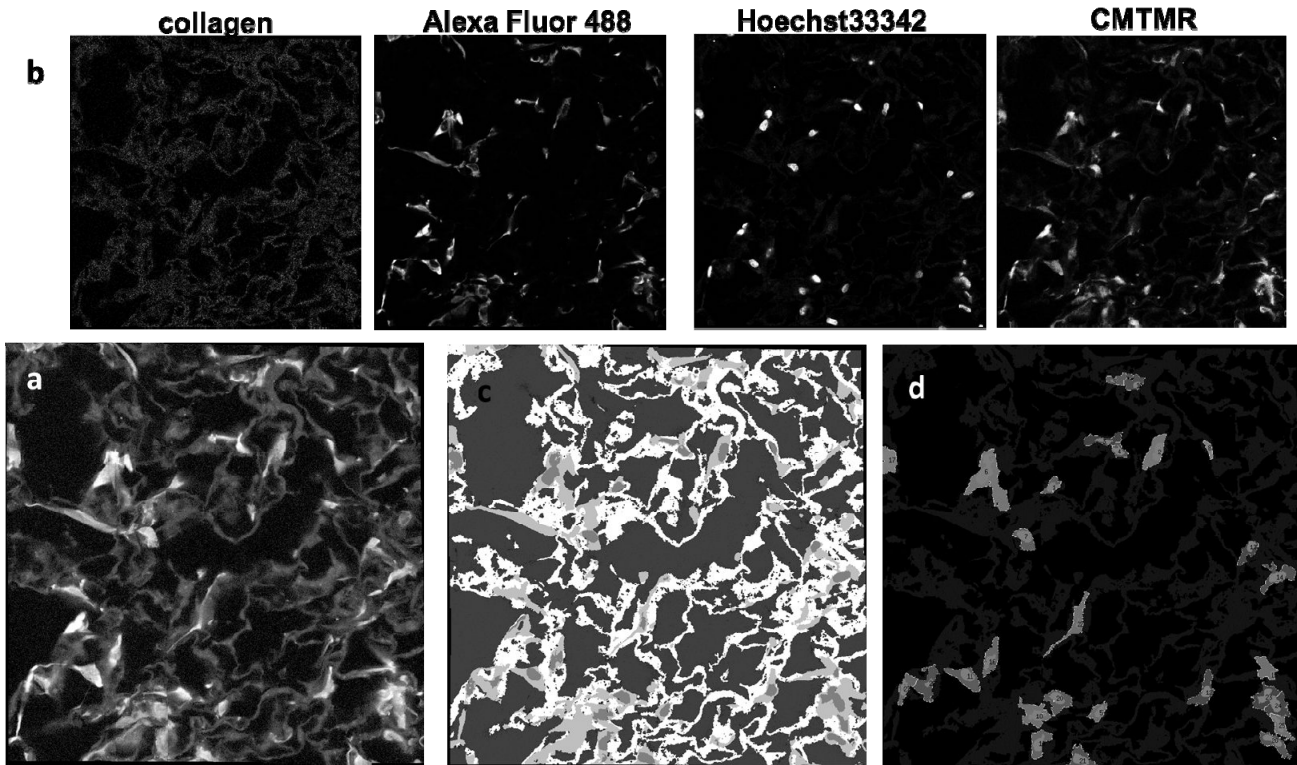


Figure 3. Representative images of the various steps in the image processing pipeline of the image informatics study. a: intensity data (sum of the signal from all 16 channels of the MA-PMT detector). Data are normalized between 0 and 100 photons per pixel and processed by gamma correction (set to 0.5 in order to enhance visualization), b: Outcome of spectral unmixing. Images show the emission of collagen (0-4 photons/pixel), alexa fluor 488 (0-80 photons/pixel), hoechst33342 (0-20 photons/pixel), and CMTMR (0-20 photons/pixel). c: outcome of the image segmentation algorithm. Black: void, Dark gray: nuclei, Light gray: cytoplasm, White: scaffold. d: Selected cell objects to be quantified. The field of view of each image is approximately 400×400 μm.

Figure 4 shows the Mahalanobis distance between different treatment conditions (TC) as calculated based on the 45 single-cell features after normalization and dimension reduction. Figure 4a displays the effect of knocking down a single rSMAD component on the response of cells to the two TGFβ isoforms (compared to WT cells). It reveals that in the absence of stimulation (No TGFβ), the Mahalanobis distance between WT fibroblasts and SMADx⁻ fibroblasts is small. This result suggests that knocking down a rSMAD component does not affect the cell state significantly. Figure 4a reveals that knocking down SMADs affects cell response to both TGFβ isoforms. The response to TGFβ1 is mostly affected by SMAD3 and SMAD1. The response to TGFβ3 is less affected by SMAD knock-down, mostly SMAD1 and SMAD5. Figure 4b shows the effect of different TGFβ isoforms to different kinds of cells. It shows that in WT fibroblasts, the response of cells to TGFβ1 or TGFβ3 is much more different compared to the response when there is no TGFβ compared to SMADx⁻ fibroblasts. Figure 4b also suggests that SMAD5⁻ fibroblasts are much less responsive to TGFβ isoforms, SMAD1⁻ and SMAD2⁻ fibroblasts are less responsive to both TGFβ isoforms, while SMAD3⁻ fibroblasts are much more responsive to TGFβ1 and less to TGFβ3.
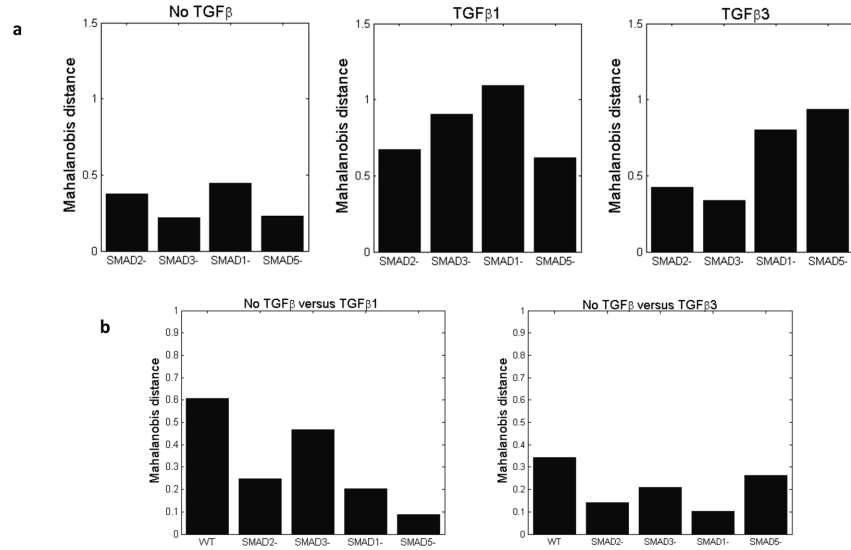
Figure 4. Mahalanobis distance between treatment conditions reveals how knocking down SMAD components affect cell response to TGFβ isoforms. **a**: Mahalanobis distance between WT fibroblasts and SMADx⁻ fibroblasts for the three kinds of stimulation considered in this study. **b**: Mahalanobis distance between different stimulations for the five kinds of cells (WT fibroblasts and SMADx⁻ fibroblasts) considered in this study.

## 4. DISCUSSION

This study describes an implementation of the image informatics methodology for studying signal transduction pathways in cells that interact with 3D matrices. This is an advancement compared to published implementations that study cells cultured on flat plastic surfaces that lack a physiologically relevant ECM. Due to the complexity and the 3D nature of the cell-matrix system, imaging features need to be calculated based on images acquired by 3D imaging modalities. Specifically, this study utilizes spectral multi-photon microscopy in order to increase the information content of the acquired images and therefore reduce the number of necessary experiments (spectral imaging enables to use the same image in order to quantify both cell morphology features and cell-matrix adhesion features). Due to the presence of the matrix, the signal transduction pathway is considered to be a part of an integrated cell-matrix system, whose state is described using imaging-based features that describe both cells (morphology, αSMA expression), and cell-matrix interactions (extent, topology). In contrast, previous image informatics implementations use only features that describe cell morphology.

The developed methodology has been applied in a pilot study of TGFβ isoform signaling via the SMAD pathway in fibroblasts cultured inside porous collagen scaffolds. This pilot study is of significant biological relevance because TGFβ isoforms are known to be involved in several processes (particularly wound contraction) that affect wound healing, and because the porous collagen-GAG scaffolds used in this study are similar to biomaterial grafts used clinically to induce skin regeneration [14]. Preliminary results based on simple statistical analysis reveal that knocking down different rSMAD components affect how fibroblasts respond to TGFβ1 and TGFβ3 isoforms. Interestingly, while it is believed that the "canonical" SMAD2 and SMAD3 transmit TGFβ isoform signaling while SMAD1 and SMAD5 transmit BMP signaling, preliminary data show that SMAD1 and SMAD5 are involved in TGFβ isoform signaling and in particular in different ways for each isoform. These observations need to be further verified by some kind of proteomic analysis.

Image informatics can provide powerful tools for studying signal transduction in cells, however they suffer from several limitations. In particular, they quantify the response of the pathway indirectly (through cell morphology) and therefore the provide evidence and not solid proof (unlike some methods of genetics and biochemistry). They also require careful fluorescent staining (in order to get consistent events) and challenging image processing. Finally, the sensitivity of immunofluorescence detection is not great, limiting the use of phosphorylation-specific antibodies that can provide direct quantification of downstream network activation. These weaknesses can be compensated by combining image informatics with proteomic approaches that provide large number of sensitive measurements of protein activation at various nodes of the signaling network [6]. In this case, image informatics can contribute data of high spatial and

temporal resolution, *in situ* measurements inside a physiologically relevant 3D matrix, and structural information. Nevertheless, there is still a long way to go in order to integrate the experimental and analytical tools of imaging-based and proteomic-based signal transduction analysis.

## ACKNOWLEDGEGMENTS

## REFERENCES

[1]  Gomperts, B. D., Kramer, I. M., and Tatham, P. E. R., [Signal Transduction], Academic Press (2009).

[2]  Scott, J. D., and Pawson, T., "Cell signaling in space and time: where proteins come together and when they're apart," Science 326(5957), 1220-1224 (2009).

[3]  Gough, N. R., and Yaffe, M. B., "Focus issue: conquering the data mountain," Sci. Signal. 4(160), eg2 (2011).

[4]  Yaffe, M. B., "Signaling networks and mathematics," Sci. Signal. 1(43), eg7 (2008).

[5]  Zhang, G., and Neubert, T. A., "Use of stable isotope labeling by amino acids in cell culture (SILAC) for phosphotyrosine protein identification and quantitation," Methods Mol. Biol. 527, 79-92 (2009).

[6]  Mitsos, A., Melas, I. N., Siminelakis, P., Chairakaki, A. D., Saez-Rodriguez, J., and Alexopoulos, L. G., "Identifying drug effects via pathway alterations using an integer linear programming optimization formulation on phosphoproteomic data," PLoS Comput Biol. 5(12), e1000591 (2009).

[7]  Zhang, S., Binari, R., Zhou, R., and Perrimon, N. A., "Genomewide RNA interference screen for modifiers of aggregates formation by mutant Huntingtin in Drosophila," Genetics 184(4), 1165-1179 (2010).

[8]  Perlman, Z. E., Slack, M. D., Feng, Y., Mitchison, T. J., Wu, L. F., and Altschuler, S. J., "Multidimensional drug profiling by automated microscopy," Science 306(5699), 1194-1198 (2004).

[9]  Bakal, C., Aach, J., Church, G., and Perrimon, N., "Quantitative Morphological Signatures Define Local Signaling Networks Regulating Cell Morphology," Science 316(5832), 1753-1756 (2007).

[10] Wollman, R., and Stuurman, N., "High throughput microscopy: from raw images to discoveries," J Cell Sci. 120(21), 3715-3722 (2007).

[11] Bakal, C., Linding, R., Llense, F., Heffern, E., Martin-Blanco, E., Pawson, T., and Perrimon, N., "Phosphorylation networks regulating JNK activity in diverse genetic backgrounds", Science 322(5900), 453-456 (2008).

[12] Nir, O., Bakal, C., Perrimon, N., and Berger, B., "Inference of RhoGAP/GTPase regulation using single-cell morphological data from a combinatorial RNAi screen," Genome Res. 20(3), 372-380 (2010).

[13] Griffith, L. G., and Swartz, M. A., "Capturing complex 3D tissue physiology in vitro," Nat Rev Mol Cell Biol. 7(3), 211-224 (2006).

[14] Yannas, I. V., [Tissue and organ regeneration in adults], Springer, (2001).

[15] Desmoulière, A., Chaponnier, C., and Gabbiani, G., "Tissue repair, contraction, and the myofibroblast," Wound Repair Regen. 13(1), 7-12 (2005).

[16] Hinz, B., Phan, S. H., Thannickal, V. J., Galli, A., Bochaton-Piallat, M. L., and Gabbiani, G., "The myofibroblast: one function, multiple origins," Am J Pathol. 170(6), 1807-1816 (2007).

[17] Massagué, J., "How cells read TGF-beta signals," Nat Rev Mol Cell Biol. 1(3), 169-178 (2000).

[18] Buehler, C., Kim, K. H., Greuter, U., Schlumpf, N., and So, P. T., "Single-photon counting multicolor multiphoton fluorescence microscope," J Fluoresc. 15(1), 41-51 (2005).

[19] Cheng, J., and Rajapakse, J. C., "Segmentation of Clustered Nuclei With Shape Markers and Marking Function," IEEE Trans. Biomedical Engineering 56(3), 741-748 (2009).

[20] Jones, T. R., Carpenter, A. E., and Golland, P., "Voronoi-based segmentation of cells on image manifolds," ICCV Workshop Comput. Vision Biomed. Image Appl., 535-543 (2005).