3D CELL NUCLEAR MORPHOLOGY: MICROSCOPY IMAGING DATASET AND VOXEL-BASED MORPHOMETRY CLASSIFICATION RESULTS

Alexandr A. Kalinin[1,2], *Ari Allyn-Feuer*[1], *Alex Ade*[1], *Gordon-Victor Fon*[1], *Walter Meixner*[1], *David Dilworth*[1], *Jeffrey R. de Wet*[1], *Gerald A. Higgins*[1], *Gen Zheng*[3], *Amy Creekmore*[3], *John W. Wiley*[3], *James E. Verdone*[4], *Robert W. Veltri*[4], *Kenneth J. Pienta*[4], *Donald S. Coffey*[4], *Brian D. Athey*[1,5], *Ivo D. Dinov*[1,2,4]

[1]University of Michigan Medical School, Department of Computational Medicine and Bioinformatics, Ann Arbor, MI [2]University of Michigan School of Nursing, Statistics Online Computational Resource (SOCR), Ann Arbor, MI [3]University of Michigan Medical School, Department of Internal Medicine, Ann Arbor, MI [4]University of Michigan, Michigan Institute for Data Science (MIDAS), Ann Arbor, MI [5]Johns Hopkins School of Medicine, James Buchanan Brady Urological Institute, Baltimore, MD

ABSTRACT

Cell deformation is regulated by complex underlying biological mechanisms associated with spatial and temporal morphological changes in the nucleus. Quantitative analysis of changes in size and shape of nuclear structures in 3D microscopic images is important not only for investigating nuclear organization, but also for detecting and treating pathological conditions such as cancer. Multiple methods have been proposed to classify cell and nuclear morphological phenotypes in 3D, however, there is a lack of publicly available 3D data for the evaluation and comparison of such algorithms. To address this problem, we present a dataset containing a of total of 1,433 segmented nuclear and 3,282 nucleolar binary masks. We also provide a baseline evaluation of a number of popular classification algorithms using voxel-based morphometric measures. Original and derived imaging data are made publicly available for downloading on the project web-page: `http://www.socr.umich.edu/projects/3d-cell-morphometry/data.html`.

***Index Terms*—** Cell nuclear morphology, nucleus shape classification

1. INTRODUCTION

Morphology of a cell nucleus and its compartments is regulated by complex biological mechanisms related to cell differentiation, development, proliferation, and disease [1]. Changes in the nuclear form are associated with reorganization of chromatin architecture and related to altered functional properties such as gene regulation. Conversely, external geometric constraints and mechanical forces that deform the cell nucleus affect chromatin dynamics and gene and pathway activation [2]. Thus, nuclear morphological quantification becomes of major relevance as the studies of the reorganization of the chromatin and DNA architecture in the spatial and temporal framework, known as the 4D nucleome, emerge [3]. Quantitative analyses of nuclear and nucleolar morphological changes also have medical implications, for example, in detection and treatment of pathological conditions such as cancer [4, 5].

Many algorithms have been proposed to classify cell and nuclear morphological phenotypes using 3D representations [6] that are often more informative as nuclear shape descriptors than 2D features [7, 8]. However, there is a lack of publicly available 3D cell image datasets that could serve for the comparison of various tools for the 3D cell nuclear morphometry. This limitation becomes of great importance in the modern reality of big data microscopy, when the ability to evaluate different approaches on benchmark data is needed for better dissemination of the current state of the art methods for bioimage analysis [9].

In order to objectively evaluate methods for 3D cell nuclear morphometry, we built and made publicly available a dataset of segmented and curated nuclear and nucleolar masks from fluorescence microscopy images. We then extracted common voxel-based measures of mask morphology in 2D and 3D, and evaluated a number of classification algorithms to provide performance baselines.

2. DATASET PREPARATION

The dataset is composed of two different cell collections.

The first collection is of Fibroblast cells (newborn male) that were purchased from ATCC (BJ Fibroblasts CRL-2522 normal) and subjected to a G0/G1 Serum Starvation Protocol [10], which has been shown to affect nuclear morphology [11]. This protocol provided us with images of the following conditions or phenotypes: cell cycle synchronized by serum-starvation (SS) and proliferating (PROLIF).

The second collection contains human prostate cancer cells (PC3). Through the course of progression to metastasis, malignant cancer cells undergo a series of reversible transitions between intermediate phenotypic states bounded by pure epithelium and pure mesenchyme [5]. These transitions in prostate cancer are associated with quantifiable changes in nuclear and nucleolar structure [12, 13]. Microscope slides of prostate cancer cell line PC3 were cultured in epithelial (EPI) and mesenchymal transition (EMT) phenotypic states.

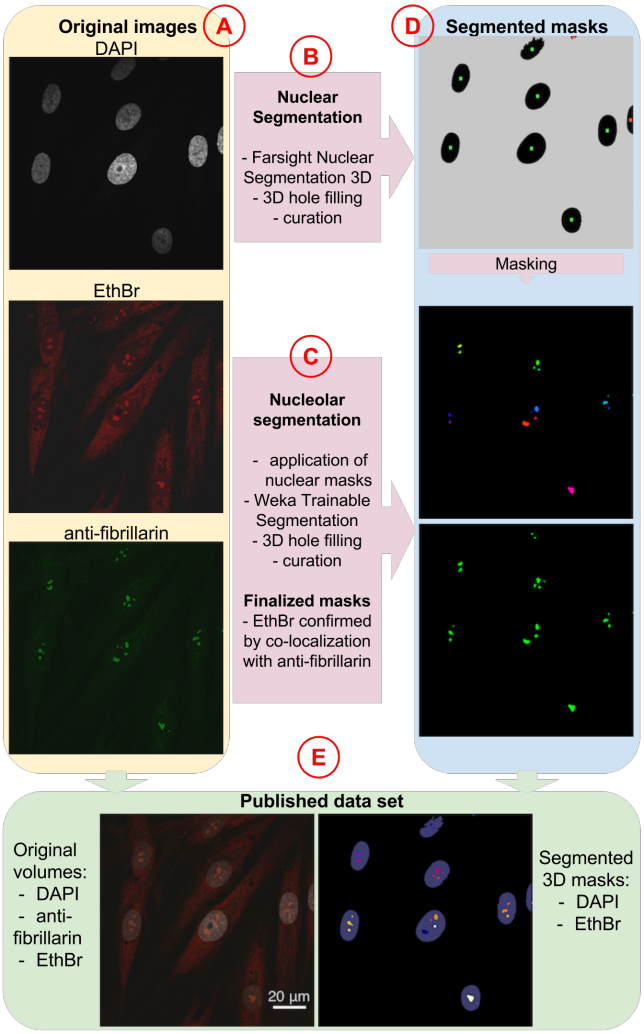

Fig. 1. A schematic view of the dataset preparation protocol and exemplar 2D slices of image volumes: (A) original Fibroblast data in three channels, see section 2.1; (B and C) main steps for nuclear and nucleolar segmentation, see section 2.2; (D) segmented images: nuclear masks after hole-filling (color-coded by quality control filter), segmented nucleolar masks from EtBr channel (colored by connected component labeling) and from anti-fibrillarin channel; (E) composite images of original and segmented data, see section 2.3.

2.1. Image acquisition

Cells in both collections were labeled with 3 different fluorophores: DAPI (4',6-diamidino-2-phenylindole), a common stain for the nuclei, fibrillarin antibody (anti-fibrillarin) and ethidium bromide (EtBr), both used for nucleoli staining. Although anti-fibrillarin is a commonly used nucleolar label, we found it to be too specific, which made extraction of a shape mask problematic. It has been shown that EtBr can be used for staining dense chromatin, nucleoli, and ribosomes [14]. We found that it provides better overall representation of nucleolar shape. Anti-fibrillarin was combined with EtBr by co-localization to confirm correct detection of nucleoli locations as described below. 3D imaging used a Zeiss LSM 710 laser scanning confocal microscope with a 63x PLAN/Apochromat 1.4NA DIC objective.

For multichannel vendor data, the channels were separated and saved as individual volumes labeled as c0, c1, c2, representing the DAPI, anti-fibrillarin, and EtBr channels, respectively, **Fig. 1A**. Each channel-specific volume was then re-sliced into a $1,024 \times 1,024 \times Z$ lattice ($Z = \{35, 50\}$), where regional sub-volumes facilitated the alignment with the native tile size of the microscope. All sub-volumes were saved as multi-image 3D TIFF volumes. For every sub-volume, accompanying vendor meta-data was extracted from the original data.

2.2. Segmentation

We performed the automatic 3D segmentation of nuclei using the Farsight toolkits Nuclear Segmentation [15], **Fig. 1B**. This tool was created specifically to segment DAPI-stained nuclei in 2D/3D and it does not require a labeled training set, and demonstrated stable results on these data. After segmentation, each segmented nucleus was represented as a mask, **Fig. 1D**. Post-segmentation processing of nuclear masks included 3D hole filling and a filtering step that removed the objects if they span the edge of a tile, are connected to other objects, or their voxel count was lower or higher than the empirically estimated threshold value.

Since the Farsight toolkits Nuclear Segmentation is not well suited for segmentation of cellular components other than nuclei, the classification of objects within the nucleus was performed using the Trainable Weka Segmentation [16]. Nuclear masks were used to isolate sub-nuclear segmentations in the EtBr and anti-fibrillarin channels to objects within a nucleus. An individual classifier model was created for each channel by using a random selection of 10% of the sub-volumes within that channel for training **Fig. 1C**. Trained models were then applied to all sub-volumes and nucleolar masks were created from the resulting probability maps and labeled as connected components. The nucleolar mask quality control protocol was similar to that of the nuclear masks with the additional filter for the spherical compactness of identified objects [17]. Finally, both EtBr and anti-fibrillarin

segmented volumes were used as input to a co-localization algorithm to validate the segmented EtBr-stained nucleoli based on the presence of anti-fibrillarin, **Fig. 1D**.

2.3. Dataset structure and archiving

Each collection in the final dataset consisted of sets of 3D segmented sub-volumes in the TIFF format from channel c0 (DAPI), representing binary nuclear masks, accompanied by a set of 3D binary masks of nucleoli from channel c2 (EthBr) per nucleus. To expand the potential use of the dataset, we also included original unsegmented $1024 \times 1024 \times Z$ TIFF sub-volumes in all 3 channels with corresponding meta-data, **Fig. 1E**. Both collections were grouped and archived per original volume for easier downloading, and made publicly available on the project web-page: `http://www.socr.umich.edu/projects/3d-cell-morphometry/data.html`. Each archive also contains a README file with the text description of the file structure.

3. MORPHOMETRIC CLASSIFICATION

To establish baseline morphometry classification results, we extracted multiple voxel-based morphometric characteristics from 3D binary masks and their 2D maximum intensity projection (2D masks). We used features to evaluate the performance of a number of widely used classification algorithms. We also assessed possible batch effects in data by comparing two different cross-validation techniques.

3.1. Voxel-based morphometry

We used image processing library, scikit-image [18], to extract sets of 2D and 3D voxel-based morphometric features from both nuclear and nucleolar binary masks.

The set of 3D morphometry features included: object volume, volume of the 3D bounding box, volume of the filled region, diameter of a sphere with the same volume as the object, and ratio of the object volume to the bounding box volume.

The 2D feature set included 2D analogs of all 3D measures and was supplemented by: convex hull area, eccentricity, Euler number, filled region area, eigenvalues of the inertia tensor of the region, major and minor axis of an ellipse fitted to the region, the angle between the X-axis and the major axis of the fitted ellipse, perimeter of an object which approximates the contour of the region, the ratio of the region area to the convex hull area.

In oder to aggregate the nucleolar features per nucleus we computed median, minimum, maximum, and standard deviation for each morphometry measure across the nucleoli within the nucleus. Correspondingly, nuclei that did not have any internally positioned nucleoli were excluded from further analysis. The number of detected nucleoli per nucleus was included as an individual feature.

3.2. Classification

We compared various supervised classification algorithms from scikit-learn, a popular Python machine learning toolkit [19], including Gaussian Naive Bayes (NB), Linear Discriminant Analysis (LDA), k nearest neighbors classifier (kNN), support vector machines with linear (SVM) and Gaussian kernels (RBF), Random Forest (RF), Extremely Randomized Trees (ET), and Gradient Boosting (GBM). All classifiers used default hyper-parameters. Feature preprocessing included feature standardization by removing the mean and scaling to unit variance of the training set.

To evaluate the possible batch effect that could occur during the image acquisition, we compared traditional k-fold cross-validation (CV) scheme with the suggested Leave-2-Opposite-Groups-Out (L2OGO) scheme. L2OGO ensures that: (1) all masks derived from one image fall either in the training or testing set, and (2) testing set always contains masks from 2 images of different classes. Unlike Leave-One-Group-Out CV, L2OGO enables per-split evaluation of performance metrics such as the Area under the Precision-Recall curve (AUPR) and the Area Under the Receiver Operating Characteristic curve (AUROC). Since original volumes were of different size and contained different number of nuclei, we joined smaller volumes into common groups to address class imbalance in testing sets and reduce the variance of the performance metric estimates. Given the class imbalance in L2OGO, we used AUC, AUPR, and F1 score to compare algorithms [20].

4. RESULTS

4.1. Fibroblast cells classification

After segmentation and exclusion of nuclei without detected nucleoli, the full collection of fibroblasts consisted of total 965 nuclei (498 SS and 470 PROLIF) and 2,181 nucleoli (1,151 SS and 1,030 PROLIF) extracted from the total of 11 volumetric images (7 SS and 4 PROLIF). 2D and 3D morphometric measures of nuclear and nucleolar masks were merged into per-nucleus feature vectors as described above.We evaluated the performance of algorithms for Fibroblast morphometric classification on 2 different CV schemes: 20 splits in L2OGO and a 7 times repeated 4-fold CV.

Results shown in **Fig. 2** do not demonstrate any batch effects in 2D classification. However, 3D performance of all classifiers using L2OGO was lower compared to 4-fold CV, which indicates the possibility of batch effects and overoptimistic classification results in 4-fold CV. As expected, L2OGO led to a large variance of metrics, especially in the F1 score, which can be a result of classifiers' sensitivity to the class imbalance. Using L2OGO, a number of algorithms showed higher performance on 3D morphometry with the best result by a Gaussian SVM (RBF) classifier with

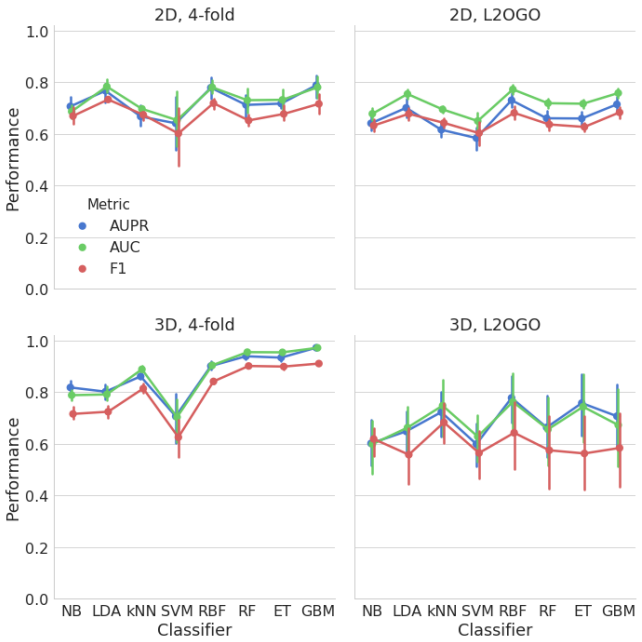

Fig. 2. Comparison of cross-validation strategies to evaluate batch effects in Fibroblasts using morphometry of 3D nuclear and nucleolar masks.

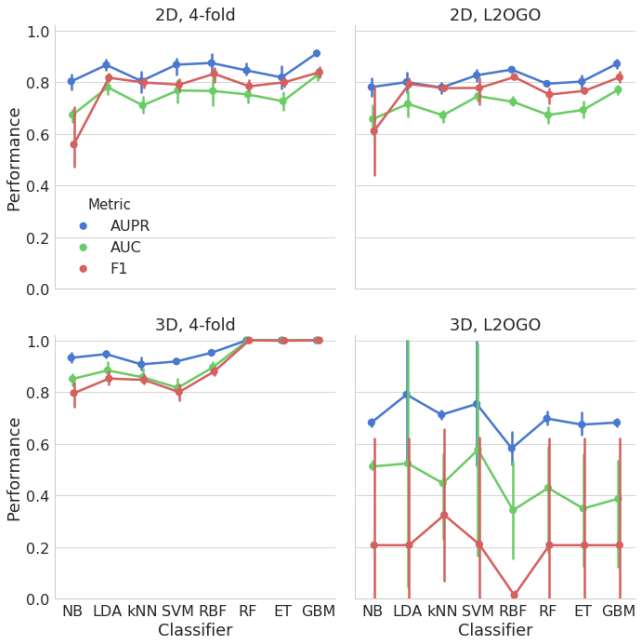

Fig. 3. Comparison of cross-validation strategies to evaluate batch effects on Fibroblasts 2D nuclear masks.

median $AUC = 0.814 \pm 0.245$, $AUPR = 0.724 \pm 0.206$, $F1 = 0.709 \pm 0.185$).

4.2. PC3 cells classification

After exclusion of nuclei without detected nucleoli, the segmented PC3 collection consisted of 458 nuclear (310 EPI and 148 EMT) and 1,101 nucleolar (649 EPI and 452 EMT) masks extracted from the total of 6 volumetric images (2 EPI and 4 EMT). After merging smaller EMT groups, L2OGO scheme produced 4 pairs of groups as training and testing sets. We compared L2OGO to 4-fold CV repeated 2 times.

Similar to the previous experiment, 2D morphometry classification performance was quite similar for both CV schemes, see **Fig 3**. However, in 3D, the performance of algorithms degraded remarkably when using L2OGO CV such that no methods performed better than in 2D. In this case, the best classification by single classifier was the result of applying a Gradient Boosting classifier (GB) with median $AUC = 0.774 \pm 0.017$, $AUPR = 0.875 \pm 0.019$, $F1 = 0.818 \pm 0.018$.

Results of classification on both collections suggest that the combination of the voxel-based morphometry and common algorithms with default parameters can provide a good baseline performance, especially in 2D. Using 3D masks can improve the performance as it did in Fibroblast classification. However, it suggests that having the third dimension sometimes can intensify batch effects and, thus, require more complex validation schemes.

5. DISCUSSION

3D cell microscopy is a powerful technique that enables investigation of biological mechanisms related to morphological changes in cell nucleus through quantitative analysis of changes in its size and shape. However, a lack of publicly available 3D cell image datasets limits the comparison of various 3D cell nuclear morphology analysis solutions. To address this limitation, we presented a new dataset that consists of two collections of 3D segmented binary masks from 2 different cell types and contains a of total of 1,433 segmented nuclear and 3,282 nucleolar 3D binary masks that can be used for evaluation of cell nuclear and nucleolar morphological classification methods. To account for batch effects, while enabling calculations of AUROC and AUPR performance metrics, we proposed a specific cross-validation scheme (L2OGO). We compared a number of commonly used machine learning classification algorithms on both collections of data using voxel-based morphometric measures extracted from original 3D binary masks as well as their 2D maximum intensity projections. Classification results provide a baseline that can be used for future comparison of morphological classifiers. As a limitation of this work, the microscope settings did not meet the Nyquist sample rates and may have created distortions in the digitized images [21].

Imaging protocols, original and segmented data, and the source code are made publicly available on the project webpage: `http://www.socr.umich.edu/projects/3d-cell-morphometry/data.html`

6. REFERENCES

[1] Predrag Jevtić, Lisa J Edens, Lidija D Vuković, and Daniel L Levy, "Sizing and shaping the nucleus: mechanisms and significance," *Current opinion in cell biology*, vol. 28, pp. 16–27, 2014.

[2] Caroline Uhler and GV Shivashankar, "Regulation of genome organization and gene expression by nuclear mechanotransduction," *Nature Reviews Molecular Cell Biology*, pp. nrm–2017, 2017.

[3] Gerald A Higgins, Ari Allyn-Feuer, Samuel Handelman, Wolfgang Sadee, and Brian D Athey, "The epigenome, 4d nucleome and next-generation neuropsychiatric pharmacogenomics," *Pharmacogenomics*, vol. 16, no. 14, pp. 1649–1669, 2015.

[4] Daniel Zink, Andrew H Fischer, and Jeffrey A Nickerson, "Nuclear structure in cancer cells.," *Nature reviews cancer*, vol. 4, no. 9, 2004.

[5] Robert W Veltri and Christhunesa S Christudass, "Nuclear morphometry, epigenetic changes, and clinical relevance in prostate cancer," in *Cancer Biology and the Nuclear Envelope*, pp. 77–99. Springer, 2014.

[6] Alexandre Cecilien Dufour, Tzu-Yu Liu, Christel Ducroz, Robin Tournemenne, Beryl Cummings, Roman Thibeaux, Nancy Guillen, Alfred O Hero, and Jean-Christophe Olivo-Marin, "Signal processing challenges in quantitative 3-d cell morphology: More than meets the eye," *IEEE Signal Processing Magazine*, vol. 32, no. 1, pp. 30–40, 2015.

[7] Hyun-Ju Choi and Heung-Kook Choi, "Grading of renal cell carcinoma by 3d morphological analysis of cell nuclei," *Computers in Biology and Medicine*, vol. 37, no. 9, pp. 1334–1341, 2007.

[8] Michael G Meyer, Mark Fauver, J Richard Rahn, Thomas Neumann, Florence W Patten, Eric J Seibel, and Alan C Nelson, "Automated cell analysis in 2d and 3d: A comparative study," *Pattern Recognition*, vol. 42, no. 1, pp. 141–146, 2009.

[9] Erik Meijering, Anne E Carpenter, Hanchuan Peng, Fred A Hamprecht, and Jean-Christophe Olivo-Marin, "Imagining the future of bioimage analysis," *Nature biotechnology*, vol. 34, no. 12, pp. 1250–1255, 2016.

[10] Thomas J Langan and Richard C Chou, "Synchronization of mammalian cell cultures by serum deprivation," *Cell Cycle Synchronization: Methods and Protocols*, pp. 75–83, 2011.

[11] Laura Seaman, Walter Meixner, John Snyder, and Indika Rajapakse, "Periodicity of nuclear morphology in human fibroblasts," *Nucleus*, vol. 6, no. 5, pp. 408–416, 2015.

[12] James E Verdone, Princy Parsana, Robert W Veltri, and Kenneth J Pienta, "Epithelial–mesenchymal transition in prostate cancer is associated with quantifiable changes in nuclear structure," *The Prostate*, vol. 75, no. 2, pp. 218–224, 2015.

[13] Lorenzo Montanaro, Davide Treré, and Massimo Derenzini, "Nucleolus, ribosomes, and cancer," *The American journal of pathology*, vol. 173, no. 2, pp. 301–310, 2008.

[14] MARCO Biggiogera and F Flach Biggiogera, "Ethidium bromide-and propidium iodide-pta staining of nucleic acids at the electron microscopic level.," *Journal of Histochemistry & Cytochemistry*, vol. 37, no. 7, pp. 1161–1166, 1989.

[15] Yousef Al-Kofahi, Wiem Lassoued, William Lee, and Badrinath Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 841–852, 2010.

[16] Ignacio Arganda-Carreras, Verena Kaynig, Curtis Rueden, Kevin W Eliceiri, Johannes Schindelin, Albert Cardona, and H Sebastian Seung, "Trainable weka segmentation: a machine learning tool for microscopy pixel classification," *Bioinformatics*, p. btx180, 2017.

[17] Raul S Montero and Ernesto Bribiesca, "State of the art of compactness and circularity measures," in *International mathematical forum*, 2009, vol. 4, pp. 1305–1335.

[18] Stefan Van der Walt, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu, "scikit-image: image processing in python," *PeerJ*, vol. 2, pp. e453, 2014.

[19] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al., "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[20] Takaya Saito and Marc Rehmsmeier, "The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets," *PloS one*, vol. 10, no. 3, pp. e0118432, 2015.

[21] Richard W Cole, Tushare Jinadasa, and Claire M Brown, "Measuring and interpreting point spread functions to determine confocal microscope resolution and ensure quality control," *Nat Protoc*, vol. 6, no. 12, pp. 1929–1941, 2011.