3D Shape Modeling for Cell Nuclear Morphological Analysis and Classification

Alexandr A. Kalinin[1,2*], Ari Allyn-Feuer[1*], Alex Ade[1], Gordon-Victor Fon[1], Walter Meixner[1], David Dilworth[1], Syed S. Husain[2], 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[1,2,5§]

[1]Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, USA

[2]Statistics Online Computational Resource (SOCR), Department of Health Behavior and Biological Sciences, University of Michigan School of Nursing, Ann Arbor, MI, USA

[3]Division of Gastroenterology, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI, USA

[4]Department of Urology, James Buchanan Brady Urological Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[5]Michigan Institute for Data Science (MIDAS), University of Michigan, Ann Arbor, MI, USA

[*]These authors contributed equally to this work

[§]Corresponding author

Abstract

Quantitative analysis of morphological changes in a cell nucleus is important for understanding of nuclear architecture and their relationship with pathological conditions such as cancer.

However, dimensionality of imaging data, together with a great variability of nuclear shapes present challenges for 3D morphological analysis. Thus, there is a compelling need for robust 3D nuclear morphometric techniques to carry out population-wise analysis. We propose a new approach that combines modeling, analysis, and interpretation of morphometric characteristics of cell nuclei and nucleoli in 3D. We use robust surface reconstruction that allows accurate approximation of 3D object boundary. Then, we compute geometric morphological measures characterizing the form of cell nuclei and nucleoli. Using these features, we compare over 450 nuclei with about 1,000 nucleoli of epithelial and mesenchymal prostate cancer cells, as well as 1,000 nuclei with over 2,000 nucleoli from serum-starved and proliferating fibroblast cells. Classification of sets of 9 and 15 cells achieves accuracy of 95.4% and 98%, respectively, for prostate cancer cells, and 95% and 98% for fibroblast cells. To our knowledge, this is the first attempt to combine these methods for 3D nuclear shape modeling and morphometry into a highly parallel pipeline workflow for morphometric analysis of thousands of nuclei and nucleoli in 3D.

Introduction

Motivation

Cell nuclear morphology is regulated by complex underlying biological mechanisms related to cell differentiation, development, proliferation, and disease [1-3]. Changes in the nuclear form are associated with reorganization of chromatin architecture related to altered functional properties such as gene regulation and expression [1,3]. Moreover, many studies in mechanobiology show that geometric constraints and mechanical forces applied to a cell deform it and, conversely, affect nuclear and chromatin dynamics and gene and pathway activation [4,5]. 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 [6,7]. Cellular structures of interest in the context of the 4D nucleome include not only the nucleus itself, but also the nucleolus and nucleolar-associating domains, chromosome territories, topologically associating domains, lamina-associating domains, and loop domains in transcription factories [6,8]. Furthermore, understanding of these processes through quantitative analysis of morphological changes also has many medical implications, for example, in detection, understanding, and treatment of pathological conditions such as cancer [7-10].

While efforts have been made to develop cell and nuclear shape characteristics in 2D or pseudo-3D [11,12], several studies have demonstrated that 3D morphometric measures provide better results for nuclear shape description and discrimination [13-15]. However, 3D shape descriptors that permit robust morphological analysis and facilitate human interpretation are still under active investigation [16]. Additionally, the dimensionality and volume of acquired data, various image acquisition conditions, and great variability of cell shapes in a population present challenges for 3D shape analysis methods that should be scalable, robust to noise, and specific enough across cell populations at the same time. Thus, there is a compelling need for robust 3D nuclear morphometric techniques to carry out population-wise analysis [17].

3D shape representation and morphometric measures

The way cell nuclear shapes can be measured depends on their representation extracted from image data [11]. Many 3D morphometric measures are applied "as is" to 3D geometric objects represented by volumetric data [18]. However, voxels-based shape representations are noisy, and

they may lose fine geometric details and even break object's topological structure. Moreover, these representations are not intrinsic, and vary when changing pose or deforming the object. A recent review of approaches to 3D cell shape description in [16] separated them into three categories in increasing order of complexity: landmark-based, graph-based, and moment-based. This last category includes approaches that are widely used in cellular morphology and allow the user to obtain a global representation that combines low-order moments describing the coarse conformation with high-order moments retaining information at higher frequency. Typically, before applying these methods a binary mask or outline of the shape (surface) is first extracted from image data, which is done by most segmentation methods. These masks are assumed to have a sphere-like topology and can be projected onto an appropriate basis. Two popular approaches of this type are spherical harmonics (SPHARM) [19] and spherical wavelets [20]. Both methods first map the surface of interest onto the sphere using appropriate spherical parameterization techniques and then project it onto a reference function basis living on the sphere. SPHARM is arguably one of the most widely applied cell morphology modeling approaches [21-24]. In SPHARM the spherical signal is projected onto a basis of Legendre polynomials, extending the classical Fourier analysis to signals on the two-sphere. SPHARM coefficients describe general conformation of the shape of interest at different spatial scales, are rotation invariant, and can be directly used as features for further analysis [25]. However, SPHARM methods are most appropriate when low order approximation is satisfactory and become less effective in preserving surface details as artificial oscillations start to appear when higher order basis functions are incorporated [26]. More robust smooth surface reconstruction can be obtained from a 3D binary mask via Laplace-Beltrami (LB) eigen-projection followed by topology-preserving boundary deformation to remove various artifacts [26]. On a unit sphere, the

LB eigen-functions correspond to spherical harmonics, so overall they can be viewed as a generalization of the SPHARM to the complex geometry manifold with local adaptation of the basis to the dataset at hand [27]. The method proposed in [26] has been demonstrated to produce smooth and more detailed surfaces compared to the SPHARM and the topology preserving level set [28]. Extracted surfaces are smooth, accurately represent the shape of an object, and can be further used for morphometric analysis.

In order to extract shape geometric characteristics, boundary surfaces of binary masks are typically reconstructed from voxel data and discretized as meshes. At the next step, various useful morphometric descriptors can be computed based on this representation. Useful extrinsic and intrinsic geometric descriptors aim to distinguish between global and local shape features. Intrinsic measures capture shape properties that are invariant under transformations (e.g., affine: rotation, translation and scaling). Various shape morphometry measures, like surface area and Gaussian curvature, represent invariant metrics of complexity, which are stable under special transformations of the surface (e.g., bending) that do not affect the inner geometry of the boundary of the 3D volume [29]. Alternatively, shape metrics, e.g., mean L^2-norm and the extrinsic curvature index are sensitive to affine transformation and other shape morphology in the ambient space. Shape index and curvedness are morphometric descriptors that can capture local shape features, independently or in relation to the size of an object [30]. Combination of the object surface reconstruction with the extraction of such shape measures demonstrated high performance in recent neuroimaging studies for discriminatory morphometric analysis of complex 3D shapes of cortical and subcortical brain areas [31-33].

Technical capabilities and interoperability of tools

When it comes to a choice of tools for 3D cell nuclear morphometrics, reproducibility and implementation availability are among major concerns in the field of bioimage analysis [16]. To date, many of the widely available software tools for cell shape morphometry were either developed for the analysis of 2D [11,34-38] or pseudo-3D images [39]. Other tools only implement slice-by-slice or voxel-based morphometry [40-43], providing a coarse approximation of the global cell shape that is sensitive to increasing amounts of noise and usually fails to characterize morphological variations occurring at different spatial scales. Other common limitations of many 3D cell morphology solutions include a lack of high-throughput processing capabilities and a restriction to the specific programming language or platform that dictate principles of a tool implementation [44-46]. Implementations of various methods in a bioimage analysis landscape are highly diverse. They range across programming languages, software libraries and file formats, which raises module interoperability issues and makes code reuse extremely difficult. Re-implementing underlying methods is often very challenging as well as being time-consuming and error prone [47]. Some of the existing bioimage analysis frameworks, including ImageJ [48], rely on a plugin architecture, which allows their extension via third-party contributions [40,41,43]. High-throughput capabilities of some of these tools are limited to processing of multiple objects simultaneously within its graphical user interface (GUI), for example, Tango [43]. More advanced packages, such as CellProfiler 2.0 [37], BioimageXD [42], and Icy [41] provide a basic graphical interface to assemble elementary tasks into reusable pipelines that are possible to execute in GUI and batch modes. However, these solutions are still limited to specific scripting languages and libraries supported by the main software package. They also don't provide a straightforward way

to take advantage of the growing number of parallel hardware configurations such as clusters, clouds, and high-performance computing, which limits the scalability of these solutions.

An alternative to plugin-based solutions, software platforms with modular design allow integration of already existing solutions into workflows without re-implementing them in a specific language and provide methods for optimizing module interaction, re-usage, and extension [49]. An example of an extensive and feature rich solution for building and executing complex workflows is the LONI Pipeline [31,50]. This client-server platform enables users to efficiently describe atomic modules and end-to-end protocols in a graphical canvas using a large library of powerful computational tools. The Pipeline back-end server has extensive support for parallel execution on a grid cluster, including automated data converting, formatting, and transfer, optimal job submission and management, pausing execution, combining local and remote software and data sources. Most importantly for this work, it makes it very easy to create new custom modules from any software that supports a command line interface (CLI). This allows one to take advantage of a highly diverse set of tools and connect them together as steps of a computational protocol that is then executed in a high-throughput, parallel fashion. Validated individual modules and end-to-end workflows may be saved, reused in other workflows, easily modified and repurposed. Additionally, the LONI Pipeline saves information about executed steps (such as software origin, version, and architecture) providing provenance information [50,51].

Study aims

This study has two complementary aims. The *first aim* is to assess and validate 3D morphometry metrics for nuclear and nucleolar shape description and classification. Improving the discriminative performance in terms of statistical metrics has been driving our methodological efforts and selection of specific tools in this work. First, surfaces of 3D masks extracted from the microscopy data are reconstructed using Laplace-Beltrami eigen-projection and topology-preserving boundary deformation [26]. Then we compute intrinsic and extrinsic geometric metrics, that are used as derived signature vectors (shape biomarkers) to characterize the complexity of the 3D shapes and discriminate between observed clinical and phenotypic traits. These metrics include volume, surface area, mean curvature, curvedness, shape index, and fractal dimension [30,52,53]. Although these methods were previously used in recent neuroimaging studies [31-33], this is the first attempt, to our knowledge, to apply robust smooth LB-based surface reconstruction with intrinsic and extrinsic morphometric measure extraction to 3D cell nuclear and nucleolar shape modeling and morphometry. Suggested modeling and analysis methods are not restricted to nuclear and nucleolar shapes and can be used for the shape quantification of other cellular compartments, depending on their topology.

The *second aim* is to develop a reproducible pipeline workflow implementing the entire process that can be customized and expanded for deep exploration of associations between 3D nuclear and nucleolar shape phenotypes in health and disease. High-throughput imaging (HTI) can include automatization of liquid handling, microscopy-based image acquisition, image processing, and statistical data analysis [17]. Our work focuses on last two aspects of this definition. We implemented a streamlined multi-step protocol using a diverse set of tools to achieve optimal performance compared to alternatives at each step of analysis. These tools are

represented as individual modules seamlessly connected in the LONI Pipeline workflow. This workflow meets modern standards for high-throughput imaging processing and analysis and is mostly automated with a focus on validity and reproducibility. Our implementation is massively parallel, customizable, and provides fully automated execution and data provenance out-of-the-box. At a final step of the workflow we employ machine learning methods to investigate the associations between cell phenotypes and treatment conditions using cell shape morphometric measures as features. We show that using a combination of 3D nuclear and nucleolar morphometry improves the discrimination between *in vitro* cell conditions of human fibroblast and human prostate cancer (PC3) cell lines.

To promote the reproducibility of results, facilitate open-scientific development, and enable collaborative validation we will make the pipeline workflows, together with underlying source code, documentation, and derived data from this study available online [54]. The workflow will be made available via the LONI Pipeline together with publicly available computational resources to showcase an online demonstration.

Methods

Fig. 1 shows a high-level view of the end-to-end protocol. We start with a dataset of 3D binary nuclear and nucleolar masks. We model 3D nuclear and nucleolar boundaries by their surface reconstruction and extracted derived morphometry measures. Finally, we compute statistical differences, identify shape morphometry-phenotype associations, and evaluate the results.

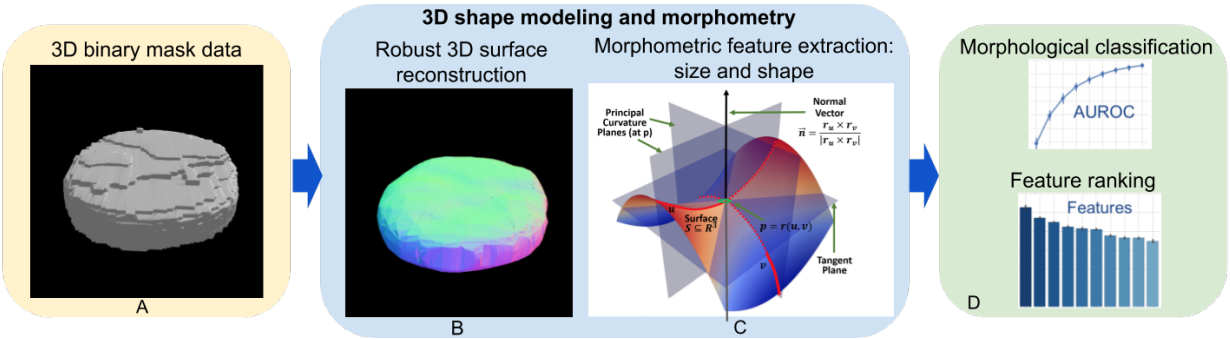

Figure 1. High-level schematic flow of the 3D image processing protocol. A: 3D binary mask data; B: mathematical representation and modeling of shape and size; C: calculation of derived intrinsic and extrinsic geometric measures; and D: machine learning based classification and analysis.

Dataset description

In this study we use 3D Cell Nuclear Morphology Microscopy Imaging Dataset, one of the biggest publicly available 3D cell imaging dataset to the date [18]. This dataset consists of two collections of 3D volumetric microscopic cell images with corresponding nuclear and nucleolar binary masks. Each collection includes images of cells in two phenotypic states, and thus this poses a binary classification problem with image-level labels that can be used for the assessment of cell nuclear and nucleolar morphometric analysis. Binary masks in each collection are obtained by segmentation of the original data. Nuclear masks are extracted from a DAPI (4',6-diamidino-2-phenylindole) channel, while fibrillarin antibody-stained (anti-fibrillarin) and ethidium bromide-stained (EtBr) channels are both used for nucleolar binary mask extraction (see [18] for details). Segmented binary masks are represented by 1024×1024×Z 3D TIFF sub-volumes. For every mask sub-volume, accompanying vendor metadata extracted from the original data are available for analysis as well.

Robust smooth surface reconstruction

To model the 3D shape of cell nuclei and nucleoli, boundaries of their 3D masks extracted from the microscopy data are modeled as genus zero two-dimensional manifolds (homeomorphic to a 2-sphere S^2) [55] that are embedded as triangulated surfaces in $\mathbb{R}^3$, Fig. 1B. Our approach uses iterative Laplace-Beltrami eigen-projection and topology-preserving boundary deformation algorithm [26]. This algorithm performs robust reconstruction of the objects' surfaces from their segmented masks using iterative mask filtering process. First, a mesh representation is constructed from the boundary of a binary mask of an object. Then, the boundary is projected onto the subspace of its Laplace–Beltrami eigen-functions [27], which allows the algorithm to automatically locate the position of spurious features by computing the metric distortion in eigen-projection. LB eigen-functions are intrinsically defined and can be easily computed from the boundary surface with no need of any parameterizations. They are also isometry invariant, and thus, are robust to the jagged nature of the boundary surface, which is desired for biomedical shape analysis [56]. In our prior experience [26], the discretized LB spectrum captures intrinsic shape characteristics (e.g., global shape transformations will preserve the spectral signature). The magnitude of the eigenvalues of the LB operator intuitively corresponds to the frequency in Fourier analysis, thus it provides a convenient mechanism to control the smoothness of the reconstructed surface. Using this information, the second step is a mask deformation process that only removes the spurious features while keeping the rest of the mask intact, thus preventing unintended volume shrinkage. This deformation is topology-preserving and well-composed such that the boundary surface of the mask is a manifold. The last two steps iterate until convergence and the method generates the final surface as the eigen-projection of the mask boundary, which

is a smooth surface with genus zero topology [26]. These properties allow application of this algorithm to any shape, including, for example, crescent-shaped, multi-lobed, and folded, as long as shape topology is homeomorphic to a sphere. The exemplar results of this step performed on nuclear and nucleolar masks are shown in Fig. 2.

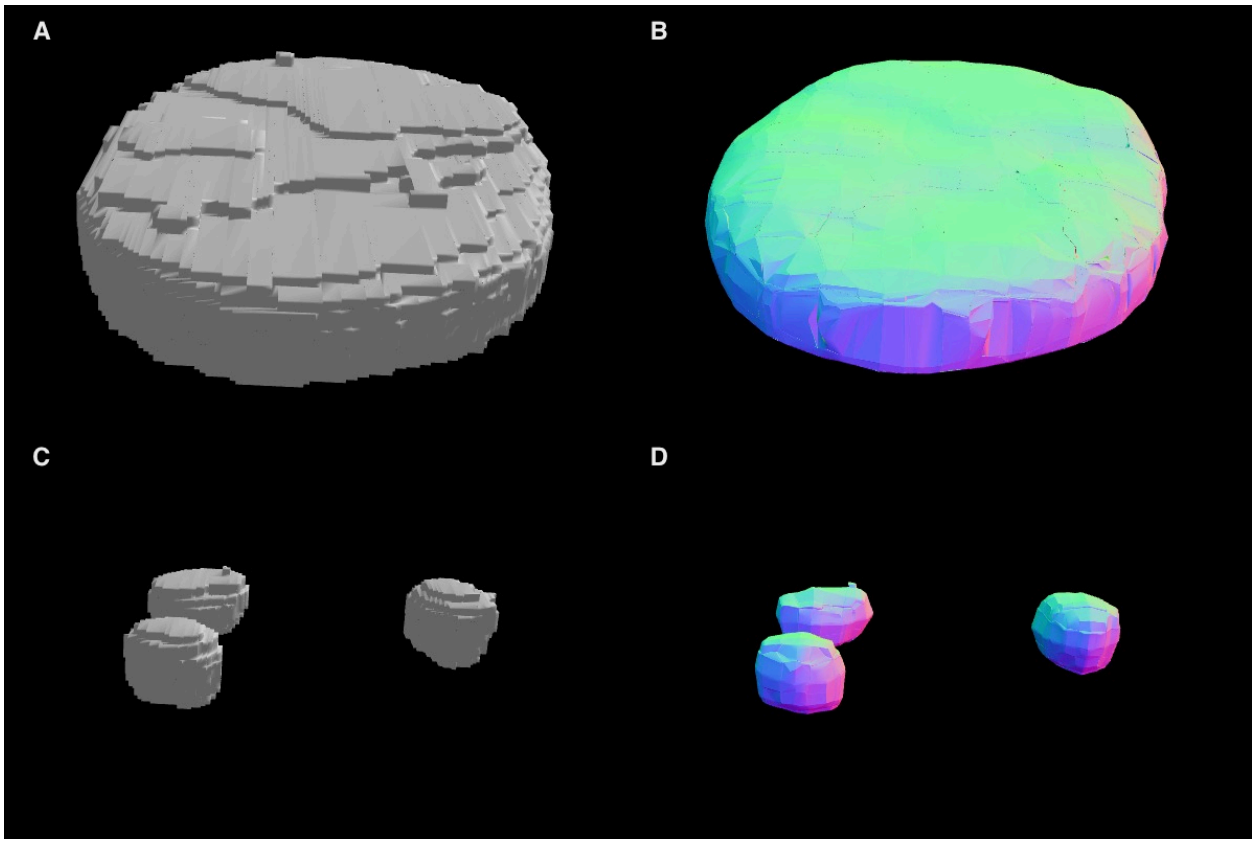

Figure 2. Robust smooth surface reconstruction. 3D visualization of: (A) a binary mask representation of a nucleus segmented from a Fibroblast cell image; (B) a mesh representation of a reconstructed smooth surface of a nucleus; (C) three binary masks for nucleoli segmented within this nucleus; and (D) three mesh representations of nucleolar surfaces, color-coded along the *Z* axis. Visualizations are produced with SOCR Dynamic Visualization Toolkit web application [57].

Morphometric measures

In this study, we use six shape measures as features quantifying geometric characteristics of the 3D surfaces, Fig. 1C. To calculate these measures, first the principal (min and max) curvatures ($\kappa_1 \leq \kappa_2$) were compute using triangulated surface models representing the boundaries of genus zero solids [58]. Then, shape morphometry measures can be expressed in terms of principal curvatures: mean curvature as $MC = \frac{\kappa_1+\kappa_2}{2}$, shape index as $SI = \frac{2}{\pi}\arctan\left(\frac{\kappa_1+\kappa_2}{\kappa_2-\kappa_1}\right)$, and curvedness as $CV = \sqrt{\frac{\kappa_1^2+\kappa_2^2}{2}}$. The principal curvatures of a surface are the eigenvalues of the Hessian matrix (second fundamental form), which solve for $k\ |H - kI| = 0$, where I is the identity matrix. If S is a surface with second fundamental form $H(X, Y)$, $p \in M$ is a fixed point, and we denote an orthonormal basis u, v of tangent vectors at p, then the principal curvatures are the eigenvalues of the symmetric Hessian matrix, $H = \begin{bmatrix} H_{u,u} & H_{u,v} \\ H_{v,u} & H_{v,v} \end{bmatrix} = H_{uu}\partial u^2 + 2H_{u,v}\partial u\partial v + H_{v,v}\partial v^2$, a.k.a. shape tensor. Let $r = r(u, v)$ be a parameterization of the surface $S \subseteq R^3$, representing a smooth vector valued function of two variables with partial derivatives with respect to u and v denoted by r_u and r_v, Fig. 3. Then, the Hessian coefficients $H_{i,j}$ at a given point (p) in the parametric u, v-plane are given by the projections of the second partial derivatives of r at that point onto the normal to S, $n = \frac{r_u \times r_v}{|r_u \times r_v|}$, and can be computed using the dot product operator: $H_{u,u} = r_{u,u} \cdot n$, $H_{u,v} = H_{v,u} = r_{u,v} \cdot n$, $H_{v,v} = r_{v,v} \cdot n$, Fig. 3.

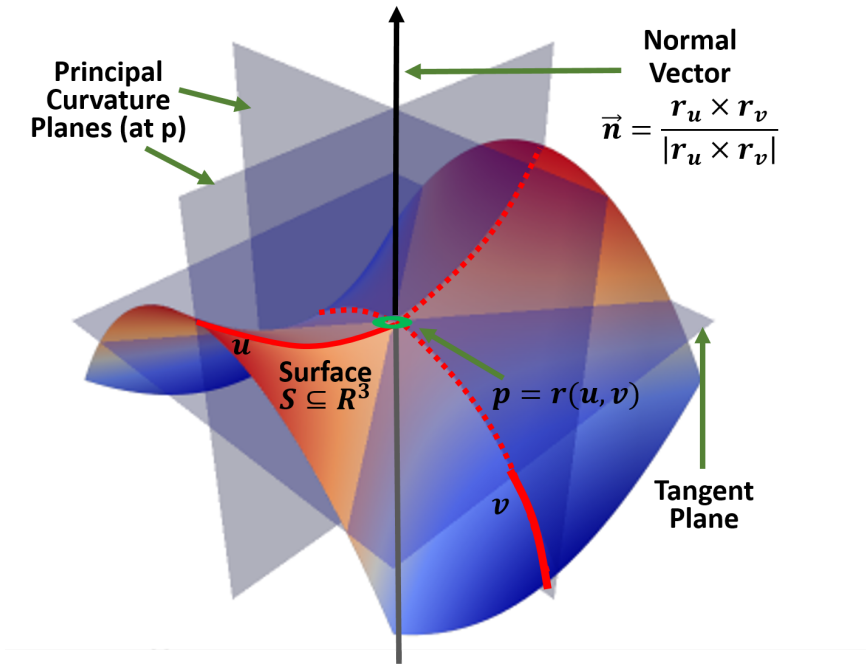

Figure 3. The (local) geometry of 2-manifolds. Per vertex definitions of curvature, relative to a local coordinate framework.

Volume is the amount of 3D space enclosed by a closed boundary surface and can be expressed as $V = \iiint_{\mathbb{R}^3} I_D(x, y, z)dxdydz$, where $I_D(x, y, z)$ represents the indicator function of the region of interest (D) [59]. If $r(u, v)$ is a continuously differentiable function and the normal vector to the surface over the appropriate region D in the parametric u, v plane is denoted by $\vec{r}_u \times \vec{r}_v$, then $S_\Omega : r = r(u, v),\ (u, v) \in \Omega$, is the parametric surface representation of the region boundary [60]. Then surface area can be expressed as $SA = \iint_\Omega |\vec{r}_u \times \vec{r}_v| dudv$. The fractal dimension calculations are based on the fractal scaling down ratio, ρ, and the number of replacement parts, N [61]. Accurate discrete approximations of these metrics are used to compute them on a mesh-represented surfaces as described in [53,62]. These discrete metrics were first introduced as a part of the shape analysis protocol [31] and were further applied in neuroimaging studies [32,33].

The extracted 3D morphometric measures serve as features for training a number of machine learning algorithms in order to assess classification performance, Fig. 1D. The number of detected nucleoli per nucleus is included as an individual feature. We merge nucleoli-level features within each nucleus by computing sample statistics (e.g., average, minimum, maximum, and higher moments) for each morphometry measure, similarly to [18]. These statistics are used to augment the signature feature vectors of the corresponding parent nuclei such that all feature vectors are of the same length. Correspondingly, nuclei that do not have any automatically detected internally positioned nucleoli were excluded from further analysis, such that for each nucleus there was at least one nucleolus.

Visual analytics and machine learning for morphometric analysis

We perform exploratory visual analysis of extracted features using SOCRAT [49], a web platform for interactive visual analytics. The goal of visual analytics is to support analytical reasoning and decision making with a combination of highly interactive visualizations and data analysis techniques [49,63]. SOCRAT implements a visual analytics workflow that encompasses an iterative process, in which data analysts can interactively interrogate extracted morphometric measures in the form of interactive dialogue supported by visualizations and data analysis components. In order to assess the variability of extracted morphometry data, we include t-Distributed Stochastic Neighbor Embedding (t-SNE) [64] visualizations of the feature space generated by SOCRAT [49]. We also use SOCRAT to demonstrate interactions between the top-3 important features according to the best-performing classification algorithm. All derived morphometric datasets are made available within SOCRAT Web Demo application [65]. Finally, we provide an ability to

visualize volumetric images and extracted meshes online via SOCR Dynamic Visualization Toolkit web application [57].

In general, correct classification of every single cell (type, stage, treatment, etc.) is a challenging task due to significant population heterogeneity of the observed cell phenotypes. For example, the same sample may contain a close mixture of intertwined "cancerous" and "non-cancerous" cells phenotypes or both classes may include apoptotic cells exhibiting similar shapes or sizes. Given the nature of cell samples, culturing, preparation and collection, we have considered classification of cell sets rather than single cells. The idea of classifying sets of cells, rather than individual samples, is not new and has been used in recent biomedical image classification studies [12,66]. The rationale behind this is based upon the observation that even if an algorithm misclassifies a few cells in a sample, the final (cell set) label will still be assigned correctly, as long as majority of cells are classified correctly. Using this strategy, we performed classification on small groups of cells, ranging from 3 to 19 cells per set. During each fold of the internal cross-validation, these small cell sets are randomized by bootstrapping procedure with 1,000 repetitions. Due to the possible presence of batch effects in data, we employ the Leave-2-Opposite-Groups-Out (L2OGO) cross-validation scheme [18]. 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. We use scikit-learn, a popular Python machine learning toolkit [67], to evaluate a number of supervised classification algorithms.

High-throughput workflow protocol

While the LONI Pipeline is a popular tool in neuroimaging and bioinformatics, it has been so far overlooked by the bioimage analysis community. In this work, we utilized the LONI Pipeline for the implementation of a streamlined multi-step protocol that relies on a diverse set of tools and solutions seamlessly connected in the LONI Pipeline workflow, Fig. 4. From a high-level perspective, every step of data processing and analysis protocol is wrapped as an individual module in the workflow that provides input and output specifications that allow the Pipeline to automatically connect and manage atomic modules. The modular structure of our implementation makes it highly flexible and not limited to specific tools included in the workflow. It can be repurposed for a wide range of different experiments by adjusting parameters, adding, removing, or replacing individual modules, while preserving high-throughput capabilities as presented in the Discussion section. Every module represents an independent component that can be used in a stand-alone fashion. As a result, a distributed, massively parallel implementation of our protocol makes it possible to easily process thousands of nuclei and nucleoli simultaneously. The workflow does not depend on the total number of 3D objects, biological conditions, or a number of running instances since its execution is completely automated once the workflow configuration is fixed, including job scheduling and resource allocation. During the execution, our workflow provides a researcher with real-time information about progress, allows the viewing of intermediate results at every individual step and failed modules may easily be restarted.

The workflow is configured in a way that it can consume data in the format that we use to share it, i.e. a $1024\times1024\times Z$ 3D volumes in different channels as 16-bit 3D TIFF files. Each volume is processed independently, in parallel fashion, such that workflow automatically defines how

many processes are needed to analyze all of the input data. 3D shape modeling and morphometric feature extraction are performed on individual masks independently, which allows us to simultaneously run up to 1,200 jobs on the cluster during our experiments, effectively reducing the computing time. Finally, the workflow collects morphometry information from each individual mask and combines them in the results table that is further used as an input to classification algorithm. These capabilities allow the user to take advantage of modern computational resources, lift the burden of low-level configuration from researchers, make it easier to control the execution process, and improve reproducibility of the whole process.

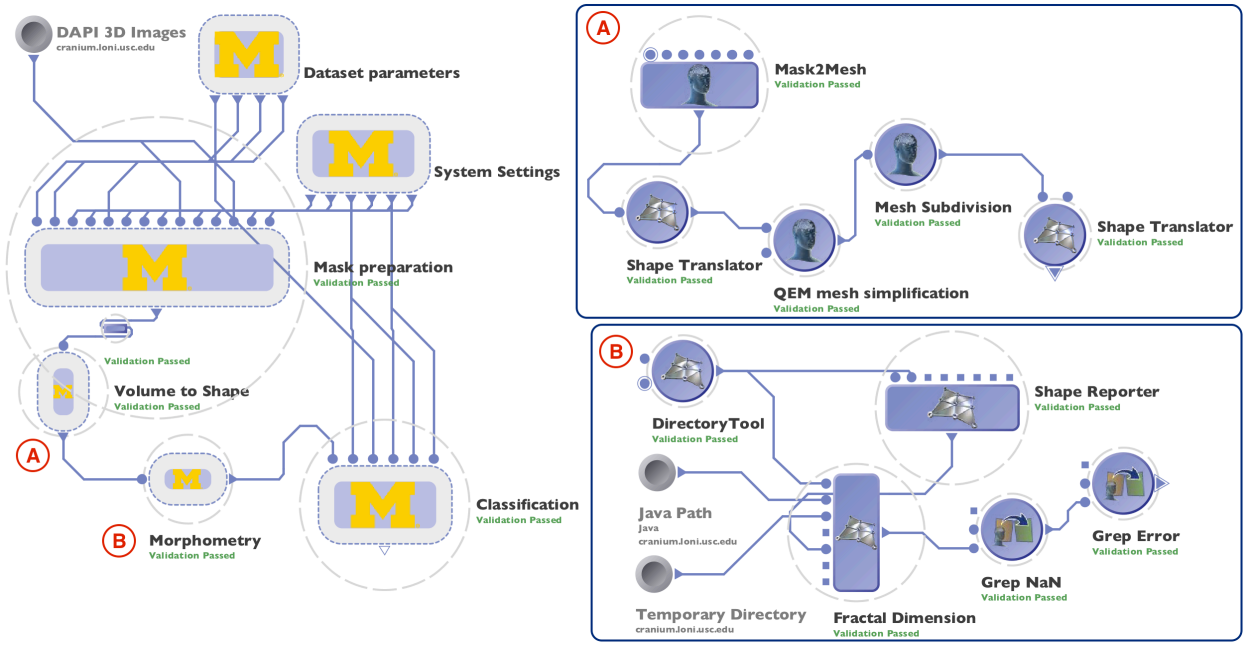

Figure 4. Screenshots of the exemplar graphical workflow in the LONI Pipeline client interface that includes: (left) overview of the validated workflow protocol showing nested groups of modules; (A) expanded Volume to Shape group that includes modules that perform 3D shape modeling refinement; and (B) expanded Morphometry group that includes a module that performs morphological measure extraction.

Results

Validation on synthetic data

To validate the shape morphometry metrics, we first apply them to synthetically generated 3D masks. We use the scikit-image Python library [68] to create 3D solids representing cubes, octahedra, spheres, ellipsoids, and 3 overlapping spheres with linearly aligned centers (Supplementary Fig. S1 online). We process these objects and compare the resulting shape morphometry measures. Specifically, we aim to confirm the expected close relation between the analytically derived measures of volume and surface area computed using the corresponding shape parameters (e.g., radius, size), and their computationally derived counterparts reported by the processing pipeline workflow. Our results illustrate that for nucleus-like shapes, e.g., spheres and ellipsoids, the computational error is within 2%. For faceted objects, e.g., cubes and octahedrons, the calculation error is within 6%. The increased error in the latter case can be explained by the mesh smoothing the surface reconstruction algorithm applies at the shape vertices to resolve points of singularity (e.g., smooth but non-differentiable surface boundaries). To demonstrate the detection of shape differences between different types of 3D objects, we also compare overlapping spheres against circumscribed ellipsoids. As expected, the average mean curvature and curvedness measures are lower and shape index values are higher for spheres compared to ellipsoids. We observe the similar trend when comparing changes in these shape morphometry measures for spheres, ellipsoids, and overlapping spheres. For example, average mean curvature and curvedness are highest for overlapping spheres and lowest for spheres, which is expected based on definitions of these measures (see Supplementary Table S2). This simulation confirms our ability to accurately measure size and shape characteristics of 3D

objects, which forms the basis for machine-learning based object classification based on boundary shapes. Exemplar results of synthetic data morphometry are available in Supplementary Table S1.

Comparison with SPHARM for fibroblast nuclei shape classification

To assess chosen shape morphometry metrics as discriminatory features we compare them and SPHARM coefficients [16,25] in classification of single cell nuclei from the fibroblast collection of the 3D Cell Nuclear Morphology Microscopy Imaging Dataset [18]. It includes images of primary human fibroblast cells that were subjected to a G0/G1 Serum Starvation Protocol that is used for cell cycle synchronization [69] and has previously been shown to alter nuclear organization, which was reflected in morphology changes, for example, nuclear size and shape [70]. As a result, this collection contains 962 3D nuclear binary masks in the following phenotypic classes: (1) proliferating fibroblasts (PROLIF), and (2) cell cycle synchronized by the serum-starvation protocol cells (SS). We use these binary nuclear masks to calculate both SPHARM and morphometric features.

To obtain SPHARM coefficients we use the popular SPHARM-MAT toolbox [71] that implements surface reconstruction and spherical parametrization using the CALD algorithm [25], followed by the expansion of the object surface into a complete set of spherical harmonic basis functions of degree $l = 13$ (default setting). Finally, SHREC method [72] is used to minimize the mean square distance between corresponding surface parts. SPHARM shape descriptors are computed as described in Ducroz *et al.* [25] and used as feature vectors for classification.

We employ the open-source Python package scikit-learn 0.17.0 [67] to test a number of commonly used machine learning classification methods on derived feature vectors with default parameters for each method. Performance is compared using the L2OGO cross-validation scheme and the area under the receiver operating characteristic curve (AUC) as a performance metric. As shown in Table 1, 3D shape morphometric measures not only demonstrate comparable discriminative performance to SPHARM coefficients, but outperform them using all tested algorithms.

Table 1. Comparison of SPHARM coefficients and our morphometry descriptors for single cell fibroblast nuclei classification

Classification algorithm	**SPHARM coefficients, mean AUC**	**Surface morphometry measures, mean AUC**
k-Nearest Neighbors	0.556±0.103	**0.629±0.204**
Linear SVM	0.593±0.165	**0.677±0.354**
Gaussian SVM	0.677±0.354	**0.682±0.264**
Random Forest	0.619±0.175	**0.645±0.200**
AdaBoost	0.612±0.246	**0.663±0.252**
Gradient Boosting	0.620±0.234	**0.674±0.229**

Fibroblast cell classification

The full collection of fibroblast masks for binary classification consists of total 965 nuclei (498 SS and 470 PROLIF) and 2,181 nucleoli (1,151 SS and 1,030 PROLIF). Figure 5A demonstrates the variability of the extracted morphometry measures in a t-SNE projection visualized in

SOCRAT. Although there is a small degree of grouping, there is no clear separation between classes.

Table 2. Fibroblast single cell and 9-cell sets classification accuracy

Measure	**Single cell, mean (± SD)**	**9 cells set, mean (± SD)**
Accuracy	0.754 (± 0.037)	0.951 (± 0.029)
Precision	0.769 (± 0.047)	0.968 (± 0.035)
Sensitivity	0.731 (± 0.055)	0.935 (± 0.049)
AUC	0.754 (± 0.037)	0.951 (± 0.029)

The best result by a single classifier is achieved using a stochastic gradient boosting classifier with 1,500 base learners, maximum tree depth 8, learning rate 0.01, subsampling rate 0.5, and minimum number of samples at a leaf node 3. Hyper-parameters are fine-tuned using a cross-validated grid search. To evaluate these classification results, we measured accuracy, precision, sensitivity and AUC over L2OGO cross-validation, which are presented in Table 2 for single cell and 9-cell-set classifications. Figure 5B shows mean AUC values for set sizes from 3 to 19 cells. A 95% accuracy is reached when classifying sets with 9 cells and 98% for sets with 15 or more cells.

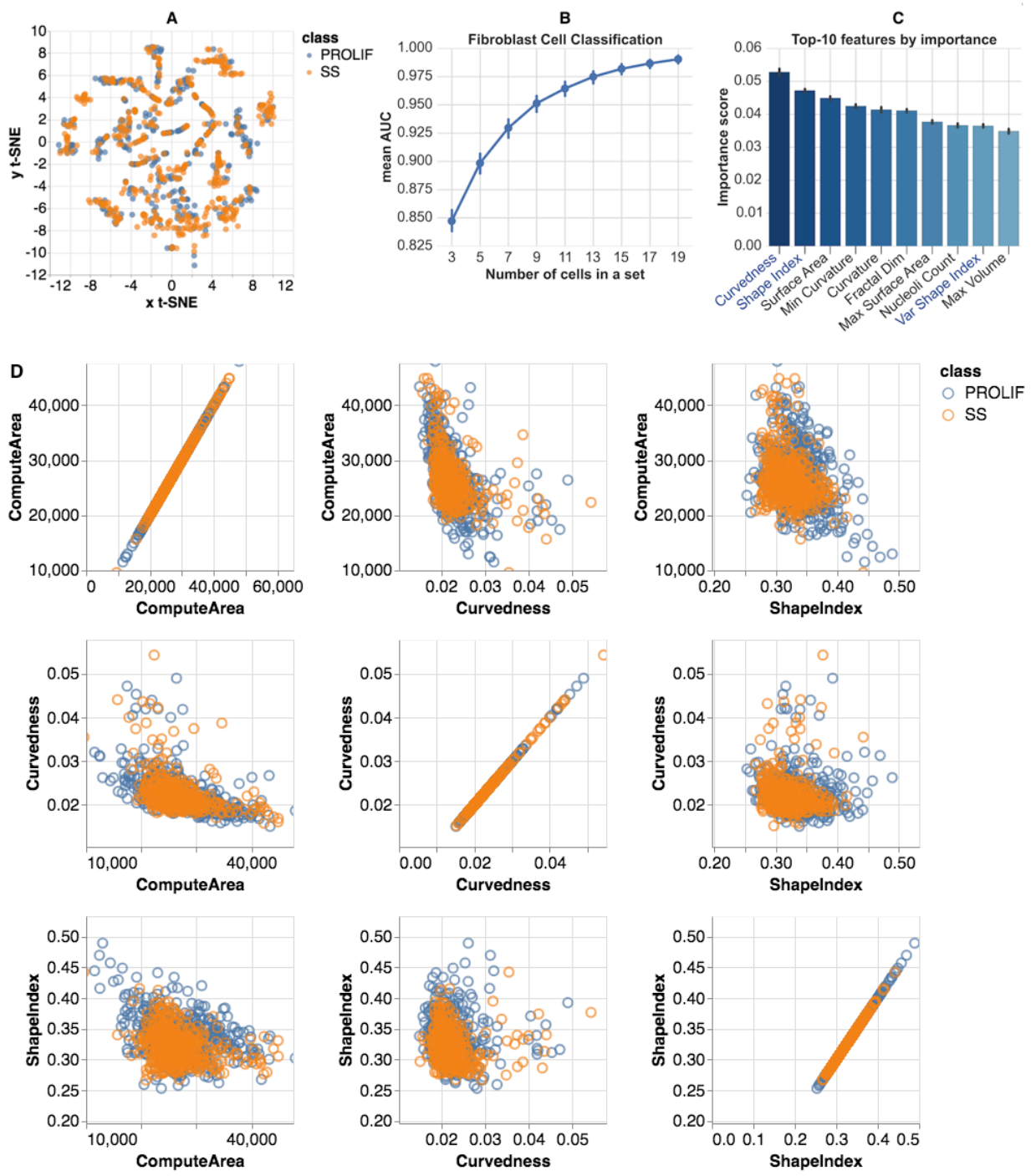

Figure 5. Fibroblast morphometric analysis: (A) SOCRAT visualization of t-SNE projection of morphometric feature space; (B) mean AUC for various cell set sizes; (C) top-10 features for classification by importance score (right, nucleolar feature names start with Avg, Min, Max or

Var, feature names that were also reported in top-10 for PC3 cells are shown in blue font); and (D): SOCRAT visualization of interactions between top-3 features.

The gradient boosting classifier also computes and reports cross-validated feature importance (Fig. 5C). These allow us to evaluate which measures differ between two cell conditions, and potentially propose novel research hypotheses that can be testing using prospective data. Previous analysis has reported quantifiable changes in both nuclear size and shape under serum-starvation[70]. In our results, both nuclear (top-3, 5 out of top-10) and nucleolar (4 of top-10) morphometric size and shape features are reported to be of high importance for distinguishing SS fibroblasts from PROLIF (Fig. 5C). We also visualize the relationship between top-3 features using SOCRAT, see Fig. 5D. Visualizations suggest the smaller variation of morphometric measures in SS fibroblast nuclei compared to their PROLIF counterparts. This result may provide insights in further downstream analysis of potential underlying mechanisms that lead to these morphometric changes. We have made the fibroblast morphometry data publicly available within SOCRAT for further analysis and validation [65].

PC3 EPI/EMT cell classification

The second collection contains images of 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[9]. These transitions in prostate cancer are associated with quantifiable changes in both nuclear and nucleolar structure[10,73]. PC3 cells were cultured in: (1) epithelial (EPI), and (2) mesenchymal

transition (EMT) phenotypic states. The collection includes 458 nuclear (310 EPI and 148 EMT) and 1,101 nucleolar (649 EPI and 452 EMT) 3D binary masks. Random uniform sub-sampling is used to resolve the large sample-size imbalance between the 2 classes. Figure 6A demonstrates the variability of the extracted morphometry measures in a t-SNE projection visualized in SOCRAT. Similar to fibroblasts, the projection of the PC3 morphometric feature space does not demonstrate clear separation between classes.

Table 3 PC3 single cell and 9-cell sets classification accuracy

Measure	**Single cell, mean (± SD)**	**9 cells set, mean (± SD)**
Accuracy	0.764 (± 0.059)	0.954 (± 0.059)
Precision	0.761 (± 0.080)	0.943 (± 0.085)
Sensitivity	0.787 (± 0.080)	0.978 (± 0.043)
AUC	0.764 (± 0.059)	0.954 (± 0.059)

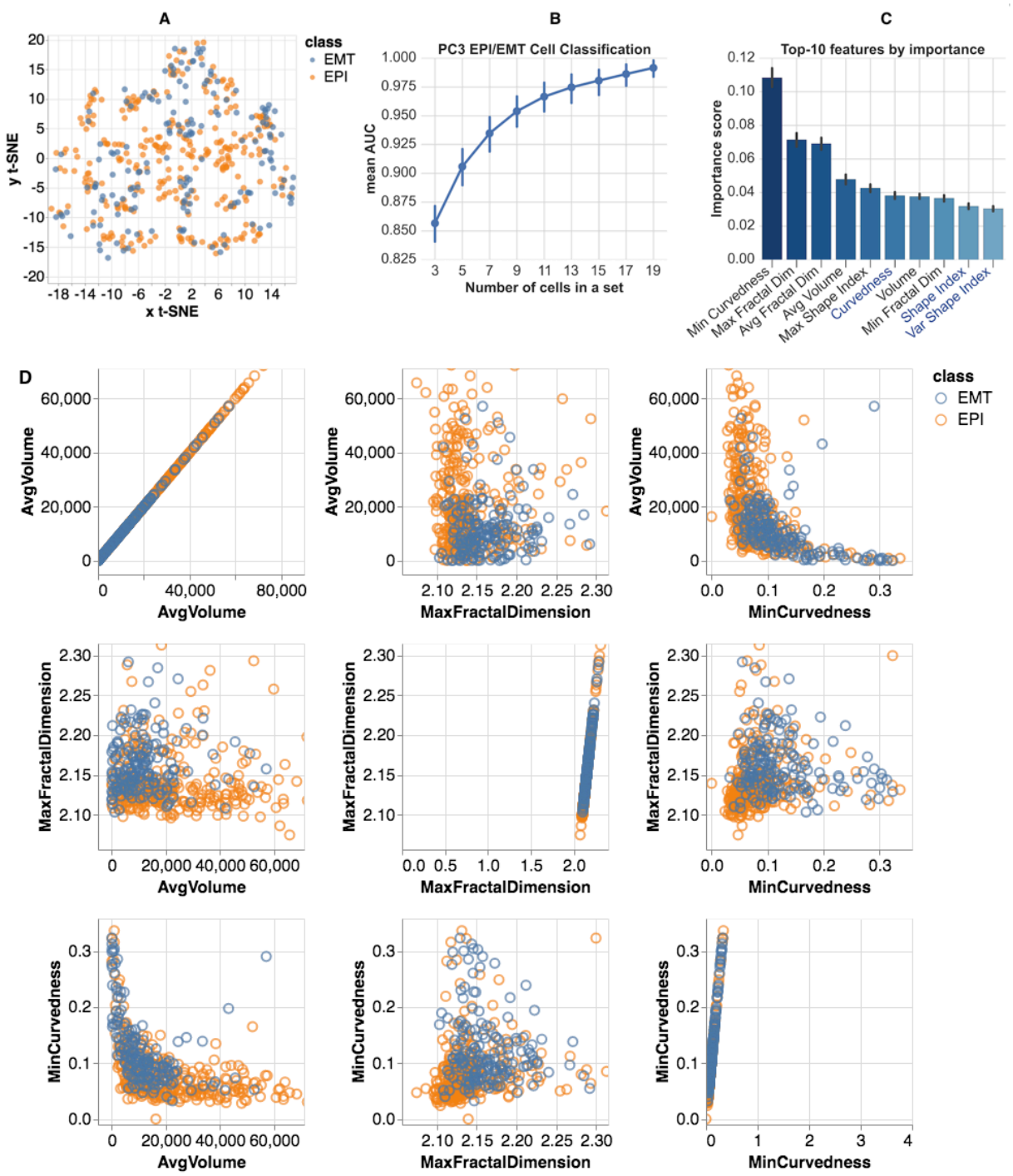

Figure 6. PC3 morphometric analysis: (A) SOCRAT visualization of t-SNE projection of morphometric feature space; (B) mean AUC for various cell set sizes; (C) top-10 features for classification by importance score (right, nucleolar feature names start with Avg, Min, Max or

Var, feature names that were also reported in top-10 for Fibroblast cells are shown in blue font); and (D): SOCRAT visualization of interactions between top features (1st, 2nd, and 4th).

In this case, the best classification by single classifier is the result of applying a random forest model (1,000 trees, maximum tree depth 12, maximum number of features for the best split 40%). Hyper-parameters fine-tuning, accuracy metrics, and cross-validation procedures are identical to the ones reported in the previous fibroblast experiment. Similar to the fibroblast cell classification, classification of sets of 9 cells achieves a mean accuracy of 95.4%, which increases to 98% for sets of 15 or more cells, Table 3. Figure 6B reports the AUC for different group sizes to show how the classification accuracy increases with the cell-set size and reaches 98% for sets of 13 cells. In this experiment, we also examine the classifier-reported feature importance, Fig. 6C. The top-10 important features in this classification included nuclear (3 of top-10, 2 of which are also Fibroblast top-2) and nucleolar (top-5) shape morphometry features. Top feature interactions visualized using SOCRAT demonstrate the important changes in distributions of nucleolar morphometric measures, Fig. 6D. For example, it seems that the EPI nucleoli tend to have a higher average volume, while also having more variability in minimal curvedness and maximum fractal dimension, compared to EMT nucleoli. Previously reported PC3 morphological analyses73 only used simple 2D nuclear form measures, such as diameter and the size of the bounding box. While we also confirm the importance of nuclear form in our results, we also suggest that further investigation of other highly ranked features, such as nucleolar curvedness, shape index and fractal dimension, may provide interesting mechanistic

insights. PC3 morphometry data are made publicly available within SOCRAT for further analysis and validation 65.

Discussion

In this study, we propose, implement and validate a solution for 3D modeling, morphological feature extraction, analysis and classification of cells by treatment conditions. Compared to other studies using 2D projections, this approach operates natively in 3D space and takes advantage of extrinsic and intrinsic morphometric measures that are more representative of the real, underlying nuclear and nucleolar geometry and allow easy human interpretation. Given the limitations of using 3D voxels for accurate shape representation, we employ 3D surface models to extract more informative size and shape measures to improve the morphology classification performance. Robust surface reconstruction allows accurate approximation of 3D object boundaries that was validated on synthetic data. Suggested shape morphometric measures outperforms another popular approach and demonstrated their universality across different cell types, conditions, and even domains [31-33].

Our computational protocol implementation is highly parallel with throughput limited only by the number of available computing nodes, and it can process thousands of objects simultaneously with minimal human intervention. This pipeline workflow integrates a number of open-source tools for different steps of data processing and analytics. Every module in our workflow represents an individual component that can be easily modified, removed, or replaced by an alternative. Such modular software platform architectures have been shown to enable high reusability and ease of modification [49]. This allows the user to use the same workflow or

customize and expand it (e.g., specification of new datasets, swapping of specific atomic modules) for other purposes that require the analysis of a diverse array of cellular, nuclear, or other studies. The live demo available via the LONI Pipeline demonstrates the simplicity of use and high efficiency of parallel data processing. LONI also provides guest access (see Supplementary Information) and an opportunity to utilize a 4,500-core LONI cluster after applying for a collaboration account.

We test our approach on the 3D Cell Nuclear Morphology Microscopy Imaging Dataset [18], which includes a total of ~1,500 nuclear and ~2700 nucleolar masks. The classification results on these data comparing epithelial vs. mesenchymal human prostate cancer cell lines and serum-starved vs. proliferating fibroblast cell lines demonstrate the high accuracy of cell type prediction using 3D morphometry, especially when applied to sets of cells. Although different classification algorithms appear to be optimal for different experiments, we observe that both nuclear and nucleolar morphometric measures are important features for discriminating between treatment conditions or cell phenotypes. In the case of fibroblast classification, the results show the importance of nuclear morphometry, the number of nucleoli per nucleus, and various internal nucleolar morphometric measures. These observations confirm and extend previously reported results. For PC3 cells, the most important classification features are the moments of the distributions of various nucleolar morphometric measures, along with nuclear size and shape. Interestingly, there were 3 common morphometric features among the top-10 most important ones for both cell lines. This confirms previously reported observations [73], suggests new important morphological characteristics, and demonstrates that our method extracts relevant information from cell forms to successfully classify cells using a combination of criteria. In

addition, this also demonstrates the importance of sophisticated shape metrics, compared to volume and surface area that alone, were not the most informative features for the classification results. The use of SOCRAT enables interactive interrogation of morphometric data in a visual manner, supported by analytical tools. This method of interactive 'visual analytics' provides insights into feature dependencies and interactions, and can be used for result interpretation. We also provide an ability to visualize 3D volumetric images and extracted meshed online using SOCR Dynamic Visualization Toolkit web application [57].

Our computational approach is scalable and capable of processing various complex big 3D imaging data, and is not limited to nuclear and nucleolar shapes. With some changes, it can be applied to other cellular and nuclear compartments of interest. More specifically, the robust smooth surface reconstruction algorithm can be directly applied to any 3D shapes as long as their topology is sphere-like. Together with molecular level techniques, such as Hi-C, our 3D shape morphometry workflow can form a powerful combination for the investigation of DNA architecture in the spatial and temporal framework of the 4D nucleome [6,74]. One example of the many possible future applications of this workflow is to study asymmetric cell division. Stem and progenitor cells are characterized by their ability to self-renew and produce differentiated progeny. A balance between these processes is achieved through controlled asymmetric divisions and is necessary to generate cellular diversity during development and to maintain adult tissue homeostasis. Disruption of this balance may result in premature depletion of the stem/progenitor cell pool, or abnormal growth [75,76]. In many tissues, dysregulated asymmetric divisions are associated with cancer. Whether there is a causal relationship between asymmetric cell division

defects and cancer initiation is unknown. We propose that our shape analysis pipeline will be useful in studying the 4D nucleome topology of morphogenesis and cancer initiation.

As one of the approach limitations we point out that other geometric measures can be used to characterize shapes of interest, such as intrinsic shape context, compactness, symmetry, smoothness, convexity, etc. In the current representation, analyzable shapes are limited to genus zero surfaces, which is a fair assumption when modeling objects like nuclei or nucleoli. However, it might be not trivial when considering other nuclear structures, for example, chromosome territories or interchromosomal loops, since their topologies may not be homeomorphic to a sphere, or may not appear to be genus zero under some imaging conditions and modalities. It is also conceivable, yet not very likely for the discretized LB, that 2 different shapes may have the same spectra, in which case, we may fail to detect the intrinsic differences between them due to false-negative error. Even though our workflow only requires little intervention (classifier selection and tuning), further improvements would involve adaptive implementations with even less manual intervention, as well as extraction of additional features. Another option is to use deep learning-based methods that alleviate the need to define features and allow to learn relevant patterns directly from data [77,78]. For example, textural features could possibly increase discriminatory power of the method and provide more information on chromatin reorganization. Since nuclear deformation serves as a proxy to underlying processes, the importance of particular features and the method's ability to classify nuclei does not provide direct insights into fundamental biological mechanism driving the observed morphometric differences between cell phenotypes or environmental conditions. The computational results

should be further tested and externally validated using other experimental conditions and prospective data.

Conclusions

Quantification of cell nuclear morphology enables more subtle characterization of cellular phenotypic traits, which can be associated with functional changes coupled to underlying biological processes. Using the new methodology described in this paper, we compared the morphology of serum-starved vs. proliferating fibroblast cells as a control, follow by a comparison of epithelial with mesenchymal human prostate cancer cell lines. In the case of fibroblast classification, our results show the importance of nuclear morphometric change, along with the number of detected nucleoli per nucleus, and various internal nucleolar morphometric measures. Results for PC3 cells demonstrate that the changes in nucleolar morphology are the most informative. However, in both cell lines both nuclear and nucleolar morphometric measures contribute to the discriminative power of the classification algorithms. To the best of our knowledge, this study is the first where a 3D morphometric assay could easily distinguish between the epithelial and mesenchymal cell nuclei. We also suggest that further investigation of highly ranked features that were not previously reported, such as nucleolar curvedness, shape index and fractal dimension, may provide interesting mechanistic insights.

The ability to automate the processes of specimen collection, image acquisition, data pre-processing, computation of derived biomarkers, modeling, classification, and analysis can significantly impact clinical decision-making and fundamental investigation of cell deformation. To our knowledge, this is the first attempt to combine 3D cell nuclear shape modeling by robust

smooth surface reconstruction and extraction of shape morphometry measures into a highly parallel pipeline workflow protocol for morphological analysis of thousands of nuclei and nucleoli in 3D. This approach allows efficient and informative evaluation of cell shapes in the imaging data and represents a reproducible technique that can be validated, modified, and repurposed by the biomedical community. This facilitates result reproducibility, collaborative method validation, and broad knowledge dissemination.

Availability of materials and data

The documentation supporting the conclusions of this article together with the pipeline workflows and underlying source code are made available online on the project webpage SOCR 3D Cell Morphometry Project, http://socr.umich.edu/projects/3d-cell-morphometry [54].

Competing interests

The authors declare that they have no competing interests.

Funding

This work was partially supported by the National Science Foundation grants 1636840, 1023115, 1022560, 1022636, 0089377, 9652870, 0442992, 0442630, 0333672, 0716055, 1734853; the National Institutes of Health grants P20 NR015331, U54 EB020406, P50 NS091856, P30 DK089503, P30 AG053760, T32 GM070449, UL1TR002240; the Department of Computational Medicine and Bioinformatics Research Account, University of Michigan; the James Buchanan Brady Urological Institute, John Hopkins University within the contract "Promoting Scientific Progress through Biomedical Research, Biomedical Informatics and the Development Of

Ontological and Biomedical Informatics Tools that enable Collaborative Biomedical Research" (Sponsor ID 420900), and the Elsie Andresen Fiske Research Fund.

Author contributions

AA and DD designed and built modules for data pre-processing and curation. AAK, AA, DD, AAF, GVF, WM, JRD, GAH, GZ, AC, JWW iteratively evaluated various image analysis approaches. AAK, AAF, SSH and IDD designed and built modules for 3D shape modeling, morphometric feature extraction, and visual analytics. AAK, AAF, AA, and IDD designed, built, and executed high-throughput Pipeline workflows. AAK implemented and performed statistical analysis and result interpretation. GVF, WM, JRD, GAH, GZ, AC, JWW, JEV, RWV, KJP and DSC contributed to conceptualization of study. IDD and BDA conceived the study. AAK, AAF, AA, IDD and BDA wrote the manuscript with input from other authors. All authors participated in numerous project discussions and decision making. All authors read and approved the final manuscript.

Acknowledgements

Many colleagues at the Michigan Institute for Data Science (MIDAS), the Department of Computational Medicine and Bioinformatics, and the Statistics Online Computational Resource (SOCR) provided contributions including ideas, pilot testing, improvement suggestions and other assistance in the development and validation of these methods. We also would like to thank colleagues at the Laboratory of Neuro Imaging (LONI), at the Keck School of Medicine, University of Southern California, for providing technical support for the LONI Pipeline environment.

References

1 Jevtic, P., Edens, L. J., Vukovic, L. D. & Levy, D. L. Sizing and shaping the nucleus: mechanisms and significance. *Curr Opin Cell Biol* **28**, 16-27, doi:10.1016/j.ceb.2014.01.003 (2014).

2 Bustin, M. & Misteli, T. Nongenetic functions of the genome. *Science* **352**, aad6933, doi:10.1126/science.aad6933 (2016).

3 Dundr, M. & Misteli, T. Biogenesis of nuclear bodies. *Cold Spring Harb Perspect Biol* **2**, a000711, doi:10.1101/cshperspect.a000711 (2010).

4 Fedorchak, G. R., Kaminski, A. & Lammerding, J. Cellular mechanosensing: getting to the nucleus of it all. *Prog Biophys Mol Biol* **115**, 76-92, doi:10.1016/j.pbiomolbio.2014.06.009 (2014).

5 Uhler, C. & Shivashankar, G. V. Regulation of genome organization and gene expression by nuclear mechanotransduction. *Nat Rev Mol Cell Biol* **18**, 717-727, doi:10.1038/nrm.2017.101 (2017).

6 Higgins, G. A., Allyn-Feuer, A., Handelman, S., Sadee, W. & Athey, B. D. The epigenome, 4D nucleome and next-generation neuropsychiatric pharmacogenomics. *Pharmacogenomics* **16**, 1649-1669, doi:10.2217/pgs.15.111 (2015).

7 Meaburn, K. J. Spatial Genome Organization and Its Emerging Role as a Potential Diagnosis Tool. *Front Genet* **7**, 134, doi:10.3389/fgene.2016.00134 (2016).

8 Zink, D., Fischer, A. H. & Nickerson, J. A. Nuclear structure in cancer cells. *Nature reviews. Cancer* **4**, 677-687, doi:10.1038/nrc1430 (2004).

9 Veltri, R. W. & Christudass, C. S. Nuclear morphometry, epigenetic changes, and clinical relevance in prostate cancer. *Advances in experimental medicine and biology* **773**, 77-99, doi:10.1007/978-1-4899-8032-8_4 (2014).

10 Montanaro, L., Trere, D. & Derenzini, M. Nucleolus, ribosomes, and cancer. *The American journal of pathology* **173**, 301-310, doi:10.2353/ajpath.2008.070752 (2008).

11 Pincus, Z. & Theriot, J. A. Comparison of quantitative methods for cell-shape analysis. *Journal of microscopy* **227**, 140-156, doi:10.1111/j.1365-2818.2007.01799.x (2007).

12 Huang, H. *et al.* Cancer diagnosis by nuclear morphometry using spatial information. *Pattern Recognit Lett* **42**, 115-121, doi:10.1016/j.patrec.2014.02.008 (2014).

13 Caserta, F. *et al.* Determination of fractal dimension of physiologically characterized neurons in two and three dimensions. *J Neurosci Methods* **56**, 133-144 (1995).

14 Choi, H. J. & Choi, H. K. Grading of renal cell carcinoma by 3D morphological analysis of cell nuclei. *Comput Biol Med* **37**, 1334-1341, doi:10.1016/j.compbiomed.2006.12.008 (2007).

15 Meyer, M. G. *et al.* Automated cell analysis in 2D and 3D: A comparative study. *Pattern Recogn* **42**, 141-146 (2009).

16 Dufour, A. C. *et al.* Signal Processing Challenges in Quantitative 3-D Cell Morphology: More than meets the eye. *IEEE Signal Processing Magazine* **32**, 30-40, doi:10.1109/MSP.2014.2359131 (2015).

17 Pegoraro, G. & Misteli, T. High-throughput Imaging as a versatile and unbiased discovery tool. *Methods (San Diego, Calif.)* **96**, 1-2, doi:10.1016/j.ymeth.2016.01.003 (2016).

18 Kalinin, A. A. *et al.* 3D Cell Nuclear Morphology: Microscopy Imaging Dataset and Voxel-Based Morphometry Classification Results. *bioRxiv*, doi:10.1101/208207 (2018).

19 Brechbuhler, C., Gerig, G. & Kubler, O. Parametrization of Closed Surfaces for 3-D Shape-Description. *Comput Vis Image Und* **61**, 154-170 (1995).

20 Antoine, J. P. & Vandergheynst, P. Wavelets on the 2-sphere: A group-theoretical approach. *Appl Comput Harmon A* **7**, 262-291 (1999).

21 Khairy, K., Foo, J. & Howard, J. Shapes of Red Blood Cells: Comparison of 3D Confocal Images with the Bilayer-Couple Model. *Cell Mol Bioeng* **1**, 173-181, doi:10.1007/s12195-008-0019-5 (2010).

22 Singh, S. *et al.* Non-parametric population analysis of cellular phenotypes. *Med Image Comput Comput Assist Interv* **14**, 343-351 (2011).

23 Ducroz, C., Olivo-Marin, J. C. & Dufour, A. Characterization of cell shape and deformation in 3D using Spherical Harmonics In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI),* 848-851 (IEEE, 2012).

24 Du, C. J., Hawkins, P. T., Stephens, L. R. & Bretschneider, T. 3D time series analysis of cell shape using Laplacian approaches. *BMC bioinformatics* **14** (2013).

25 Shen, L. & Makedon, F. Spherical mapping for processing of 3D closed surfaces. *Image Vision Comput* **24**, 743-761 (2006).

26 Shi, Y. G. *et al.* Robust Surface Reconstruction via Laplace-Beltrami Eigen-Projection and Boundary Deformation. *IEEE transactions on medical imaging* **29**, 2009-2022 (2010).

27 Levy, B. Laplace-Beltrami Eigenfunctions Towards an Algorithm That "Understands" Geometry In *IEEE International Conference on Shape Modeling and Applications 2006 (SMI'06),* 13-13 (2006).

28 Han, X., Xu, C. Y. & Prince, J. L. A topology preserving level set method for geometric deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**, 755-768 (2003).

29 Batchelor, P. G. *et al.* Measures of folding applied to the development of the human fetal brain. *IEEE transactions on medical imaging* **21**, 953-965, doi:10.1109/TMI.2002.803108 (2002).

30 Koenderink, J. J. & Vandoorn, A. J. Surface Shape and Curvature Scales. *Image Vision Comput* **10**, 565-565 (1992).

31 Dinov, I. D. *et al.* Efficient, Distributed and Interactive Neuroimaging Data Analysis Using the LONI Pipeline. *Front Neuroinform* **3**, 22, doi:10.3389/neuro.11.022.2009 (2009).

32 Fani, N. *et al.* FKBP5 and attention bias for threat: associations with hippocampal function and shape. *JAMA Psychiatry* **70**, 392-400, doi:10.1001/2013.jamapsychiatry.210 (2013).

33 Moon, S. W. *et al.* Structural Brain Changes in Early-Onset Alzheimer's Disease Subjects Using the LONI Pipeline Environment. *Journal of neuroimaging : official journal of the American Society of Neuroimaging* **25**, 728-737, doi:10.1111/jon.12252 (2015).

34 Ramo, P., Sacher, R., Snijder, B., Begemann, B. & Pelkmans, L. CellClassifier: supervised learning of cellular phenotypes. *Bioinformatics (Oxford, England)* **25**, 3028-3030, doi:10.1093/bioinformatics/btp524 (2009).

35 Held, M. *et al.* CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. *Nature methods* **7**, 747-754, doi:10.1038/nmeth.1486 (2010).

36 Pau, G., Fuchs, F., Sklyar, O., Boutros, M. & Huber, W. EBImage--an R package for image processing with applications to cellular phenotypes. *Bioinformatics (Oxford, England)* **26**, 979-981, doi:10.1093/bioinformatics/btq046 (2010).

37 Kamentsky, L. *et al.* Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics (Oxford, England)* **27**, 1179-1180, doi:10.1093/bioinformatics/btr095 (2011).

38 Chiang, M. *et al.* Analysis of in vivo single cell behavior by high throughput, human-in-the-loop segmentation of three-dimensional images. *BMC bioinformatics* **16**, 397, doi:10.1186/s12859-015-0814-7 (2015).

39 Peng, T. & Murphy, R. F. Image-derived, three-dimensional generative models of cellular organization. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* **79**, 383-391, doi:10.1002/cyto.a.21066 (2011).

40 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).

41 de Chaumont, F. *et al.* Icy: an open bioimage informatics platform for extended reproducible research. *Nature methods* **9**, 690-696, doi:10.1038/nmeth.2075 (2012).

42 Kankaanpaa, P. *et al.* BioImageXD: an open, general-purpose and high-throughput image-processing platform. *Nature methods* **9**, 683-689, doi:10.1038/nmeth.2047 (2012).

43 Ollion, J., Cochennec, J., Loll, F., Escude, C. & Boudier, T. TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization. *Bioinformatics (Oxford, England)* **29**, 1840-1841, doi:10.1093/bioinformatics/btt276 (2013).

44 Eliceiri, K. W. *et al.* Biological imaging software tools. *Nature methods* **9**, 697-710, doi:10.1038/nmeth.2084 (2012).

45 Peng, H., Bria, A., Zhou, Z., Iannello, G. & Long, F. Extensible visualization and analysis for multidimensional images using Vaa3D. *Nature protocols* **9**, 193-208, doi:10.1038/nprot.2014.011 (2014).

46 Li, L., Zhou, Q., Voss, T. C., Quick, K. L. & LaBarbera, D. V. High-throughput imaging: Focusing in on drug discovery in 3D. *Methods (San Diego, Calif.)* **96**, 97-102, doi:10.1016/j.ymeth.2015.11.013 (2016).

47 Ince, D. C., Hatton, L. & Graham-Cumming, J. The case for open computer programs. *Nature* **482**, 485-488, doi:10.1038/nature10836 (2012).

48 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **9**, 671-675 (2012).

49 Kalinin, A. A., Palanimalai, S. & Dinov, I. D. SOCRAT Platform Design: A Web Architecture for Interactive Visual Analytics Applications In *Proceedings of the 2nd Workshop on Human-In-the-Loop Data Analytics, HILDA'17* 1-6 (ACM, 2017).

50 Dinov, I. D. *et al.* Neuroimaging study designs, computational analyses and data provenance using the LONI pipeline. *PloS one* **5**, doi:10.1371/journal.pone.0013070 (2010).

51 Dinov, I. D. *et al.* Applications of the pipeline environment for visual informatics and genomics computations. *BMC bioinformatics* **12**, 304, doi:10.1186/1471-2105-12-304 (2011).

52 Thompson, P. M., Schwartz, C., Lin, R. T., Khan, A. A. & Toga, A. W. Three-dimensional statistical analysis of sulcal variability in the human brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **16**, 4261-4274 (1996).

53 Meyer, M., Desbrun, M., Schröder, P. & Barr, A. H. Discrete Differential-Geometry Operators for Triangulated 2-Manifolds In *Visualization and Mathematics III,* (eds Hans-Christian Hege & Konrad Polthier) 35-57 (Springer Berlin Heidelberg, 2003).

54 *SOCR 3D Cell Morphometry Project*, <http://socr.umich.edu/projects/3d-cell-morphometry> (2018).

55 Ferri, M. & Gagliardi, C. The only genus zero *n*-manifold is S^n. *Proc. Amer. Math. Soc.* **85**, 638-642, doi:10.1090/S0002-9939-1982-0660620-5 (1982).

56 Niethammer, M. *et al.* Global medical shape analysis using the Laplace-Beltrami spectrum. *Med Image Comput Comput Assist Interv* **10**, 850-857 (2007).

57 *SOCR Dynamic Visualization Toolkit* <http://socr.umich.edu/HTML5/DViewer/> (2018).

58 Terzopoulos, D. The computation of visible-surface representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **10**, 417-438, doi:10.1109/34.3908 (1988).

59 Larson, R. & Edwards, B. *Calculus*. 10th edn, (Cengage Learning, 2009).

60 Santaló, L. A. *Integral geometry and geometric probability*. 2nd edn, (Cambridge University Press, 2004).

61 Mandelbrot, B. B. *The fractal geometry of nature*. 1st edn, Vol. 173 (W. H. Freeman and Company, 1982).

62 Jagannathan, A. *Segmentation and Recognition of 3D Point Clouds within Graph-theoretic and thermodynamic frameworks : a thesis*, Northeastern University, 2005., (2005).

63 Husain, S. S., Kalinin, A., Truong, A. & Dinov, I. D. SOCR data dashboard: an integrated big data archive mashing medicare, labor, census and econometric information. *J Big Data* **2**, doi:10.1186/s40537-015-0018-z (2015).

64 van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *J Mach Learn Res* **9**, 2579-2605 (2008).

65 *SOCRAT Web Demo*, <http://socr.umich.edu/HTML5/SOCRAT/> (2018).

66 Cheplygina, V., Tax, D. M. J. & Loog, M. On classification with bags, groups and sets. *Pattern Recognition Letters* **59**, 11-17, doi:10.1016/j.patrec.2015.03.008 (2015).

67 Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research* **12**, 2825-2830 (2011).

68 van der Walt, S. *et al.* scikit-image: image processing in Python. *PeerJ* **2**, e453, doi:10.7717/peerj.453 (2014).

69 Langan, T. J. & Chou, R. C. Synchronization of mammalian cell cultures by serum deprivation. *Methods Mol Biol* **761**, 75-83, doi:10.1007/978-1-61779-182-6_5 (2011).

70 Seaman, L., Meixner, W., Snyder, J. & Rajapakse, I. Periodicity of nuclear morphology in human fibroblasts. *Nucleus* **6**, 408-416, doi:10.1080/19491034.2015.1095432 (2015).

71 *SPHARM-MAT*, <http://www.iu.edu/~spharm/> (2010).

72 Shen, L., Farid, H. & McPeek, M. A. Modeling Three-Dimensional Morphological Structures Using Spherical Harmonics. *Evolution* **63**, 1003-1016 (2009).

73 Verdone, J. E., Parsana, P., Veltri, R. W. & Pienta, K. J. Epithelial-mesenchymal transition in prostate cancer is associated with quantifiable changes in nuclear structure. *The Prostate* **75**, 218-224, doi:10.1002/pros.22908 (2015).

74 Zheng, G. *et al.* Rotational 3D mechanogenomic Turing patterns of human colon Caco-2 cells during differentiation. *bioRxiv*, doi:10.1101/272096 (2018).

75 Morrison, S. J. & Kimble, J. Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature* **441**, 1068-1074, doi:10.1038/nature04956 (2006).

76 Gomez-Lopez, S., Lerner, R. G. & Petritsch, C. Asymmetric cell division of stem and progenitor cells during homeostasis and cancer. *Cell Mol Life Sci* **71**, 575-597, doi:10.1007/s00018-013-1386-1 (2014).

77 Ching, T. *et al.* Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface* **15**, doi:10.1098/rsif.2017.0387 (2018).

78 Kalinin, A. A. *et al.* Deep learning in pharmacogenomics: from gene regulation to patient stratification. *Pharmacogenomics* **19**, 629-650, doi:10.2217/pgs-2018-0008 (2018).