POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Deep learning-driven image analysis of epithelial tissues structure and organization

Tesi di Laurea Magistrale in
Ingegneria Matematica

Author: **Federico Carrara**

Student ID: 970583
Advisor: Prof. Francesca Ieva
Co-advisors: Prof. Dagmar Iber
Academic Year: 2022-23

Abstract

Epithelial tissues are widely present across all organs and organisms, serving as barriers and boundaries between biological means. Yet, their 3D organization differs widely. Epithelia can organize in single or multiple layers, and cell shapes range from cuboidal to highly elongated shapes, whose long axis runs either along the apico-basal axis (columnar) or parallel to the epithelial surface (squamous). Those different organizations must reflect differences in the biophysical cell parameters, but their type and magnitude have remained elusive. In this study, we leverage deep learning methods for the segmentation of 3D images of epithelial samples acquired at the light sheet microscope. Then, we present EpiStats, a library for the analysis of morphological tissue features. By applying EpiStats to the segmented epithelial samples, we provide a complete experimental characterization of the tissues, highlighting common traits as well as discrepancies between them. Finally, we employ SimuCell3D, a 3D cell-based simulation framework, to perform an exploratory investigation of the mechanical and biophysical cell parameters. This preliminary research highlights aspects of the current state of SimuCell3D that need to be improved in future iterations to provide more conclusive results about epithelial tissues' mechanical properties.

Keywords: Epithelial tissues, Deep learning, Image analysis, Numerical simulations

Abstract in lingua italiana

I tessuti epiteliali sono ampiamente diffusi in ogni organo e organismo. Nonostante in ogni parte compiano la stessa funzione di barriera tra mezzi, la loro organizzazione tridimensionale è estremamente varia. In particular, gli epiteli possono organizzarsi in uno o più strati, e le forme delle cellule vanno da cuboidali a forme maggiormente allungate, come quelle colonnari, il cui asse lungo si estende lungo l'asse apico-basale e quelle squamose, disposte parallelamente alla superficie epiteliale. Le diverse conformazioni di tali tessuti epiteliali si riflettono in differenze nei parametri biofisici delle cellule. Tuttavia, la specificità e l'entità di tali discrepanze rimangono sfuggenti. In questo studio, utilizziamo metodi di apprendimento profondo per la segmentazione di immagini 3D di campioni epiteliali acquisite al microscopio light sheet. Successivamente, presentiamo EpiStats, una libreria per l'analisi delle caratteristiche morfologiche dei tessuti. Applicando EpiStats ai campioni epiteliali raccolti, produciamo una caratterizzazione sperimentale dei vari tessuti, evidenziando tratti comuni e difference salienti. Infine, utilizziamo SimuCell3D, un software per la simulazione di modelli basati su singola cellular in 3D. Lo scopo è quello di condurre un'indagine esplorativa dei parametri meccanici e biofisici delle cellule. Questa ricerca preliminare mette in risalto alcuni aspetti dello stato attuale di SimuCell3D che richiedono miglioramenti per le future versioni dello studio, al fine di ottenere risultati più conclusivi sulle proprietà meccaniche dei tessuti epiteliali.

Parole chiave: Tessuti epiteliali, Apprendimento profondo, Analisi di immagini, Simulazioni numeriche

Contents

Introduction

An introduction on epithelial tissues

Covering epithelial tissues are sheets of tightly packed, closely connected cells that cover the external surfaces (e.g., epidermis) and line the internal cavities of the body (e.g., membranes of tubular organs) (Fig. 1a). Their main functions are to provide a barrier between the interior and exterior of the body and to separate internal environments (Pacheco et al., 2008).

Epithelial tissues are distributed across the majority of the body's organs, and as a result, they exhibit diverse arrangements depending on each organ's specific function. Structurally (Fig. 1b), epithelia can be composed of a single layer of cells (simple epithelium) or multiple layers of cells (stratified epithelium). For what concerns their shape, single epithelial cells are classified depending on their elongation. On the one hand, squamous cells present a major axis parallel to the epithelial surface and have flattened nuclei. On the other hand, in columnar cells, the major axis runs along the apical-basal axis and nuclei are usually cylindrical. Finally, cuboidal cells are not elongated and are characterized by round-shaped centrally located nuclei. In most cases, simple epithelia are made by cells of the same shape. Therefore, we can have simple squamous, simple columnar, and simple cuboidal epithelial tissues. On the contrary stratified epithelia are classified according to the shape of cells in the upper layer, while the underlying layers are usually composed of cuboidal (polygonal) cells. Thus, we define stratified squamous, stratified cuboidal, and stratified columnar epithelia. Among stratified epithelia, a peculiar structure is the one shown in the transitional epithelium. In this case, the shape of cells can change due to stretching of the tissue. An example of transitional epithelial tissue is the Urothelium, located in the urinary system. The shape of its cells depends on the filling of the urinary tract. A specialized form of simple epithelia is the pseudostratified epithelium. Its peculiarity is given by the fact that its columnar cells are all in contact with the basal membrane, but only some of them have a free apical surface. This specific organization is caused by the nuclei arrangement, which are not aligned at the same level (Dudas, 2023).

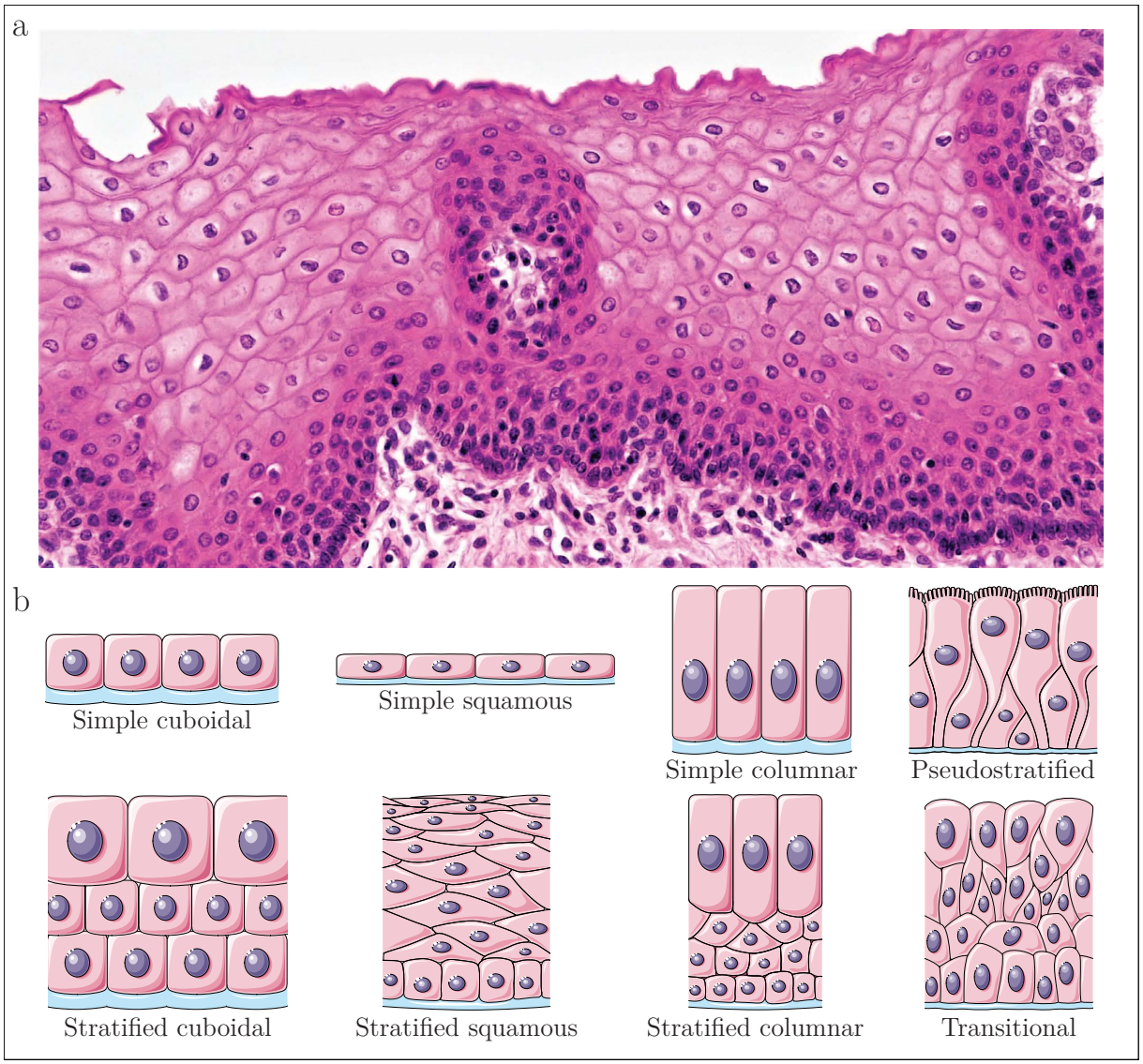

Figure 1: **Epithelial tissues structure and classification.** **a,** A microscope image of a squamous stratified epithelium *(Berkshire Community College Bioscience Image Library, CC0, via Wikimedia Commons)*. **b,** Classification of epithelial tissues is based on the number of layers (simple vs. stratified epithelia) and the shape of the cells in the most superficial layer (cuboidal, squamous, columnar). Pseudostratified and Transitional epithelia are particular structures of, respectively, simple and stratified tissues.

A review on methods and discoveries in epithelial biology

The driving factors that lead to such a variety of epithelia organizations and structures are still elusive and subject to research. Since the beginning of the 20th century, scientists

have developed mathematical models to describe the organization of epithelia.

Most of the early models were based on the reductionist assumption of prismatic-shaped epithelial cells (Fig. 2a). According to this assumption, cells keep the same neighbor relationships along the entire apical-basal axis. At the same time, prism-like cells are allowed to reduce one of the polygonal surfaces (apical or basal) to accommodate the curvature of the tissue, assuming a shape defined as *frustum* (Fig. 2b, 2d). Another consequence of the model is that 2D slices of epithelial cells are modeled as convex tessellations of the plane (Fig. 2c). An important advantage of the *prismatic approximation* is the possibility of investigating tissue packing by resorting to common tools of mathematical topology, as well as borrowing results from other physical processes, for example, the growth and organization of crystals and foams, as the founding geometrical model is essentially the same.

In this setting, Reinhart formally deduced that, according to *Euler's principle for convex polyhedrons*, the average number of sides of the cells in a plane should be six (Reinhardt, 1918).
Later, Lewis established the existence of a linear relationship between the average cell areas and the number of neighbors (Lewis, 1928):

$$\frac{\overline{A_n}}{\overline{A}} = \frac{n-4}{2} \tag{1}$$

In the expression above n is the number of neighbors, $\overline{A_n}$ is the average area of cells having n neighbors, and $\overline{A}$ is the average area among all the cells. *Lewis' Law* was subsequently demonstrated as a direct consequence of the maximum entropy principle (Rivier and Lissowski, 1982) and at the same time of the minimization of the lateral cell-cell contact surface energy (Kokic et al., 2019), which drives cells to regular polygonal shapes.

Another mathematical principle observed in convex tessellations of the plane is *Aboav-Weaire's law* (Aboav, 1970; Chiu, 1995), which states that the average number of sides of the neighbors of a cell is inversely related to its number of neighbors:

$$m_n = 5 + \frac{8}{n} \tag{2}$$

where m_n is the average number of neighbors computed over the adjacent cells to a given cell having n neighbors. Originally deduced for growing polycrystals, this law was afterward confirmed for apical planes of growing epithelia (Bi et al., 2014; Sánchez-Gutiérrez et al., 2016). Similar to Lewis' law, it has been proven that Aboav-Weaire's law results from the minimization of lateral contact surface energy (Vetter et al., 2019). However,

whereas the former regulates the cells' side lengths, the latter controls the cells' contact angles.

The mathematical principles and properties described above were assumed to be valid in a 3D context in epithelial monolayers given the prismatic simplification. However, this approximation started to be challenged after discovering that cells did not preserve the same number of neighbors on the apical and basal surfaces (Fig. 2e) (Condic et al., 1991). *Apical-basal intercalations* (i.e., cell-neighboring changes between apical and basal surfaces) (Fig. 2f) were also reported in mouse epithelia (Xu et al., 2016). These discoveries led to the formulation of a new 3D geometrical representation of cells defined as *scutoid* (Fig. 2c) (Gómez-Gálvez et al., 2018). This model allows cells to have apical-basal intercalations and provides a way to achieve epithelial bending by minimizing tissue energy and stabilizing 3D cell packing.

These new discoveries, together with massive advances in microscopy and machine learning methods for automated cell segmentation, spurred the development of further studies about 3D epithelial organization. In Escudero et al. (2011) the author produces a characterization of the epithelial structure considering building a cellular network in which cells are nodes and contacts are edges. Improvements of this approach are presented in Kursawe et al. (2016) and in Vicente-Munuera et al. (2020). An alternative method involves the application of modern image analysis techniques to extract morphological statistics from epithelial cells and the subsequent analysis of such statistics over the tissues. An example is Gómez et al. (2021), in which the author shows that epithelial organization emerges from simple physical principles, such as the minimization of cell-cell surface energy. Using similar techniques, in Rupprecht et al. (2017) it is shown that geometric constraints can influence three-dimensional cell morphology and packing within epithelial tissues.

In parallel with the development of mathematical models and computational analysis tools, researchers started to implement in-silico models for the simulation of the epithelial tissues' dynamics. Cell-based models simulate virtual tissue by modeling cells as individual agents, each possessing unique mechanical characteristics and behaviors. These models provide an in-silico environment in which stress patterns and mechanical attributes of cells can be manipulated to investigate their influence on tissue structure and functionality (Runser et al., 2023).

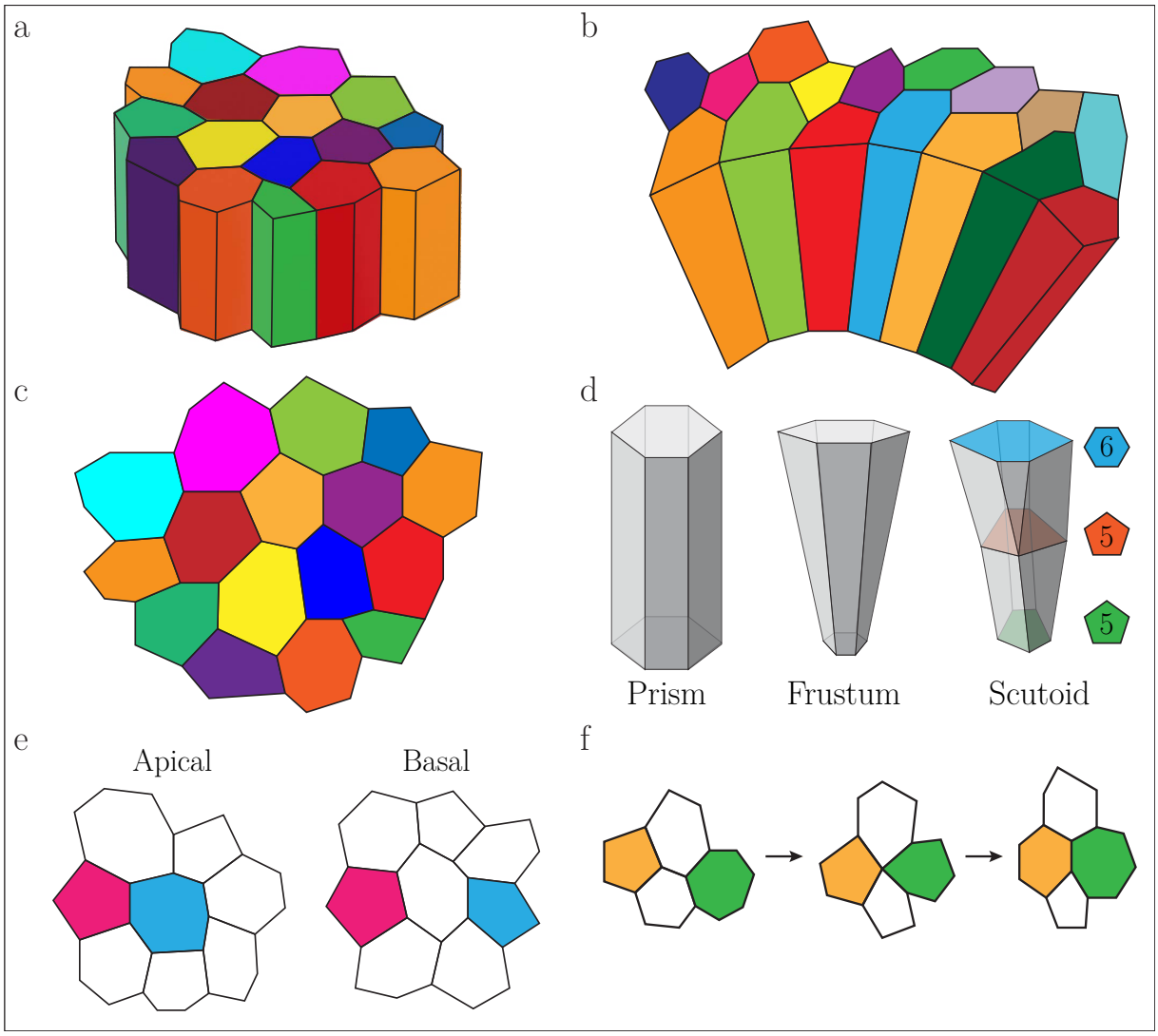

Figure 2: **Schematical representations of epithelial tissues and cells. a,** Illustration of a planar simple epithelium under prismatic approximation. **b,** Cells can adapt their conformation to the tissue curvature by adopting the shape of a truncated pyramid (i.e. frustum). **c,** When cells are modeled as prisms, 2D slices of epithelial tissues become convex tessellations of the plane. **d,** Depending on the chosen approximation, an epithelial cell can be represented as a prism, a frustum, or a scutoid. **e,** Prismatic approximation started to be challenged after the discovery of changes in cells' neighbors between apical and basal planes (Condic et al., 1991; Xu et al., 2016). **f,** Schematic 2D depiction of an apical-basal intercalation.

All numerical models are the result of a trade-off between the level of resolution and computational cost. This choice usually depends on the specific research problem the investigator wants to address. For instance, when the quantity of cells to simulate is more

important than their individual characteristics, center-based models (i.e., a model in which cells are represented as simple spheres) are the most efficient solution. Vertex models, instead, allow increasing the simulations' likelihood by approximating cell shapes with polygons in 2D and polyhedrons in 3D, where forces are applied at their vertices (Nagai and Honda, 2001; Farhadifar et al., 2007; Canela-Xandri et al., 2011; Osterfield et al., 2013; Monier et al., 2015; Hirashima and Adachi, 2019). Finally, when cells have complex shapes and providing a geometrically realistic representation of tissues is important Deformable Cell Models (DCMs) are a suitable choice. In these models, cells are regarded as closed loops of connected points in 2D (Rejniak, 2005; Merks et al., 2011; Tanaka et al., 2015; Ataei et al., 2021; Kim et al., 2021; Brown et al., 2021), or a closed triangulated surface in 3D (Brakke, 1992; Madhikar et al., 2018; Liedekerke et al., 2020; Wang et al., 2021; Torres-Sánchez et al., 2022; Okuda and Hiraiwa, 2023; Runser et al., 2023). DCMs are well-suited to realistically represent a huge number of biological processes, clearly at the expense of a remarkable computational cost and more frequent numerical stability. For these reasons, such models are frequently used for highly detailed simulations of a limited number of cells.

Objectives and aims

As stated in the initial paragraph, epithelial tissues are undeniably intriguing in nature. Despite their main functions remaining somewhat consistent across the body, they exhibit diverse structural arrangements and cell configurations. Therefore, delving into the exploration of the mechanisms and influences that drive the development of a variety of structures within epithelial tissues unquestionably attracts one's curiosity.

In the second paragraph, we aimed to present a summary of the current research status in the field of epithelial biology. We highlighted the development and application of numerous methods over the past decades for analyzing the recurring morphological structures of epithelia. Despite these advancements, a persistent challenge exists in achieving a comprehensive understanding of the properties that characterize various tissues. Specifically, gaining insight into the unique features that differentiate epithelia from a structural and organizational standpoint is crucial. This understanding can illuminate the underlying reasons guiding tissues to adopt specific configurations to fulfill distinct biological requirements. Unraveling these principles may serve as a crucial element in uncovering the holy grail of developmental biology, namely the comprehension of the biophysical and physiological mechanisms governing the morphogenesis of epithelial and other biological tissues.

Clearly, the possibility of providing sufficient experimental evidence to unveil the biological principles goes hand in hand with the development of powerful, yet efficient technological solutions. Within the context of analyzing the morphology and mechanics of epithelial tissues, it becomes evident that several critical technological tools are indispensable. Firstly, the availability of high-resolution 3D microscope images is crucial for capturing the intricate relationships among epithelial cells. Consequently, the implementation of a precise 3D cell segmentation algorithm becomes essential to identify individual cell entities within the imaged tissue, facilitating the study of both their individual and collective features. Subsequently, sophisticated methods are fundamental for conducting a comprehensive analysis of the properties of the segmented cells. In this respect, opting for 3D image analysis software becomes an apparent choice for accurately describing tissue morphological characteristics, encompassing both cell shape and tissue organization. Conversely, frameworks based on cell simulations emerge as the ideal option for exploring the mechanical interactions and stress distribution within biological tissues. Despite the considerable effort and remarkable advancements showcased in numerous works among the biology research community, our review underscores a prevalent deficiency in robust, cohesive approaches from a technological standpoint. These approaches are necessary to enable a high-quality, end-to-end analysis of epithelial tissue properties.

In this context, our project aims to conduct a comprehensive characterization study of 3D epithelial tissues, providing a detailed comparison of various organizations and structures. To achieve this objective, our focus is twofold: on the one hand, we develop potent and efficient technological tools tailored to address the current gaps in epithelial biology. On the other hand, we exploit their capabilities to extract meaningful insights that can offer an answer to our research question. Starting with 3D light sheet microscope images of diverse mouse epithelia, we perform high-quality 3D cell segmentation to identify individual cells. Subsequently, we pursue a dual approach. Firstly, we implement an optimized 3D image analysis pipeline to extract information regarding epithelial organization and single-cell morphology. This allows us to discern differences among tissues and empirically evaluate the consistency of mathematical and geometrical results, both on 2D slices and in the overall 3D tissues. Secondly, we delve into the mechanical structure of epithelial tissues using a 3D cell-based simulation framework.

The technological stack involved in this project massively relies on Python programming language and its libraries. In particular, for what concerns image preprocessing and analysis we relied on *Numpy* (Harris et al., 2020), *Scikit-image* (van der Walt et al., 2014), *SciPy* (Virtanen et al., 2020), and *napari* (Chiu and Clack, 2022). In the framework of cell-based simulations, we employed the *Trimesh* (Dawson-Haggerty, 2019), *VTK* (Schroeder et al.,

2006), *PyMeshLab* (Muntoni and Cignoni, 2021), *PyMeshFix* (Attene, 2010), and *Open3D* (Zhou et al., 2018) Python libraries for the generation, refinement and handling of triangular cell meshes. We gathered the algorithms for the computation of morphological features from segmented epithelial tissues in an open-source Python library called *EpiStats*. The codebase contains also a pipeline that performs the refinement of triangular meshes and it is available on GitHub at `https://github.com/AntanasMurelis/EpiStats`. For the sake of 3D cell segmentation, we resorted to existing software tools implemented in Python, namely *micro-SAM* (Archit et al., 2023) and *Plantseg* (Wolny et al., 2020). Additionally, we employed *Fiji* (Schindelin et al., 2012) and *Ilastik* (Sommer et al., 2011; Berg et al., 2019) programs for image preprocessing, *Paraview* (Ahrens et al., 2005) for the visualization of triangular meshes and *SimuCell3D* (Runser et al., 2023) software for the simulation of tissue mechanics.

In the following, we outline the thesis's structure, where we will delve into greater detail on how our research objective was actively pursued. In Chapter 1 we present the microscope images employed for the study and we explain how they were collected and selected. In Chapter 2 we describe the 3D cell segmentation pipeline we chose to utilize, and we illustrate enhancements we implemented to improve its overall performance. In Chapter 3 we first present EpiStats, an efficient Python library to collect morphological statistics from 3D cells, and afterward, we compare the extracted features among different epithelia and discuss the consistency with theoretical laws. Finally, in Chapter 4 we describe the application of the SimuCell3D software for the inference of mechanical properties in epithelial tissues via numerical simulations. In particular, we detail its underlying model and the technical setup necessary for the framework to deal with the dynamical simulation of epithelial tissues.

1 | Data

1.1. Introduction

The ability to create more comprehensive models for epithelial tissues, leading to more precise outcomes and descriptions, is strictly linked to the availability of high-resolution microscope images, in which salient cell features (membrane, nuclei, boundaries) are well-defined. Specifically, if the research goal is to understand the overall arrangement of these tissues and cell interactions, 3D imaging proves to be a significantly more potent tool, offering extensive insights into both epithelial structure and individual cell shapes.

Light sheet fluorescence microscopy (LSFM) provides an intermediate-to-high spatial resolution, but remarkable sectioning capabilities and high speed. For this reason, it is a suitable technique for in-depth analyses of rather thick, optically cleared samples (Stelzer et al., 2021). Thanks to the use of a thin sheet of light (Fig. 3b), LSFM can be employed for long-term 3D observations of biological specimens at high temporal resolution.

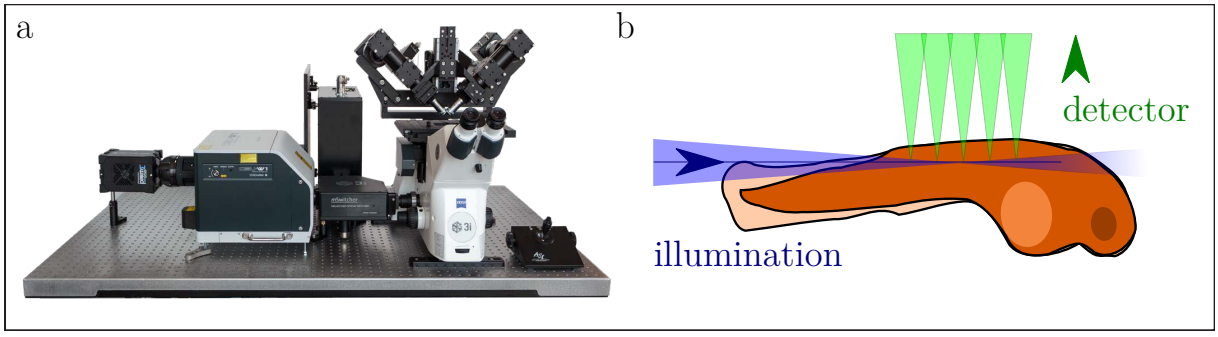

Figure 3: **Light sheet fluorescence microscopy offers high-quality 3D imaging for biological samples. a,** A picture of a modern Light sheet microscope. **b,** LSFM uses a thin sheet of light to excite fluorophores within the focal volume.

Light sheet microscopes use a thin sheet of light to excite fluorophores within the focal volume (Stelzer et al., 2021). Fluorophores are particular proteins that, when exposed to light in the blue to ultraviolet range, exhibit fluorescence at different wavelengths, thus

resulting in different colors in the microscope images. In order to amplify the fluorescence signals, fluorophore proteins are stained by using specific antibodies, by a procedure that takes the name of immunostaining (Liu et al., 2011).

In the following paragraphs, we describe the protocol employed to collect 3D images from samples of embryonic (stage E17.5) mouse epithelial tissues. Then, we present the images selected for the study, motivating our choices.

1.2. Methods

In this section, we report a summary of the procedure followed to obtain high-resolution 3D light sheet microscope images from a variety of epithelial tissues. As we mentioned earlier, epithelial samples were collected from dissected organs of mice at the embryonic stage E17.5. In particular, the specimens were stained for GFP, DAPI, tdTomato, and laminB1 proteins. Specifically, for the sake of our analysis, we are mainly interested in the GFP and DAPI signals. The former highlights cell membranes in bright green, whereas the latter marks the cell nuclei with a blue/cyan color. The choice of primarily focusing on those signals attains to the fact that to obtain high-quality cell segmentation the presence of sharp, well-defined cell boundaries is key. Moreover, knowing the nuclei's position is extremely helpful within the segmentation process. Indeed, it eases the identification of single cells, since in healthy conditions they possess only one nucleus.

The microscope images were acquired with a *Zeiss Z.1* light sheet microscope. The whole procedure of specimen preparation and image collection was carried out by Dr. Laura Schaumann. An extensive and detailed description of the employed protocol is reported in Appendix A.

1.3. Results

As discussed in the introductory chapter the ultimate goal of this study is to provide a complete characterization of epithelial tissues. We also mentioned that, even though they approximately serve the same functions, epithelia show diverse configurations and geometries. Therefore, to provide a thorough and detailed analysis, one needs to take into consideration a variety of epithelial tissues, each one possibly showing a different global organization and peculiar shape of its cells. Nevertheless, the choice of 3D images to include in the dataset must also take into consideration their quality and accessibility. Indeed, higher sharpness and definition of images naturally lead to greater accuracy in the following cell segmentation.

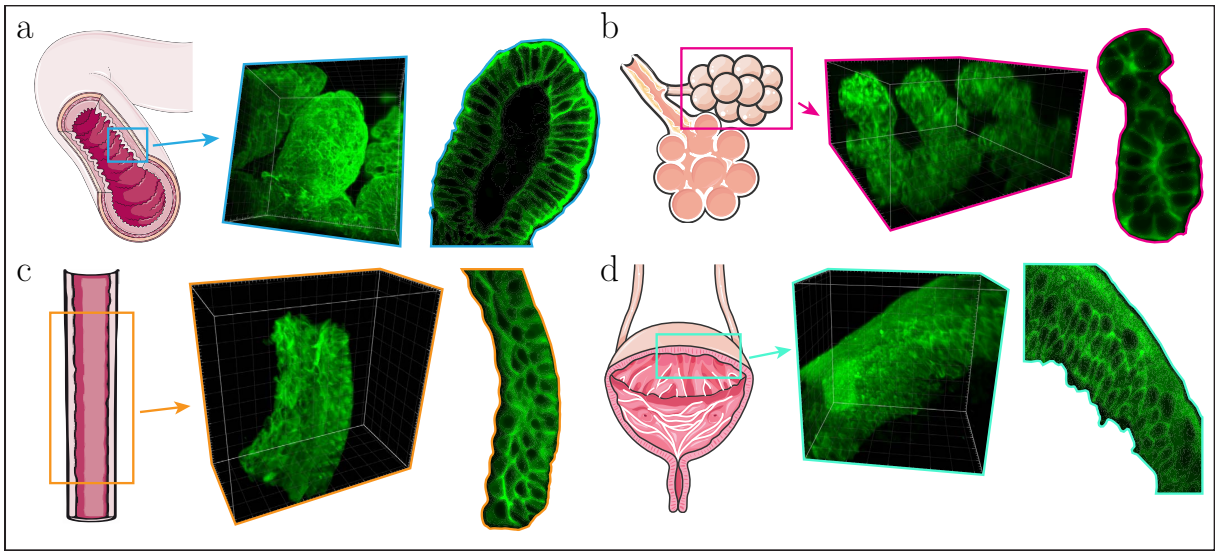

Figure 4: **Schematic illustration, 3D microscope image, and 2D projection of the epithelial tissue samples employed in our study. a,** Section of the small intestinal tube highlighting villi, finger-like projections that extrude from the intestine walls inside its lumen. **b,** Illustration of lung bronchioles. They extend from the lung tube constituting its terminal part. From a 2D slice, it is possible to observe that cells are organized in a single layer. **c,** The sample relative to the esophageal epithelium is constituted by a section of the tube. The inner layer of cells clearly shows a squamous morphology. **d,** The bladder epithelium outlines a stratified structure. The external surface is entirely covered by large umbrella cells. *Note:* For what concerns the lung tube epithelial sample please refer to Gómez et al. (2021).

For these reasons, given a set of imaged mouse epithelia at embryonic stage E17.5 acquired with the aforementioned protocol, we selected the following:

1. A sample of *Intestinal villus* epithelium,
2. A sample of *Lung bronchiole* epithelium,
3. A sample of *Esophagus* epithelium.

The intestinal villi are small, finger-like projections that extend into the lumen of the small intestine (Dudas, 2023). The covering epithelial tissue exhibits a simple columnar organization, i.e., epithelial cells are organized on a single layer and are rather elongated along the direction perpendicular to the villus surface (Fig. 4a). The bronchioles are the smaller terminal branches of the bronchial airways in the lower respiratory tract. Their walls are enveloped by a monolayer cuboidal epithelium (Tortora and Derrickson, 2017) (Fig. 4b). The esophagus is a fibro-muscular tube in the upper digestive system. The

internal surface of the esophagus is covered by mucosa, which is a stratified epithelial tissue of around three layers of cells (Patti et al., 1997) (Fig. 4c). The esophageal epithelium is classified as squamous, as the most internal layer of cells in contact with the lumen presents squamous cells.

To increase the variety of mouse epithelia in the study, we decided to include two more samples collected and processed for previous studies and publications. In particular, we utilized:

4. A sample of *Bladder* epithelium (*urothelium*) from an adult mouse,
5. A sample of *Lung tube* from embryonic (E12.5) mouse.

The bladder epithelial tissue is composed of three layers of cells (Fig. 4d), namely basal, intermediate, and umbrella cells. While basal and intermediate cells have a size in the range of 5-20 μm and exhibit a rather regular (cuboidal) shape, the superficial umbrella cells are circa 10 times larger and their shape depends on the degree of tissue stretch (Lewis, 2000). This leads to the classification of the urothelium as a transitional epithelial tissue. The lung tube is a tubular structure located in the bronchial airways (Fig. 4e). It is covered by pseudostratified cells that extrude towards the lumen of the tube. Specifically, we re-used the sample analyzed in our publication (Gómez et al., 2021).

High-quality renderings of the 3D light sheet microscope images and the associated 2D slices for all the tissues involved in the study are reported in Appendix B.

2 | 3D Cell Segmentation

As we discussed within the objective and aims, a key step to provide an accurate characterization of epithelia is to produce high-quality 3D cell segmentation. The ability to accurately identify the locations and shapes of individual cells provides us with valuable insights into the morphological and mechanical characteristics at both cell and tissue levels.

In this chapter, we first discuss existing methods to produce accurate 3D cell segmentation of biological tissues, highlighting the pros and cons. Following that, we present the pipeline we decided to employ for the segmentation of our light sheet microscope images of mouse epithelial tissues, explaining its main features and motivating our choice. Later on, we outline the improvements we made on the original version of the segmentation workflow that allowed us to get higher-quality results in a shorter time. Finally, we discuss the resulting segmented images, which will serve as inputs for the following steps of our analysis.

2.1. Related Works

Recently, biological microscopy experienced a period of remarkable advancements, due to the emergence of new technologies that enable the acquisition of high-resolution and high-depth fluorescence images. These improvements resulted in both the expansion of image size and scale and the growth in data quality. Certainly, the availability of larger datasets allows researchers to perform more thorough and precise studies of biological phenomena, capturing more detailed patterns and delving into more detailed analyses. At the same time, the extent and detail of these images made the employment of automated image-processing techniques essential to efficiently perform quantitative cell-based analysis on them. In fact, segmentation of large 3D images performed manually by common image analysis tools (Schindelin et al., 2012; Hanslovsky et al., 2018) requires many weeks of work and is no longer feasible at this scale.

Segmentation of an image volume into cell bodies can be rather straightforward when cells

are sufficiently separated from one other. In such cases, classical segmentation approaches based on edge detection algorithms (Canny, 1986) or the direct application of the watershed algorithm (Li et al., 2008; Dougherty, 1992) have proved to be a valid tool. However, in the majority of animal and plant tissues, cells are tightly packed together, with cell membranes in close contact with each other. Hence, in such cases standard methods fail to isolate single entities, providing unsatisfactory outcomes.

Over the last decade, Deep Learning (DL) models have established themselves as the state-of-the-art for computer vision tasks, such as image recognition and instance segmentation. Thanks to the availability of larger amounts of data and following the success of the U-Net model (Ronneberger et al., 2015) for the 2D cell segmentation task, artificial neural networks started to be heavily employed for image analysis tasks in bioscience. In particular, many researchers developed deep learning-driven methods to tackle the instance segmentation problem in biology and biomedicine (Moen et al., 2019). For instance, the algorithms illustrated in Dunn et al. (2019) and in Wu et al. (2023) focus on providing accurate 3D nuclei segmentation as an initial step for a subsequent cell segmentation task. Using similar methods, in the article from Isensee et al. (2021) the author proposes a self-configuring DL model for the segmentation of 2D and 3D biomedical images. Another example can be found in Wang et al. (2022), where the authors introduce an innovative U-Net-inspired architecture designed for the precise segmentation of densely packed tissues, demonstrating encouraging outcomes. However, the absence of a complete end-to-end segmentation pipeline (as the model is composed of separated blocks) and the absence of publicly available code hinder both the reproducibility of the findings and accessibility for external users.

In general, a common challenge with the majority of these studies is that they tend to be tailored for specific tasks and are trained using specific datasets. Hence, they sometimes struggle to generalize on unseen data or on slightly different image analysis tasks. Furthermore, a significant number of these models are highly complex, and the lack of open-source code or user-friendly interfaces reduces their reproducibility and adaptability for other researchers and users.

To overcome the problem of accessibility, in the past few years, some DL-driven software platforms for end-to-end 3D cell segmentation provided with user interfaces have been released (Carpenter et al., 2006; Sommer et al., 2011; Schindelin et al., 2012; de Reuille et al., 2015; McQuin et al., 2018; Berg et al., 2019; Stringer et al., 2021; Eschweiler et al., 2021). These tools offer various levels of customization and different kinds of interaction with the user. For instance, some of them provide dropdown menus from which the user can choose the algorithms to apply, while some others are capable of interpreting

user-drawn strokes on images as segmentation instructions. Anyway, all of these software tools offer a robust and easy-to-use 3D segmentation solution that performs effectively across a wide range of data and tasks. However, there are also some drawbacks associated with these platforms. On one hand, having a user interface is advantageous, especially for beginners, but it can restrict the extent to which segmentation algorithms can be customized. Furthermore, a common issue of low-code programs is an increased image processing time and limitations in scaling to handle larger inputs.

Given the necessity for a flexible tool, that offers improved accuracy and customization with respect to generalist platforms, while guaranteeing a fair degree of usability and reproducibility, for our study we decided to employ *Plantseg* (Wolny et al., 2020). Plantseg is a software for DL-driven cell instance segmentation in densely packed 3D volumetric images. It provides a wide array of publicly accessible, pre-trained convolutional neural networks (CNN) for the cell boundaries detection task. Each model has been trained on different plant tissue images (e.g., confocal vs. light sheet microscopes, different voxel sizes, cell shapes, and kinds of tissues). Moreover, the possibility of fine-tuning these neural networks on external datasets from a command line interface (CLI) brings additional versatility to the Plantseg software. Furthermore, the package also includes a user-friendly interactive interface in *napari* (Chiu and Clack, 2022) that allows instantaneous visualization of segmentation outcomes. In the next section, we will describe the architecture behind Plantseg and its capabilities.

2.2. Methods

2.2.1. Image Preprocessing

Deep learning models are potent feature extractors, that are able to recover patterns and provide excellent performance also from raw, noisy, unstructured data (Goodfellow et al., 2016). Nevertheless, image preprocessing techniques are an essential stage for most DL pipelines as they have been proven to sharpen model efficiency. As an example, standardization of inputs can enhance the numerical stability of the networks leading to faster model training. Similarly, image enhancement by means of noise removal and masking of superfluous information allows the model to focus on more meaningful content and obtain better accuracy overall. In particular, noise removal is a crucial step for fluorescence microscopy images. In fact, background fluorescence is very frequent in these images, since staining often targets unwanted structures and features of the samples under analysis.

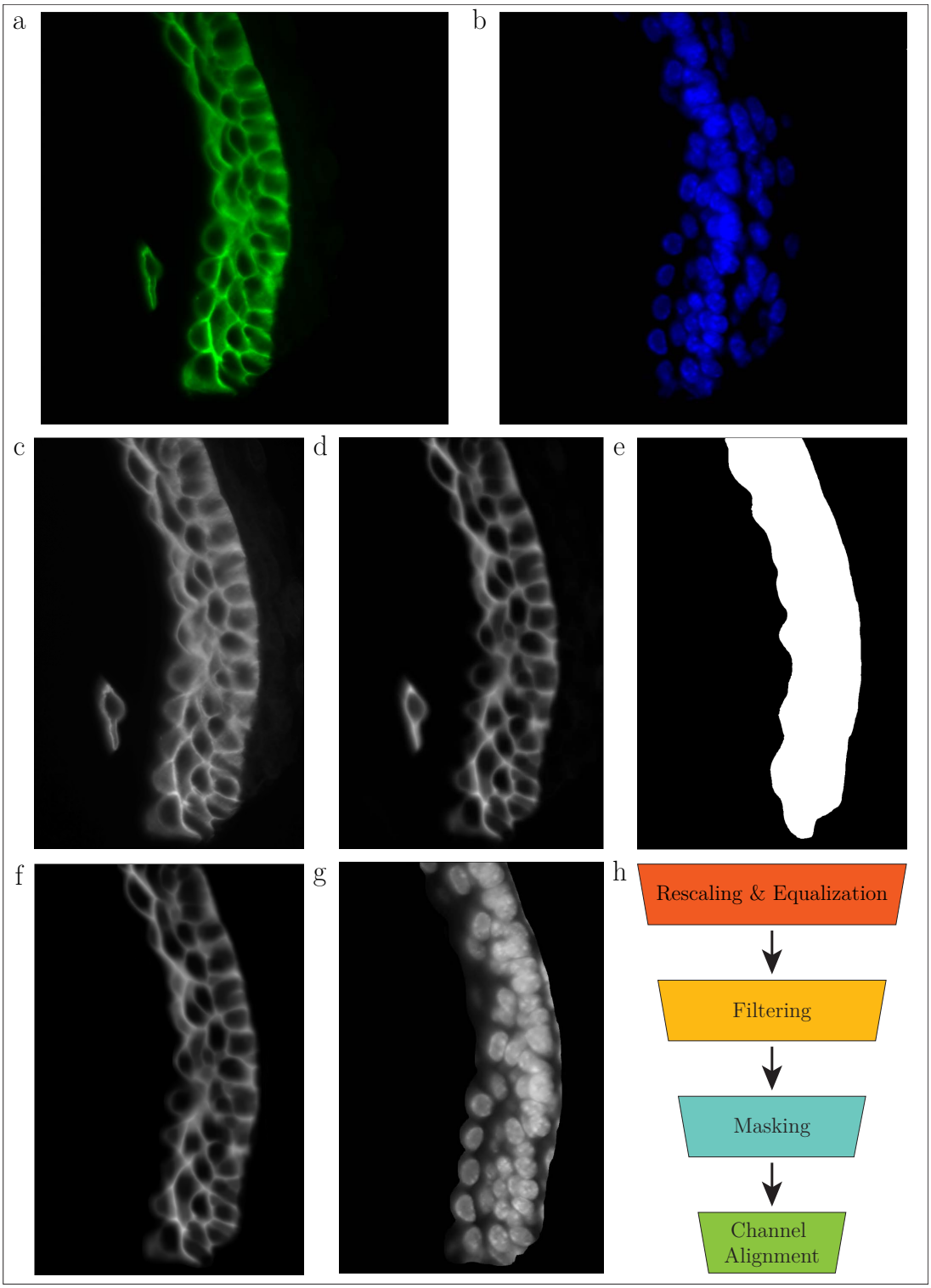

Figure 5: **Visualization of the Image Preprocessing steps on the Esophagus sample.** **a, b,** A 2D slice of the cell membrane (GFP) and nuclei (DAPI) from the esophagus epithelium as taken at the light sheet microscope. **c,** The esophagus 2D slice after voxel rescaling and intensity equalization. **d,** The esophagus 2D slice after applying noise filters on Fiji. **e,** The thresholded binary background mask obtained using micro-SAM. **f, g,** The final esophagus 2D slices of both cell membrane and nuclei after masking and channel alignment. **h,** A schema of the image preprocessing workflow.

For these reasons, in our project, we decided to complement the Plantseg model with a tailored image preprocessing pipeline. It is worth mentioning that some steps of this pipeline were inherited from previous studies conducted internally in the lab. However, to address the absence of a unified and well-structured workflow and the overall need for optimization both in terms of time and performance, we were prompted to develop an improved end-to-end image preprocessing pipeline. This pipeline leverages *Python* scripting and libraries (such as *Numpy* (Harris et al., 2020), *Scikit-image* (van der Walt et al., 2014), *SciPy* (Virtanen et al., 2020), and *napari* (Chiu and Clack, 2022)), as well as specialized image analysis programs (such as Fiji (Schindelin et al., 2012) and (Berg et al., 2019)). The same procedures are applied both to the cell boundaries (GFP channel) and the nuclei (DAPI channel) files. The workflow is visually described in Figure 5 and it is composed of the following ordered steps:

- First we rescale the raw microscope images in order to get the same voxel size across all the items in the input dataset. Namely, we selected a voxel size of [0.1625, 0.1625, 0.25]μm. Indeed, the performance of CNNs is sensitive to changes in voxel size (and hence object size) due to the fact that convolutional filters learn to detect features at a specific resolution. Afterward, we perform histogram equalization of the intensity to enhance the sharpness of the fluorescence signal.

- Then we resort to *Fiji* software tool (Schindelin et al., 2012) to perform image cropping, filtering, and, if necessary, downsizing. Cropping plays a crucial role in the preprocessing pipeline by enabling the removal of uninteresting or noisy areas from images. This not only helps preserve higher quality and more meaningful features but also reduces significantly the file size. The latter aspect is particularly important as we deal with high-resolution and large-scale 3D images.
  Image filtering consists of an ensemble of techniques to effectively remove a substantial amount of noise from the original image. Its primary purpose is to enhance the contrast between cell signals and the background, resulting in crisper and more distinct boundaries. Specifically, in our case, we employed the following processes: background subtraction via *rolling ball algorithm* (Sternberg, 1983), gaussian blurring to filter out high-frequency noise components, morphological closing to remove small noisy structures, and *white top-hat transform* to homogenize the background and highlight bright features (Serra, 1983; Serra and Vincent, 1992).
  Finally, the size of a 3D light sheet microscope image can easily exceed ten gigabytes. Handling these large images can impact the performance of image analysis software, leading to unsustainable computation time. As a result, there were instances where we had to reduce the size of the image files by doubling the voxel

size in the X-Y dimensions. While downsizing the voxels does diminish the level of detail in the images, the original resolution of the images was sufficiently high to prevent a significant deterioration in segmentation outcomes. Anyway, as we progressed in our work, we iteratively refined the preprocessing pipeline by replacing less computationally efficient tools. Consequently, the need for downsizing became obsolete.

- In addition, we decided to include in the pipeline a background masking step, in order to completely erase uninteresting and noisy features from the image. The goal is to produce a probability mask that marks all the non-informative structures as background. Then, the mask is thresholded (at a background probability of 0.9) and it is superimposed on the filtered image from the previous step. In the original image preprocessing pipeline, the probability mask was computed using the interactive 3D segmentation tool provided by *Ilastik* (Berg et al., 2019). Even though the mask produced by this software was of good quality, we faced some computational issues while working with images at full resolution. In general, the masking process with Ilastik could take up to a few hours even on a machine endowed with a good GPU, and require extensive prompting from the user. As a consequence, in a later iteration of the image preprocessing pipeline, we decided to replace Ilastik with more efficient masking performed by the *Segment Anything Model for Microscopy* (micro-SAM) software (Archit et al., 2023). micro-SAM is built on top of the *Segment Anything Model* (SAM) (Kirillov et al., 2023), the state-of-the-art tool for user-prompted 2D instance segmentation. Specifically, micro-SAM is fine-tuned on microscopy data and extends SAM functionalities by enabling interactive 3D segmentation. Overall, micro-SAM provides extremely accurate background predictions and it only requires a few minutes to run on large, high-resolution 3D images. It is worth mentioning that we also tried to employ micro-SAM to perform 3D cell segmentation. However, the tool still struggles to identify large amounts of instances, such as the tightly packed cells in an epithelial tissue.

- As we pointed out in chapter 1, our light sheet microscope images include two channels, respectively obtained from GFP staining (highlighting cell membranes), and DAPI staining (highlighting the nuclei). Since the two fluorescence signals present different wavelengths, they are subject to the dispersion phenomenon. As a consequence, when the different channels are combined in a single image, it may happen that the same features do not properly overlap (e.g., nuclei are shifted with respect to the relative cell boundaries) (Scalettar et al., 1996). Therefore, a necessary step in the image preprocessing pipeline is the alignment of fluorescence

signals. We implemented this procedure in a *Jupyter Notebook*, in which the user can interactively shift the channels using the *napari* viewer to monitor the outcome.

2.2.2. The Plantseg pipeline

The preprocessed 3D microscope images of epithelial tissues are then fed to the Plantseg pipeline for automated cell segmentation. The Plantseg workflow comprises the following blocks:

1. A U-Net-like convolutional neural network (CNN) to predict cell boundaries;
2. *Graph Partitioning* (GP) to perform segmentation based on the cell boundary predictions.

The combination of CNN boundary prediction with GP postprocessing is quite common for the 3D cell segmentation task of tightly packed tissues (Eschweiler et al., 2018; Funke et al., 2019a; Greenwald et al., 2022). In fact, CNN models operate on a per-pixel basis and can easily introduce biologically implausible segmentation, such as unconnected regions or far-fetched cell shapes. Conversely, GP algorithms are inspired by physical principles (e.g., the watershed phenomenon), and hence guarantee consistent and reliable results even from a biological perspective.

Convolutional Neural Network for boundary prediction

For what concerns the CNN for cell boundary predictions, Plantseg implements two different kinds of architectures: a standard *3D U-Net* (3D-Unet) (Özgün Çiçek et al., 2016) and a *Residual 3D U-Net* (3D-ResUnet) (Lee et al., 2017). For both architectures, pretrained weights are provided for different voxel sizes, types of microscope (light sheet vs. confocal), and loss functions. Since the use of the residual version did not bring any relevant advantage in terms of segmentation performance, we decided to employ the standard 3D-Unet for our analyses.

A schematic illustration of the 3D-Unet model is reported in Fig.6b. The network consists of a contracting encoder that extracts deep features from the input images (analysis path) and an expanding decoder that produces a full-resolution segmentation (synthesis path). In the Plantseg implementation, both the encoder and the decoder include four resolution levels. In the encoder, each level comprises two blocks composed of a Group Normalization (GN) layer followed by a 3D convolution layer and a ReLU activation. Downsizing to lower resolution is performed via a Max-Pooling layer, placed at the end of each level. On the contrary, the decoder levels are made by an Upconvolution layer, followed by Group

Normalization, a stack of two 3D convolution layers, and a ReLU activation on top. In the last upsampling layer, a 1x1x1 convolution reduces the number of output channels to 2 for the binary segmentation task. Contrary to the original U-Net implementation, the proposed architecture employs *same* convolutions, and the upscaling (resp. downscaling) only happens in the max-pooling (resp. upconvolution) layers. Finally, shortcut connections are placed between downscaling and upscaling paths at the same resolution, to provide useful high-resolution features in the output.

For memory-related reasons, the available pre-trained models were trained on a single patch of size $[170 \times 170 \times 80]$ (in voxels). Therefore, Group Normalization is chosen over Batch Normalization, as the latter would essentially function like Instance Normalization, leading to less reliable batch statistics.

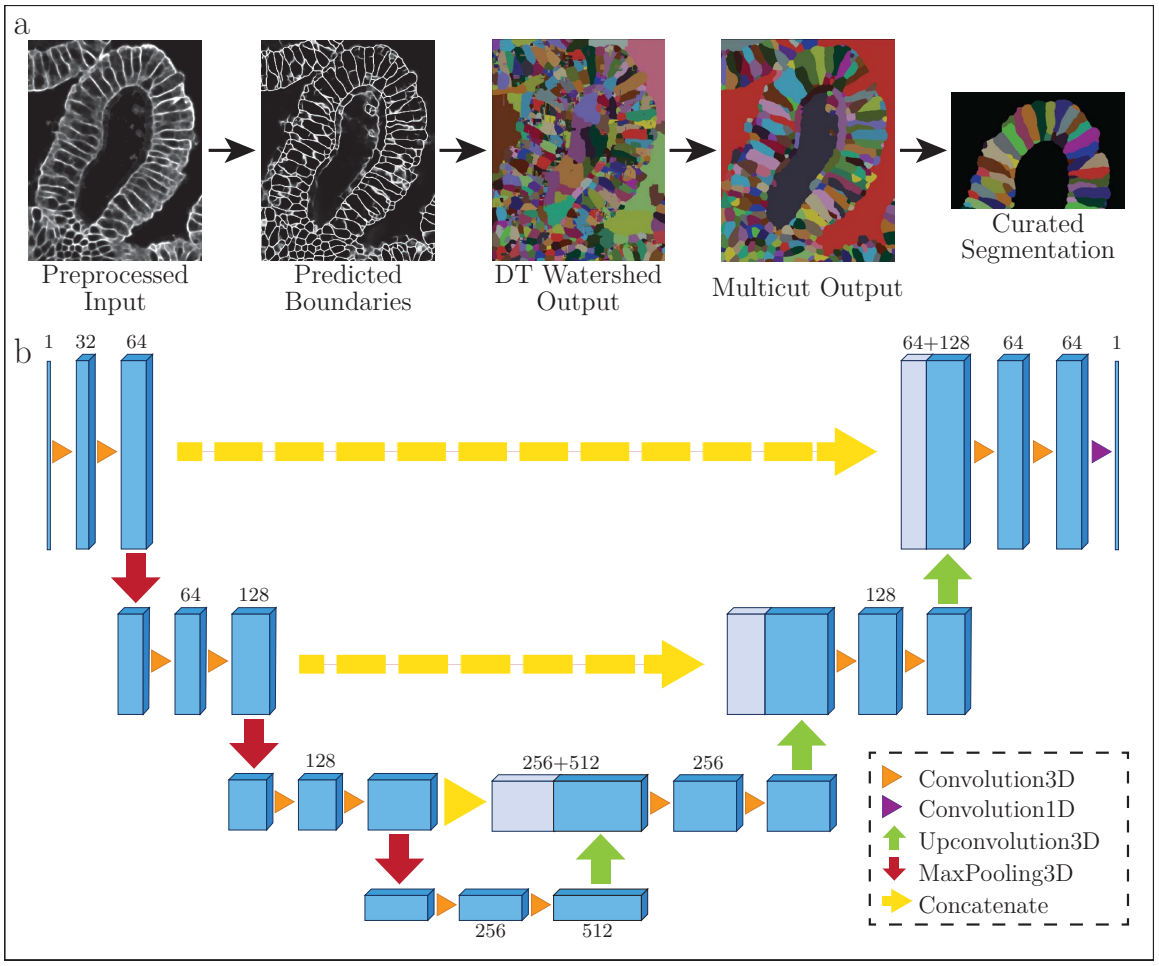

Figure 6: **The 3D cell segmentation workflow. a,** An example of a 3D cell segmentation workflow using Plantseg. Given the preprocessed input image, first, a CNN predicts the cell boundaries, and then GP algorithms like DT Watershed and Multicut are applied to produce the segmentation. Finally, manual curation is performed to correct segmentation errors. **b,** The architecture of the 3D-Unet implemented in Plantseg.

The loss function selected for training the released networks is given by the sum of binary cross-entropy and Dice loss:

$$\mathcal{L} = \mathcal{L}_{BCE} + \mathcal{L}_{Dice}$$

Lastly, at inference time *mirror padding* is applied on the input image to improve the prediction accuracy on the edges. Moreover, in order to reduce the *checkerboard effect*, the volume is parsed with patches overlapping for 50% of their volume. The overlapping probability maps are then averaged out to produce the final prediction.

Graph Partitioning for voxel clustering

Given the predicted cell boundaries, a region adjacency graph $G(V, E)$ can be constructed from the image voxels and cell segmentation can be reduced to a graph partitioning problem. Specifically, each voxel is a vertex V of the graph and edges E connect adjacent voxels. Edges are weighted by scalars $w \in \mathbb{R}^+$ derived from the boundary probability map.

Since solving the partitioning problem at voxel-level would be computationally prohibitive for volumes of biologically relevant size, in the first place single voxels are clustered into so-called *supervoxels* by running the Distance-Transform (DT) Watershed algorithm (Roerdink and Meijster, 2000). This algorithm consists of the following steps:

1. Thresholding of the boundary probability map is performed at probability level d in order to obtain a binary image.
2. The distance transform is computed on the binary boundary image and Gaussian smoothing is applied to the result.
3. A source is assigned to every local minimum in the resulting distance transform map.
4. From each source a given label is propagated in every direction until it encounters a different one. The formed contact surface is the so-called watershed and represents the boundary between different segmented instances.

At this point, a new region adjacency graph is then built on the supervoxels obtained from the first DT Watershed iteration. The latter graph is then partitioned into sections to deliver the actual cell segmentation. For this task, in addition to the already mentioned DT Watershed approach, Plantseg provides three alternative, more sophisticated partitioning strategies. Namely:

- The *Multicut* algorithm (Kappes et al., 2011),

- The *Mutex Watershed* algorithm (Wolf et al., 2018),
- The *GASP* algorithm (Bailoni et al., 2022),

Among the four aforementioned methods, we empirically assessed that Multicut provides the best outcomes on the samples involved in our study.

Nuclei Segmentation

In addition to the functionalities mentioned above, Plantseg also offers tools for volumetric segmentation of cell nuclei. In this case, a pre-trained 3D U-Net enables the prediction of the binary nuclei mask. The final segmentation is subsequently obtained by thresholding the probability maps and applying a post-processing algorithm to merge connected components.

Although segmented nuclei are not directly employed in our analysis of epithelial tissues, the information about nuclei position, combined with the knowledge that, in healthy tissues, each cell has a single nucleus, can be used to perform more informed cell segmentation. For instance, *Lifted Multicut* (Horňáková et al., 2017; Pape et al., 2019) is an enhanced version of the Multicut algorithm that leverages additional domain-specific information, like the nuclei location in the case of cell segmentation, to provide a more efficient solution for the graph partitioning problem. Plantseg provides an optimized implementation of the Lifted Multicut algorithm, that exploits nuclei position to separate incorrectly merged cells. Moreover, in the paper the author shows that Lifted Multicut outperforms all the other graph partitioning algorithms when applied to their data.

Given the availability of a nuclei channel (DAPI) in our image dataset and the advantage of using the Lifted Multicut approach, for all the epithelial tissues in our analysis, we tried to segment the nuclei employing the Plantseg tools. However, this procedure failed for some of the samples. Indeed, in both the bronchiole and the esophagus samples, nuclei are so close to one another that their signal overlaps and they are hardly distinguishable even to the human eye (Fig.7a). Therefore, for this purpose, a more efficient algorithm should be designed.

2.2.3. Manual Curation

As we discussed, our segmentation pipeline is capable of automatic, high-quality 3D cell segmentation. Nevertheless, in order to run detailed morphological and mechanical analysis on both the cell shape and the overall tissue organization, we need to make sure that in the automatic segmentation mask, there are no major segmentation errors. The most

common mistakes are over-segmentation (i.e., two or more cells are wrongly merged in a single one) and under-segmentation (i.e., two or more "pieces" belonging to the same cell are associated with different labels). Additionally, the pipeline can also introduce segmentation artifacts, for example, misinterpreting noise in the input image as an actual feature in the sample.

For the reasons above, a post-processing step on the initial segmentation mask is essential to enhance the quality of our analysis. We call this post-processing of the segmented image *curation*. The curation of the segmented image is performed manually cell-by-cell, through a *napari* plug-in called *Morphometrics* [1]. Morphometrics curation plug-in offers a variety of tools to correct segmentation errors and remove artifacts, such as merging labels, deleting labels, and painting and erasing labels both on 2D slices and on 3D volumes.

With the current state of the 3D cell segmentation software, manual curation is an indispensable process to get high-quality outcomes for subsequent analyses. That said, this image post-processing step presents two main drawbacks, that both limit the segmentation quality and the time required to obtain results.
On the one hand, correcting segmentation errors by hand using Morphometrics tools is often a complex task and can potentially introduce noise and inaccuracies into the final results. Painting 3D boundaries between cells with voxel-level precision is a highly challenging endeavor. Consequently, the resulting segmentation may exhibit irregular cell borders whenever manual painting is required.
On the other hand, manual curation is an extremely time-consuming operation, as it requires the user to manually check and, if necessary, clean one cell at a time. To provide the reader with an example, the post-processing of a volume of roughly 200 cells could take up to 4 days of full-time work for a user who is already experienced in using the curation tools.

Therefore, for the moment, the manual curation process is by far the biggest bottleneck in our workflow. Clearly, the better the segmentation provided automatically by the pipeline, the lesser the time required to manually correct it. We leave the discussion about possible improvements in the segmentation pipeline in the final section of this chapter.

2.2.4. 3D U-Net Fine-tuning

When we introduced Plantseg we mentioned that all the available models were trained on microscope images of *plant* tissues. Although the author claims that the Plantseg pipeline is capable of generalizing to epithelial animal samples (e.g., the *Drosophila* wing

[1] Available at https://github.com/morphometrics/morphometrics

disc dataset (Funke et al., 2019b)), we decided to fine-tune a pre-trained CNN on an internal dataset of segmented animal epithelia to enhance the model performance.
Specifically, this dataset includes:

- Four segmented volumes cropped from a healthy *mouse bladder* sample, that was employed in a previous bladder cancer study.
- The *intestine villus* sample mentioned in section 1.3, which we segmented before fine-tuning the model.

One of the bladder volumes was held out for validation, while all the other segmented images were used for training.
The use of a such limited dataset should not surprise the reader. In fact, since biological images usually comprise repetitive structures and close slices of the volumes frequently bring redundant information, just a few annotated samples are enough to obtain reasonably accurate outcomes (Özgün Çiçek et al., 2016).

We performed fine-tuning on the 3D-UNet trained on light-sheet images of Arabidopsis lateral root at half-resolution[2] (voxel size of $[0.325 \times 0.325 \times 0.25]\mu m$). The choice of this resolution is linked to the fact that, due to the aforementioned memory and computation constraints in the segmentation pipeline, we downscaled the intestine villus image to half-resolution.
We ran multiple trainings with different combinations of *learning rate* (LR) (namely, 4e-6, 2e-5, 1e-4) and *batch size* (BS) (namely, 1, 3). We evaluated the different configurations by computing both the *Precision-Recall curve* (and the associated *Area Under the Curve* (AUC)) and the *Dice Coefficient*, which is computed as:

$$Dice = \frac{2\sum_{n=1}^{N} p_i g_i}{\sum_{n=1}^{N} p_i + \sum_{n=1}^{N} g_i}$$

where p_i and g_i are, respectively, the predicted and ground-truth binary voxel values and N is the total number of voxels in the image.

A comparison of the Precision-Recall curves for the different fine-tuning routines is reported in Fig.7b and the results are presented in 1. Despite the metrics for [LR=2e-5; BS=3], [LR=1e-4; BS=1], and [LR=1e-4; BS=3] are very similar, with $AUC \approx 0.71$ and $Dice \approx 0.67$, the configuration with [LR=1e-4; BS=1] attains the highest AUC score and the second highest $Dice$ score. Therefore, we selected the fine-tuned weights obtained with those parameters as the coefficients for subsequent 3D-Unet predictions.

[2]Pre-trained weights available at this https://zenodo.org/record/7774122/files/unet3d-lateral-root-lightsheet-ds2x.pytorch

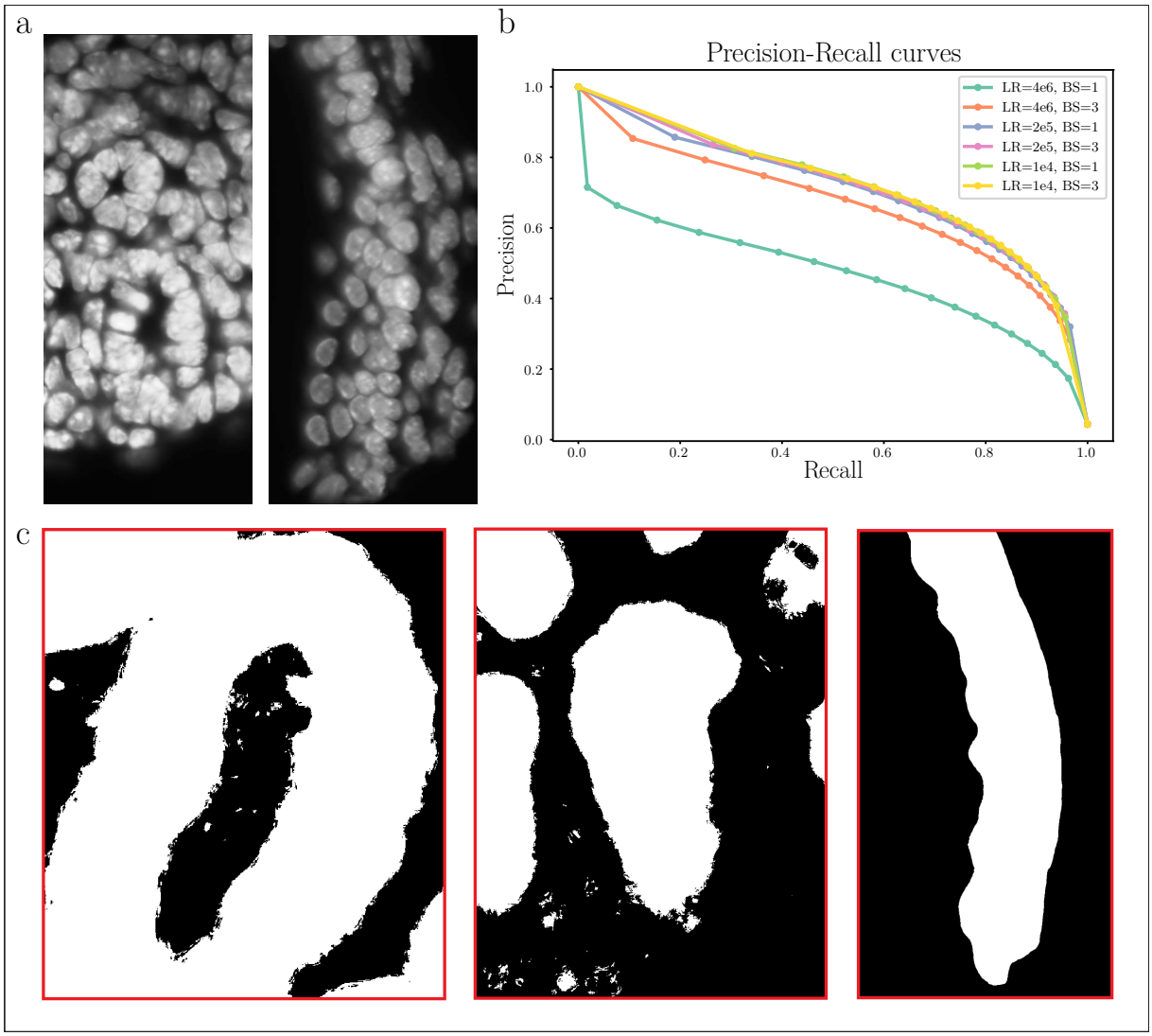

Figure 7: **The 3D cell segmentation workflow. a,** Bronchiole (left) and esophagus (right) epithelia show tightly packed nuclei. The presence of overlapping signals makes nuclei segmentation rather challenging. **b,** The Precision-Recall curves for different learning rate and batch size configurations used for 3D-Unet fine-tuning. **c,** Binary background masks for intestine villus (left), bronchiole (center), and esophagus (right) epithelia. The mask made with micro-SAM for the esophagus presents sharper and clearer boundaries with respect to the others made with Ilastik.

To assess the goodness of the fine-tuning, we compared the predictions of the original CNN versus the fine-tuned CNN on the boundary prediction task. Specifically, we used as test data the segmented bladder image that was kept out for validation and the bronchiole image that was segmented in our study. Once again, we resorted to the *Dice* score as our evaluation metric. In particular, we computed both the per-image *Dice* coefficient

and the global *Dice* coefficient. The reason for this choice is that the network was fine-tuned on other bladder samples, and hence global metric could have been biased by the score obtained for the bladder. Results reported in Table 1, show that, after fine-tuning, for the bronchiole epithelium the *Dice* score increases by almost 20%. As expected, for the bladder epithelium the score increases dramatically, almost doubling. Overall, the aggregated metric grows by more than 50%.

Table 1: **Fine-tuning results.** Evaluation of fine-tuning routines for different parameter configurations.

Learning Rate	Batch Size	AUC	Dice score
4×10^{-6}	1	0.470	0.513
4×10^{-6}	3	0.661	0.641
2×10^{-5}	1	0.702	0.666
2×10^{-5}	3	0.708	0.670
1×10^{-4}	1	**0.711**	0.672
1×10^{-4}	3	0.706	**0.676**

Table 2: **Dice Score results.** Comparison of Dice score for the cell boundaries prediction task before and after CNN fine-tuning.

Sample	Dice original	Dice fine-tuned
Bladder	0.487	0.585
Bronchiole	0.347	0.676
Global	0.413	0.631

2.3. Results

Following the 3D cell segmentation workflow above, we segmented the cells of three samples of mouse epithelial tissues, namely:

- The intestine villus epithelium,
- The lung bronchiole epithelium,
- The esophagus epithelium.

It is important to remark that, since the segmentation of these three tissues happened at different times, we employed slightly different versions of the pipeline for each of them.

Namely, for the intestine tissue we used the original pre-trained 3D-UNet provided in Plantseg to perform boundary prediction, while for the bronchiole and esophagus epithelia, we employed a model fine-tuned on some mouse epithelia. As we discussed in section 2.2.4, the use of a fine-tuned CNN allows us to obtain way more accurate boundary predictions, leading to higher quality in the final automatic 3D cell segmentation. This fact consistently reduces the manual curation required.
An additional difference resides in the fact that, for the intestine villus and the lung bronchiole samples background masking was performed with Ilastik, whereas for the esophagus we exploited micro-SAM. The use of micro-SAM offers three main advantages:

1. It reduces the computation time to produce a background mask from a few hours to a few minutes.
2. It increases the sharpness of the contours of the mask itself (see Fig.7c).
3. It requires fewer inputs by the user, as it handles prompts more efficiently.

We would like to note that, before curation, we extracted smaller sections of the automatically segmented tissues to expedite the manual review process of individual cells.

Furthermore, to enrich the variety of cell shapes and tissue organizations in our study, we also considered previously segmented samples of mouse epithelial tissues, specifically:

- The bladder epithelium,
- The lung bronchiole epithelium,

The urothelium sample was segmented through more or less the same pipeline as us. On the contrary, the whole segmentation of the lung tube epithelium was carried out by manually labeling the image slice by slice using a computer-aided annotator.

Table 3: **Segmentation outcomes.** Summary of the segmented 3D microscope images of the epithelial tissues included in the study.

Tissue	Voxel size (μm)	Nr. cells	Pipeline	Time req.
Intestine	$[0.325 \times 0.325 \times 0.25]$	322	Plantseg (original)	$\sim$2 weeks
Bronchiole	$[0.1625 \times 0.1625 \times 0.25]$	180	Plantseg (fine-tuned)	$\sim$1 week
Esophagus	$[0.1625 \times 0.1625 \times 0.25]$	129	Plantseg (fine-tuned)	$<$1 week
*Bladder**	$[0.21 \times 0.21 \times 0.39]$	638	Plantseg (original)	$\sim$3 weeks
*Lung Tube**	$[0.1 \times 0.1 \times 0.1]$	147	Manual annotation	$>$1 month

* *Already segmented for other studies and publications.*

Table 3 reports, for each tissue involved in the study, the voxel size, the number of segmented cells, the employed segmentation pipeline, and the time required to segment (an estimate, in the case it was not directly done by us).
For what concerns the intestine, bronchiole, and esophagus samples, the iterative refinement of the cell segmentation pipeline provided a twofold enhancement. In fact, on the one hand, it enabled us to progressively reduce the time spent to obtain the outcomes. For instance, the approximate time required for the segmentation of each epithelial cell, went from about 15 minutes in the case of the intestine, to less than 10 minutes in the case of the esophagus. On the other hand, the enhanced workflow version enabled us to handle larger images, eliminating the need for voxel downscaling.

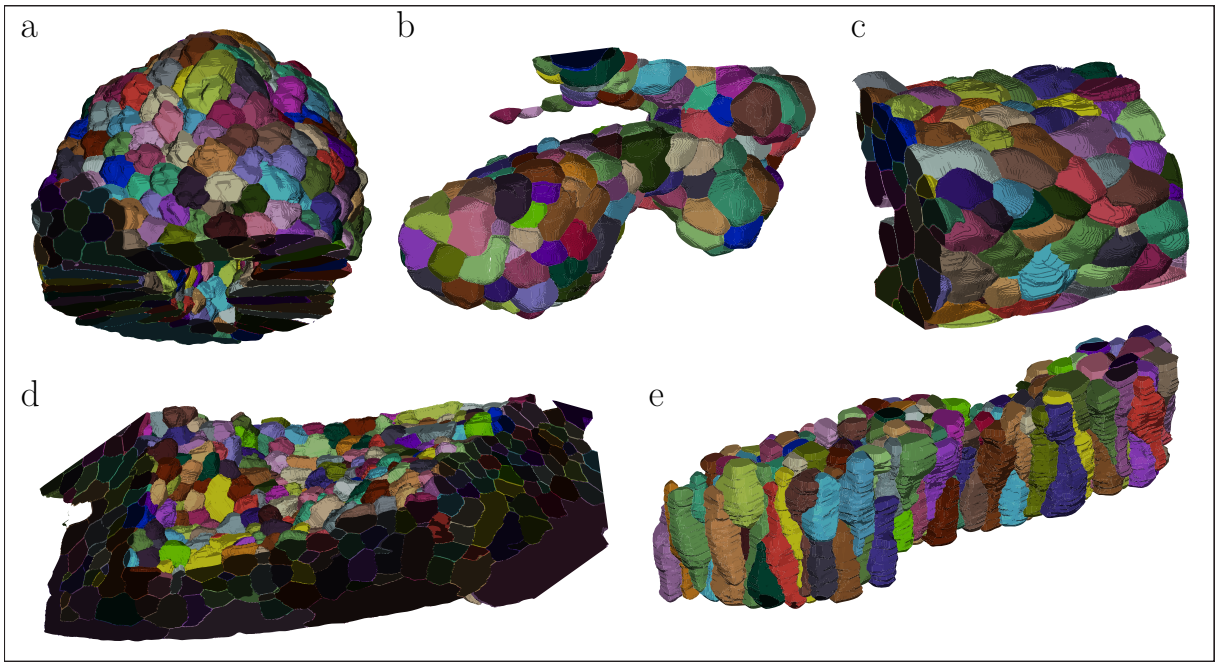

Figure 8: **A rendering of the final segmented samples of epithelial tissues. a,** Intestine villus epithelium. **b,** Bronchiole epithelium. **c,** Esophagus epithelium. **d,** Bladder epithelium. **e,** Lung tube epithelium.

2.4. Discussion

In this chapter, we illustrated a pipeline for 3D cell segmentation of light sheet microscope images. Specifically, we described the components and their functions, from image preprocessing, through automatic segmentation, and, eventually, to manual curation. Moreover, we presented the results of our epithelial tissue sample and highlighted the strengths and drawbacks of the current workflow.
It turned out that the most evident bottleneck of the actual implementation was the

extensive requirement of manual curation to overcome the limits and errors of the automated processes. Curation is not only an extremely time-consuming operation, but it is also a source of segmentation artifacts, due to the difficulty of painting cell labels in 3D by hand.

Future Developments

Intuitively, a way to reduce the manual work in the segmentation process is to improve the performance of the automated part of the existing pipeline. Out of the various potential areas for improvement, we are convinced that enhancing the deep learning model employed for boundary prediction holds the most promise. Indeed, as noted in the preceding section, more accurate predicted boundaries overall lead to a more precise segmentation by the graph partitioning algorithms.

In particular, for future works, we propose the following modifications:

- Following the success in Nautural Language Processing, *transformer*-based architectures like the *Vision Transformer* (ViT) (Dosovitskiy et al., 2020) have become the state-of-the-art also for computer vision tasks. Specifically, U-Net-inspired models like the *U-Net Transformer* (UNETR) (Hatamizadeh et al., 2021) have established themselves as the current cutting-edge solutions for the instance segmentation task in biomedical and biological settings. Therefore, we propose to replace the current 3D-Unet with the UNETR to obtain an improved 3D cell segmentation pipeline.

- With the recent advancements in microscopy, the availability of 3D biological images increased dramatically. However, the scarcity of annotated datasets still limits the possibility of training deep learning models to analyze those images. *Self-supervised learning* enables the superation of these limitations by learning meaningful representations of the input data even in the case the latter is not labeled (Pathak et al., 2016; Jaiswal et al., 2020). Because of the aforementioned annotations scarcity, self-supervised learning found several applications in the field of biomedical and biological imaging (Chen et al., 2019; Krishnan et al., 2022; Tang et al., 2022).
  Given the abundance of light sheet microscope images of epithelia in our databases, for the future improvement of the segmentation model, we propose *Contrastive learning* as a method to exploit the informative content from unlabeled images (Chen et al., 2020). Given an input image, the rationale behind this approach is to create an array of perturbed images (e.g., by applying blurring, elastic deformations, patch removal, ...) and train an encoder-decoder architecture by evaluating its ability to reconstruct the input image from the perturbed ones.

Lastly, we plan to implement a better algorithm for the volumetric segmentation of cell nuclei. In fact, the availability of precise information about nuclei position would enable us to perform the graph partitioning step using the Lifted Multicut algorithm, which has been shown to outperform the other methods when fed with properly segmented nuclei.

3 | Morphological analysis of epithelial tissues

In the Introduction we presented epithelial tissues with their functions and classification. We mentioned that the driving factors that give rise to such a wide array of diverse tissue configurations and cell shapes are still unclear, and, hence, a potentially fruitful research subject. In this context, we formulated the main objective for this study, namely the in-depth characterization of 3D epithelia, by analyzing and comparing different tissue organizations and structures using advanced morphometrics methods.
In this chapter, we first present *EpiStats*, an image analysis pipeline that we implemented with the goal of gathering morphological statistics from epithelial tissues. Specifically, EpiStats is an efficient, user-friendly Python library that enables the extraction of meaningful features concerning cell morphology and tissue organization from segmented epithelia. Then, we delve into the analysis of the morphological statistics with a twofold approach. On the one hand, we showcase a detailed comparison of epithelial features across the different samples involved in the study aimed at providing a detailed description of key similarities and differences among epithelia. On the other hand, we use the collected morphological statistics to assess whether the theoretical laws outlined in the introductory sections hold in the case of our samples.

3.1. Related Works

The scientific literature in biology comprises a plethora of studies concerning the morphology of epithelial tissues. As a matter of fact, there are several articles that leverage the investigation of epithelial properties in order to unveil the mechanism that regulates the morphogenesis of biological tissues (Odell et al., 1981; Gumbiner, 1992; Gibson and Gibson, 2009; Miklius and Hilgenfeldt, 2011; Wang et al., 2012; Guirao et al., 2015; Miao and Blankenship, 2020). However, as far as we are concerned, there is a general lack of studies in which the morphology of epithelial samples from different organs are compared in order to draw insights about their different organizations and development.

Furthermore, we also noticed the scarcity of accessible, open-source software that directly aims at the morphological analysis of biological images. With a few exceptions (Heller et al., 2016), most of the available image analysis libraries (van der Walt et al., 2014; Virtanen et al., 2020; Haase and Rigaud, 2020), offer yet extremely powerful, but general-purpose tools, that are not explicitly designed for the study of biological patterns. Additionally, all of these software do not provide an easy-to-use, end-to-end pipeline. As a consequence, the user would need to manually select and combine some of their functionalities to obtain the desired results.

With the goal of trying to overcome these difficulties and offering a complete and efficient tool for morphological image analysis, in our study, we developed EpiStats, a Python library for the comprehensive analysis of epithelial tissues. Thanks to this software we were able to perform refined morphological analyses of different epithelia with the aim of comparing their cell organization and tissue structure

3.2. Methods

As already mentioned above, in order to compute morphological statistics from segmented 3D images of epithelial tissues we developed a Python library called EpiStats.
EpiStats is composed of three main blocks:

1. A set of image processing tools, in order to improve the quality of segmented cell volumes by reducing the extent of segmentation errors and artifacts.
2. A list of efficient image analysis algorithms to compute and analyze cell statistics from labeled images and triangular meshes of segmented cells.
3. A mesh refinement pipeline to produce regular cell meshes for numerical simulation.

In this chapter, we will focus on the first two functionalities, since they are directly utilized for the computation of morphological cell statistics from segmented tissues (Fig. 9a). On the contrary, we leave the discussion of the mesh refinement workflow in the following chapter which focuses on numerical simulations.

With the aim of creating a user-friendly tool, we developed EpiStats as a comprehensive pipeline for gathering cell statistics. With this approach, users only need to provide 3D segmented images and define a small set of customization parameters, such as the voxel size of the input sample and the name of the statistics they want to compute. By doing so, they can effortlessly obtain a dataset containing morphological features for each cell of the analyzed tissues. Specifically, EpiStats supports the computation of the following

statistics: *surface area*, *volume*, *elongation*, *neighbors indices* and *counts*, *contact area*, and *cell area* and *neighbors on 2D slices* of the segmented volume. Lastly, the library offers a range of functions for generating analytical insights and visualizations regarding cell shape and tissue organization.

EpiStats is an open-source project developed in Python and the code is available on GitHub at `https://github.com/AntanasMurelis/EpiStats`.

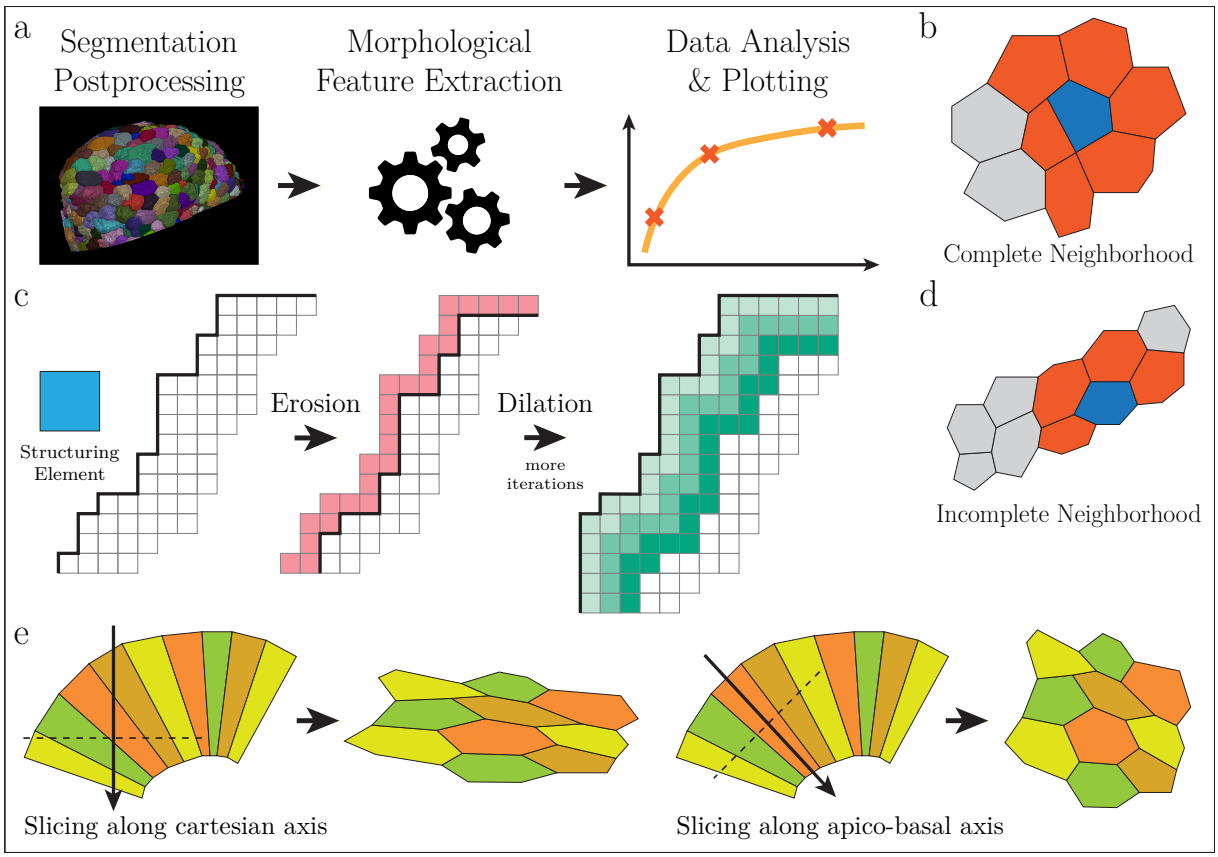

Figure 9: **The EpiStats pipeline for morphological cell statistics computation and analysis.** **a,** EpiStats software is composed by three main parts: *Segmentation Postprocessing*, *Morphological Feature Extraction*, and *Data Analysis and Plotting*. **b,** A cell (in blue) is said to have a *complete neighborhood* if it is entirely surrounded by other cells (in red), i.e., any of its lateral surfaces are in contact with the external side of the tissue. **c,** Schematic representation of morphological erosion and dilation. **d,** A cell (in blue) is said to have an *incomplete neighborhood* if at least one of its lateral surfaces is in direct contact with the exterior. **e,** Sampling 2D slices from a segmented cell volume along one of the cartesian axes may lead to distorted planar projections, and, hence, to unreliable resulting data. A more accurate approach is to slice the volume along the directions determined by each cell's apico-basal axis.

Regarding the computation of cells' morphological statistics, we implemented part of our algorithms by leveraging existing tools from *Scikit-Image* (van der Walt et al., 2014), *SciPy ndimage* package (Virtanen et al., 2020), *Numpy* (Harris et al., 2020), and *Morphosamplers*[1] Python packages. For the generation, manipulation, and refinement of 3D triangular meshes, we mostly relied on the following libraries: *Trimesh* (Dawson-Haggerty, 2019), *VTK* (Schroeder et al., 2006), and *Open3D* (Zhou et al., 2018).

In the following paragraphs, we provide a more comprehensive description of EpiStats' wide array of capabilities, offering a detailed explanation of the implementation and the intended use of the diverse algorithms, both regarding segmentation post-processing and morphological statistics computation.

3.2.1. Post-processing of 3D Segmented Images

The availability of a reliable method for the volumetric segmentation of cell tissues is an essential starting point for accurate and meaningful morphological and organizational analyses of biological tissues. Although the pipeline we described in the previous chapter enables us to obtain high-quality 3D cell segmentation, the outcomes are far from being error-free. Indeed, as previously mentioned, images obtained using a light sheet microscope often exhibit noisy features, which can easily result in segmentation errors. The manual curation process is employed to identify and rectify some of these errors, yet it may also introduce minor imprecisions and artifacts (e.g., bumpy boundaries between neighboring cells). Therefore, we implemented a range of post-processing operations on the segmented images in order to reduce the extent of segmentation imperfections in the analyzed samples. In the following, we will refer to the single instances in the segmented masks as *labels*.

The post-processing stack of EpiStats comprises the following sub-modules:

- A function to *remove unconnected regions* for a given label.
- A set of functions to detect and filter out cells that are either truncated or have an *incomplete neighborhood*.
- A set of functions to smooth cell boundaries and fill holes between labels by using *morphological erosion* followed by *dilation*.

Removal of unconnected regions is needed when the segmentation mask presents two or more areas marked with the same label that are not physically attached. Manual curation is often the cause of such errors. For instance, it may happen that a newly painted cell

[1] GitHub repository available https://github.com/kevinyamauchi/morphosamplers.

is mistakenly annotated with an already existing label, or that the user unintentionally draws an unconnected stroke without noticing it.
The algorithm for removing unconnected regions works as follows: for each label in the segmented image, the algorithm creates a binary mask. Then, a function computes the number of separated positive regions and the relative size. If the number of such regions exceeds one, only the largest is retained, while the rest are designated as background (i.e., labeled as 0). Finally, in case the algorithm detects a positive region whose size is comparable to the largest one, it triggers a warning message, as this scenario potentially indicates that the same label has been assigned to two or more distinct cells. The average time complexity for this algorithm is $\mathcal{O}(N \times M)$, where N and M are respectively the number of labels and the total number of voxels in the segmented image.

In our research, it's frequent to encounter images featuring cells that are partially cut off. This is frequently a result of how the sample was positioned within the image during the microscopy acquisition process, or due to some cropping operations that were performed to reduce the amount of cells to focus on. Clearly, since our goal is to obtain reliable morphological statistics from the segmented tissues, we want these truncated cells to be filtered out from our computation. In order to retrieve and exclude these cells we can simply build *bounding boxes* around each single cell and check which of these bounding boxes are in direct contact with the image edges. This algorithm has an approximate time complexity of $\mathcal{O}(M)$.

For the sake of producing accurate cell statistics, we also implemented an algorithm that retrieves cells that have an incomplete neighborhood. We say that a cell has an incomplete neighborhood if at least one of its lateral surfaces is a free surface, namely it is not in contact with other cells (see Fig. 9b, d). The exclusion of cells with an incomplete neighborhood is essential for our analysis, as otherwise, it would heavily bias the distribution of the number of neighbors among cells. For this particular task, we developed the following algorithm: focusing on one segmented cell at a time, we crop its upper and lower portions along its longest axis, effectively eliminating the apical and basal surfaces from our calculations. Subsequently, we tally the number of background voxels that come into contact with the specific cell. If this count surpasses a predefined threshold, which is determined in relation to the cell's size, we classify the cell as having an incomplete neighborhood, resulting in its exclusion from further computations. This implementation has an approximate time complexity of $\mathcal{O}(M)$.
In the context of the algorithms above, it is important to remark that, while morphological statistics are not computed on truncated cells or cells with incomplete neighborhoods, these cells are still considered in the counts of neighbors for cells close by.

The last set of operations carried out on the segmented samples is intended to obtain smoother cell boundaries and fill any gaps between them. Both of these processes contribute to achieving more reliable results when computing the cell contact area. For what concerns the refinement of cell borders, we resort to morphological erosion and dilation, which are the fundamental operations in morphological image processing (Serra, 1983; Dougherty, 1992). From the mathematical perspective, let us consider the integer grid $\mathbb{Z}^3$ and $A \in \mathbb{Z}^3$ 3D binary image. We define the *structuring element* $B \in \mathbb{Z}^3$ as a binary probe that is applied to the binary image. Erosion and dilation are performed by checking how this probe, respectively, fits or hits the shapes in the image. Namely, we can define erosion ($\ominus$) and dilation ($\oplus$) operations respectively as:

$$A \ominus B = \bigcap_{b \in B} A_{-b}$$
$$A \oplus B = \bigcup_{b \in B} A_b$$

where A_b and A_{-b} correspond, respectively, to a translation of A by an amount b or $-b$, with $b \in B$.
An illustration of the impact of morphological erosion and dilation on image boundaries can be found in Figure 9c. When utilizing a regular structuring element (such as a sphere in the 3D context), it becomes possible, for instance, to smooth the uneven cell boundaries often visible in the initial segmented image. In the post-processing pipeline of EpiStats, we execute a specific quantity of erosion iterations, which are then followed by a set number of dilation iterations. When erosion and dilation are applied in this sequence, it is referred to as *morphological opening*. Performing more dilation iterations than erosion iterations helps to bridge gaps between cells. Lastly, if any gaps persist between adjacent labels, we assign the corresponding voxels to the nearest label.

At the end of this image post-processing step, EpiStats automatically generates triangular meshes from the 3D segmented cells using the *marching cubes* algorithm (Lorensen and Cline, 1987). At this stage, only *Laplacian* smoothing is applied to the constructed meshes, as we do not require extremely regular meshes to run morphological analyses.

3.2.2. Computation of Morphological Cell Statistics

As discussed above, EpiStats enables the extraction of several morphological statistics from segmented volumes of biological tissues. In particular, it incorporates algorithms for the computation of 3D cells' features, including surface area, volume, elongation, neighbors, and contact area. Additionally, it calculates area and neighbors on 2D slices of the

three-dimensional image. Depending on what is more convenient, these algorithms operate either directly on the labeled voxel image or on the 3D triangular mesh of individual cells, which are generated in the previous step of the pipeline.
In the following, for the computation of each morphological feature, we explain the relative algorithmic implementation, discussing its approximate computational complexity.

Algorithms for 3D morphological statistics

The surface area of segmented cells is trivially computed by summing the area of the triangular faces that define the 3D mesh. Therefore, the time complexity is $\mathcal{O}(N \times V)$, where V is the (average) number of vertices in each mesh.

On the contrary, the cells' volume is computed by counting the number of voxels associated with each label and multiplying this number by the voxel size of the image. In this case, the complexity is $\mathcal{O}(M)$.

Cell elongation describes the ratio of the cell's length to its width. Practically, it is obtained as the ratio of the principal components of inertia from the inertia tensor computed for cell meshes. In this case, we rely on functions provided in the *trimesh* Python library, whose time complexity is not directly accessible.

For what concerns the computation of the cell neighbors within the tissues, we implemented an algorithm that operates on the segmented image. Specifically, given a certain cell label, we expand the associated binary image by two voxels in all directions employing binary dilation. Then, we subtract this expanded image from the original binary image to identify the positions of neighboring voxels that are in close contact with the boundary of that label. By examining the labels of these neighboring voxels, we obtain a list of neighbors for the given cell. The approximate time complexity for this algorithm is $\mathcal{O}(M)$.

The algorithm to compute the contact area between adjacent cells is slightly more involved and requires the use of triangular meshes. For each given cell, it iterates over the list of its neighbors to find which triangular facets are in close contact with any of the facets of the neighbors' meshes. This last operation is performed by checking whether the minimum distance between the face centroid of the considered cell and any of the vertices of the neighboring cells falls below a predefined contact cutoff. This threshold is determined as twice the maximum dimension of a voxel. To speed up this calculation, a *KDTree* (Friedman et al., 1977) is initially constructed from each cell mesh vertices, enabling efficient retrieval of the minimum distance vertex as described above. The overall time complexity of this algorithm is $\mathcal{O}(N^2 \times V \times \log V)$. However, since for each cell, we limit

the algorithm to the analysis of the neighboring cells, the actual complexity is less than quadratic with N.

Algorithms for morphological statistics on 2D slices

The rationale behind the computation of cell statistics on 2D slices of the segmented volume is to assess the theoretical laws concerning tightly packed cell tissues that we described in the introduction. In particular, our goal is to use the data coming from our epithelial samples to empirically evaluate Lewis' law and Aboav-Weaire's law. The former states that the average cell area of cells having n neighbors is linearly related the the number of neighbors n itself through the relationship in (1). The latter, instead, claims that the average amount of cells adjacent to the neighbors of a given cell is inversely related to the number of neighbors of the cell itself (2). For clarity, when we talk about Aboav-Weaire's law we refer to the cell for which neighbors of neighbors are computed as the *central cell*. Therefore, to experimentally assess these laws we need to extract the following statistics from the 2D slices of the segmented tissue image:

- The cross-sectional area of cells.
- The number of *first-order* neighbors of cells.
- The number of *second-order* neighbors of cells (i.e., the number of neighbors relative to each neighbor of a given central cell).

The reader should notice that in our study we decided to assess the laws globally on the entire tissues, whereas in literature they are usually computed solely for the apical and basal surfaces. Our choice resides in the fact that we believe that the epithelial organization remains coherent on all the slices of a given tissue. As a consequence, we expect that the theoretical laws that were proven for some of the surfaces, actually hold for all of them at once.

The main challenge related to the computation of these 2D statistics is given by the fact that due to the curvature of the epithelial tissues' surface, the cells are commonly oriented in different directions. As a result, slicing the segmented volume along any of the axes of the canonical cartesian reference system would lead to crooked 2D projections and, hence, to inaccurate results (see Figure 9e). To address this challenge, within EpiStats, we've introduced an efficient algorithm for calculating both 2D area and neighbor statistics along the apico-basal axis of each cell. This axis defines the cell's orientation and it is computed as the principal axis in the inertia tensor. For the sake of clarity, we've broken down the algorithm explanation into distinct sections:

- First, we iterate across all the cells in the labeled image, computing and storing for each one its centroid, apico-basal axis, and length in the axis direction.
- Then, for each cell, we sample a certain number of 2D slices along its principal direction. To perform this task, for each given cell:
  I. We take a 2D evenly-spaced grid, whose size is able to accommodate a cell's neighbors and adjacent cells.
  II. We compute the rotation matrix that aligns the initial grid orthogonally to the cell's principal axis.
  III. We place the rotated grid at different locations along the cell's axis, each time sampling label values from the segmented volume at grid coordinates.
- Now, given each 2D slice, we compute the cells' area and number of first and second-order neighbors. The area is trivially computed by multiplying the number of pixels belonging to each label by the effective pixel size on the 2D slice. The number of neighbors for each given cell is obtained by adapting the algorithm for the 3D case to the 2D case. On the contrary, the calculation of the number of second-order neighbors is more involved. Namely, the algorithm reads as follows:
  For each 2D slice along the central cell's apico-basal axis and for each one of the cell's neighbors:
  I. We find the intersection between the grid itself and the principal axis of the neighboring cell.
  II. We sample a new 2D slice by placing a new grid orthogonally to the neighbor's apico-basal axis and centered at the intersection point above following the procedure described above.

Clearly, the leading terms in the algorithm's overall time complexity analysis are the extraction of the 2D slices along each cell's principal axis and the computation of the number of second-order neighbors for a given central cell. Let S be the number of 2D slices taken for each cell. Then, the approximate time complexity of sampling S slices is $\mathcal{O}(N \times S \times M)$. In turn, the calculation of the second-order neighbors requires $\mathcal{O}(N \times M)$, since the neighbors of a given cell can be as many as N and the complexity of finding the intersection and placing a new grid is proportional to the number of voxels in the image. Therefore, the total complexity is $\mathcal{O}(N^2 \times S \times M^2)$. Nevertheless, just like in previous algorithms, the actual number of neighbors of a given cell is way less than N. Moreover, the computation of second-order neighbors is skipped whenever at least one of the first-order neighbors has an incomplete neighborhood. In practice, this additional

check enables us to bypass calculations for many slices and cells, resulting in a substantial acceleration of the algorithm. To provide an example, the computation of 2D statistics for a sample of one hundred cells takes approximately 10 minutes on an Intel Core i7-9750H CPU.

3.3. Results

In this section, we report the outcomes of our comparative analysis of the morphological cell statistics computed from the segmented images of epithelial tissues using EpiStats. To ensure the consistency of the extracted features, in the segmentation post-processing pipeline, we applied six erosion iterations and eight dilation iterations for morphological smoothing. At the same time, we polished the triangular meshes built upon the cell volumes, by subjecting them to ten passes of the Laplacian filter. For all the samples in our study, we detected all the cells that were either truncated or had an incomplete neighborhood. In this way, we excluded from the subsequent computations 237 bladder cells, 41 bronchiole cells, 77 esophagus cells, 83 intestine cells, and 74 lung tube cells. Furthermore, in the context of morphological statistics collection, we decided that a good trade-off between obtaining detailed results and computational efficiency was to limit the number of slices used for computing 2D cell features to 50.

An important step before the statistical analysis of morphological cell features is the removal of cells that present outlying values in any of the features. Such outliers are most often related to segmentation errors and artifacts that were not detected during the previous steps. For instance, in some cases, the segmented image presents a large labeled region, presumably corresponding to two or more actual cells, that was, however, impossible to split due to a noisy boundary signal. In some other cases, it is possible that during curation the user unintentionally drew a tiny stroke, which is interpreted as a small cell. Overall, in the majority of the situations, the outlying cells are characterized by an anomalous size or peculiar neighboring relationships. Following these ideas, we perform outlier detection in a twofold approach:

- We mark as an outlier any cell that had less than three neighbors within the tissue volume since this case is not physically possible for cells with a complete neighborhood.
- We check for anomalous values of cell volume and of number of neighbors. In this context, we label cells as outliers if the *robust Z-Score* for their volume or number of neighbors is greater than 3 in absolute value. The robust Z-Score is computed by subtracting to a given feature the median and then normalizing by the median

absolute deviation. As a result, the outcomes are less influenced by the presence of extreme values within the observations.

Overall, this procedure led us to the detection of 80 outliers, among which 51 from the bladder, 4 from the bronchiole, 1 from the esophagus, 20 from the intestine, and 4 from the lung tube.

In the following, we report a two-sided analysis of the epithelial tissues' morphology and organization. First, we illustrate a comparison of the different features among the variety of our epithelial samples, to highlight the traits that allow us to characterize them the most. Afterward, considering the cell statistics collected from 2D slices of the segmented volumes, we check whether our data fits with Lewis' and Aboav-Weaire's laws.

Table 4: **Amounts of cell labels per tissue.** Number of initially available, excluded (since truncated or with incomplete neighborhood), and outlying cell labels for each tissue.

Tissue	Initial	Excluded	Outlying	For analysis
Intestine	322	83	20	**219**
Bronchiole	180	41	4	**135**
Esophagus	129	77	1	**51**
Bladder	638	237	51	**350**
Lung Tube	147	91	4	**52**

3.3.1. Comparative analysis of morphological cell statistics among epithelial tissues

In order to compare the cell shape and tissue organization across the range of epithelial samples involved in our study, we selected a subset of features that we deemed relevant. In particular, we included the following:

- *Surface area*,
- *Volume*,
- *Isoperimetric ratio*,
- *Number of neighbors*,
- *Elongation*,
- *Contact area fraction*.

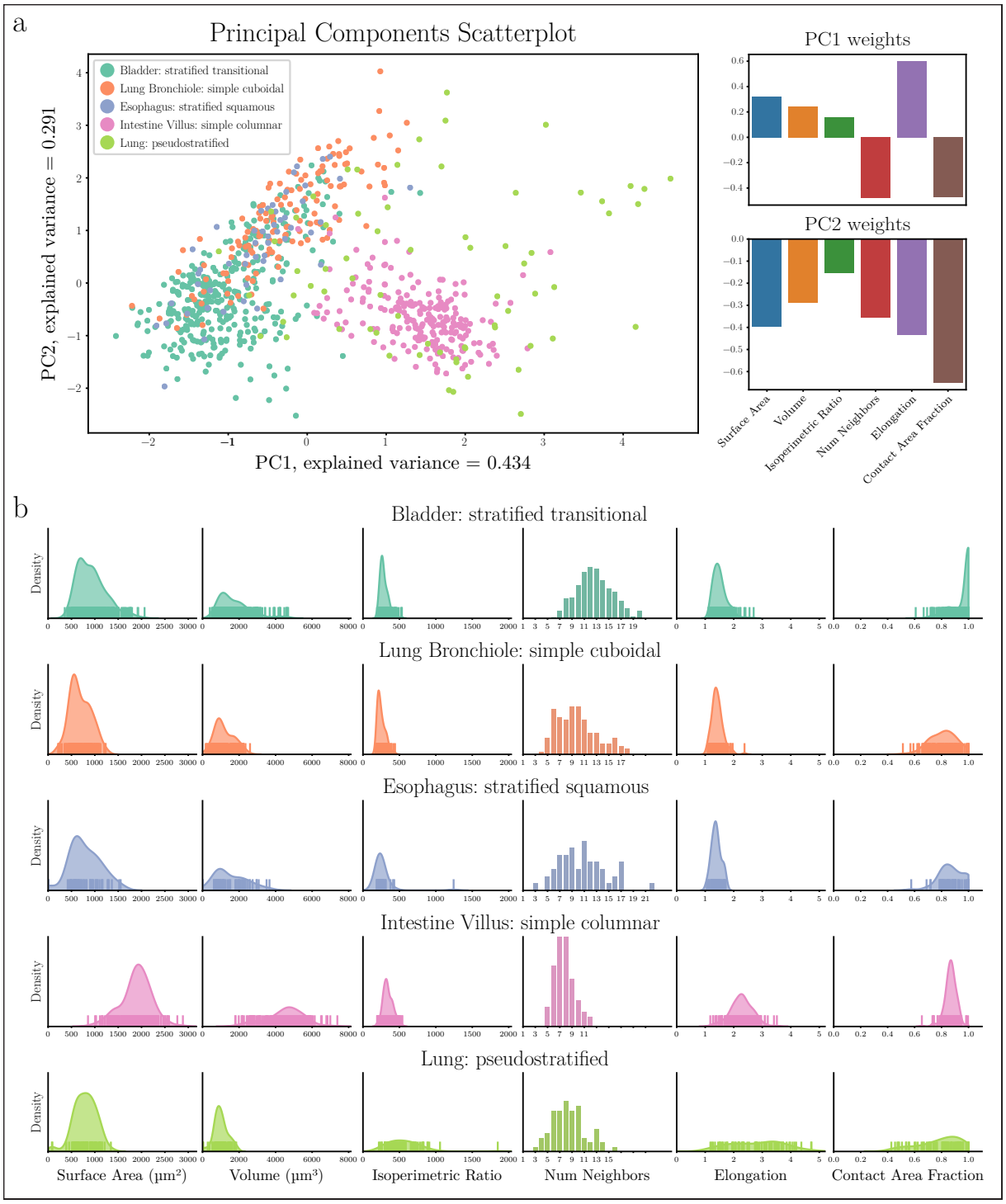

Figure 10: **Comparative analysis of morphological cell statistics a,** Principal Components scatter plot for the epithelial tissues involved in the study (left) and barplots reporting the weights of principal components. **b,** Kernel density estimate plots for all the epithelial tissues and the numerical features selected for the analysis. Each row outlines the estimated distributions for each tissue, while each column is associated with a different morphological cell feature.

The isoperimetric ratio is a measure of sphericity and it is computed as $S^3 \,/\, V^2$, where S and V are, respectively, the cell's surface area and volume. In 3D, the lower bound for the isoperimetric ratio is given by the sphere, namely equal to $(9/4)\pi$ (≈ 7.065). In general, the larger the isoperimetric ratio, the less spherical the shape. On the contrary, the contact area fraction is computed as the portion of each cell's surface area that is in contact with other neighboring cells. Hence, it takes values in $[0, 1]$ and provides a quantification of how tight is the epithelial packing. Additionally, the contact area fraction is equal to 1 in stratified tissues, as the cells do not have free apical and basal surfaces.

As a first mean of comparison of the epithelial tissues, we resorted to the analysis of the two *Principal Components* (PCs) computed over the set of standardized cell statistics. The scatter plot of the PCs together with the relative weights is shown in Fig.10a. Before delving into the discussion, we provide an interpretation of the PCs by looking at their coefficients. In this case, the first two PCs account for roughly 72.5% of the total variability of the data. Specifically, the first PC (PC1) explains 43.4% of variance and it presents marked positive weights for surface area, volume, and, especially, elongation. As a result, a positive PC1 score is associated with cells that are larger in size or particularly elongated. Vice versa, the number of neighbors and the contact area fraction are provided with negative weights, and, thus, a PC1 score below zero is likely related to a tightly packed or stratified epithelium. The second PC (PC2) accounts for 29.1% of the variability and it is defined as the (negative) weighted sum of all the considered features, where the bigger contribution is associated with the contact area fraction. Therefore, we can conclude that a positive PC2 score corresponds to smaller, more spherical cells in a more loosely arranged tissue, and vice versa for a negative value.
As we examine the scatter plot along the x-axis, we immediately notice that the intestine villus and the lung tube epithelia are associated with higher scores of PC1. This observation confirms the fact that the columnar and pseudostratified cells in these tissues exhibit significantly greater elongation compared to others. Additionally, the elevated PC1 scores may also suggest that these epithelial cells are generally larger in size, although we plan to validate this hypothesis through further analyses.
Conversely, drawing meaningful insights from the distribution of data points along the PC2 axis proves to be more challenging. The only discernible observation is that the point clouds related to the esophagus and bronchiole epithelial cells tend to skew toward higher PC2 values. Given our previous interpretation of PC2, this shift along the y-axis might imply that these epithelia typically comprise smaller cells that are less densely packed and tend to have a more spherical shape. This interpretation aligns well with the classification of esophagus and bronchiole epithelia as simple cuboidal and squamous tissues.

To support the observations drawn from the analysis of the PCs scatter plot we directly compared the distributions of each feature among different tissues. Figure 10b reports a matrix of kernel density estimate plots where rows are associated with epithelial tissues and columns with the morphological statistics. A first observation is that the distribution differs across the different samples for each one of the selected features. This claim is statistically confirmed by the fact that by running an ANOVA test separately on each variable, we obtain a p-value of 0 in all the cases.
Moreover, to refine our analysis, we took into consideration one feature at a time and we compared its distribution among different tissues. In this context, we made further observations and verified them by hypothesis testing at 95% confidence level with Bonferroni correction (see Fig. 11 for the results). Specifically, we deduced the following:

- The size of intestine cells is larger than all the other tissues' cells. This is proven by the fact that both the mean volume and surface area of intestinal cells are statistically larger than the ones of every other epithelium.
- The epithelial cells in the lung tube and in the intestine villi are, on average, more elongated than in the rest of the tissues. This result complies with the fact that both pseudostratified and columnar cells show a rather elongated shape.
- The exceptional elongation together with the peculiar shape of the pseudostratified lung cells establish their dominance also in the context of isoperimetric ratio.
- Regarding the number of neighbors, it is observable that in the bladder and esophagus epithelial tissues, the distribution tends toward larger values. This observation is expected since these epithelia are the only ones displaying a stratified organization. Consequently, the majority of their cells are entirely enclosed by neighboring cells, lacking free apical and basal surfaces.
- Similar reasoning applies to the values of the contact area fraction. Notably, we observe that, on average, the bladder epithelium maintains larger values in this aspect. Simultaneously, we observe that the contact area fraction for esophageal cells doesn't rank among the highest, despite the tissue's stratified structure.

3.3.2. Assessment of theoretical laws for 2D slices of epithelial tissues

As previously mentioned, one of the goals of our analysis was to assess the validity of Lewis' and Aboav-Weaire's laws on 2D slices of our segmented epithelial samples. To this end, we computed each cell's cross-sectional area and number of first and second-order

neighbors for each slice within the segmented volume. Specifically, we extracted such 2D slices by sampling the segmented volumes along the apico-basal axis of each cell. This choice of slicing method, as opposed to a more direct approach along one of the Cartesian axes, aims to reduce the chance of getting distorted 2D volume sections, which would lead to poor estimation of cross-sectional area and neighboring relationships.

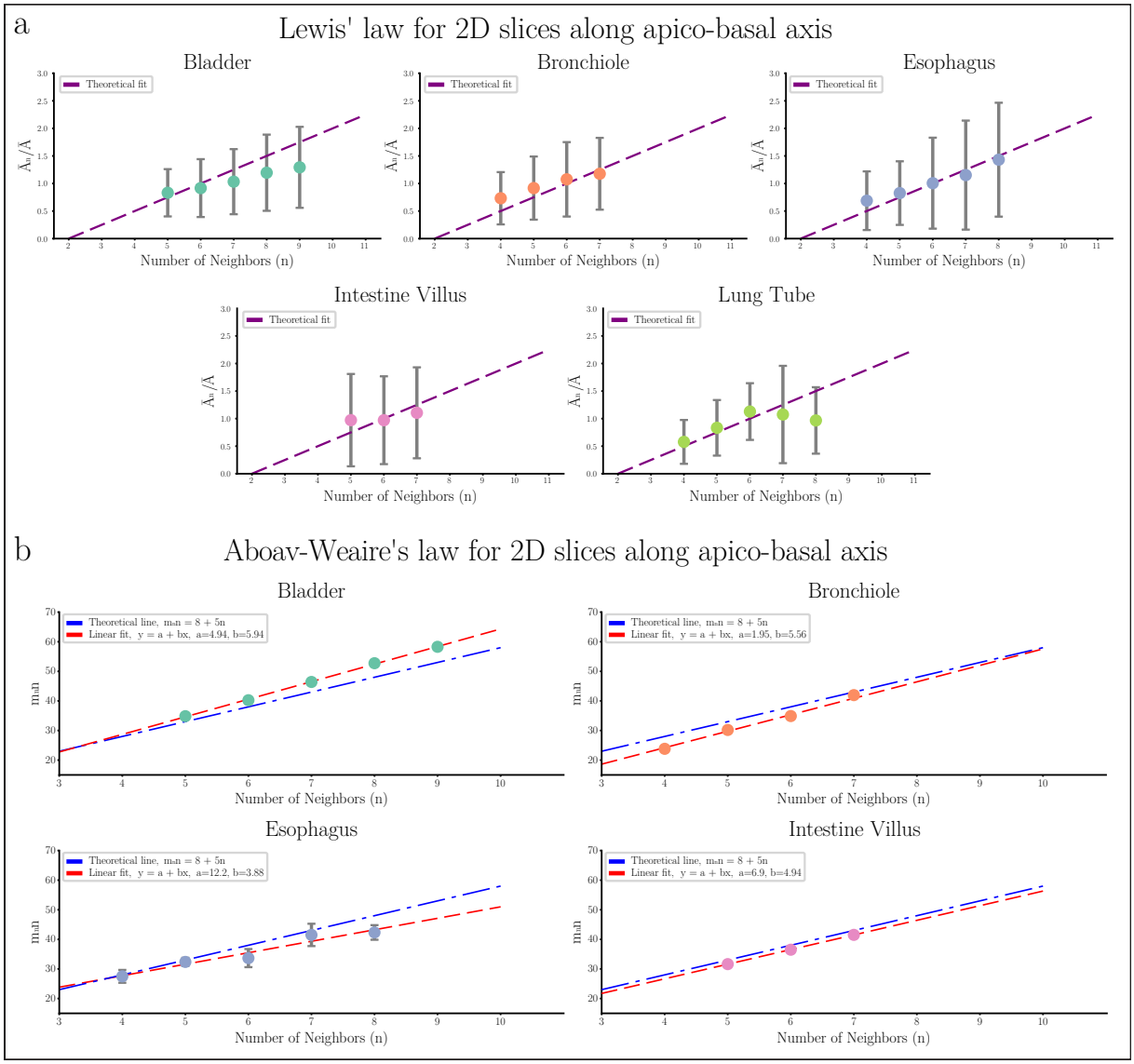

Figure 11: **Experimental assessment of Lewis' and Aboav-Weaire's laws on 2D slices of segmented epithelial tissues** Empirical values regarding the Lewis' law (**a**) and the Aboav-Weaire's law (**b**) were collected on 2D slices sampled from the segmented epithelial volumes along each cell's apico-basal axis. Aboav-Weaire's law data are not available for the lung tube sample, since the sample does not contain any complete second-order neighborhood.

It is important to remark that, within each tissue, we rejected the 2D statistics associated with infrequent numbers of first-order neighbors. Specifically, we marked as infrequent any value with relative frequency in 2D slices below a 1% threshold. The rationale behind this decision is that both Lewis' and Aboav-Weaire's laws are based on statistical moments, namely the average cross-sectional area and the average number of second-order neighbors. Therefore, it is essential that the sample size on which these averages are computed is sufficiently large to guarantee a certain degree of robustness.
Additionally, we removed from Lewis' law outcomes all the cells that lacked a complete first-order neighborhood, and from Aboav-Weaire's law outcomes also the cells without a complete second-order neighborhood. As a consequence, we couldn't assess Aboav-Weaire's law for the lung tube epithelial sample, as the limited extent of the segmented volume does not allow us to get complete second-order neighborhoods.

The Lewis' law experimental results are reported in Figure 11a. We can immediately observe that in the case of the esophagus, bronchiole, and lung tube samples, the obtained empirical points are consistent with the theory. On the contrary, in the other instances, we have discrepancies with the theoretical expectations. In particular, in the bladder sample, the values of $\bar{A}_n/\bar{A}$ follow a linear relationship, albeit with a smaller slope than the one described by Lewis. This slope approaches null within the intestine sample, suggesting that the cell cross-sectional area shows almost no correlation with the number of direct neighbors. However, it is crucial to note that drawing definitive conclusions warrants careful observation. As a matter of fact, while our algorithm for obtaining 2D slices is accurate enough in the case of elongated cells where the apico-basal axis is well-defined (e.g., the columnar epithelial cells of the intestine), it can struggle when dealing with samples with a wide range of cells' shapes and orientations. This observation holds true in the case of the bladder epithelium, where cells adapt their shape to the tissue movement, resulting in a variety of tissue geometries. Consequently, the observed behavior in the bladder epithelium could, in part, be attributed to the algorithm being not specifically designed to handle such kind of epithelium.

In Figure 11b we outline the results of Aboav-Weaire's law assessment. For each tissue, we plot the empirical values accompanied by the theoretical line and an empirical fit. The latter enables us to properly check how the experiment compares with the theory. Furthermore, it is important to note that in order to visualize a linear rather than an inverse relationship, we plot Aboav-Weaire's law in the following alternative form:

$$m_n n = 5n + 8$$

On the contrary, the line fitted on the experimental outcomes has equation $y = a + bx$, with the values of coefficients a and b reported in the plots' legends. Thus, we expect $a \approx 8$ and $b \approx 5$ for the empirical results to be coherent with the theoretical ones. In this context, we see that for all the epithelial tissues, the experimental slope b is actually rather close to the expected value of 5, as the coefficients are in the range $[3.88, 5.96]$. In contrast, except for the epithelial cells in the intestine villus, we observe that the value of the fitted intercept a is significantly shifted from the supposed value. Once again we can attribute this discrepancy to the fact that the algorithm used to sample the 2D slices from the segmented volume may struggle to deal with non-elongated cells.

Table 5: **Epithelial tissues summary statistics.** Mean, Standard Deviation, and Coefficient of Variation computer for each feature and each tissue involved in the study.

Tissue	Surf. Area	Volume	Isop. Ratio	Nr. Neigh.	Elong.	C.A.F*
Mean						
Intestine	1891.27	4510.80	342.71	7.05	2.27	0.87
Bronchiole	691.62	1189.72	251.52	9.21	1.43	0.81
Esophagus	793.86	1532.98	265.31	10.51	1.39	0.87
Bladder	932.30	1744.95	289.43	11.45	1.51	0.95
Lung Tube	792.89	999.74	565.03	8.35	2.81	0.79
Standard Deviation						
Intestine	328.39	1033.77	66.81	1.48	0.39	0.04
Bronchiole	226.73	531.64	54.54	3.06	0.20	0.10
Esophagus	333.82	871.17	148.7	3.76	0.17	0.09
Bladder	313.00	866.55	59.22	2.79	0.25	0.08
Lung Tube	238.84	378.49	244.99	2.55	0.92	0.16
Coefficient of Variation						
Intestine	0.17	0.23	0.19	0.21	0.17	0.05
Bronchiole	0.33	0.45	0.22	0.33	0.14	0.12
Esophagus	0.42	0.57	0.56	0.36	0.12	0.10
Bladder	0.34	0.50	0.20	0.24	0.17	0.08
Lung Tube	0.30	0.38	0.43	0.31	0.33	0.20

* *Contact Area Fraction*

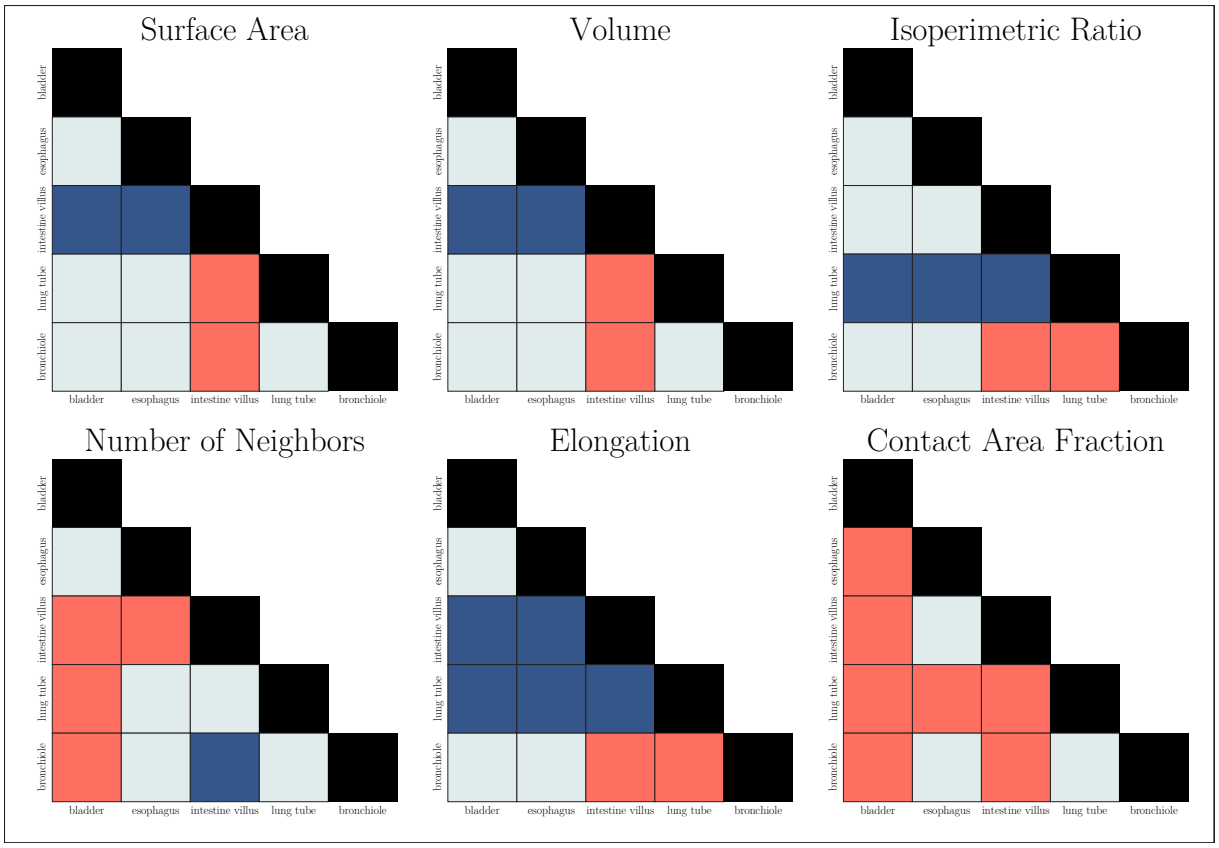

Figure 12: **Results of Bonferroni-corrected statistical tests on the mean difference of each feature among every pair of tissues.** Each matrix reported in the plot corresponds to a different morphological feature. Each matrix entry reports the outcome of the test performed between the tissues associated with the rows and columns of the matrix. A blue square signifies that there is statistical evidence to state that the particular feature is on average higher for the tissue listed in the row. Conversely, a red cell indicates that the tissue listed in the column has been proven to surpass the one in the row for that specific characteristic. Finally, a grey entry means that the test does not allow us to provide any specific claim.

3.4. Discussion

In this chapter, we presented the method we used to compute morphological cell statistics from the segmented volumes of epithelial tissues. With the aim of comparing such features among different tissues to get insights about cells' shape and tissues' organization, we developed EpiStats. This Python library performs efficient morphological statistics extraction from segmented samples of cellular tissues. Moreover, EpiStats not only provides fast algorithms for the computation of cell statistics both in the segmented volume

and on 2D slices sampled from it, but it also carries out the postprocessing of the segmented labels to increase the accuracy in the outcomes, and the generation of triangular meshes for each cell. Overall, one of the main strengths of this software is that it offers an accessible end-to-end pipeline, that, given a segmented tissue as input, enables the user to automatically obtain as output a polished dataset containing the desired morphological cell features. Additionally, EpiStats includes a wide array of functionalities for the analysis of the resulting statistics, such as functions to detect outliers, or to produce informative plots from the data. Finally, as the name suggests, EpiStats was originally designed for the morphological analysis of epithelial tissues. However, all of its capabilities have been purposely implemented to generalize well to other kinds of cellular patterns. For instance, the software has been employed by other researchers of our laboratory in the morphological analysis of early-stage embryos, producing robust and accurate results even in that case.

Thanks to the EpiStats library, we have been able to compute several morphological features from the different segmented epithelial tissues involved in our study. We then utilized such information to perform a comprehensive comparative analysis of the different cells' shapes and tissues' organization, with the aim of characterizing the diverse epithelia. In particular, for this task, we considered some salient cell statistics, such as their volume, surface area, isoperimetric ratio, number of neighbors, elongation, and contact area fraction. The choice of the aforementioned features is based on the fact that they provide a thorough description of a variety of epithelial characteristics, such as the cells' size (volume and surface area), their shape (isoperimetric ratio and elongation), and the coordination among contiguous cells (number of neighbors and contact area fraction).
Using the morphological statistics introduced above, we derived meaningful insights about our epithelial samples by firstly evaluating the outcomes of the Principal Components computed on the cell features and secondly comparing their distribution among the different tissues (see Table 5 for the mean values of the morphological features across tissues). This analysis led us to the following characterization of the analyzed tissues:

- The cells in the intestine villus epithelium are remarkably bigger, with their volume and surface area being between 2 and 3 times larger than the ones of the other tissues included in our study. Moreover, as expected, the columnar intestinal cells present an elongation that is only matched by the pseudostratified cells of the lung tube epithelium. Despite the exceptional elongation, their isoperimetric ratio is not statistically larger than the one in the other samples, probably due to the degree of regularity that is observable in these cells' shape. Regarding the number of neighbors, we observe that the intestinal epithelium stands at the bottom of

the ranking. However, from the statistical point of view, only the bladder and the bronchial epithelia can claim to have a larger number of neighbors. An interesting point concerning the cells' contact area fraction is that it is generally larger than in the other mono-layer tissues (namely bronchiole and lung tube), but statistically smaller than in the stratified tissue present in the bladder. This fact probably hints that the epithelial cells of the intestine villi are quite tightly packed. Additionally, it is meaningful to observe that the coefficient of variation of most features computed on the intestinal cells is lower compared to the other tissues. This detail might suggest that the villus epithelial tissue outlines a striking regularity, with the columnar cells packed in a well-organized structure. A further result supporting this speculation can be found in the fact that in the assessment of Lewis' law for 2D slices, we saw that the value of $\overline{A}_n/\overline{A}$ is almost independent of the number of neighbors n. Once again, this insight suggests that the epithelial cells of the intestinal villi are organized following a more regular pattern than the other epithelia.

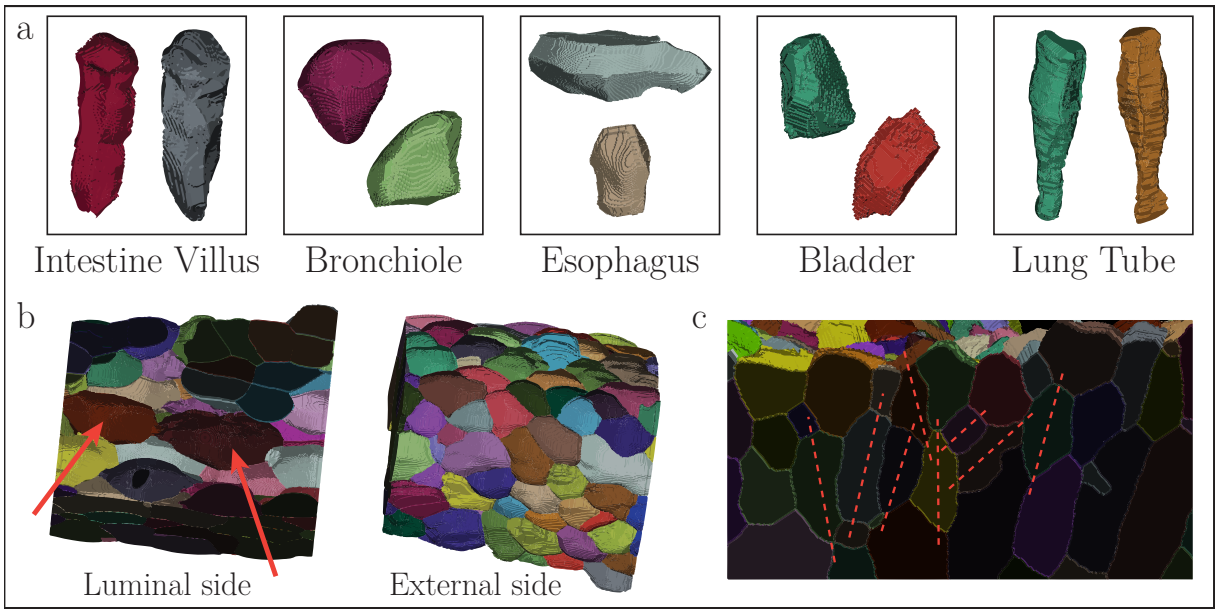

Figure 13: **Characteristics of the epithelial tissues. a,** Typical shape of cells from the different epithelia involved in our study. It is interesting to notice that the cells within mono-layer epithelia tend to have, more or less, the same shape throughout the sample. In contrast, cells from stratified epithelia show diverse morphology depending on their location in the tissue. **b,** The esophageal epithelium presents elongated squamous cells on the luminal surface of the tissue (left). In contrast, the rest of the cells are characterized by a cuboidal-like morphology (right). **c,** Apico-basal axis definition as each cell's principal axis may be misleading in the case of complex tissues like the urothelium, where neighboring cells are oriented along a wide range of different directions.

- The bladder epithelial tissue presents, on average, a consistently larger cell contact area fraction and a number of neighbors that is comparable only to the one of the esophageal epithelial cells. As we mentioned in the Results section above, the natural explanation of these results is given by the fact that the bladder possesses a stratified epithelium. As a result, all of the internal epithelial cells are surrounded by neighbors in all directions. The size of epithelial cells in our bladder sample is considerably smaller than in the intestinal villus, but slightly larger than the rest of the tissues. However, this difference is not wide enough to derive any statistically relevant conclusion. Finally, regarding elongation and isoperimetric ratio, we observe that the urothelial cells are quite similar to the esophageal and bronchial epithelial cells. This fact suggests that even though the cells in the bladder epithelium can be stretched according to the tissue movements, they are generally close to spherical or cuboidal shapes.

- We already mentioned that the cells within the esophageal epithelial tissue are characterized by a generally large number of neighbors. Similar to what was discussed for the bladder epithelium, in this case, these features can be imputed to the esophagus being a stratified epithelium. However, the average value of cells' contact area fraction is not comparable to the bladder, and it is, instead, in the same range of single-layer tissues like the one surrounding the intestinal villi. As we previously discussed, we can attempt to justify this fact by observing that the squamous cells in the luminal side of the esophagus have a rather extended free surface (Fig. 13b). As a result, the global contact area fraction for the cells in this tissue is heavily reduced due to this fact. We can further notice that the epithelial cells in the esophagus tend to show a smaller elongation with respect to the other tissues. While this is not surprising if we consider the columnar and pseudostratified cells of the intestine and the lung tube, we would expect the squamous cells in the esophagus to be more lengthened in comparison with the rather cuboidal cells of the bronchiole and the bladder. Nevertheless, we should note that the esophageal epithelium outlines squamous cells only on the luminal surface of the tissue, whereas the majority of the cells in the bulk of the sample show a more spherical shape. As a result, these epithelial cells are, on average, not particularly elongated. Lastly, we can state that the size of the cells in the esophagus sample is rather similar to the other tissues in terms of size.

- Epithelial cells in the lung tube are characterized by extraordinary values of elongation and isoperimetric ratio. This result is not surprising as the pseudostratified cells are particularly stretched along their apico-basal axis. Moreover, the greater

isoperimetric ratio with respect to the columnar cells of the intestine can be attributed to the fact that pseudostratified cells present a less regular organization within the tissue, with the cell membrane peculiarly modeled to accommodate the nucleus. For what concerns all the other features, we can conclude that the lung tube does not show peculiar values, meaning that lung tube epithelial cells are not exceptional in terms of size and overall tissue organization.

- The bronchiole epithelial tissue does not show particularly exceptional values in any of the analyzed features. Overall, we can observe that cells' size stands similar to the one of all tissues but the intestine. Additionally, despite being statistically equivalent to the one of the esophagus and the bladder, the isoperimetric ratio of bronchial cells is the lowest. This detail highlights the fact that the cells exhibit a particularly spherical shape, which agrees with the classification of the bronchiole epithelial cells as cuboidal. Finally, the values attained by the remaining features are aligned with the ones found in tissues with similar characteristics.

Future Developments

In this chapter, we extensively discussed the contributions that the Python library EpiStats can bring to the characterization of epithelial tissues. Nonetheless, the current version of the pipeline still presents some criticalities that could be improved in future works.
First of all, we already discussed that although the algorithm employed to sample 2D slices from segmented volumes certainly constitutes a huge improvement from the trivial sampling along the cartesian axes, it is far from being ideal for every kind of tissue and cell. Indeed, the fact that the planar slices are collected along the cells' apico-basal axis poses some limitations in cases where the latter is not clearly defined, which can sometimes be the case for non-elongated cells or tissues with a complex organization (Fig. 13c). A potential solution to this problem for future versions of the library could be to define the slicing direction depending on the estimate of the local curvature of the tissue.
An additional improvement would be the implementation of an efficient and accurate algorithm for the identification of apical, basal, and lateral surfaces of the cells. This would allow us, for instance, to draw more refined conclusions about the contact area distribution among the different surfaces of a given cell, and, in general, about the orientation of cells within the tissues.
Lastly, our observations revealed that, especially within stratified epithelia, the tissues exhibit diverse cell shapes and characteristics across different layers. Therefore, developing an algorithm to recognize cell types within a tissue would enable us to separately analyze distinct features for subgroups of cells within the same sample.

4 | Inference of mechanical properties of epithelial tissues

In the previous chapter, we extensively discussed the morphological features that influence the characterization and diversification of epithelia. In this regard, we observed that different tissues outline diverse cell shapes and organization among the cells. At this point, we aim to move forward and focus on the underlying mechanisms that generate such an extraordinary variety of structures. Indeed, as epithelial morphology is tightly linked to the physiological functioning of the tissue, we believe that the comprehension of the principles that regulate the packing of tissues would allow researchers to provide a more accurate interpretation of the causes that lead to pathological conditions, such as cancer.

From a mechanical perspective, the structure of tissues is determined by the dynamical interaction of their constituting units, the cells. Specifically, the distribution of stresses within the tissue dictates both the configuration adopted by groups of cells and the shape of individual cells. Furthermore, in order to thoroughly comprehend how tissues acquire and maintain their structure it is necessary to study the interplay of forces between the tissue and the extracellular matrix (ECM).

In the Introduction we discussed that cell-based in-silico models are currently recognized by the research community as the preferred tool for the mechanical analysis of biological tissues. By simulating the evolution of the virtual environments reproduced by these models, researchers can investigate the influence of inter-cell stress patterns (adhesion, repulsion) and intra-cell mechanical properties (stiffness, surface tension) on the overall characterization and arrangement of tissues. In this context, 3D Deformable Cell Models (3D-DCM) emerged as the state-of-the-art technology, primarily due to their capability to illustrate not just the interactions among various cells but also their internal dynamics. This possibility is allowed by the representation of cells as 3D triangular meshes, with forces acting upon their vertices. Among the wide array of simulation frameworks based on 3D-DCMs (Brakke, 1992; Madhikar et al., 2018; Liedekerke et al., 2020; Wang et al., 2021;

Torres-Sánchez et al., 2022; Okuda and Hiraiwa, 2023; Runser et al., 2023), SimuCell3D (Runser et al., 2023) stands out for its versatility and efficiency. Due to its optimized design, SimuCell3D successfully addresses the traditional trade-off between the number of simulated cells and the spatio-temporal resolution, a limitation commonly found in many existing cell-based models. For instance, this software enables the simulation of hundreds of cells within hours or, at most, a few days of computation. Moreover, it is specifically implemented with the aim of handling the simulation of cells characterized by a wide variety of morphologies, as well as other entities such as nuclei, lumens, and ECM. For these reasons, SimuCell3D constitutes an optimal software for the inference of the mechanical properties of the epithelial tissue involved in our study.

Within this context, we pursue our objective of inferring epithelial tissues' mechanical properties as follows: starting from the initial tissue geometry experimentally computed from the segmented microscope images, we use SimuCell3D to simulate the dynamical process that brings such initial geometry to an equilibrium state. In this framework, in order to find the model configuration that best characterizes the experimental tissue organization, we need to find a parameter set such that the morphological discrepancy between the initial and the steady-state geometries is the minimum possible or, ideally, none. Consequently, the problem can be interpreted as parameter estimation.

In the following, we will start by illustrating a robust triangular mesh refinement pipeline to improve the numerical stability of SimuCell3D when used on our epithelial samples. Then we will discuss the functionalities and the biophysical model defined within SimuCell3D. Afterward, we will showcase a Python module to compute significant metrics for monitoring simulation progress and results. After that, we will outline the parameter estimation process that we performed in order to retrieve the mechanical configuration that characterizes tissues at the equilibrium. Finally, we will discuss the current limitations of our approach and the possible solutions to obtain more accurate results.

4.1. Related Works

In the introductory paragraph, we mentioned that DCMs are currently the state-of-the-art technology for the computational analysis of biological tissues' mechanics. With regard to epithelial tissues, several simulation studies were conducted to uncover the principles underlying the dynamics and the organization of such tissues (Nagai and Honda, 2001; Smallwood, 2009; González-Valverde and García-Aznar, 2017; Barton et al., 2017; Mazarei et al., 2022; Runser et al., 2023). However, from our perspective, there's a noticeable scarcity of studies that engage in comparing cellular mechanical properties across various

epithelial structures. This subject holds particular relevance for us as the dynamic interaction among cells is the pivotal factor directing tissues to adopt a particular structure. Within the context of epithelial tissues, despite their shared fundamental function, cells attain different configurations to suit the specific demands of the hosting organ. Thus, comprehending the mechanical aspects that either unify or differentiate epithelia would significantly enhance our understanding of the mechanisms that let tissues acquire and maintain their structure during growth and morphogenetic deformation.

4.2. Methods

4.2.1. Mesh Refinement Pipeline

Earlier, we mentioned that SimuCell3D operates as a 3D Deformable Cell Model using 3D triangular meshes to represent cells. Therefore, the spatial resolution of the model relies on the length of edges between successive vertices. To ensure that the input meshes reach a specific resolution and lack irregularities that might compromise numerical stability, the SimuCell3D includes a predefined mesh refinement step before the beginning of simulation runs. In the case of our segmented epithelial samples, some cells exhibited notable imperfections, such as sharp angles and irregular boundaries, largely due to segmentation artifacts. Consequently, the SimuCell3D's mesh refinement tool alone was insufficient to create sufficiently regular meshes. To address this issue, we introduced a customized mesh refinement pipeline, allowing us to generate smoother meshes and, thus, ensure improved numerical stability during simulations.

The aforementioned mesh refinement pipeline is implemented in Python and it is part of the EpiStats we introduced a customized mesh refinement pipeline library. Most of the functions rely on existing tools from *PyMeshFix* (Attene, 2010) and *PyMeshLab* (Muntoni and Cignoni, 2021). Specifically, the pipeline is composed of the following steps:

1. A first cleaning iteration aimed at removing *non-manifold* faces and closing major holes in the input mesh.
2. A subsequent re-meshing step that relies on the *Ball Pivoting* algorithm (Bernardini et al., 1999) followed by *Taubin smoothing* (Taubin, 1995).
3. The application of predefined filters to find and delete duplicate faces, edges, and vertices.
4. A second cleaning iteration to remove further non-manifold faces and close any remaining hole in the re-meshed geometry.

5. A procedure to consistently orient the normal vectors to the mesh faces.

4.2.2. SimuCell3D Model

Biophysical Model

In the SimuCell3D biophysical model, the mechanical state of each cell is determined, other than the viscous damping, by two main components. These components are the forces that act on each cell's membrane and the cell-cell interactions. Specifically, the contribution related to the cell membrane can be summarized by the following energy potential:

$$U = KV\left(\ln\frac{V}{V_0} - 1\right) + \frac{k_a}{2}\left(\frac{A}{A_0} - 1\right)^2 + \int_{\partial\Omega}\left(\gamma + \frac{k_b}{2}(2H)^2\right)dS \tag{3}$$

The first term reported in the expression represents the energy associated with the cell's internal pressure, $p = dW/dV = -K\ln(V/V_0)$. This contribution is generated by the volumetric strain of the cell cytoplasm. In the formulas above V and V_0 stand for the current and target cell volumes, while K corresponds to the cytoplasmic bulk modulus. The second term in the energy potential defines the process that each cell undertakes to regulate its membrane area A. Deviations from the target value A_0 are penalized according to the parameter k_a, which is defined as the effective isotropic membrane elasticity. The integrand of the surface integral over the cell surface $\partial\Omega$ is composed of two additive factors. On the one hand, the isotropic cortical tension γ models the surface tension generated by the cell cortex. On the other hand, the second factor models the resistance of the cell cortex to bending, with H denoting the local mean curvature of the cell membrane, and k_b its bending rigidity.

Within SimuCell3D the intercellular interactions are modeled as local elastic contact forces that act at the level of their triangular faces. In this context, the contributions to the interaction model are described by a pair of parameters, which are the adhesion strength ω and the repulsion strength ξ.

Table 6 reports a list of the salient mechanical parameters that specify the SimuCell3D biophysical model. Each parameter is presented with the associated unit of measure, the default values in SimuCell3D, and a plausible range for parameter screening. In general, the values are obtained as the combination of experimental measures reported in the literature and computational assessments made by the developers of SimuCell3D.

Table 6: **Mechanical Parameters**. List of mechanical parameters that characterize the properties of biological tissue as described by the SimuCell3D biophysical model. Plausible ranges are computed combining experimental measures and numerical outcomes.

Symbol	Description	Unit	Default	Plausible Range
K	*Bulk Modulus*	Pa	2500	$[2 \times 10^3 - 1 \times 10^5]$
V_0	*Target Volume*	m^3	–	–*
A_0	*Target Surface Area*	m^2	–	–*
Q_0	*Target Isoperimetric Ratio*	–	250	–**
k_a	*Area Elasticity Modulus*	J	1×10^{-15}	–**
γ	*Surface Tension*	N/m	0.007	$[1 \times 10^{-5} - 1 \times 10^{-2}]$
k_b	*Bending Stiffness*	J	2×10^{-18}	$[1 \times 10^{-16}$-$1 \times 10^{-19}]$
ω	*Adhesion Strength*	Pa/m	1×10^9	$[1 \times 10^8$-$1 \times 10^{11}]$
ξ	*Repulsion Strength*	Pa/m	1×10^9	$[1 \times 10^8$-$1 \times 10^{10}]$
ρ	*Mass Density*	kg/m^3	1000	$[1000 - 1100]$**

* Value determined depending on the input sample.

** Kept to default value during our analysis

Numerical Methods

We already mentioned that SimuCell3D is an example of 3D-DCM. This entails that the forces involved in the model act directly on the vertices of the triangular meshes. As a consequence of the biophysical model described by the energy potential in (3), we can describe the overall force vector that acts on each mesh node as follows:

$$\vec{f_i} = \vec{f_{s,i}} + \vec{f_{m,i}} + \vec{f_{p,i}} + \vec{f_{b,i}} + \vec{f_{c,i}} \tag{4}$$

The components of the expression above read, respectively, as the surface tension forces ($\vec{f_{s,i}}$), the membrane area elasticity forces ($\vec{f_{m,i}}$), the pressure forces exerted by the cytoplasm ($\vec{f_{p,i}}$), the bending forces ($\vec{f_{b,i}}$), and contact forces due to adhesion and repulsion ($\vec{f_{c,i}}$). Figure 14a schematically illustrates how forces act on the faces of the triangular meshes associated with the cells.

Given the schema of forces as mentioned above, SimuCell3D implements time propagation by solving the dynamic equation of motion:

$$m\ddot{\vec{r}_i} + \zeta\dot{\vec{r}_i} = \vec{f_i}$$

In the expression above $\vec{f_i}$ is the nodal force vector in (4) and ζ the viscous damping

coefficient. The nodal mass m is computed as $m = \rho V/N_n$, where V is the cell volume, ρ is the cell's mass density, and N_n is the number of vertices in the cell's triangular mesh. Within the aforementioned simulation framework, the nodal positions $\vec{r_i}$ and linear momenta $\vec{p_i} = m\dot{\vec{r_i}}$ are integrated according to the semi-implicit Euler scheme in the following way:

$$\vec{p_i} \leftarrow \vec{p_i} + \Delta t \vec{f_i} - \zeta \vec{p_i}/m$$
$$\vec{r_i} \leftarrow \vec{r_i} + \Delta t \vec{p_i}/m$$

where Δt is the time step employed for the increment. The tuning of this time step parameter plays a crucial role. In fact, the value of the time increment determines a trade-off between the numerical stability of the simulation, and the computational time required to reach a steady state. For instance, opting for a smaller time step enhances the robustness of the simulation, given that cell deformation between successive time points tends to be more limited. Conversely, models employing a shorter time step necessitate more time to simulate the entire evolution of the biological tissue structure. Consequently, the computational time needed for the analyzed tissue to attain equilibrium may render the selection of such a time step impractical.

Additional Technical Aspects

Earlier, we discussed the importance of providing high-quality 3D triangular meshes as input to the SimuCell3D model in order to guarantee a certain degree of numerical stability during simulations. In the case of our study, the consistency of the input cell meshes is ensured by the application of our custom refinement pipeline. Furthermore, with the aim of preserving the regularity of triangular meshes throughout the subsequent simulation iterations, SimuCell3D implements a local mesh adaptation algorithm. Namely, this algorithm checks that edge lengths are maintained within the range $[l_{min}, l_{max}]$. Specifically, the minimum edge length l_{min} is a user-defined model parameter, whereas is empirically set to $l_{max} = 3l_{min}$. The local mesh adaptation method works as follows:

- When the length of an edge exceeds l_{max}, the local mesh adaptation method splits it in two, adding one node and two faces to the mesh. Then, the original momentum is distributed on the new vertices to ensure continuity.
- In the case an edge shrinks to a length below l_{min}, it is collapsed into a node. Now, the new momentum is computed as the sum of the ones corresponding to the merged nodes.

Another relevant model parameter is the contact cutoff distance, which arises from the definition of the intercellular interaction model. In this context, the contact forces are supposed to act between pairs of faces a and b. In particular, an actual force is exchanged if and only if the shortest signed distance d_{ab} within the facets is in the range $[-c, c]$, where c is the contact cutoff distance mentioned earlier.

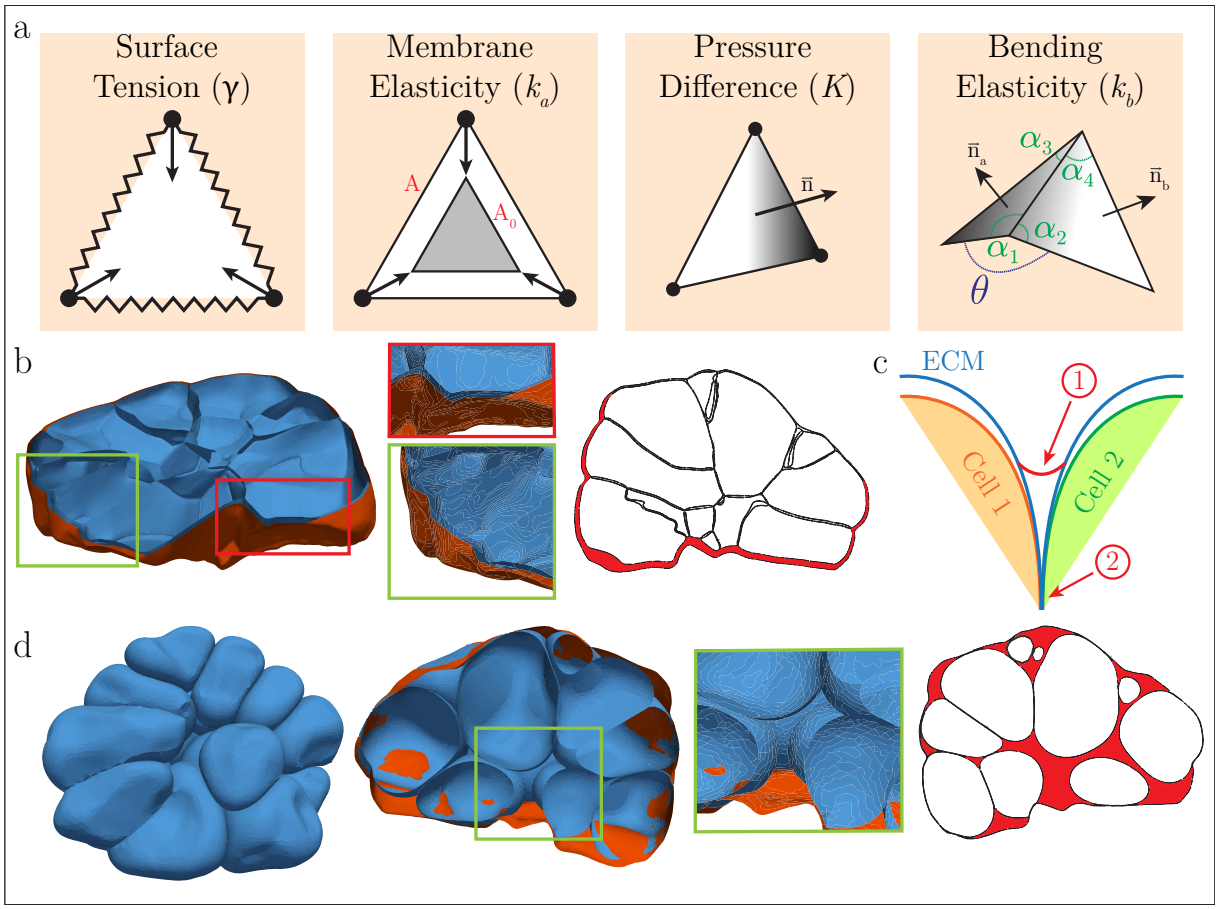

Figure 14: **SimuCell3D schema of forces and generation of the triangular mesh for representing the ECM shell. a,** Summary of different forces acting on the triangulated cell membranes. **b,** Generating the ECM shell triangular mesh from the binary mask computed by simply merging the cell labels leaves a non-negligible gap between the ECM and the cell meshes, as depicted in the clipped volume (ECM shell in orange, cells in blue) and the 2D slice (gap highlighted in red). **c,** The algorithm presented for the generation of a tighter ECM shell mesh performs a smoothing of the mesh surface in correspondence with cell junctions (1). This approach is designed to avoid the possibility of having sharp junctions (2), a feature that would cause major numerical problems during the simulations. **d,** During the simulation run, cells tend to adhere to the ECM leaving a wide gap in the middle of the sample, which has no biological explanation.

A complete list of all the parameters associated with numerical and technical aspects of the SimuCell3D model is reported in Table 7, together with the associated unit of measure, default values, and a plausible working range for generic cases.

Table 7: **Numerical Parameters**. List of numerical and technical parameters that complement the SimuCell3D model. Plausible ranges are provided by empirical assessments carried out by the software developers.

Symbol	Description	Unit	Default	Plausible Range
ζ	*Viscous Damping Coefficient*	kg/s	1×10^{-9}	$[1 \times 10^{-10} - 1 \times 10^{-7}]$
Δt	*Time Step*	s	1×10^{-7}	$[1 \times 10^{-8} - 1 \times 10^{-6}]$
l_{min}	*Minimum Edge Length*	m	2×10^{-7}	$-$*
c	*Contact Cutoff Distance*	m	5×10^{-7}	$-$*
T	*Simulation Duration*	s	0.001	$-$

* Kept to default value during our analysis

Finally, it is important to mention that for the sake of our simulations, we decided to encapsulate the epithelial cells composing our samples within an outer triangular mesh representing the ECM. This outer mesh, also termed the ECM *shell*, serves the purpose of applying an external constraint to the cells, hence preventing excessive extrusion of the meshes from the original tissue geometry. This choice is taken to both reproduce the biophysical presence of the ECM and avert unrealistic deformations and volume alterations during simulations.

From the technical perspective, the straightforward approach to obtain the ECM shell for a given clump of segmented cells is to first merge all the segmented cell labels in a single binary mask and, then, generate a triangular mesh from the merged image using one of the standard meshing algorithms and applying some smoothing to remove irregularities. Although this method is relatively robust and easy to implement, it presents one remarkable issue. Namely, the application of a global smoothing filter on the ECM shell mesh leads to the formation of a relevant gap between the cell and ECM meshes(Fig. 14b). During simulations, the cells usually rearrange their configuration by adhering to the outer ECM surfaces and filling these initial gaps. However, since the total volume of the cells is approximately constant throughout simulations, it happens that the empty spaces that were originally located at the junctions of cells are redistributed in the middle of the tissue, resulting in a lumen-like geometry that has no biophysical explanation (Fig. 14d). An intuitive way to reduce the extent of this would be to exclude any smoothing from the mesh generation process. However, this solution is not viable, as the triangular meshes obtained without any sort of smoothing may present irregular boundaries and sharp angles

at the cells' junctions, undermining the numerical stability of simulations.

Therefore, in order to overcome this issue, we decided to implement an alternative, more efficient method for the generation of a triangular mesh representing the ECM shell which is tighter on the internal cells. The idea is to first generate separate meshes for each cell and then construct the ECM shell mesh by merging the vertices of the cell meshes. The algorithm reads as follows:

1. Generate triangular meshes separately for each segmented cell label in the analyzed tissue sample.
2. For each mesh, compute the minimum distance between each one of its nodes and any neighboring mesh. Discard nodes whose distance exceeds a predefined threshold.
3. Estimate the normals to the vertices according to the local curvature of the point clouds.
4. Displace nodes by an amount equal to the average edge length along the directions of the normal vectors to make sure that the outer mesh encapsulates all the epithelial cells in the sample.
5. Given the resulting point cloud with the associated normals, generate a tight ECM shell triangular mesh by applying the *Screened Poisson Surface Reconstruction* algorithm (Kazhdan and Hoppe, 2013).

The procedure at point 2 offers a way to smooth the resulting ECM shell at the cell junctions to avoid major numerical issues while maintaining a tight enough envelope (Fig. 14c). In our study, we experimentally set the threshold to five times the average edge length across all the cells in the sample. The algorithm above is implemented in the EpiStats library and the triangular mesh representing the ECM shell is automatically generated within the mesh refinement pipeline illustrated above.

4.3. Results

In this section, we will report the outcomes of the application of SimuCell3D on the segmented epithelial tissue samples involved in our study. First of all, we will observe more in detail how the problem of inferring the mechanical properties characterizing epithelial tissues can be interpreted as a parameter estimation problem and will present our current approach to tackle it. Then, we will outline the status reached by our study, highlighting the outcomes related to some of our epithelial structures and the current challenges. In this regard, it is crucial to remark that our research constitutes one of the first instances

of the application of this simulation software on a wide variety of real biological samples. As a consequence, our project served to provide a first, realistic exploratory analysis of the mechanical properties of epithelial tissues, and a general assessment of the SimuCell3D pipeline, in order to find aspects that can be improved in future versions of the software. For this reason, all the results discussed below are referred to a pair of cell clumps extracted from the intestinal and bronchial epithelia that have been used as test cases. The samples are composed, respectively, of 16 and 19 cells to reduce the computational burden of simulations during these exploratory steps and are shown in Figure 15.

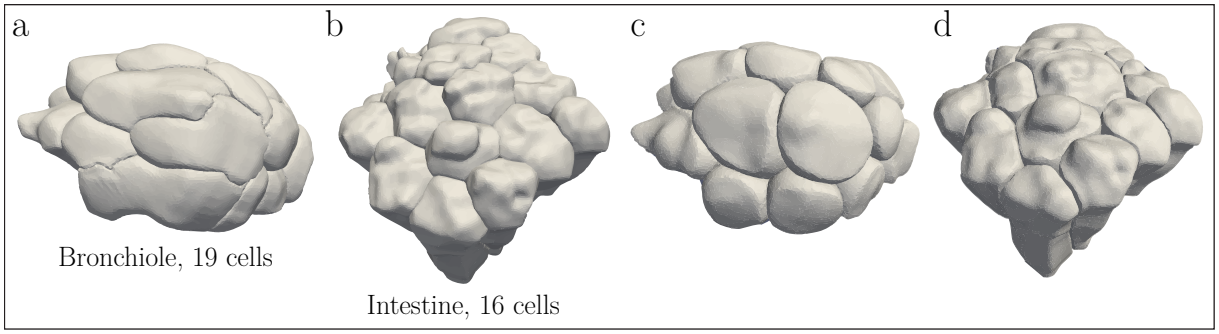

Figure 15: **Epithelial tissue cell samples used for the exploratory analysis on mechanical properties with SimuCell3D. a,** A cell clump from the segmented bronchiole epithelium, containing 16 cells. **b,** A cell clump from the segmented intestine epithelium, containing 19 cells. **c,** The geometry shown by the bronchiole sample after a simulation run using SimuCell3D. **d,** The geometry shown by the intestine sample after a simulation run using SimuCell3D.

4.3.1. Parameter Estimation

In the introductory section of this chapter, we proposed a numerical simulation framework based on cell-based models as a way to infer information about the mechanical configuration within epithelial tissues. Specifically, SimuCell3D simulates the evolving arrangement of epithelial cells, progressing from an initial empirical geometry obtained through the segmentation of an actual sample to a steady state. This simulation is conducted based on a specified set of mechanical and numerical parameters. Therefore, to identify the mechanical configuration that accurately describes our input tissues, it is necessary to determine the parameter set for which the simulated cellular geometry at the equilibrium closely resembles the initial structure. As a result, we can rephrase the research question of inferring the mechanical properties that characterize the organization of epithelial tissues as a parameter estimation problem.

In the context of our study, we attempt to solve the parameter estimation problem by

employing a grid search over the plausible ranges. More precisely, the parameter screening approach can be divided into two consecutive phases. First, starting from the default values and the predefined plausible ranges, we aim to provide a set of refined intervals for the model parameters such that simulations are numerically stable and the resulting steady-state geometries exhibit a somewhat similar structure to the initial one. Afterward, we repeatedly shrink the novel parameter ranges up to the point in which we find a single parameter configuration that closely describes the experimental cellular geometry. This last procedure is performed by empirically maximizing the IoU computed between the initial and the steady-state geometries for a given sample.

In practice, since the screening of parameter ranges involves launching a simulation run for each parameter configuration, it is preferable to investigate a few parameters at a time to reduce the overall computational resource requirements and to facilitate the analysis of the outcomes. A reasonable choice is to screen for a group of no more than three parameters at a time. Within the SimuCell3D model, parameters are not independent, and, hence, their interaction must be taken into account during screening. For instance, an increase in surface tension necessitates an increase of the bulk modulus for stability, given its role in counterbalancing the compressing pressure. As a result, we observed that it is better to first screen for the parameters that are more influential on the simulation outcomes (e.g., the surface tension), and then refine the search on less impactful ones.

In order to facilitate the analysis of the outcomes of parameter screening runs we developed some complementary tools in Python. In particular, we implemented the following:

- A function that monitors the total *volume loss/gain* with respect to the initial geometry over the iterations of the simulation runs. As we mentioned before, we generally expect the total volume of cells in the sample to remain approximately constant during the simulation runs.
- A function that at each simulation step computes the variation of average cell *sphericity* with respect to the initial geometry. Monitoring sphericity is a crucial aspect to assess the validity of the simulated geometries since for some parameter configurations cells tend to assume rather spherical shapes.
- A function that checks whether, at any point of the simulations, there is at least a pair of cells that *interpenetrate* each other. Interpenetration is common when the repulsion forces among cells are not sufficient to keep volumes apart. Clearly, geometries presenting any pair of interpenetrating cells are invalid. Hence, the associated parameter sets must be rejected.

- A function that computes the value of $(1 - IoU)$ between the cell geometry at any iteration of the simulation process and the initial geometry. Specifically, the IoU between geometries is obtained as the average of the values computed separately for each cell.

With the aim of visualizing the results obtained from the function above, we additionally developed the code for generating a dashboard that reports the values corresponding to each combination of parameters in a screening. An example of such a dashboard is reported in Figure 16.

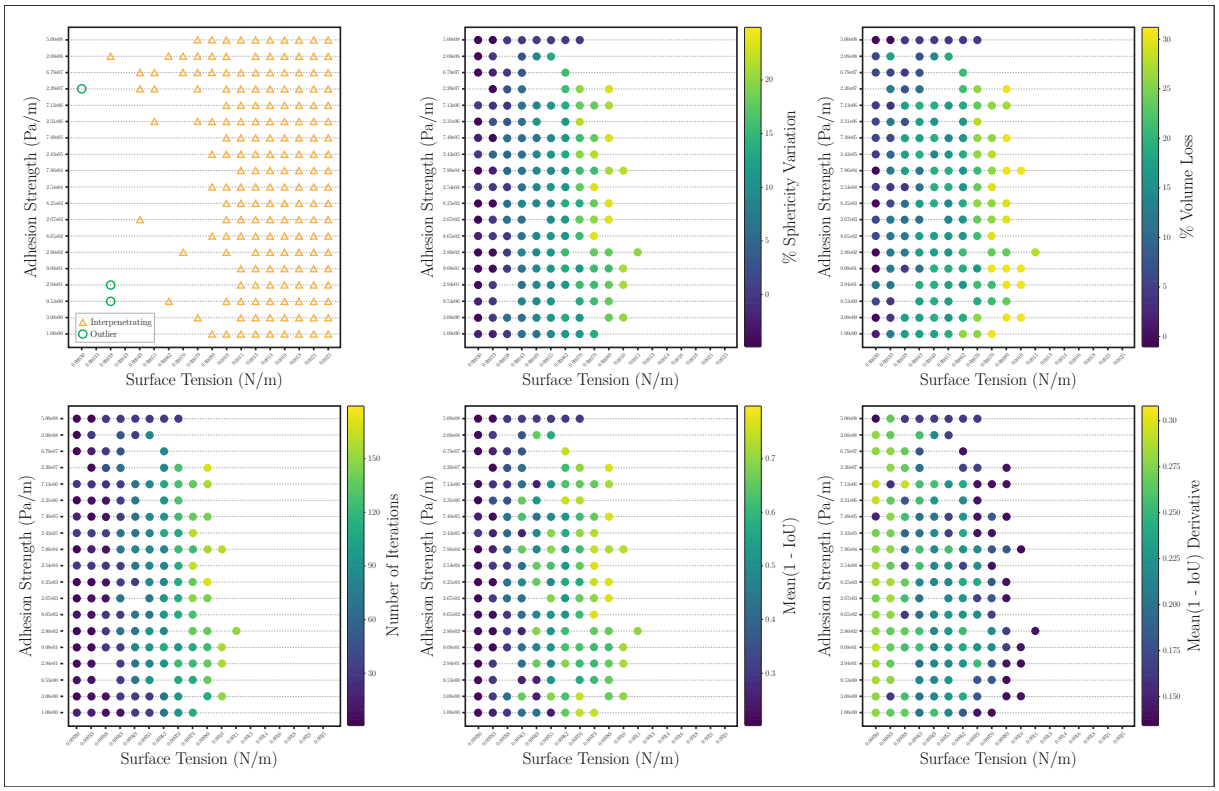

Figure 16: **Example of dashboard illustrating meaningful metrics to assess the goodness of a simulation run.** The dashboard reports the values of an array of metrics with the aim of facilitating the analysis of simulation outcomes. For instance, knowing that the outcome of a given simulation presents any interpenetrating cell, or a geometry characterized by an outlying volume enables us to directly exclude that parameter set from further analyses. Additionally, the percent variation of sphericity and volume loss provide a measure of the deformation that the cells were subjected to over the simulation. The number of iterations performed together with the derivative of the $mean(1 - IoU)$ metric give information about the advancement of a simulation run and whether it is close to the equilibrium. Lastly, the $mean(1 - IoU)$ is the objective score that we want to minimize during our parameter estimation task.

Step-by-step parameter screening approach

In the following, we describe a common step-by-step approach to perform a complete screening of the SimuCell3D model parameters. We urge to remark that this is an empirical approach, designed upon experience and advice from the SimuCell3D developers. Therefore, the same procedure might not generalize well for other samples.

As we mentioned earlier in the chapter, our parameter screening approach is composed of two stages:

1. In the first stage, from the default parameter ranges we extract sub-intervals for which simulation outcomes are reasonable. Specifically, we proceed as follows:

    - We begin to screen for surface tension γ, repulsion strength ξ, and bulk modulus K parameters in the ranges defined in Table 6. The goal, in this case, is to find sub-intervals for which simulations are stable and the cell geometries at the equilibrium do not show extreme behaviors. For instance, for low surface tension cell meshes might get noticeably irregular, while for large values they might assume spherical shapes. At this stage, it is recommended to consider as wide as possible parameter ranges. This is because adjusting subsequent parameters may impact those that were previously screened.

    - Considering the combinations of boundary values of the working sub-ranges for surface tension, repulsion strength, and bulk modulus found at the previous step, we now screen for adhesion strength ω and bending stiffness k_b in the predefined ranges. In this way, we obtain working sub-intervals for the entire range of mechanical parameters.

    - At this point we need to tune the numerical parameters of SimuCell3D, namely the time increment Δt and the viscous damping coefficient ζ. Selecting these parameters involves finding a compromise between numerical stability and the time required to reach the equilibrium. In other terms, our objective here is to determine an appropriate range for the viscous damping coefficient that enables the use of a large enough time step. Practically, in our simulation of epithelial tissues, we could not use a time increment Δt larger than $1 \times 10^{-7} s$.

    - Given all the computed sub-ranges for the model parameters, we now run some simulations in order to find the approximate simulation time T required to reach a steady state. This is useful in order to reserve an appropriate computational time for the later simulations. It is useful to observe that model configurations characterized by smaller values of surface tension generally take

more time to converge to the equilibrium state and, thus, determine the upper bound for the total computational time. In our case, we set $T = 0.1s$, which roughly corresponds to one week of computation.

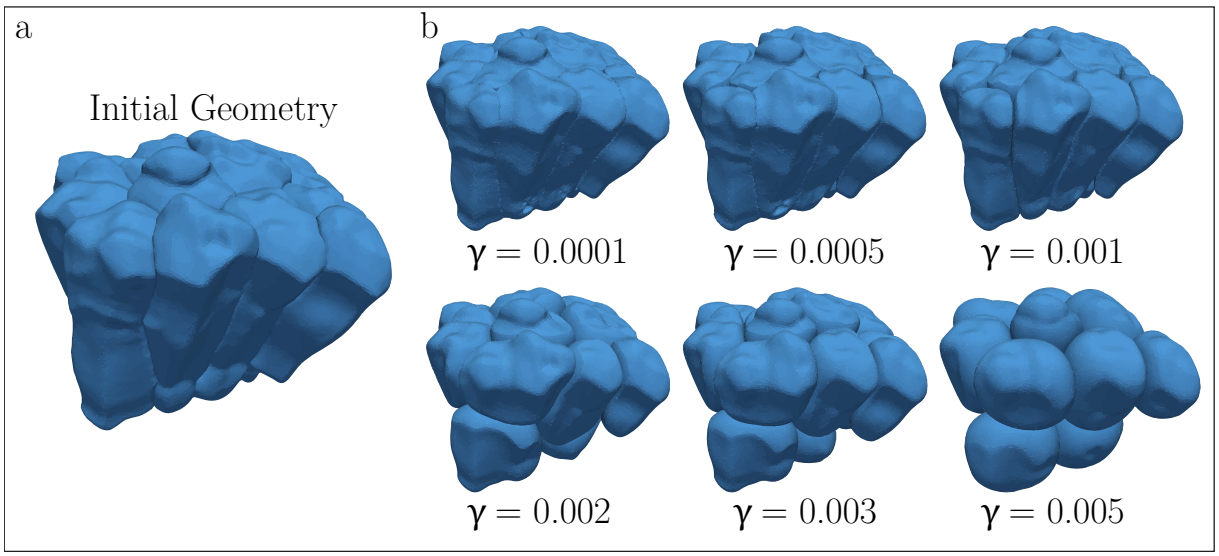

Figure 17: **Influence of surface tension parameter on simulation outcomes. a,** The initial geometry for the intestinal epithelial tissue sample. **b,** The simulated geometries for increasing values of surface tension γ. It is evident that as the surface tension increases, cells tend to assume more and more spherical shapes, resulting in biologically implausible structures.

Table 8: **Results of the first iteration of parameter screening**. List of working subranges and values found by applying the first stage of the step-by-step parameter screening approach on the intestinal and bronchial cell clumps selected for our preliminary analysis.

		Working ranges	
Parameter	**Unit**	**Bronchiole**	**Intestine**
Time Step	s	1×10^{-7}	1×10^{-7}
Viscous Damping Coefficient	kg/s	$[5 \times 10^{-10} - 2.5 \times 10^{-9}]$	$[5 \times 10^{-10} - 2.5 \times 10^{-9}]$
Simulation Duration	s	0.1	0.1
Surface Tension	N/m	$[3 \times 10^{-4} - 2.5 \times 10^{-3}]$	$[1 \times 10^{-4} - 1.5 \times 10^{-3}]$
Repulsion Strength	Pa/m	$[2.5 \times 10^{8} - 5 \times 10^{8}]$	$[1 \times 10^{8} - 2.5 \times 10^{8}]$
Adhesion Strength	Pa/m	$[0 - 5 \times 10^{8}]$*	$[0 - 1 \times 10^{9}]$*
Bulk Modulus	kg/s	$[2 \times 10^{4} - 6 \times 10^{4}]$	$[8 \times 10^{3} - 3 \times 10^{4}]$
Bending Stiffness	J	0**	0**

* *To better assess the influence of adhesion strength on the simulation outcomes, we tested a wider range.*

** *Resistance to bending was excluded from the model in our case due to numerical stability issues.*

The results regarding the first iteration of parameter screening for the intestine and bronchiole epithelial samples mentioned above are reported in Table 8. In this phase, we observed that the mechanical parameter that mostly influences the cell geometries is the surface tension (Fig. 17). Furthermore, since the contribution of the adhesion strength to the simulation outcomes was unclear, we decided to explore a wider range than the plausible one described above. Finally, we decided to exclude the contribution related to the resistance to bending from our model, as it was a source of numerical instability within our simulations.

2. In the second stage of our parameter screening approach, we aim to retrieve the parameter configuration that best characterizes the input geometry. Namely, starting from the working ranges above, we want to find the set of values for the mechanical parameters such that the value of $(1 - IoU)$ computed between the initial and the steady-state geometries is minimum, i.e., ideally equal to 0. Earlier we observed that the shape of cells attained during simulations is dominated by the surface tension value. Moreover, from previous studies in the field, we expect adhesion strength to be the other most influential parameter (Runser et al., 2023). Therefore, in this phase, our goal becomes to minimize $(1 - IoU)$ as a function of the surface tension and the adhesion strength. For this reason, we repeatedly screen for surface tension and adhesion strength, each time using a finer grid of values.

The results of the second screening stage are reported in Figure 18a, where the average value of $(1 - IoU)$ computed over the cells of the initial and final geometries are reported for the different combinations of surface tension and adherence strength. From a first look, we can immediately observe that the value of $mean(1 - IoU)$ shows a clear dependence on the surface tension and little to none on the adhesion strength. In particular, smaller values of $(1 - IoU)$ are attained for lower values of surface tension.

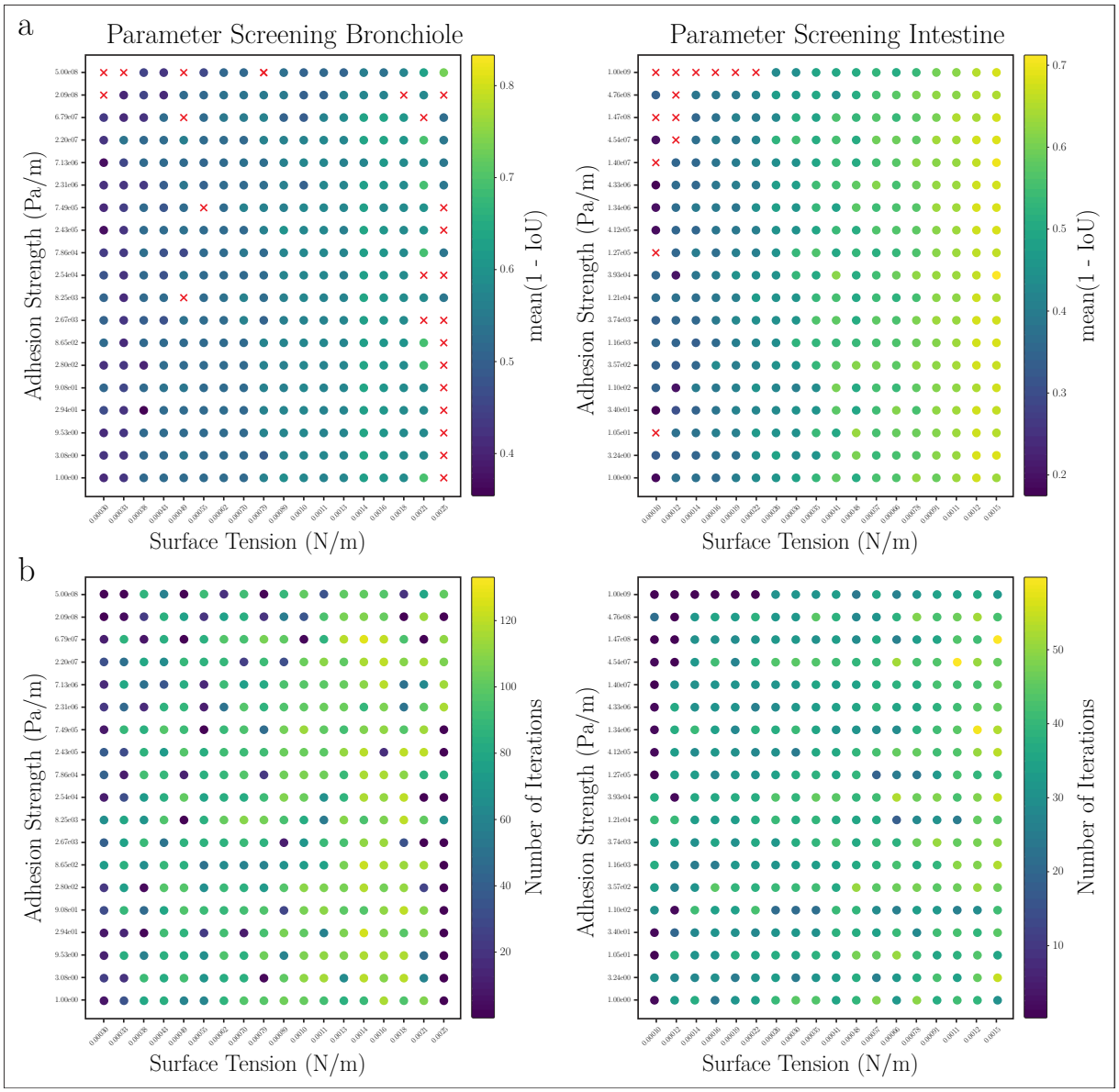

Figure 18: **Results of the exploratory parameter screening on the bronchiole and intestine cell clumps. a,** Values of $mean(1 - IoU)$ for the different combinations of adhesion strength and surface tension for the bronchiole (left) and intestine (right) cell clumps. Red crosses correspond to simulation runs that failed due to numerical errors. **b,** The distribution of the number of iteration values (bottom) shows a similar pattern of $mean(1 - IoU)$ values (top). Indeed, it is possible to observe that both values grow as the surface tension parameter increases. This fact hints that the resulting $mean(1 - IoU)$ could be related to simulations that did not reach a proper steady state.

4.4. Discussion

In this chapter, we discussed the problem of inferring the mechanical properties that characterize diverse epithelial tissues. Since the modality of packing of cellular tissue is deeply related to the stress distribution among cells, the possibility of developing a biophysical model that is capable of truthfully depicting the intra-cellular and inter-cellular interplay of forces is key to the understanding of the fundamental principles that regulate epithelial morphogenesis. For this purpose, we employed SimuCell3D, a powerful DCM that efficiently simulates growth and morphogenetic deformation for a wide array of tissue types. In order to improve its accessibility, we developed some Python modules that complement SimuCell3D functionalities. Specifically, we implemented a mesh refinement pipeline and an algorithm to generate tight triangular meshes for the ECM shell encapsulating a given group of cells. Moreover, we developed functions to compute meaningful metrics to monitor the quality of simulation outcomes and a dashboard to enable their direct inspection.

In the previous sections, we explained that the inference of the mechanical properties regarding the epithelial tissues involved in our study can be translated into a parameter estimation problem. Specifically, our goal is to identify a collection of model parameters such that the cellular structure, once simulations have reached an equilibrium state, closely mirrors the initially segmented structure. The similarity between tissue geometries is obtained by means of the value of $(1 - IoU)$ averaged over each cell within the sample. In our research, we performed parameter estimation following the approach presented in the earlier paragraphs for a patch of 16 epithelial cells from the intestine villus sample and 19 epithelial cells from the bronchiole sample.

The results of the parameter screening are reported in Figure 18a. We already mentioned that the mean values of $(1-IoU)$ are heavily dependent on the surface tension parameter, while there is almost no influence from the value of the adhesion strength parameter. The little to no impact of the adherence forces could imply that the adhesion model that is currently implemented within SimuCell3D is not sufficient to portray the actual forces interplay in epithelial tissues.

For both the tissues subjected to simulations, we see that the minimum for the $(1 - IoU)$ metric is attained for small values of surface tension (Fig. 18a). However, comparing the values obtained for $mean(1-IoU)$ with the number of simulation iterations performed in the corresponding simulations (Figure 18b), we notice that the two quantities are strongly correlated. Namely, the value of $mean(1 - IoU)$ increases with the number of simulation steps performed. This detail can be explained by hypothesizing that the simulation runs

related to small values of surface tension have not reached a proper steady state at the time the metric was computed. Therefore, it is not possible to exclude that by letting these simulations run for a longer time, we could reach different steady-state geometries, farther from the initial one, and, hence, associated with larger $(1-IoU)$ values. To support this hypothesis we can analyze in more detail the tissue geometries for simulation steps closer to the previously supposed steady state. We notice that cell meshes tend to slightly oscillate among iterations, meaning that a proper equilibrium is not reached. A possible reason to interpret the presence of such oscillation is that the current choice of viscous damping coefficient renders the system *underdamped*. A way to exit the underdamped regime is to employ a larger damping coefficient. However, this is not a possibility in our study, as we demonstrated that for $\Delta t \geq 1 \times 10^{-7} s$ simulations become numerically unstable if we employ a larger damping coefficient than the one we utilized.

To conclude, the current status of our research does not provide any undeniable evidence about the characterization of the mechanical properties associated with epithelial tissues. Nevertheless, we should acknowledge our study as the first iteration toward the achievement of a more detailed comprehension of epithelial dynamics using the SimuCell3D simulation model. In these regards, the contributions brought by our project constitute valuable bricks for the future development of the research in this framework.

Future Developments

As we mentioned above, our study should be regarded as an exploratory step in broader research about the mechanical characterization of epithelial tissues. For this reason, the potential enhancements that can be implemented in order to overcome the current challenges are manifold.

First of all, an improved version of the SimuCell3D adhesion model is essential to provide a more realistic representation of the stress distribution and a better assessment of the role of cell adhesion in the tissue forces interplay. As far as we are concerned, the implementation of this enhanced model is underway and, hence, we expect to employ it in future iterations of the project.

An alternative way to reduce the extent of the oscillation in the underdamped system discussed above could be to limit the value for the bulk modulus K. In fact, a model characterized by lower cells' internal pressure generally possesses less energy. Therefore, the system is inherently more stable, allowing for larger viscous damping. As we mentioned earlier, reducing the bulk modulus could impact also other parameters. Therefore, it would require to re-run the entire parameter screening from scratch.

Finally, in the previous sections, we extensively showed that finding the optimal ranges for the model parameters poses a significant challenge, as it requires massive computational time and manual work to analyze the results. As a consequence, the effort required for parameter screening represented the main bottleneck in the project's progression, and, due to time constraints, it prevented us from performing further experiments with different samples. A possible way to overcome these limitations is to replace the current grid search approach with a more efficient parameter estimation pipeline. Given an appropriate loss function and a set of metrics to evaluate the goodness of the simulation outputs, the pipeline would optimize the set of model parameters by minimizing the loss. Furthermore, such a parameter estimation algorithm would enable us to define and tune a larger parameter collection. For instance, it would provide the possibility of defining a distinct set of properties for each cell, increasing the detail of the tissues' mechanical description.

Conclusions and future developments

Throughout this thesis, we extensively explored the fact that, despite their shared role as protective barriers and separators, epithelial tissues exhibit a diverse range of organizations across various organs in living organisms. More specifically, different types of epithelia manifest varying numbers of cell layers (simple vs. stratified) and diverse cell shapes (cuboidal, squamous, columnar, pseudostratified, transitional). Despite extensive studies within the biological community, the understanding of the underlying factors that govern the formation of such a variety of structures remains an unresolved question. To unravel these mechanisms, it is crucial to pinpoint the distinctive features that characterize the broad spectrum of epithelia. Thus, we defined this study with the objective of providing a comprehensive characterization of epithelial tissues by conducting a thorough comparison of their different structures. In particular, our aim was twofold: firstly, to outline the morphological features that either unify or diversify epithelial tissues. Secondly, to explore the mechanical properties linked to different epithelial structures.

With the aim of achieving this objective, we acquired 3D images of various mouse epithelial samples using the light sheet microscope. By combining these images with others gathered for prior publications, we assembled a dataset encompassing five types of epithelia. Specifically, our dataset comprises the simple columnar epithelium from an intestine villus, the simple cuboidal epithelium from a bronchiole, the stratified squamous epithelium from the esophagus, the transitional epithelium from the bladder, and the pseudostratified epithelium found in the lung tube.

As a first step in our pursuit of the objective, we established an end-to-end pipeline for the high-quality 3D cell segmentation of epithelial tissues captured in the microscope images. This segmentation workflow is achieved through the integration of existing image preprocessing tools (Fiji, micro-SAM), pre-trained segmentation software based on a 3D U-Net model (Plantseg), and a collection of complementary enhancements developed by our team. Although a significant portion of the pipeline operates automatically, the outputs of the segmentation software necessitate a manual curation step. The purpose of this step is to identify and rectify significant segmentation errors and artifacts.
The application of this pipeline to our epithelial samples led to the segmentation of 322

cells from the intestine villus, 180 cells from the bronchiole, 129 cells from the esophagus, 638 cells from the bladder, and 147 cells from the lung tube. The time required for segmenting different tissues varied based on both the sample size and the version of the pipeline used. In fact, through successive enhancements to our pipeline, we managed to reduce the overall segmentation time for a fixed number of cells by approximately one-third. Despite these significant improvements, the time required to segment images remains a major constraint, preventing the inclusion of more samples in our analysis. In particular, the extent of manual curation necessary to produce accurate results is the main bottleneck of the current version of the workflow.
To potentially expedite the processes and enhance outcome quality in the future, we suggest implementing a more sophisticated model based on Vision Transformers (e.g., UNETR) for predicting cell boundaries and exploring the use of unlabeled images in the training phase through self-supervised learning. Both these solutions would certainly enable us to obtain more precise automatic 3D cell segmentation. As a consequence, we expect that the necessity of extensive manual curation would radically decrease.

In the second part of our research, we focused on studying the salient features that discriminate epithelial tissues from a morphological perspective. In this framework, we developed a Python library called EpiStats to extract, gather, and analyze morphological statistics regarding cell shape and organization. Within EpiStats, we implemented a series of efficient image analysis algorithms that enable the computation of the cells' surface area, volume, elongation, number of neighbors, and contact area with adjacent cells. The analysis of these morphological features led us to the following observations:

- Intestinal epithelial cells are remarkably larger in size and, as the columnar classification might suggest, they exhibit a greater elongation. Compared to other tissues, the intestine villus epithelium shows striking regularity, implying that its columnar cells are packed in a well-organized structure.

- Due to its stratified nature, epithelial cells from the bladder sample are surrounded by neighbors in all directions. As a result, they exhibit a larger number of neighbors and contact area fractions with respect to simple epithelia. Despite its classification as a transitional tissue, from the morphological perspective, the urothelium presents cells with similar characteristics to cuboidal ones.

- The esophageal epithelial tissue, characterized by a stratified structure, exhibits a substantial number of neighbors for its cells. However, the average contact area fraction of esophageal cells falls within the range of single-layer tissues, due to the extended free surface of squamous cells on the esophagus luminal side. The

distribution of cell features related to cell shape and size suggests that the inner layers of the esophageal epithelium are composed of cuboidal-like cells.

- The unique geometry of the pseudostratified epithelial cells in the lung tube is evident in their exceptional elongation and isoperimetric ratio values.
- The morphological features of epithelial cells in the bronchiole do not show any significantly different value compared to other tissues. This, coupled with the observation of bronchial cells having a relatively lower average elongation, is consistent with the classification of bronchial tissue as simple cuboidal.

In addition, EpiStats enables the calculation of area and the count of first and second-order neighbors on 2D slices extracted from segmented volumes along the apico-basal axes of the cells. For what concerns our epithelial samples, we compared the experimental 2D statistics with the theoretical expectations derived from Lewis and Aboav-Weaire's laws. In this way, we assessed whether the epithelial tissues under investigation followed common organizational patterns. The analysis revealed deviations from theoretical results in some epithelial samples.
Although the causes of these discrepancies are not fully understood, we hypothesized that one possible explanation is the current algorithm's difficulty in cases where cells lack a clear apico-basal axis. Hence, future iterations of the study plan to develop an enhanced algorithm that determines the slicing direction based on the local curvature of the tissue. Additionally, we realized the need to implement algorithms for the identification of different cell types within stratified tissues. This enhancement would enable us to separately analyze distinct features for subgroups of cells within the same sample, allowing us to draw more refined conclusions about tissue structure and organization in stratified epithelia.

The last step of our study focused on attempting to characterize epithelial tissues from a mechanical perspective. Namely, as the tissue structure is heavily impacted by the distribution of stresses within and among cells, being able to model such a forces interplay is crucial. To achieve this goal, we employed SimuCell3D, an efficient and versatile 3D-DCM-based simulation software. SimuCell3D describes the interactions between cells in tissue by means of a set of mechanical and numerical parameters. Therefore, the problem of the characterization of epithelial tissue mechanics translated into a parameter estimation problem. Specifically, the objective is to define a collection of mechanical parameters such that the morphological discrepancy between the experimental and the simulated steady-state geometry is minimal. In our case, this discrepancy was assessed in terms of the mean value of $(1 - IoU)$ computed over the cells in the selected sub-samples.

As our study represented one of the initial applications of SimuCell3D to diverse and intricate biological samples, we confined our investigation to an exploratory analysis involving two subsets of cells from the intestinal and bronchial epithelia. In addition to this investigation, we developed several complementary tools, including a pipeline for the refinement of 3D triangular meshes employed in the simulations and functions to analyze and visualize the parameter screening outcomes.

In this project, we addressed the challenge of parameter estimation through a two-step grid search approach. The initial stage involved screening default parameter ranges to identify sub-intervals ensuring numerical stability and reasonable cell geometries in simulation outcomes. Subsequently, in the second stage, the objective was to iteratively refine the range of the most influential mechanical parameters through sequential screenings, aiming to identify the parameter set that best aligns with the initial tissue geometry. This process, applied to selected epithelial samples, helped identify suitable sub-ranges for the numerical and mechanical parameters of the SimuCell3D model. Despite these efforts, we were unable to obtain a single parameter configuration that accurately described the structure of our experimental samples.

The potential reasons for this issue are diverse. Firstly, we found that the model describing cell-cell adhesion in the current version of SimuCell3D lacks the ability to offer a comprehensive depiction of the interplay of forces. As a matter of fact, our screenings indicated that the adhesion strength parameter had no impact on determining the final simulation outcomes. Therefore, in future iterations of the project we plan to employ an enhanced version of the adhesion model, whose implementation is currently underway by the SimuCell3D developers. Secondly, to ensure the convergence of simulations to equilibrium within a reasonable time frame (5-6 days) while preserving numerical stability, our parameter screening revealed the necessity of using a relatively small viscous damping coefficient. However, this resulted in an underdamped system, introducing oscillations that hinder reaching a proper steady state. To address this issue, we propose rerunning the entire parameter screening process anew, aiming to identify a parameter configuration linked to lower overall energy within the system. This lower energy regime can be attained, for instance, by limiting the range for the bulk modulus parameter K. Lastly, the absence of an automated parameter estimation pipeline makes the screening process lengthy and demands substantial manual effort for the analysis of the outcomes. This bottleneck hinders the exploration of more extensive combinations of parameters and the use of larger epithelial samples. As an improved strategy for parameter estimation, we recommend the deployment of a novel approach based on operation research. This approach should incorporate a fitting loss function and a suite of metrics designed to evaluate the fidelity of simulation outputs, with the ultimate goal of fine-tuning the model parameters

through the minimization of the loss.

In conclusion, the current status of our research offers meaningful outcomes relative to the morphological characterization of the different epithelial tissues involved in our study. Our analysis not only validated expected observations about tissue structure and organization but also highlighted non-trivial findings that set epithelia configuration further apart. Moreover, the EpiStats library developed to answer our research questions constitutes a versatile and powerful tool that answers the need for a unified, end-to-end pipeline for the morphological analysis of 3D segmented tissues. Due to its accessibility and accuracy, we expect EpiStats to be a pivotal component of many future studies in the field of epithelial biology.

On the contrary, the project does not provide any definitive evidence regarding the mechanical properties associated with epithelial tissues. The current simulation workflow still outlines some major bottlenecks and weaknesses that must be overcome in future versions of the work. Nonetheless, we proudly consider our study as a fundamental milestone toward the implementation of enhanced versions of the SimuCell3D simulation model for a more detailed understanding of epithelial mechanics.

Bibliography

D.A Aboav. The arrangement of grains in a polycrystal. *Metallography*, 3:383–390, 12 1970. ISSN 00260800. doi: 10.1016/0026-0800(70)90038-8.

James Ahrens, Berk Geveci, and Charles Law. ParaView: An end-user tool for large data visualization. In *Visualization Handbook*. Elesvier, 2005. ISBN 978-0123875822.

Anwai Archit, Sushmita Nair, Nabeel Khalid, Paul Hilt, Vikas Rajashekar, Marei Freitag, Sagnik Gupta, Andreas Dengel, Sheraz Ahmed, and Constantin Pape. Segment anything for microscopy. *bioRxiv*, page 2023.08.21.554208, 1 2023. doi: 10.1101/2023.08.21.554208. URL `http://biorxiv.org/content/early/2023/08/22/2023.08.21.554208.abstract`.

Mohammadmehdi Ataei, Vahid Shaayegan, Franco Costa, Sejin Han, Chul B. Park, and Markus Bussmann. Lbfoam: An open-source software package for the simulation of foaming using the lattice boltzmann method. *Computer Physics Communications*, 259: 107698, 2 2021. ISSN 00104655. doi: 10.1016/j.cpc.2020.107698.

Marco Attene. A lightweight approach to repairing digitized polygon meshes. *The Visual Computer*, 26:1393–1406, 11 2010. ISSN 0178-2789. doi: 10.1007/s00371-010-0416-3.

Alberto Bailoni, Constantin Pape, Nathan Hutsch, Steffen Wolf, Thorsten Beier, Anna Kreshuk, and Fred A. Hamprecht. Gasp, a generalized framework for agglomerative clustering of signed graphs and its application to instance segmentation. pages 11635–11645. IEEE, 6 2022. ISBN 978-1-6654-6946-3. doi: 10.1109/CVPR52688.2022.01135.

Daniel L. Barton, Silke Henkes, Cornelis J. Weijer, and Rastko Sknepnek. Active vertex model for cell-resolution description of epithelial tissue mechanics. *PLOS Computational Biology*, 13:e1005569, 6 2017. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1005569.

Stuart Berg, Dominik Kutra, Thorben Kroeger, Christoph N. Straehle, Bernhard X. Kausler, Carsten Haubold, Martin Schiegg, Janez Ales, Thorsten Beier, Markus Rudy, Kemal Eren, Jaime I Cervantes, Buote Xu, Fynn Beuttenmueller, Adrian Wolny, Chong Zhang, Ullrich Koethe, Fred A. Hamprecht, and Anna Kreshuk. ilastik: interactive ma-

chine learning for (bio)image analysis. *Nature Methods*, 16:1226–1232, 12 2019. ISSN 1548-7091. doi: 10.1038/s41592-019-0582-9.

F. Bernardini, J. Mittleman, H. Rushmeier, C. Silva, and G. Taubin. The ball-pivoting algorithm for surface reconstruction. *IEEE Transactions on Visualization and Computer Graphics*, 5:349–359, 10 1999. ISSN 1077-2626. doi: 10.1109/2945.817351.

Dapeng Bi, Jorge H. Lopez, J. M. Schwarz, and M. Lisa Manning. Energy barriers and cell migration in densely packed tissues. *Soft Matter*, 10:1885, 2014. ISSN 1744-683X. doi: 10.1039/c3sm52893f.

Kenneth A. Brakke. The surface evolver. *Experimental Mathematics*, 1:141–165, 1 1992. ISSN 1058-6458. doi: 10.1080/10586458.1992.10504253.

Phillip J. Brown, J. Edward F. Green, Benjamin J. Binder, and James M. Osborne. A rigid body framework for multicellular modeling. *Nature Computational Science*, 1: 754–766, 11 2021. ISSN 2662-8457. doi: 10.1038/s43588-021-00154-4.

Oriol Canela-Xandri, Francesc Sagués, Jaume Casademunt, and Javier Buceta. Dynamics and mechanical stability of the developing dorsoventral organizer of the wing imaginal disc. *PLoS Computational Biology*, 7:e1002153, 9 2011. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002153.

John Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8:679–698, 11 1986. ISSN 0162-8828. doi: 10.1109/TPAMI.1986.4767851.

Anne E Carpenter, Thouis R Jones, Michael R Lamprecht, Colin Clarke, In Kang, Ola Friman, David A Guertin, Joo Chang, Robert A Lindquist, Jason Moffat, Polina Golland, and David M Sabatini. Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, 7:R100, 2006. ISSN 14656906. doi: 10.1186/gb-2006-7-10-r100.

L. Chen, Paul Bentley, Kensaku Mori, Kazunari Misawa, Michitaka Fujiwara, and Daniel Rueckert. Self-supervised learning for medical image analysis using image context restoration. *Medical Image Analysis*, 58, 12 2019. ISSN 13618423. doi: 10.1016/j.media.2019.101539.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. 2 2020.

Chi-Li Chiu and Nathan Clack. napari: a python multi-dimensional image viewer platform

for the research community. *Microscopy and Microanalysis*, 28:1576–1577, 8 2022. ISSN 1435-8115. doi: 10.1017/S1431927622006328.

S.N. Chiu. Aboav-weaire's and lewis' laws—a review. *Materials Characterization*, 34: 149–165, 3 1995. ISSN 10445803. doi: 10.1016/1044-5803(94)00081-U.

M. L. Condic, D. Fristrom, and J. W. Fristrom. Apical cell shape changes during drosophila imaginal leg disc elongation: a novel morphogenetic mechanism. *Development*, 111:23–33, 1 1991. ISSN 0950-1991. doi: 10.1242/dev.111.1.23.

Michael Dawson-Haggerty. trimesh, 2019. URL `https://trimsh.org/`.

Pierre Barbier de Reuille, Anne-Lise Routier-Kierzkowska, Daniel Kierzkowski, George W Bassel, Thierry Schüpbach, Gerardo Tauriello, Namrata Bajpai, Sören Strauss, Alain Weber, Annamaria Kiss, Agata Burian, Hugo Hofhuis, Aleksandra Sapala, Marcin Lipowczan, Maria B Heimlicher, Sarah Robinson, Emmanuelle M Bayer, Konrad Basler, Petros Koumoutsakos, Adrienne HK Roeder, Tinri Aegerter-Wilmsen, Naomi Nakayama, Miltos Tsiantis, Angela Hay, Dorota Kwiatkowska, Ioannis Xenarios, Cris Kuhlemeier, and Richard S Smith. Morphographx: A platform for quantifying morphogenesis in 4d. *eLife*, 4, 5 2015. ISSN 2050-084X. doi: 10.7554/eLife.05864.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. 10 2020. URL `http://arxiv.org/abs/2010.11929`.

Edward R. Dougherty. *An Introduction to Morphological Image Processing*, volume 9. University of Michigan, 1992.

Bertalan Dudas. Human histology: A text and atlas for physicians and scientists, 2023.

Kenneth W. Dunn, Chichen Fu, David Joon Ho, Soonam Lee, Shuo Han, Paul Salama, and Edward J. Delp. Deepsynth: Three-dimensional nuclear segmentation of biological images using neural networks trained with synthetic data. *Scientific Reports*, 9:18295, 12 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-54244-5.

Dennis Eschweiler, Thiago V. Spina, Rohan C. Choudhury, Elliot Meyerowitz, Alexandre Cunha, and Johannes Stegmaier. Cnn-based preprocessing to optimize watershed-based cell segmentation in 3d confocal microscopy images. 10 2018.

Dennis Eschweiler, Richard S. Smith, and Johannes Stegmaier. Robust 3d cell segmentation: Extending the view of cellpose. 5 2021.

Luis M. Escudero, Luciano da F. Costa, Anna Kicheva, James Briscoe, Matthew Freeman, and M. Madan Babu. Epithelial organisation revealed by a network of cellular contacts. *Nature Communications*, 2:526, 11 2011. ISSN 2041-1723. doi: 10.1038/ncomms1536.

Reza Farhadifar, Jens-Christian Röper, Benoit Aigouy, Suzanne Eaton, and Frank Jülicher. The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. *Current Biology*, 17:2095–2104, 12 2007. ISSN 09609822. doi: 10.1016/j.cub.2007.11.049.

Jerome H. Friedman, Jon Louis Bentley, and Raphael Ari Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software*, 3:209–226, 9 1977. ISSN 0098-3500. doi: 10.1145/355744.355745.

Jan Funke, Fabian Tschopp, William Grisaitis, Arlo Sheridan, Chandan Singh, Stephan Saalfeld, and Srinivas C. Turaga. Large scale image segmentation with structured loss based deep learning for connectome reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41:1669–1680, 7 2019a. ISSN 0162-8828. doi: 10.1109/TPAMI.2018.2835450.

Jan Funke, Lisa Mais, Andrew Champion, Natalie Dye, and Dagmar Kainmueller. A benchmark for epithelial cell tracking. In Laura Leal-Taixé and Stefan Roth, editors, *Computer Vision – ECCV 2018 Workshops*, pages 437–445, Cham, 2019b. Springer International Publishing. ISBN 978-3-030-11024-6.

William T. Gibson and Matthew C. Gibson. *Chapter 4 Cell Topology, Geometry, and Morphogenesis in Proliferating Epithelia*, pages 87–114. 2009. doi: 10.1016/S0070-2153(09)89004-2.

Ismael González-Valverde and José Manuel García-Aznar. A hybrid computational model to explore the topological characteristics of epithelial tissues. *International Journal for Numerical Methods in Biomedical Engineering*, 33, 11 2017. ISSN 2040-7939. doi: 10.1002/cnm.2877.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

Noah F. Greenwald, Geneva Miller, Erick Moen, Alex Kong, Adam Kagel, Thomas Dougherty, Christine Camacho Fullaway, Brianna J. McIntosh, Ke Xuan Leow, Morgan Sarah Schwartz, Cole Pavelchek, Sunny Cui, Isabella Camplisson, Omer Bar-Tal, Jaiveer Singh, Mara Fong, Gautam Chaudhry, Zion Abraham, Jackson Moseley, Shiri Warshawsky, Erin Soon, Shirley Greenbaum, Tyler Risom, Travis Hollmann, Sean C. Bendall, Leeat Keren, William Graf, Michael Angelo, and David Van Valen. Whole-cell

segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature Biotechnology*, 40:555–565, 4 2022. ISSN 1087-0156. doi: 10.1038/s41587-021-01094-0.

Boris Guirao, Stéphane U Rigaud, Floris Bosveld, Anaïs Bailles, Jesús López-Gay, Shuji Ishihara, Kaoru Sugimura, François Graner, and Yohanns Bellaïche. Unified quantitative characterization of epithelial tissue development. *eLife*, 4, 12 2015. ISSN 2050-084X. doi: 10.7554/eLife.08519.

Barry M. Gumbiner. Epithelial morphogenesis. *Cell*, 69:385–387, 5 1992. ISSN 00928674. doi: 10.1016/0092-8674(92)90440-N.

Harold F. Gómez, Mathilde S. Dumond, Leonie Hodel, Roman Vetter, and Dagmar Iber. 3d cell neighbour dynamics in growing pseudostratified epithelia. *eLife*, 10, 10 2021. ISSN 2050084X. doi: 10.7554/eLife.68135.

Pedro Gómez-Gálvez, Pablo Vicente-Munuera, Antonio Tagua, Cristina Forja, Ana M. Castro, Marta Letrán, Andrea Valencia-Expósito, Clara Grima, Marina Bermúdez-Gallardo, Óscar Serrano-Pérez-Higueras, Florencia Cavodeassi, Sol Sotillos, María D. Martín-Bermudo, Alberto Márquez, Javier Buceta, and Luis M. Escudero. Scutoids are a geometrical solution to three-dimensional packing of epithelia. *Nature Communications*, 9:2960, 7 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-05376-1.

Robert Haase and Stephane Rigaud. clesperanto: Gpu-accelerated image processing across languages and platforms, 3 2020.

Philipp Hanslovsky, Vanessa Leite, Stephan Saalfeld, Tobia Pietzsch, Ulrik Günther, John Bogovic, and Juan Nunez-Iglesias. Paintera, 2018.

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with numpy. *Nature*, 585:357–362, 9 2020. ISSN 0028-0836. doi: 10.1038/s41586-020-2649-2.

Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger Roth, and Daguang Xu. Unetr: Transformers for 3d medical image segmentation. 3 2021. URL `http://arxiv.org/abs/2103.10504`.

Davide Heller, Andreas Hoppe, Simon Restrepo, Lorenzo Gatti, Alexander L. Tournier,

Nicolas Tapon, Konrad Basler, and Yanlan Mao. Epitools: An open-source image analysis toolkit for quantifying epithelial growth dynamics. *Developmental Cell*, 36: 103–116, 1 2016. ISSN 15345807. doi: 10.1016/j.devcel.2015.12.012.

Tsuyoshi Hirashima and Taiji Adachi. Polarized cellular mechanoresponse system for maintaining radial size in developing epithelial tubes. *Development*, 1 2019. ISSN 1477-9129. doi: 10.1242/dev.181206.

Andrea Horňáková, Jan-Hendrik Lange, and Bjoern Andres. Analysis and optimization of graph decompositions by lifted multicuts. 2017.

Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18:203–211, 2 2021. ISSN 1548-7091. doi: 10.1038/s41592-020-01008-z.

Ashish Jaiswal, Ashwin Ramesh Babu, Mohammad Zaki Zadeh, Debapriya Banerjee, and Fillia Makedon. A survey on contrastive self-supervised learning. *Technologies*, 9:2, 12 2020. ISSN 2227-7080. doi: 10.3390/technologies9010002.

Jörg Hendrik Kappes, Markus Speth, Björn Andres, Gerhard Reinelt, and Christoph Schn. *Globally Optimal Image Partitioning by Multicuts*, pages 31–44. 2011. doi: 10.1007/978-3-642-23094-3_3.

Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics*, 32:1–13, 6 2013. ISSN 0730-0301. doi: 10.1145/2487228.2487237.

Sangwoo Kim, Marie Pochitaloff, Georgina A. Stooke-Vaughan, and Otger Campàs. Embryonic tissues as active foams. *Nature Physics*, 17:859–866, 7 2021. ISSN 1745-2473. doi: 10.1038/s41567-021-01215-1.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. 4 2023. URL `http://arxiv.org/abs/2304.02643`.

Marco Kokic, Antonella Iannini, Gema Villa-Fombuena, Fernando Casares, and Dagmar Iber. Minimisation of surface energy drives apical epithelial organisation and gives rise to lewis' law. 2019. doi: 10.1101/590729. URL `https://doi.org/10.1101/590729`.

Rayan Krishnan, Pranav Rajpurkar, and Eric J. Topol. Self-supervised learning in

medicine and healthcare. *Nature Biomedical Engineering*, 6:1346–1352, 8 2022. ISSN 2157-846X. doi: 10.1038/s41551-022-00914-1.

Jochen Kursawe, Rémi Bardenet, Jeremiah J. Zartman, Ruth E. Baker, and Alexander G. Fletcher. Robust cell tracking in epithelial tissues through identification of maximum common subgraphs. *Journal of The Royal Society Interface*, 13:20160725, 11 2016. ISSN 1742-5689. doi: 10.1098/rsif.2016.0725.

Kisuk Lee, Jonathan Zung, Peter Li, Viren Jain, and H. Sebastian Seung. Superhuman accuracy on the snemi3d connectomics challenge. 5 2017. URL `http://arxiv.org/abs/1706.00120`.

Frederic T. Lewis. The correlation between cell division and the shapes and sizes of prismatic cells in the epidermis of cucumis. *The Anatomical Record*, 38:341–376, 5 1928. ISSN 0003-276X. doi: 10.1002/ar.1090380305.

Simon A. Lewis. Everything you wanted to know about the bladder epithelium but were afraid to ask. *American Journal of Physiology-Renal Physiology*, 278:F867–F874, 6 2000. ISSN 1931-857X. doi: 10.1152/ajprenal.2000.278.6.F867.

G. Li, T. Liu, J. Nie, L. Guo, J. Chen, J. Zhu, W. Xia, A. Mara, S. Holley, and S.T.C. Wong. Segmentation of touching cell nuclei using gradient flow tracking. *Journal of Microscopy*, 231:47–58, 7 2008. ISSN 00222720. doi: 10.1111/j.1365-2818.2008.02016.x.

Paul Van Liedekerke, Johannes Neitsch, Tim Johann, Enrico Warmt, Ismael Gonzàlez-Valverde, Stefan Hoehme, Steffen Grosser, Josef Kaes, and Dirk Drasdo. A quantitative high-resolution computational mechanics cell model for growing and regenerating tissues. *Biomechanics and Modeling in Mechanobiology*, 19:189–220, 2 2020. ISSN 1617-7959. doi: 10.1007/s10237-019-01204-7.

E. Liu, S. Vega, M.D. Treiser, H.-J. Sung, and P.V. Moghe. *Fluorescence Imaging of Cell–Biomaterial Interactions*, pages 291–303. Elsevier, 2011. doi: 10.1016/B978-0-08-055294-1.00101-X.

William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. *ACM SIGGRAPH Computer Graphics*, 21:163–169, 8 1987. ISSN 0097-8930. doi: 10.1145/37402.37422.

Pranav Madhikar, Jan Åström, Jan Westerholm, and Mikko Karttunen. Cellsim3d: Gpu accelerated software for simulations of cellular growth and division in three dimensions. *Computer Physics Communications*, 232:206–213, 11 2018. ISSN 00104655. doi: 10.1016/j.cpc.2018.05.024.

Mahmood Mazarei, Jan Åström, Jan Westerholm, and Mikko Karttunen. In-silico testing of the universality of epithelial tissue growth. *Physical Review E*, 106:L062402, 12 2022. ISSN 2470-0045. doi: 10.1103/PhysRevE.106.L062402.

Claire McQuin, Allen Goodman, Vasiliy Chernyshev, Lee Kamentsky, Beth A. Cimini, Kyle W. Karhohs, Minh Doan, Liya Ding, Susanne M. Rafelski, Derek Thirstrup, Winfried Wiegraebe, Shantanu Singh, Tim Becker, Juan C. Caicedo, and Anne E. Carpenter. Cellprofiler 3.0: Next-generation image processing for biology. *PLOS Biology*, 16: e2005970, 7 2018. ISSN 1545-7885. doi: 10.1371/journal.pbio.2005970.

Roeland M.H. Merks, Michael Guravage, Dirk Inzé, and Gerrit T.S. Beemster. Virtualeaf: An open-source framework for cell-based modeling of plant tissue growth and development . *Plant Physiology*, 155:656–666, 2 2011. ISSN 1532-2548. doi: 10.1104/pp.110.167619.

Hui Miao and J. Todd Blankenship. The pulse of morphogenesis: actomyosin dynamics and regulation in epithelia. *Development*, 147, 9 2020. ISSN 1477-9129. doi: 10.1242/dev.186502.

M. P. Miklius and S. Hilgenfeldt. Epithelial tissue statistics: Eliminating bias reveals morphological and morphogenetic features. *The European Physical Journal E*, 34:50, 5 2011. ISSN 1292-8941. doi: 10.1140/epje/i2011-11050-7.

Erick Moen, Dylan Bannon, Takamasa Kudo, William Graf, Markus Covert, and David Van Valen. Deep learning for cellular image analysis. *Nature Methods*, 16: 1233–1246, 12 2019. ISSN 1548-7091. doi: 10.1038/s41592-019-0403-1.

Bruno Monier, Melanie Gettings, Guillaume Gay, Thomas Mangeat, Sonia Schott, Ana Guarner, and Magali Suzanne. Apico-basal forces exerted by apoptotic cells drive epithelium folding. *Nature*, 518:245–248, 2 2015. ISSN 0028-0836. doi: 10.1038/nature14152.

Alessandro Muntoni and Paolo Cignoni. Pymeshlab, jan 2021.

Mandar Deepak Muzumdar, Bosiljka Tasic, Kazunari Miyamichi, Ling Li, and Liqun Luo. A global double-fluorescent cre reporter mouse. *genesis*, 45:593–605, 9 2007. ISSN 1526954X. doi: 10.1002/dvg.20335.

Tatsuzo Nagai and Hisao Honda. A dynamic cell model for the formation of epithelial tissues. *Philosophical Magazine B*, 81:699–719, 7 2001. ISSN 1364-2812. doi: 10.1080/13642810108205772.

G.M. Odell, G. Oster, P. Alberch, and B. Burnside. The mechanical basis of mor-

phogenesis. *Developmental Biology*, 85:446–462, 7 1981. ISSN 00121606. doi: 10.1016/0012-1606(81)90276-1.

Satoru Okuda and Tetsuya Hiraiwa. Modelling contractile ring formation and division to daughter cells for simulating proliferative multicellular dynamics. *The European Physical Journal E*, 46:56, 7 2023. ISSN 1292-8941. doi: 10.1140/epje/s10189-023-00315-5.

Miriam Osterfield, XinXin Du, Trudi Schüpbach, Eric Wieschaus, and Stanislav Y. Shvartsman. Three-dimensional epithelial morphogenesis in the developing drosophila egg. *Developmental Cell*, 24:400–410, 2 2013. ISSN 15345807. doi: 10.1016/j.devcel.2013.01.017.

Manuel Megías Pacheco, Pilar Molist García, and Manuel Ángel Diego Pombal. Atlas of plant and animal histology, 2008.

Constantin Pape, Alex Matskevych, Adrian Wolny, Julian Hennies, Giulia Mizzon, Marion Louveaux, Jacob Musser, Alexis Maizel, Detlev Arendt, and Anna Kreshuk. Leveraging domain knowledge to improve microscopy image segmentation with lifted multicuts. *Frontiers in Computer Science*, 1, 10 2019. ISSN 2624-9898. doi: 10.3389/fcomp.2019.00006.

Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros. Context encoders: Feature learning by inpainting. 4 2016.

Marco G. Patti, Walter Gantert, and Lawrence W. Way. Surgery of the esophagus. *Surgical Clinics of North America*, 77:959–970, 10 1997. ISSN 00396109. doi: 10.1016/S0039-6109(05)70600-9.

Karl Reinhardt. Über die zerlegung der ebene in polygone terms and conditions, 1918.

Katarzyna A. Rejniak. A single-cell approach in modeling the dynamics of tumor microregions. *Mathematical Biosciences and Engineering*, 2:643–655, 2005. ISSN 1551-0018. doi: 10.3934/mbe.2005.2.643.

N Rivier and A Lissowski. On the correlation between sizes and shapes of cells in epithelial mosaics. *Journal of Physics A: Mathematical and General*, 15:L143–L148, 3 1982. ISSN 0305-4470. doi: 10.1088/0305-4470/15/3/012.

Jos B.T.M. Roerdink and Arnold Meijster. The watershed transform: Definitions, algorithms and parallelization strategies. *Fundamenta Informaticae*, 41:187–228, 2000. ISSN 01692968. doi: 10.3233/FI-2000-411207.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. 5 2015.

Steve Runser, Roman Vetter, and Dagmar Iber. 3d simulation of tissue mechanics with cell polarization. 2023. doi: 10.1101/2023.03.28.534574. URL `https://doi.org/10.1101/2023.03.28.534574`.

Jean-Francois Rupprecht, Kok Haur Ong, Jianmin Yin, Anqi Huang, Huy-Hong-Quan Dinh, Anand P. Singh, Shaobo Zhang, Weimiao Yu, and Timothy E. Saunders. Geometric constraints alter cell arrangements within curved epithelial tissues. *Molecular Biology of the Cell*, 28:3582–3594, 12 2017. ISSN 1059-1524. doi: 10.1091/mbc.e17-01-0060.

B. A. Scalettar, J. R. Swedlow, J. W. Sedat, and D. A. Agard. Dispersion, aberration and deconvolution in multi-wavelength fluorescence images. *Journal of Microscopy*, 182: 50–60, 4 1996. ISSN 00222720. doi: 10.1046/j.1365-2818.1996.122402.x.

Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, Jean-Yves Tinevez, Daniel James White, Volker Hartenstein, Kevin Eliceiri, Pavel Tomancak, and Albert Cardona. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9:676–682, 7 2012. ISSN 1548-7091. doi: 10.1038/nmeth.2019.

Will Schroeder, Ken Martin, and Bill Lorensen. *The Visualization Toolkit (4th ed.)*. Kitware, 2006. ISBN 978-1-930934-19-1.

Jean Serra. *Image Analysis and Mathematical Morphology*. Academic Press, Inc., USA, 1983. ISBN 0126372403.

Jean Serra and Luc Vincent. An overview of morphological filtering. *Circuits Systems and Signal Processing*, 11:47–108, 3 1992. ISSN 0278-081X. doi: 10.1007/BF01189221.

Rod Smallwood. Computational modeling of epithelial tissues. *WIREs Systems Biology and Medicine*, 1:191–201, 9 2009. ISSN 1939-5094. doi: 10.1002/wsbm.18.

Christoph Sommer, Christoph Straehle, Ullrich Kothe, and Fred A. Hamprecht. Ilastik: Interactive learning and segmentation toolkit. pages 230–233. IEEE, 3 2011. ISBN 978-1-4244-4127-3. doi: 10.1109/ISBI.2011.5872394.

Ernst H. K. Stelzer, Frederic Strobl, Bo-Jui Chang, Friedrich Preusser, Stephan Preibisch, Katie McDole, and Reto Fiolka. Light sheet fluorescence microscopy. *Nature Reviews Methods Primers*, 1:73, 11 2021. ISSN 2662-8449. doi: 10.1038/s43586-021-00069-4.

Sternberg. Biomedical image processing. *IEEE Computer*, 16:22–34, 1 1983. ISSN 0018-9162. doi: 10.1109/MC.1983.1654163.

Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu. Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods*, 18:100–106, 1 2021. ISSN 1548-7091. doi: 10.1038/s41592-020-01018-x.

Daniel Sánchez-Gutiérrez, Melda Tozluoglu, Joseph D Barry, Alberto Pascual, Yanlan Mao, and Luis M Escudero. Fundamental physical cellular constraints drive self-organization of tissues. *The EMBO Journal*, 35:77–88, 1 2016. ISSN 0261-4189. doi: 10.15252/embj.201592374.

Simon Tanaka, David Sichau, and Dagmar Iber. Lbibcell: a cell-based simulation environment for morphogenetic problems. *Bioinformatics*, 31:2340–2347, 7 2015. ISSN 1367-4811. doi: 10.1093/bioinformatics/btv147.

Yucheng Tang, Dong Yang, Wenqi Li, Holger R Roth, Bennett Landman, Daguang Xu, Vishwesh Nath, and Ali Hatamizadeh. Self-supervised pre-training of swin transformers for 3d medical image analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20730–20740, 2022.

G. Taubin. Curve and surface smoothing without shrinkage. pages 852–857. IEEE Comput. Soc. Press, 1995. ISBN 0-8186-7042-8. doi: 10.1109/ICCV.1995.466848.

Alejandro Torres-Sánchez, Max Kerr Winter, and Guillaume Salbreux. Interacting active surfaces: A model for three-dimensional cell aggregates. *PLOS Computational Biology*, 18:e1010762, 12 2022. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1010762.

Gerard J. Tortora and Bryan Derrickson. *Principles of anatomy & physiology / Gerard J. Tortora, Bryan Derrickson.* Wiley, Hoboken, NJ, 15th edition, global edition edition, 2017. ISBN 9781119399933.

Stéfan 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*, 2:e453, 6 2014. ISSN 2167-8359. doi: 10.7717/peerj.453.

Roman Vetter, Marco Kokic, Harold Gómez, Leonie Hodel, Bruno Gjeta, Antonella Iannini, Gema Villa-Fombuena, Fernando Casares, and Dagmar Iber. Aboave-weaire's law in epithelia results from an angle constraint in contiguous polygonal lattices. 2019. doi: 10.1101/591461. URL `https://doi.org/10.1101/591461`.

Pablo Vicente-Munuera, Pedro Gómez-Gálvez, Robert J Tetley, Cristina Forja, Antonio Tagua, Marta Letrán, Melda Tozluoglu, Yanlan Mao, and Luis M Escudero. Epigraph:

an open-source platform to quantify epithelial organization. *Bioinformatics*, 36:1314–1316, 2 2020. ISSN 1367-4803. doi: 10.1093/bioinformatics/btz683.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, Aditya Vijaykumar, Alessandro Pietro Bardelli, Alex Rothberg, Andreas Hilboll, Andreas Kloeckner, Anthony Scopatz, Antony Lee, Ariel Rokem, C. Nathan Woods, Chad Fulton, Charles Masson, Christian Häggström, Clark Fitzgerald, David A. Nicholson, David R. Hagen, Dmitrii V. Pasechnik, Emanuele Olivetti, Eric Martin, Eric Wieser, Fabrice Silva, Felix Lenders, Florian Wilhelm, G. Young, Gavin A. Price, Gert-Ludwig Ingold, Gregory E. Allen, Gregory R. Lee, Hervé Audren, Irvin Probst, Jörg P. Dietrich, Jacob Silterra, James T Webber, Janko Slavič, Joel Nothman, Johannes Buchner, Johannes Kulick, Johannes L. Schönberger, José Vinícius de Miranda Cardoso, Joscha Reimer, Joseph Harrington, Juan Luis Cano Rodríguez, Juan Nunez-Iglesias, Justin Kuczynski, Kevin Tritz, Martin Thoma, Matthew Newville, Matthias Kümmerer, Maximilian Bolingbroke, Michael Tartre, Mikhail Pak, Nathaniel J. Smith, Nikolai Nowaczyk, Nikolay Shebanov, Oleksandr Pavlyk, Per A. Brodtkorb, Perry Lee, Robert T. McGibbon, Roman Feldbauer, Sam Lewis, Sam Tygier, Scott Sievert, Sebastiano Vigna, Stefan Peterson, Surhud More, Tadeusz Pudlik, Takuya Oshima, Thomas J. Pingel, Thomas P. Robitaille, Thomas Spura, Thouis R. Jones, Tim Cera, Tim Leslie, Tiziano Zito, Tom Krauss, Utkarsh Upadhyay, Yaroslav O. Halchenko, and Yoshiki Vázquez-Baeza. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, 17: 261–272, 3 2020. ISSN 1548-7091. doi: 10.1038/s41592-019-0686-2.

Andong Wang, Qi Zhang, Yang Han, Sean Megason, Sahand Hormoz, Kishore R. Mosaliganti, Jacqueline C. K. Lam, and Victor O. K. Li. A novel deep learning-based 3d cell segmentation framework for future image-based disease detection. *Scientific Reports*, 12:342, 1 2022. ISSN 2045-2322. doi: 10.1038/s41598-021-04048-3.

Chun-Chao Wang, Leen Jamal, and Kevin A. Janes. Normal morphogenesis of epithelial tissues and progression of epithelial tumors. *WIREs Systems Biology and Medicine*, 4: 51–78, 1 2012. ISSN 1939-5094. doi: 10.1002/wsbm.159.

Dong Wang, John D. Treado, Arman Boromand, Blake Norwick, Michael P. Murrell,

Mark D. Shattuck, and Corey S. O'Hern. The structural, vibrational, and mechanical properties of jammed packings of deformable particles in three dimensions. *Soft Matter*, 17:9901–9915, 2021. ISSN 1744-683X. doi: 10.1039/D1SM01228B.

Steffen Wolf, Constantin Pape, Alberto Bailoni, Nasim Rahaman, Anna Kreshuk, Ullrich Köthe, and Fred A. Hamprecht. *The Mutex Watershed: Efficient, Parameter-Free Image Partitioning*, pages 571–587. 2018. doi: 10.1007/978-3-030-01225-0_34.

Adrian Wolny, Lorenzo Cerrone, Athul Vijayan, Rachele Tofanelli, Amaya Vilches Barro, Marion Louveaux, Christian Wenzl, Sören Strauss, David Wilson-Sánchez, Rena Lymbouridou, Susanne S. Steigleder, Constantin Pape, Alberto Bailoni, Salva Duran-Nebreda, George Bassel, Jan U. Lohmann, Miltos Tsiantis, Fred A. Hamprecht, Kay Schneitz, Alexis Maizel, and Anna Kreshuk. Accurate and versatile 3d segmentation of plant tissues at cellular resolution. *eLife*, 9:1–34, 7 2020. ISSN 2050084X. doi: 10.7554/eLife.57613.

Liming Wu, Alain Chen, Paul Salama, Seth Winfree, Kenneth W. Dunn, and Edward J. Delp. Nisnet3d: three-dimensional nuclear synthesis and instance segmentation for fluorescence microscopy images. *Scientific Reports*, 13:9533, 6 2023. ISSN 2045-2322. doi: 10.1038/s41598-023-36243-9.

Bingfang Xu, Angela M. Washington, Raquel Fantin Domeniconi, Ana Cláudia Ferreira Souza, Xiaowei Lu, Ann Sutherland, and Barry T. Hinton. Protein tyrosine kinase 7 is essential for tubular morphogenesis of the wolffian duct. *Developmental Biology*, 412: 219–233, 4 2016. ISSN 00121606. doi: 10.1016/j.ydbio.2016.02.029.

Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3d: A modern library for 3d data processing. 1 2018.

Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation. 6 2016. URL `http://arxiv.org/abs/1606.06650`.

A | Appendix A

A.0.1. Ethics statement

The use of animals was approved by Canton Basel-Stadt (license number 2777/33495). Experimental procedures were performed in accordance with the Guide for the Care and Use of Laboratory Animals and approved by the Ethics Committee for Animal Care of ETH Zurich. All animals were housed at the D-BSSE/UniBasel facility under standard water, feeding, enrichment, and 12-hour light/dark cycles.

A.0.2. Mouse model

We isolated the epithelial tissues from E17.5 mice homozygous for Rosa26mTmG and heterozygous for the ShhGFP-Cre allele (ShhGFP-Cre/+; Rosa26mTmG). The epithelial cell membranes are labeled with GFP following CRE-mediated excision (Muzumdar et al., 2007). Offspring from Shhcre x mT/mG breedings show membrane-targeted tdTomato (mT) in all tissues, but have the mT cassette deleted for tissues expressing shh and therefore express membrane-targeted EGFP (mG).

A.0.3. Immunofluorescence staining and optical clearing

The isolated tissues were fixed in 4% paraformaldehyde (PFA) in PBS (28908, Thermo Fisher, brand name) for 1.5 hours at 4 deg. C. Samples were washed 4 times with PBSTx (0.25% Triton X-100 in PBS) following fixation. Samples were incubated with CUBIC-1 clearing solution for 3 days at 37 deg C. Following the first clearing step, samples were blocked for 2 days in 1% BSA (catalog number, brand name A7906, Sigma), 10% goat serum (S26-100ml, Sigma) in PBSTx at 4 deg. C on a shaker. The samples were incubated with the primary antibodies, anti-GFP (1:500, GFP-1020, aves) and anti-laminB1 (1:200, 702972, Thermo Fisher), for 2 days at 4 deg. C. The primary antibodies were diluted in a blocking buffer. After primary antibody incubation, the samples were washed 3 times with PBSTx. The samples were incubated with secondary antibodies, goat anti-chicken conjugated with Alexa Fluor 488 (1:200, A11039, Invitrogen), donkey anti-rabbit

conjugated with Alexa Fluor 647 (1:200, A32795, Invitrogen), and DAPI (1:2000, 1mg/ml stock solution, D9542, Sigma) for 3 days at 4 deg. C. The secondary antibodies were diluted in a blocking buffer. After secondary antibody incubation, the samples were washed 3 times for 15 minutes in PBSTx at 4 deg. C. Samples were fixed a second time in 4% PFA in PBS for 30 minutes at 4 degrees C. Samples were washed 4 times in PBSTx for 15 minutes at 4 deg. C. The samples were incubated in 50% CUBIC-2 solution diluted in ddH2O for 1 day. Then, they were then incubated in 100% CUBIC-2 until fully cleared (depending on sample size). During the 100% CUBIC-2 incubation, the CUBIC-2 solution was exchanged every other day. Finally, they were immersed in a freshly prepared CUBIC-2 solution. All CUBIC-2 incubations were performed at room temperature.

A.0.4. Imaging of whole mount epithelial tissues

Cleared and stained samples were embedded in 2% agarose cylinders and incubated in CUBIC-2 for 2 days. Images were acquired with a Zeiss Z.1 light sheet microscope using a 20X 1.0NA Plan Neofluar objective (421459-9970, Zeiss).

B | Appendix B

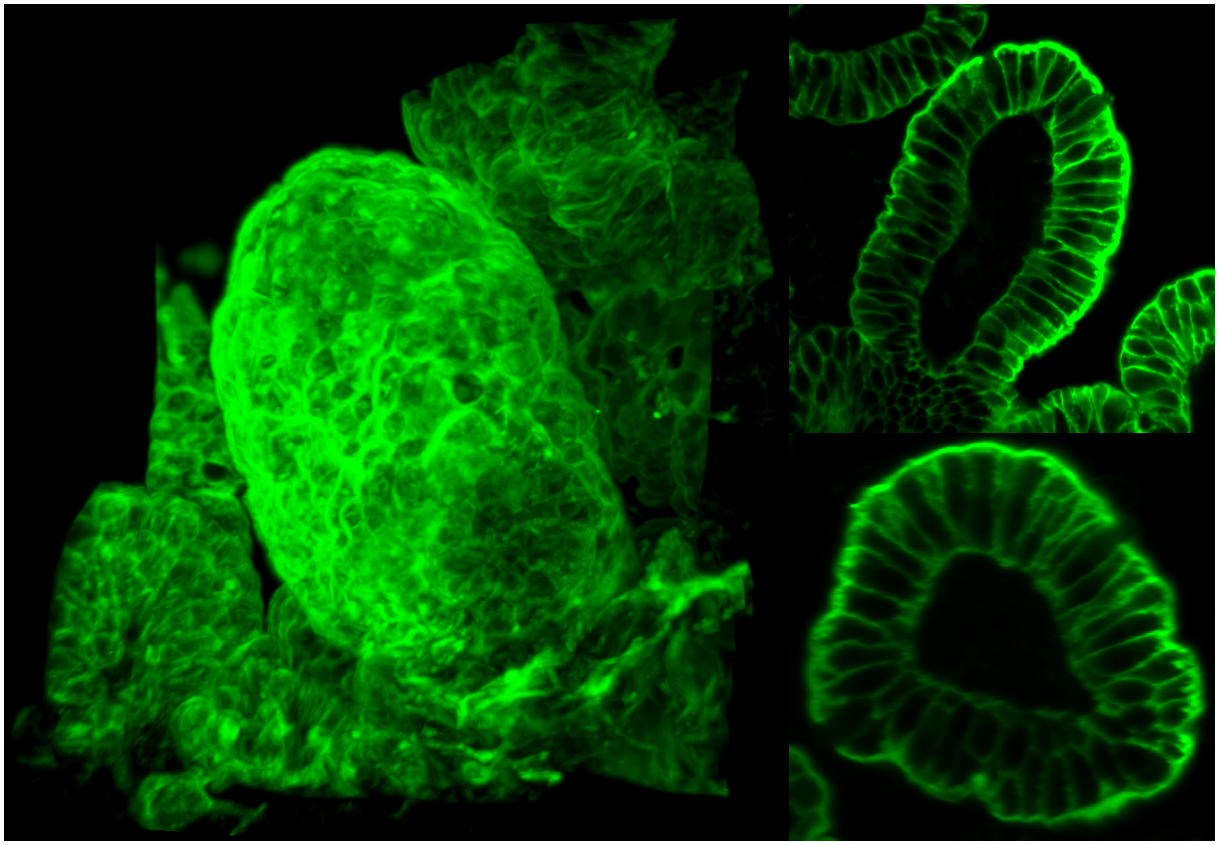

Figure 19: **3D rendering and 2D slices of the intestine villus epithelial tissue sample.**

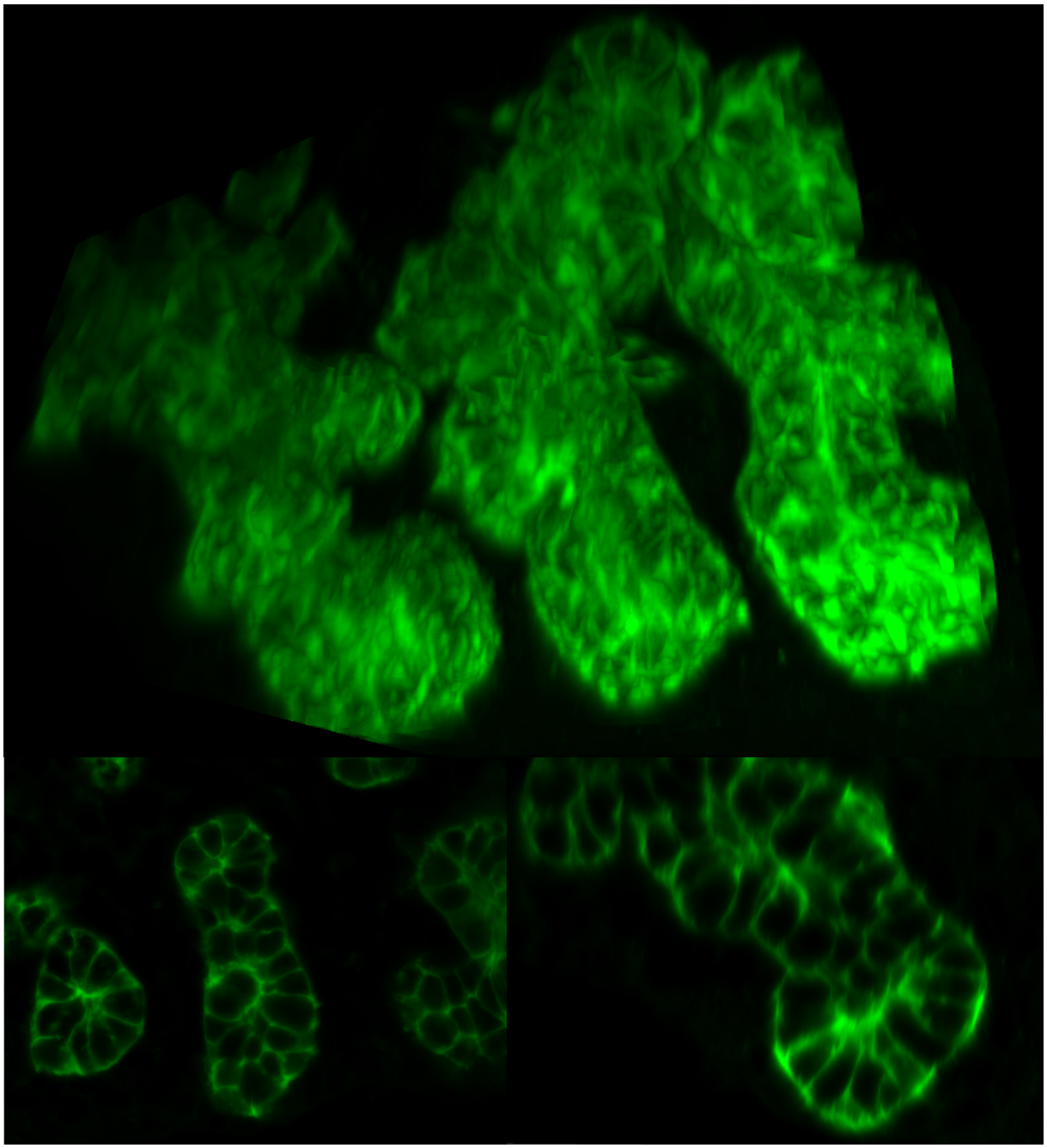

Figure 20: **3D rendering and 2D slices of the bronchiole epithelial tissue sample.**

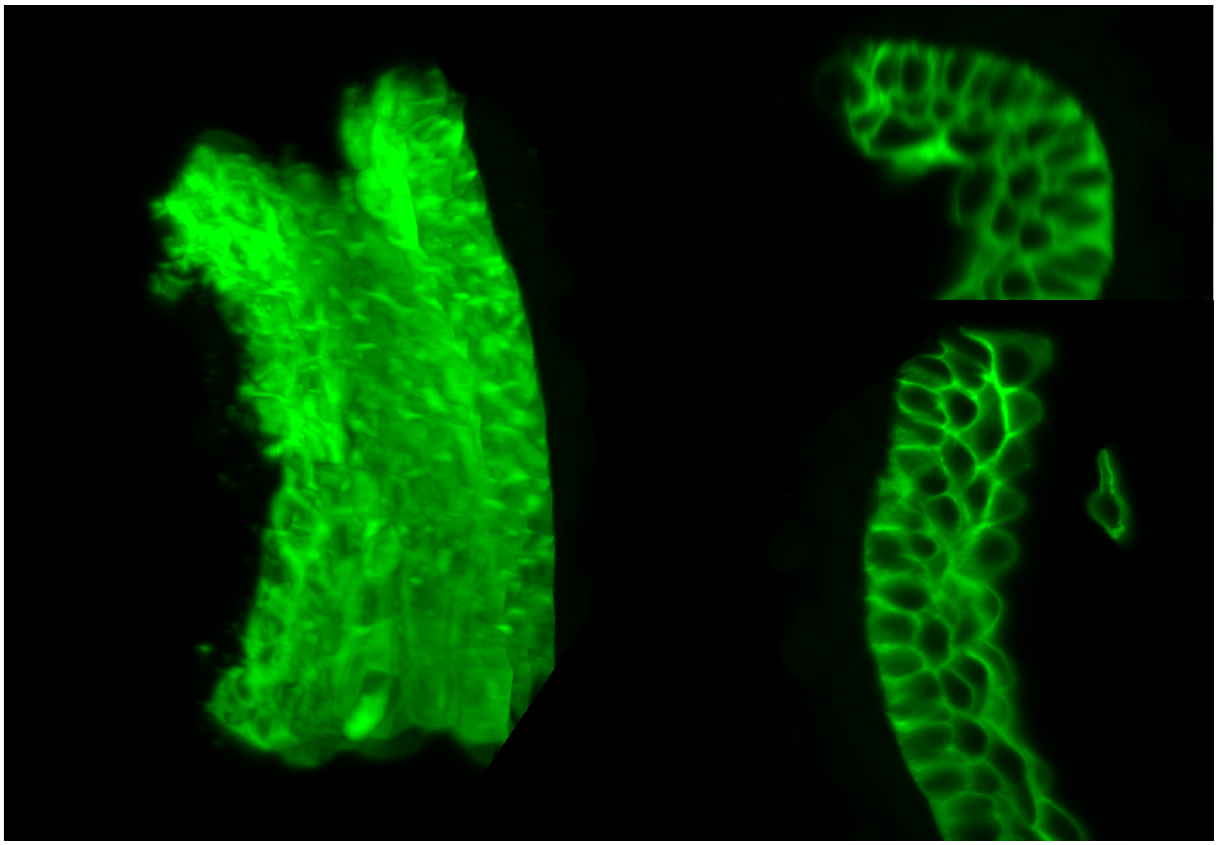

Figure 21: **3D rendering and 2D slices of the esophagus epithelial tissue sample.**

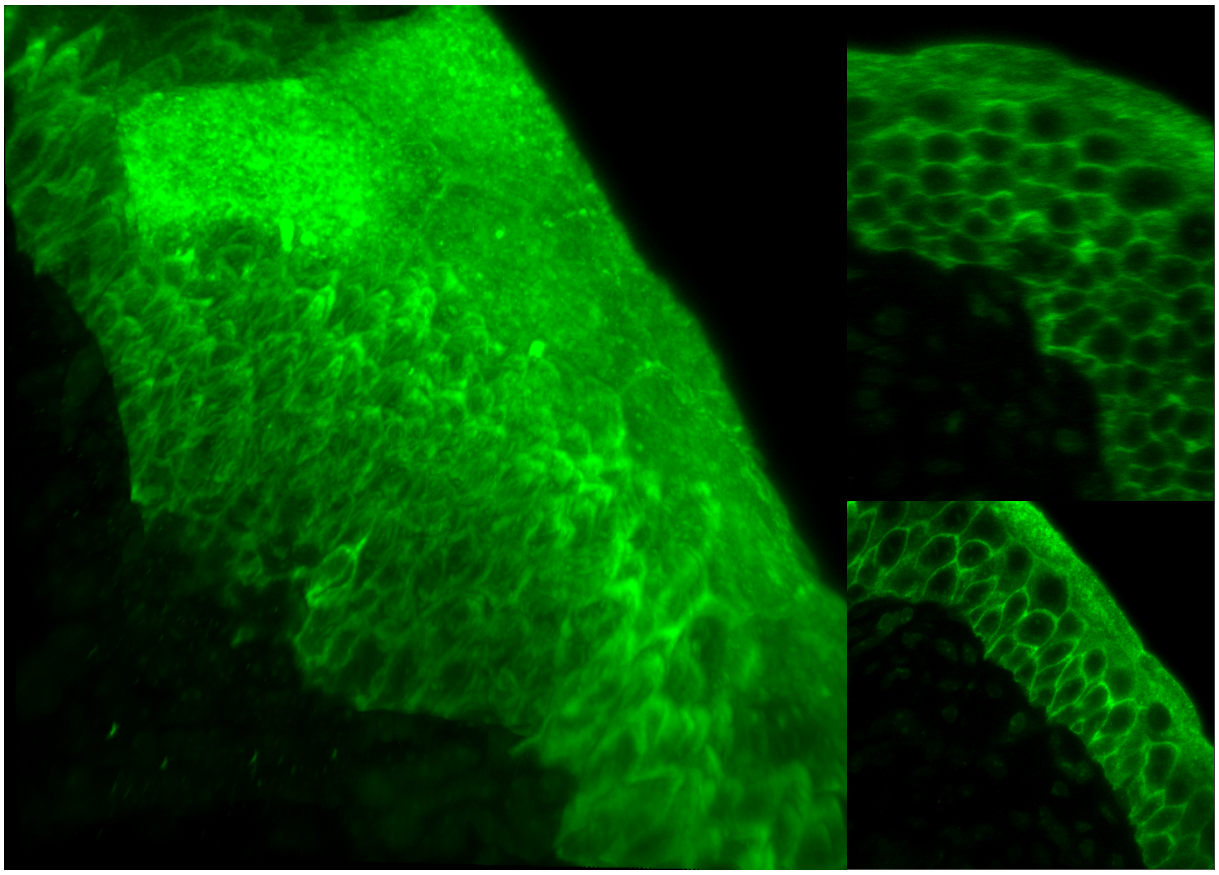

Figure 22: **3D rendering and 2D slices of the bladder epithelial tissue sample.**

List of Figures

List of Tables

List of Symbols

Variable	Description	SI unit
n	Number of neighbors	–
$\overline{A_n}$	Average area for cells with neighbors	m^2
$\overline{A}$	Average area of cells	m^2
m_n	Average number of second-order neighbors	–
$\mathcal{L}$	Loss function	–
$\mathcal{L}_{BCE}$	Binary Cross-Entropy Loss	–
$\mathcal{L}_{Dice}$	Dice Loss	–
p_i	Predicted voxel binary value	–
g_i	Ground-truth voxel binary value	–
N	Number of segmented cells in image	–
M	Total number of voxels in image	–
V	Average number of vertices in cell meshes	–
U	Potential energy	J
$\partial\Omega$	Cell surface	–
H	Local mean curvature of cell membrane	–
K	Bulk modulus	Pa
V_0	Target volume	m^3
A_0	Target surface area	m^2
Q_0	Target isoperimetric ratio	–
k_a	Area elasticity modulus	J
γ	Surface tension	N/m
k_b	Bending stiffness	J
ω	Adhesion strength	Pa/m
ξ	Repulsion strength	Pa/m

Variable	Description	SI unit
m	Mass	kg
ρ	Mass density	kg/m^3
$\vec{f_i}$	Force acting on node i	N
$\vec{r_i}$	Position of node i	m
$\vec{p_i}$	Linear momentum of node i	$kg \times m/s$
ζ	Viscous damping coefficient	kg/s
Δt	Time step	s
l_{min}	Minimum edge length	m
l_{max}	Maximum edge length	m
c	Contact cutoff distance	m
T	Simulation duration	s
IoU	Intersection over union	$-$

Acknowledgements

As I type these words, I am marking the conclusion of my master's study. First and foremost, I want to thank my advisor Prof. Dagmar Iber for having hosted me in the Computational Biology Lab at ETH D-BSSE during my master's thesis project. I will take her precious feedback as well as the words of esteem that she said to me at the end of my project as fuel for the rest of my career. My deepest gratitude goes also to my amazing thesis supervisor Dr. Kevin Yamauchi. The pieces of advice that I received by working alongside such a talented, well-organized, and sympathetic person will certainly help me through the challenges that I will face during my professional life. A special thanks to Steve Runser, whose calm and steady support allowed me to preserve my sanity in periods of frustration due to things not working. In general, I express my sincere appreciation to all the members of Prof. Iber's team for the inspiring and collaborative working environment they fostered. Thanks to their kindness, encouragement, and sympathy they made me feel at ease from the first day.

Special acknowledgments go to my advisor at Politecnico di Milano, Prof. Francesca Ieva. I genuinely appreciate the fact that she managed to take the time to supervise the writing process during a special period of her life.

To my parents, I express my profound gratitude for the fundamental support they gave me to reach this achievement. I (literally) would not be here if it wasn't for you.

I sincerely thank all my family members and friends, who always gave me the motivation to reach more and more ambitious goals. In particular, my most sincere gratitude goes to my aunt Lice for the invaluable support and time that she dedicated to me throughout these years and to my lovely grandmother Angela, who has been an outstanding role model for me since I was a little kid.

Finally, to my girlfriend Chicca, who always encouraged me in my decisions and backed me in challenging times with sweet words and warm gestures. I hope I can offer you the same support you showed me and that we can spend the next chapters of our lives together.