

Queen Mary
University of London

**Tissue architecture and cell division patterns in human
brain organoids**

Thomas Bennett

August 2024

Supervisors: Natalia Bulgakova & Isabel Palacios

This report was submitted as partial fulfilment of a MSc degree in
Bioinformatics

Abstract

To address the increasing prevalence of neurological disorders, a deeper understanding of early brain development is crucial for identifying underlying causes and developing effective treatments. Brain organoids have become an invaluable tool for modelling early human brain development with greater accuracy than traditional models. However, the quantification of these organoids using 2D histology images presents several challenges. Manual quantification is labour-intensive, time-consuming, and prone to bias, often failing to capture the full complexity of the data. To overcome these challenges, this study developed an automated pipeline that integrates traditional descriptors with deep learning methods, streamlining the analysis and enhancing quantification accuracy and reliability. All code for this pipeline can be found at <https://github.com/TomBennett2202/NeuroOrganoidAnalysis>.

The pipeline extracts a comprehensive range of parameters, including area, eccentricity, circularity, solidity, and mean intensity for both whole organoids and nuclei, as well as the number of nuclei and their distance from the centre of the organoid. It then conducts extensive statistical analyses. First, a pairwise correlation analysis reveals significant relationships among organoid and nuclei metrics. Second, a comparison is made between organoid and nuclei metrics in organoids with and without a lumen in present in the image. Finally, principal component analysis (PCA) is applied to reduce the dimensionality of the organoid and nuclei metrics, aiding in the identification of distinct patterning. Furthermore, the analysis focuses on exploring cell proliferation within the organoids by examining differences among metrics for mitotic nuclei, non-mitotic nuclei, and recognised artefacts, which are hypothesised to be apoptotic cells due to their small size and brightness—characteristics of chromatin condensation during apoptosis. A CNN was employed to identify mitotic nuclei, non-mitotic nuclei, and artefacts. While the CNN model demonstrated promising accuracy at 68.18% in distinguishing between these categories, there remains significant room for improvement.

Future work should prioritise expanding the image dataset for both mitotic training and exploratory data analysis, as this study utilised only 18 organoids and 41 images per nuclei class. Increasing the number of images will significantly improve model accuracy and the reliability of identified trends. Additionally, advancing lumen classification using deep learning techniques, transitioning from 2D to 3D analysis, and applying the pipeline to organoids modelled after neurological disorders are crucial next steps.

Acknowledgements

I would like to extend my deepest gratitude to my supervisors, Dr Natalia Bulgakova and Dr Isabel Palacios, for their invaluable guidance and support throughout my dissertation. Their expertise, insights, and encouragement have been pivotal to the success of this project. I am also immensely grateful to Raeesah Khan for her unwavering support and encouragement through every challenge, and without her I would have not made it this far. Finally, I wish to express my profound thanks to my family for their constant love and encouragement throughout this journey.

Abstract.....	2
Acknowledgements	2
1.0 Introduction.....	5
1.1 Burden of Neurological Disorders	5
1.2 Early Brain Development	5
1.3 Previous Model Systems.....	7
1.4 Brain Organoids: A Novel Model System.....	7
1.5.0 Quantitative Analysis of Organoids.....	9
1.5.1 Manual Quantification	10
1.5.2.0 Automated Quantification.....	10
2.0 Aims.....	12
3.0 Methods.....	12
3.1 Image preprocessing	12
3.2 Segmentation.....	13
3.3.0 Analysis of Mitotic Nuclei.....	13
3.3.1 Dataset extraction.....	13
3.3.2 Model training.....	14
3.4.0 Pipeline for Quantification and Analysis of Organoids	14
3.4.1 Comparative Analysis of Organoids Based on Lumen Presence.....	14
3.4.2 Quantification of Nuclei	14
3.4.3 Quantification of Whole Organoids	15
3.4.4.0 Exploratory data analysis	15
4.0 Results	16
4.1 Segmentation.....	16
4.3 Training model.....	18
4.4 Optimising the pipeline	19
4.5.0 Exploratory data analysis	22
4.5.1 Pairwise Correlation Analysis.....	22
4.5.2 Analysis of Organoids Based on Lumen Presence	24
4.5.3 Analysis of Mitotic Nuclei	26
4.5.4 Principal Component Analysis	27
5.0 Discussion.....	28
5.1 Segmentation.....	29
5.2.0 Pipeline for Quantification and Analysis of Organoids	29
5.2.1 Normalisation.....	29
5.2.2 Pairwise Correlation Analysis.....	30
5.2.3 Comparative Analysis of Organoids Based on Lumen Presence.....	31
5.2.4 Principal Component Analysis	31
5.2.5.0 Analysis of Mitotic Nuclei.....	31
5.3.0 Future Work	33
5.3.1 Principal Component Analysis	33

5.3.2 Future Directions for Mitotic Model Training	33
5.3.3 Advancing Lumen Classification with Deep Learning.....	33
5.3.4 Transitioning from 2D to 3D Analysis	34
5.3.5 Extending Analysis to Neurological Disorder Models	34
6.0 Conclusion	34
7.0 References	35
8.0 Appendix.....	38
8.1 Appendix A	38
8.2 Appendix B	68

1.0 Introduction

1.1 Burden of Neurological Disorders

Neurological disorders are the leading cause of disabilities and the second largest cause of deaths globally (Feigin *et al.*, 2020). Over the last 30 years, the absolute number of deaths has increased by 39%, and disability-adjusted life-years (DALYs) have risen by 15% (Feigin *et al.*, 2020). This significant and growing burden is driven by global population growth and increasing life expectancy and is projected to continue to rise (Gu *et al.*, 2021). This increase suggests that current preventative and treatment measures are insufficient to counteract these global demographic changes.

1.2 Early Brain Development

There is growing evidence that abnormal brain development during gestation contributes to many lifelong neurological disorders. For instance, schizophrenia, one of the most debilitating mental disorders affecting approximately 1% of the population, cannot be explained solely by genetic inheritance (Rees and Inder, 2005). It has been proposed that prenatal insults, combined with genetic risk factors, can lead to the development of schizophrenia symptoms in young adults (Rees and Inder, 2005).

Foetal brain development begins after the third week of gestation with the expansion of the neuroepithelium, which originates from the ectoderm (Konkel, 2018). Neuroepithelial (NE) cells will divide symmetrically to form the neural plate, which then invaginates to create the neural tube (see Figure 1). The neural tube is lined by a polarised pseudostratified neuroepithelium, with its basal side attached to the pial surface and its apical side contacting a fluid-filled lumen (Barry *et al.*, 2014). Like most developing epithelia, mitosis takes place at the surface, neighbouring the ventricle. To reach the right location for cell division, the nuclei of neuroepithelial cells must undergo a directed movement towards the apical side of the neuroepithelium, known as interkinetic nuclear migration (INM), a process that is essential throughout neurogenesis (Benito-Kwiecinski *et al.*, 2021). However, before neurogenesis begins, NE cells undergo a transformation into radial glia RG), shifting from symmetric, proliferative division to asymmetric, self-renewing division while maintaining vertical division planes (Benito-Kwiecinski *et al.*, 2021).

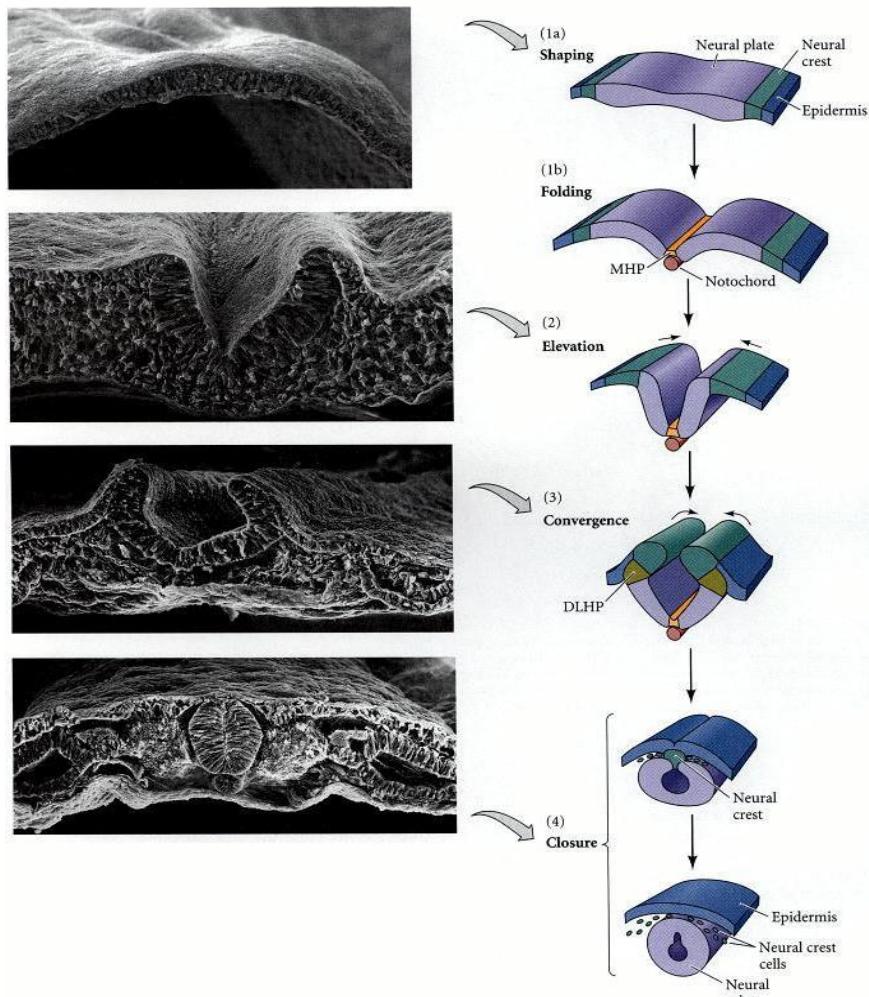


Figure 1. Neural tube formation in the chick embryo (taken from Gilbert, 2000). The neural plate cells in the dorsal ectoderm elongate and fold as medial neural hinge point (MHP) cells anchor to the notochord. Neural folds converge as dorsolateral hinge point (DLHP) cells become wedge-shaped and epidermal cells push inward. The folds meet, neural crest cells link the neural tube to the epidermis, and then disperse.

These early stages of CNS development are crucial, as they set the foundation for future brain growth and function. Disruptions in these processes can have significant implications for brain architecture and behaviour. In particular, the transition from NE cells to RG cells is very important for the development of the human brain. In humans and apes, this process takes several days, while in mice it occurs rapidly, within only a few hours (Benito-Kwiecinski *et al.*, 2021). This difference in timing results in a significantly higher production of neurons, contributing to the human brain's 1,000-fold increase in neuron number compared to that of mice (Benito-Kwiecinski *et al.*, 2021). During this transition in humans and apes, there is an additional intermediate cell morphological state known as transitioning NE (tNE) (see Figure 2). In this state, cell morphology changes precede the change in cell identity and the onset of neurogenesis. Consequently, it is suggested that these changes in cell morphology are linked to cell cycle, through INM, and a delay in this transition in humans results in a larger founder progenitor pool and thus an increase in neuron number (Benito-Kwiecinski *et al.*, 2021). This disparity between species underscores the necessity of understanding the cellular and molecular mechanisms that drive early brain development, as such knowledge is essential for advancing our understanding of neurological disorders.

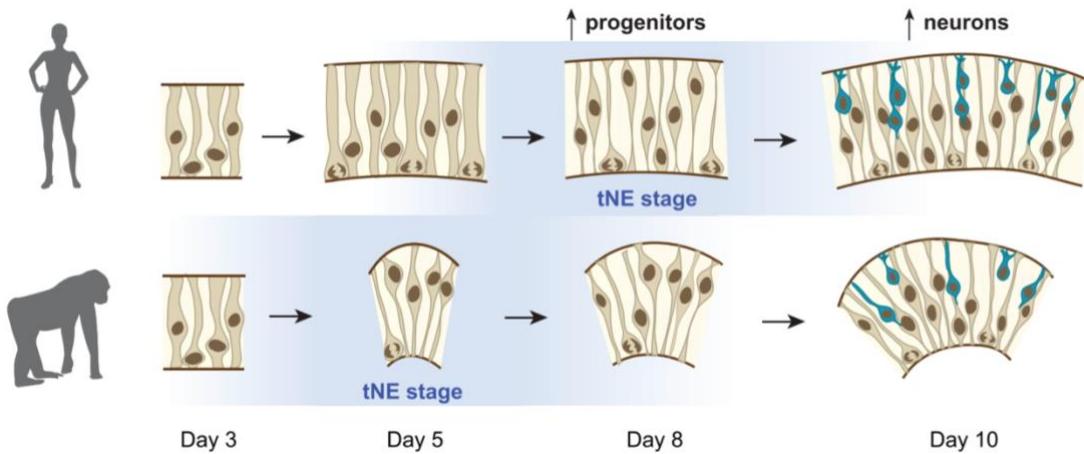


Figure 2. Schematic overview of the morphological changes in neural progenitor cells (taken from Benito-Kwiecinski *et al.*, 2021). Nonhuman ape cells exhibit tNE morphologies (blue background) earlier than human cells, which instead have shorter cell cycles, resulting in an increased number of neurons.

1.3 Previous Model Systems

Current knowledge on the human brain and neurological disorders primarily stems from post-mortem examinations and animal models. Post-mortem studies allow researchers to directly examine the brain in detail (Kim and Chang, 2023). However, this approach has significant limitations. Obtaining CNS tissue from individuals with certain disorders or at specific developmental stages is not always feasible. Moreover, post-mortem studies only offer a snapshot of the brain at the time of death, missing the opportunity to observe the temporal sequence of developmental changes and pathological events.

To address the limitations of studying post-mortem human CNS tissue, animal models have played a vital role in neuroscience research. These models have provided invaluable insights into the genetic origins of certain diseases and have enhanced our understanding of pathological processes at the cellular level (Kim and Chang, 2023). Animal studies have been instrumental in uncovering the mechanisms of neurodevelopment and neurodegeneration, identifying potential therapeutic targets, and testing the efficacy and safety of new drugs. For instance, researchers at the German Centre for Neurodegenerative Diseases in Berlin have developed a promising gene therapy for Alzheimer's disease. Using a gene editing tool called "zinc fingers" they successfully reduced levels of tau, a key protein that accumulates in the brain during Alzheimer's, potentially paving the way for effective treatments (Wegmann *et al.*, 2021).

Animal models, while invaluable, often fail to fully recapitulate the complexity of human brain development and disease (Adams *et al.*, 2019). There are significant differences in brain architecture, function, and development between humans and commonly used animal models, such as mice. These disparities have led to numerous failures in CNS drug development, as drugs that showed promise in animal studies often did not achieve the same in human clinical trials (Kim and Chang, 2023). Furthermore, the ethical implications of using animals in research cannot be overlooked.

1.4 Brain Organoids: A Novel Model System

Given the limitations of previously used models, developing methods that accurately recapitulate paradigms of human brain development *in vitro* holds immense potential. As a result, recent advancements in stem cell technology have led to the development of new models for studying the human brain. Human embryonic stem cells (hESCs) can be induced into neural stem cells and subsequently differentiated into neurons and glia (Adams *et al.*, 2019). However, due to ethical

concerns surrounding their origin, their use has been limited. Alternatively, cellular reprogramming offers a solution to this by converting a patient's somatic cells into a pluripotent state, creating induced pluripotent stem cells (iPSCs) (see Figure 3) (Adams *et al.*, 2019). iPSCs retain the patient's genotype and, like hESCs, can differentiate into various cell types, including neurons and glia. Consequently, iPSCs offer the opportunity for direct in vitro manipulation of cellular phenotypes relevant to nervous system disorders.

Two-dimensional (2D) cultures of these cells are not indicative of 'real-world' environments. It does not provide a comprehensive understanding of their development and functionality within the human body, where they neighbour other cells in three dimensions (3D). Comparatively, 3D cell culture involves enhanced cell-to-cell interactions, cell division, and morphology, which are far more closely resembling those *in vivo*. This led to the initial discovery of what would become organoids by the Sasai laboratory as they observed that ESCs and iPSCs could self-organise replicating *in vivo* development (Nakano *et al.*, 2012). This was initially observed in the formation of optic tissue; however, organoids of other neural fates display similar organisation, gene expression profiling, topographical induction, and temporal development to that seen during foetal neural development (Adams *et al.*, 2019).

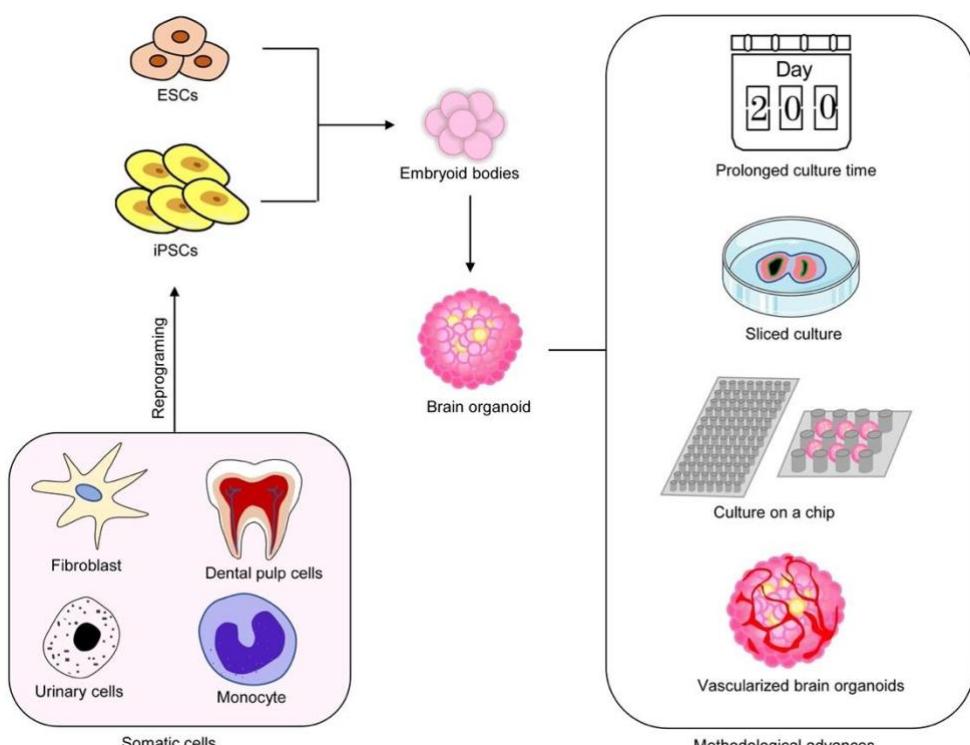


Figure 3. Schematic of the methods used for brain organoid creation (taken from Shou *et al.*, 2020). The process involves cellular reprogramming from somatic cells to induced pluripotent stem cells (iPSCs), which, along with embryonic stem cells (ESCs), can be utilised to generate brain organoids.

Early-stage brain organoids are highly effective at modelling the early stages of human brain development. In a cross section or z-stack of a brain organoid, cells organise into neural rosettes, with a central lumen surrounded by cells that undergo mitosis at the luminal surface. This organisation closely mimics the structure of *in vivo*, where NE cells and RG are arranged radially around a fluid-filled ventricle (see Figure 4). Additionally, neural progenitors within organoids exhibit a temporally regulated pattern of neuronal production, demonstrated by the sequential appearance of molecular markers that define distinct neuronal layers which can generate action potentials, form synapses, and, when grafted into the cortex of mice, they integrate successfully with existing cortical neurons in a

layer-specific manner (Kelava and Lancaster, 2016). This finding underscores an intrinsic mechanism within *in vitro*-derived neural progenitor cells that allows them to replicate critical developmental milestones in neocortical neuron generation. Moreover, brain organoids also display INM and a complex configuration of neural progenitors, including bRG-like cells characterised by typical morphology and molecular signatures such as PAX6 expression (Kelava and Lancaster, 2016). By accurately mimicking the early stages of human brain development, these models offer a valuable opportunity to investigate how disruptions in the foundational processes that take place may impact cell behaviour and organisation. This, in turn, enables researchers to gain critical insights into the potential causes of neurological disorders, as well as use organoids modelled after neurological disorders to track and evaluate treatment efficacy.

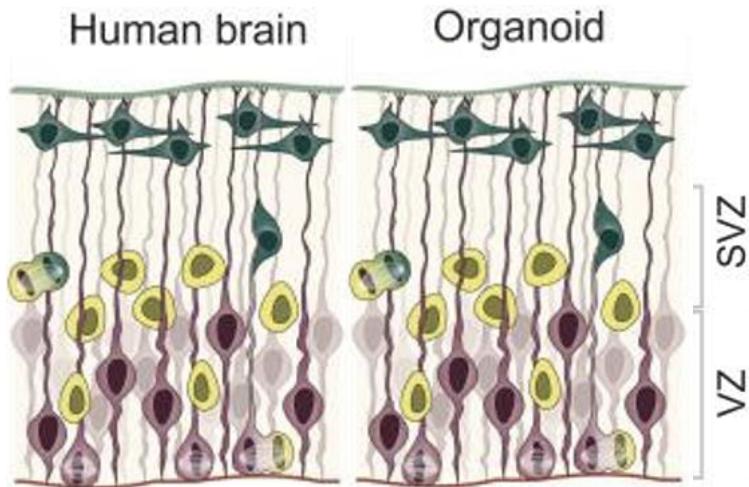


Figure 4. A comparison of in vivo and in vitro brain development. This schematic provides a simplified representation illustrating the cellular complexity in the developing brain *in vivo* and in brain organoids generated *in vitro* (adapted from Kelava and Lancaster, 2016).

1.5.0 Quantitative Analysis of Organoids

The analysis of brain organoids predominantly relies on single-cell transcriptome analysis and 2D histology. Recent advancements in single-cell sequencing technologies have expanded the ability to analyse organoid cultures, enabling detailed identification of cell-type composition and offering new opportunities to investigate brain development, cellular diversity, and gene regulation in otherwise inaccessible human cells (Patikas *et al.*, 2023). 2D histology involves measuring various morphological, structural, and functional properties to understand their development, differentiation, and response to experimental conditions. It is common that researchers investigate organoid morphology and size as these are the first signs of health and disease. To help assess the growth and expansion rate of an organoid culture, the size of the organoids can be tracked over time (Haja *et al.*, 2023). Monitoring this proliferative status can offer valuable information into the developmental state of the organoid, its regenerative capacity, and its response to pharmacological interventions (Haja *et al.*, 2023).

It is important to note that the way these structures are cultured often results in significant variability in the size and morphology of organoids within the same dome. The main factor contributing to this variability is the distribution of organoids within the hydrogel domes (see Figure 5). Since these domes are three-dimensional, not all organoids receive equal access to nutrients from the medium (Haja *et al.*, 2023). Organoids located near the edge of the hydrogel have easier access to nutrients, allowing them to grow larger than those situated toward the centre of the dome.

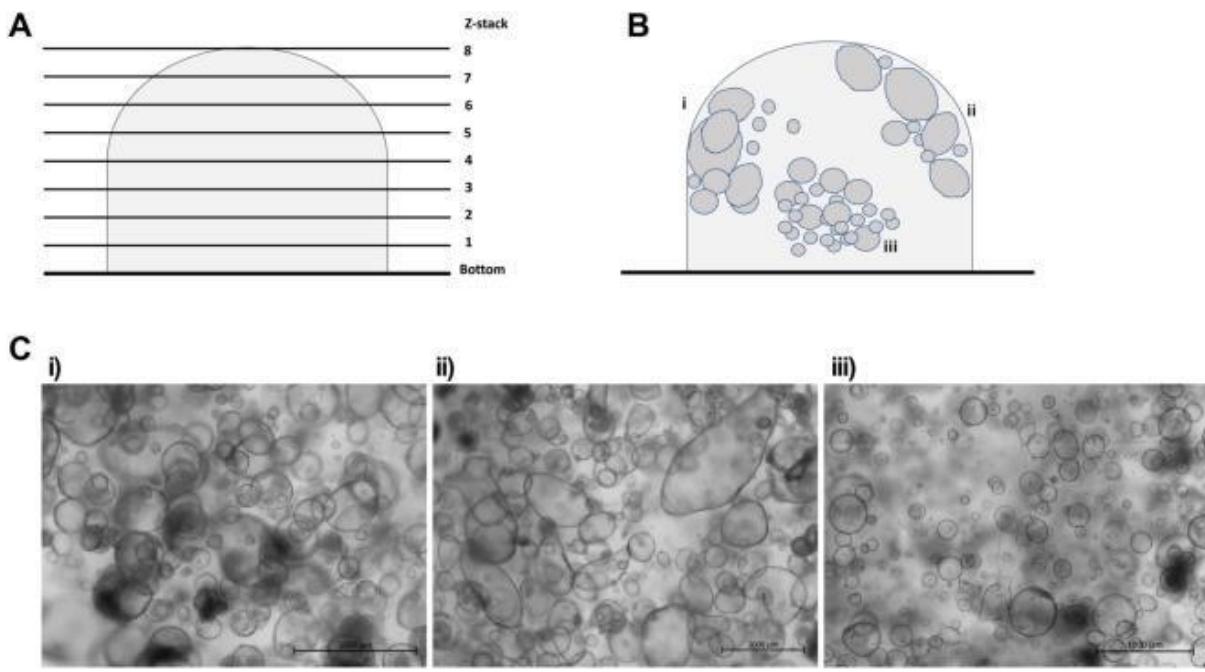


Figure 5. Organoids exhibit high variability in size and morphology within the same culture (taken from Haja *et al.*, 2023). (A) Schematic of a hydrogel dome with line bars representing different Z-stacks within the dome. (B) Schematic illustrating the distribution of organoids within the hydrogel dome: (i) Organoids accumulate in one area of the dome, resulting in overlap; (ii) Organoids near the edge of the dome develop into larger, more complex morphologies; (iii) Organoids in the centre of the dome tend to be smaller and more spherical. (C) Representative brightfield images corresponding to the different distributions described in (B).

1.5.1 Manual Quantification

Due to the high variability between organoids within the same dome, researchers must quantify a large number of organoids per well to get an accurate depiction of what is happening in the culture. Manual evaluation is typically conducted through a visually guided, labour-intensive, and time-consuming process, taking several hours depending on density (Haja *et al.*, 2023). As a result, this can lead to significant variability between observers often due to tiredness and personal interpretation. Andriom *et al.* (1995) assessed the consistency of histopathological diagnoses in 88 cases of pleural malignant mesothelioma by comparing the observations of five pathologists. The study found that only 70% of the diagnoses were consistently reproduced by the panel review.

1.5.2.0 Automated Quantification

Alternatively, an automated method for quantifying organoids offers a standardised approach, eliminating the potential for any human bias. This will enable rapid and reliable analysis of organoid images, facilitating precise modelling of disease phenotypes, offering detailed insights into organ development that may be overlooked by manual evaluation, and tracking the effects of drug treatments on diseased organoids. High-throughput analysis can allow for the rapid screening of multiple drug candidates, enabling the identification of effective treatments quicker.

1.5.2.1 Segmentation

The first step in imaging pipelines is image segmentation, a crucial process that isolates specific areas of interest from the background and other non-relevant parts of the image. By accurately delineating these segments, researchers can then quantify attributes, such as size and shape. In the context of

organoids this could entail identifying and delineating the boundaries of individual organoids or nuclei within an organoid.

Various techniques are employed for image segmentation, ranging from traditional methods like thresholding and edge detection to more advanced approaches such as machine learning and deep learning algorithms, depending on the complexity and requirements of the analysis. One notable example is Cellpose, a versatile deep-learning-based segmentation algorithm (Stringer *et al.*, 2021). Cellpose's popularity is primarily due to its ability to accurately segment a wide range of image types "right out-of-the-box", without the need for model retraining or parameter adjustments. Cellpose was trained on an extensive dataset comprising highly varied images of cells, which includes over 70,000 segmented objects, and periodically is retrained using community-contributed data to ensure continuous improvement and adaptability (Stringer *et al.*, 2021).

1.5.2.2 Traditional Descriptors

When quantifying complex biological samples such as organoids, researchers use either traditional descriptors or machine learning techniques. Traditional descriptors are quantitative measures that derive from analysing image features using established mathematical and statistical methods. These include metrics such as size, shape, intensity, and texture, computed through predefined algorithms and rules. They are straightforward to compute, requiring neither extensive training data nor complex algorithms, and are easy to interpret as they directly relate to the physical properties of the objects being analysed.

However, selecting the appropriate descriptors can be challenging and often requires domain expertise. It may involve trial and error to determine which features best characterise different object classes, and incorrect feature selection can lead to suboptimal analysis outcomes (O'Mahony *et al.*, 2020). As the number of classes increases, feature extraction becomes increasingly cumbersome (O'Mahony *et al.*, 2020). Moreover, traditional descriptors tend to offer a limited view, focusing on basic geometric and intensity properties while potentially overlooking more complex patterns and relationships. They may not be well-suited for analysing intricate data patterns that require a more holistic approach.

1.5.2.3 Machine Learning

Machine learning has revolutionised image analysis by automating and refining pattern recognition. This technology involves training systems on extensive datasets with known outputs, enabling them to learn and recognise patterns. Once trained, these systems can analyse and interpret new, unseen data with high accuracy. A key advancement in this field is deep learning, a specialised branch of machine learning that leverages Artificial Neural Networks (ANNs). Modelled after the human brain, these networks consist of numerous interconnected 'neurons' that perform simple tasks and work together to make complex decisions (Voulodimos *et al.*, 2020). Through end-to-end learning, deep learning excels at processing information across multiple layers of neural networks. Each layer in this hierarchical structure refines and builds upon the features identified by the previous layer (Elgendi, 2020). Initial layers may detect basic features, such as edges and textures, while deeper layers capture more complex patterns and higher-level abstractions, like shapes and objects (Elgendi, 2020). This layered approach enables the network to develop a progressively detailed understanding of the input data, homing in on the most relevant features. Each layer enhances its ability to recognise increasingly intricate features, ultimately leading to a comprehensive representation of the data.

By learning directly from the data, deep learning significantly improves both the accuracy and efficiency of image analysis. Additionally, as neural networks are trained rather than programmed, unlike traditional descriptors, deep learning requires less expert intervention and can be re-trained for various applications with custom datasets (O'Mahony *et al.*, 2020). On the other hand, deep learning

may be excessive for some tasks. Traditional descriptors often provide a more efficient solution, requiring fewer lines of code and less computational power (O’Mahony *et al.*, 2020). While deep learning thrives with large datasets, typically involving millions of records and significant computational resources, traditional methods can be highly effective when extensive data or high computing facilities are unavailable (Hassaballah and Awad, 2020). When training datasets are limited, deep learning models may be overfit to the training data, struggling to generalise effectively. Additionally, fine-tuning a deep neural network can be challenging due to its millions of parameters and their complex interrelationships. This has led to criticisms of deep learning models as ‘black boxes,’ where understanding and adjusting the model’s behaviour can be particularly difficult (O’Mahony *et al.*, 2020).

Hybrid approaches, which combine traditional descriptors with deep learning techniques, harness the strengths of both methodologies and address the limitations inherent in using either approach in isolation. By integrating the efficiency and simplicity of traditional descriptors with the powerful, data-driven insights of deep learning, these approaches can improve computer vision performance.

2.0 Aims

As global population growth and increasing life expectancy continue, the prevalence of neurological disorders is projected to rise significantly. To combat this trend, a deeper understanding of early brain development is essential for identifying the underlying causes and developing effective treatments. Brain organoids have emerged as a valuable tool for modelling early human brain development more accurately than traditional models. However, quantifying these organoids through 2D histology images poses challenges. Manual quantification is labour-intensive, time-consuming, prone to bias, and often fails to capture the full complexity of the organoid data.

This study aims to overcome these limitations by developing a high-throughput automated approach that combines traditional descriptors with deep learning methods to quantify organoid architecture and behaviour. This robust and standardised process will eliminate human bias, significantly enhancing the accuracy and efficiency of organoid quantification. Furthermore, the pipeline will facilitate subsequent analyses, enabling the accurate modelling of even the smallest disease phenotypes and providing detailed insights into organ development that may be overlooked through manual evaluation.

3.0 Methods

3.1 Image preprocessing

Microscopic images of brain organoids stained with Phalloidin and DAPI were provided by the Palacios lab in Leica Image File (LIF) format. Each image in the LIF file contains several channels, with each channel highlighting different components of the brain organoid. One channel represents the DNA in the brain organoid, while another represents actin (see Figure 6). Using Fiji (ImageJ) (<https://imagej.net/software/fiji/>), the DNA and actin channels were merged and saved as a new JPEG image (see Figure 6). The magnification of each image was also noted using Fiji. This, along with the corresponding image file name, was recorded in a CSV file.

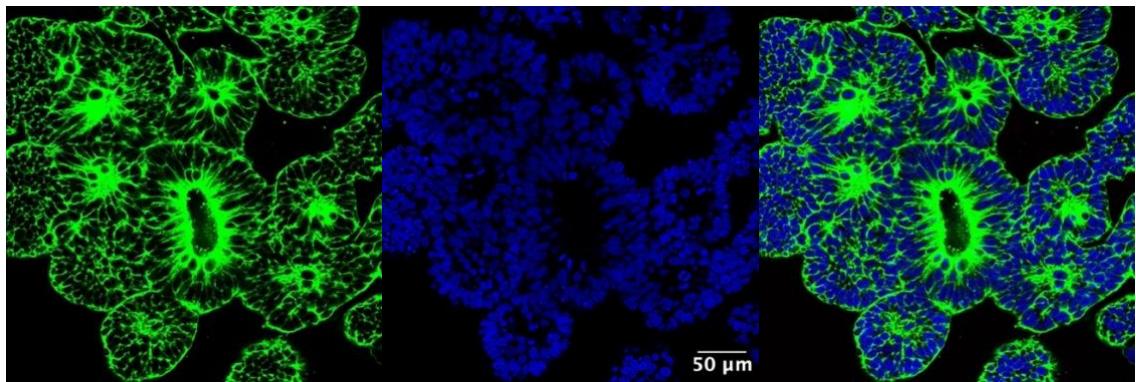


Figure 6. Channels of stained organoids. The left panel displays the actin staining in green. The middle panel shows the DNA staining in blue. The right panel presents the merged image.

3.2 Segmentation

The images were loaded into Cellpose (<https://github.com/mouseland/cellpose>), where all pre-trained models were tested to identify the most effective one for organoid segmentation. After determining that the “cyto2_cp3” model performed best, the images were segmented using this model. To address any initial inaccuracies, the segmented organoids were manually corrected, and the model was retrained using the corrected segmentation. The refined segmentation was saved as a PNG organoid mask, with each label representing an individual organoid. This process was repeated for each image.

For nuclei segmentation, as with organoid segmentation, pre-trained models were evaluated to identify the most effective one for nuclei segmentation. After determining “cyto3” was the best performing model, the images were segmented using this model. Initial inaccuracies were manually corrected, and the model was retrained using the corrected segmentation. The refined segmented nuclei were saved as a PNG nuclei mask, with each label representing an individual nucleus.

3.3.0 Analysis of Mitotic Nuclei

3.3.1 Dataset extraction

To explore nuclei within the organoids, two scripts were created. The first script stores images of the nuclei in a directory called “nuclei_training,” which contains three subdirectories: “mitotic,” “non_mitotic,” and “miscellaneous”. The script iterates through the images in the “images” directory, extracting each file name to retrieve the corresponding nuclei mask from the “nuclei_masks” directory. Nuclei touching the border are removed, and objects in the mask relabelled. The user is then presented with a pop-up displaying the DNA channel and asked to select nuclei that are dividing. The coordinates of the selected points are used to extract the nucleus index from the nuclei mask and if the selected point was not part of the background, the index was stored in an array.

Using the nucleus index, a binary mask was generated to isolate the specific nucleus of interest by setting pixels corresponding to that index to “1” and all other pixels to “0.” The bounding box of this nucleus mask was extracted using regionprops, and the image was cropped to isolate the nucleus. The cropped image was resized to 22x22 pixels and saved with a timestamp in the “mitotic” subdirectory. This process was repeated, ensuring an equal number of selections, for non-mitotic nuclei, saving them in the “non_mitotic” subdirectory, and for miscellaneous objects, saving them in the “miscellaneous” subdirectory. After saving the images, they were reviewed to ensure accuracy, with any segmentation errors identified and removed.

3.3.2 Model training

The second script was created to train a network on the previous images. The dataset for training and evaluating the neural network is organised in an imageDatastore object that reads images from the “nuclei_training” directory, including subdirectories, and uses these subdirectories names as labels for classification. The dataset is divided into training, validation, and test sets using a stratified split. This process randomly allocates 70% of the images to the training set and the remaining 30% to a combined validation and test set. The latter is then split evenly into validation and test sets.

The deep learning model consists of several layers. The architecture begins with an image input layer configured to accept images with dimensions of 22x22 pixels and a single colour channel (22x22x1). This is followed by a convolution 2D layer with 20 filters of size 5x5, which extracts features from the input images. A batch normalisation layer is included to stabilise and accelerate training by normalising the activations. The relu layer introduces non-linearity to the network, allowing it to learn complex patterns. The network ends with a fully connected layer that outputs predictions for the nuclei classes, a soft max layer that converts these predictions into probability distributions, and a classification layer for evaluating classification performance.

For training, the model employs stochastic gradient descent with momentum (SGDM) as the optimisation algorithm. The training is configured with a maximum of 10 epochs, and the learning rate is initially set to 1e-3. Training progress is monitored through visual plots, and validation is performed at a frequency of every 3 iterations using the validation dataset. The training process is executed with an automatically chosen execution environment and utilises mini batches of size 16.

Upon completion of the training, the model's accuracy is evaluated by classifying the test set and the model is saved as “trainedNetwork.mat”.

3.4.0 Pipeline for Quantification and Analysis of Organoids

The main pipeline starts by creating a structured array called “all_data” where image file names are used as fields to store the corresponding data. It then iterates through each microscopic image in the “images” directory, loading each one and extracts the file name of the current image to load their corresponding organoid and nuclei masks from the “organoid_masks” and “nuclei_masks” directories. Additionally, it accesses the corresponding magnification of the image from the “magnifications.csv” file. For each image the pipeline removes organoids and nuclei that touch the borders of their respective masks and then relabel the remaining objects in the masks.

3.4.1 Comparative Analysis of Organoids Based on Lumen Presence

To explore differences between organoids with and without a lumen within the z-slice, the pipeline provides two options. First, it checks for a CSV file for the current image in the "lumen_classification" directory. If found, it loads the organoid labels from the file. Alternatively, if no CSV file is found for the image, the user is prompted with a pop-up displaying the current microscopic image to manually select organoids with a lumen. The coordinates of the selected organoids are recorded and used to identify the corresponding organoid labels in the organoid mask. These labels are then saved in a CSV file within the "lumen_classification" directory.

3.4.2 Quantification of Nuclei

An empty array called “nuclei_data” is created to store a series of tables, where each cell represents an organoid and contains the data on the nuclei within it. The pipeline then iterates through each organoid within the organoid mask. For each organoid, it identifies the nuclei that fall within the

organoid by locating the pixels where the nuclei mask intersects with the organoid. It also identifies any nuclei outside the current organoid. The pipeline then iterates through each of these nuclei, adding it to a new mask if it falls within the organoid and does not overlap outside the organoid by verifying the label's absence in the nuclei outside the organoid. Additionally, artefacts smaller than 3 pixels are removed. This process continues until all relevant nuclei are included in the new mask and remaining nuclei are relabelled to be specific for each organoid. Additionally, as the pipeline iterates through the nuclei, it extracts the bounding box of each nucleus, which is then used to crop both the image and the nuclei mask. The cropped mask is applied to the corresponding cropped image, isolating only the nucleus. The image is then resized to 22x22 pixels and classified by the trained network, with the classification results saved to an array.

To ensure accurate measurements, the image intensity is first normalised for each organoid using the imadjust function before any data extraction. Then using the regionprops function, properties of the nuclei within the current organoid are extracted and saved in a table. The table includes the following nuclei properties: area, centroid, eccentricity, circularity, solidity, and mean intensity of DNA. To determine the distance of each nucleus from the centre of the organoid, regionprops is used to find both the centroid of the current organoid and the length of its minor axis. The Euclidean distance between each nucleus's centroid and the organoid's centroid is calculated using the formula in Figure 7. To normalise the distance per organoid, it is divided by the organoid's radius, which is derived from the length of the minor axis. To convert the area and the distance from the centre of the organoid from pixels to micrometres, the magnification factor for the current image is applied, and the measurements are updated in the table. The previously created array containing nuclei classifications is added to the table. Additionally, if the current organoid is among the lumen-containing organoid labels extracted from the CSV, it is classified accordingly within the table. This table is then added to the empty array “nuclei_data”. Moreover, the number of labels within the organoid’s nuclei mask is extracted to represent the total number of nuclei within the current organoid. After iterating through all the organoids, the array of tables is stored in the structured array “all_data” under the field corresponding to the current file name.

$$Distance = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Figure 7. Euclidean distance formula.

3.4.3 Quantification of Whole Organoids

The regionprops function is then applied to the entire organoid mask to extract whole organoid properties including area, centroid, eccentricity, circularity, solidity, and mean intensity of the actin. It is important to note prior to extraction, the image intensity of the whole image was also normalised using the imadjust function. These properties are compiled into a table called “organoid_data”. The area measurements are converted to micrometres from pixels, and the table is updated with these converted values. Additionally, the number of nuclei within each organoid, previously calculated, is recorded in the table. Furthermore, if the current organoid is among the lumen-containing organoid labels extracted from the CSV, it is classified accordingly within the table. Finally, this table is saved in the structured array “all_data” under the field corresponding to the current file name.

3.4.4.0 Exploratory data analysis

To explore the data, all data must first be consolidated. This initial step involves merging the data on whole organoids from all images into a comprehensive table. This table includes properties: area, eccentricity, circularity, solidity, mean intensity, number of nuclei, and lumen classification for each organoid. Subsequently, data from all nuclei across the images are combined into a separate table. This

table features metrics: area, eccentricity, circularity, solidity, mean intensity, and the distance from the organoid's centre, along with the lumen classification of the corresponding organoid and the mitotic classification of each nucleus. Additionally, the average nuclei data per organoid is calculated and aggregated across all images, then merged with the consolidated organoid table. This process results in a comprehensive table of organoid data, as well as average nuclei metrics for each the corresponding organoid.

3.4.4.1 Pairwise Correlation Analysis

Pairwise correlations are performed on the data, plotting each parameter against every other parameter. This analysis is conducted separately for whole organoid data, nuclei data, and the combined whole organoid data with nuclei averages, ensuring that comparisons are made appropriately without directly comparing organoids to organoids or nuclei to nuclei. For each pairwise correlation, Pearson correlation coefficient and corresponding p-value is calculated. If the correlation is found to be statistically significant ($p < 0.05$), the corresponding graphs, along with the correlation coefficient and p-value, are saved in a subdirectory named "correlated." Conversely, if the correlation was not significant ($p \geq 0.05$) the results are saved in a subdirectory named "uncorrelated."

3.4.4.2 Lumen analysis

For each parameter in the consolidated organoid and nuclei data, comparisons are visualised using swarm plots that separate the data by lumen classification. Additionally, p-values are calculated by a two-sample t-test and accompany the graphs in the title. These swarm plots are then saved in a directory named "lumen_graphs".

3.4.4.3 Mitotic analysis

For each parameter in the consolidated nuclei data, comparisons are visualised using swarm plots that separate the data based on their nucleus class (mitotic nuclei, non-mitotic nuclei, and miscellaneous). To assess differences between groups, a one-way Analysis of Variance (ANOVA) is conducted on the data, resulting in a p-value to assess overall nuclei class differences. Post-hoc comparisons are then performed using the Tukey's Honestly Significant Difference (HSD) test to identify which specific nuclei classes differed significantly from each other. The swarm plots, which include the p-value in the title, were saved in a directory named "mitotic_graphs." Additionally, the results of the post-hoc tests were documented in a CSV file within the same directory, detailing the mean differences, confidence intervals, and p-values for each class comparison.

3.4.4.4 Principal Component Analysis

For the combined whole organoid data, nuclei data, and combined organoid data with nuclei averages, principal component analysis (PCA) is conducted. The pipeline generates both 2D and 3D graphs, with the variance contribution percentage labelled on the axes. If the variance contribution for the second and third principal components is less than 0.01%, it is labelled as "< 0.01%". These graphs are saved in a directory named "pca_graphs".

4.0 Results

4.1 Segmentation

As seen in Figure 8, the initial segmentation using the model cyto2_cp3 demonstrates several issues. Multiple organoids were completely missed, and some segmentations did not accurately encompass the entire organoid or included extraneous parts of the image. The manual correction addressed these shortcomings, successfully segmenting the previously missed organoids and accurately refining the segmentations to exclude any non-organoid regions while fully containing each organoid. For

example, within the red box in Figure 8, the initial segmentation extends beyond the boundaries of the organoid and does not fully encompass it. Through manual correction, the segmentation was refined to accurately contain the entire organoid, ensuring no extraneous areas were included.

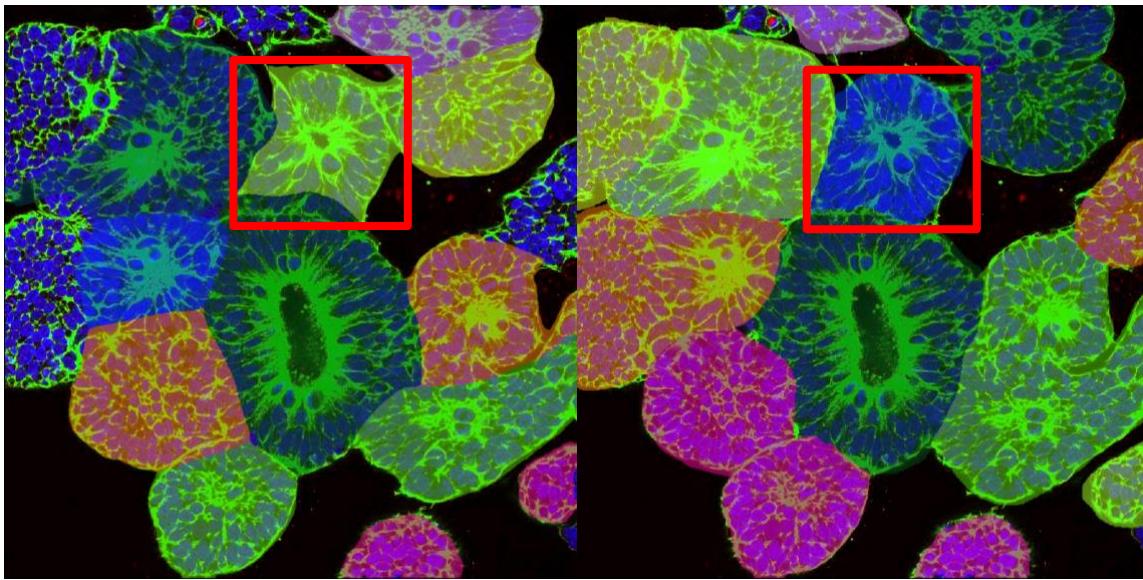


Figure 8. Manual correction of segmented organoids. The left image displays the initial segmentation, with a red box highlighting an organoid that has not been accurately refined. The right image shows the segmentation after manual correction, with the red box highlighting the refined organoid.

As seen in Figure 9, the initial segmentation using the cyto3 model demonstrates several issues. Multiple nuclei were completely missed, and some segmentations did not accurately encompass the entire nuclei or included extraneous parts of the image. This included instances where multiple nuclei were combined due to failure to identify boundaries between them. The manual correction addressed these shortcomings by successfully segmenting the previously missed nuclei and accurately refining the segmentations to correct the boundaries. For instance, within the red box in Figure 8, the model initially segmented what appeared to be two nuclei. However, upon closer examination, it was revealed to be a single mitotic nucleus, which was then correctly segmented manually.

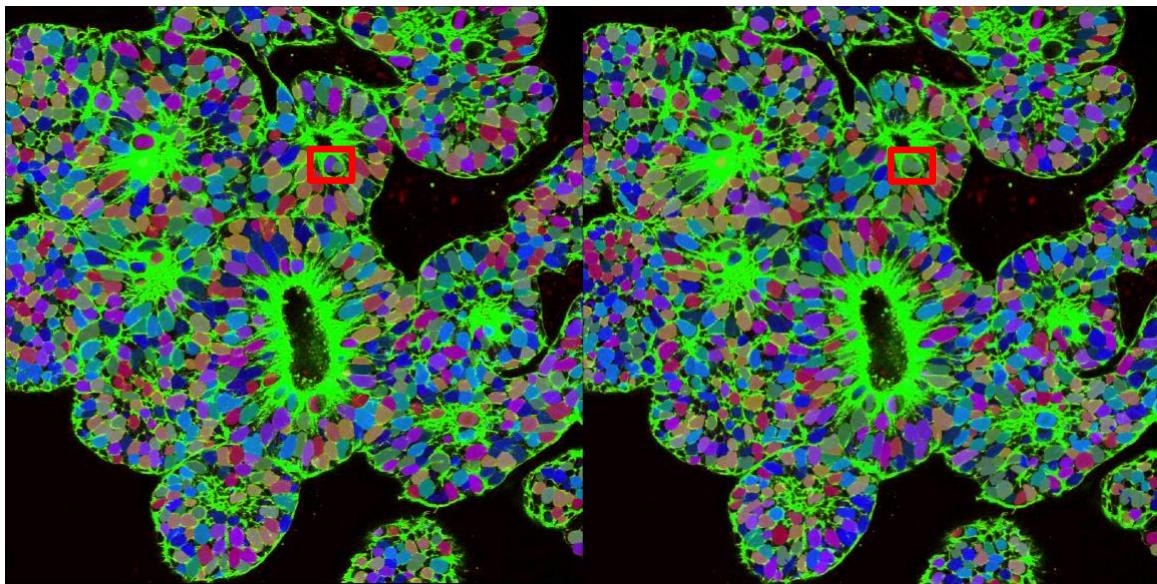


Figure 9. Manual correction of segmented nuclei. The left image displays the initial segmentation, with a red box highlighting two nuclei that had not been accurately segmented. The right image shows the segmentation after manual correction, with the red box highlighting the accurately segmented nucleus.

4.3 Training model

Figure 10 illustrates the training progress of the network designed to differentiate between mitotic and non-mitotic nuclei and other objects. The training was conducted over 10 epochs, with a total of 60 iterations, and completed in approximately 5 seconds using a single CPU.

The training accuracy improved steadily with each epoch, stabilising around epoch 8 at nearly 100%. In contrast, validation accuracy initially reached about 70% by the end of epoch 2 but then fluctuated between 60% and 70% for the rest of the training, gradually approaching the final validation accuracy of 69.57%.

Training loss decreased consistently, reaching approximately 0.2 by the end of the training. Validation loss also decreased but at a slower rate, stabilising around epoch 4 at approximately 0.6. When applied to the test dataset, the model achieved an accuracy of 68.18%.

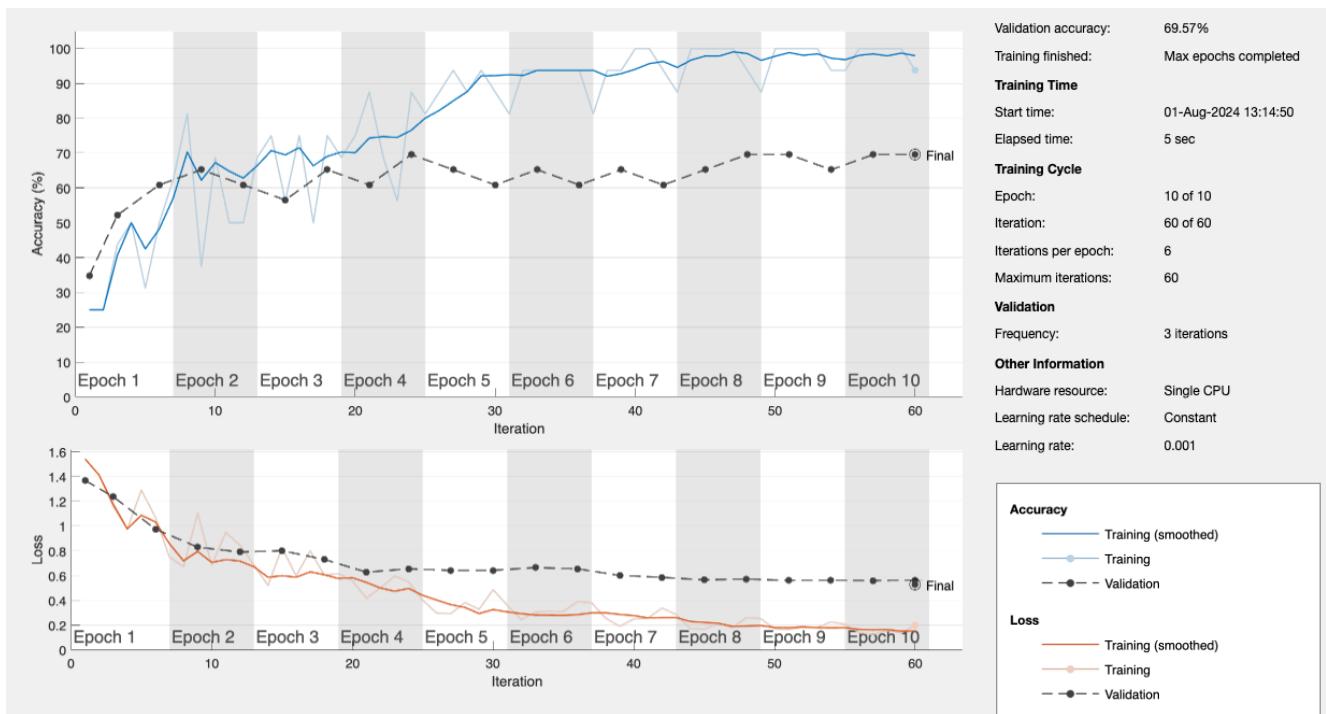


Figure 10. Training progress of the network for differentiating between mitotic and non-mitotic nuclei and miscellaneous objects. The training was conducted over 10 epochs, with metrics for accuracy and loss recorded at every iteration, while validation metrics were recorded every 3 iterations.

4.4 Optimising the pipeline

For all normalisations, a Mann-Whitney U test was conducted to compare the measurements before and after the normalisation process.

The intensity of the actin signal was adjusted, using imadjust, to ensure consistency and comparability across images. As you can see in Figure 11, there was no difference in normalising the mean intensity of the images.

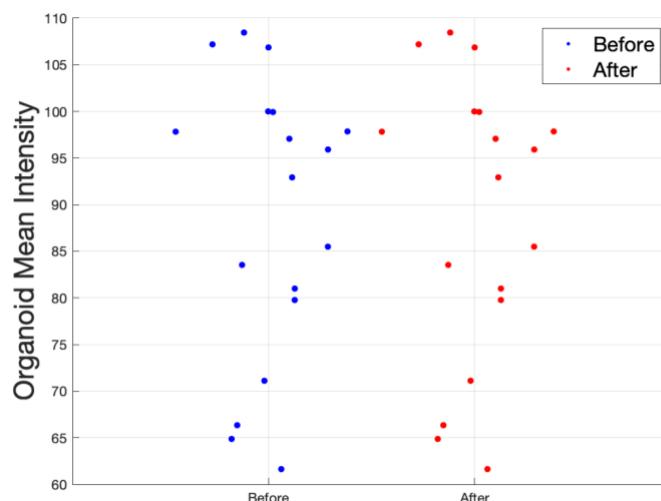


Figure 11. Normalisation of organoid image intensity. The figure displays the mean intensity of organoids before and after normalisation.

The intensity of the DNA signal for each organoid was adjusted using imadjust to ensure consistent and comparable nuclei intensity across all organoids. Figure 12 showcases the effect of normalising the nuclei intensity per organoid. In Figure 12a, it is evident that there is no difference in the intensity of image 1 before and after normalisation. Conversely, Figure 12b highlights a noticeable increase in intensity following normalisation, further supported by a highly significant p-value of 1.26e10–87. Moreover, the p-value between the two images before adjustment was 1.01e-88, while after adjustment, it improved to 0.049, indicating a more comparable mean intensity (see Figure 12c).

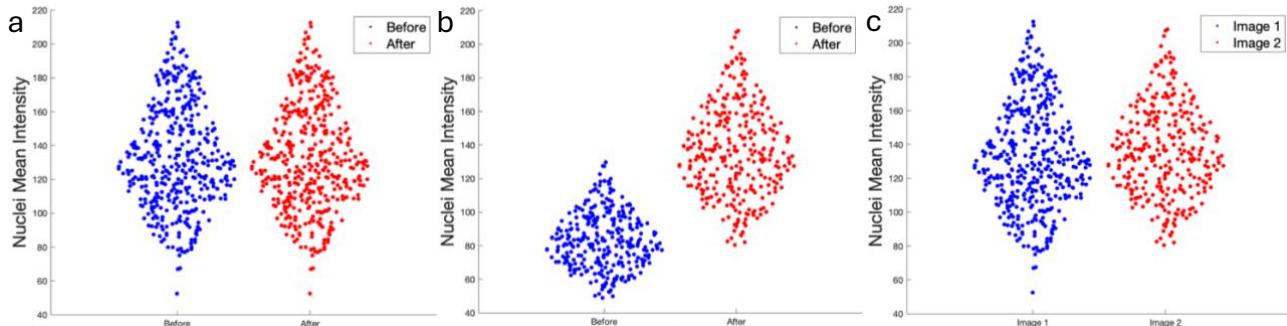


Figure 12. Normalisation of nuclei intensity per organoid. (a) Displays the nuclei intensity per organoid for Image 1 before and after normalisation. (b) Illustrates the nuclei intensity per organoid for Image 2 before and after normalisation. (c) Illustrates the nuclei mean intensity for Image 1 and 2 after normalisation.

To ensure consistent measurement of organoid area regardless of the magnification at which the image was captured, a normalisation process was applied. Specifically, the area of each organoid was adjusted by multiplying it by a factor corresponding to the magnification level. Figure 13 illustrates the differences in organoid area between Images 1 and 2 before and after conversion to micrometres. Initially, the area of Image 2 was significantly smaller than that of Image 1, with a p-value of 6.28e-05. After normalisation, Image 2 remained smaller, but the difference was less pronounced, reflected by a p-value of 0.011.

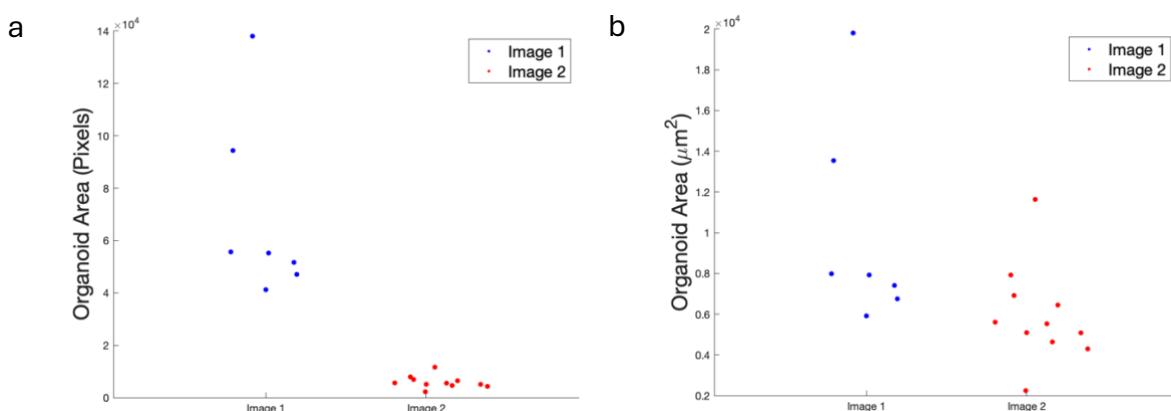


Figure 13. Normalisation of organoid area when converted to micrometres (a) Displays the organoid area, measured in pixels, for Image 1 and 2 before normalisation. (b) Illustrates the organoid area, measured in μm^2 , for Image 1 and 2 after normalisation.

A similar normalisation approach was applied to the nuclei area. As shown in Figure 14, the nuclei area in Images 1 and 2 differed significantly, before normalisation, with Image 2 being considerably smaller, as indicated by a p-value of 2.16e-103. Following normalisation, Image 2 remained smaller, though the difference was less pronounced, with a p-value of 1.31e-14.

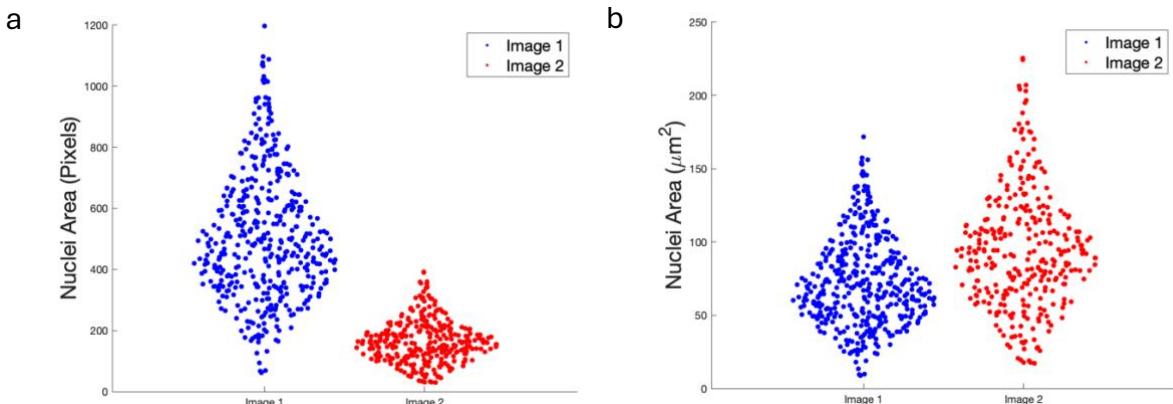


Figure 14. Normalisation of nuclei area when converted to micrometres (a) Displays the nuclei area, measured in pixels, for Image 1 and 2 before normalisation. (b) Illustrates the nuclei area, measured in μm^2 , for Image 1 and 2 after normalisation.

Furthermore, the same approach was applied to the distance of the nuclei from the centre of the organoid. As shown in Figure 15, the distance of nuclei from the centre of the organoid was significantly different between Images 1 and 2 before normalisation, with Image 2 showing a notably shorter distance, supported by a p-value of 1.76e-90. After normalisation, this difference persisted, though it was less pronounced, with a p-value of 1.80e-18.

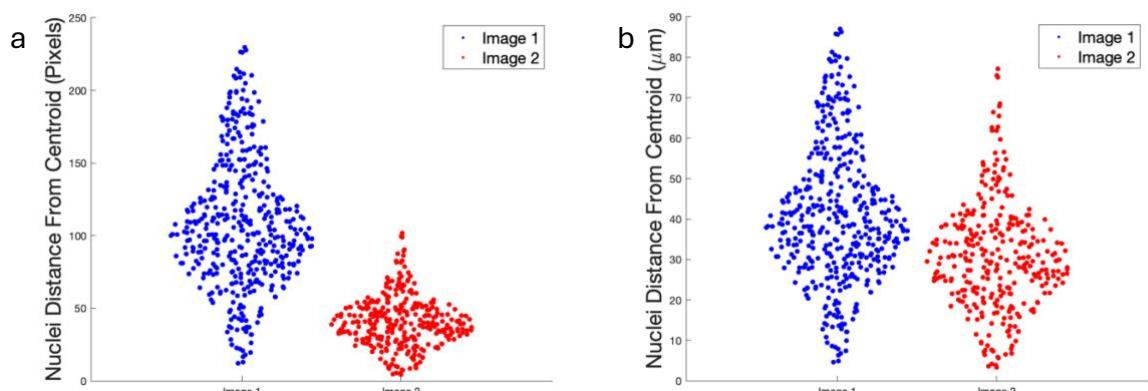


Figure 15. Normalisation of the distance of nuclei from the centre of the organoid, when converted to micrometres. (a) Displays the nuclei distance of from the centre of the organoid, measured in pixels, for Image 1 and 2 before normalisation. (b) Illustrates the nuclei distance from the centre of the organoid, measured in μm , for Image 1 and 2 after normalisation.

Larger organoids naturally exhibited greater distances from the centre, making it difficult to assess true spatial distribution, therefore, the distance was divided by the radius of the respective organoid. Figure 16 illustrates the variation in nuclei distance from the centre of the organoid before and after normalisation. In Figure 16a, normalisation significantly reduced the distance in Image 1, with the mean distance decreasing from 41.30 to 0.28, supported by a highly significant p-value of 9.33e-144. Similarly, in Figure 16b, the distance in Image 2 decreased from 30.85 to 0.61, with a significant p-value of 5.74e-103.

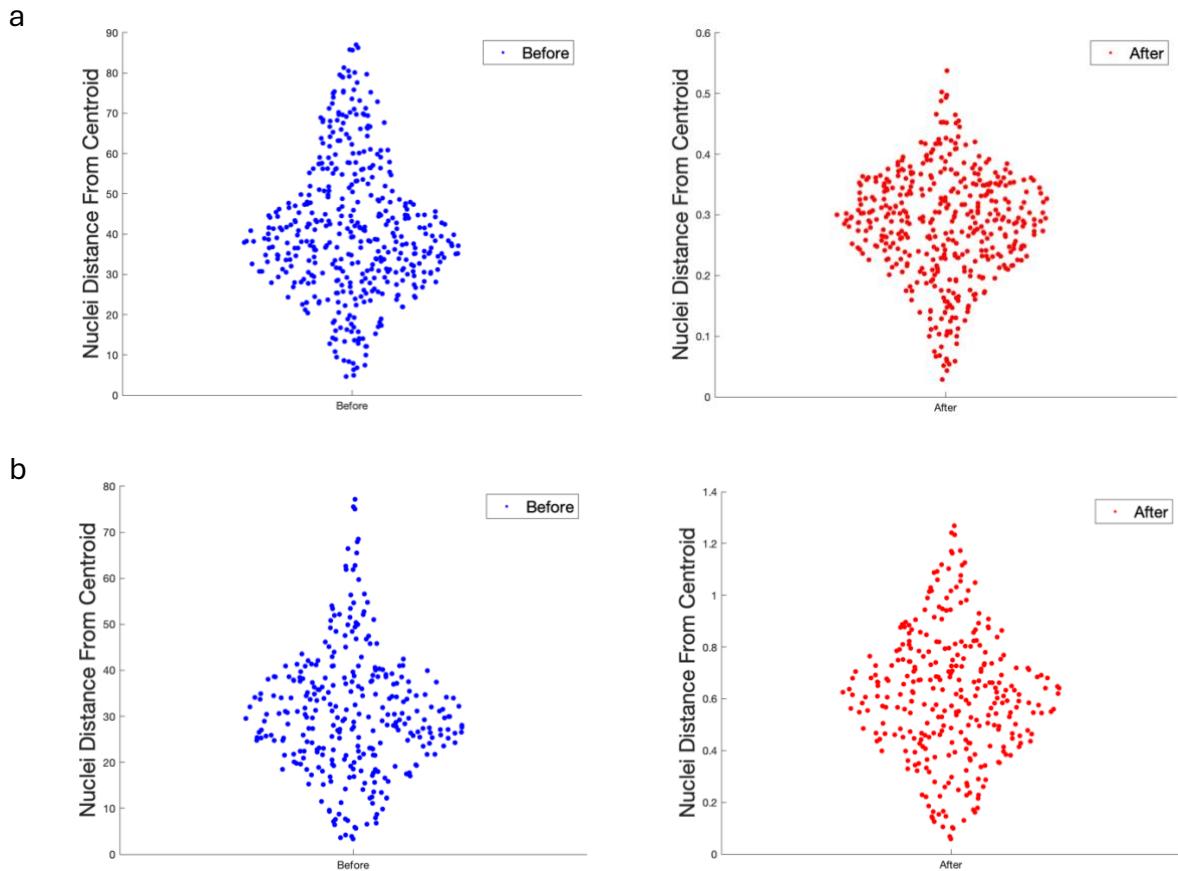


Figure 16. Normalisation of the distance of nuclei from the centre of the organoid. (a) Displays the nuclei distance of from the centre of the organoid for Image 1 before and after normalisation. (b) Illustrates the nuclei distance from the centre of the organoid for Image 2 before and after normalisation.

4.5.0 Exploratory data analysis

4.5.1 Pairwise Correlation Analysis

The pipeline generated 66 graphs, illustrating pairwise correlations between parameters. Among these, 23 graphs displayed significant correlations, while 43 did not show any correlation (see Appendix A).

Among the pairwise correlations performed on whole organoid parameters, 5 were found to be significantly correlated, all of which had an absolute correlation coefficient (r) over 0.4, while 10 were not (Figure 17a). The most significant and strongest positive correlation was observed between organoid solidity and organoid circularity ($p\text{-value} = 3.52\text{e-}08$, $r = 0.93$), as highlighted in Figure 17.b. This was followed by a significant positive correlation between organoid area and the number of nuclei within the organoid ($p\text{-value} = 1.32\text{e-}06$, $r = 0.88$). In contrast, the strongest negative correlation was between organoid mean intensity and organoid circularity ($p\text{-value} = 0.0062$, $r = -0.62$). Additionally, a significant but weaker correlation was noted between organoid solidity and organoid mean intensity ($p\text{-value} = 0.039$, $r = -0.49$).

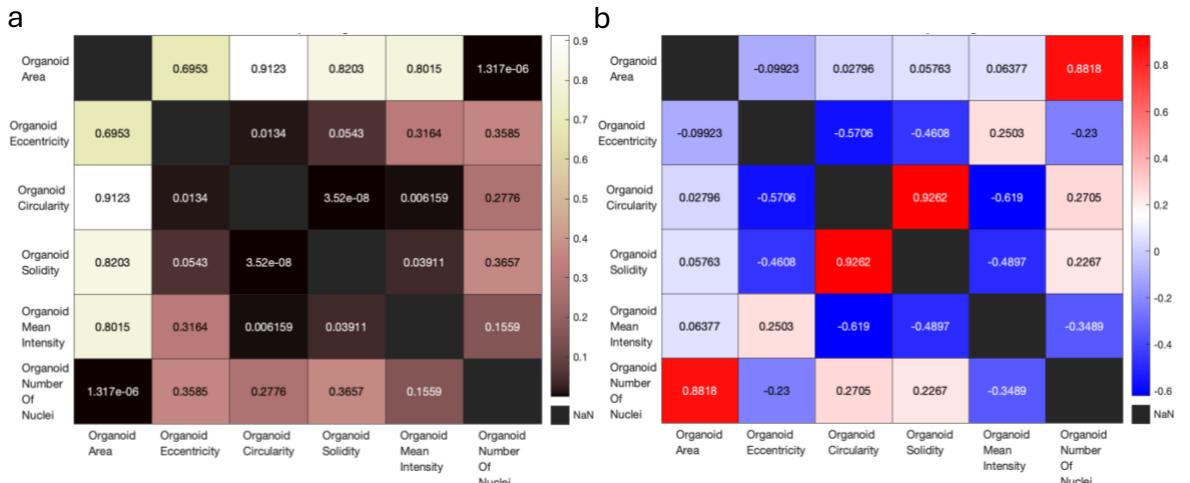


Figure 17. Pairwise correlations between organoid metrics. (a) Displays the p-values for the pairwise correlations. (b) Shows the correlation coefficients for the pairwise correlations.

Among the pairwise correlations analysed for nuclei parameters, 8 were found to be significant, while 7 were not (Figure 18a). Of these 8 significant correlations, only 2 had an absolute r value greater than 0.3, as shown in Figure 18b. The strongest positive correlation with the highest significance was observed between nuclei solidity and circularity ($p\text{-value} = 4.62\text{e-}143$, $r=0.76$). Conversely, the second strongest correlation, which was negative and also highly significant, was found between nuclei circularity and eccentricity ($p\text{-value} = 2.70\text{e-}84$, $r=-0.63$). There were several significant correlations with weak associations, such as between nuclei mean intensity and area ($p\text{-value} = 0.011$, $r=0.093$).

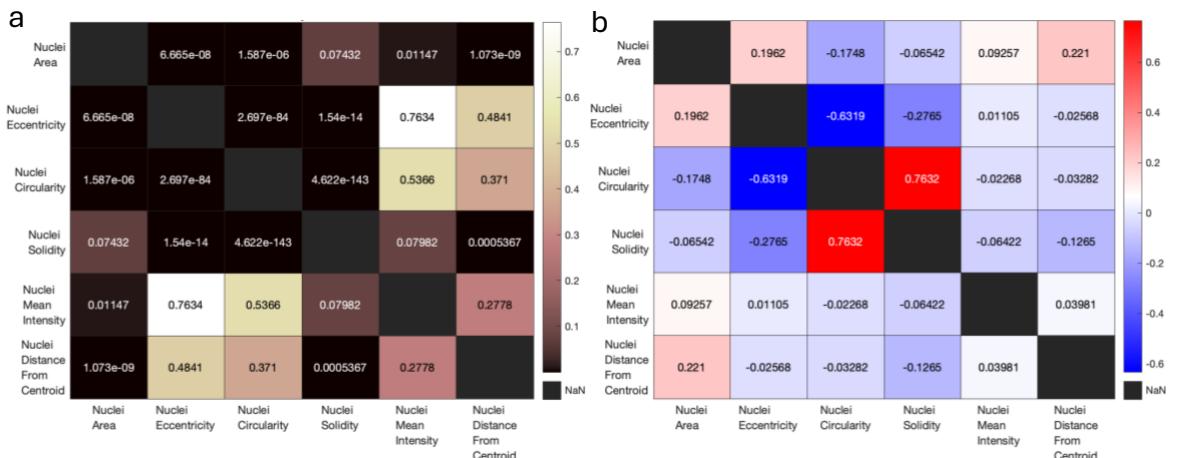


Figure 18. Pairwise correlations between nuclei metrics. (a) Displays the p-values for the pairwise correlations. (b) Shows the correlation coefficients for the pairwise correlations.

Among the pairwise correlations analysed for combined whole organoid and average nuclei parameters, 10 were found to be significant, while 26 were not (Figure 19a). Of the 10 that were significant, they all had an r value greater than 0.4, as shown in Figure 19b. The most significant correlation, and also the strongest positive correlation, was between organoid eccentricity and the distance of nuclei from the centre of the organoid ($p\text{-value} = 5.7\text{e-}05$, $r=0.80$). The second most significant correlation, which was the second strongest positive correlation, was observed between organoid mean intensity and nuclei area ($p\text{-value} = 0.0012$, $r=0.70$). The third most significant correlation, which was the highest negative correlation, was between organoid circularity and nuclei distance from the centre of the organoid ($p\text{-value} = 0.0030$, $r=-0.66$). The least correlated but still significant relationship was between organoid circularity and nuclei area ($p\text{-value} = 0.035$, $r=-0.50$).

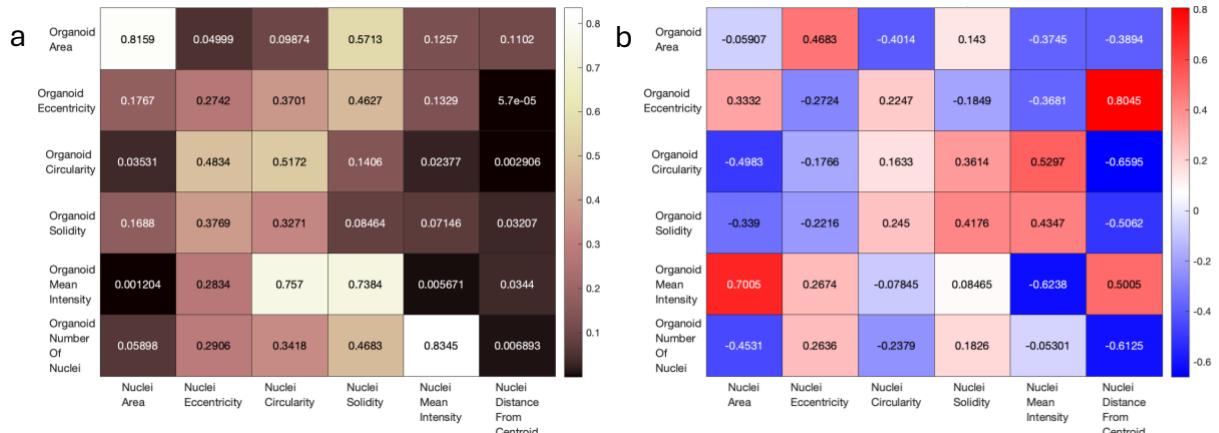


Figure 19. Pairwise correlations between organoid and nuclei metrics. (a) Displays the p-values for the pairwise correlations. (b) Shows the correlation coefficients for the pairwise correlations.

4.5.2 Analysis of Organoids Based on Lumen Presence

The pipeline generated 12 graphs comparing data from lumen-containing organoids and those without. 6 graphs depicted nuclei data (see Figure 20), and six represented whole organoid data (see Figure 21).

In the analysis of whole organoid data, two parameters showed a significant difference (p-value < 0.05) between organoids with and without a lumen within the z-slice. The first was circularity, with organoids containing a lumen being significantly less circular. The second was mean intensity, where organoids with a lumen exhibited a significantly higher mean intensity. Conversely, several parameters did not show significant differences between organoids with and without a lumen. These include organoid area, eccentricity, solidity, and the number of nuclei within the organoid, all with p-values greater than 0.05.

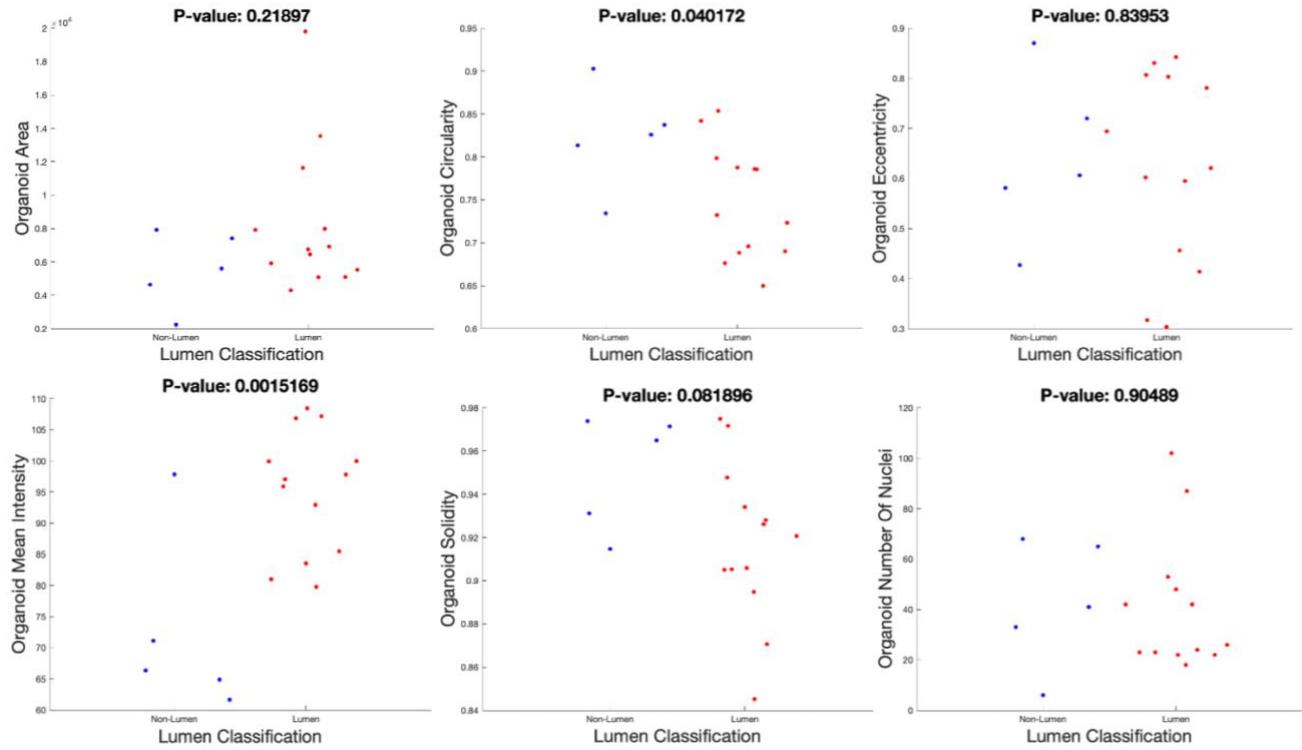


Figure 20. Comparison of organoid metrics between those with a lumen and those without.

Metrics for those organoids with a lumen are red and those without are blue. Each graph is accompanied by their corresponding p-value.

In the analysis of nuclei data, all parameters except one showed significant differences between organoids with and without a lumen (see Figure 21). The exception was nuclei solidity, which did not differ significantly between the two groups. However, nuclear area, eccentricity, and distance from the centre of the organoid were all significantly increased in organoids with a lumen. In contrast, nuclear circularity and mean intensity were significantly decreased in organoids containing a lumen.

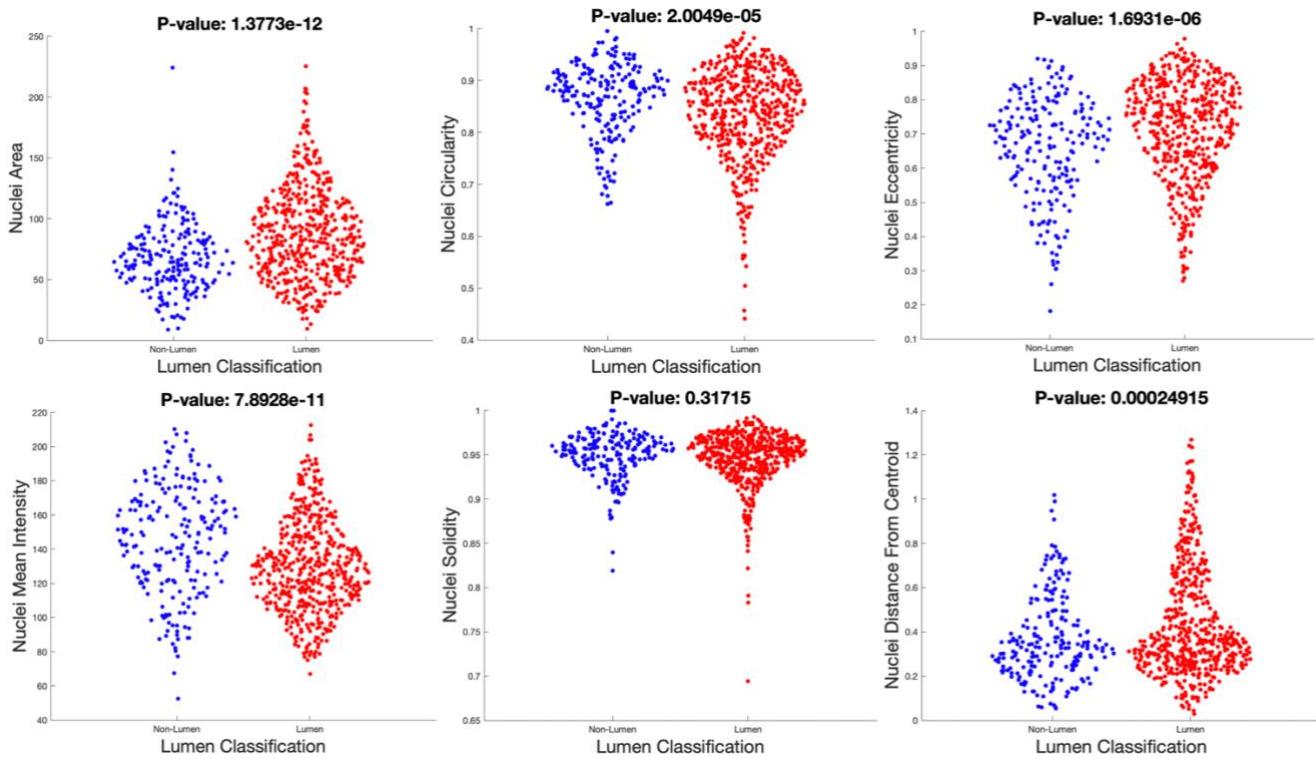


Figure 21. Comparison of nuclei metrics between organoids with a lumen and those without.

Metrics for those nuclei within an organoid containing a lumen are red and those within an organoid without are blue. Each graph is accompanied by their corresponding p-value.

4.5.3 Analysis of Mitotic Nuclei

The pipeline generated six graphs (Figure 22) and corresponding CSV files (see Appendix B) that compared mitotic nuclei, non-mitotic nuclei, and the miscellaneous objects. Among the various parameters analysed, all but one exhibited a significant difference between the groups ($p\text{-value} < 0.05$). The exception was the distance from the centre of the organoid. In terms of area, there was no significant difference between mitotic and non-mitotic nuclei ($p\text{-value} = 0.99$). However, the miscellaneous objects had a significantly larger area compared to non-mitotic nuclei ($p\text{-value} = 0.002$) and mitotic nuclei ($p\text{-value} = 0.006$). For circularity, mitotic nuclei were significantly less circular than non-mitotic nuclei ($p\text{-value} < 0.0001$), while the miscellaneous objects were significantly more circular than both mitotic ($p\text{-value} < 0.0001$) and non-mitotic nuclei ($p\text{-value} = 0.0006$). Regarding eccentricity, mitotic nuclei were significantly more elongated than non-mitotic nuclei ($p\text{-value} = 0.0035$). The miscellaneous objects were significantly less elongated than both non-mitotic ($p\text{-value} < 0.0001$) and mitotic nuclei ($p\text{-value} < 0.0001$). For mean intensity, mitotic nuclei and the miscellaneous objects exhibited significantly higher mean intensity compared to non-mitotic nuclei (mitotic nuclei: $p\text{-value} < 0.0001$; miscellaneous objects: $p\text{-value} < 0.0001$). However, no significant difference was observed between the miscellaneous objects and mitotic nuclei ($p\text{-value} = 0.30$). Finally, in terms of solidity, non-mitotic nuclei had significantly higher solidity than mitotic nuclei ($p\text{-value} < 0.0001$). There was no significant difference in solidity between non-mitotic nuclei and miscellaneous objects ($p\text{-value} = 0.15$), but the miscellaneous objects exhibited significantly higher solidity compared to mitotic nuclei ($p\text{-value} < 0.0001$).

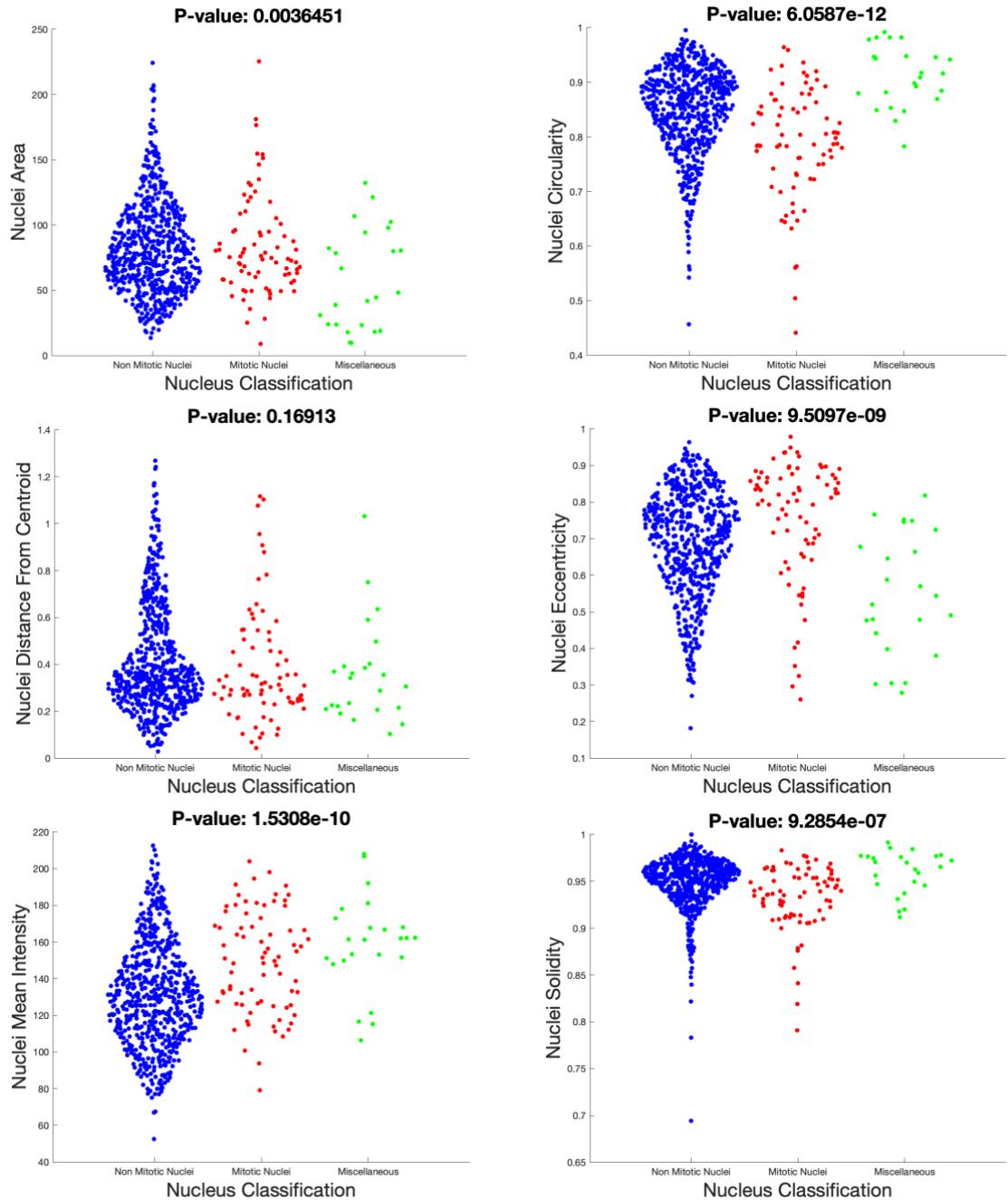


Figure 22. Comparison of various metrics for different nuclei within organoids. Non-mitotic nuclei are represented in blue, mitotic nuclei in red, and miscellaneous objects in green. Each graph includes the corresponding p-value.

4.5.4 Principal Component Analysis

The pipeline generated six graphs for the PCA: a two-dimensional and three-dimensional graph for each: whole organoid data, nuclei data, and the combined whole organoid data with nuclei averages. The PCA plot for both the organoid data (Figure 23a), as well as the organoid and average nuclei data (Figure 23c) reveals that the first principal component (PC1) accounts for almost all the variance (~100%), while the second principal component (PC2) and third principal component (PC3) explain a negligible amount of the variance (<0.01%). The spread of data points in the 2D plot indicates a significant variance along PC1, with minimal variation along PC2. For the nuclei data (Figure 23b), both the first two principal components together explain a substantial proportion of the variance, with

PC1 accounting for 61.67% and PC2 accounting for 38.32%, and PC3 with negligible amount of the variance. Also, the 2D scatter plot shows a more even spread of data points across both PC1 and PC2.

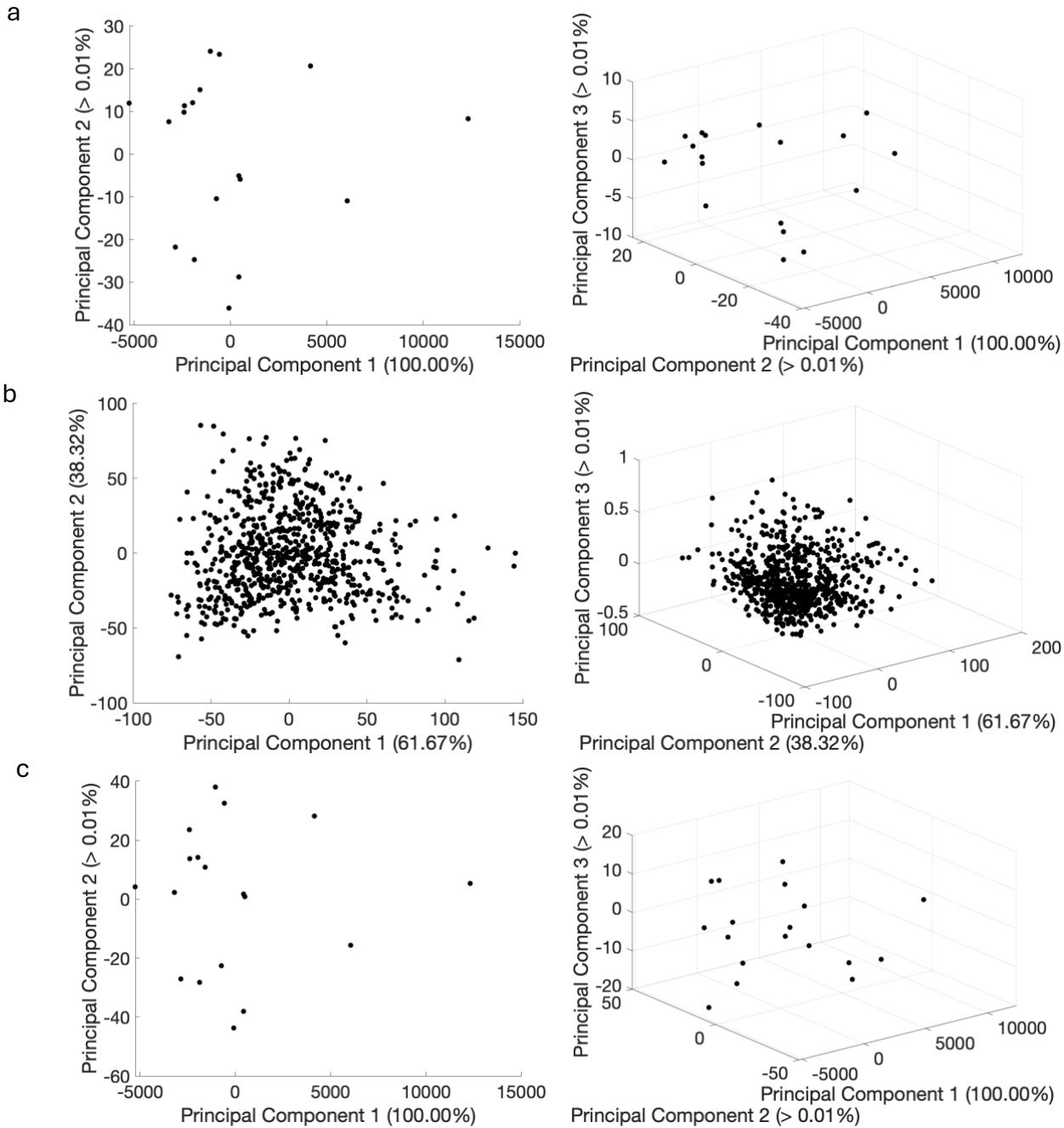


Figure 23. Principal Component Analysis (PCA) of organoid and nuclei metrics. (a) PCA for whole organoid metrics, (b) PCA for nuclei metrics, and (c) PCA combining organoid metrics with nuclei averages.

5.0 Discussion

Due to the limitations of traditional models in studying neurological disorders in the human brain, brain organoids are gaining traction for their ability to more accurately model early brain development than previously used animal models. 2D histology images are commonly used to study these organoids, allowing researchers to evaluate various morphological, structural, and functional properties. Quantification of these properties can be done manually or through automated methods. However,

manual quantification is labour-intensive, time-consuming, prone to bias, and often fails to capture the full complexity of the organoid data. In contrast, this study successfully implements an automated approach that combines traditional descriptors with deep learning techniques to standardise and streamline the analysis process. The methodology comprises three key components: first, segmenting images of stained organoids; second, training a convolutional neural network (CNN) to differentiate dividing nuclei; and third, developing a pipeline that processes the segmented masks and images to quantify organoid features. This pipeline also includes generating visualisations and conducting statistical analyses to effectively explore and interpret the generated data.

5.1 Segmentation

The initial step involved segmenting the organoids and nuclei within the images. For this task, Cellpose was utilised. Cellpose was first used to identify the premade models that provided the most accurate segmentation for organoids (cyt02_cp3) and nuclei (cyt03). However, it was observed that even the most accurate models produced segmentations that either missed organoids or nuclei or included extraneous parts of the image. To address these inaccuracies, the segments that were not accurately identified were manually corrected, and the revised segmentation masks were saved as PNG images. To enhance the accuracy of future segmentations, the premade models were then retrained using the corrected segmentations. This process was repeated for another image.

5.2.0 Pipeline for Quantification and Analysis of Organoids

The pipeline takes the segmentation masks and the stained organoid images as input and successfully extracts a range of parameters. For the entire organoid, the pipeline calculates metrics such as area, eccentricity, circularity, solidity, actin mean intensity, and the number of nuclei present. Additionally, for the nuclei located within these organoids, it extracts parameters including area, eccentricity, circularity, solidity, DNA mean intensity, and the distance of each nucleus from the centre of the organoid.

5.2.1 Normalisation

Before delving into data analysis, it was crucial to normalise the data to ensure consistency across images. Variations in image intensity can occur due to differences in imaging conditions, such as varying signal strength or inconsistencies in the amount of dye the culture absorbed. To account for these variations, the signal intensity of each image was adjusted, standardising the intensity levels across all images, regardless of their origin. Although normalisation did not affect the intensity, this is likely because the images provided were already normalised. Additionally, organoids within the same culture can exhibit variability in dye absorption. This is due to the dome shape of the organoids, which can result in inconsistent nuclei intensity across different organoids. To address this, the intensity of the nuclei was adjusted successfully ensuring uniformity, allowing for more accurate comparisons and analyses across different organoids.

Another crucial step of normalisation involved the measurement scales of the images. The apparent size of an organoid can vary depending on the magnification level at which the image was captured. For instance, an organoid might appear larger in one image than in another, but this could be due to higher magnification rather than an actual size difference. To address this, a CSV file containing the magnification details of each image was used to convert all measurements to micrometres. This adjustment was applied to parameters such as organoid area, nuclei area, and the distance of nuclei from the centre of the organoid. Further normalisation was required for the distance of nuclei from the centre of the organoid. Without normalisation, larger organoids would naturally exhibit greater distances from the centre, skewing the results, thus making it difficult to assess true spatial distribution.

The distance was normalised by dividing it by the radius of the respective organoid. This step successfully ensured that the distance is comparable across different organoids, regardless of their size.

5.2.2 Pairwise Correlation Analysis

To explore the relationships among various organoid metrics, pairwise comparisons were performed. Each metric was compared against every other to identify significant relationships, visualise their interconnections, and calculate correlation coefficients to assess the strength of these relationships. The analysis revealed five significant relationships. The strongest correlation existed between organoid solidity and organoid circularity. This finding aligns with expectations, as a more solid organoid often indicates a uniform and compact structure, naturally contributing to a more circular appearance. Following this, the next notable correlation was between organoid area and the number of nuclei within the organoid. This correlation is also expected, as larger organoids typically contain a greater number of cells. Additionally, the analysis uncovered a significant negative relationship between the mean intensity of actin staining and organoid circularity. This suggests that as organoid circularity decreases, the mean intensity of actin staining increases. A possible explanation is that less circular, more irregularly shaped organoids are more likely to contain lumens, where actin is concentrated on the apical side of cells (Doctor, 2006), leading to higher overall actin intensity. To confirm this hypothesis, it would be beneficial to compare the circularity or eccentricity of organoids with and without lumens. Furthermore, elongated organoids may have more extensive or uneven surfaces, allowing greater interaction and accumulation of Phalloidin dye, which binds to actin filaments, thereby increasing its intensity.

For nuclei metrics, a pairwise correlation analysis was also performed, revealing only two strong correlations, indicating limited relationships among the nuclei metrics. The first correlation was between nuclei solidity and circularity, a finding that mirrors the organoid data, as a more solid nucleus also tends to be more circular. The other significant relationship was found between nuclei circularity and eccentricity, with results showing that as nuclei become less circular, their eccentricity increases. This relationship is expected, as circularity and eccentricity are geometrically related properties: circularity measures how closely a shape resembles a perfect circle, while eccentricity indicates deviation from circularity, leading to more elongated shapes. Although there were statistically significant relationships between other metrics, such as mean intensity and area, no meaningful correlation ($r < 0.04$) was found between them.

When both organoid and nuclei metrics were analysed together, there were ten that were found to be significant. The most significant correlation identified was between organoid eccentricity and the distance of nuclei from the centre of the organoid. As organoids increase in size, nuclei tend to be situated further from the centre, which is an expected outcome given that larger organoids would naturally have a greater spatial distribution of nuclei. Another key finding was as nuclei area increases, the actin mean intensity of the organoid also raised. This relationship may be due to the observation that nuclei closer to the lumen tend to be larger, and as there are higher concentrations of actin neighbouring the lumen (Doctor, 2006), the overall mean intensity of that area is elevated. To confirm this hypothesis, future research could involve comparing nuclei area and actin intensity in organoids with and without lumens. Furthermore, the strongest negative correlation was observed between organoid circularity and the distance of nuclei from the centre of the organoid. This might be due to less circular organoids having lumens, resulting in nuclei being located further from the centre as the lumen occupies central space. To further explore this rationale, future investigations should compare the circularity of organoids with and without lumens.

5.2.3 Comparative Analysis of Organoids Based on Lumen Presence

To investigate the differences in metrics between organoids with and without a lumen in z-slice images, the pipeline offers the functionality to select and classify organoids based on the presence of a lumen. Comparing these classifications is crucial, as organoids are three-dimensional structures, and understanding how metrics differ in various cross-sections is essential. In the analysis of whole organoid data, two metrics showed significant differences. First, organoids with a lumen were significantly less circular, supporting the earlier rationale that less circular organoids have a greater average nuclei distance from the centre due to the empty space created by the lumen. Second, organoids with a lumen exhibited significantly higher mean intensity, likely because the organoids apical side contains a higher concentration of actin, resulting in increased actin mean intensity across the organoid. This also confirms the hypothesis that actin intensity in organoids with a lumen is increased.

In the analysis of whole organoid data, all metrics except nuclei solidity showed significant differences between organoids with and without a lumen. Specifically, nuclei area, eccentricity, and distance from the centre of the organoid were all significantly increased in organoids with a lumen. The increase in nuclei area supports the hypothesis that organoids with a lumen exhibit larger nuclei, likely due to a higher presence of dividing nuclei near the lumen. This mitotic activity can lead to nuclei elongation and stretching, resulting in larger nuclei with more empty spaces devoid of DNA. Additionally, the observations of reduced nuclear circularity and DNA mean intensity in organoids with a lumen further support this. To validate these findings and explore nuclei metrics related to mitotic activity in greater detail, further research is warranted.

5.2.4 Principal Component Analysis

The pipeline successfully generated both 2D and 3D PCA graphs for whole organoid data, nuclei data, and combined data integrating the whole organoid data with the average nuclei data. The whole organoid data and the combined data show a dominant variance concentrated along a single principal component. In contrast, the nuclei data exhibits significant contributions from two principal components, indicating a more complex variance pattern.

5.2.5.0 Analysis of Mitotic Nuclei

In many organs, variations in cell number can be offset by changes in cell size. For instance, in the fruit fly *Drosophila melanogaster*, a decrease in the total number of cells in the wing disc is offset by an increase in cell size, ensuring that the overall size of the organ remains unchanged (Homem *et al.*, 2015). However, in the brain, the precise number of neurons produced during development is crucial for proper circuit formation, and any variation in cell size will disrupt their function (Homem *et al.*, 2015). Since neurons originate from a relatively small pool of neural progenitor cells, the regulation of these cells' division is particularly critical in the CNS.

To investigate neural progenitor cell proliferation in human brains during early development, the division of nuclei within organoids at the tNE stage was analysed. To achieve this, it was necessary to identify mitotic nuclei and collect and assess data related to them. Traditional descriptors were insufficient as no single descriptor could accurately differentiate mitotic nuclei. For instance, intensity alone was not reliable, as dividing nuclei could have similar intensities to non-mitotic nuclei. Additionally, while cells undergoing mitosis are often found surrounding the central lumen, using distance from the lumen as a criterion proved inadequate, as non-mitotic nuclei can also be located near the lumen (Barry *et al.*, 2014). Consequently, a more effective approach involved analysing the patterns of the nuclei using deep learning. This method's ability to recognise complex patterns is particularly valuable for identifying the intricate features of mitotic nuclei that traditional approaches will overlook.

5.2.5.1 Model training

A CNN was trained on images of both mitotic and non-mitotic nuclei. The images also contained small bright artefacts of unknown origin, and the model was additionally trained to recognise these artefacts. The training accuracy of the model showed a significant improvement early on, stabilising around 70-80%. This rapid increase in accuracy during the initial epochs indicates that the model learnt well from the training data and the steady decrease in training loss further supports this, as it shows the model is minimising the error on the training set. Additionally, the final validation accuracy of 69.57% indicates that the model is generalising fairly well to the validation data. However, the training accuracy is higher than the validation accuracy, and the training loss is consistently decreasing while the validation loss shows less consistency, stabilising far higher. These are classic signs that the model may be overfitting. When applied to the test set, the model achieved an accuracy of 68.18%. While this result indicates that the model performs better than random guessing and represents a promising start, there remains significant room for improvement. For example, in recent years, top models like EfficientNet and Vision Transformers (ViTs) achieve accuracies around 80-90% on the ImageNet dataset, which consists of over a 1,000 classes 1.3 million images (Tan, 2019; Dosovitskiy *et al.*, 2020).

5.2.5.2 Statistical Analysis of Mitotic Nuclei

The model was used to classify nuclei within the images. There was no significant difference in area between mitotic and non-mitotic nuclei, suggesting that the size of these nuclei does not change substantially during mitosis. However, both types of nuclei had a significantly larger area compared to the miscellaneous objects. This reduced size can be attributed to chromatin condensation, which is observed during apoptosis, where the nucleus becomes compacted and smaller (Dmitrieva and Burg, 2008). Additionally, the model's prediction of the small area for the artefacts in unseen data further validates its effectiveness. This success demonstrates that the model has effectively learned to identify and differentiate these artefacts based on their distinctive characteristics, rather than relying on their size, as all images were made the same size.

Mitotic nuclei were significantly less circular and more eccentric than both non-mitotic nuclei and the miscellaneous objects. This increased elongation reflects the changes in nuclear shape during mitosis, where nuclei often become stretched as they two daughter nuclei are splitting apart. The miscellaneous objects were significantly more circular than non-mitotic nuclei. This high circularity has been reported as a typical feature of early cell death, adding to the belief that these are apoptotic cells (Daniel and DeCoster, 2004; Helmy and Abdel Azim, 2012).

There was no significant difference in the distance from the centre of the organoid among mitotic nuclei, non-mitotic nuclei, and the artefacts. This uniform distribution indicates that the positioning of these objects within the organoid is consistent across different types and stages, suggesting that their placement is not influenced by their specific characteristics. This finding is somewhat unexpected, as mitotic activity is typically observed near the surface adjacent to the lumen in many developing epithelia. However, the inclusion of both lumen-containing and non-lumen organoids in the study may have introduced variability, masking the typical spatial patterns. Therefore, the uniform distribution observed reflects the mixed nature of the organoids used, rather than a consistent biological pattern.

Mitotic nuclei and the miscellaneous objects exhibited significantly higher mean intensity compared to non-mitotic nuclei. This increase in intensity can be attributed to the substantial changes in chromatin structure during mitosis and apoptosis and DNA content that occur during mitosis. As the cell cycle progresses from interphase to metaphase, chromatin undergoes progressive condensation, ultimately reaching its most compact state in metaphase (Estandarte *et al.*, 2016). This condensation results in a higher density of DNA, which, when stained with DAPI, leads to increased fluorescence.

Additionally, mitosis involves the duplication of the genome. For example, cells in the G2 phase of mitosis will have twice the amount of DNA as cells in the G1 phase (Israels and Israels, 2000). This further contributes to the elevated mean intensity observed in mitotic nuclei, as both the increased DNA content and the compacted chromatin structure enhance the DAPI staining signal. However, there is no significant difference in intensity between the artefacts and mitotic nuclei. This suggests that the level of chromatin compaction in apoptotic cells may be comparable to that in mitotic nuclei.

Non-mitotic nuclei and the artefacts both exhibited higher solidity compared to mitotic nuclei. This suggests that both non-mitotic nuclei and the miscellaneous objects tend to have more well-defined, solid shapes, whereas mitotic nuclei, undergoing active cell division, display a more irregular and fragmented nature of the nuclear envelope during cell division. However, there was no significant difference in solidity between non-mitotic nuclei and the miscellaneous objects, suggesting that both share similarly compact and well-defined shapes. In non-mitotic nuclei, this solidity likely reflects the stable and intact structure of the nuclear envelope, which maintains the nucleus's consistent shape during the interphase. On the other hand, the high solidity of the miscellaneous objects, hypothesised to be apoptotic cells, is indicative of the condensed and compacted nature of chromatin during apoptosis.

5.3.0 Future Work

5.3.1 Principal Component Analysis

The limited variance observed in the whole organoid and combined datasets is likely due to the small number of organoids analysed. For instance, a study by Shaukat *et al.* (2016) investigated the impact of sample size on PCA ordination and found that a sample size of 40 to 50 is necessary to reliably recover the first few components essential for exploring and summarising multivariate data. Therefore, it is crucial that future research includes additional data collection.

To further uncover potential underlying patterns and groupings within the data, future studies should integrate clustering techniques as a complementary approach to PCA. Clustering can be used to identify distinct subgroups or trends that might not be immediately apparent through PCA alone. This approach will be particularly valuable when analysing diverse types of organoids, such as organoids modelled after neurological disorders, which may exhibit unique clustering.

5.3.2 Future Directions for Mitotic Model Training

To improve the model accuracy and decrease overfitting it is vital that it is trained on more images. It was only trained on 123 images 41 for each class. When training with deep learning millions of data records is required. For example, Microsoft Common Objects in Context (COCO) consists of 2.5 million images with 91 object classes (Lin *et al.*, 2014). Increasing the size of the training dataset would help the model better capture the variability within each class and improve its ability to generalise to new, unseen images.

5.3.3 Advancing Lumen Classification with Deep Learning

Manually selecting organoids with a lumen is highly inefficient and becomes increasingly impractical when dealing with hundreds, let alone millions, of images. This process is not only laborious but also time-consuming, making it a significant bottleneck in large-scale analyses. To address this challenge, automating the classification of organoids based on the presence of a lumen is crucial. Deep learning techniques, similar to those used for nucleus classification, offer an effective solution. By harnessing deep learning's ability to recognise complex patterns, these methods can be trained to holistically

identify the features of organoids with and without a lumen, which would be difficult to capture with traditional methods.

5.3.4 Transitioning from 2D to 3D Analysis

Current methods, including those used in this study, primarily focus on 2D segmentation and subsequent analysis. However, 2D analysis inadequately captures the complexity of 3D spatial relationships in cells and organoids (Albanese *et al.*, 2020). Additionally, previous computational approaches for analysing organoids in 3D have centred on qualitative visualisation rather than quantitative cellular analysis, with existing 3D cell segmentation methods being restricted to specialised contexts (Mukashyaka *et al.*, 2023). For example, Boutin *et al.* (2018) developed a 3D spheroid and nuclei segmentation approach for optically cleared images of a single spheroid per well, while more recently, Beghin *et al.* (2022) introduced a segmentation technique for the Jewell system, again focusing on one organoid per well. However, typical assays involve analysing large numbers of organoids within each well (Mukashyaka *et al.*, 2023). This underscores the need for a flexible 3D approach to accurately identify and quantify individual cells and their morphologies. To achieve this, z-stack images can be stitched together, and the process used in this study can be adapted for 3D analysis, allowing for the retrieval of additional metrics such as volume. This approach could be implemented similarly to the Cellos pipeline (<https://github.com/TheJacksonLaboratory/Cellos>), although Cellos is developed in Python. In contrast, the pipeline in this study is built in MATLAB, which offers superior capabilities for high-throughput image data processing and analysis due to its efficiency in handling matrices and arrays.

5.3.5 Extending Analysis to Neurological Disorder Models

While this project has focused exclusively on the analysis of healthy brain organoids, it is crucial to extend this research to organoids modelled after neurodevelopmental disorders in order to explore their morphological and behavioural differences. For instance, multiple studies have demonstrated that brain organoids exposed to the Zika virus exhibit significant defects in neural progenitors, leading to abnormal neurogenesis and even cell death (Cugola *et al.*, 2016; Garcez *et al.*, 2016; Qian *et al.*, 2017; Qian *et al.*, 2016). To further investigate these phenomena, the pipeline developed in this study can be utilised to compare apoptotic cells, previously hypothesised to be present as image artefacts, by comparing healthy brain organoids with those exposed to those modelled after neurodevelopmental disorders.

In addition to investigating disease mechanisms, the pipeline holds significant potential for monitoring and testing the efficacy of drug treatments on disease-modelled brain organoids. By applying this tool to organoids modelled after neurodevelopmental disorders, researchers can track changes in cellular morphology and behaviour over time potentially aiding the development of targeted and more personalised treatments.

6.0 Conclusion

To overcome the challenges presented by manual quantification of brain organoids, this study successfully developed an automated pipeline that integrates traditional descriptors with deep learning methods, streamlining the analysis and enhancing quantification accuracy and reliability.

The pipeline extracts a comprehensive range of parameters, including area, eccentricity, circularity, solidity, and mean intensity for both whole organoids and nuclei, along with the number of nuclei and their distance from the centre of the organoid. To ensure data consistency and reliability, normalisation techniques were applied to account for variations in image intensity and scale. The pipeline then

conducts extensive statistical analyses, including pairwise correlation analysis, differences between organoids with and without lumens, and PCA. Moreover, the analysis also concentrated on exploring cell proliferation within the organoids by examining key differences among metrics for mitotic nuclei, non-mitotic nuclei, and recognised artefacts, which were hypothesised to be apoptotic cells due to their small size and brightness—indicators of chromatin condensation during apoptosis. To achieve this, a CNN was employed to identify mitotic nuclei, non-mitotic nuclei, and artefacts. While the CNN model showed promising accuracy (68.18%) in distinguishing between mitotic non-mitotic nuclei, and the artefacts, there is still significant room for improvement.

Future work should focus on expanding the image dataset for both mitotic training and exploratory data analysis, as this study only utilised 18 organoids and 41 images for each nuclei class. Increasing the number of images will significantly enhance model accuracy and the reliability in trends identified. Additionally, advancing lumen classification with deep learning techniques, transitioning from 2D to 3D analysis, and applying the pipeline to organoids modelled after neurological disorders are essential next steps. These improvements will enhance the pipeline's accuracy and versatility in studying brain organoids and ultimately aid in the identification of effective treatments for neurological disorders.

7.0 References

- Adams, J.W., Cugola, F.R. and Muotri, A.R., 2019. Brain organoids as tools for modeling human neurodevelopmental disorders. *Physiology*, 34(5), pp.365-375.
- Albanese, A., Swaney, J.M., Yun, D.H., Evans, N.B., Antonucci, J.M., Velasco, S., Sohn, C.H., Arlotta, P., Gehrke, L. and Chung, K., 2020. Multiscale 3D phenotyping of human cerebral organoids. *Scientific reports*, 10(1), p.21487.
- Andrion, A., Magnani, C., Betta, P.G., Donna, A., Mollo, F., Scelsi, M., Bernardi, P., Botta, M. and Terracini, B., 1995. Malignant mesothelioma of the pleura: interobserver variability. *Journal of clinical pathology*, 48(9), pp.856-860.
- Barry, D.S., Pakan, J.M. and McDermott, K.W., 2014. Radial glial cells: key organisers in CNS development. *The international journal of biochemistry & cell biology*, 46, pp.76-79.
- Beghin, A., Grenci, G., Sahni, G., Guo, S., Rajendiran, H., Delaire, T., Mohamad Raffi, S.B., Blanc, D., de Mets, R., Ong, H.T. and Galindo, X., 2022. Automated high-speed 3D imaging of organoid cultures with multi-scale phenotypic quantification. *Nature Methods*, 19(7), pp.881-892.
- Benito-Kwiecinski, S., Giandomenico, S.L., Sutcliffe, M., Riis, E.S., Freire-Pritchett, P., Kelava, I., Wunderlich, S., Martin, U., Wray, G.A., McDole, K. and Lancaster, M.A., 2021. An early cell shape transition drives evolutionary expansion of the human forebrain. *Cell*, 184(8), pp.2084-2102.
- Boutin, M.E., Voss, T.C., Titus, S.A., Cruz-Gutierrez, K., Michael, S. and Ferrer, M., 2018. A high-throughput imaging and nuclear segmentation analysis protocol for cleared 3D culture models. *Scientific reports*, 8(1), p.11135.
- Cugola, F.R., Fernandes, I.R., Russo, F.B., Freitas, B.C., Dias, J.L., Guimarães, K.P., Benazzato, C., Almeida, N., Pignatari, G.C., Romero, S. and Polonio, C.M., 2016. The Brazilian Zika virus strain causes birth defects in experimental models. *Nature*, 534(7606), pp.267-271.

Daniel, B. and DeCoster, M.A., 2004. Quantification of sPLA₂-induced early and late apoptosis changes in neuronal cell cultures using combined TUNEL and DAPI staining. *Brain Research Protocols*, 13(3), pp.144-150.

Di Lullo, E. and Kriegstein, A.R., 2017. The use of brain organoids to investigate neural development and disease. *Nature Reviews Neuroscience*, 18(10), pp.573-584.

Dmitrieva, N.I. and Burg, M.B., 2008. Analysis of DNA breaks, DNA damage response, and apoptosis produced by high NaCl. *American Journal of Physiology-Renal Physiology*, 295(6), pp.F1678-F1688.

Doctor, R.B., 2006. The actin cytoskeleton in the apical domain of epithelial cells. *Advances in Molecular and Cell Biology*, 37, pp.25-47.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. and Uszkoreit, J., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Elgendi, M., 2020. *Deep learning for vision systems*. Simon and Schuster.

Estandarte, A.K., Botchway, S., Lynch, C., Yusuf, M. and Robinson, I., 2016. The use of DAPI fluorescence lifetime imaging for investigating chromatin condensation in human chromosomes. *Scientific reports*, 6(1), p.31417.

Feigin, V.L., Vos, T., Nichols, E., Owolabi, M.O., Carroll, W.M., Dichgans, M., Deuschl, G., Parmar, P., Brainin, M. and Murray, C., 2020. The global burden of neurological disorders: translating evidence into policy. *The Lancet Neurology*, 19(3), pp.255-265.

Garcez, P.P., Loiola, E.C., Madeiro da Costa, R., Higa, L.M., Trindade, P., Delvecchio, R., Nascimento, J.M., Brindeiro, R., Tanuri, A. and Rehen, S.K., 2016. Zika virus impairs growth in human neurospheres and brain organoids. *Science*, 352(6287), pp.816-818.

Gilbert, S.F., 2000. An introduction to early developmental processes. In *Developmental Biology. 6th edition*. Sinauer Associates.

Gu, D., Andreev, K. and Dupre, M.E., 2021. Major trends in population growth around the world. *China CDC weekly*, 3(28), p.604.

Haja, A., Horcas-Nieto, J.M., Bakker, B.M. and Schomaker, L., 2023. Towards automatization of organoid analysis: A deep learning approach to localize and quantify organoid images. *Computer Methods and Programs in Biomedicine Update*, 3, p.100101.

Hassaballah, M. and Awad, A.I. eds., 2020. *Deep learning in computer vision: principles and applications*. CRC Press.

Helmy, I.M. and Abdel Azim, A.M., 2012. Efficacy of ImageJ in the assessment of apoptosis. *Diagnostic pathology*, 7, pp.1-6.

Homem, C.C., Repic, M. and Knoblich, J.A., 2015. Proliferation control in neural stem and progenitor cells. *Nature Reviews Neuroscience*, 16(11), pp.647-659.

- Israels, E.D. and Israels, L.G., 2000. The cell cycle. *The oncologist*, 5(6), pp.510-513.
- Kelava, I. and Lancaster, M.A., 2016. Dishing out mini-brains: Current progress and future prospects in brain organoid research. *Developmental biology*, 420(2), pp.199-209.
- Kim, S.H. and Chang, M.Y., 2023. Application of Human Brain Organoids—Opportunities and Challenges in Modeling Human Brain Development and Neurodevelopmental Diseases. *International Journal of Molecular Sciences*, 24(15), p.12528.
- Konkel, L., 2018. The brain before birth: using fMRI to explore the secrets of fetal neurodevelopment.
- Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C.L., 2014. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13* (pp. 740-755). Springer International Publishing.
- Mukashyaka, P., Kumar, P., Mellert, D.J., Nicholas, S., Noorbakhsh, J., Brugio, M., Anczukow, O., Liu, E.T. and Chuang, J.H., 2023. Cellos: High-throughput deconvolution of 3D organoid dynamics at cellular resolution for cancer pharmacology. *bioRxiv*.
- Nakano, T., Ando, S., Takata, N., Kawada, M., Muguruma, K., Sekiguchi, K., Saito, K., Yonemura, S., Eiraku, M. and Sasai, Y., 2012. Self-formation of optic cups and storable stratified neural retina from human ESCs. *Cell stem cell*, 10(6), pp.771-785.
- O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G.V., Krpalkova, L., Riordan, D. and Walsh, J., 2020. Deep learning vs. traditional computer vision. In *Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 11* (pp. 128-144). Springer International Publishing.
- Patikas, N., Ansari, R. and Metzakopian, E., 2023. Single-cell transcriptomics identifies perturbed molecular pathways in midbrain organoids using α -synuclein triplication Parkinson's disease patient-derived iPSCs. *Neuroscience Research*, 195, pp.13-28.
- Qian, X., Nguyen, H.N., Jacob, F., Song, H. and Ming, G.L., 2017. Using brain organoids to understand Zika virus-induced microcephaly. *Development*, 144(6), pp.952-957.
- Qian, X., Nguyen, H.N., Song, M.M., Hadiono, C., Ogden, S.C., Hammack, C., Yao, B., Hamersky, G.R., Jacob, F., Zhong, C. and Yoon, K.J., 2016. Brain-region-specific organoids using mini-bioreactors for modeling ZIKV exposure. *Cell*, 165(5), pp.1238-1254.
- Rees, S. and Inder, T., 2005. Fetal and neonatal origins of altered brain development. *Early human development*, 81(9), pp.753-761.
- Shaukat, S.S., Rao, T.A. and Khan, M.A., 2016. Impact of sample size on principal component analysis ordination of an environmental data set: Effects on eigenstructure. *Ekológia (bratislava)*, 35(2), pp.173-190.
- Shou, Y., Liang, F., Xu, S. and Li, X., 2020. The application of brain organoids: from neuronal development to neurological diseases. *Frontiers in cell and developmental biology*, 8, p.579659.

Steiner, T.J. and Stovner, L.J., 2023. Global epidemiology of migraine and its implications for public health and health policy. *Nature Reviews Neurology*, 19(2), pp.109-117.

Stringer, C., Wang, T., Michaelos, M. and Pachitariu, M., 2021. Cellpose: a generalist algorithm for cellular segmentation. *Nature methods*, 18(1), pp.100-106.

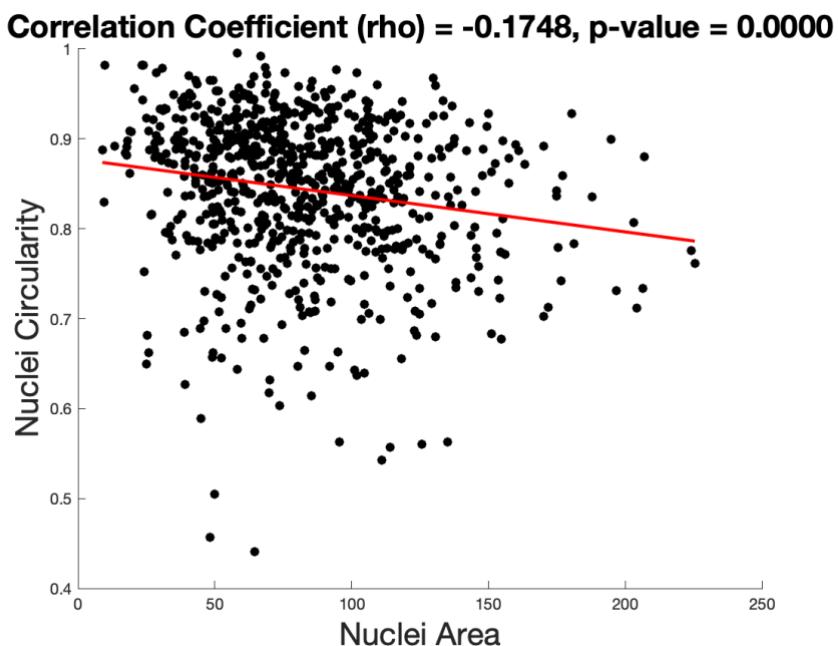
Tan, M., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*.

Voulodimos, A., Doulamis, N., Doulamis, A. and Protopapadakis, E., 2018. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018(1), p.7068349.

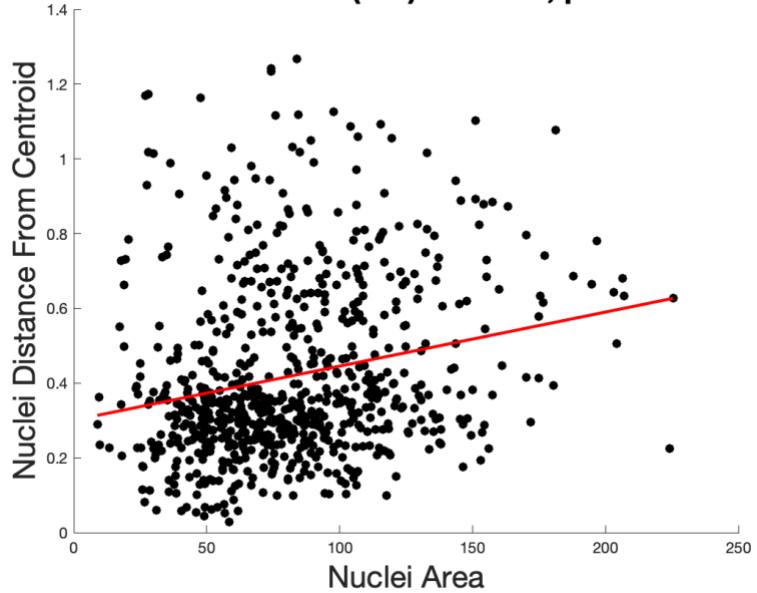
Wegmann, S., DeVos, S.L., Zeitler, B., Marlen, K., Bennett, R.E., Perez-Rando, M., MacKenzie, D., Yu, Q., Commins, C., Bannon, R.N. and Corjuc, B.T., 2021. Persistent repression of tau in the brain using engineered zinc finger protein transcription factors. *Science Advances*, 7(12), p.eabe1611.

8.0 Appendix

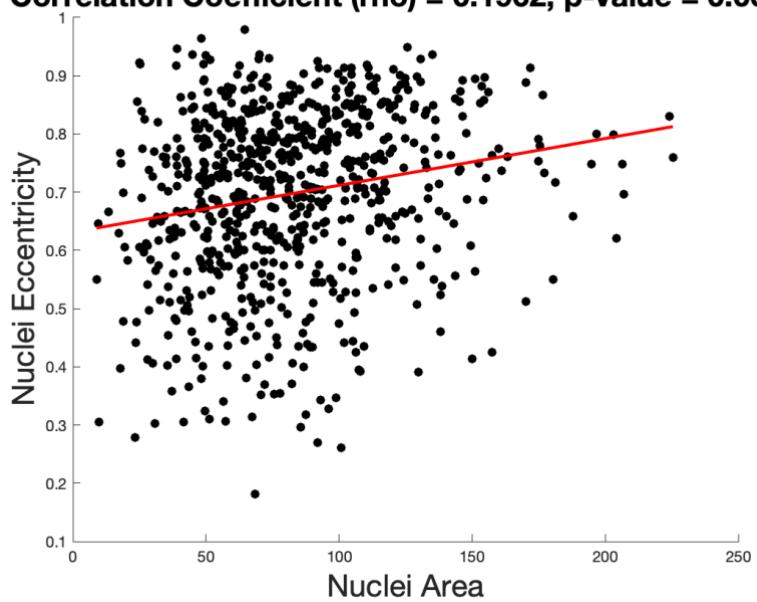
8.1 Appendix A



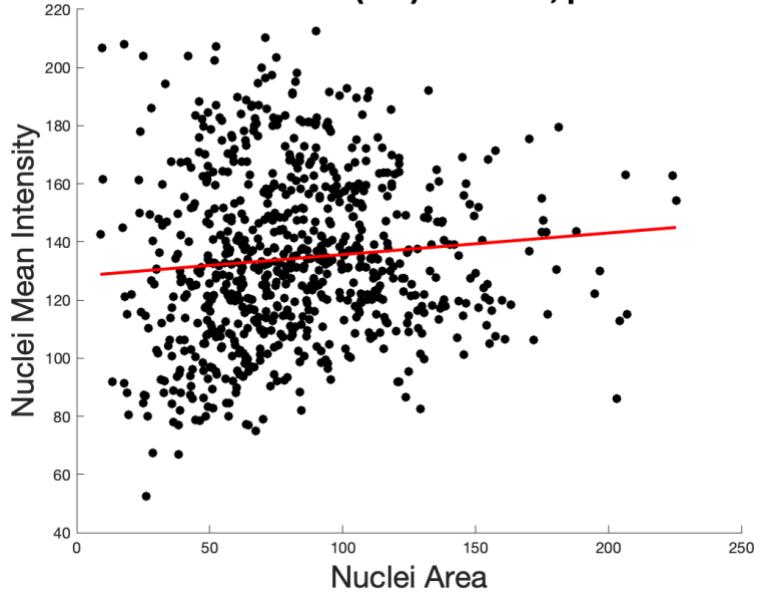
Correlation Coefficient (rho) = 0.2210, p-value = 0.0000



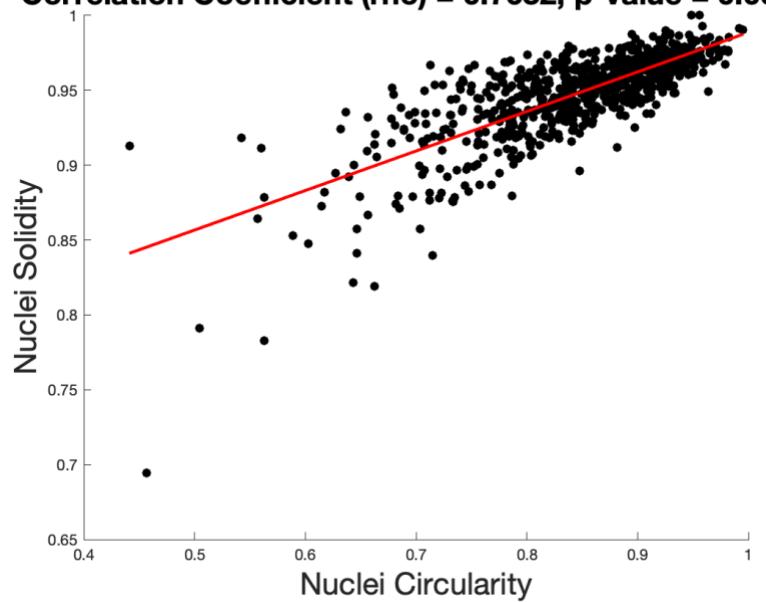
Correlation Coefficient (rho) = 0.1962, p-value = 0.0000



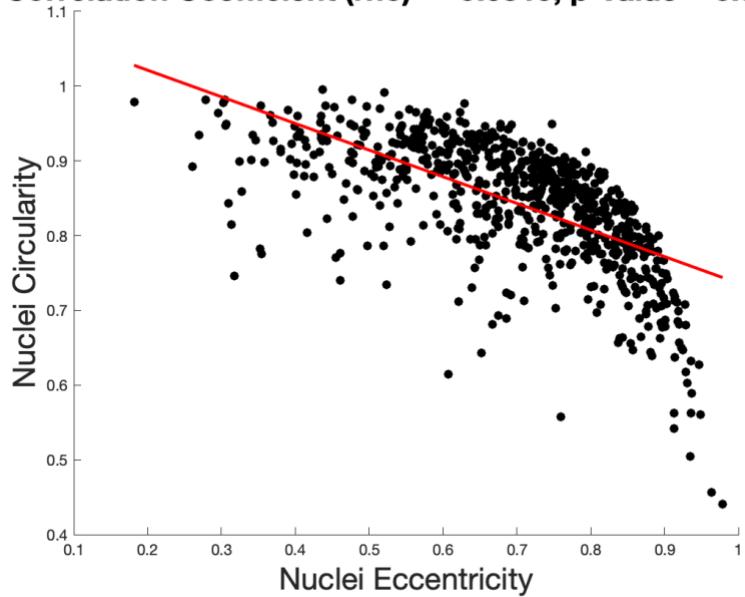
Correlation Coefficient (rho) = 0.0926, p-value = 0.0115



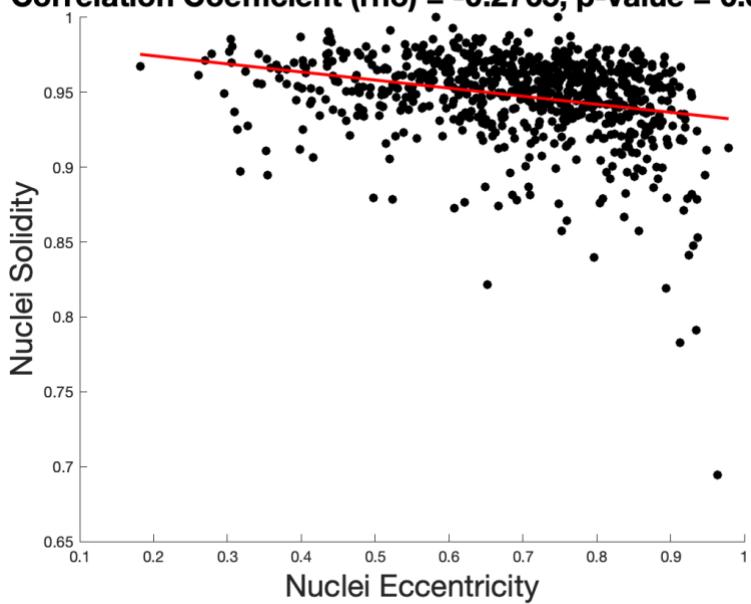
Correlation Coefficient (rho) = 0.7632, p-value = 0.0000



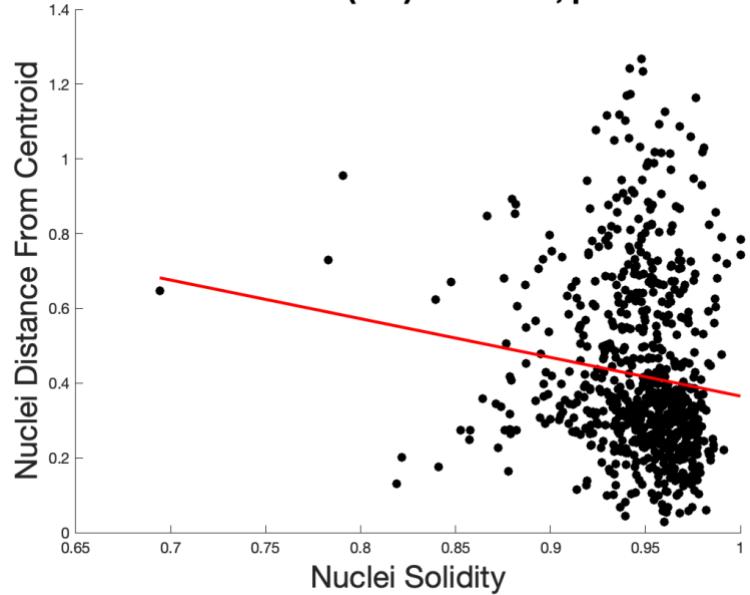
Correlation Coefficient (rho) = -0.6319, p-value = 0.0000



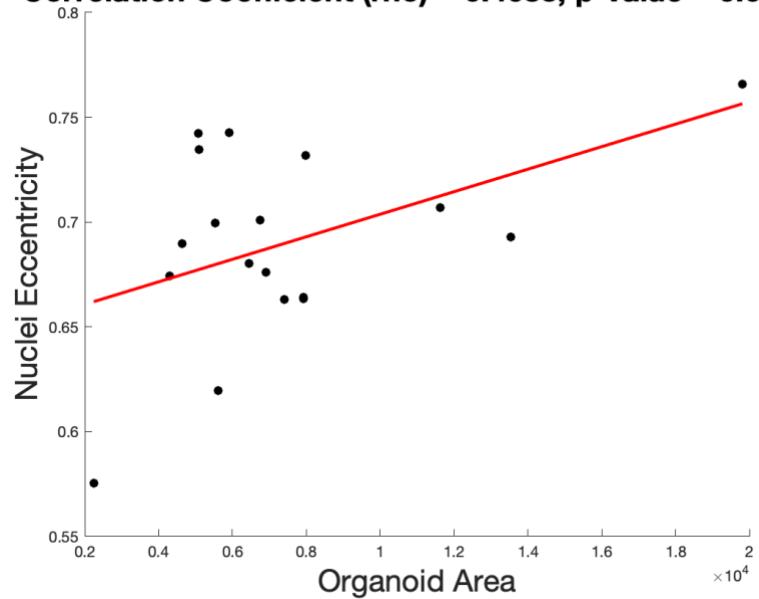
Correlation Coefficient (rho) = -0.2765, p-value = 0.0000



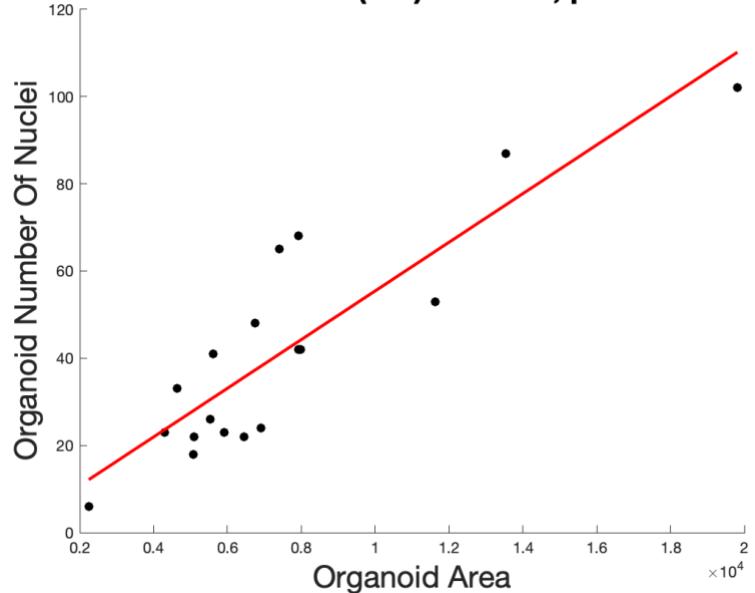
Correlation Coefficient (rho) = -0.1265, p-value = 0.0005



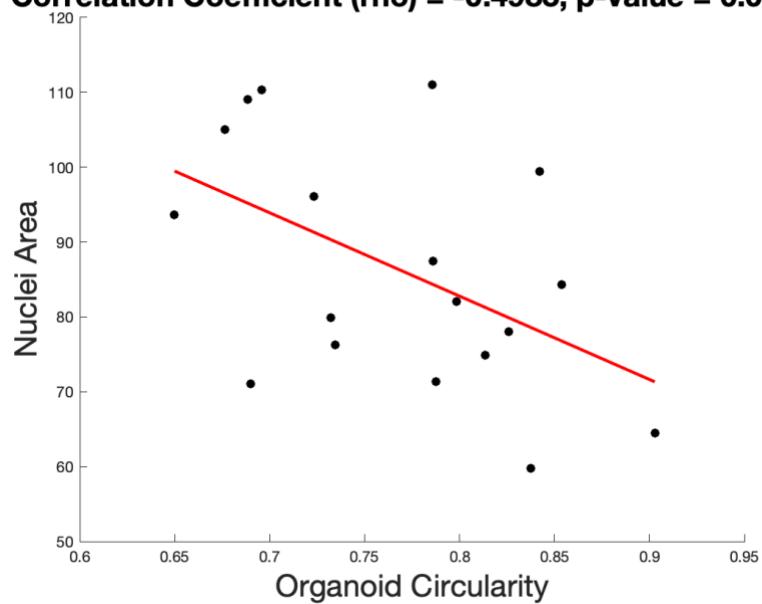
Correlation Coefficient (rho) = 0.4683, p-value = 0.0500



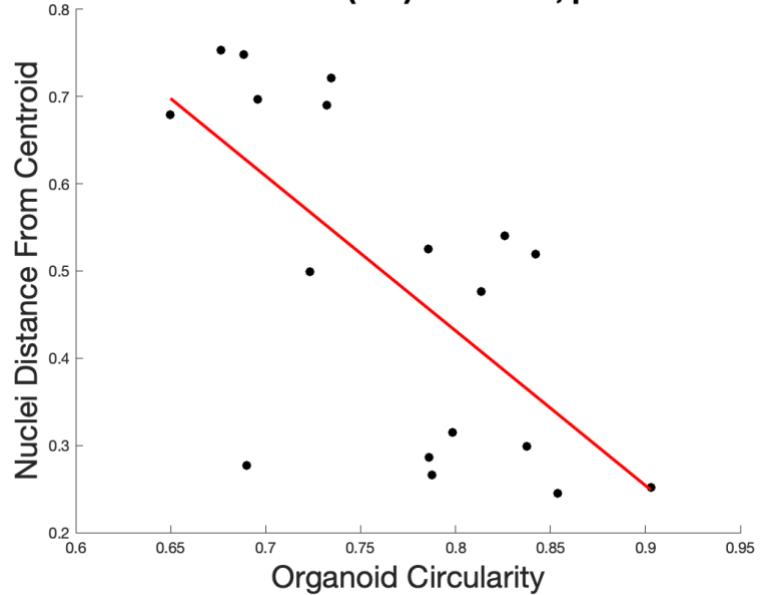
Correlation Coefficient (rho) = 0.8818, p-value = 0.0000



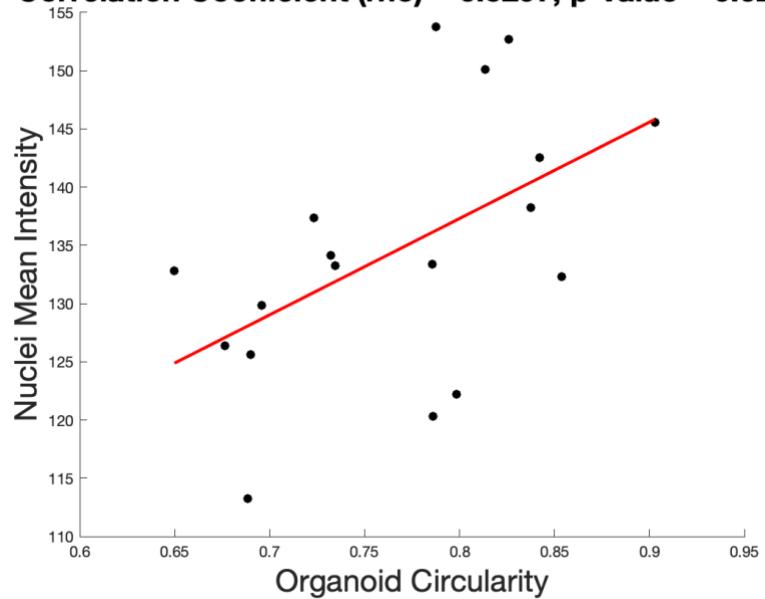
Correlation Coefficient (rho) = -0.4983, p-value = 0.0353



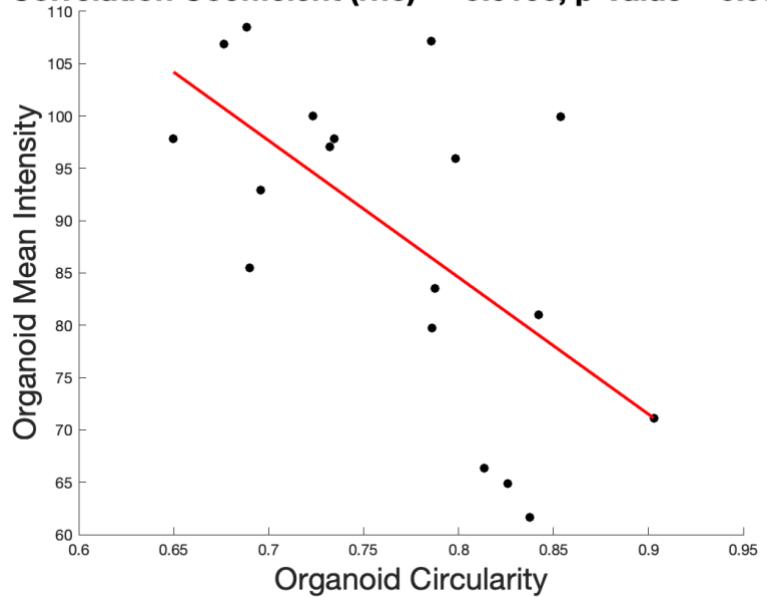
Correlation Coefficient (rho) = -0.6595, p-value = 0.0029



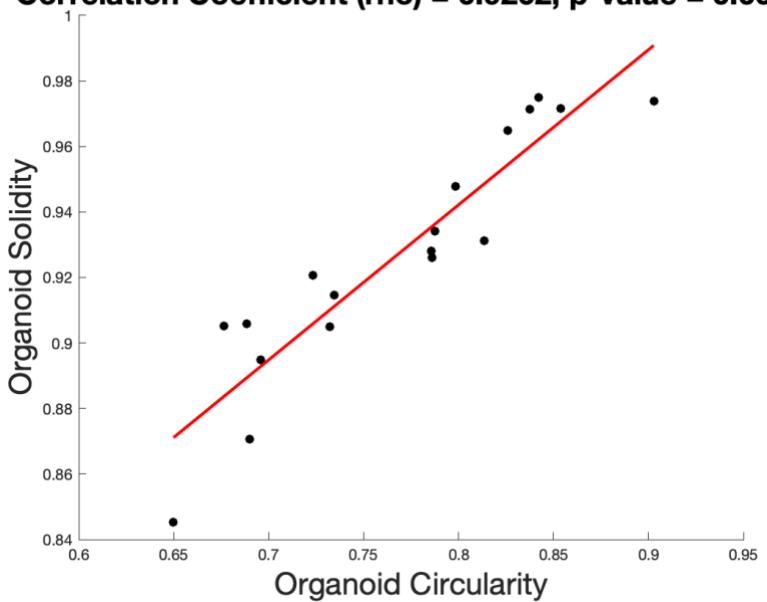
Correlation Coefficient (rho) = 0.5297, p-value = 0.0238



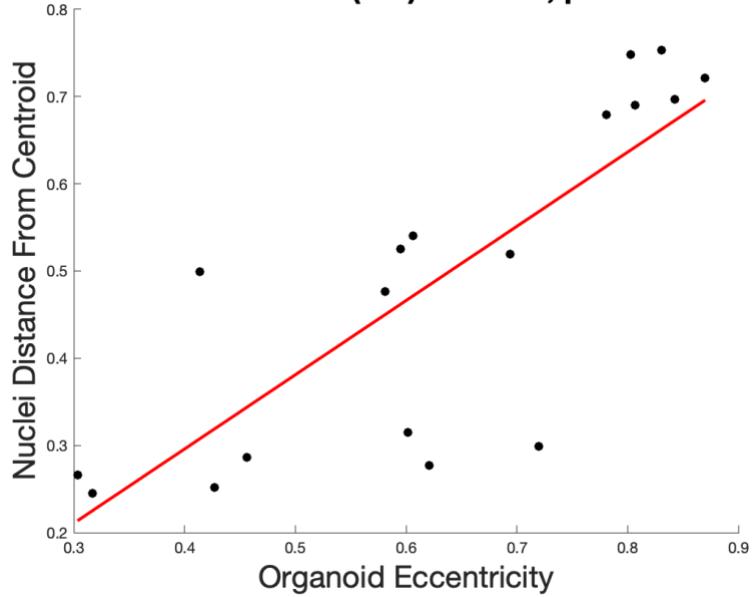
Correlation Coefficient (rho) = -0.6190, p-value = 0.0062



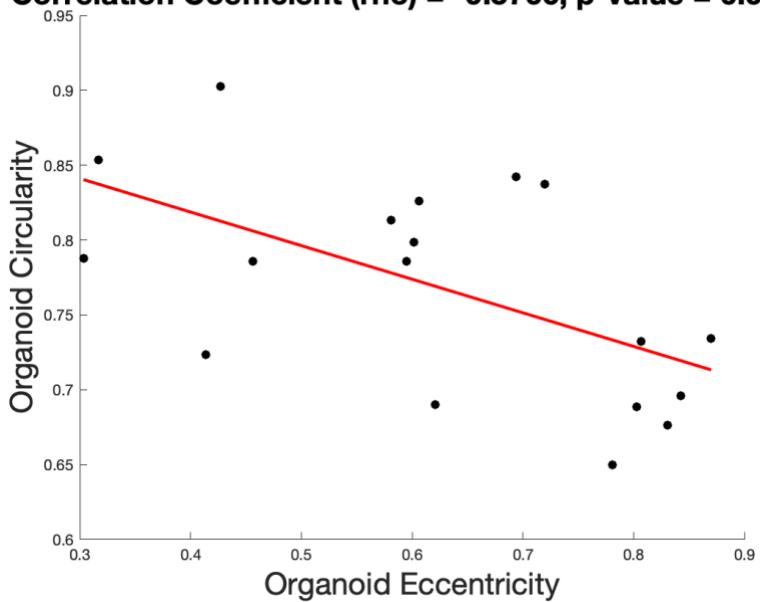
Correlation Coefficient (rho) = 0.9262, p-value = 0.0000



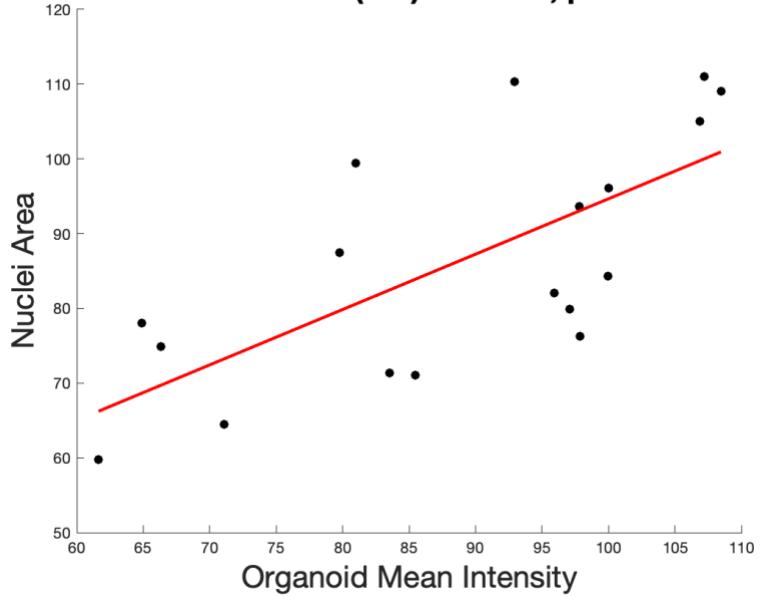
Correlation Coefficient (rho) = 0.8045, p-value = 0.0001



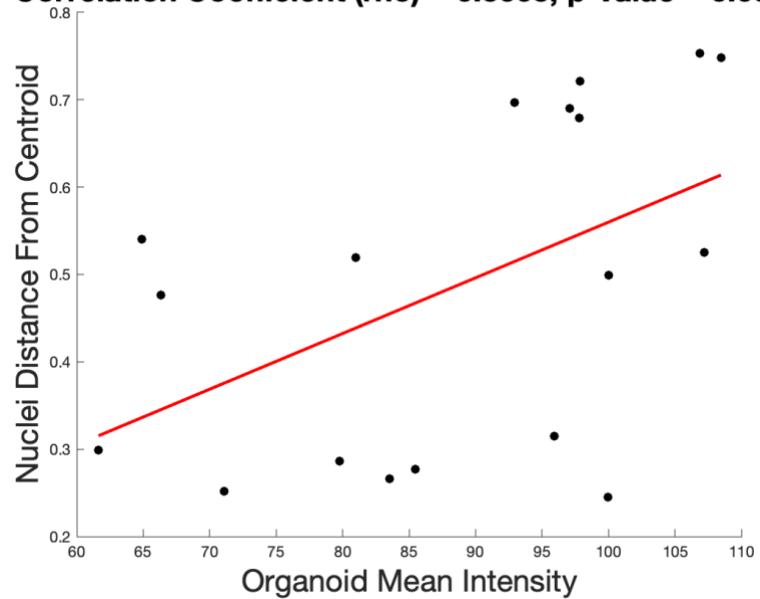
Correlation Coefficient (rho) = -0.5706, p-value = 0.0134



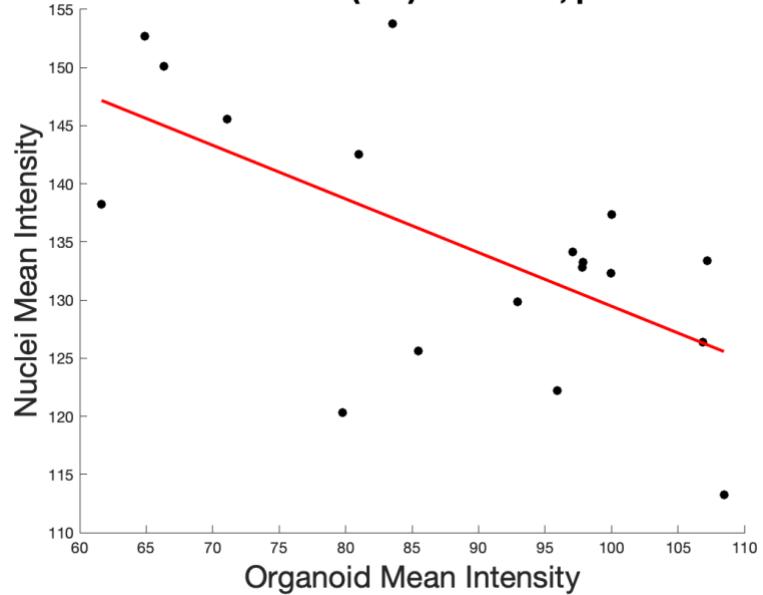
Correlation Coefficient (rho) = 0.7005, p-value = 0.0012



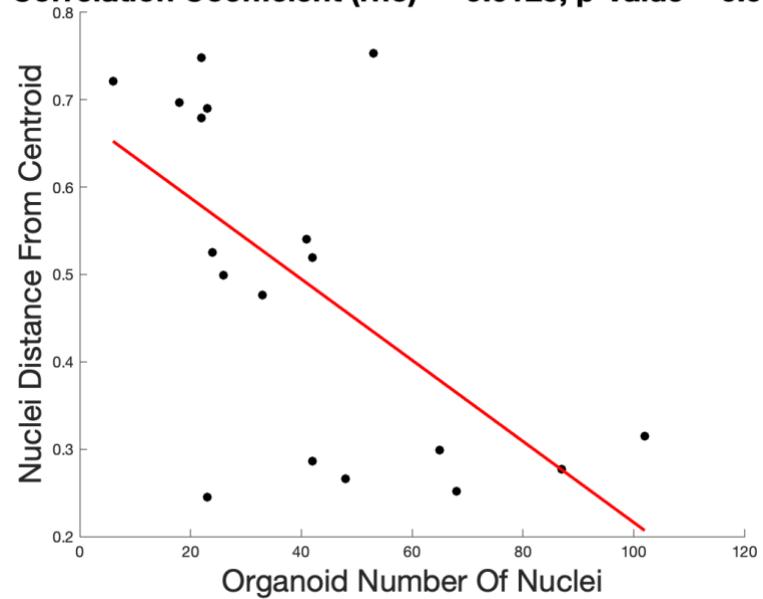
Correlation Coefficient (rho) = 0.5005, p-value = 0.0344



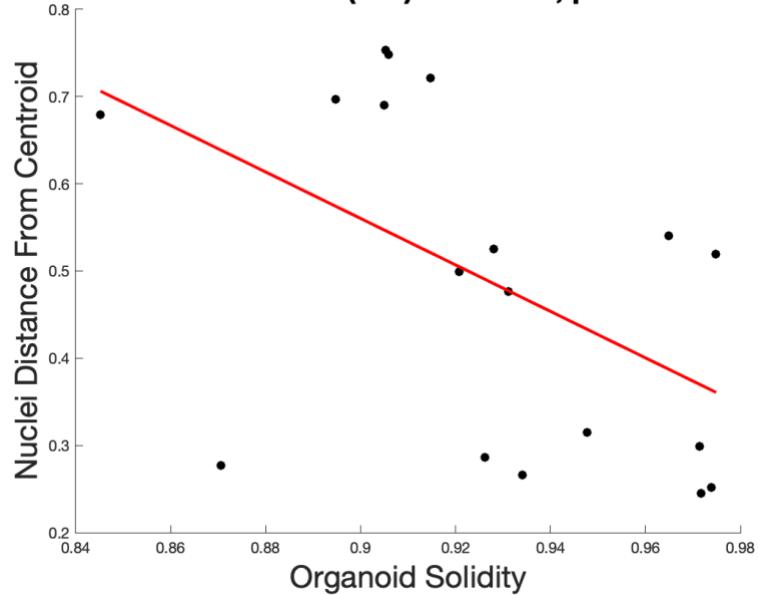
Correlation Coefficient (rho) = -0.6238, p-value = 0.0057



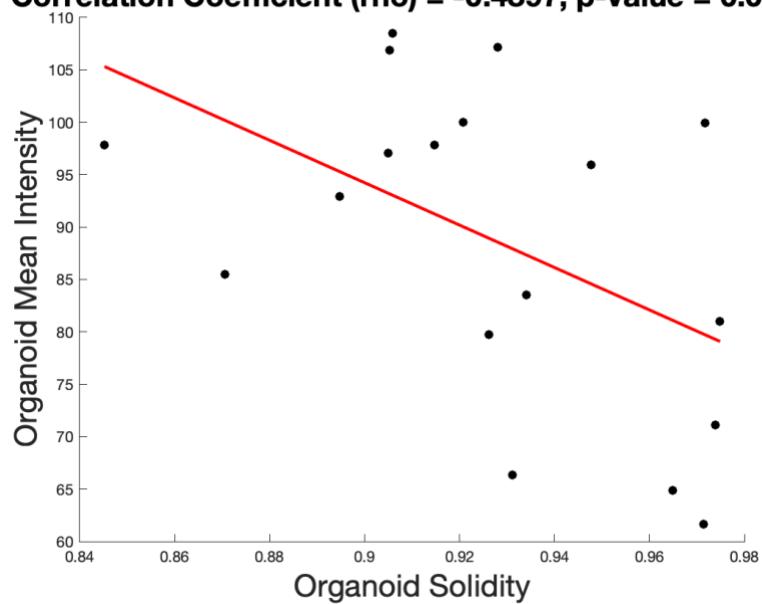
Correlation Coefficient (rho) = -0.6125, p-value = 0.0069



Correlation Coefficient (rho) = -0.5062, p-value = 0.0321

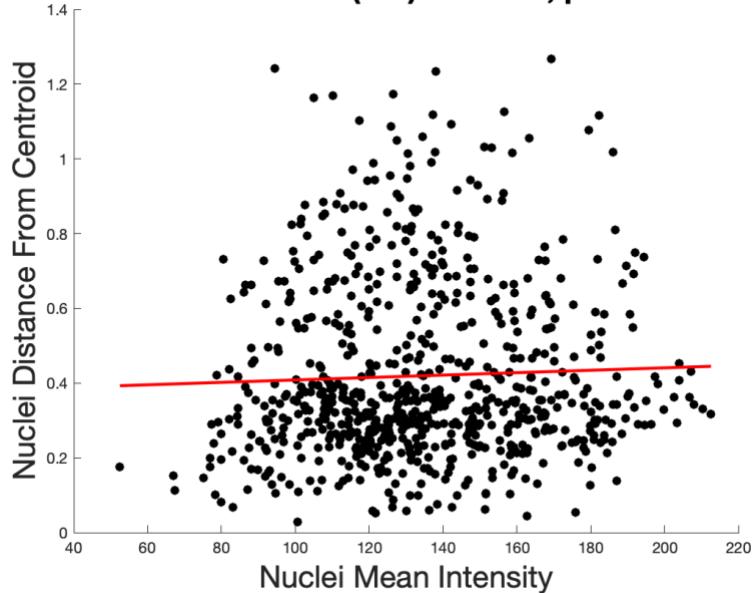


Correlation Coefficient (rho) = -0.4897, p-value = 0.0391

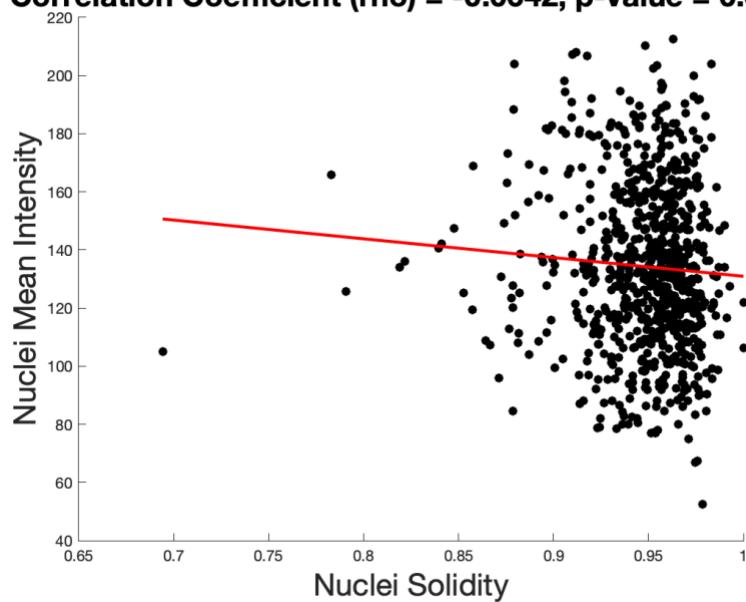


Uncorrelated

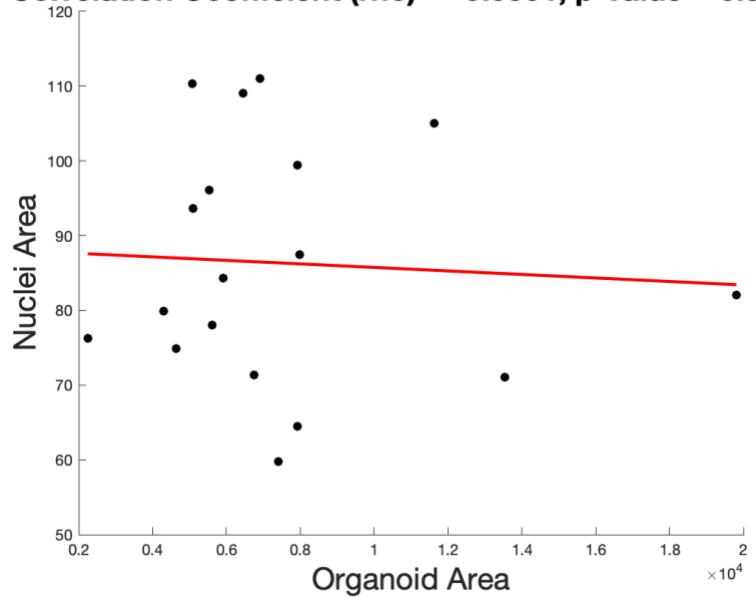
Correlation Coefficient (rho) = 0.0398, p-value = 0.2778



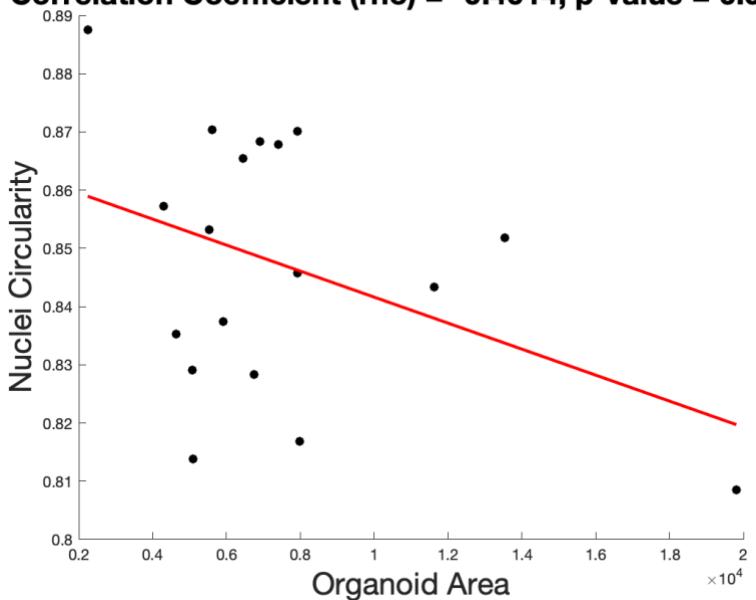
Correlation Coefficient (rho) = -0.0642, p-value = 0.0798



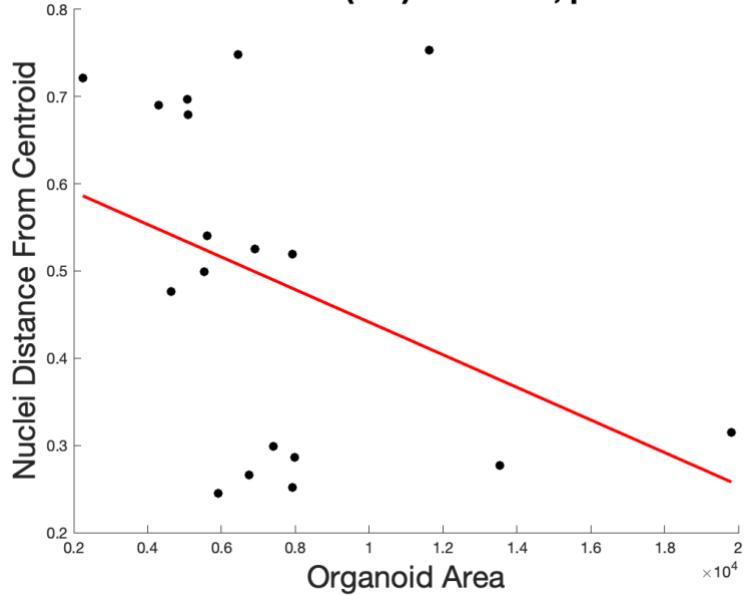
Correlation Coefficient (rho) = -0.0591, p-value = 0.8159



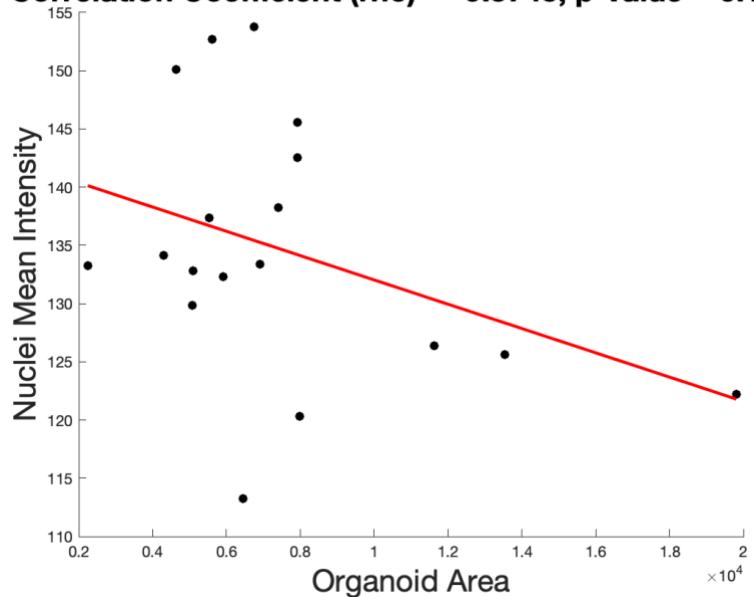
Correlation Coefficient (rho) = -0.4014, p-value = 0.0987



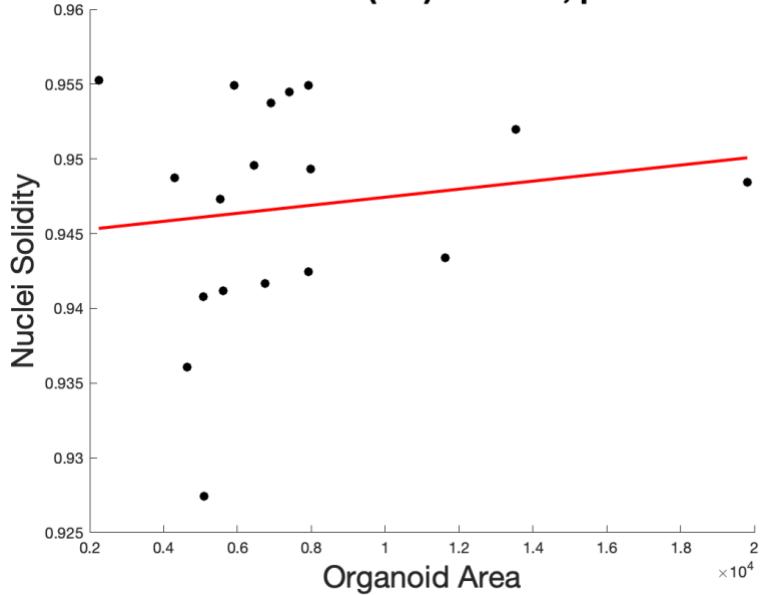
Correlation Coefficient (rho) = -0.3894, p-value = 0.1102



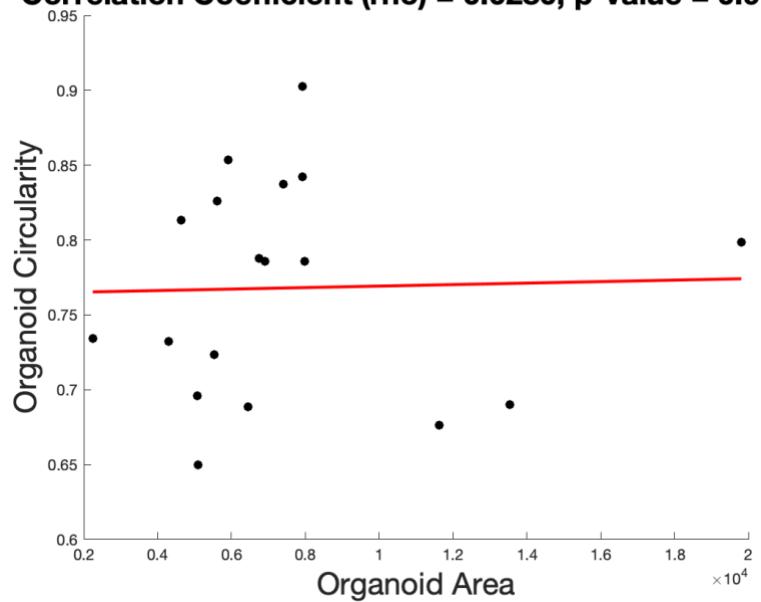
Correlation Coefficient (rho) = -0.3745, p-value = 0.1257



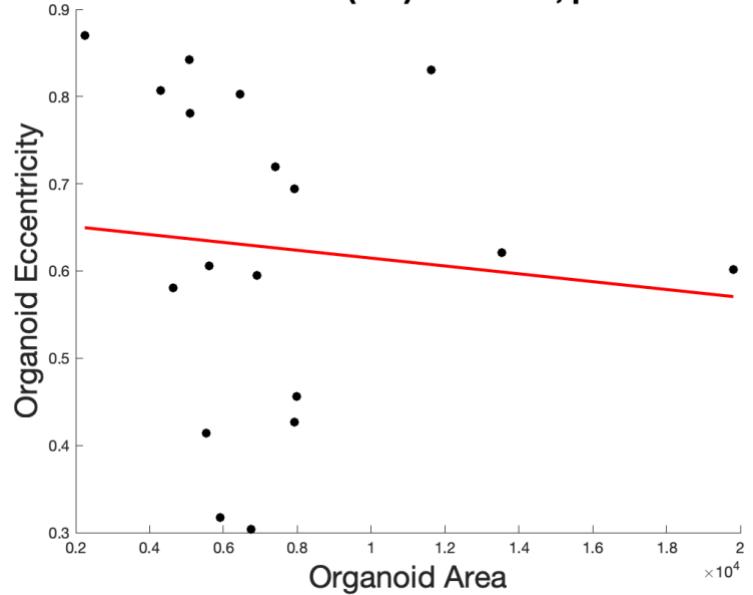
Correlation Coefficient (rho) = 0.1430, p-value = 0.5713



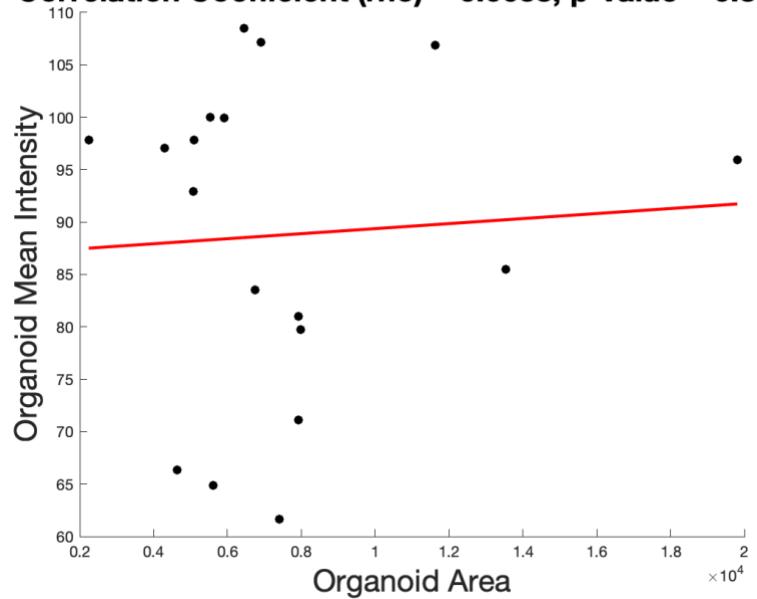
Correlation Coefficient (rho) = 0.0280, p-value = 0.9123



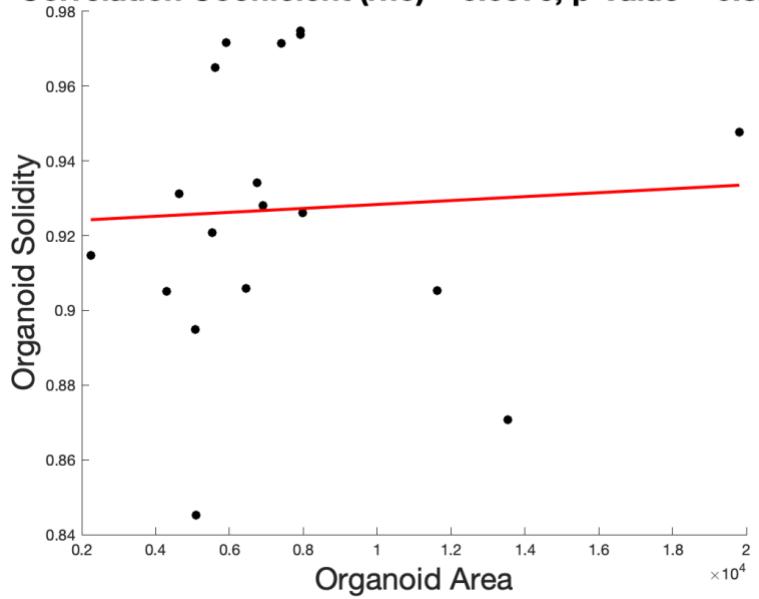
Correlation Coefficient (rho) = -0.0992, p-value = 0.6953



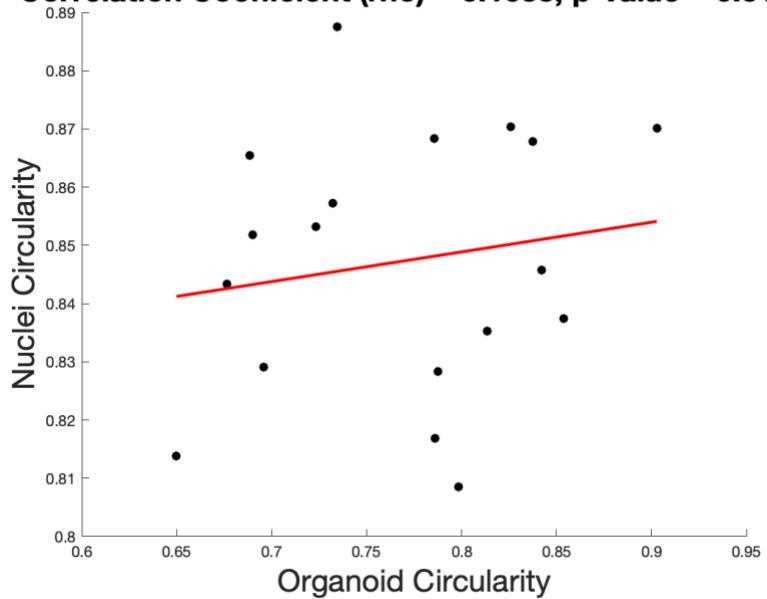
Correlation Coefficient (rho) = 0.0638, p-value = 0.8015



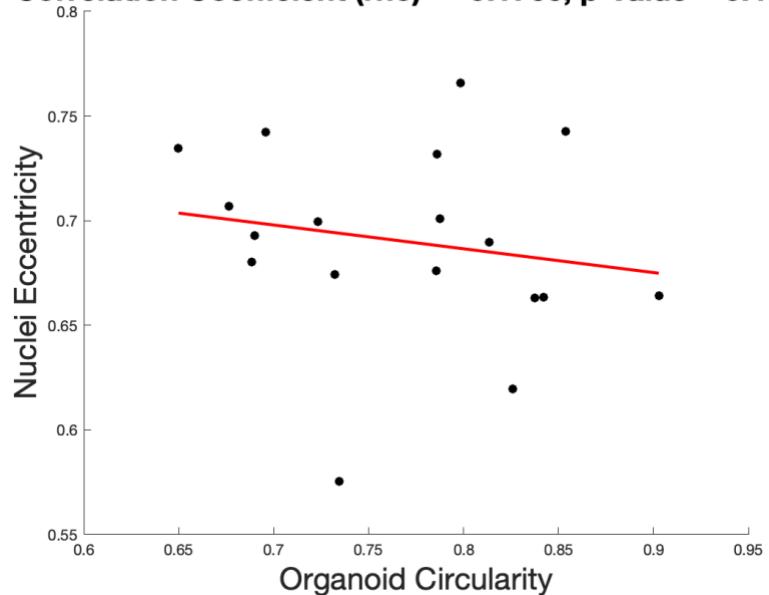
Correlation Coefficient (rho) = 0.0576, p-value = 0.8203



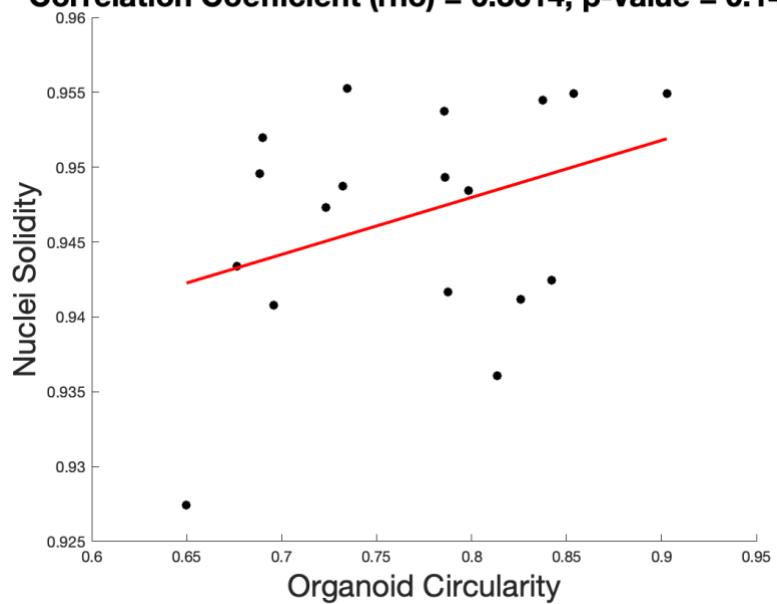
Correlation Coefficient (rho) = 0.1633, p-value = 0.5172



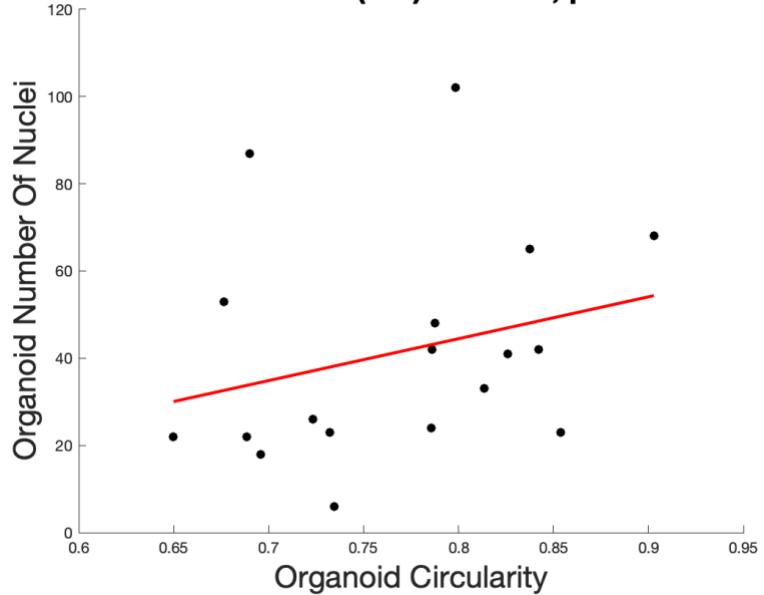
Correlation Coefficient (rho) = -0.1766, p-value = 0.4834



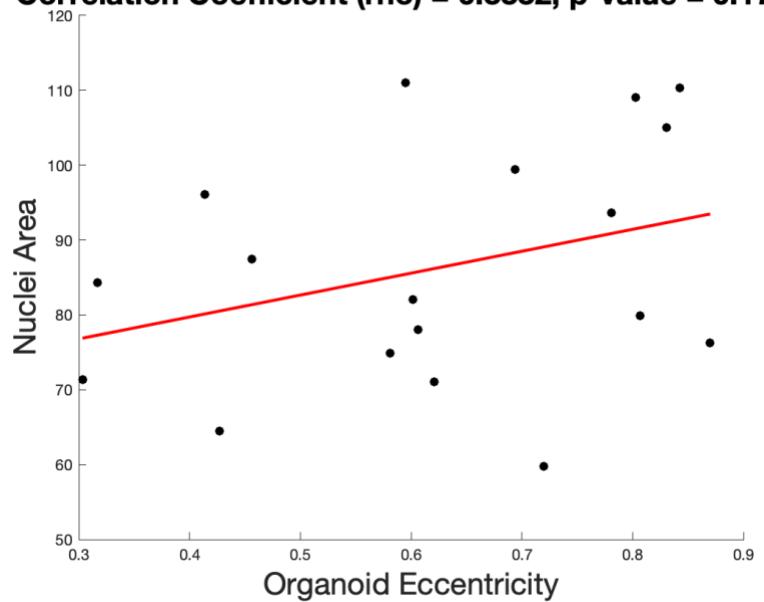
Correlation Coefficient (rho) = 0.3614, p-value = 0.1406



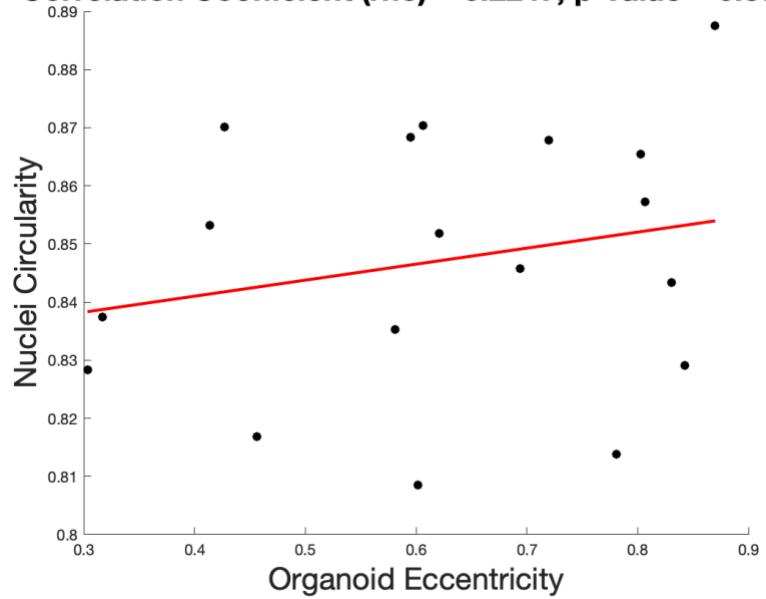
Correlation Coefficient (rho) = 0.2705, p-value = 0.2776



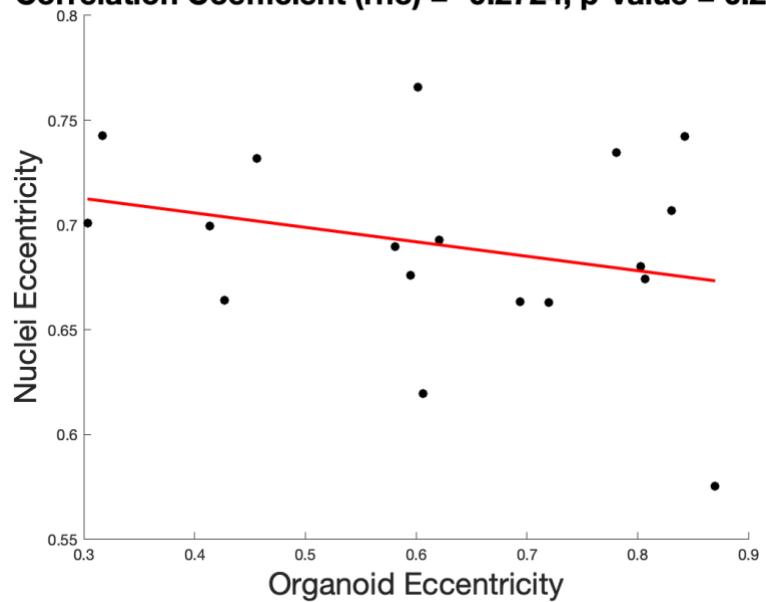
Correlation Coefficient (rho) = 0.3332, p-value = 0.1767



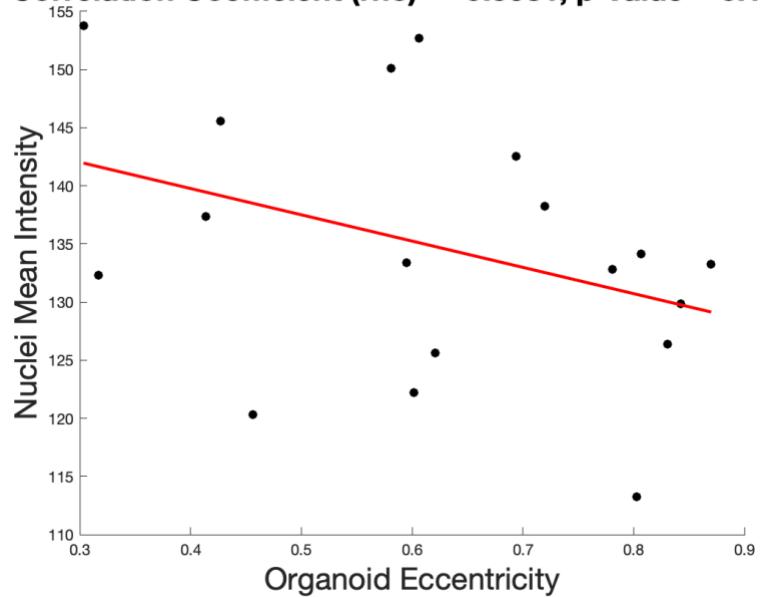
Correlation Coefficient (rho) = 0.2247, p-value = 0.3701



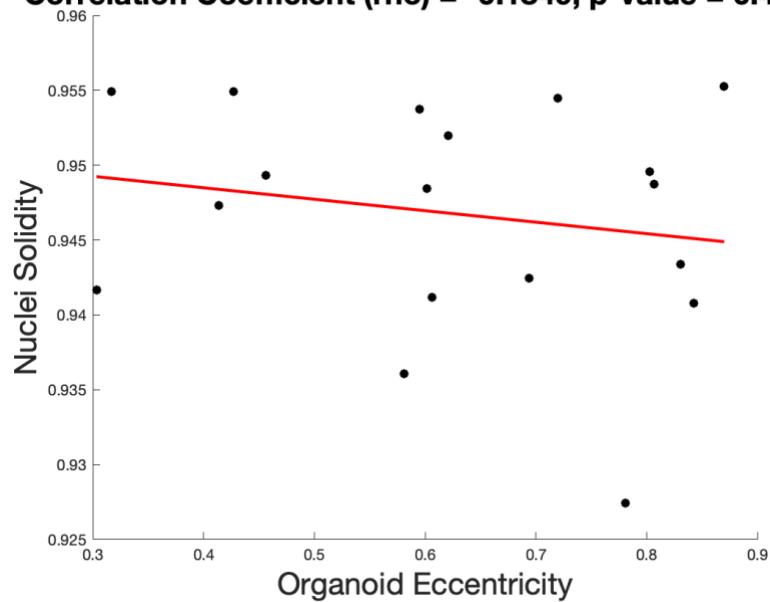
Correlation Coefficient (rho) = -0.2724, p-value = 0.2742



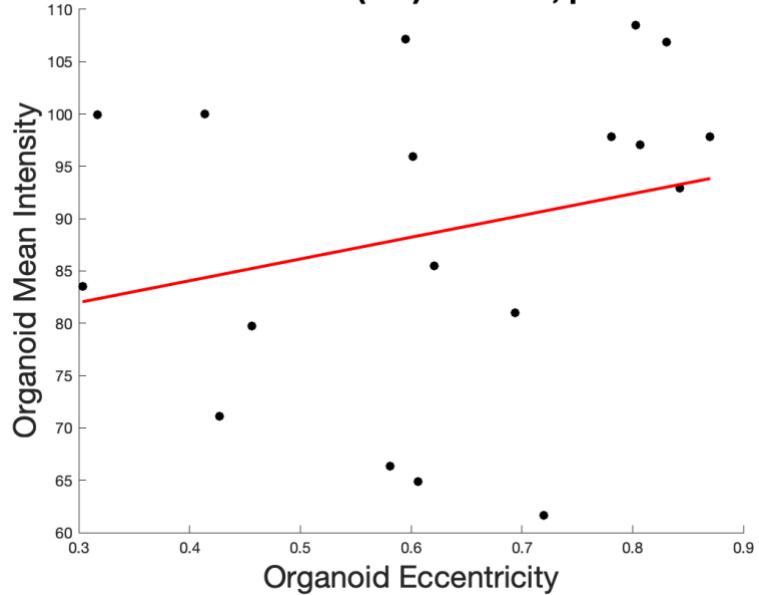
Correlation Coefficient (rho) = -0.3681, p-value = 0.1329



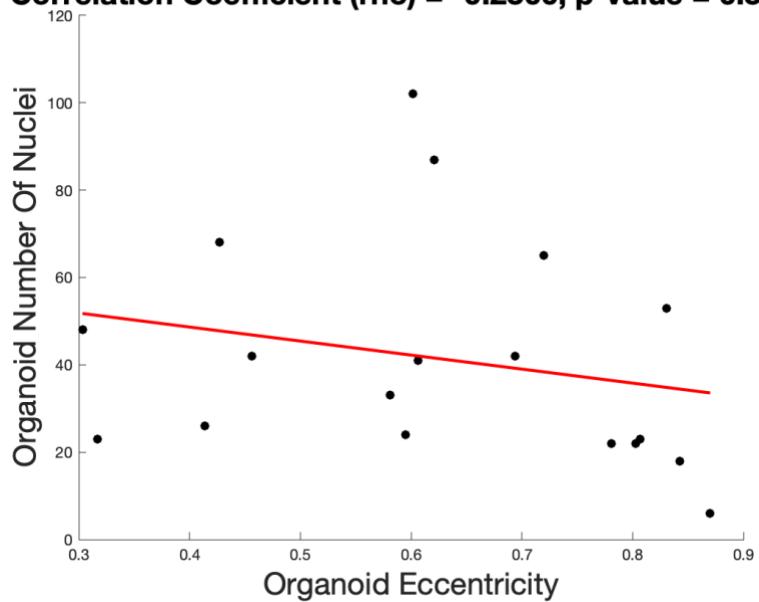
Correlation Coefficient (rho) = -0.1849, p-value = 0.4627



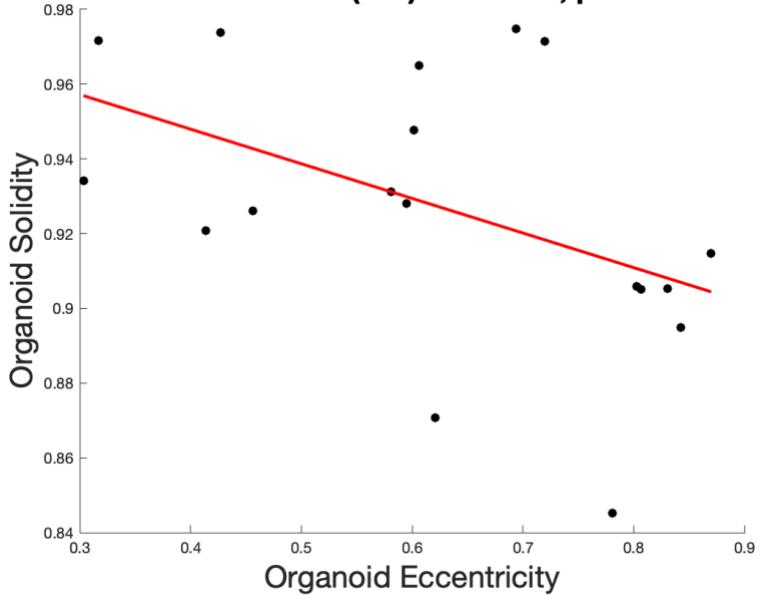
Correlation Coefficient (rho) = 0.2503, p-value = 0.3164



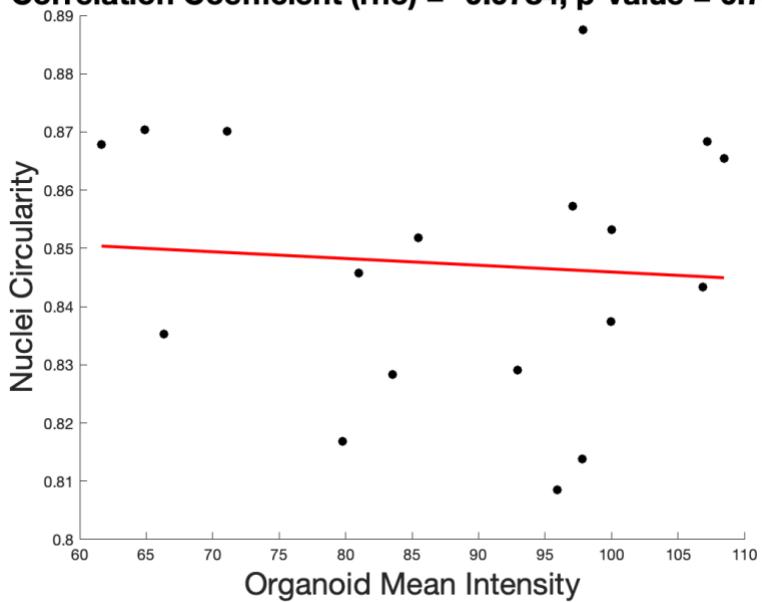
Correlation Coefficient (rho) = -0.2300, p-value = 0.3585



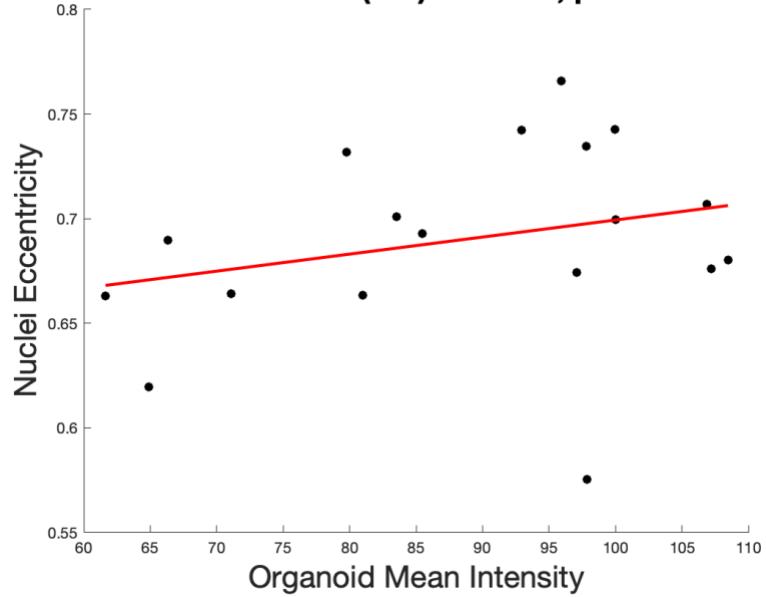
Correlation Coefficient (rho) = -0.4608, p-value = 0.0543



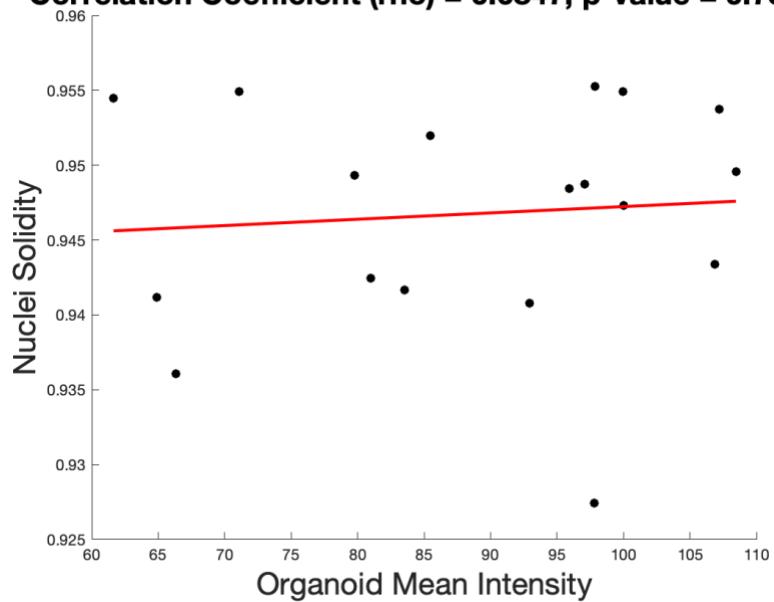
Correlation Coefficient (rho) = -0.0784, p-value = 0.7570



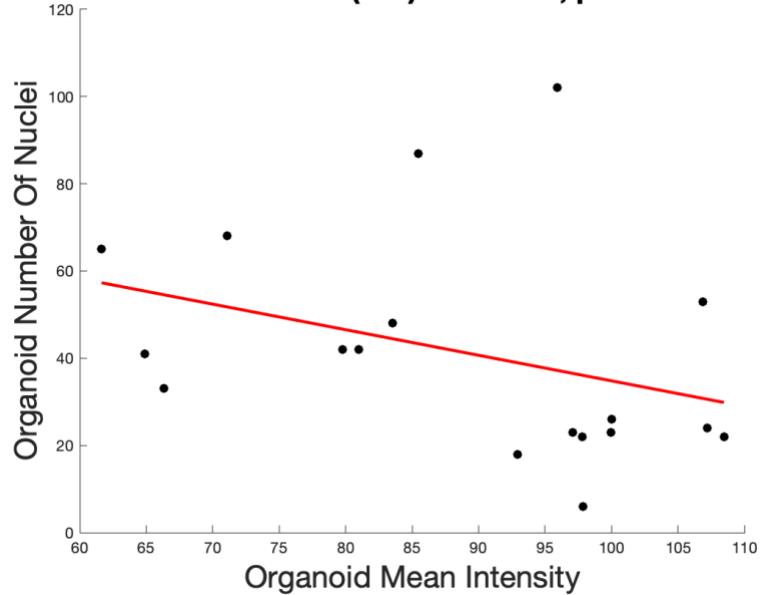
Correlation Coefficient (rho) = 0.2674, p-value = 0.2834



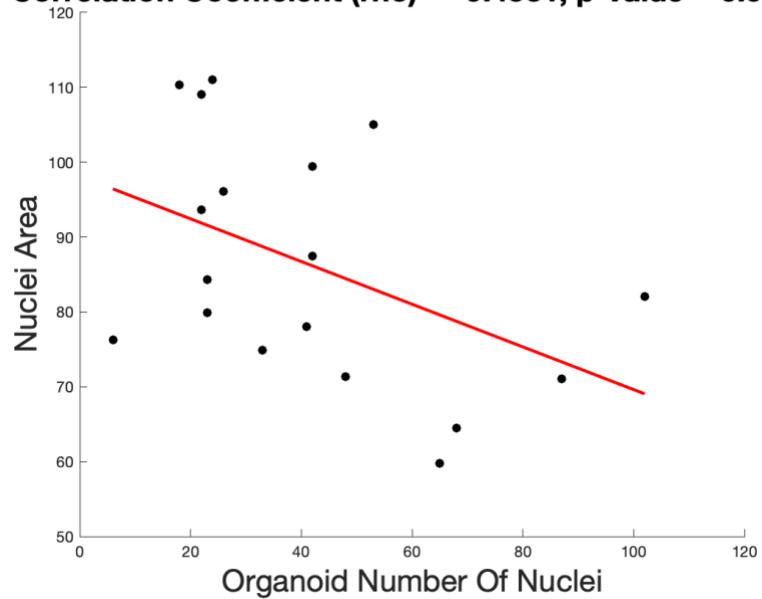
Correlation Coefficient (rho) = 0.0847, p-value = 0.7384



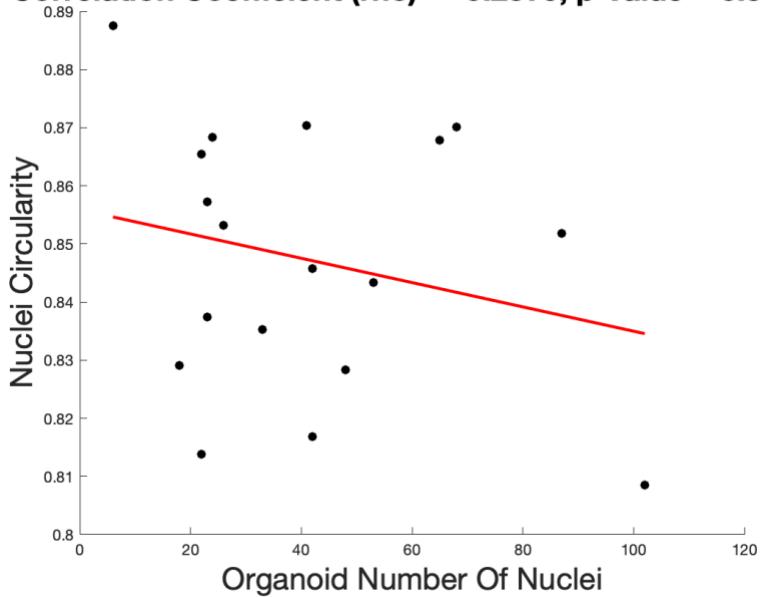
Correlation Coefficient (rho) = -0.3489, p-value = 0.1559



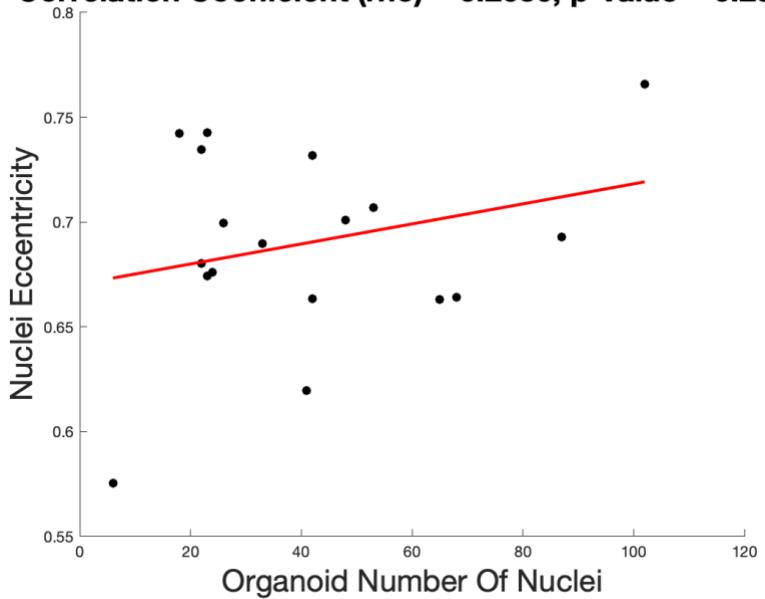
Correlation Coefficient (rho) = -0.4531, p-value = 0.0590



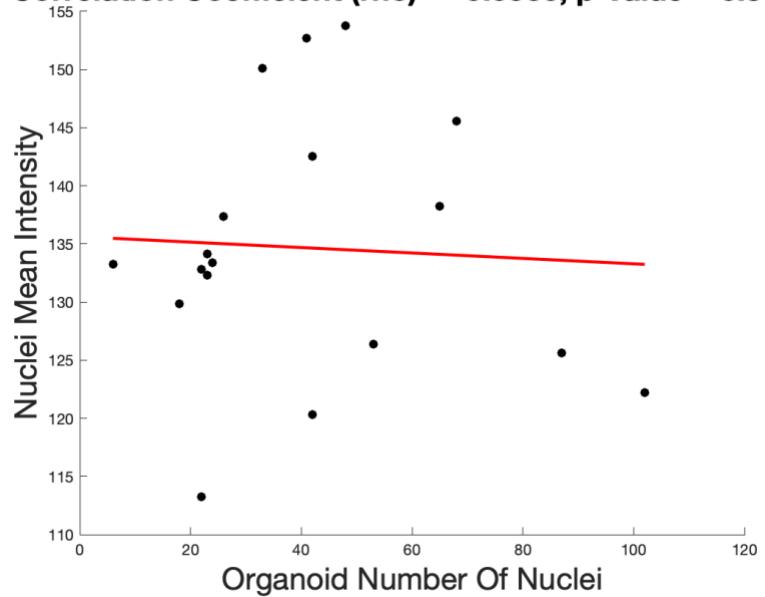
Correlation Coefficient (rho) = -0.2379, p-value = 0.3418



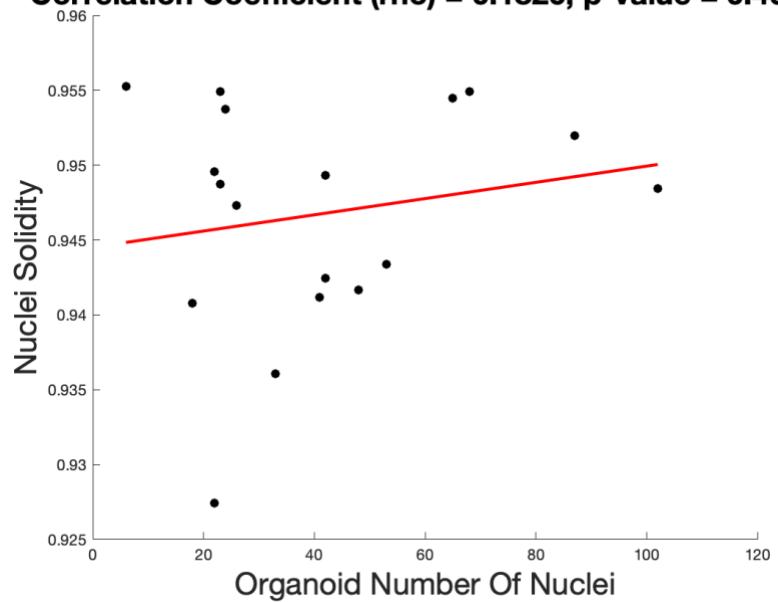
Correlation Coefficient (rho) = 0.2636, p-value = 0.2906



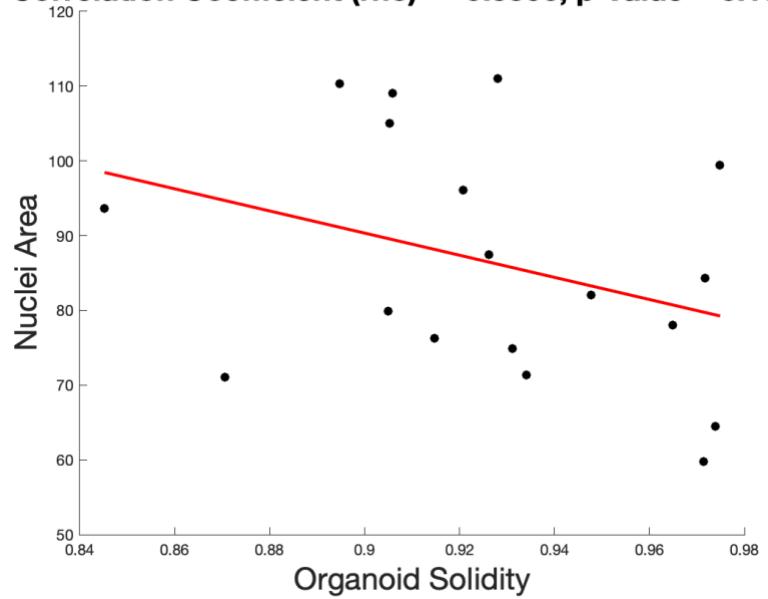
Correlation Coefficient (rho) = -0.0530, p-value = 0.8345



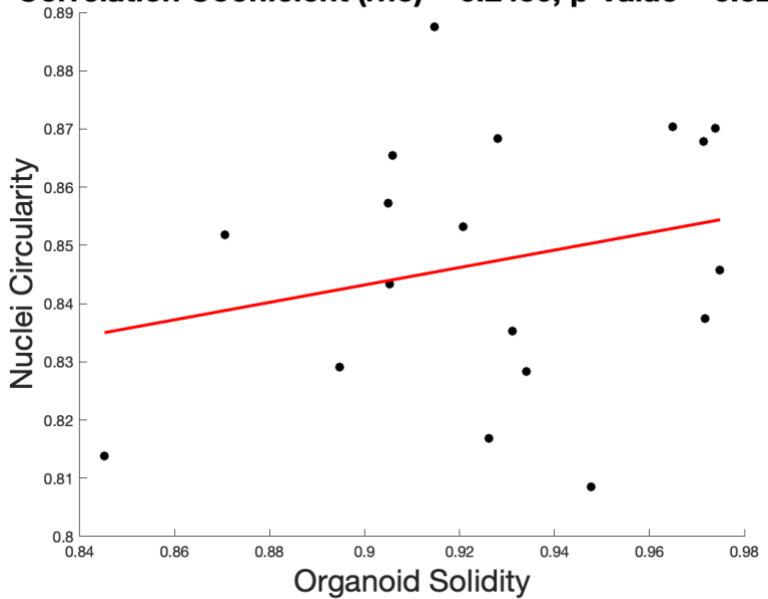
Correlation Coefficient (rho) = 0.1826, p-value = 0.4683



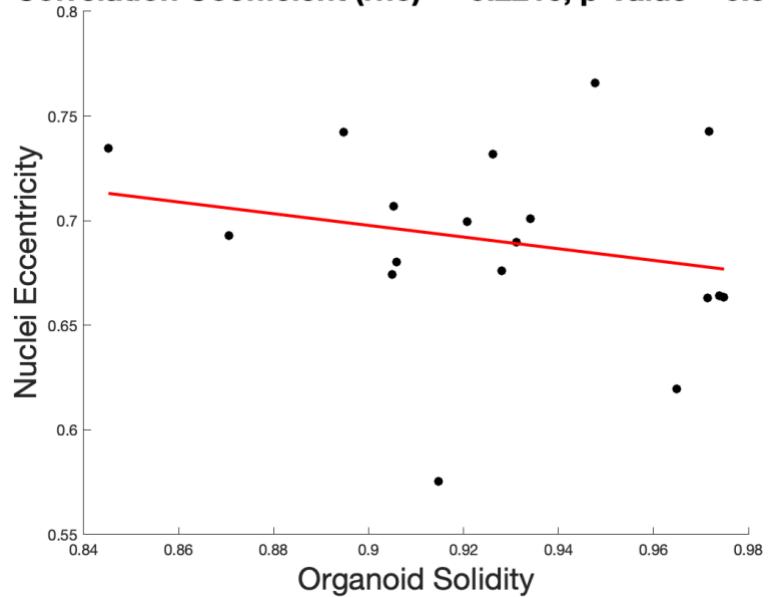
Correlation Coefficient (rho) = -0.3390, p-value = 0.1688



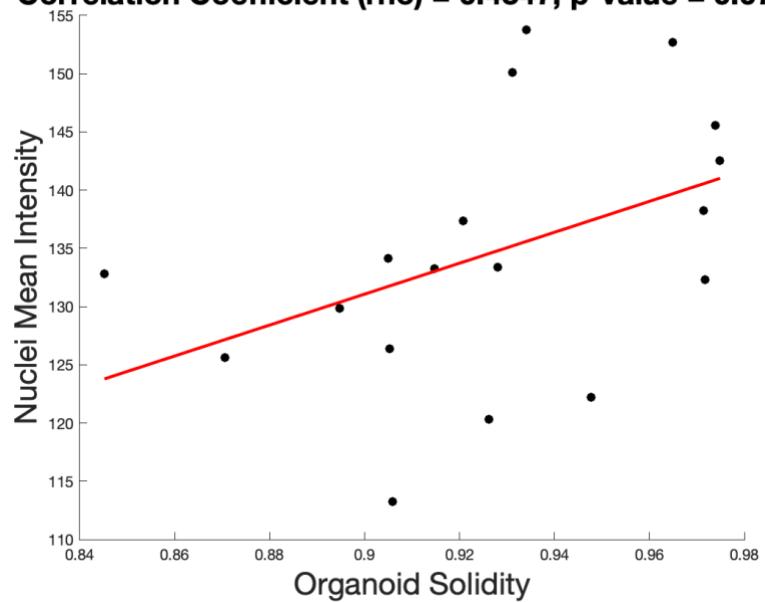
Correlation Coefficient (rho) = 0.2450, p-value = 0.3271

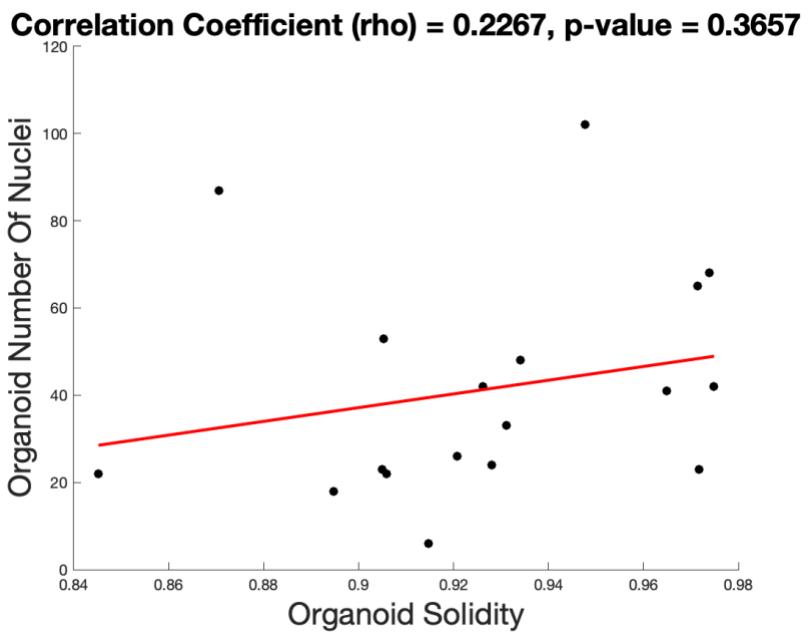
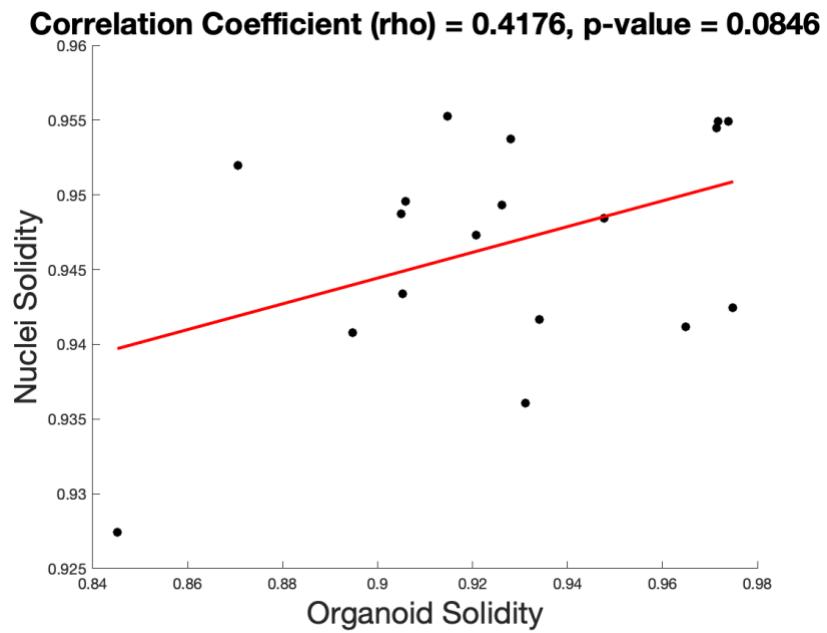


Correlation Coefficient (rho) = -0.2216, p-value = 0.3769



Correlation Coefficient (rho) = 0.4347, p-value = 0.0715





8.2 Appendix B

Area

Group1	Group2	LowerCI	MeanDifference	UpperCI	PValue
1	2	-10.757892	-0.6568505	9.44419109	0.9872762
1	3	7.43506221	24.8501833	42.2653045	0.0023779
2	3	5.92325363	25.5070338	45.0908141	0.00641969

Circularity

Group1	Group2	LowerCI	MeanDifference	UpperCI	PValue
1	2	0.03598995	0.05871577	0.08144158	4.17E-09
1	3	-0.1013684	-0.062187	-0.0230056	0.00058417
2	3	-0.1649633	-0.1209027	-0.0768422	3.73E-10

Distance from centroid

Group1	Group2	LowerCI	MeanDifference	UpperCI	PValue
1	2	-0.0270174	0.03944649	0.10591039	0.34559507
1	3	-0.0481156	0.06647429	0.18106415	0.36231463
2	3	-0.1018316	0.02702781	0.15588724	0.87532257

Eccentricity

Group1	Group2	LowerCI	MeanDifference	UpperCI	PValue
1	2	-0.0968554	-0.056152	-0.0154486	0.00350581
1	3	0.08032627	0.15050267	0.22067906	1.49E-06
2	3	0.12773944	0.20665471	0.28556998	2.49E-09

Mean intensity

Group1	Group2	LowerCI	MeanDifference	UpperCI	PValue
1	2	-25.629243	-17.696329	-9.7634159	5.12E-07
1	3	-41.135702	-27.458633	-13.781563	7.57E-06
2	3	-25.142542	-9.7623033	5.61793556	0.29684436

Solidity

Group1	Group2	LowerCI	MeanDifference	UpperCI	PValue
1	2	0.00856941	0.01656216	0.02455491	3.57E-06
1	3	-0.0247613	-0.0109811	0.00279916	0.14815377
2	3	-0.0430395	-0.0275432	-0.012047	9.20E-05