



**HAL**  
open science

# Analysis and modelling of spatial organization of 3D tissues in biological images

Hoa Tran Thi Nhu

► **To cite this version:**

Hoa Tran Thi Nhu. Analysis and modelling of spatial organization of 3D tissues in biological images. Bioinformatics [q-bio.QM]. Sorbonne Université, 2018. English. NNT : 2018SORUS457 . tel-02924852

**HAL Id: tel-02924852**

**<https://theses.hal.science/tel-02924852v1>**

Submitted on 28 Aug 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE DE DOCTORAT DE  
L'UNIVERSITÉ PIERRE ET MARIE CURIE**

Spécialité

**Informatique**

École doctorale Informatique, Télécommunications et Électronique (Paris)

Présentée par

**Hoa TRAN THI NHU**

Pour obtenir le grade de

**DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE**

Sujet de la thèse :

**Analyse et modélisation 3D de l'organisation spatiale  
des tissus dans des images biologiques**

soutenue le 30 janvier 2018

devant le jury composé de :

Dr. Jean-Louis DILLENSEGER	Rapporteur
Dr. Xavier DESCOMBES	Rapporteur
Dr. Sylvie SCHNEIDER-MAUNOURY	Examinatrice
Dr. Cristina MANFREDOTTI	Examinatrice
Dr. Germain TRUGNAN	Examinateur
Dr. Thomas BOUDIER	Directeur de thèse



**DOCTOR OF PHILOSOPHY (PhD) THESIS OF  
PIERRE AND MARIE CURIE UNIVERSITY**

Specialization

**Computer Science**

Doctoral school : EDITE 130 (Paris)

Presented by

**Hoa TRAN THI NHU**

Submitted for partial requirement of

**Doctor of Philosophy from  
UNIVERSITY OF PIERRE AND MARIE CURIE**

Thesis subject :

**Analysis and modelling of spatial organization of 3D  
tissues in biological images**

defended on January 30 2018

comity composed of :

Dr. Jean-Louis DILLESEGER	Reviewer
Dr. Xavier DESCOMBES	Reviewer
Dr. Sylvie SCHNEIDER-MAUNOURY	Examiner
Dr. Cristina MANFREDOTTI	Examiner
Dr. Germain TRUGNAN	Examiner
Dr. Thomas BOUDIER	Supervisor

# *Abstract*

Cells are the basic units of living organisms that make up a tissue. Cells within the tissue preferentially form a network or a community that works together to carry out a specific function. Thus, the role of a tissue is affected by its cell types as well as the architecture of cellular interactions. A fundamental research question that can be raised is at what degree the spatial organization of these cells affects the function of the tissue. Recent research shows that the study of cells within a tissue in three dimensional (3D) space provides a better and more realistic way to explore the spatial organization with regard to the physical and biochemical properties of the entire micro-environment compared to other two dimensional approaches.

Confocal microscopy techniques is a standard method developed to capture the 3D fluorescent image volume of the tissue. It can provide high resolution images to visualize the structure of cells within the tissue. Nevertheless, existing analysis working on the spatial organization of tissues are quite limited and often provide only some very basic processing and analysis features. They also lack the capability to analyze the interaction among cells and their organization within the tissue.

In this thesis, we first propose a set of methodologies to analyze the multi-cellular structure of tissues at both local and global scale. The goal is to analyze, formalize, and model the spatial organization of the tissue captured by fluorescence microscopy images. At the local scale, we investigate the spatial relationship of several structures with both direct and indirect cellular interactions. At the global scale, we apply spatial statistic approaches to investigate the degree of randomness of the cell distribution.

In addition, an open source toolbox is developed to support all the computations in this thesis. Our toolbox allows researchers to perform investigations of the position of different cells within a 3D multicellular structure. In our case, we apply the toolbox to study of the spatial organization of the islet of Langerhans, a special kind of tissue that plays an important role in regulating the blood glucose level. Its spatial organization has been hypothesized to have a big impact upon the islet function. With a good segmentation accuracy, we have been able to perform our analysis of the islets of Langerhans on several different species such as mouse and monkey. We also utilize our toolbox to explore the structural-functional mechanism of the delta cell, a specific kind of cell within the islet whose role has not yet been determined, but could potentially influence the islet function, in mouse and human. Our generic toolbox is implemented with unbiased analytical capabilities in the widely used biomedical software platform ImageJ. We believe that it has the potential to help researchers explore the function of other complex structures.

*Keywords* : tissue spatial organization, 3D segmentation, islet of Langerhans, cell phenotype detection, spatial statistic.

# *Acknowledgements*

I would like to thank my supervisor, Prof. Thomas Boudier, for his guidance, his support, and his valuable advices throughout the whole period of my PhD. He has been a source of motivation for me to push myself go beyond my limits. I also want to give my kindest appreciations to all members of my laboratory IPAL, especially Prof. Mounir Mokhtari, Dr. Ludovic Roux for their warm welcome, the friendly working atmosphere, and for helping me with all the difficult administrative step to the completion of my study.

A special acknowledgment to Dr. Rafael Arrojo E. Drigo from Institute of Biological Engineering, Nanyang Technological University, Singapore who provided me the microscopic images and helped me with all his biological/medical knowledge. Without his help and encouraging discussions during my works, this thesis would not have been possible.

Special thanks to Prof. Lee Hee Kwan, Prof. Loo Lit Hsin from Bioinformatics Institute, Singapore for welcoming me as a student in bio-imaging department and giving me a good research environment.

I am also very thankful to all the members of Bioinformatics group for their interesting and useful discussions as well as for their very funny daily jokes. I have had a very good time with many special moments with them here in Singapore.

I would express my gratitude to my dear mother, father, brother, sister, and husband for supporting me with their love during my journey.

Finally, I would like to thank all my professors and my friends of promo 17 at IFI - Institute Francophone of Informatics who gave me a strong knowledge foundation for my current research.

# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgements</b>	<b>ii</b>
<b>Abbreviations</b>	<b>xvi</b>
<b>1 Résumé en Français</b>	<b>1</b>
<b>2 Introduction &amp; Literature Review</b>	<b>4</b>
2.1 Biological Background . . . . .	5
2.1.1 Tissue Organization . . . . .	5
2.1.2 Islets of Langerhans . . . . .	7
2.1.3 Methods to Study the Islet of Langerhans . . . . .	10
2.2 Image Analysis . . . . .	14
2.2.1 Confocal Microscopy . . . . .	14
2.2.2 Methods for Studying Tissue Organization . . . . .	16
2.2.3 Image Enhancement . . . . .	18
2.2.4 Nuclei Segmentation . . . . .	20
2.2.5 Mathematical Morphology . . . . .	24
2.2.6 Cell Segmentation . . . . .	28
2.2.7 Cell Phenotype Detection . . . . .	30
2.3 Spatial Tissue Organization Analysis . . . . .	31
2.3.1 Analysis at Local Scale . . . . .	32
2.3.2 Analysis at Global Scale . . . . .	33
2.3.3 Modeling of Spatial Organization . . . . .	36
2.4 Objectives of the Thesis . . . . .	37
<b>3 Materials and Methods</b>	<b>39</b>
3.1 Materials . . . . .	39
3.2 Methodology Overview . . . . .	40
3.3 Segmentation and Cell Phenotype Detection . . . . .	42
3.3.1 Nuclei segmentation . . . . .	42

3.3.2	Cell zone computation . . . . .	46
3.3.2.1	Cell membrane segmentation . . . . .	46
3.3.2.2	Cell zone determination - Watershed separation . . . . .	47
3.3.3	3D cell phenotype computation . . . . .	48
3.4	Analysis of Spatial Organization . . . . .	57
3.4.1	Cells composition and cellular interaction . . . . .	58
3.4.2	Analysis of spatial organization at global scale . . . . .	64
3.4.3	Cluster analysis . . . . .	65
3.5	Delta Cell Analysis and Modeling . . . . .	67
3.5.1	Analysis of spatial organization of delta-cells . . . . .	67
3.5.2	Modeling and simulation of dynamic morphology of delta cell . . . . .	68
3.6	Modeling of Spatial Organization . . . . .	71
3.6.1	Modeling of the spatial organization at the local scale . . . . .	73
3.6.2	Modeling of the spatial organization at global scale . . . . .	76
3.7	Developing Environment . . . . .	78
<b>4</b>	<b>Results</b>	<b>80</b>
4.1	Program Description . . . . .	80
4.1.1	Experimental datasets . . . . .	80
4.1.2	Developed toolbox to investigate tissue spatial organization . . . . .	82
4.2	Tissue Quantification and Analysis . . . . .	85
4.2.1	Nuclei Segmentation Result . . . . .	85
4.2.1.1	Parameters selection . . . . .	86
4.2.1.2	Nuclei segmentation result . . . . .	88
4.2.1.3	Comparison of nuclei segmentation methods . . . . .	90
4.2.2	Cell Zone Computation Results . . . . .	93
4.2.2.1	Cell segmentation using plasma membrane labeling . . . . .	93
4.2.2.2	Cell zone computation using watershed separation . . . . .	94
4.2.3	3D Cell Phenotype Computation . . . . .	98
4.2.3.1	Parameter selection . . . . .	98
4.2.3.2	Results of cell type identification . . . . .	101
4.2.4	Analysis of Cellular Interaction . . . . .	102
4.2.5	Analysis of Spatial Organization . . . . .	108
4.3	Delta Cell Analysis and Modeling . . . . .	114
4.3.1	Quantification and analysis of delta cells . . . . .	115
4.3.2	Results of modeling and simulation of dynamic morphology of delta cell . . . . .	116
4.4	Modeling Spatial Organization . . . . .	117

---

<b>5 Conclusion &amp; Discussion</b>	<b>123</b>
5.1 Discussion . . . . .	123
5.2 Conclusion . . . . .	128
<b>A Toolbox to Investigate Tissue Spatial Organization (TITSO)</b>	<b>130</b>
<b>B Tissue Quantification and Analysis Results</b>	<b>145</b>
<b>Bibliography</b>	<b>149</b>



# List of Figures

2.1	Pancreas in situ and zoom in of one islet of Langerhans. The figure is taken from the journal of microscopy [1]. . . . .	7
2.2	Example of different islets: mouse islet sample(A), monkey islet sample (B), human islet sample (C) with different regions of core region, mantle region, near blood vessels region (see the arrows) (A-, B-, C- scale bar, 50 $\mu\text{m}$ ). Zoom in different markers and cell type (scale bar, 5 $\mu\text{m}$ ): (D) alpha cell and glucagon cytoplasmic marker, (E) beta cells and insulin cytoplasmic marker, (F) delta cell and somatostatin marker. . . . .	9
2.3	Different strategies to use distance map. (A) Input image consists of two objects, (B) Euclidean distance map, measure from center of each object, (C) Normalized Euclidean distance map, measure from center of each object, (D) Euclidean distance map, measure from boundaries to the center of object. (E) Normalized Euclidean distance map, measure from boundaries to center of object, (F) Euclidean distance map, measure from boundaries of each object to the background. . . . .	25
2.4	Workflow of cell phenotypic profiling process. Figure is taken from the publication of Grys et al.(2016)[2]. . . . .	31
2.5	The RCC8D relations. The eight RCC8D relations. In each case, the relation shown is that of the square with NW–SE shading to the square with SW–NE. shading. Figure is taken from the publication of Randell et al.(2013)[3]. . . . .	33
2.6	Various types of spatial distribution. Positions can be independently distributed (completely random pattern), or attraction (aggregated pattern) or repulsion (regular pattern). Figure is taken from the publication of Andrey et al.(2010) [4]. . . . .	34
2.7	Illustrating of three specific distributions such as ‘polar’, ‘center’, ‘boundary’. Figure is taken from the publication of Weston et al.(2012)[5]. . . .	35
2.8	F-function, cumulative distribution function of the empty region . . . . .	37

3.1	Workflow of the processing. First, nuclei is segmented and cell zone is computed. From three markers glucagon, insulin, somatostatin as input image and detected cell zone, cell phenotype is identified for each cell. Based on the cell profiles from first stage, we investigate the spatial organization of islet of Langerhans. . . . .	42
3.2	Diagram of 3D spot segmentation. . . . .	43
3.3	Band pass filtering and its help to facilitate the nuclei seed detection. . .	44
3.4	Proposed workflow for cell segmentation which divide into nuclei segmentation and membrane segmentation and combine them together. (A) Nuclei is segmented using our spot segmentation method. (B) Raw membrane image is filtered using 3D median filter, after that apply 3D Canny Deriche Filter and normalization to capture the edges of membrane, and applied 3D mean filter to smooth intensity inside the edges and then applied 3D minimum filter to reduce the size of edge. (C) The watershed intensity driven approach is adopted using segmented nuclei as seed in the region of filtered membrane image. We obtained the cell region result. From cell region, we can get the contours of region as the cell wall.	47
3.5	The association between alpha- and beta- cells in cultured human islet cells, insulin (red) and glucagon (green). A cell pair composed of one alpha-cell (A) surrounded by one beta-cell (B). Merged image is shown in (C). The cell pair shown here is representative of cell pairs observed in different human cell preparations from at least 10 different pancreas. Figure is taken from the work of Bosco et al. (2010) [6]. . . . .	50
3.6	SLIC Clustering. (A-C) original data, scale bar $10\mu$ (D) overlay, arrow: the part containing multi markers. (E-H) Result of SLIC in 3D Viewer, E - 3D visualization gray scale, F - slice at background and noise, G - slice contained 3 markers, SLIC median value correspond to each channels. (I-K) volume test, correspond to small[50-100], medium[100-300], big[300-500], (L) overlay of 3 channels. (M-O) compactness test, (M) small compactness $m=1000$ , (N) medium compactness $m=5000$ , (O) big compactness $m=10000$ , (P) 3D view with different colors correspond to different supervoxels regions. . . . .	53
3.7	Scheme for 3D cell phenotype computation. (1)(2) Normalization of markers and SLIC Supervoxels clustering. (3) Labeling a cell phenotype or unlabeled for each voxel. (4) Defining a study area. (5) Detect cell type	56

3.8	Different strategies to determine a study area (A1, A2, A3, A4). (A)overlay of alpha-, beta-,delta-channel and nucleus boundary, and cell zone boundary. (B) result of SLIC supervoxels, nucleus boundary, and cell zone boundary. (C) Study area A1 (Euclidean distance inside nucleus boundary d1, Euclidean distance outside nucleus boundary d2), nucleus boundary, and cell zone boundary. (D)Study area A2 (Euclidean distance outside nucleus boundary d2), nucleus boundary, and cell zone boundary. (E) Study area A3 (Euclidean distance inside nucleus boundary d1), nucleus boundary, and cell zone boundary. (F) Study area A4 (entire watershed area), nucleus boundary, and cell zone boundary. Scale 2 $\mu\text{m}$ . . . . .	58
3.9	Cellular interaction computation with different input parameters of $NB_{neivox}$ neighborhood voxels of the current voxel $P(x, y, z)$ at the boundary. Three options of $NB_{neivox}$ is 6-, 18-, 26-neighboring voxels of the current voxel $P$ . Image result of cell zone determination-watershed separation of a mouse islet in our datasets, visualization in XY, YZ, XZ directions of slide 43 in total of 86 slices, each cell zone have one specific label. . . . .	59
3.10	The formulate of computation of cell network. Using our tools visual display of 3D network is done within ImageJ using the ImageJ 3D viewer plugin [7]. . . . .	62
3.11	Result of cell distance computation in Mouse1-b and Monkey1 islet, one slice is displayed, the distance is expressed as cell distance between all cells and a reference cell (displayed in white). Reference cell has distance 0, a given islet-cell “X” contacts directly a reference cell, “X” (dark purple cells) will be at distance equal to 1 from the reference cell. If another islet-cell “Y” contacts islet cell “X”, “Y” will then be at a distance value of 2 from the “reference” cell, and so on. Scale bar, 50 $\mu\text{m}$ . . . . .	63
3.12	Clusters of alpha cells in the example of mouse4 and monkey1 dataset (Table 4.1) (each cluster is marked by a circle) . . . . .	66
3.13	Delta cell excess to modulate beta cell function, example of the clear mouse4 islet. (A) somatostatin marker, (B) somatostatin with the tendency to contact beta cells at the center region of tissue (C-D) and (E-F) zoom in the label of delta cell that look like a filopodia-like structure (body and an arm) in a slide and 3D visualization. Scale bar, 20 $\mu\text{m}$ . . . . .	68
3.14	Delta cell and modeling of physical arm by two different methods: (1) using an ellipse and a vector, (2) using an affected range . . . . .	69

3.15	The diagram of delta cell protrusion simulation. The output of tissue quantification using our toolbox TITSO is used as the input for delta cell modeling and analysis. There are three options for establish a simulation: (i) all delta cells reach out to modulate the beta cells function at the same time, (ii) each delta cell reach out to modulate the beta cells function (with $n$ delta cells - program produce $n$ simulations), (iii) a specific delta cell reach out to modulate the beta cells function. . . . .	70
3.16	Workflow of modeling of the spatial organization. (i) Construct the cell model (ii) Define the study area (iii) Distribute cells into the study area by applying a modeling procedure (iv) Apply statistic tests to prove that the generated pattern follows a certain distribution (v) Visualization and spatial analysis. . . . .	72
3.17	The spatial relationship between the objects $S_1, S_2, \dots, S_n$ and between these objects with the domain research $P$ are represented in the table of relationship. Our goal is to generate randomly many objects in the domain research $P$ which satisfy the list of relationship rules. . . . .	73
3.18	Computation of possible zone for the center coordinate $C_i(x, y, z)$ of a generated object $S_i$ in the study area $P$ , arrow: the possible zone in one slice, and in 3D. . . . .	74
3.19	Steps to modeling of spatial organization at local scale. . . . .	75
3.20	Distribution function using Euclidean distance map with input parameter alpha in $[0, 1]$ . Results shown the possible zone obtained from distribution function with alpha = 0.9 - region is close to the existing object, and alpha = 0.2 - region is far from the existing object. . . . .	76
3.21	Algorithm to modeling the spatial organization at global scale. . . . .	77
4.1	Workflow of the analysis. Nuclei are segmented and cell boundaries computed from the DNA-label channel. The images corresponding to the different labels of the different cell types are combined together in a SLIC image (see text for details). This combined image is used to compute each cell type. From the interactions between cells, computed using the cell boundaries information, the cells network is built. . . . .	83
4.2	TITSO (Toolbox to Investigate Tissue Spatial Organization) - Plugin developed under ImageJ-image analysis platform for biomedical image. Figure show the graphical interface of TITSO under ImageJ (left side) and the dependancy libraries that TITSO used. . . . .	84

- 4.3 Comparison of 3D Nuclei Segmentation Methods. 1st column: Raw data showed one slice, 2nd: Results of 3D Iterative Thresholding method, 3rd: Segmented results of Farsight toolkit, 4th: Results of MINS platform, 5th: Results of our Spot Segmentation method. 1st row: Segmentation nuclei result of mouse sample, visualize in 3D using 3D Viewer plugin, under ImageJ platform. 2nd row: Segmentation nuclei result in monkey sample, visualize segmented contours of object using 3D ROI plugin under ImageJ platform. . . . . 91
- 4.4 Results of membrane processing, slide 19 of 53 slices, using Image Edge plugin, under ImageJ platform. Red arrow: blood vessel area, blue arrow: cell contours area. (A) original membrane labeling, (B) result of membrane contours, (C) result of normalized membrane contours, (D) result of membrane filter using normalized contours + mean filter radius 2-2-1 + 3D minimum filter, (E) result of membrane segmentation with a low threshold, (F) result of membrane segmentation with a high threshold. . . 94
- 4.5 Comparison of cell zone obtained from membrane detection + segmented nuclei and our segmented nuclei + 3D watershed separation. Nuclei DAPI original image (A), nuclei segmentation result (B), our computation of cell zones based on watershed and maximal radius of average cell size (here 11 pixels, 4.16 microns (C)). Membrane original image(D), merge processed membrane (see Section 4.2.2.1) + nuclei segmentation result (E). Intensity driven watershed - region growing using segmented nuclei as seed based on processed membrane labeling (F). . . . . 96
- 4.6 Result of the Voronoi with three different methods :(A) Method 1, region growing determined by centers of cell nuclei as seed, (B) Method 2, region growing determined by cell nuclei regions as seeds, (C) region growing determined by cell nuclei regions as seed, with constant speed - radius  $R_{cell}$  from boundaries of nuclei, slice 23 of 45, mouse islet sample. . . . . 97
- 4.7 Result of the computation of the cell zones around each nuclei with two different values of fixed maxima radius  $R_{cell}$  around each nuclei, 3.78 microns (A, B) and 7.56 microns (C, D). Result of computation of the average number of cellular interaction based on maxima radius (radius from 3 to 30 pixels). The number of cell to cell contacts average is quite stable in the interval [10, 15] pixels corresponding to the average radius of nucleus size. . . . . 98

- 4.8 Results of nuclei segmentation and cell type computation for the monkey1 dataset. (A) Merged image of raw data, only one slice is displayed (red=insulin, green=glucagon, blue=somatostatin, cyan=DAPI). (B) DAPI channel with overlaid contours of detected nuclei. (C) 3D visualization of segmented nuclei. (D) Merged raw image with overlaid contours of computed cell boundaries. (E) 3D SLIC supervoxels clustering result (red=insulin, green=glucagon, blue=somatostatin) (F) 3D visualization of computed cell types (red=beta-cells, green=alpha-cells, blue=delta-cells). Scale bar,  $50\mu\text{m}$ . . . . . 99
- 4.9 Result of the computation of the cell zones around each nuclei with 4 different values of fixed maximal radius around each nuclei ( $r = 4, 6, 8, 10$  pixels). A-, F-, K- overlay of 4 channels (glucagon-green, insulin-red, somatostatin-blue, and dapi-white) and nuclei ROI result. First row show experiment for  $\alpha$  cell, second row show experiment for  $\beta$  cell, third row show experiment for  $\delta$  cell which consist of overlay image, watershed ROI with radius 4, 6, 8, 10 pixels. . . . . 101
- 4.10 Cell-cell contact probabilities. Top: an example of result of cell phenotype computation (colored nuclei area, green - alpha cell, red - beta cel, blue - delta cell) and cell network (green alpha-alpha, red beta-beta, blue delta - delta, black - heterotypic interaction between alpha, beta and delta) at mouse 4 and monkey 1 dataset. Bottom: results of frequency of interaction between cells (6 mice,  $n=22$  datasets; 6 monkeys,  $n=12$  datasets) done on observed islet dataset, using random procedure or theoretical model presented in Hoang et al.(2014)[8]. . . . . 104
- 4.11 Relative proportions of direct cellular interactions between the two main cellular types alpha- and beta-cells for the extended datasets (6 mice,  $n=22$  datasets; 6 monkeys,  $n=12$  datasets). (A) Homotypic contacts in mouse and monkeys datasets (\*) denotes significant difference. (B-G) comparison of cellular interactions between extended mouse and monkey datasets and random models (\*) denotes significant difference. All statistics were done using ANOVA method with the Graphpad Prism V6.0 software. . . . . 107

- 4.12 The corresponding F-function, G-function cumulative distribution functions of the observed data and simulated random organization for alpha-cells within mouse4 and monkey1 islets, using Euclidean distance. F-function (A) and G-function (B) of observed mouse4 islet, F-function (C) and G-function (D) of simulated random organization of mouse4 islet. F-function (E) and G-function (F) of observed monkey1 islet, F-function (G) and G-function (H) of simulated random organization of monkey1 islet. In black (low) low envelope of 5% confidence interval, in black (high) high envelope of 95% confidence interval, in orange (theory) average value of 100 simulated random organization, in violet (observed), our observed data. Note that for observed data the curve is outside interval delimited by the 5%-95% confidence interval. . . . . 112
- 4.13 Cyto-architecture of delta cells in rodents and human islets. Delta cells send out cellular processes to increase their intra-islet reach. (A) and (B), maximum projection images of representative mouse (A) and a human islet (B) glucagon, insulin and somatostatin. Cell nuclei are shown in light blue. (C) Cellular composition of mouse (n=24 islets total, n=7 mice) and human islets (n=21 islets total, n=3 donors). (D) Maximum projection of a mouse islet and (E) a human islet stained with somatostatin labeling. The yellow arrowheads indicate cellular extensions from islet delta cells. (F) human delta cell image with somatostatin labeling). Yellow arrowheads indicate the delta cell extensions. (G) Scanning electron microscopy of a human islet. Somatostatin - delta cell is highlighted by a purple shade. (H) Physical length of the delta cell extension in its longest axis in vivo in mouse islets (green circles, n=47 cells) and in mice (black circles, n=31) and humans (purple squares, n=57) samples. (I) Relative gain in delta cell contact with islet beta cells and (J) alpha cells in mouse (pink line) and human islets (black lines). In (A) and (B), scale bar  $50\mu\text{m}$ . In (D), scale bar  $25\mu\text{m}$ . In (E) and (F), scale bar  $10\mu\text{m}$ . (Figure is taken from Fig. 1 of our revision publication [9]) . . . . . 114

4.14	Mapping delta cell contacts in mouse and human islets. (A) Overlay of insulin, glucagon and somatostatin marker in a human islet. (B) 3D mapping of beta, alpha and delta cells using the algorithm described in our plugin[10]. (C) cell distance map of beta cells from the nearest delta cell (purple). (D-I) result image showing the result of simulation of increasing lengths of delta cell arms and its influence on the relative beta cell (D-F) or alpha cell (G-I) contacts in human islets. White arrows indicate cells that are not contacted directly by a target delta cell but become within reach by increasing arm length of 15 and 27 $\mu\text{m}$ . (Figure is take from from Supplementary Fig. 3 of our revision publication [9]) . . . . .	115
4.15	Results of modeling & simulation of spatial organization of vesicles within cell area, visualize in 3D (Use case 1). Each column represent one type of distribution, followed by result of F-function (blue curve: our observed model, green curves: envelopes of 5% and 95% of 100 random models, red curve: average of 100 random models). (A)random distribution SDI = 0.65, (B) cluster distribution (attraction) SDI=0.02, (C) uniform distribution (repulsion) SDI=1.0. . . . .	118
4.16	Results of modeling & simulation of spatial organization of vesicles within cell area at multi-scales (Use case 2). (A) visualization of generated pattern in 3D, followed by result of F-function, SDI=0.01 (B), and G-function, SDI=0.73 (C). . . . .	119
4.17	Results of modelling & simulation of spatial organization of cells within monkey islet (Use case 3). (A)raw data, (B) random pattern, (C) clustered pattern(attraction), (D)regular pattern (repulsion). . . . .	120
A.1	Workflow of the TITSO (the Toolbox to Investigate Tissue Spatial Organization). . . . .	133
A.2	Pre-filtering and normalize labelling image sub menu. . . . .	134
A.3	Nuclei segmentation sub menu. . . . .	136
A.4	Cell zone detection sub menu. . . . .	136
A.5	3D SLIC Clustering sub menu. . . . .	137
A.6	3D SLIC 3 Channels sub menu. . . . .	138
A.7	Cell Type Detection sub menu. . . . .	139
A.8	Cellular interaction computation sub menu. . . . .	140
A.9	Random Organization sub menu. . . . .	141
A.10	Random Organization sub menu. . . . .	142
A.11	Spatial Statistic Analysis sub menu. . . . .	143
A.12	Modeling and simulation of dynamic morphology of a cell sub menu. . . . .	144



# List of Tables

2.1	Comparison of different related works on islet of Langerhans. The two first works are on the biological investigation. The four final works are on computational biology. $\checkmark$ : shows that the work took into account this analysis, X : it did not take into account this analysis. . . . .	13
2.2	Comparison between histological and fluorescence imaging systems . . . .	16
2.3	Comparison of different methods for analyzing spatial tissue organization	17
2.4	Review of 3D available nuclei segmentation methods . . . . .	22
3.1	Execution time of each experiments with small, medium, and large size of cluster size. Testing with monkey islet sample, $203 \times 203 \times 40 \mu\text{m}$ , $1024 \times 1024 \times 40$ pixels . . . . .	55
4.1	Origin and characteristics of mouse, monkey, and human pancreas datasets used in this study. * indicates serial histological sections. . . . .	82
4.2	Impact of nuclei diameter $[D_{min} - D_{max}]$ to segmentation result. Testing dataset of mouse3 and monkey6 islet (Table 4.1. . . . .	88
4.3	Nuclei segmentation results for the observed mouse and monkey datasets. * indicates serial histological sections. . . . .	89
4.4	Origin and characteristics of mouse7 and monkey7 nuclei labeling image using as testing data for the comparison of different nuclei segmentation methods. . . . .	91
4.5	The output of different methods(3D Iterative Thresholding, Farsight, MINS, our Spot Segmentation) for 3D nuclei segmentation of a nuclei labeling image of a crop-mouse7, and a nuclei labeling image of crop-monkey7 (Table 4.4). The actual number of nuclei with strong intensity level in mouse nuclei labeling image is approximately $600 \pm 20$ nuclei, and $700 \pm 20$ nuclei. . . . .	93

4.6	Effect of average cell size (for cell zone computation) to cell phenotype computation. When increasing the radius of cell size $R_{cell}$ , and use obtained cell zone as study area (strategy $S4$ ), number of cell phenotype is increased ('unlabeled' cell detected as a cell phenotype alpha, beta, delta) and also there are a shift of cell type between alpha-, beta-, and delta-cells (wrong identification of cell type). . . . .	100
4.7	Actual and predicted classifications result and overall accuracy (number in bold) of cell type computation for mouse and monkey datasets. . . . .	102
4.8	Cell composition of observed mouse and monkey islets. . . . .	103
4.9	Number of direct interactions, per cell, for all datasets. . . . .	106
4.10	Number of direct interactions between different islet-cell types in observed mouse and monkey islets. . . . .	106
4.11	Histogram of cell distances between alpha and beta cells in observed and simulated random organization - the average value of 100 simulated random organization for mouse and monkey islets, $d$ : distance unit. . . . .	109
4.12	Spatial descriptor index (SDI) for F-function and G-function, calculated using Euclidean distance and cell distance. . . . .	111
4.13	Cluster analysis for all datasets, with number of clusters in each dataset, average number of cells in clusters and maximum number of cells in a cluster. . . . .	111
B.1	Results of cell composition, cellular interaction, homotypic, and heterotypic contact on C57/BL6 mouse islets (8 weeks of age, $n = 6$ animals and 8 stacks) . . . . .	146
B.2	Result of cellular interaction of observed models, theoretical models and simulated random models on 7 monkey islets. . . . .	147
B.3	Results of cell composition, cellular interaction, homotypic, and heterotypic contact on human islets ( $n = 21$ islets total, $n = 3$ donors) . . . . .	148

# Abbreviations

<b>2D</b>	two-Dimensional space
<b>3D</b>	three-Dimensional space
<b>DAPI</b>	4, 6-diamino-2-phenylindole
<b>TITSO</b>	Toolbox to Investigate Tissue Spatial Organization

# Chapter 1

## Résumé en Français

### L'objectif

#### **Analyse et modélisation de l'organisation spatiale des tissus en 3D**

Au cours des dernières années, de nombreux travaux ont été menés pour analyser automatiquement des cellules dans des images de microscopie. Néanmoins, ces travaux se concentrent principalement sur l'application de traitement et d'analyse d'images sur des cellules individuelles. Il y a un manque d'un outil générique pour analyser les interactions entre les cellules et leur organisation spatiale au sein de tissus biologiques.

De plus, il existe une demande pour une approche qui est plus efficace pour gérer des images de cellules en 3D à haut débit. Cela vient du fait que l'étude des cellules dans l'espace tridimensionnel (3D) donne une meilleure impression et est plus réaliste en ce qui concerne les propriétés physiques et biochimiques du micro-environnement des cellules par rapport à des approches bidimensionnelles.

#### **Cyto-architecture des îlots de Langerhans**

L'îlot de Langerhans est un sujet de recherche ouvert en biologie. Il y a un intérêt considérable à analyser la structure multicellulaire de ce type de structures à l'échelle locale et à l'échelle globale. Cependant, il n'y a pas beaucoup de travaux sur l'analyse, la formalisation et la modélisation de leur organisation spatiale.

Par conséquent, ma thèse se concentre sur ces objectifs. De plus, nous avons collaboré avec les experts du domaine, Dr Rafael Arrojo E. Drigo, et Dr. Per-Olof Berggren qui ont beaucoup d'expérience sur l'étude des îlots de Langerhans.

## Contributions

Nous avons proposé un ensemble de méthodologies et un outil pour l'analyse de l'organisation des tissus dans des images biologiques. Cette combinaison d'outils logiciels se propose d'algorithmes pour i) l'identification automatique de plusieurs noyaux individuels et ii) des marqueurs cytoplasmiques, iii) la prédiction de la position de la membrane cellulaire, et enfin iv) la reconstruction du réseaux cellulaire. Nous avons appliqué notre outil pour étudier l'organisation spatiale de l'îlot de Langerhans en essayant de comprendre son mécanisme interne.

Nous avons aussi appliqué notre outil pour explorer la fonction des cellules de type delta cell dans l'îlot, dont le rôle n'est pas encore déterminé. Nous avons introduit une procédure générique pour modéliser l'organisation spatiale des cellules dans un tissu, qui peut être utilisée pour créer des modèles virtuels de tissus et créer des données à tester.

## Résultats

### Modèle de données expérimentales

Afin de tester la performance de notre outil, nous avons utilisé notre outil d'abord pour analyser des images confocales en 3D des îlots de Langerhans dans des coupes histologiques complètes ou en série de pancréas de deux différentes espèces de souris et de singe.

En plus des sections de tissus épais, nous avons analysé également l'îlot en entier, obtenu à partir des tissus optiquement éclaircis de deux souris.

- C57 / souris BL6 (8 semaines d'âge,  $n = 6$  animaux et  $n = 22$  îlots au total).
- Singe (*Macaca Fascicularis*, 9-12 ans,  $n = 6$  animaux et 12 stacks).

Enfin, afin d'explorer le mécanisme des cellules delta, nous avons testé aussi notre framework sur deux espèces différentes :

- C57 / souris BL6 (8 semaines d'âge,  $n = 7$  souris,  $n = 24$  îlots au total).
- Humains ( $n = 21$  îlots au total,  $n = 3$  donneurs).

Ici, nous démontrons que notre algorithme de détection de noyaux a donné de bons résultats avec plus de 98 % de précision. Notre outil peut être appliqué sur des sections épaisses (30 micromètres) de pancréas de souris et de singe, et a également été validé avec succès dans un pancréas entier de souris.

En plus d'être capable de détecter et d'identifier avec succès les trois principaux types de cellules dans les îlots de souris et de singes, nous avons identifié pour la première fois un modèle de distribution en agrégats (c'est-à-dire une organisation non aléatoire) de cellules alpha chez la souris et chez le singe. Nos résultats favorisent l'hypothèse que les interactions entre les cellules endocriniennes dans les îlots de Langerhans ne se produisent pas au hasard, supportant l'idée que l'organisation spatiale des cellules est importante pour contrôler le bon fonctionnement des îlots et les fonctions des îlots. Il est important de noter que toute détection et analyse dans notre outil est effectuée de manière non biaisée et automatisée.

Par ailleurs, nous avons opté pour la mise en œuvre de nos algorithmes dans une plateforme logicielle largement utilisée (ImageJ). Nous pensons que notre outil a le potentiel pour être utilisé largement pour l'étude de la cyto-architecture des îlots de Langerhans et des structures multicellulaires en général.

## Conclusion

Nous avons proposé le développement d'un nouvel ensemble de méthodes pour visualiser, analyser et comparer les différentes organisations de structures multicellulaires dans les îlots tissulaires de souris, de singes et d'humains. Notre outil a fourni un moyen automatique pour analyser des îlots de manière non biaisée.

Les différences fondamentales entre les îlots de souris et de primates supportent l'hypothèse selon laquelle la physiologie et les architectures des îlots de souris et de primates sont différentes.

Cette étude met en évidence les différences entre les îlots de souris et de primates. Dans les études d'îlots, au lieu d'utiliser des îlots de souris, il faut utiliser des îlots de singes et d'humains directement pour une compréhension plus significative de la patho-physiologie du diabète.

## Chapter 2

# Introduction & Literature Review

In the field of bioimage-informatics, the study of living organism tissues is one of the most important topic that have spawned many researches on their underlying principles, such as pathogenesis, diabetes, and cancer progression. To facilitate these study, the development of imaging technologies has enabled researchers to capture biological samples better with large spatial resolutions and in high dimensions.

In the literature review part of this thesis, we will first provide a basis biological knowledge of tissue organization. Then, we focus on a special type of tissue, the focus of our research, called ‘Islet of Langerhans’. This tissue contains the hormone cells and its main function is to maintain and to regulate proper blood glucose levels. We will also present briefly the outline of some latest works with regard to the study of tissue spatial organization using both biological investigations and bio-informatics computation techniques.

Because we utilize the imaging technologies to capture the tissue information required for our works, we will provide a sketch of some image processing methods to extract information from the image structure such as image enhancement, segmentation, mathematical morphology, cell segmentation, cell phenotype detection. Because cells in a tissue do not work in isolation but collaborate with each other to achieve their goal, we present several mathematical models to formalize their relations and to quantify the information and the structure of cells within tissue.

## 2.1 Biological Background

### 2.1.1 Tissue Organization

#### Introduction

The cell is a complex structure and the basic unit of any living organisms. A proven fact is that cells in multicellular organisms interact and work with each other to perform their functions [11]. The group of similar cells and their extra-cellular matrix is called a tissue. In general, a tissue can be classified into four categories based on their functions: epithelial tissue, connective tissue, muscle tissue, and nervous tissue [12].

In this thesis, we focus on epithelial tissues that cover body surfaces, organ systems, as well as perform various functions for the organism such as protection, absorption, filtration, excretion, and secretion. As such, this group of tissue plays a very important role and has been extensively covered in numerous studies.

#### Epithelial tissue

Unlike other types of tissues with large amount of extracellular material, epithelial tissue is densely packed with cells [12]. In general, there are two ways to classify epithelial, either using the shape of the cells or using the degree of layering: simple or stratified.

Simple epithelial tissues have only a single layer of epithelial cells. They can be found inside the body and inside the blood vessels and the heart chambers. On the other hand, stratified epithelial tissues have several layers of cells which give additional level of protection. Due to the complex structure of stratified epithelial, cells on different layers can have different shapes.

Traditionally, researchers capture epithelial tissues in two dimensional (2D) instead of 3D images. It is easy to figure out that there are much less information about the tissue in such case. It's especially true for stratified epithelial due to its multiple layers. We won't be able to study its characteristics without capturing it in a 3D image. In our work, we performed both the analysis with sections of tissue in one and multiple layers.

#### Glandular epithelia

One of the main function of epithelial tissue is secretion, which is done by glands containing one or multiple cells. In general, we can separate glands into two categories: endocrine and exocrine glands. Endocrine glands are ductless and their products are



released into the bloodstream or lymph. Exocrine glands charge their products onto the surface of the cell body or into the body cavities. In this work, we focus on the endocrine glands and their characteristics.

### **Dynamic architecture of cell organization**

A normal cell consists of several components: plasma membrane, nucleus, cytoplasm, vesicles and others. Membranes are made of lipids and proteins and act as barriers for cells and intracellular components. These barriers, separating the cell and its outside environment, allow only special molecules such as vesicles to cross and relay messages. Inside the cell, there is a nucleus at the centre surrounding by cytoplasm. There are many other molecules in the cytoplasm area. To study a tissue, we need to take into account an ensemble of cells and all their components including membranes, cytoplasm, and the very important nucleus.

### **Architecture of tissue organization**

A fundamental question in our work is how a group of cells works together to make the tissue function as it supposed to. There are two major aspects [11] to look at:

1. The cell type. A cell type (or cell phenotype) is the composite of cell characteristics such as its morphology, biochemical, physiological properties, and behaviours. The cell type decides the cell function, thus, has direct influence upon the function of the tissue.
2. The spatial organization of the cells that form the tissue. The spatial organization of cells includes both the cell's position, its orientation inside the tissue (mantle, head, core of tissue) and all external signals from its neighbouring cells.

All of these aspects show that cells within the tissue form a special mechanism to carry out their functions and our work focuses exclusively on formalizing the relationship between cells to understand such mechanisms.

### **The need to study tissue organization**

Due to the interactions among cells in the tissue, the study of tissue organization is the key to understand its evolutionary mechanisms and the foundation for the comparison between diabetic and normal tissues [13]. In addition, by studying tissues from different species, we can also understand cross-specie architectures, both similarities and differences. In our work, we look at the structures of rodent, i.e. mouse, non human primate

(NHP), i.e. monkey, and human (HP). Using the information about cross-specie structures, we can develop and test new therapeutic approaches. For example, in biology, we often study the structures of rodent tissues and hypothesize that human tissues have the same structures and behaviors. However, by proving that they are different, we can justify the need to obtain and do experiments directly upon human sample tissues.

### 2.1.2 Islets of Langerhans

#### Introduction

In 1869, German physician Paul Langerhans detected small groups of cells - patches of endocrine tissues in the pancreas of the rabbit. Later in 1893, Edouard Laguesse named these groups of cells 'Islets of Langerhans' and showed their role in the control of secretion and regulation of blood glucose levels. As such, the islet of Langerhans plays an important role in the pathogenesis of diabetes.

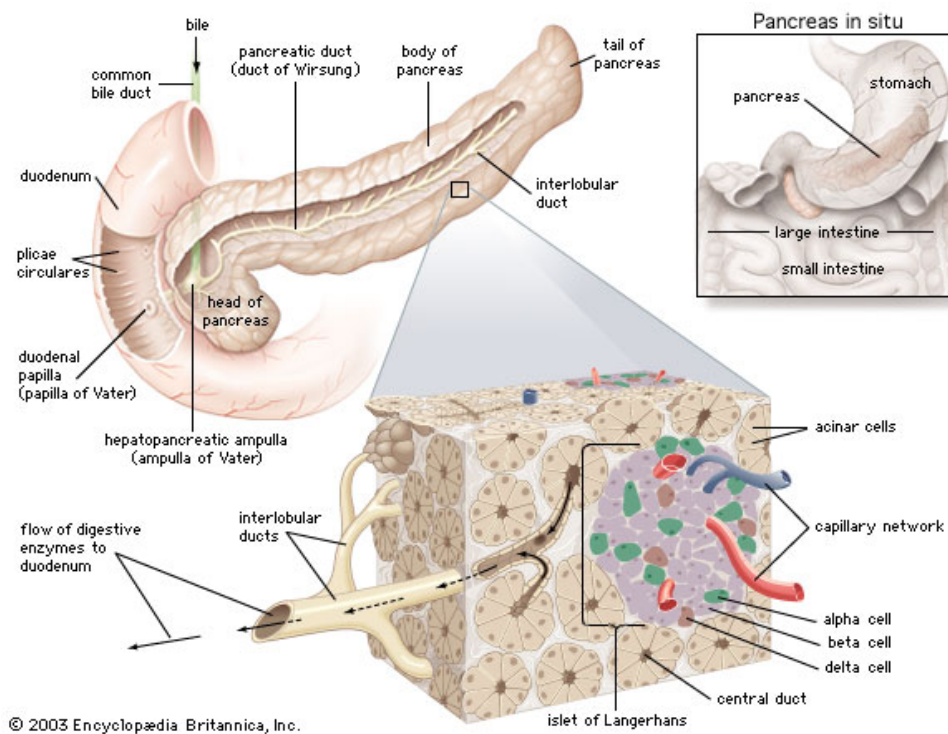


FIGURE 2.1: Pancreas in situ and zoom in of one islet of Langerhans. The figure is taken from the journal of microscopy [1].

#### Multi-cellular structure

The islet is a multi-cellular structure that consists of three major types of secretory endocrine cells:

1. The majority of cells within the islet of Langerhans are beta cells which produce insulin, the major hormone in the regulation of carbohydrate, fat, and protein metabolism. Beta cells make up approximately 80% of the rodent islet and 65% of the primate islet [14].
2. The second, in term of number, is alpha cells which produce the hormone glucagon. Alpha cells make up approximately 15% of the rodent islet and 30% of the primate islet [14].
3. The last group is delta cells which produce the hormone somatostatin. Delta cells make up approximately 5% of both rodent and primate islet [14].

All these three cell types cooperate with each other to perform the islet's main function which is essential to maintain a proper blood glucose level at all the times. They affect the capability of each other, for example, the release of insulin from beta cells can be triggered by glucagon or somatostatin. There are existing works in the field trying to understand their relationships.

### **Cells composition and cellular interactions**

There are many related works exploring the islet spatial organization to study its cells composition and cellular interactions. Some of the most important works will be shown in more details in Chapter 4 [6, 15–17].

By definition, the cells composition is the proportion of each cell types within a islet while cellular interactions or cell-to-cell contact denote their relationships. A simple way to represent such information is to show the proportion of all possible combinations, for example  $P_{\alpha\alpha}$ ,  $P_{\alpha\beta}$ ,  $P_{\alpha\delta}$ ,  $P_{\beta\beta}$ ,  $P_{\beta\delta}$ , and  $P_{\delta\delta}$ . As such, we distinguish these cellular interactions as homotypic interactions among cells of the same type and heterotypic interactions among different cell types.

### **Different regions of the islet**

Bosco et al. (2010) [6] divided each islet by different subregions including whole, mantle, core, and vessels to quantify the cell distribution and cellular interactions.

1. Whole region: the whole islet region.
2. Mantle subregion: all cells at the rear regions or near the outside environment.

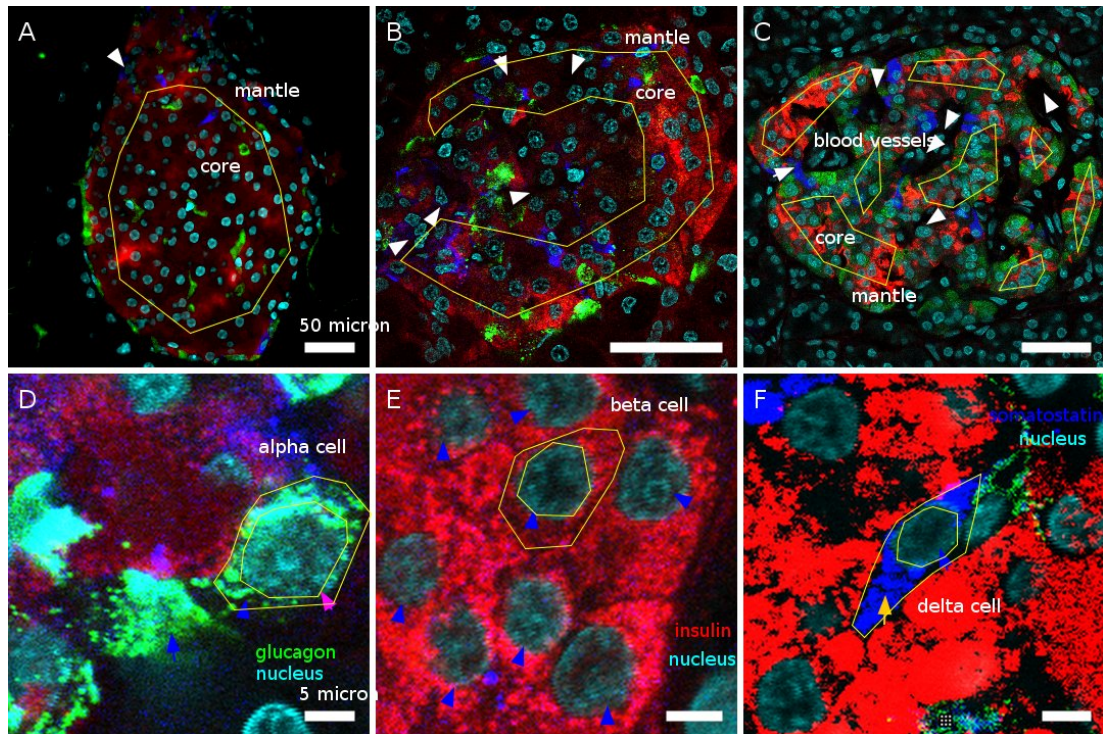


FIGURE 2.2: Example of different islets: mouse islet sample (A), monkey islet sample (B), human islet sample (C) with different regions of core region, mantle region, near blood vessels region (see the arrows) (A-, B-, C- scale bar, 50  $\mu\text{m}$ ). Zoom in different markers and cell type (scale bar, 5  $\mu\text{m}$ ): (D) alpha cell and glucagon cytoplasmic marker, (E) beta cells and insulin cytoplasmic marker, (F) delta cell and somatostatin marker.

3. Core subregion: all cells which are near to centre region of the islet.
4. Vessels: all cells which are located near to blood vessels regions.

### The architecture of the islet in different species

In rodent islets, the majority of cells are beta cells which are located at the core of the round islet. They are surrounded by a mantle of alpha and delta cells. On the other hand, alpha and delta cells in non-human and human islets are more evenly distributed throughout the islets region, thus, increasing their contact with beta cells. In our work, we develop a set of methods to analyze this spatial organization and create a toolbox for it. The toolbox is then used to investigate the spatial organization islet of Langerhans to understand its role in diabetes. It could serve as an useful tool to help biologist test new therapeutic approaches. We will come back with more details in the next chapter.

### Diabetes and the islet of Langerhans

Related works on the islet of Langerhans have shown its direct involvement with pathogenesis of diabetes [14]. So, a question raised is whether there are any differences in the

spatial organization between normal and diabetic tissues.

In general, there are two types of diabetes:

1. In type 1 diabetes, there are less beta cells than as usual because the body immune system mistakenly destroys them [18]. The function of beta cells is to produce hormone insulin. Therefore, in this case, the pancreas produces less and less insulin and there is not enough of it around to maintain all the normal activities. Insulin will need to be taken from outside.
2. In type 2 diabetes - insulin-resistance, the body does not utilize insulin properly. At the beginning, beta cells react and make more insulin to make up for it. However, over a long period of time, the pancreas still cannot make enough insulin to go around.

In both cases, there are changes in the proportion of each cell type [14, 16] and this leads directly to corresponding changes in the tissue spatial organization.

### 2.1.3 Methods to Study the Islet of Langerhans

There has been an increasing number of studies on the islet of Langerhans covering many different aspects such as providing a qualitative description of cells within the islet. In this part, we present briefly some latest works in the field focusing mainly on their biological investigations and bio-informatics computations. Their comparison is shown in Table 2.1.

For the analysis of islets of Langerhans, works done by Cabrera et al. (2016) [16], Bosco et al. (2010) [6], and Drigo et al. (2015) [14] provide a comprehensive survey.

On the specific topic of using computational biology approaches to study the tissue organization, there are works done by Kilimnik et al. [19–21], Striegel et al. (2015) [22], and Hoang et al. (2014) [8].

*Confocal microscopy.* There are two main imaging methods to capture the tissue: histological imaging technique [17] and confocal microscopic imaging technique [6, 16, 19]. The majority of works on islets of Langerhans are based on confocal microscopy. The confocal microscopy and fluorescent technologies also increase the ability to observe different markers corresponding to different cell types (Section 2.2.1).

*Area of study.* A popular approach in other works is to cut off the entire islet by different sections: head, body, and tail [6, 14, 16] because it is much easier to prepare and process the sample that way. However, its major limitation is the missing cell-to-cell contacts between these different sections.

*Cross-species study.* Researches in this topic normally divide the studying species into three main levels: rodent such as mouse, non human primate (NHP) such as monkey, and human primate (HP) such as human. Early study normally used rodent islet to make assumptions about the equivalent human islet functions. However, more recent studies have shown that such assumptions are not always true. As such, there is a need to expand the study to other species as well. As it is more difficult to extract human samples, we could choose to study non human primates which is closer to human and link the result to human islets (see the survey of different approaches at [14]).

*Study of different cell types (alpha, beta, delta) and the blood vessel.* Early works only focused on two main cell types (alpha, beta) and skipped the role of delta cells. This has been changed and more recent works now show the important role of delta cells to the islet function as well. In addition, blood vasculature has also been suggested to play an important role in maintaining islet function.

*Cell composition and cell morphological computation.* Existing works show differences of cell composition across different species such as rodent and primate [6, 14, 16, 19]. For non-human primate (monkey) and human islets, it has been more alpha and delta contacts than in rodent [14].

*Homotypic and heterotypic contacts.* Rodent islets are shown to have more homotypic contacts whereas non human primate(NHP) and human(HP) islets have more heterotypic contacts.

*Diabetic study.* There are not many works on this topic mostly due to the amount of available data, which is difficult to extract. Some works show the differences in cell compositions between normal and control diabetes [6].

*3D study.* Bosco et al. (2010) [6] reconstruct the islet in 3D using information from serial human pancreatic sections. The islet is reported to consist of a beta-cell core surrounding by mantles of alpha cells and vessels. The same paper also suggests that, similar to rodents, human islet cells are purposely arranged in a complex trilaminar

epithelial plate fold that favours heterologous contacts and paracrine signalling among cells. Such results can only be achieved by analysing sample data in 3D.

*The analysis of tissue spatial organization.* The works in Table 2.1 analyze the spatial organization of tissue using various methods but they all have certain limitations. None of them looks at cell distribution or compare the observed patterns in different species. In addition, they don't have any tools that can help other researchers to perform similar tasks.

*Modeling of tissue spatial organization.* There are currently very few works that touch the topic of modeling the tissue spatial organization. Striegel et al. (2015) [22] focuses on the modeling of beta cells in their clusters within the human islet. Hoang et al. (2014) [8] generates a 3D spatial distribution of alpha and beta cells in mouse, pig, and human islets. However, these models represent each cell using cell-like shapes (sphere, ellipsoid), thus, can only compute cellular interactions using basic geometry. This leads to an incorrect estimation of the number of cellular interactions. Therefore, there is a need for better modeling processes that take into account actual shapes of different cell as well as their volume and physical interactions.

Related work	Study area	Study species	Confocal, multi photon	Study alpha cell	Study beta cell	Study delta cell	Study blood vessel	Cell composition	Homo-, Hetero- typic	Diabetic	3D	Tissue organization analysis	Tissue organization modeling
Cabrera et al (2006)[16]	head, neck, body, tail	human, mouse, monkey, pig	✓	✓	✓	✓	✓	✓	✓	X	X	✓	X
Bosco et al (2010)[6]	head, body, tail	pig, human	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	X
Ionescu et al (2015)[17]	head, neck, body, tail	human	X	✓	✓	✓	✓	✓	✓	✓	X	✓	X
Kilimnik et al (2012)[19, 21, 23]	head, body, tail	mouse, human	✓	✓	✓	✓	X	✓	X	✓	X	X	X
Striegel et al (2015)[22]	head, body, tail	human	✓	✓	✓	✓	X	✓	X	✓	✓	X	✓
Hoang et al (2014)[8]	head, body, tail	mouse, pig, human	X	✓	✓	✓	X	✓	✓	X	✓	✓	✓

TABLE 2.1: Comparison of different related works on islet of Langerhans. The two first works are on the biological investigation. The four final works are on computational biology. ✓: shows that the work took into account this analysis, X : it did not take into account this analysis.



## 2.2 Image Analysis

### 2.2.1 Confocal Microscopy

#### Fluorescence Imaging

We begin this part by providing a quick description of the fluorescence imaging technique, in which specimens are labelled with fluorophores. The distribution of fluorescence are then observed under exciting illumination and captured by photosensitive detectors that measure the intensity of the emitted light and create a digital image of the sample.

In addition, to capture the thick sample and dynamic live cell imaging, confocal fluorescence microscopy is the commonly used technology. Many focal planes have been designed for labeling biological molecules. Depending on the type of the fluorescent marker used on tissues, different biological components (proteins) are highlighted. One of the common marker is Di Aminido Phenyl Indol (DAPI) used to visualize the DNA and cell nuclei. Confocal fluorescence microscopy technique plays a central role in our research and are covered in the next section.

#### Confocal microscopy and image acquisition

The last decade has seen a huge development in the number of different methods used to capture 3D images as well as to process and analyze them. One of the most common methods is confocal microscopy. It is a powerful technique that can provide high resolution images with depth discrimination ability. It significantly improves the ways in which cells and tissues can be studied with two different aspects:

1. Captured images are in high resolution.
2. Captured images have a higher level of contrast because out-of-focus information has been removed.

The output of confocal microscope consists of a stack of optical sections called z-series. The optical sections are obtained at fixed intervals along the z-axis. Each 2D image in the stack is called an optical slice. All slices together comprise a volume dataset. Building up the z-series in depth allows the 3D image representation of the sample to be reconstructed.

To facilitate its processing, a 3D image obtained from the microscope will be represented as a 3D array whose elements are called voxels. An image  $f(x, y, z)$  is represented as a 3D matrix of dimensions  $L \times M \times N$ , where  $x, y, z$  denote column, row, and slice coordinates respectively. Each voxel has a physical size  $D_x, D_y, D_z$  in units, commonly measured in micrometers. In our case, the image usually has a different calibration of voxel size between (xy) and z planes because the later dimension is much thicker.

### **Limitations of confocal microscopy**

Although confocal microscopy technique has been widely adopted in the field, it still has several limitations that researchers need to be aware of. Three main drawbacks are (i) slow scanning speed, (ii) lower axial resolution along the z-axis comparing to the lateral resolution along the x and y-axis, and (iii) the potential damage to the sample in the long scanning process.

### **Imaging systems for islet of Langerhans**

For studying the tissue organization in general and the islet of Langerhans in particular, there are two common imaging techniques, histological imaging [17] and fluorescence imaging [14, 16]. Between them, the light microscope is widely adopted in biological applications.

In histological imaging technique, tissues are fixed, and embedded in wax, which makes them hard but much easier to trim into sections. These sections are then stained and examined with the light microscope. Histological stains like H&E (Haemotoxylin and Eosin) contain the two dyes Haemotoxylin and Eosin. Haemotoxylin binds to DNA (Heterochromatin and the nucleolus) in the nucleus, and RNA in ribosomes and stains them purple. Eosin binds to the proteins in the cytoplasm and stains them pink.

The key differences between histological method and fluorescence method are the staining process and the thickness of sample, in which fluorescence imaging can capture a thick tissue sample while histological technique could not. Table 2.2 shows a detail comparison between histological and fluorescence imaging systems with criteria such as the staining process, the reliability of color, the cutting method, the size of sample, the laser technique used, the number of possible markers, the complexity, and the capacity to capture 3D thick samples.

System & Characteristics	staining	color	cutting	sample size	laser	number markers	complexity	3D
Histological imaging technique	yes	not stable, different each time	cut by many small section, damage	big, overview	no	small number	simple	no
Fluorescent imaging technique	yes	quite robust	thick section, not damage	smaller	yes	many markers	difficult	yes

TABLE 2.2: Comparison between histological and fluorescence imaging systems

In our case, cells and tissues of the islet of Langerhans are three dimensional entities and most of their cellular activities occur in 3D space. Thus, confocal microscopy and fluorescence imaging allow us to observe the entire tissue sample in 3D and is the most suitable method to acquire our samples.

## 2.2.2 Methods for Studying Tissue Organization

### Existing frameworks

In recent years, there has been an increase in the number of tools to analyze cells in tissues including both commercial platforms like Imaris(Bitplane), Image-Pro(MediaCybernetics), Amira and Metamorph (Molecular Devices), and open source equivalents such as CellProfiler [24], BioImage-XD [25]. These tools focus mostly on morphological analysis of cells in biological imaging while commercial ones also give a friendly graphical interface that supports good visualization. However, they all have some major drawbacks where only basic processing and analysis features are provided. Specifically, they all lack of ability to analyze 3D structures or to model spatial tissue organization in 3D, not to mention the high costs of purchase and the proprietary license of commercial softwares which prevents additional features to be added by the community.

**Existing approaches for analysing spatial organizations** Although there are several existing algorithms for analysing cells and nuclei in microscopic images, those that can deal with 3D spatial organizations of tissues are limited and their measurements are mostly done on each cell individually. In our case, analysing the tissue requires the

Criteria/Methods	CellProfiler[24]	Drasdo[26]	Morales[27]	Andreas[28]
Confocal Microscopy	✓	✓	✓	X
High throughput	✓	✓	✓	X
Nuclei Segmentation	✓	✓	✓	✓
Cytoplasm/ Membrane Detection	✓	✓	✓	✓
Cell Zone Detection (Region Growing)	✓	✓	✓	✓
Cell Phenotype Detection	✓	X	✓	X
Cells Morphology	✓	✓	X	✓
Colocalization Analysis	✓	✓	X	X
Statistical Analysis	✓	X	X	X
Tissue Organization	X	✓	X	X
Nuclei/Cell Classification	✓	X	✓	✓
Tissue Modeling, Simulation	X	✓	X	X

TABLE 2.3: Comparison of different methods for analyzing spatial tissue organization

identification of different cellular components, including cells, as well as the physical interactions among them.

Therefore, the first step towards this goal is to identify the location and identity of all cells that make up a given tissue. Since a clear cytoplasmic or membrane labelling is usually difficult to obtain in thick tissue samples, most studies rely on nuclear labeling or DAPI to identify these cells. However, nuclei segmentation, especially in a large 3D image, is not trivial and remains an active research area among bio-imaging researchers. Furthermore, analysing the whole tissue poses an additional challenge in which cells are normally crowded together, thus makes it more difficult to separate them. Techniques for segmenting nuclei or cells in such cases are commonly based on the ‘region-growing approach’ [24, 29, 30].

Once the primary segmentation step is done, the identity of the segmented cells will need to be determined. Depending on the type of markers available, it can either be achieved by manual annotation of images, or by setting a simple threshold for nuclear and cytoplasmic content, or by using a more complex supervised machine learning approach. Several existing methods are CellProfiler [24], Dirk Drasdo et al. [26], Morales et al. [27],

Andreas et al. [28]. Table 2.3 gives an overview of these existing frameworks. However, they all share some limitations in their abilities to observe a large population of cells or to analyze cellular interaction and spatial organization of tissue.

### Existing approaches for nuclei and cell segmentation in 3D

Approaches for analyzing the tissue organization of microscopic image can be generally divided into two groups:

1. Segmentation and quantification:
  - Nuclei segmentation.
  - Cytoplasm/cell membrane segmentation or cell zone detection if there is no membrane image.
  - Cell phenotype detection if input data includes marker channels.
2. Spatial Organization Analysis:
  - At local scale including cellular interaction.
  - Global spatial organization.
  - Spatial organization modeling and simulation.

In the next section, we will review state-of-the-art image processing methods which can resolve these problems as well as be able to work with thick tissue samples captured by confocal microscopy.

### 2.2.3 Image Enhancement

Image enhancement enhance the values of pixel in image is the first step to prepare the input data. In the case of confocal microscopy, although it has several advantages over other microscopic techniques (section 2.2.1), it still has some limitations when it is used to capture three dimensional images, especially the low resolution along the z-axis. Image enhancement can help in this case separating noise from the foreground.

Enhancement algorithms can generally be grouped into two categories: spatial domain methods and transform domain methods. Spatial domain methods work on direct manipulation of pixels and include the set of operations processed on the whole image or a local region such as histogram equalization, sharpening, or nonlinear filtering. On the other hand, transform domain processing techniques process image information based

on the Fourier transform domains of the image. We will present several methods on both categories.

### Spatial domain methods

The median filter is one of the most powerful method in this category. It scans each voxel in the image and takes into account the neighbor voxels and replace the voxel value by the median of its neighbours.

*Median filter 3D:* Define an observed voxel  $v_{curr}$  and a filter window of size  $M = X \times Y \times Z$  surrounding this voxel. Considering all the neighboring voxels  $n_1, n_2, n_3, \dots, n_m$  sorted by their values  $n_1 < n_2 < n_3 < \dots < n_m$ , we have the median value  $n_{med} = n_{(m+1)/2}$  if  $m$  is odd or  $n_{med} = (n_{m/2} + n_{m/2+1})/2$  if  $m$  is even.

In practice, the window size that we often use for 3D median filter is  $3 \times 3 \times 1$  or  $4 \times 4 \times 2$  because the output of confocal microscopy usually has different calibrations of voxel size between (xy) and z plane in which the z dimension can be much thicker, thus the inversely proportional size of z comparing with x and y.

The main advantage of median filter is that an unrepresentative pixel in a neighborhood will not affect the median value significantly, so the median filter is quite robust in practice. This filter can also preserve the edge better than other filtering method because it uses the values of the neighboring voxels to represent the value of the current one.

### Transform domain methods

In many cases, noise can be remove more easily in the frequency domain than in the spatial domain. In the frequency domain, low frequency part correspond to smooth area and large structures in the image while high frequency is noise and medium frequency parts correspond to image features. Some examples of this type of methods are Viener Filtering, Deconvolution, low-pass filtering, high-pass filtering, and band-pass filtering. Some of the most common transformations are Fourier transform, Gaussian transform, and wavelet transform.

Let  $f(x, y, z)$  be the raw image and  $F(x, y, z)$  be the input image in the transform domain (transformed image). This image is then multiplied with the filter function voxel-by-voxel:  $G(x, y, z) = H(x, y, z) \times F(x, y, z)$  (\*), in which  $H(x, y, z)$  is the filter function, and  $G(x, y, z)$  is the filtered image. In order to turn the result image back into the spatial domain,  $G(x, y, z)$  has to re-transformed using the inverse function.

*Low-pass filter:* This filter allows to remove a big part of noise in an image. The main idea of the low-pass filter is the frequency cutoff value. Let  $D(x, y, z)$  be the distance of the point from the origin in the transform domain (ex: in the Fourier domain).

$$H(x, y, z) = \begin{cases} 1 & \text{if } D(x, y, z) \leq D_0 \\ 0 & \text{otherwise} \end{cases} \quad (2.1)$$

*High-pass filter:* This filter allow to highlight edge information and sharpen the image. Main idea of high-pass filter is frequency cutoff  $D_1$ :

$$H(x, y, z) = \begin{cases} 1 & \text{if } D(x, y, z) \geq D_1 \\ 0 & \text{otherwise} \end{cases} \quad (2.2)$$

*Band-pass filters:* Band-pass filters are a combination of both low-pass and high-pass filters. They attenuate all frequencies smaller than a frequency  $D_0$  and higher than a frequency  $D_1$ , while the frequencies between the two cut-offs remain in the output image. We obtain the filter function of a band-pass filter by multiplying the filter functions of a low-pass and of a high-pass in the frequency domain.

$$H(x, y, z) = \begin{cases} 1 & \text{if } D(x, y, z) > D_0 \ \& \ D(x, y, z) < D_1 \\ 0 & \text{if } D(x, y, z) < D_0 \cup D(x, y, z) > D_1 \end{cases} \quad (2.3)$$

## 2.2.4 Nuclei Segmentation

Thanks to confocal fluorescence microscopy techniques, especially DAPI staining, researchers can capture thick tissue samples in much higher resolutions with less out-of-focus defects. However, most of the existing methods in the field only work with 2D images. The few methods that can handle 3D images suffer from high complexity, over-segmentation, and the inability to work with thick samples. Therefore, there is a need for more efficient and accurate methods for 3D tissue images. Such methods influent all the subsequence quantification and analysis steps.

### Nuclei segmentation methods

In general, nuclei segmentation methods can be put into two categories: methods using machine learning techniques for training and methods utilizing cell shapes.

In the first group, the expert marks the correct nuclei region in the captured image for training. A program can then learn from this and predict the cell nuclei region in other images [31]. However, it is notorious difficult for human to prepare the ground truth data in 3D because the cell nuclei is elongated through many slices. Moreover, due to the variability of the cell's volume and its shape in high throughput images, a large amount of training data is required to achieve good accuracy.

The methods in the second group take into account the cell's shape and can be applied directly to the input image without any training, for example, graph cut [32, 33], seed detection [33, 34].

Implementation-wise, there are some existing platforms for 3D nuclei segmentation of fluorescent microscopic images including Nuclear Segmentation of FARSIGHT toolkit [32], Modular Interactive Nuclear Segmentation (MINS)[34], 3D Multiple Level Sets (3D MLS) [35], 3D Iterative Thresholding [36, 37]. Table 2.4 compares these platforms in terms of usability and performance including (i) capability to process a 3D confocal microscopic image, (ii) implemented language, (iii) level of friendly user interface, (iiii) OS system, (v) the difficulty of selected input parameters, (vi) performance on detection of flexible nuclei shape.

1. *FARSIGHT toolkit (Al-Kofahi et al. (2010)[32])*. The automatic nuclear segmentation method implemented in FARSIGHT toolkit is one of the fastest and most efficient method for 3D nuclei segmentation. It first creates an initial binarization of the image using Graph-Cuts algorithm. Seeds are then detected on the transformed image using a multi-scales Laplacian of Gaussian LoG and nuclei are represented by blobs. The seed here is a sub-region inside each cell nuclei region and is used to segment the nuclei. Also, a graph-cuts approach with alpha-expansions can be used to obtain the best contour of each nuclei region.
2. *Modular Interactive Nuclear Segmentation (MINS)[34]*. This is an accurate and user-friendly nuclear segmentation framework written in Matlab. The main technique of MINS is to detect seeds using Hessian of Gaussian detectors.
3. *3D Multiple Level Sets (3D MLS)[35]*. When the nuclei intensity level varies, a single level set functions is not enough to detect nuclei, this framework uses multiple level sets surface modeling to detect the nuclei.



Criteria/Methods	Farsight	MINS	3D MLS	3D Iterative Thresholding
Confocal Microscopy, 3D	Yes	Yes	Yes	Yes
Language	C, C++	matlab	C++	java
Friendly User Interface	medium	high	high	medium
OS System	linux, mac windows	windows	windows	linux, mac windows
Parameters Difficulty	easy	easy	difficult	difficult
Flexible Shape Detector	high	high	high	high

TABLE 2.4: Review of 3D available nuclei segmentation methods

4. *3D Iterative Thresholding* [36, 37]. This framework was first developed by Mohammed et al. (2014) [36, 37] to detect nuclei in early embryos. Dr Boudier implements it in Java under ImageJ software platform. This method tests all the thresholds and detects objects at each threshold. The objects at different thresholds are then link together. This method can handle the intensity variances but its drawback is the amount of time required to test all the thresholds.

### Evaluation of segmentation accuracy(SEG)

The accuracy of nuclei segmentation affects all the subsequent stages. Therefore, it should be as stable, robust, and accurate as possible. In order to evaluate the accuracy of different segmentation methods, we compare the matching of segmented region with the actual object region [38]. Thus, the evaluation steps are:

1. Create ground truth(GT) using manual segmentations by experts in biology domain. Each manually segmented object in GT is called a reference object.
2. Compute the Jaccard similarity index (J) of the set of voxels of matching objects (Equation 2.4).
3. Check the matching condition and decide whether a segmented object match a manual segmented object. Suppose that  $R$  denotes the set of voxels belonging to a reference object and  $S$  is the set of voxels belonging to its segmented object. The Jaccard index measures the significant overlap of reference and segmented objects. A segmented object and a ground truth object are considered matching

if and only if equation 2.4 is satisfied, with  $\lambda$  - a predefined parameter which denote the level of matching. Normally  $\lambda$  equal to 0.5 this means 50% overlap.

$$\begin{aligned} \text{Jaccard - index : } J(S, R) &= \frac{|R \cap S|}{|R \cup S|} \\ \text{Condition : } |R \cap S| &> \lambda \cdot |R| \end{aligned} \tag{2.4}$$

### Ground truth for SEG measurement

We rely on domain experts to prepare the ground truth for SEG measurement. The annotated task is to mark voxels belonging to objects accurately. The experts should ensure that the contours of the ground truth objects are as close as possible to real visible object contours. In three dimensional space, this task is even harder because each object is located in several slices, which makes it difficult to mark the area and combine the marked areas of many slices together. Certain tools have been built to facilitate this function such as Segmentation Editor [39], a plugin under ImageJ platform. This tool helps you annotate the ground truth objects by going through the slices and selecting the region you want to label. It offers some useful features which can help users in their tasks. When the difference of consecutive slices is slight, user can use the interpolation function of the tool to segment them automatically. Moreover, 3D ROI Manager is a useful tool to observe regions of interest, which helps to project the 3D contours of segmented objects into the real objects [37].

### Error Detection

After preparing the ground truth, in order to count the detection errors, we use mis-detection and wrong detection as simple metric measurements to evaluate the performance. The detection errors are classified in two groups, False Positive (FP) and False Negative(FN). FP happens in case when a sample is negative but the algorithm declares as a positive object. On the other hand, FN happens in case when a sample is positive but the algorithm erroneously declares it as a negative one.

In the context of biological image segmentation, a false positive(FP) happens when noise is wrongly detected as an object or an object is split into several ones due to the presence of strong edges within it. On the other hand, when the intensity of an object is lower than the predefined threshold, the program could not detect it, which causes a false negative.

Using these concepts, precision and recall are two common metrics to estimate the segmentation accuracy. Precision measures the fraction of objects declared positive and the total of truth positive  $Precision = TP/(TP + FP)$ . Recall, on the other hand, measures the fraction of objects that have been retrieved correctly over the total amount of existing objects in the image  $Recall = TP/(TP + FN)$ . Finally, we can combine them to get the F-measure defined as  $F_{measure} = 2 \times (Precision * Recall) / (Precision + Recall)$ .

### 2.2.5 Mathematical Morphology

By its definition, mathematical morphology is a set of nonlinear filters used together to filter binary and gray-scale images, which we then extend to apply to our 3D context.

Morphological operators consist of two basic operations ‘dilation’ and ‘erosion’. Dilation expands the features of the shape whereas erosion suppresses them. The basic effect of dilation on a binary image is to gradually enlarge the boundaries of regions of foreground pixels, typically white pixels. Thus areas of foreground pixels grow in size while holes within those regions become smaller. On the other hand, the basic effect of erosion is to erode away the boundaries of regions of foreground pixels. Thus areas of foreground pixels shrink in size while holes within those areas become larger.

Morphology operators are based on set theory. Assuming that a binary 3D image  $X$  and  $B$  is its set of coordinates for the structuring elements. A structuring element, also known as kernel, determines the effect of the operator upon the image. Several popular structuring elements in 3D images are a  $3 \times 3 \times 3$  structure with the central pixel and its direct 26 neighbors or a disk of a given radius.

Given  $B^S$  is the symmetric of  $B$ , we have the following definition:

$$\begin{aligned}
 \text{dilation} : \quad X \oplus B^S &= \{x \in Z^3 : B_x \cap X \neq \emptyset\} \\
 \text{erosion} : \quad X \ominus B^S &= \{v \in Z^3 : B_v \subset X\} \\
 \text{closing} : \quad (X \oplus B^S) \ominus B & \\
 \text{opening} : \quad (X \ominus B^S) \oplus B &
 \end{aligned} \tag{2.5}$$

Closing and opening are also two operators in the field of mathematical morphology. They could be derived from the basic operations of erosion and dilation. Closing fills small holes and connects objects that are close to each other. Closing is similar in some

ways to dilation in that it tends to enlarge the boundaries of the foreground (bright) regions in an image and shrink background color holes in such regions, but it causes less destruction to the original boundaries.

Opening smooths the surface and its effect is somewhat like erosion in that it tends to remove some of the foreground (bright) pixels from the edges of regions of foreground pixels. Again, it is less destructive than erosion in general.

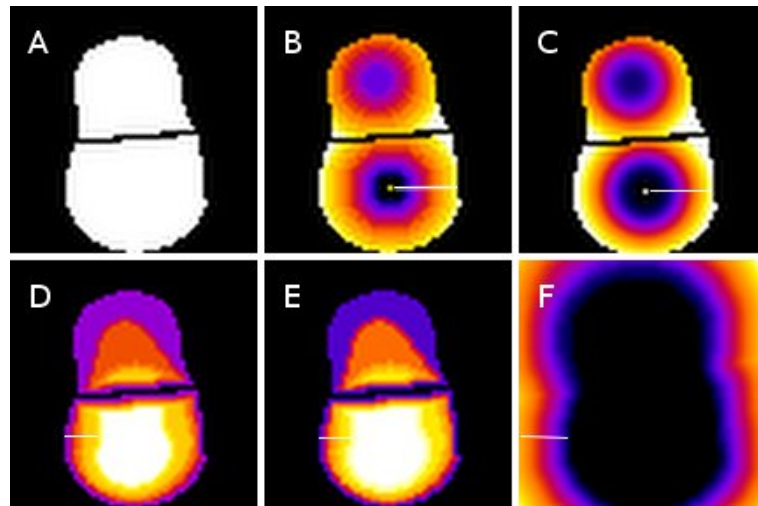


FIGURE 2.3: Different strategies to use distance map. (A) Input image consists of two objects, (B) Euclidean distance map, measure from center of each object, (C) Normalized Euclidean distance map, measure from center of each object, (D) Euclidean distance map, measure from boundaries to the center of object. (E) Normalized Euclidean distance map, measure from boundaries to center of object, (F) Euclidean distance map, measure from boundaries of each object to the background.

Distance maps are data structures commonly used in image analysis. They provide various types of distances, for example Euclidean distance, of background pixels to the nearest object pixel. Distance maps can be used to describe the internal structure of objects.

If distance metric is Euclidean, we have Euclidean distance map (EDM). If normalization is adopted, we have normalized distance map. Although this structure has been adapted to 3D, its computation complexity is much higher. So, Schouten et al. (2006) [40] proposed a new adaptation of the algorithm to reduce the complexity of distance transformation in 3D.

In Fig. 2.3, we show different strategies to use distance maps. Using the same image, but based on different set of voxels as objects (center of objects or object boundaries), we can obtain different results.

1. Measure based on the center of object. The value of each voxel is the distance from this voxel to the object center (see Fig. 2.3B). We can also normalize these distances between 0 and 1 by dividing the value of each voxel to the maximum distance value and obtain the normalized distance map (see Fig. 2.3C). The advantage of this approach is that it is not sensible to the precision of the segmentation method. However, its drawback is the underestimate of the center positions and the overestimate of points closer to the periphery. Thus, the layer closed to periphery is thicker than at the center.
2. Measure based on the distance from the boundaries to center of object. Biological events are dynamic and certain components have tendency to move closely to the periphery and then go out. A measure based on the distance from a voxel to its boundaries makes biological sense and is commonly used. Fig. 2.3D shows the Euclidean distance map and Fig. 2.3E shows the normalized Euclidean distance map of this approach. Per observations, the layer closed to the periphery is thinner than at the center. They indicate that this measure varies quickly at the periphery and underestimate the number of voxels there.

### **The watershed transformation & its application to image segmentation**

Automatic segmentation of clustered objects requires the implementation of specific image segmentation tools. Among them, Watershed transform is a powerful region-based segmentation approach. The idea underlying this method comes from the domain of geography (Beucher et al. (1992)[29]).

1. Landscape with holes pierced in local minima.
2. Basins will fill up with the water starting at these local minimal.
3. Where water coming from different basis would meet, dams are built.
4. Where the water level has reached the highest peak, the process is stopped.
5. Basins separated by dams, called watershed lines or watersheds.

#### *Watershed transform algorithm*

There are many versions of watershed transform algorithm such as watershed by flooding, by topographic distance, by the drop of water principle, or inter-pixel watershed. The detail of different implementations is described in the work of Roerdink et al.(2000) [30].

The basic steps of watershed transform algorithm are the following:

1. Compute  $g = \text{gradient}(I)$  or  $m = \text{distance transform}(I)$ .
2. Extract local minima( $g$ ) or local maxima( $m$ ).
3. Dam is constructed to prevent the merging of regions.
4. Watershed transform algorithm like flooding algorithm is performed in each catchment basin area.
5. The output of algorithm returns label image  $L$ , identify the watershed regions of  $I$ .

Here we present three main applications of watershed algorithm in image segmentation.

*Application 1: splitting of merged objects*

The Watershed algorithm is based on the computation of the distance map inside the mask of the merged object[41]. The distance map is calculated from the boundary to the center of the mask. We then detect the seed point as the local maxima of the distance map, the farthest point from the boundary, corresponding to the centre of the object. Finally, we apply the watershed algorithm to obtain the region of each object.

*Application 2: computing the Voronoi zones*

A Voronoi region  $V_{obj}$  associated with the object  $obj$  of a study area is the set of points that are closer to this object than to any other objects of the study area. The Voronoi regions partition the study area, except for boundaries between them which are of measure 0. The Watershed algorithm is based on the computation of the distance map outside of the object but inside the study area. We detect the seed point as the local maxima of the distance map. Finally, we apply the Watershed algorithm to obtain the Voronoi region of each object. This watershed algorithm draws lines between objects at the equal distance from the boundaries of the different objects.

*Application 3: spot segmentation*

In this application, the Watershed algorithm can be applied even before the segmentation process. The objective is to define the area around each seed [29]. Then the segmentation method is applied within the region defined by the Watershed procedure. To do so, we first compute the local maxima of each object and consider them as seeds. Next, we compute the Voronoi zones surrounding each objects using a watershed algorithm. The segmentation method is then applied within the Voronoi zone. Each Voronoi zone contains one object zone and a part of background. This approach allows the partition

of a big image into many small regions. It reduces the computational time and increases the accuracy of segmentation by examining objects using a local threshold.

### 2.2.6 Cell Segmentation

#### Cell segmentation overview

A common approach to cell segmentation involves the following three steps:

1. Perform nuclei segmentation to identify the nuclei region.
2. Detect cell components located outside of the nuclei but inside of the cell borders.
3. Detect the cell region using the informations from these previous steps. However, depending on what kind of cell components obtained from the second step, different algorithms are proposed, which we will describe into details.

*The cytoplasm exists.* The cell region is determined by the union of nuclei (Nuc) and cytoplasm region (Cyt) (Equation 2.6).

$$\begin{aligned}
 & \textit{Suppose} : \textit{Structure}(\textit{Nuc}), \textit{Structure}(\textit{Cell}), \textit{Structure}(\textit{Cyt}) \\
 & \textit{Nuclei, Cyt are part of Cell, Nuclei disconnect Cyt} \\
 & \textit{Segmentation}(\textit{Nuc}), \textit{Segmentation}(\textit{Cyt}) \\
 \implies & \textit{Cell region} = \textit{Nuc region} \cup \textit{Cyt region}
 \end{aligned} \tag{2.6}$$

*One of these components exist: cell cortex, tubular structures, plasma membrane (Mem).* Cell segmentation is achieved by expanding the region from the nuclei(Nuc) as the seed to the cell surface. The growing area is limited to the cell cortex, or to the cell plasma membrane, or to the border of neighboring cells, or to other tubular structures. An example of cell detection using membrane image is in equation 2.7.

Keraudren et al. (2011) [42] proposed an automatic approach to cell segmentation on epithelial cells and cells-based screens in confocal microscopy. This method estimates the cell-to-cell contact areas and their thickness by computing an edge map on the membrane channel and then performs adaptive watershed on the edge map to obtain a segmentation of cell-to-cell contact. This method achieves good accuracy but only works well in two dimensional space. Its capability to streamline in three dimensions is

still an open question. Stegmaier et al. (2016) [43] proposed a new method to perform this task in 3D real time. It first processes a slice based image segmentation and then combines 2D cell segments to complete the 3D cell shapes.

$$\begin{aligned}
 & \textit{Suppose} : \textit{Structure}(Nuc), \textit{Structure}(Cell), \textit{Structure}(Mem) \\
 & \quad Nuc, Mem \textit{ are part of } Cell, Nuc \textit{ disconnect } Mem \\
 & \quad \textit{Segmentation}(Nuc), \textit{Segmentation}(Mem) \\
 & \implies \textit{Cell region} = Nuc \textit{ region} \cup \textit{expand}(Nuc, Mem \textit{ as limit})
 \end{aligned} \tag{2.7}$$

*No membrane.* In many cases, it is difficult to capture cell plasma membrane because of the limited number of staining in fluorescence technique. Even when membrane channel exist, there is no existing method that could work well in case of high noise and low intensity of the membrane contours. In such case, a specific procedure is adopted to estimate the cell zone. A Voronoi-based segmentation approach is proposed by Jones et al. (2005) [44] that estimates the Voronoi regions of each nuclei as seed in their surrounding area. This approach is part of the image platform CellProfiler [24] and is discussed in more details by Kriete et al. (2006) [45].

In practical, there are different ways to implement a Voronoi based segmentation approach. This approach focuses on compute the Voronoi region surrounded by a given finite set of points or seeds  $S_1, S_2, \dots, S_n$ . And a Voronoi region of the seed  $S_i$  consists of every point in the Euclidean space whose distance to  $S_i$  is less than or equal to its distance to any other seeds  $S_j, i, j \in \{1, 2, \dots, n\}$ . One approach is Voronoi tessellation which uses a center of nuclei as seed and grows into its surrounding areas. It is commonly used to approximate the morphology of cells in epithelial tissues [46]. However, the limit of this method is the over estimation of cell region when the cell volume is small and the under estimation of the cell region when the cell volume is large.

Other strategy of Watershed is to use the nuclei region as seed [24]. This method can resolve the limitation of the one above but still suffers from the over estimation of the cell zone area.

*Conclusion:* Membrane segmentation is one of the most challenging segmentation tasks. Watershed region growing is currently one of the best solution when there is a lack of membrane images.



### 2.2.7 Cell Phenotype Detection

The development of automated fluorescent confocal microscopes allows researchers to collect high throughput biological images. Moreover, it increases the possibility of highlighting and capturing the complex sub-cellular phenotypes using specific fluorescent labeled proteins, a.k.a. staining. As a result, the door is now open to the identification of the cell biological phenotypes. However, it is also a big challenge for automated image analysis and computation.

Grys et al.(2016) [2] discuss the commonly used computer vision and machine learning methods to identify phenotypic profiles. They come up with a general two-stage workflow for the generation and classification of phenotypic profiles as shown in Fig. 2.4.

1. Generating phenotypic profiles. This step performs image analysis approaches including high throughput image acquisition, segmentation, object recognition, feature extraction, and feature selection.
2. Clustering and classifying phenotypic profiles. Depending on the characteristics of image modalities and the research goals, different strategies can be applied such as clustering, machine learning classifiers, and outlier detection.

#### **The difficulty of generating phenotypic profiles**

The first and foremost difficulty for cell phenotype detection is feature extraction. CellProfiler platform [24] can extract hundreds to thousands of different features for each object and use these information to generate the phenotypic profile. Some common features are the object volume, shape, texture, and the histogram of intensity inside segmented cells. Due to such large number of extracted features, feature selection and dimensionality reduction are a must so that only relevant features are kept [47, 48].

In addition, the high level of noise is also a major obstacle. This is especially important when the tissue consists of many cell phenotypes, each is highlighted by one marker. In such case, each pixel coordinate can be a label on one image and noise on another one.

#### **Approaches to clustering and classifying phenotypic profiles**

CellProfiler (2006) [24, 47] uses classification to detect unknown or rare phenotypes. However, it requires the preparation of training data, which is a very time consuming

and difficult process in 3D. A recent work by Oliver et al. (2016) [49] applies deep learning to the cell phenotype problem.

Grys et al. (2016) [2] applies unsupervised clustering and outlier analysis approaches to cell phenotype detection. While clustering can identify common phenotypes, outlier detection can successfully identify unknown ones. When there is no expert knowledge about phenotypes to prepare the training data or there is a lack of cell boundary information, they are suitable candidates to replace supervised machine learning approaches.

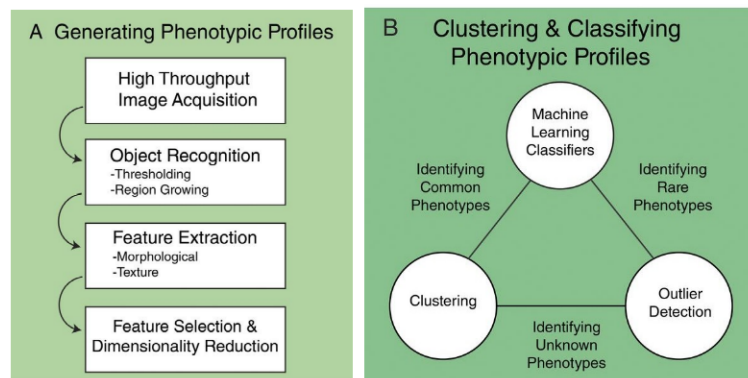


FIGURE 2.4: Workflow of cell phenotypic profiling process. Figure is taken from the publication of Grys et al.(2016)[2]

## 2.3 Spatial Tissue Organization Analysis

Past researches about the extraction of features focused mostly on a set of individual objects and did not take into account the relationship between these objects. Spatial organization is a new approach that focuses on such relations to describe, detect, and recognize these objects. It is particularly useful when objects are located in a complex and dynamic biomedical environment. In addition, spatial information can be more reliable than the characteristic of the objects themselves.

In this section, we describe some approaches to exploit the spatial organization of objects and perform a certain level of mechanical reasoning about the image contents. We observe the spatial organization at different scales, at local scale, between two objects, to global scale, between each object and a set of objects or a population.

### 2.3.1 Analysis at Local Scale

Analysis at local scale focuses on the relationship between two objects or between an object and its local surrounded area. In biological application, the relationship between two or a small number of objects can be modelled as the spatial relationship of one component to another. For example, the spatial relationship of cell to cell, cell to nucleus, cell to vesicle, nucleoli to nucleus, or between cells within a tissue. A spatial relation can be classified as a metric relation, or a topological relation, or a more complex one [50].

Metric relations consist of directional relation and distance relation. Examples of directional relations are ‘right to’, ‘left to’, ‘in front of’, ‘above’, ‘below’, ‘behind’. Examples of distance relations are ‘close to’, ‘far from’. Metric relations are quite simple to understand but play an important role in describing the object positions. More complex relations can base on this such as ‘between’, ‘surround’, ‘among’. Hudelot et al(2008)[50]. represented metric relations and the links between them as ontologies. This work combined ontology of spatial relations with fuzzy representation of distance relations in order to guide image interpretation and the recognition of its structures.

Topological relation is commonly used for qualitative spatial reasoning. Randell et al. [3] implemented a set of spatial relations such as contact, overlap, and the relation of part to whole in order to describe the topology and structural organization of biomedical images. In this work, the authors proposed the usage of Discrete Mereotopology (*DM*), a spatial logic which combines mereology, the theory of part-hood relations, and topology to model discrete spaces. Particularly, DM can be represented as Mathematical Morphology operations that enables the representation of spatial models in computer program.

These spatial relations are a set of contact and part-whole relationships in discrete 2D space between pairs of binary regions in a single image and between regions across different images, such as in multi-channel images. These models focus on ‘external contact’, ‘partial overlap’, and ‘tangential’ and ‘non-tangential’ connections of pairs of objects. The Discrete Mereotopology is a discrete variant of Region Connection Calculus (RCC) including the relations which are applied to discrete embedding space. We also include the popular sub-theory of RCC which are RCC5 and RCC8 as shown in Fig. 2.5. The RCC5 set covers five basic relationships between pairs of regions while the RCC8 set

gives more details which describes eight jointly exhaustive and pairwise disjoint relations defined on pairs of regions.

- $RCC5D = \{DR, PO, PP, PPI, EQ\}$ , with DR (‘is discrete from’) as the least upper bound of disconnection and external connections, PP (‘is a proper part of’), PPI is inverse of PP, and EQ (‘is identical with’).
- $RCC8D = \{DC, EC, PO, EQ, TPP, NTPP, TPPI, NTPPI\}$ , with DC (‘is disconnected from’), EC (‘is externally connected with’), PO (‘partially overlaps’), TPP (‘is a tangential proper part of’), NTPP (‘is a non-tangential proper-part of’), EQ (‘is identical with’), TPPI is inverse of TPP, and NTPPI is inverse of NTPP.

The model of spatial logic for DM is a promising way to extend to 3D images or even higher dimensional 3D+t biological images. Its input consists of two binary images, X and Y, that are the objects to be tested. The region can be every regions inside cells and tissues. In order to achieve this, we use the Boudier’s plugin [37] which represents Discrete Mereotopology map directly to the erosion and dilation operators of mathematical morphology. This work also extends the spatial relationship computation to 3D images.

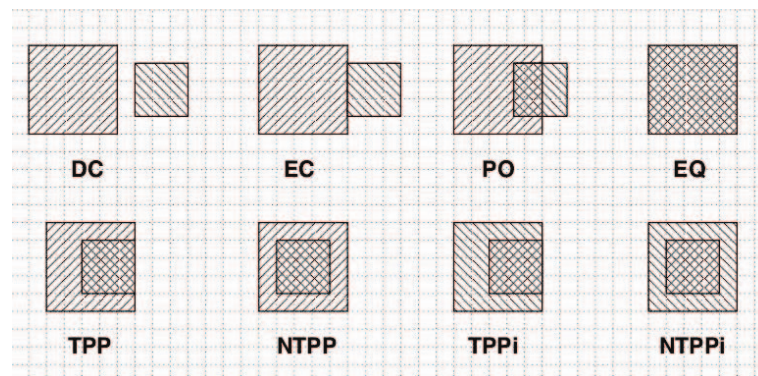


FIGURE 2.5: The RCC8D relations. The eight RCC8D relations. In each case, the relation shown is that of the square with NW–SE shading to the square with SW–NE shading. Figure is taken from the publication of Randell et al.(2013)[3].

### 2.3.2 Analysis at Global Scale

In biological environment, cells are not independent but form a tissue. Cells work together to perform their function in regulating and translating. The structure or arrangement of cells is an essential key to investigate the function of the entire tissue.

Therefore, there is an interest in determining whether the spatial organization of cells within tissue affects the tissue function. Our objective is to determine whether there is a tendency of the observed spatial organization of the cell to match a specific pattern in the study area which are not randomly distributed. The problem, then, becomes how to find a formal analysis or a descriptor which enables us to summarize the spatial information of the observed cells organization and figure out the specific relationships between them. Thus, this part focuses on the methods of quantitative spatial reasoning of cells at the global scale where the relation of an object with many objects in the study area is observed.

The point pattern analysis (PPA) is currently the best method for this problem. This method was developed originally at the ecology and epidemiology domain and has been adapted to many more. The cell coordinates can have many attributes, but at the global scale, we are only interested in their locations in the point pattern analysis. Each object can be replaced by its center position. Weston et al. [5] analyzed the point pattern of nuclear biology and determined its different distributions. Andrey et al. [4] also pointed out the distribution of centromeres and chromocenter in animal and plant nuclei.

### Different types of distribution

There are three general patterns [4] that are commonly used today as shown in Fig. 2.6).

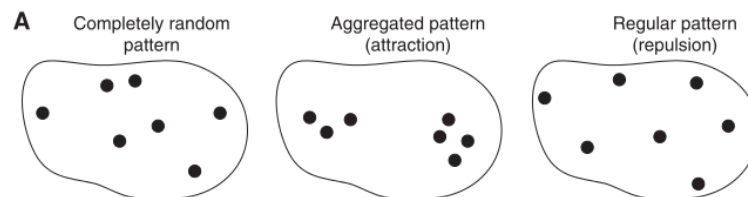


FIGURE 2.6: Various types of spatial distribution. Positions can be independently distributed (completely random pattern), or attraction (aggregated pattern) or repulsion (regular pattern). Figure is taken from the publication of Andrey et al.(2010) [4]

1. Completely random distribution : the position of any point is not affected by the position of any other point.
2. Uniform distribution, regular pattern, repulsion: every point is as far from all of its neighbors as possible.
3. Cluster distribution, attraction: many points are concentrated close together, and large areas that contain very few points.

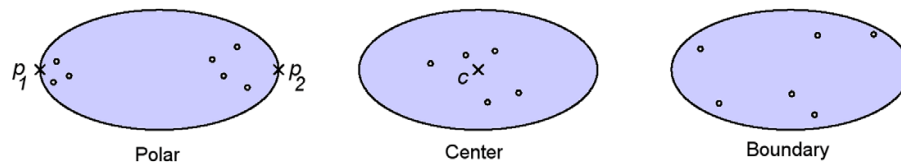


FIGURE 2.7: Illustrating of three specific distributions such as ‘polar’, ‘center’, ‘boundary’. Figure is taken from the publication of Weston et al.(2012)[5].

In the dynamic and complex biological environment, more specific distributions have been proposed such as ‘polar’, ‘center’, ‘boundary’ [5]. These patterns describe point pattern process that the components locate themselves near the poles, at the center, or close to periphery of the study area respectively. Examples of these specific patterns are that nucleoli located near the particular positions inside nucleus, vesicles are located near the cell periphery and have tendency to reach out of cell membrane and connect to outside environment.

### Spatial distribution function

In order to analyze the spatial distribution of cells and their components, we assume that the input image has been segmented beforehand and all regions are represented by their centers of gravity.

Visualizing a cell spatial organization can help to detect its pattern. However, in order to summarize the spatial information of a spatial pattern, a mathematical function is needed. Distance functions are a simple measurements and standard tools in the statistical analysis spatial point processes. Here we refer to two main distance functions which are commonly used. There are  $F$ -function which examines the cumulative frequency distribution of the empty region and  $G$ -function which allows the measurement of the cumulative frequency distribution of the nearest neighbor distances in the study area (see Andrey et al. (2010) [4], Weston et al. (2012) [5]. These functions allow us to compare an observed distribution, e.g. cells organization, with a specific distribution simulated in the same study area.

*F-function*  $F(r) = P(R < r)$ .  $F$ -function is the cumulative distribution function of the distance  $R$  between a typical position within the observed space and its closest point in the pattern. A stochastic scheme (such as Monte Carlo) was implemented to compute the  $F$ -function. A number of evaluation points, randomly chosen points or reference points, were generated uniformly at random within observed space. For each evaluation point, the distance to its closest point of the pattern was determined as

shown in Fig. 2.8. The cumulative distribution function  $F(r)$  was then estimated by the proportion of evaluation points for which this distance is below  $y$ .

*Steps to compute F-function.*

1. Assume that we have  $n$  observed points (event) in the study area:  $(e_1, e_2, \dots, e_n)$ .
2. Randomly select  $N_{ref}$  reference points  $(rf_1, rf_2, \dots, rf_N)$  in the study area  $A$ .
3. Compute the minimum distance from reference point  $rf_i$  to any observed point  $e_i$  in the point pattern:  $d_{min}(rf_i, e_i)$ .
4. Calculate the number of point pairs where the distance is smaller than or equal the distance  $d$ :  $F(d) = \frac{d_{min}(rf_i, e_i) < d}{N_{ref}}$ .

*Spatial Distribution Index (SDI)* : In order to determine whether the spatial distribution of objects matches any organizational rule for example randomly distributed, the observed distributions were compared against a completely random distribution conditioned on the observed numbers of objects (cells, vesicles, centromeres, chromocenters). A Monte-Carlo approach was applied, whereby the distance functions were computed over sets of simulated patterns. For each object such as vesicle, nucleus, telomere, chromocenter, random patterns were generated with the same number of objects as detected in the actual sample.

1. Verify the significance of any non-random pattern (either clustering or regularity) is evaluated using the theory of simulated ‘confidence envelopes’.
2. Simulate  $S$  spatial point processes and estimate the F-, G- function for each of these processes. Practically, we usually choose  $S$  equal 1000 or 100 processes.
3. Then rank all the simulations based on the results of distance.
4. Mark the 5<sup>th</sup> and 95<sup>th</sup> of  $F(r)$  or  $G(r)$  values (black and grey curves in Fig. 2.8).
5. Plot these as the 95% confidence intervals (between 5<sup>th</sup> and 95<sup>th</sup> values).
6. Compare the observed pattern with  $S$  reference patterns.

### 2.3.3 Modeling of Spatial Organization

Using 3D modeling, we can build virtual structures based on various biological hypotheses, which can help biologist to visualize in 3D their hypotheses and quantitative spatial analysis for each case. In addition, modeling is an alternative strategy in case there are

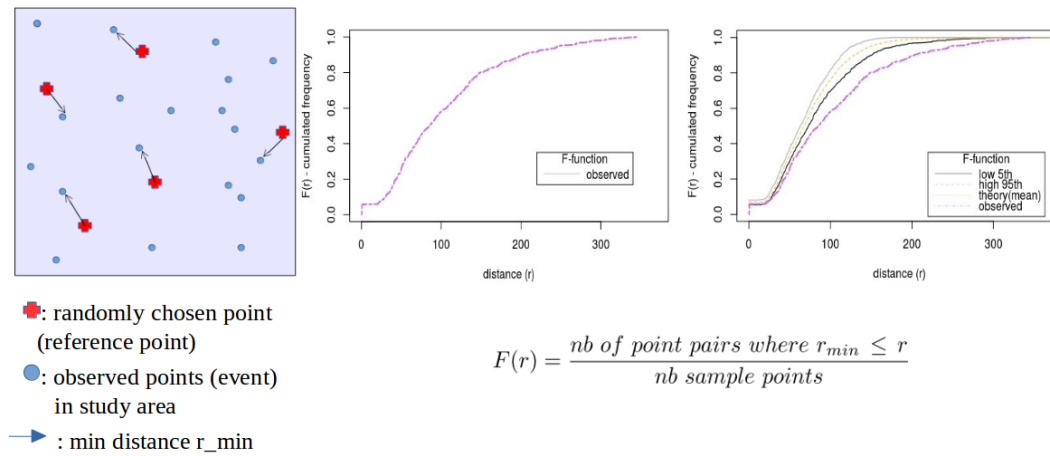


FIGURE 2.8: F-function, cumulative distribution function of the empty region

limited number of the available imaging data. The reason is the high time consuming to capture a high resolution image and some samples could take up to one to two weeks.

Moreover, due to the amount of time required to collect and prepare sample data, it is difficult to capture the expected variation in a cell structure. In the normal case, a biologist could only capture several instances. In such situation, modeling of spatial organization is very useful in generating many variances of the structure. One more advantage of the modeling approach is to be able to generate synthesis data for statistical tests. The modeling procedure is covered in Chapter 3.

## 2.4 Objectives of the Thesis

### Analysis and modeling of tissue spatial organization in 3D

Over the last decade, numerous works have been done to automate the analysis of cells in microscopic images. Nevertheless, these works focus mostly on applying image processing and analysis techniques on individual cells. There is a general lack of tool to analyze the interactions among cells and their spatial organization. In addition, there is a demand for more efficient approaches to handle high throughput 3D images of cells. This comes from the fact that the study of cells in three dimensional (3D) space provides a better and more realistic way to do researches with regard to the physical and biochemical properties of the entire micro-environment comparing to two dimensional approaches.



### **Cyto-architecture of the islets of Langerhans**

The islet of Langerhans is an open research topic in biology. There is considerable interest in analyzing the multi-cellular structure of the islet at both local and global scale. However, there are not many works focusing on formalizing and modeling its spatial organization. Therefore, this is one of the main goal of this thesis. In addition, we collaborates closely with domain experts, Dr. Rafael Arrojo E. Drigo, and Dr. Per-Olof Berggren who have many experiences working with the islet of Langerhans.

### **Contributions**

1. We provide a set of methodologies and a toolbox for the analysis of tissue organization. They include cell segmentation, analysis of multi-cellular interaction, analysis of spatial organization, and modeling of spatial organization.
2. We apply our toolbox to the investigation of the spatial organization of the islet of Langerhans trying to understand its internal mechanism.
3. We apply our toolbox to explore the function of delta cell in the islet, whose role is yet to determine.
4. We introduce the procedure to model the cell spatial organization, which can be used to create virtual models of tissues and sample data for testing.

### **Outline of the thesis**

In the next chapter, we introduce the image processing methods and mathematical models that we use to handle confocal microscopic images. After that, in chapter 4, we present our research results comparing with other methods in 3D image processing. That chapter is followed by a discussion about the implementation, perspectives, and potential future works.

The documentation of our toolbox as well as all the extra results of every datasets (mouse, monkey, human) that we used are included in the Appendices A and B.

## Chapter 3

# Materials and Methods

In this chapter, we describe our set of methods to analyze the tissue spatial organization. Its objective is to propose a generic workflow to quantify, analyze, and model the entire tissue spatial organization. Our general algorithms will be presented first, followed by their application to the islet of Langerhans. The chapter covers:

1. Our materials and the workflow to analyze tissue spatial organization.
2. The main computational methods to perform cell segmentation (nuclei segmentation, cell zone computation, 3D cell phenotype detection) and to analyze the spatial organization (cellular interaction computation, spatial statistic analysis).
3. The application to study the role of delta cell.
4. The modeling procedure to create the virtual distribution models.
5. The implementation of our algorithm.

### 3.1 Materials

#### Reconstruction data in 3D

We provide the capturing of both sections and whole slides of the tissue in our works. In order to prepare the serial section reconstruction, the tissue sample is separated in  $T$   $\mu\text{m}$  thick sections and placed on slides. The parameter  $T$ , the section thickness, is chosen depending on the type of sample and the optical property of the lenses. Normally, thick sections ( $T$  from 30 to 100  $\mu\text{m}$ ) reduces the number of gaps in the reconstruction. In

the application of the islet of Langerhans, we choose  $30\ \mu\text{m}$  as the value of  $T$  and divide each sample into 2 to 3 parts. For the capturing of the whole slides tissue, we adopt the optical clearing technique from Ke et al. (2013) [51].

In the application of islet of Langerhans, the thickness of whole slides sample is greater than  $80\ \mu\text{m}$ .

### **Confocal microscopy and image acquisition**

There are many different image modalities for the study of tissues and they need to be processed differently. Two common techniques are the fluorescent imaging and histological imaging (Section 2.2.1). In this thesis, we focus only on the fluorescent imaging technique in which images are captured by confocal microscope. Specifically, for the islet of Langerhans, we use single photon confocal microscope equipped with a white laser and hybrid detectors. More details will be covered in Chapter 4.

The inputs we need consist of a nuclei labeling channel and several marker channels  $M_1, M_2, M_3, \dots, M_n$  corresponding to different cell phenotypes. Specifically, our input has four channels, a nuclei labeling image and three markers glucagon (produced by alpha cells), insulin (produced by beta cells), and somatostatin (produced by delta cells). The nuclei is stained with DAPI (4, 6-diamino-2-phenylindole) while different marker channels are highlighted using different focal planes (staining methods).

## **3.2 Methodology Overview**

### **Generic workflow for tissue spatial organization analysis**

Based on our studies of other related works in Section 2.2.2, we divide our workflow into two stages:

1. Segmentation and cell phenotype detection. This covers all our computational methods to get the cell profile which contains all the information about the cell including its phenotype and the coordinates of the nuclei and the cytoplasmic regions.
2. Spatial organization analysis. This analyses cellular interactions at local scale and spatial organization at global scale. In addition, the modeling and simulation of the spatial organization is also done here.

(1) The segmentation and cell phenotype detection stage includes:

1. *Nuclei detection.* This is the core function to detect and segment nuclei (Section 2.2.4).
2. *Cytoplasm detection.* It is used when the input contains cytoplasm marker channels. A cytoplasm region is defined as the region surrounding each nucleus. This function uses the segmented nuclei image and the cytoplasm marker as its inputs.
3. *Membrane detection.* It detects the cell membrane surrounding the nucleus and the cell cytoplasm. If the input contains the cell membrane marker channel, it is used to segment the membrane and obtain the cell boundary.
4. *Cell zone determination.* A cell region is determined by combining the nuclei, cytoplasm, and the cell boundary regions. However, when there are no cytoplasm or membrane markers, the cell region is defined as the region surrounding each nucleus with a predefined radius  $d$ .
5. *Cell phenotype detection.* By examining the amount of markers at the cell region, we can detect the phenotype of the cell (Section 2.2.7).

(2) The second stage of spatial organization analysis includes:

1. *Analysis of cellular interaction.* We analyze the cell-to-cell contacts among neighboring cells.
2. *Analysis of spatial organization.* We determine the spatial distribution of objects within tissue. It could be a random, a clustered, or a regular distribution.
3. *Modeling and simulation of spatial organization.* This generates a virtual model of cells within the tissue and applies this virtual model to test new hypotheses.

### **Workflow for application of the islet of Langerhans**

The same two-stage workflow is used to analyze the islet of Langerhans with a small modification in the first stage because of the difficulty to produce the cytoplasm marker. Therefore, we skip the cytoplasm detection part. The detailed workflow is presented in Fig. 3.1 and Fig. 4.1.

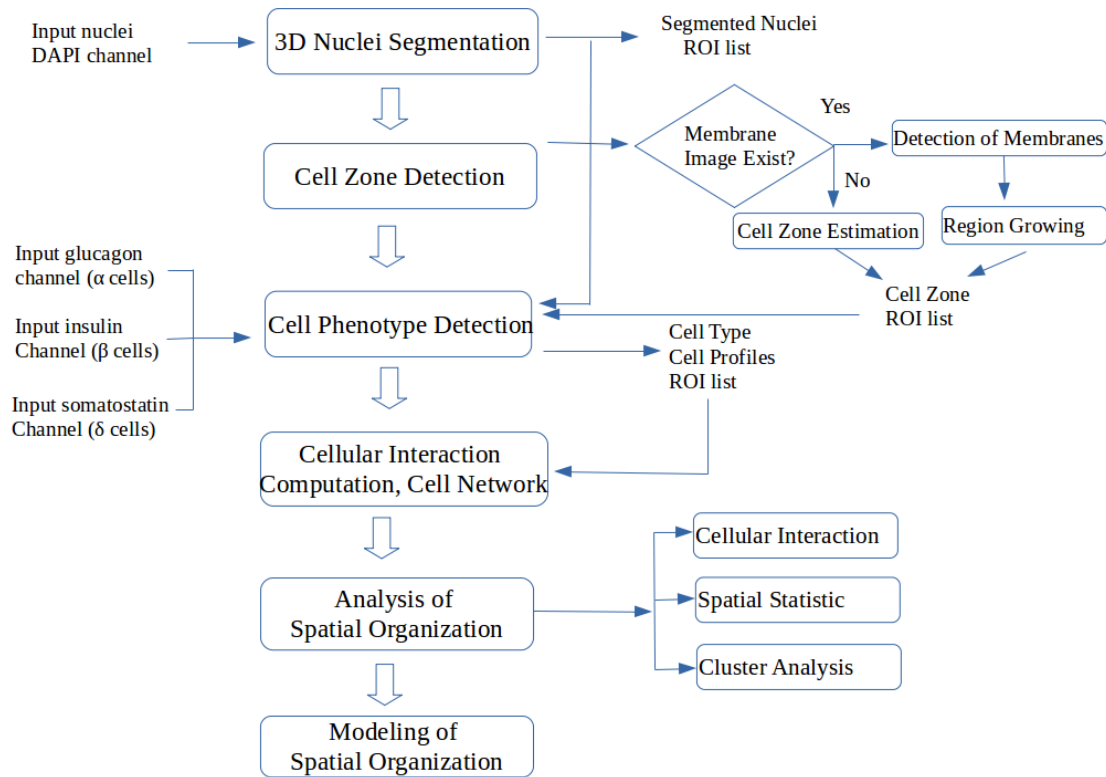


FIGURE 3.1: Workflow of the processing. First, nuclei is segmented and cell zone is computed. From three markers glucagon, insulin, somatostatin as input image and detected cell zone, cell phenotype is identified for each cell. Based on the cell profiles from first stage, we investigate the spatial organization of islet of Langerhans.

### 3.3 Segmentation and Cell Phenotype Detection

In the first stage, we present a set of algorithms to perform various image analysis tasks including those that can handle thick tissue samples. These methods all use the voxel as the basis unit.

#### 3.3.1 Nuclei segmentation

##### General introduction

Nuclei segmentation is the very first as well as the most important step in our process. It can potentially affect the results of all the subsequent analysis. The algorithm we use is first presented in Heck et al. (2015) [52] and has already been implemented as the spot segmentation feature in 3D ImageJ Suite [37] plugin by Boudier's group. We have modified this method to make it work with thick nuclei samples.

### Seed detection

The key idea of the algorithm is to use seed detection in the segmentation. The original nuclei image is transformed to remove noise and to capture different frequencies corresponding to different features in the image. We then compute the local maximal values from the transformed image and use these values as seeds to group neighboring voxels.

The input of 3D spot segmentation algorithm consists of the original image (nuclei labeling image - spots) and all the seeds obtained from the above process. We propose the use of Fourier transform as the transformation function for seed detection.

### 3D spot segmentation

The spot segmentation algorithm has five steps: (i) pre-processing, (ii) seed detection, (iii) voxel clustering, (iv) segmentation, and (v) an optional semi-automatic segmentation to correct the error of the segmentation step.

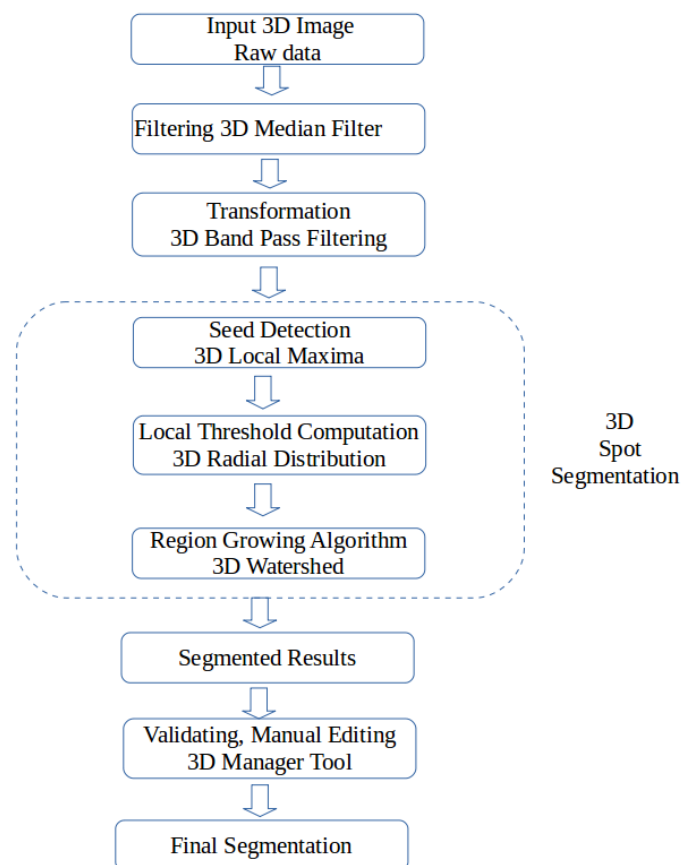


FIGURE 3.2: Diagram of 3D spot segmentation.

*i) Pre-processing.* Images with nuclear labeling (DAPI) are first processed by a 3D median filter to reduce noise and to homogenize intensities inside the nuclei. We choose to use the median filter, a simple yet powerful filtering method on noisy confocal images.

It is robust to noise level and can preserve the edge of the object (Section 2.2.3) while other methods tend to extend to the background and merge different objects together.

*ii) Nuclei seed detection.* The filtered image is then transformed one more time by a band pass filter with a specific size interval of  $[D_{min} - D_{max}]$ . It corresponds to the range of the nuclei diameter (Section 2.2.3). Here we utilize the 3D band pass filter implemented in the lipid droplet plugin written by Samuel Moll [53].

The resulting image is encoded in 32-bits which is then scaled down to a 16-bit image. Our plugin 3D spot segmentation [37, 52] takes it as input to detect the nuclei seed with a local maxima in the radius of  $r = 4$  voxels.

Because the detected seeds can be in either the noise region or in the nuclei region, a threshold  $T_{seed}$  is set to only keep the latter. Seeds whose intensity values are smaller than this threshold are excluded. Because the filtered image is encoded in 16-bit, we can choose the threshold  $T_{seed}$  of around  $30000 \pm 500$ .

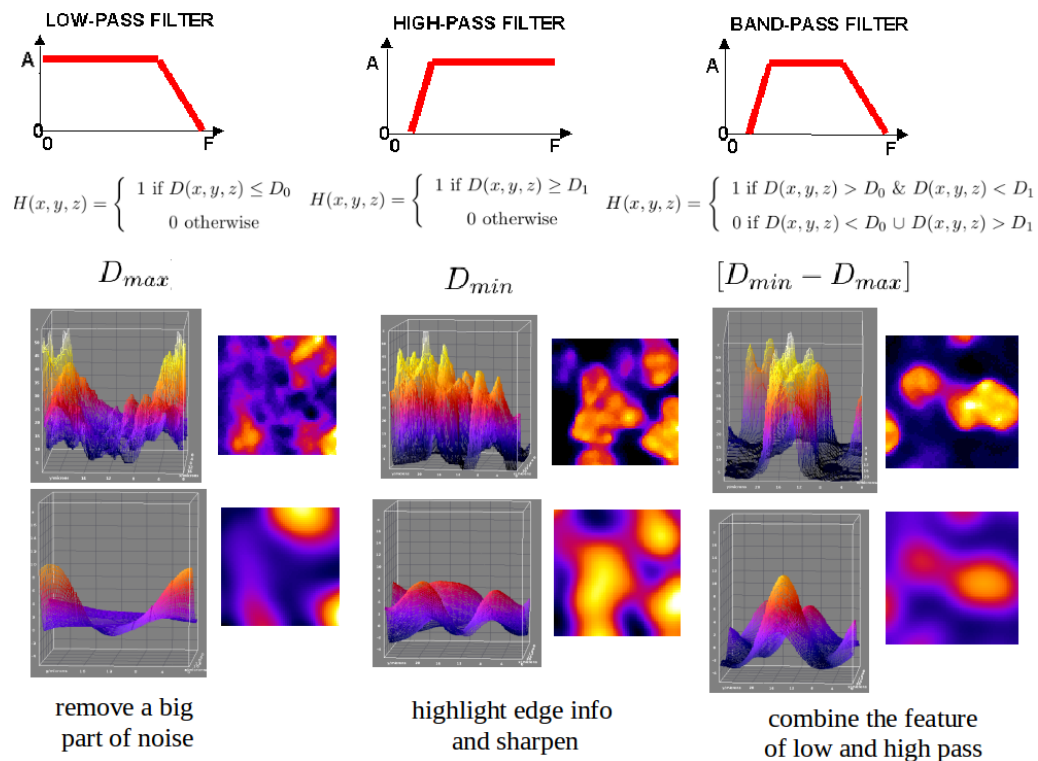


FIGURE 3.3: Band pass filtering and its help to facilitate the nuclei seed detection.

*iii) Voxel clustering - determine the extension of the object.* After we have detected all the seeds, one per nucleus, the next step is to properly delineate each nucleus around

its seed. Different techniques can be applied here such as voxels clustering used by the spot segmentation or utilizing level sets.

Voxel clustering considers the seed as the center of the nucleus and group neighboring voxels around it. Therefore, an important question is how to estimate the extension of the nuclei object as a terminating condition for the voxel clustering algorithm. We decide to use a threshold in terms of the intensity value to achieve this.

The threshold can be a global or a local threshold within a specific region. However, due to the big variance of intensity values in the nuclei image, a global constant threshold does not work well in many cases. Therefore, we stick with local thresholds defined by the Gaussian fit method in our works. We compute the 3D radial distribution of the mean intensity values in growing concentric layers around each seed. From this radial distribution, we could fit a Gaussian curve and find a cut-off on it to a specific standard deviation value (SD value). This SD value is in the range  $[-1, 3]$ . In practice, the SD value is set at 1.17 meaning that the threshold will cut the Gaussian curve surface at the half maximum. For example, if the SD value is set at 2 and 3, it will fill about 90% and 99% of the surface respectively.

*iv) Segmentation.* The computed value of the Gaussian fitting is then used as a local threshold for a region-growing algorithm around each seed in which all neighboring voxels are examined. Instead of considering one voxel at a time, a block of voxels with the intensity value greater than the local threshold is processed and added to the cluster of voxels instead. Finally, the output nuclei regions are returned where voxels in the same nuclei object share the same label.

*iv) Watershed - region growing.* As part of the segmentation process, a 3D watershed algorithm is also performed to define the area around each seed. This is done to ensure that the segmented nuclei object will not propagate to its neighboring nuclei.

*v) Semi-automatic segmentation.* As an optional step, manual editing is needed when it deems necessary to correct the few segmentation errors. We use our 3D Manager tool from the 3D ImageJ suite [37, 54] to assess the segmentation accuracy and to perform the correction. Specifically, when we observe a nuclei object split into two, we can select these two nuclei objects and apply the merge function of 3D Manager to merge them



into one. Similarly, we can also use the split function of 3D Manager to split an object into two when they are incorrectly merged together.

### 3.3.2 Cell zone computation

#### General introduction

After the nuclei segmentation, the next step in our workflow is to detect the cell membrane. Note that there are existing methods to detect the cell cytoplasm as well but it is difficult in practice to capture the cytoplasm labeling image. Therefore, we do not include cytoplasm detection in our works.

In general, there are two ways to detect the cell membrane:

1. When the cell membrane labeling image is captured perfectly, we segment this image to get the cell boundary in its closing form. We can then apply the region growing technique around each nuclei seed and inside this boundary to obtain the cell region.
2. When it is difficult to capture a good membrane image or there might be no membrane image at all, we should use an ad-hoc method to estimate a cell zone. In such cases, we use region growing technique to estimate the cell zone surround each nuclei.

In our works, we study both approaches to understand their advantages as well as limitations.

#### 3.3.2.1 Cell membrane segmentation

##### Setup experiment

We first capture the 3D membrane labeling images of several mouse and monkey islets, one image per islet. These membrane labeling images together with those nuclei labeling and glucagon marker images are what we need for the experiment. Note that the number of input images is smaller than usual because it is difficult to capture many markers at the same time for one sample.

##### Membrane detection

We follow the proposed pipeline in Keraudren et al. (2011) [42] for cell segmentation

which includes four consecutive stages: filtering, nuclei segmentation, edge-map processing of the membrane, and applying an adaptive watershed using nuclei as seeds 3.3.2.1. Specifically, we use 3D Canny Deriche method from the Edge and Symmetry filter of Image Edge plugin [37] to capture the gradients of the membrane image. The Alpha parameter controlling the level of smoothing in canny edge detection was set to 0.1. The proposed workflow for membrane detection and cell segmentation is shown in Fig. 3.4.

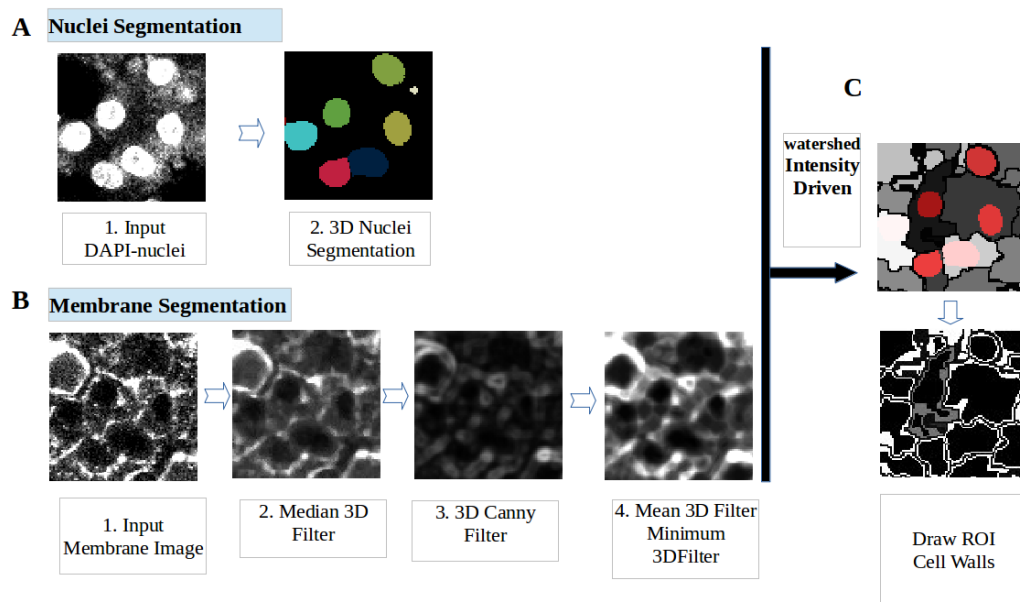


FIGURE 3.4: Proposed workflow for cell segmentation which divide into nuclei segmentation and membrane segmentation and combine them together. (A) Nuclei is segmented using our spot segmentation method. (B) Raw membrane image is filtered using 3D median filter, after that apply 3D Canny Deriche Filter and normalization to capture the edges of membrane, and applied 3D mean filter to smooth intensity inside the edges and then applied 3D minimum filter to reduce the size of edge. (C) The watershed intensity driven approach is adopted using segmented nuclei as seed in the region of filtered membrane image. We obtained the cell region result. From cell region, we can get the contours of region as the cell wall.

### Limitations

Due to the thickness of the sample, we cannot produce good membrane labeling images. As a result, there are many gaps in the membrane. This leads to both over-estimation and under-segmentation of the cell region. If the cell region could not be determined correctly, this could potentially affect all subsequent computations.

### 3.3.2.2 Cell zone determination - Watershed separation

When there is no membrane image or its quality is not good enough, we can use a watershed procedure to estimate the cell zone. This procedure computes the Voronoi

region surrounding each seed nucleus on the nuclei segmented image.

### **Our proposed method**

We define the cell zone around each nucleus using a watershed separation with a predefined radius. A 3D Euclidean distance map is computed on the background of the segmented image for all nuclei (Section 2.2.5). The output is a 3D image where the value of each voxel represents the minimal calibrated distance to the closest nucleus boundary. Note that our 3D distance map implementation works using calibrated unit and hence take the anisotropy in  $Z$  into account.

The distance map image is then inverted so highest values are close to nuclei boundaries. After that, the watershed procedure is applied on inverted distance map image. Basically, all pixels are put into an ordered list and processed sequentially. The first pixel with the highest value is observed and processed. The algorithm performs a search on the 3D neighborhood of the observed voxel to examine all voxels who are already assigned to a nucleus. Next, non-labelled voxels from the 3D neighborhood of the observed voxel are added to the list based on their distance map values. When two defined zones touch each other, the voxel at the frontier between these two zones is assigned a specific label in order to mark the boundary of the cell zone. Finally, a mask is applied and set to 0 for voxels whose values in the distance map image are greater than the predefined radius.

*Validation:* We rely on domain experts to validate the cell zone detection result.

### **3.3.3 3D cell phenotype computation**

In this section, we work with an algorithm to identify different cell types using the input obtained in the previous step. Specifically, the input consists of different markers corresponding to different cell phenotypes as well as the cell segmentation result. In our case, the markers are insulin, glucagon, and somatostatin corresponding to three cell types alpha-, beta-, and delta- respectively. The segmentation result includes information about the nucleus and the cell zone such as voxel coordinates and voxel values.

An overview of other approaches to identify phenotypic profiles is covered in Section 2.2.7. In our study, we apply 3D cell phenotype computation on the islet of Langerhans to classify cells into 4 groups: delta (label 1), beta (label 2), alpha (label 3), and ‘*unlabelled*’ (label 0) cells.

Our two-stage workflow includes:

1. Generating phenotypic profiles by nuclei segmentation and cell zone computation.
2. Clustering and classifying phenotypic profiles.

There are several reasons to support the clustering approach:

1. There are only 3 cell phenotypes in our application. Our target is to detect only alpha-, beta-, and delta- cells. Therefore, clustering is the most simple and efficient.
2. In our experiment, we could not obtain the accurate membrane labeling image and there is a genuine lack of cell boundary information. To make it worse, cells of a thick tissue sample in three dimensions are elongated across many slices and it is a challenging task to draw the cell region. As such, it is difficult to prepare the training data and to apply supervised machine learning approaches.
3. Cell characteristics. Some common features are required in order to generate phenotypic profiles, for example the object volume, texture, the histogram of intensity inside segmented cells and the shape of object. However, there is no information about the cell shape in our case. We only have information about the content within the cell region such as the histogram of intensity of the voxel at the cell region. So clustering is a suitable option in our application.

### **The difficulty of detecting the phenotype of endocrine cells**

Bosco et al. (2010) [6] detects the unique association between alpha- and beta- cells in human islet cells from at least 10 different pancreas. They detect some cell pairs which compose of one alpha-cell surrounded by one beta-cell (beta wrapping alpha), or neutral apposition between alpha- and beta-cells (alpha-beta), or a beta-cell wrapping an alpha-cell (alpha wrapping beta) (Fig. 3.5). As the result, this can lead to mis-identification of cell phenotype between alpha- and beta- cell in our application. In order to identify

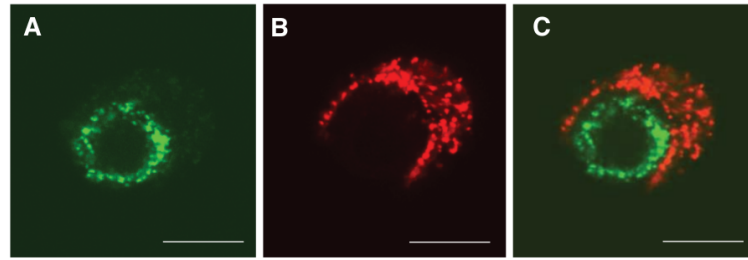


FIGURE 3.5: The association between alpha- and beta- cells in cultured human islet cells, insulin (red) and glucagon (green). A cell pair composed of one alpha-cell (A) surrounded by one beta-cell (B). Merged image is shown in (C). The cell pair shown here is representative of cell pairs observed in different human cell preparations from at least 10 different pancreas. Figure is taken from the work of Bosco et al. (2010) [6].

the cell phenotype, we assume that each cell has only one cell phenotype and we observe the cytoplasmic region that is close to nuclei boundary only.

### The algorithm

The algorithm to detect cell phenotype includes the following four steps:

1. Normalizing three marker images.
2. Supervoxels clustering. In order to reduce the complexity in the number of voxels in each dataset, we cluster voxels from different images containing specific markers into supervoxels. A supervoxels is a group of voxels which share similar intensity values.
3. Each supervoxel is assigned a cell phenotype based on the most abundant label with the highest average median intensity value.
4. Finally, we define a study area for each cell and identify the cell type based on the most abundant intensity in this study area.

#### (1) Normalization of markers

During the preparation of samples, there are many factors that can affect the quality of microscopic images. By consequence, each input dataset could have a different histogram and a different noise level. Pre-processing is, therefore, very important to enhance and normalize the intensity level. It also reduces the level of noise from the image.

In our works, information about the markers (glucagon, insulin, somatostatin) plays a very important role in deciding the type of each cell. At the same position, if the noise level is very high in one channel and the cell marker is lower in another, the cell type detection algorithm could detect wrongly. Therefore, we use the median filter to

enhance the signal of all channels and apply a small threshold to remove parts of the noise. Finally, we normalize all channels to the range  $[0, 255]$ .

## (2) Supervoxels clustering

We need an efficient supervoxels clustering algorithm due to the high number of markers in the input image. It should cluster voxels quickly and achieve a good overall accuracy.

Specifically, Simple Linear Iterative Clustering (SLIC) is the most suitable algorithm for our problem. Besides its accuracy and high performance, SLIC can also handle 3D supervoxels.

SLIC was first proposed by Achanta et al. (2010)[55] as a novel algorithm that could cluster pixels in the combined five-dimensional color and image plane space. It was shown to have greater segmentation accuracy than other state-of-the-art methods. We adapt the SLIC method to generate supervoxels from three marker channels and their spatial distances in three dimensional space.

### SLIC algorithm

We use SLIC Supervoxel with Complex Distance Measure to segment the volume of interest. The supervoxel of size  $M$  is the volume of the supervoxel corresponding to voxels within the special region of tissue such as the nuclei. The SLIC algorithm is divided into several steps as follows:

#### *Initialization*

This first step divides an image into  $k$  clusters of equal volume. We define the SLIC spacing parameter  $S$  with  $S = \sqrt[3]{N/k}$ ,  $N$  is the total number of voxels in the volume. This parameter, thus, depends on the supervoxels size.

In 3D, the study domain is a cube, not a block and the 3D confocal microscopic image has lower axial resolution along the  $z$ -axis comparing to its lateral resolution along its  $x$ - and  $y$ -axis. Therefore, we should take this into account by multiplying the size of the cube with the size of the voxel.

Let  $x, y, z$  be the coordinates of voxel  $i$ . Let the vector of signal intensities at each voxel be  $I_{x,y,z} = [I_{\alpha i}, I_{\beta i}, I_{\delta i}]^T$ , where  $I_{\alpha i}$  is the intensity value of voxel  $i$  at glucagon marker,  $I_{\beta i}$  is the intensity value of voxel  $i$  at insulin marker, and  $I_{\delta i}$  is the intensity value of voxel  $i$  at somatostatin marker.

The cluster center  $c$  is

$$I_{ci} = [I_{\alpha i}, I_{\beta i}, I_{\delta i}, x_i, y_i, z_i]^T$$

. And let  $d_{ci}$  be the distance from voxel  $i$  to  $c$  in the signals space, and  $d_{si}$  is the distance from voxel  $i$  to  $c$  in the Euclidean distance space.

A variable  $m$  is introduced as the compactness level which is used to weight the ratio of signal intensity similarity and the spatial proximity. The compactness level  $m$  decides the compactness and shape of the supervoxels.

The formula to compute the distance  $d_{ci}$ ,  $d_{si}$ , and the final distance  $D$  is presented in Equation 3.1. SLIC combines the intensity distance  $d_{ci}$  and the spatial distance  $d_{si}$  into the final distance  $D$ .

$$\begin{aligned} d_{ci} &= \sqrt{(I_{\alpha} - I_{ci})^2 + (I_{\beta} - I_{ci})^2 + (I_{\delta} - I_{ci})^2} \\ d_{si} &= \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \\ D &= d_{ci} + \frac{d_{si}}{S} \times m \end{aligned} \quad (3.1)$$

### (3) Assign a voxel based on intensity similarity and spatial proximity.

Each voxel will link to its nearest cluster center  $c$  in the search region which is extend to  $2S \times 2S \times 2S$  around the supervoxels center. Limiting the search region helps SLIC reduce its computational time comparing with the K-means approach. Specifically, K-means searches the entire image while SLIC only needs to handle a small region.

#### *Updating the cluster center assignment*

Updating the cluster center assignment is done repeatedly until a predefined number of iterations  $N_{iteration}$  is reached, this step has the goal of finding the nearest cluster center for each voxel  $i$ .

#### *Post processing, enforce connectivity*

This step is done by re-assigning any disjoint voxels to a nearby supervoxels and making sure that all the voxels within a supervoxel are connected.

#### *Implementation*

The algorithm 1\* is implemented as a Java plugin under ImageJ platform.

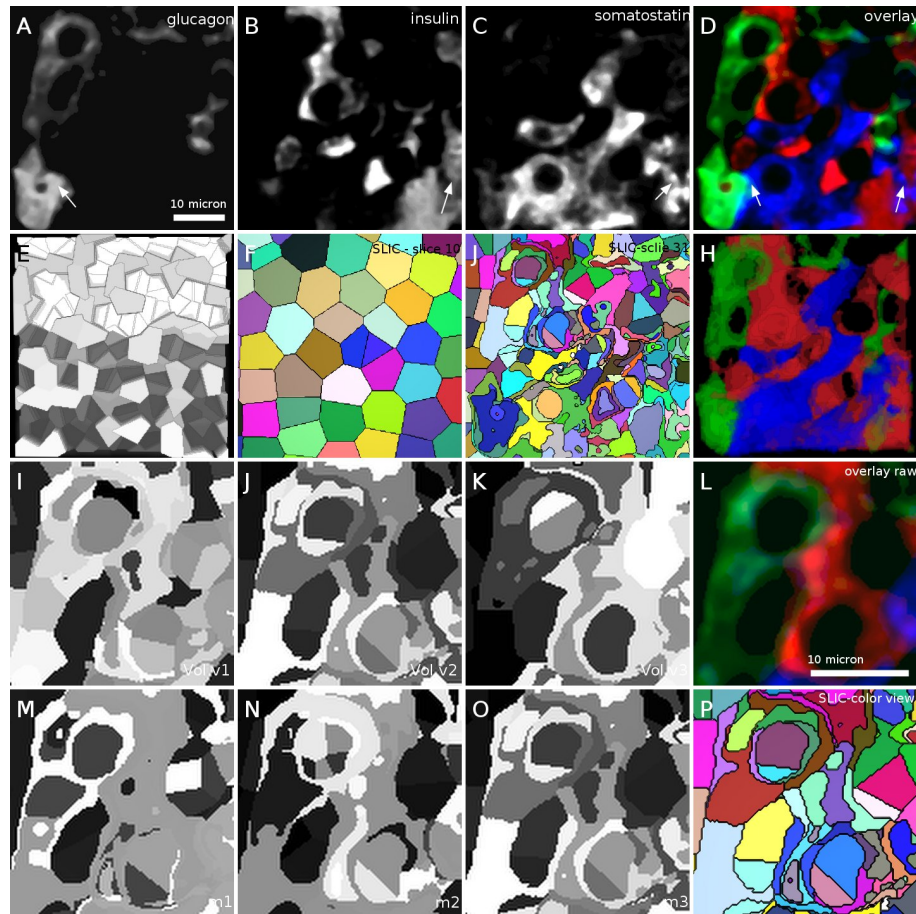


FIGURE 3.6: SLIC Clustering. (A-C) original data, scale bar  $10\mu$  (D) overlay, arrow: the part containing multi markers. (E-H) Result of SLIC in 3D Viewer, E - 3D visualization gray scale, F - slice at background and noise, G - slice contained 3 markers, SLIC median value correspond to each channels. (I-K) volume test, correspond to small[50-100], medium[100-300], big[300-500], (L) overlay of 3 channels. (M-O) compactness test, (M) small compactness  $m=1000$ , (N) medium compactness  $m=5000$ , (O) big compactness  $m=10000$ , (P) 3D view with different colors correspond to different supervoxels regions.

#### (4) Applying SLIC to 3D cell phenotype detection of endocrine cells

##### *Assigning a cell phenotype for each supervoxel*

The output of SLIC algorithm is a SLIC-supervoxels image that contains many supervoxels (Fig. 3.6E,F,J).

Let  $Z$  be a supervoxel. The median value of each channel is computed inside each zone:

1.  $Z_\alpha = \text{median}(I_\alpha)$ .
2.  $Z_\beta = \text{median}(I_\beta)$ .
3.  $Z_\delta = \text{median}(I_\delta)$ .



---

**Algorithm** (1\*) Adapted SLIC Supervoxel Segmentation with Complex Distance Measure

---

1: **Inputs:**

3 raw marker images (three channels correspond to three markers  
: somatostatin, insulin, glucagon)  $I_\alpha, I_\beta, I_\delta$   
compactness parameter  $m$   
number of iterations  $N_{iteration}$

2: **Initialize:**

Cluster centres  $C_k = [I_\alpha, I_\beta, I_\delta, x, y, z]^T$  by sampling at the grid  
intervals  $S$

3: **Set:**

label  $l(i) = 0$  for each voxel  $i$ ,  
distance  $d(i) = \infty$  for each voxel  $i$

1: **repeat**

2:   **for** each cluster center  $C_k$  **do**  
3:     **for** each voxel  $i$  in  $2S \times 2S \times 2S$  in region around  $C_k$  **do**  
4:       Calculate the distance  $D$  between  $C_k$  and  $i$   
5:       **if**  $D < d_i$  **then**  
6:         Set  $d_i = D$   
7:         Set  $l(i) = k$   
8:    Compute new cluster centers  
9:    Compute residual error  
10: **until** Number of Iterations reached  $N_{iteration}$   
    Enforce connectivity

11: **Outputs:**

One image which labeled each supervoxel

---

Each supervoxel is labelled with a cell phenotype using the most abundant label with the highest average median intensity value. Otherwise, if all median values inside the zone are lower than the predefined threshold  $T_{med}$ , this supervoxel  $Z$  is an unlabeled type (label 0).

For example, if  $Z_\delta$  is greater than  $Z_\beta$  and  $Z_\alpha$  and  $Z_\delta$  is greater than  $T_{med}$ . The supervoxel  $Z$  is assigned with the label 1. In other words, it has delta phenotype.

A 3D-median-SLIC image is then created with voxels having the value of their corresponding supervoxels type (Fig. 3.6H).

*Defining a study area for each cell and identify the cell type*

Next, we identify the cell phenotype using the cell segmentation result and the 3D-median-SLIC image from the previous step. To achieve this, we need to define the study

Nb Iterations	Compactness	Min Volume	Max Volume	Execution Time (s)	Test
10	10000	50	100	368	Cluster Size, Small, 19360 regions (Fig. 3.6I)
10	10000	100	300	376	Cluster Size, Medium, 5766 regions (Fig. 3.6J)
10	10000	300	500	383	Cluster Size, Big, 3380 regions (Fig. 3.6K)
10	1000	100	300	429	Small Compactness, 5766 regions, (Fig. 3.6M)
10	5000	100	300	375	Medium Compactness, 5766 regions, (Fig. 3.6N)
10	10000	100	300	383	Big Compactness, 3380 regions (Fig. 3.6O)
10	10000	100	300	383	Medium Iterations, 3380 regions
50	10000	100	300	1937	Big Iterations, 5766 regions
100	10000	100	300	3830	Big Iterations, 5766 regions

TABLE 3.1: Execution time of each experiments with small, medium, and large size of cluster size. Testing with monkey islet sample, 203 x 203 x 40  $\mu\text{m}$ , 1024 x 1024 x 40 pixels

area  $A$  inside each cell region and compute the volume of each marker inside  $A$ .

Let  $V_\delta$  be the volume of voxels within the study area  $A$  whose label is 1,  $V_\beta$  with the label 2, and  $V_\alpha$  with the label 3. As mentioned previously, if all the marker volumes inside the study area  $A$  are lower than predefined threshold volume  $T_{vol}$ , the cell becomes an unlabeled type (label 0 - ‘unlabelled’ cell).

We can also assign the phenotype of the cell as the one with the biggest volume. For example, if  $V_\delta$  is greater than  $V_\beta$  and  $V_\alpha$  and  $V_\delta$  is greater than  $T_{vol}$ . It is a delta cell (label 1). The same logic applies to beta cells (label 2), alpha cells (label 3), and ‘unlabelled’ (label 0).

#### *Different strategies to determine a study area*

In order to detect the cytoplasmic signal with high specificity, we need to look for markers in the SLIC image around the nuclei boundary (Algorithm 1\*\*). We define two kinds of distance as follows:

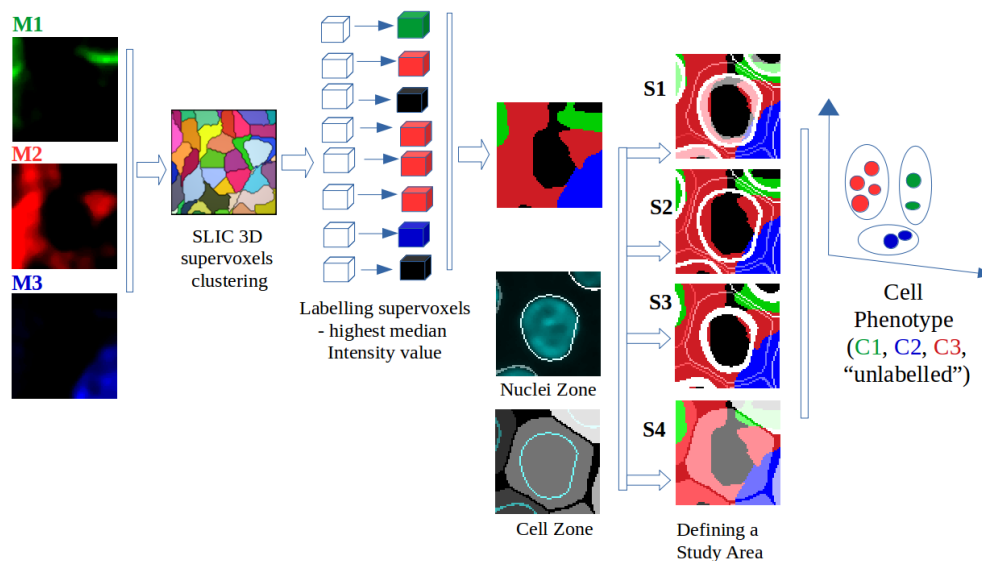


FIGURE 3.7: Scheme for 3D cell phenotype computation. (1)(2) Normalization of markers and SLIC Supervoxels clustering. (3) Labeling a cell phenotype or unlabeled for each voxel. (4) Defining a study area. (5) Detect cell type

1. The inner distance from the nucleus boundary towards the inside of the nucleus. It detects fluorescence signal that may appear inside the segmented nucleus;
2. An outer distance from the nucleus boundary towards the outside of the nucleus and inside of cell zone.

To determine the study area, we compute an Euclidean distance map inside each nuclei region and another Euclidean distance map at the region outside of the nuclei in the image space. We then select a strategy to compute the study area depending on the cell characteristic and research problems. The impact of each strategy to the accuracy of cell phenotype detection is discussed in Section 4.2.3.

1. Counting from the nucleus boundary towards the inside of the nucleus with the inner distance  $d_1$ . Counting from the nucleus boundary towards the outside of the nucleus with outer distance  $d_2$ . We call it study area A1.
2. Counting from the nucleus boundary towards the outside of the nucleus with the outer distance  $d_2$ , called study area A2.
3. Counting from the nucleus boundary towards the inside of the nucleus with the inner distance  $d_1$ , called study area A3.
4. Observing the entire cell zone obtained from watershed from the previous Section 3.3.2, called study area A4.

---

**Algorithm** (1\*\*) Scheme for 3D cell phenotype computation.

---

**Require:** (1) Raw label images (three channels correspond to three markers): alpha, beta, delta. (2) Nuclei Segmentation Result. (3) inner distance  $d_1$ , outer distance  $d_2$ .

**Ensure:** (1) Cell type (1-delta, 2-beta, 3-alpha) for each cell. (2) ROI and cell type coordinates

```

1: function NORMALIZATION( $a$ ) ▷ normalize different channels in the range of [0,255]
2:   normalize( $I_\alpha$ )
3:   normalize( $I_\beta$ )
4:   normalize( $I_\delta$ )

```

```

1: function SLIC CLUSTERING( $a$ ) ▷ find clusters with criteria of intensity similarity
   and spatial proximity
2:   Clusters = SLICClustering( $I_\alpha, I_\beta, I_\delta$ ) ▷ return a list of supervoxels
3:   SLICMedianImage = SLICMedianValue(Clusters,  $I_\alpha, I_\beta, I_\delta$ ) ▷ return label 1, 2,
   3 of cell phenotype or label 0 of unlabeled for each voxel

```

```

1: function CELL TYPE DETECTION( $a$ ) ▷ Compute the strongest marker at each cell
   profile and decide the cell type
2:   CellZone = watershed(nucleiSeed, radiusExtend  $R_{cell}$ ) ▷ cell zone detection -
   watershed separation
3:   InnerNucleiZone = erode(nuclei, innerDistance  $d_1$ )
4:   OuterNucleiZone = dilate(nuclei, outerDistance  $d_2$ )
5:   IntersectInner = intersect(CellZone, InNucleiZone)
6:   IntersectOuter = intersect(CellZone, OuterNucleiZone)
7:   if method = S1 then ▷ observe region inside and outside of nuclei
8:     observedRegion = union(IntersectInner, IntersectOuter)
9:   else if method = S2 then ▷ observe only the region outside
10:    observedRegion = IntersectOuter
11:  else if method = S3 then ▷ observe only the region inside
12:    observedRegion = IntersectInner
13:  else ▷ method = S4, observe the entire cell region
14:    observedRegion = CellZone
15:  detectCellType(observedRegion, SLICMedianImage )

```

---

### 3.4 Analysis of Spatial Organization

After acquiring the cell profile from the previous segmentation, cell zone computation, and cell phenotype computation steps, we apply various mathematical models to formalize the spatial relationship among cells and to quantify the structure of cells within tissue. Note that each cell profile contains the cell phenotype (label 1-delta, 2-beta, 3-alpha, 0-unlabeled), the coordinates of all voxels in the nuclei region, and the coordinates of all voxels in the cell region.

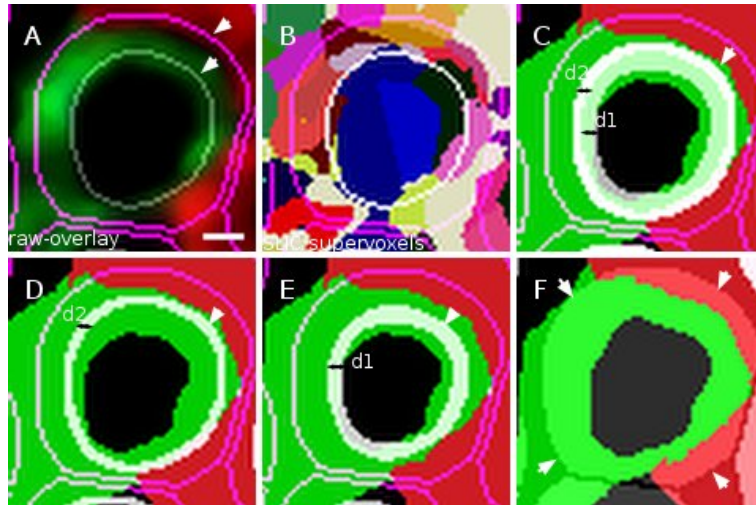


FIGURE 3.8: Different strategies to determine a study area (A1, A2, A3, A4). (A) overlay of alpha-, beta-, delta-channel and nucleus boundary, and cell zone boundary. (B) result of SLIC supervoxels, nucleus boundary, and cell zone boundary. (C) Study area A1 (Euclidean distance inside nucleus boundary d1, Euclidean distance outside nucleus boundary d2), nucleus boundary, and cell zone boundary. (D) Study area A2 (Euclidean distance outside nucleus boundary d2), nucleus boundary, and cell zone boundary. (E) Study area A3 (Euclidean distance inside nucleus boundary d1), nucleus boundary, and cell zone boundary. (F) Study area A4 (entire watershed area), nucleus boundary, and cell zone boundary. Scale  $2 \mu\text{m}$ .

### 3.4.1 Cells composition and cellular interaction

#### Endocrine cell composition

We compute the cell composition across all our mouse, monkey, and human islets and compare our results with those from other related works. The cell proportion is computed as follows:

1. The proportion of alpha cells in total number of islet cells:  $P_\alpha = n_\alpha / (n_\alpha + n_\beta + n_\delta) \times 100$ .
2. The proportion of beta cells in total number of islet cells:  $P_\beta = n_\beta / (n_\alpha + n_\beta + n_\delta) \times 100$ .
3. The proportion of delta cells in total number of islet cells:  $P_\delta = n_\delta / (n_\alpha + n_\beta + n_\delta) \times 100$ .
4. The number of *unlabelled* cells is :  $n_{unlabeled} = n_{totalCells} - n_\alpha - n_\beta - n_\delta$ .

#### Analysis of direct cell-to-cell interactions

Cellular interaction is the key to determine the function of a tissue. In our works, ‘cellular interaction’, or ‘cell-to-cell contact’, or ‘cells contact’ are used interchangeably.

Cellular interaction is defined as the number of interactions between a cell and its neighboring cells within a predefined range. For simplicity, in our implementation, two cells interact with each other when their voxels touch each other. The detail is shown in Figs.3.9

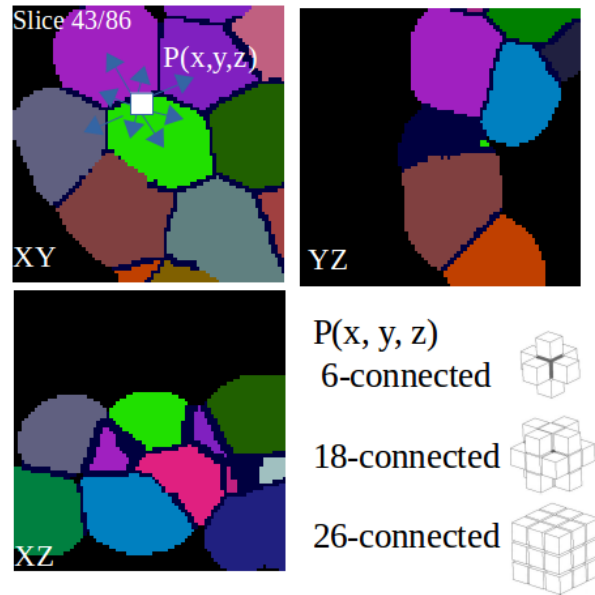


FIGURE 3.9: Cellular interaction computation with different input parameters of  $NB_{neivox}$  neighborhood voxels of the current voxel  $P(x, y, z)$  at the boundary. Three options of  $NB_{neivox}$  is 6-, 18-, 26-neighboring voxels of the current voxel  $P$ . Image result of cell zone determination-watershed separation of a mouse islet in our datasets, visualization in XY, YZ, XZ directions of slide 43 in total of 86 slices, each cell zone have one specific label.

#### *Cellular interaction computation*

The result of cell zone determination-watershed separation contains images of different cell zones. Each zone has a specific label and all voxels in the zone have the same label.

For each cell zone, we iterate over the boundary voxels and compute the labels found in  $NB_{neivox}$  neighboring voxels of the current voxel as shown in Fig. 3.9. If a neighboring voxel with a different label is found, we count this as one cellular interaction between the reference cell and the neighboring cell.

#### *3-dimensional voxel connectivity*

In three dimensional space,  $NB_{neivox}$  is commonly defined as the group of 6, 8, or 26 neighboring voxels (Fig. 3.9).

1. *6-connected.* 6 neighboring voxels of the observed voxel P  $(x, y, z)$  that are connected along one of the primary axes. It includes voxels with coordinates:  $(x \pm 1, y, z)$ ,  $(x, y \pm 1, z)$ ,  $(x, y, z \pm 1)$ .
2. *18-connected.* 18 neighboring voxels of the observed voxel P  $(x, y, z)$  that are connected along either one or two of the primary axes. It includes all voxels in 6-connected and voxels with coordinates:  $(x \pm 1, y \pm 1, z)$ ,  $(x \pm 1, y \mp 1, z)$ ,  $(x \pm 1, y, z \pm 1)$ ,  $(x \pm 1, y, z \mp 1)$ ,  $(x, y \pm 1, z \pm 1)$ ,  $(x, y \pm 1, z \mp 1)$ .
3. *26-connected.* 26 neighboring voxels of the observed voxel P  $(x, y, z)$  that are connected along either one, two, or three of the primary axes. It includes all voxels in 18-connected and voxels with coordinates:  $(x \pm 1, y \pm 1, z \pm 1)$ ,  $(x \pm 1, y \pm 1, z \mp 1)$ ,  $(x \pm 1, y \mp 1, z \pm 1)$ ,  $(x \mp 1, y \pm 1, z \pm 1)$ .

We use 26-connected voxels in our works.

#### *Cell network construction*

We model the network of cellular interactions as a graph in which the center of nuclear regions are the graph vertices. If two cells interact with each other, we create an edge between them. The cell network provides the information about neighboring cells and the spatial position of each cell in the network. In addition, the network also provides an easy way to compute the number of direct and indirect interactions among cells. It serves as the input for many graph theory algorithm in this network. For example, we can use Dijkstra's shortest path algorithm [56] to compute the cell distance.

#### *Cellular interaction between different cell types*

We apply the following computational procedure to compute the number of cellular interactions between different cell types (Fig. 3.10).

1. For each cell such as an alpha cell, we have a list of its neighboring cells. We then check the cell type of its neighbors. If the neighboring cell is a beta cell, we increase the number of interactions between alpha and beta cell,  $n_{\alpha\beta} + +$ .
2. Repeating the first step to all cells in the tissue.

The formulas to compute the cellular interaction between different cell types is as follows:

1. Total number of cellular interaction  $L = n_{\alpha\alpha} + n_{\beta\beta} + n_{\delta\delta} + n_{\alpha\beta} + n_{\alpha\delta} + n_{\beta\delta}$

2. The percentage of alpha- to alpha-cell interactions is  $P_{\alpha\alpha} = (n_{\alpha\alpha} \div L$
3. The percentage of alpha- to beta-cell interactions is  $P_{\alpha\beta} = n_{\alpha\beta} \div L$
4. The percentage of alpha- to delta-cell interactions is  $P_{\alpha\delta} = n_{\alpha\delta} \div L$
5. The percentage of beta- to beta-cell interactions is  $P_{\beta\beta} = n_{\beta\beta} \div L$
6. The percentage of beta- to delta-cell interactions is  $P_{\beta\delta} = n_{\beta\delta} \div L$
7. The percentage of delta- to delta-cell interactions is  $P_{\delta\delta} = n_{\delta\delta} \div L$

Using the cell composition and cellular interactions, we can provide the following information about cell within the tissue:

1. The total number of direct interactions.
2. The maximum number of direct interactions for one cell.
3. The average and standard deviation of number of direct interactions, per cell.
4. The average and standard deviation of number of direct interactions for alpha cells, beta cells, delta cells.

Moreover, we also have homotypic interactions among cells of the same type and heterotypic interactions among different cell types:

1. Homotypic probability computes cellular interactions between cells with the same type (alpha- and alpha- cells, beta- and beta- cells, delta- and delta-cells):  $P_{homotypic} = P_{\alpha\alpha} + P_{\beta\beta} + P_{\delta\delta}$ .
2. Heterotypic probability computes cellular interactions between different cell types:  $P_{heterotypic} = P_{\alpha\beta} + P_{\beta\delta} + P_{\alpha\delta}$ .

### Theoretical probability of a random organization

We compute the percentage of direct cellular interactions that should occur in a random organization in a similar manner to Kilimnik et al. (2012) [8, 19]. If alpha-cells are organized randomly, the theoretical probability of observing an interaction between two alpha-cells is determined by  $PT_{\alpha\alpha} = P_{\alpha} \cdot P_{\alpha} = P_{\alpha}^2$ , where  $(P_{\alpha})$  represents the probability of a cell to be an alpha-cell and  $P_{\alpha}^2$  is the probability that the first and also the second randomly chosen cell are alpha-cells.



Cellular Composition

$$N = n(\bullet) + n(\bullet) + n(\bullet)$$

$$P_\alpha = n(\bullet) / N \quad P_\beta = n(\bullet) / N \quad P_\delta = n(\bullet) / N$$

Cell-Cell Contact Probabilities

$$L = n_{\alpha\alpha}(\text{---}) + n_{\beta\beta}(\text{---}) + n_{\delta\delta}(\text{---}) \\ + n_{\alpha\beta}(\text{---}) + n_{\alpha\delta}(\text{---}) + n_{\beta\delta}(\text{---})$$

Observed & Simulated Random Model:

$$P_{\alpha\alpha} = n_{\alpha\alpha}(\text{---}) / L \quad P_{\beta\beta} = n_{\beta\beta}(\text{---}) / L$$

$$P_{\delta\delta} = n_{\delta\delta}(\text{---}) / L \quad P_{\alpha\beta} = n_{\alpha\beta}(\text{---}) / L$$

$$P_{\alpha\delta} = n_{\alpha\delta}(\text{---}) / L \quad P_{\beta\delta} = n_{\beta\delta}(\text{---}) / L$$

Theoretical Model:

$$PT_{\alpha\alpha} = P_\alpha \cdot P_\alpha$$

$$PT_{\beta\beta} = P_\beta \cdot P_\beta$$

$$PT_{\delta\delta} = P_\delta \cdot P_\delta$$

$$PT_{\beta\delta} = P_\beta \cdot P_\delta + P_\delta \cdot P_\beta$$

$$PT_{\alpha\beta} = P_\alpha \cdot P_\beta + P_\beta \cdot P_\alpha \quad PT_{\alpha\delta} = P_\alpha \cdot P_\delta + P_\delta \cdot P_\alpha$$

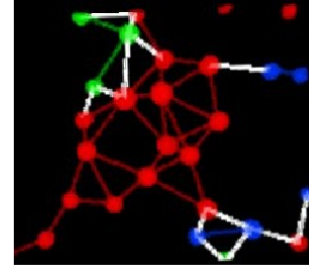
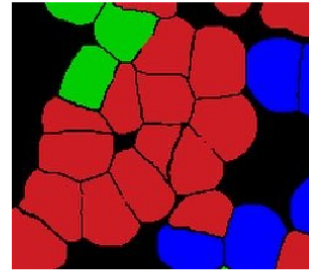


FIGURE 3.10: The formulate of computation of cell network. Using our tools visual display of 3D network is done within ImageJ using the ImageJ 3D viewer plugin [7].

Similar logic also applies to beta-cells or between alpha- and beta-cells. When two random cells are chosen in the latter case, the probability that the first cell is an alpha-cell is determined by  $(P_\alpha)$  and the probability that the second cell is a beta-cell is given by  $(P_\beta)$ .

Therefore, the theoretical probability of a random alpha- and a beta cell to interact with each other is determined by  $P_\alpha \cdot P_\beta$ . However, the possibility that the first cell is a beta-cell and the second cell is an alpha-cell must be considered as well, thus, the final equation is  $PT_{\alpha\beta} = P_\alpha \cdot P_\beta + P_\beta \cdot P_\alpha = 2 \cdot P_\alpha P_\beta$ .

1.  $PT_{\alpha\alpha} = P_\alpha \cdot P_\alpha = P_\alpha^2$ .
2.  $PT_{\beta\beta} = P_\beta \cdot P_\beta = P_\beta^2$ .
3.  $PT_{\delta\delta} = P_\delta \cdot P_\delta = P_\delta^2$ .
4.  $PT_{\alpha\beta} = P_\alpha \cdot P_\beta + P_\beta \cdot P_\alpha = 2 \cdot P_\alpha P_\beta$ .
5.  $PT_{\alpha\delta} = P_\alpha \cdot P_\delta + P_\delta \cdot P_\alpha = 2 \cdot P_\alpha P_\delta$ .
6.  $PT_{\beta\delta} = P_\beta \cdot P_\delta + P_\delta \cdot P_\beta = 2 \cdot P_\beta P_\delta$ .

*Comparing different models of cellular interaction*

We design a randomization procedure to generate cells inside a tissue and compare our randomised models and the theoretical model. The main goal is to test whether an

observed model is randomly distributed by comparing the number of cellular interactions in our observed models of mouse and monkey islets, in the simulated randomizing model, and in the theoretical model. The comparison results will be shown in the Chapter 4.

### Randomizing procedure

The randomization procedure simply shuffles the cell types and does not actually move the cells from their geographical positions. In our case study, we select the first alpha cell and then randomly choose another cell in the islet structure, for example another alpha, or a beta, or a delta-cell. If the selected cell is not an alpha-cell, we swap the two cell types. This process is repeated for all alpha-cells. Swapping the type of two cells will automatically change the number of direct interactions of the two cell types without affecting the number of interactions with a third cell type. These changes are in accordance with the theoretical values for direct interactions upon a random model.

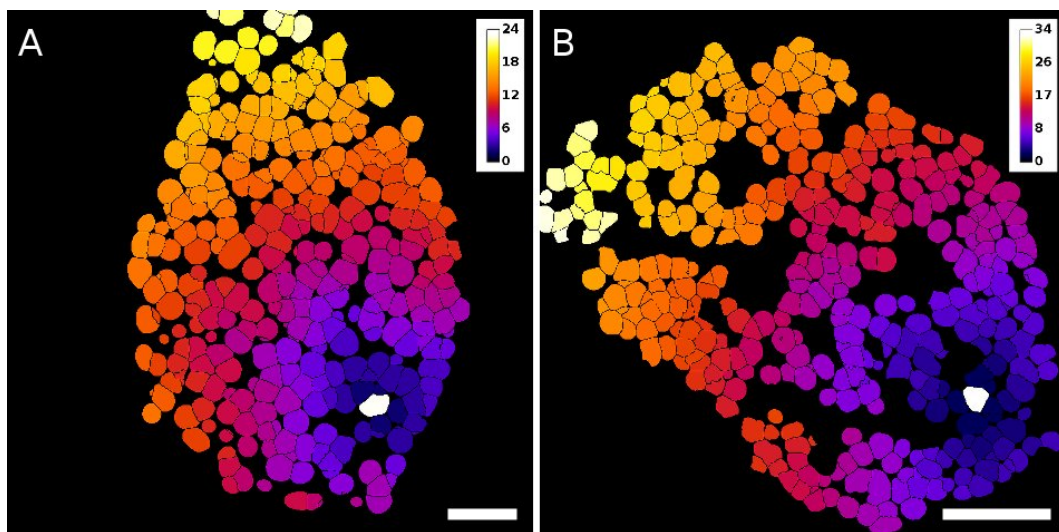


FIGURE 3.11: Result of cell distance computation in Mouse1-b and Monkey1 islet, one slice is displayed, the distance is expressed as cell distance between all cells and a reference cell (displayed in white). Reference cell has distance 0, a given islet-cell “X” contacts directly a reference cell, “X” (dark purple cells) will be at distance equal to 1 from the reference cell. If another islet-cell “Y” contacts islet cell “X”, “Y” will then be at a distance value of 2 from the “reference” cell, and so on. Scale bar, 50  $\mu\text{m}$ .

### Indirect cell-to-cell interaction computation

After analyzing the direct interaction among cells, we will continue with those indirect cellular interactions. An indirect cell-to-cell interaction is calculated using the following logic:

- When a given islet-cell ‘X’ contacts a predetermined ‘reference’ cell indirectly, ‘X’ will be at a distance equal to 1 from the ‘reference’ cell as shown in Fig. 3.11,

white cells.

- If another islet-cell ‘Y’ contacts islet cell ‘X’, ‘Y’ will then be at a distance with the value of 2 from the ‘reference’ cell as shown in Fig. 3.11, dark purple cells.

This mechanism allows the computation of the relative distance map for any cell or group of cells within our datasets. The result of indirect cellular interactions is shown in Chapter 4.

### 3.4.2 Analysis of spatial organization at global scale

#### Spatial statistics - Analysis at global scale

To start analyzing the cell spatial organization, we will first need to perform it at the global scale or, in another word, at the population level. We base our methods from the works of Andrey et al. (2010) [4, 57] and Weston et al. (2012) [5]. They have been implemented in the 2D/3D spatial statistics ImageJ plugin. We adapt this plugin and make a specific version dedicated for tissue analysis.

As a precondition of our method, we assume that the cell segmentation and cellular interactions result includes two things:

1. 3D spatial coordinates of voxels at the nucleus region and the cell zone.
2. Information about cellular interactions and the cell network.

To understand the spatial organization of cells within tissue, we compute the cumulative distribution function F-function, G-function, and Spatial Distribution Index (SDI) for all our islets (Chapter 2).

Normally, the study area  $A$  includes all points in the entire space. However, in the tissue application, cells have volume constraints inside the whole tissue and can only take predefined locations. Let  $E = \{e_1, e_2, \dots, e_n\}$  be the list of center coordinates of all nuclei in the tissue. We consider these to be the theoretical center of cells. So the study area  $A$  is the list of all centers  $E$ .

*Observed cells (event) in the study area.* We choose alpha-cells as examples of the observed cells in the study area  $AL = \{al_1, al_2, \dots, al_n\}$ .

*Distance function.* We use two kind of distance function: the standard Euclidean distance function, and cell distance function. The cell distance function is computed based on the structure of the cell network. The distance between any two cells is the shortest distance in the graph between two vertexes representing them.

*F-function computation.* The F-function is computed as follows:

1. Randomly select  $N_{ref}$  reference points  $(rf_1, rf_2, \dots, rf_N)$  in the study area  $A$ . Here we choose  $N_{ref} = 1000$ .
2. Compute the minimum distance from reference point  $rf_i$  to any observed point  $al_i$  in the point pattern:  $d_{min}(rf_i, al_i)$ ,  $al_i \in AL$ .
3. Calculate the number of pairs where the distance is smaller than or equal the distance  $d$ :  $F(d) = \frac{Nb_{d_{min}(rf_i, al_i) \leq d}}{N_{ref}}$ .

*Spatial Distribution Index(SDI) of F-function*

After getting the F-function, we generate  $S = 100$  completely random organizations and compute the cumulated distribution function (cdf) for all of them. We then rank all  $S$  simulations based on the results of their cdf. We mark the confidence envelopes of 5<sup>th</sup> and 95<sup>th</sup> of  $F(r)$ . Finally, we plot these as the 95% confidence intervals between 5<sup>th</sup> and 95<sup>th</sup> values and compare our observed organizations with the range of the two envelopes.

Visually, if the F-function curve of the observed pattern is within the two envelopes of random organizations, our observed alpha-cell pattern is random. Otherwise, we have clustered pattern if the curve is below the range and regular pattern if it is above.

In terms of SDI index, we have  $SDI - F \in [0.95, 1]$  - cluster distribution,  $SDI - F \in [0.05, 0.95]$  - random pattern, and  $SDI - F \in [0, 0.05]$  - regular pattern.

The same process is used to compute G-cumulative distribution function, and the SDI index of G-function. The results are shown in Chapter 4.

### 3.4.3 Cluster analysis

The cells network can be clustered further into a number of different groups of closely related cells or referred to clusters. Each cluster  $C$  contains a set of cells  $c_1, c_2, \dots, c_n$  where each cell  $c_i$  interact with at least one another cell  $c_j$  ( $c_j \in C$ ) of this set. Taken

the islet of Langerhans as an example, we have clusters of beta cells surrounding by alpha and delta cells. Studying cell clusters helps to understand the variability of the cell structure. Moreover, it allows the analysis of intra-cellular interactions within the cluster and between neighboring clusters.

- The cellular interaction between clusters forms the complex paracrine islet cell signaling network.
- The number of alpha-, beta-, delta- cell clutters differs across species.

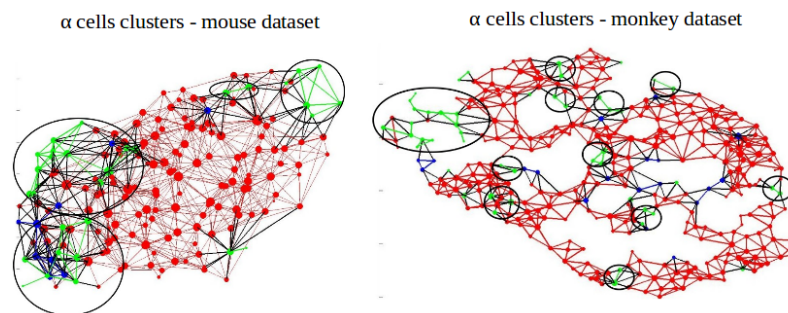


FIGURE 3.12: Clusters of alpha cells in the example of mouse4 and monkey1 dataset (Table 4.1) (each cluster is marked by a circle)

### Methods to study clustered data

We apply methods from [58] in our study as follows:

1. At the cluster level, we calculate the number of individual cells in each cluster, their spatial position and their cellular interactions.
2. At the global level, we reduce data to independent observations in a two-step approach. The first step is to reduce multiple cells in a cluster to a single observation, which is commonly the mean of the center spatial coordinates of cells in each cluster. The resulting data points (representative-cells) are all from different clusters and regarded as independent cells. This data can be analyzed in the second step using standard methods for independent observations, such as t-test, cumulative distribution function (F-, G- functions), and SDI index. The main advantage of this approach is to remove the correlation associated with clustered data and does not take into account the information within the cluster. However, the drawback is that it considers only one representative of each cluster, such as the cluster center point, regardless of the number of cells in the cluster. Cluster which contain more cells than others should have more informations and be given more weight in the analysis.

## Implementation

We choose alpha cell cluster in our study because alpha cells are located in multiple clusters and these alpha-cell clusters have direct contact with their surrounding beta cells. Our implementation is as follows:

1. We first compute a set of several characteristic for each cluster in our mouse and monkey datasets. The set includes the number of clusters, the average number of cells in the cluster, and the maximum number of cells in the cluster.
2. The global context is taken into account next. With the cell network showing the interaction between cells, we simply remove from it all cells comprising the cluster and replace them by one single node which is the center of the cluster. We run our spatial analysis procedure on the modified network and apply statistical functions similar to the previous section.

## 3.5 Delta Cell Analysis and Modeling

### 3.5.1 Analysis of spatial organization of delta-cells

Inspire of its relatively minor population, delta cells play a key role in intra-islet cell signaling and the producing of somatostatin. In mouse, monkey, and humans islets, somatostatin inhibits alpha and beta cell activities as well as repressing glucagon and insulin secretion [14].

Most works in this domain focus on the beta- and alpha- cells because they are the main compositions of the islet, leaving only fews works on delta cells [8] [14]. It leaves rooms for researchers to fill in this gap.

### Methods

To better understand the function of delta cells, we investigate delta cell structures in rodent and human islets. We use our framework to analyze the tissue organization of delta instead of alpha cells [10].

### The quantification and analysis of delta cells

Using our framework, we can choose the type of cell we want to analyze. The process is the same (i) segmentation and cell phenotype detection followed by (ii) spatial organization analysis. The result will be discussed in Chapter 4.

#### 3.5.2 Modeling and simulation of dynamic morphology of delta cell

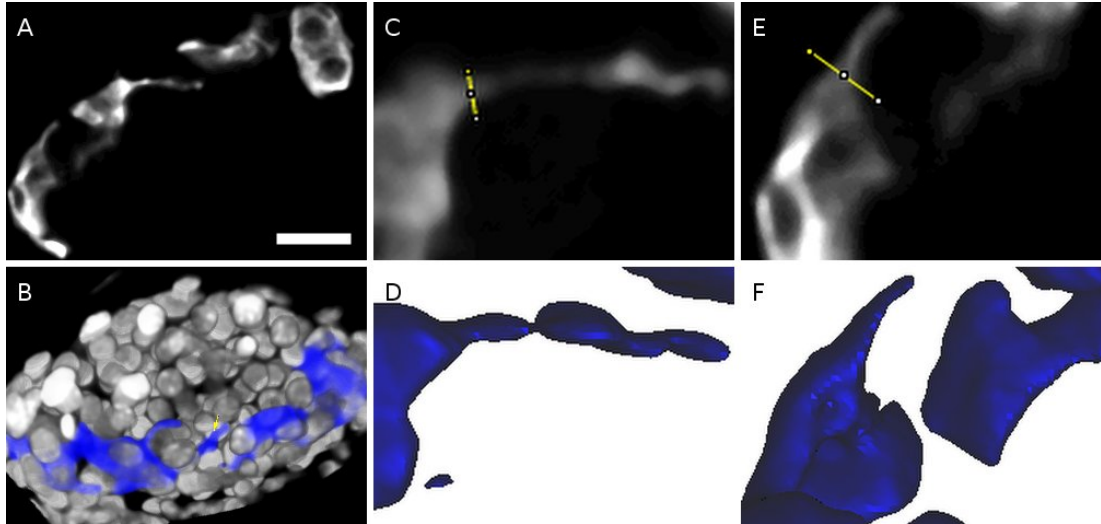


FIGURE 3.13: Delta cell excess to modulate beta cell function, example of the clear mouse4 islet. (A) somatostatin marker, (B) somatostatin with the tendency to contact beta cells at the center region of tissue (C-D) and (E-F) zoom in the label of delta cell that look like a filopodia-like structure (body and an arm) in a slide and 3D visualization. Scale bar, 20  $\mu\text{m}$ .

#### Observations

Observing the confocal microscopic images of rodent and human delta cells reveals that delta cells are elongated and, in most cases, have a well-defined filopodia-like structure as shown in Fig. 3.13C-F. We hypothesize that this filopodia-like structure enhances the range of somatostatin action in the intra islet. So, we first look at the length of delta cell filopodia structures in rodent and human pancreas sections. The result suggests that delta cells can reach out and contact numerous other cells in the islet. We use our framework [10] to identify delta cells and verify the physical advantage provided by the filopodia in a crowded islet environment. We also develop a modeling and simulation procedure for delta cells.

#### Modeling and simulation of dynamic morphology of delta cell

When a delta cell reaches out and contacts other islet cells, the extended part is called its ‘arm’ or the delta cell ‘filopodia’ (Fig. 3.13). To study this ‘arm’, we model it using

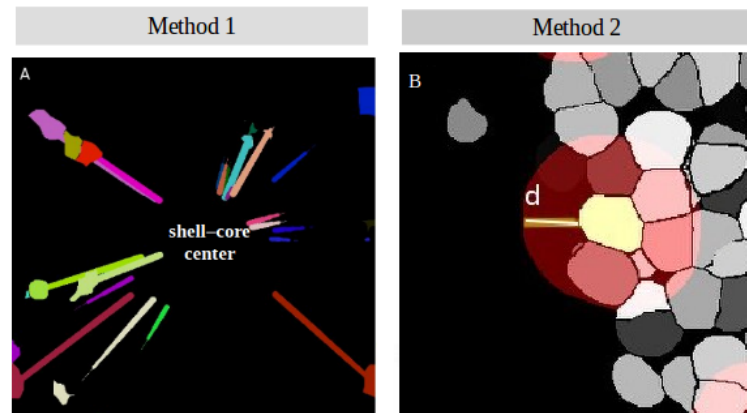


FIGURE 3.14: Delta cell and modeling of physical arm by two different methods: (1) using an ellipse and a vector, (2) using an affected range

the information about its size, volume, and direction. In general, there are two methods to achieve this:

*Method 1: Using an ellipse and a vector*

Delta- and alpha- cells have the tendency to contact with beta cells at the core of the islet (Fig. 2.2). As a result, the arm is most likely oriented towards the core (Fig. 3.13).

In our work, we define the core as the area where cells are located inside its cluster. It is at the center of the islet and potentially have a big number of cellular interactions. To determine the core area, we first compute the number of cellular interactions for each cell. We then choose the cells with a big number of cellular interactions, for example in contact with more than 4 neighboring cells. These cells should be located in a big cluster and we know the spatial coordinates of all the voxels there. After that, we compute the mean of all these cell centers and use this point as the center of core area (Fig. 3.14).

We use an ellipse to model the physical arm of a delta cell. The diameter  $d$  of this ellipse is the length of the arm (Fig. 3.14). Note that this length is counted from the cell zone boundary of the delta cell.

The direction of the arm is a vector  $V$  from the center of the delta cell region to the center of the core area.

*Method 2: Using an affected range*

In case delta cells reach out in all directions, we define the dynamic morphology of the delta cell by an affected range with radius  $r$ . The affected range is computed by dilating



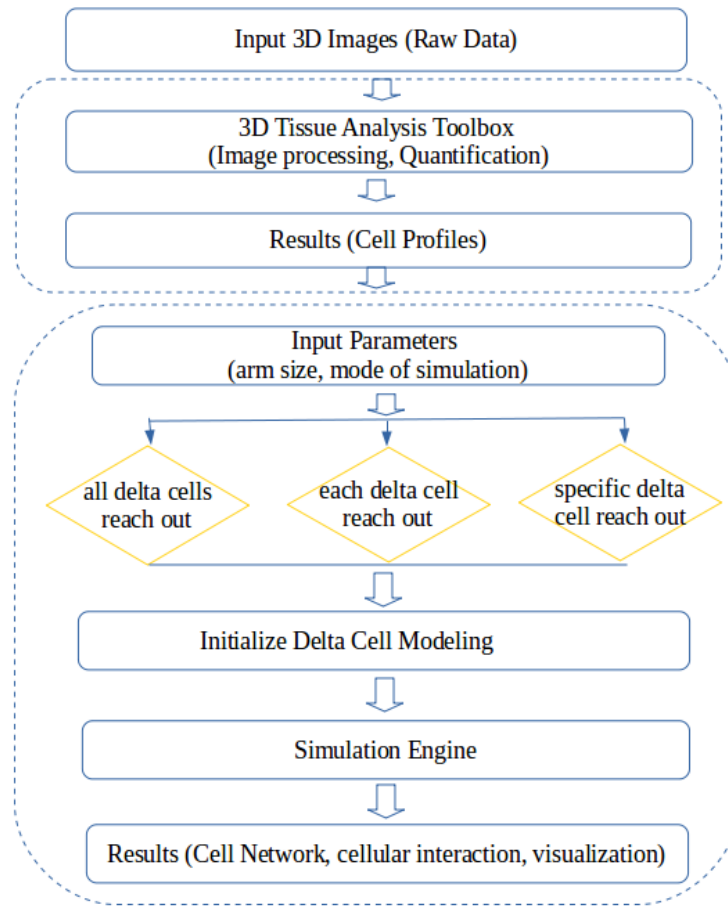


FIGURE 3.15: The diagram of delta cell protrusion simulation. The output of tissue quantification using our toolbox TITSO is used as the input for delta cell modeling and analysis. There are three options for establish a simulation: (i) all delta cells reach out to modulate the beta cells function at the same time, (ii) each delta cell reach out to modulate the beta cells function (with  $n$  delta cells - program produce  $n$  simulations), (iii) a specific delta cell reach out to modulate the beta cells function.

the delta cell zone with a predefined radius  $r$ , starting from the boundary of the delta cell zone.

We enhance the 3D dilation function of 3D mathematical morphology from the mcib3d-suite plugin developed by Boudier's group. The implementation of the dilate operator in 3D is based on the computation of Euclidean distance maps. More details are covered in Section 2.2.5.

#### *Calculate the number of cells in contact with delta cells*

A delta cell contains the body part, the delta cell zone obtained from our framework, and the physical arm obtained from modeling procedure. Thus, the total number of beta cells in contact with a delta cell includes the number of neighboring beta cells and

the number of beta cells that intersect with the delta cell arm. Again, we will come back with the result in Chapter 4.

### 3.6 Modeling of Spatial Organization

As presented in Chapter 2, modeling spatial organization helps to build virtual structures, visualize biological hypotheses, and capture the variance of cell organization. In addition, modeling processes can also generate sample data for statistic test.

Noted that here we try to generate a static model which contain a set of biological components and follow the predefined spatial rules. We can extend this work by taking into account the time notion and build a dynamic model. The modeling processes are the following:

1. Construct the cell model.
2. Define the study area.
3. Distribute cells into the study area by applying a modeling procedure.
4. Apply statistic tests to prove that the generated pattern follows a certain distribution.
5. Visualization and spatial analysis.

#### Constructing the cell model

There are three strategies to construct a virtual model of cell:

1. Use a cell-like shape to represent the cell shape in 3D such as sphere, or a more ellipsoidal shape (Weston et al. (2012)[5]).
2. Use cell profiles obtained from segmentation as the material for simulation. In our case, we can use the output of tissue quantification for this purpose.
3. Use training data to generate the cell shape. This strategy can generate a cell model close to actual cells. However, preparing the training data is a very time consuming and tedious process (Murphy et al. (2016) [59]).

In this thesis, we have been able to test the first two. The third strategy is planned for some future works.

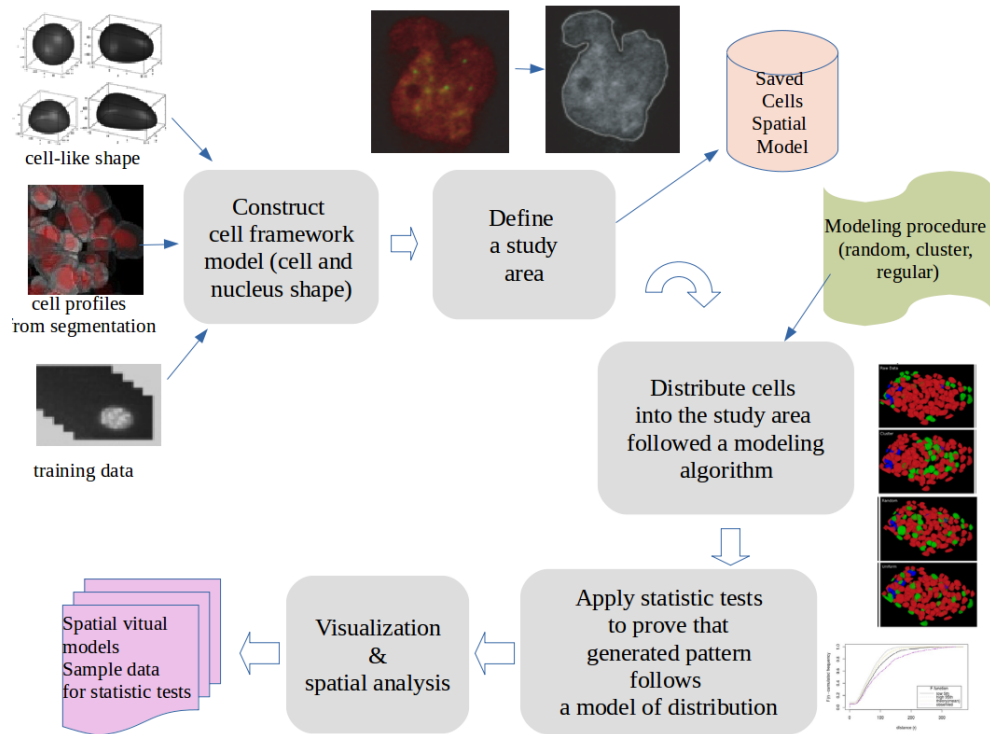


FIGURE 3.16: Workflow of modeling of the spatial organization. (i) Construct the cell model (ii) Define the study area (iii) Distribute cells into the study area by applying a modeling procedure (iv) Apply statistic tests to prove that the generated pattern follows a certain distribution (v) Visualization and spatial analysis.

### Defining the study area

The study area  $P$  is the region where the biological components are distributed. For example, the study area  $P$  of nucleolus is the entire zone inside of nucleus, the study area  $P$  of vesicles is the zone inside of the cell zone but outside of nucleus zone.

A simple way to define a study area is to use a sphere or an ellipsoidal shape. For the tissue application, the study area is the entire tissue area or the spatial coordinates of cells within tissue.

### Modeling methods

Our methods to model the spatial organization of biological components at local scale and at global scale are as follows:

1. At the local scale, we base on the work of Randell et al. (2013) [3], which proposes the use of a set of spatial relations based on discrete mereotopology. It can describe the topology and spatial organization of many objects in the image. In our works, we apply the RCC8 relation approach to describe the relationship of objects (Section 2.3.1).

2. At the global scale, we base on the work of Andrey et al. (2010) [4], and Weston et al. (2012) [5].

### 3.6.1 Modeling of the spatial organization at the local scale

Our main objective is to generate a model of spatial organization that contains  $n$  biological components  $S_1, S_2, \dots, S_n$  inside the study area  $P$ .  $S_i$  is a non-tangential proper-part of  $P$  - NTPP relation or  $S_i$  is a tangential proper part of  $P$  - TPP relation. The relation between  $S_i$  and  $S_j$   $i, j \in \{1, 2, \dots, n\}$  is that  $S_i$  is externally connected with  $S_j$  - EC relation, or  $S_i$  is disconnected from  $S_j$  - DC relation (Section 2.3.1).

The detail relationship among objects and between objects and the study area are described in Table 3.17.

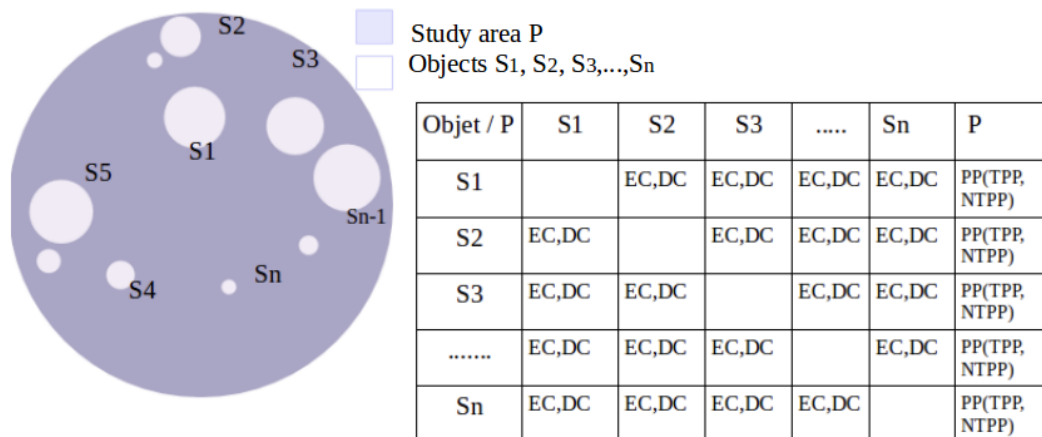


FIGURE 3.17: The spatial relationship between the objects  $S_1, S_2, \dots, S_n$  and between these objects with the domain research  $P$  are represented in the table of relationship. Our goal is to generate randomly many objects in the domain research  $P$  which satisfy the list of relationship rules.

#### Relation TPPi

When the generated object  $S_i$  is a tangential proper part inverse of the study area  $P$ ,  $S_i$  is said to have TPPi relationship with  $P$ . We use the simple strategy of applying a cell-like shape such as a sphere to represent a cell shape in 3D. The study area  $P$  is also represented as a big sphere with radius  $R$  and the center coordinate  $CP(x, y, z)$  while  $S_i$  is a sphere with the predefined radius  $r_i$ , the center coordinate  $Ci(x, y, z)$ . To generate the sphere  $S_i$ , we need to find the possible zone to distribute the center coordinate  $Ci$

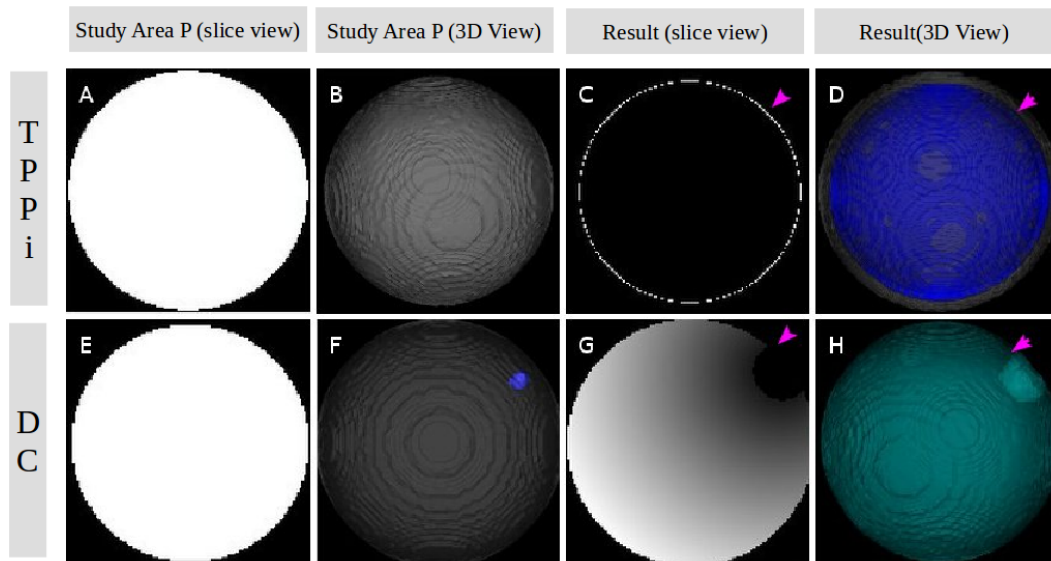


FIGURE 3.18: Computation of possible zone for the center coordinate  $Ci(x, y, z)$  of a generated object  $S_i$  in the study area  $P$ , arrow: the possible zone in one slice, and in 3D.

so that it satisfies the TPPi relationship between  $S_i$  and  $P$ . Our algorithm to determine these possible zones is as follows:

1. First, we need to compute the Euclidean distance map measured from the boundary to the center of  $P$ .
2. We then compute the possible zone whose distance is greater than  $r_i$ , denoted as  $D_1$ .
3. We randomly choose a voxel coordinate of  $Ci$  in the possible zone  $D_1$  and create the sphere  $S_i$  (Fig. 3.18).
4. The illustration of this procedure can be seen in the Figure 3.19.

#### Relation DC

When the generated object  $S_j$  is disconnected from object  $S_i$ ,  $S_j$  is said to have DC relationship with  $S_i$ . We again use the simple strategy of applying a cell-like shape such as a sphere to represent a cell shape in 3D. The study area  $P$  is also represented as a big sphere with radius  $R$  and center coordinate  $CP(x, y, z)$  while  $S_j$  is a sphere with the predefined radius  $r_j$ , the center coordinate  $Cj(x, y, z)$ . To generate the sphere  $S_j$ , we need to find the possible zone to distribute the center coordinate  $Cj$  so that it satisfies the two relationships between  $S_j$  and  $S_i$  and between  $S_j$  and  $P$  (Fig. 3.18). Our algorithm to determine these possible zones is as follows:

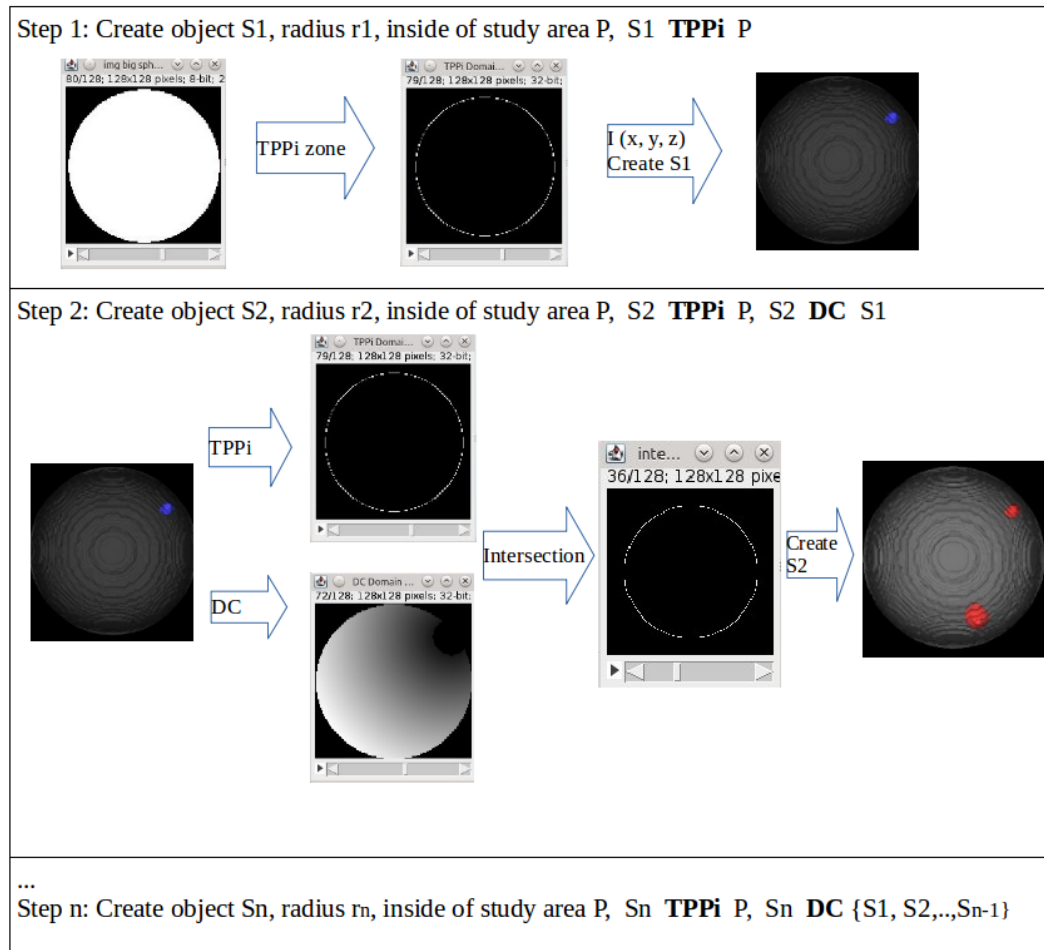


FIGURE 3.19: Steps to modeling of spatial organization at local scale.

1. First, we need to compute the Euclidean distance map measured from the boundary to the center of  $P$ .
2. We then compute the possible zone whose distance is greater than  $r_j$ , denoted as  $D_1$ .
3. Next, we compute the Euclidean distance map of the zone which is outside of the object  $S_i$  but is inside the study area  $P$ , denoted as  $D_2$ .
4. We intersect  $D_1$  and  $D_2$  to get the possible zone  $E = D_1 \cap D_2$ .
5. Finally, we randomly choose a voxel coordinate of  $C_j$  in the possible zone  $E$  and create the sphere  $S_j$ .
6. The demonstration of this procedure can be seen in the Fig. 3.19.

### 3.6.2 Modeling of the spatial organization at global scale

Section 2.3.2 shows different types of distribution such as completely random pattern, regular pattern, and cluster pattern. In this part, we introduce our procedure to generate these models. The key idea is to use Euclidean distance map to quantify the distance in the study area.

Take the regular pattern as an example, a regular pattern is a pattern where every observed point in the study area is as far from all of its neighbors as possible. We would like to generate a regular pattern that contains  $n$  biological components  $S_1, S_2, \dots, S_n$  inside the study area  $P$ .  $S_i$  is a non-tangential proper-part of  $P$  - NTPP relation or  $S_i$  is a tangential proper part of  $P$  - TPP relation. The relation between  $S_i$  and  $S_j$ ,  $i, j \in \{1, 2, \dots, n\}$  is that  $S_i$  is externally connected with  $S_j$  - EC relation or  $S_i$  is disconnected from  $S_j$  - DC relation (Section 2.3.1). At the global scale,  $S_i$  is as far from all of its neighbors  $\{S_1, S_2, \dots, S_n\}$  as possible.

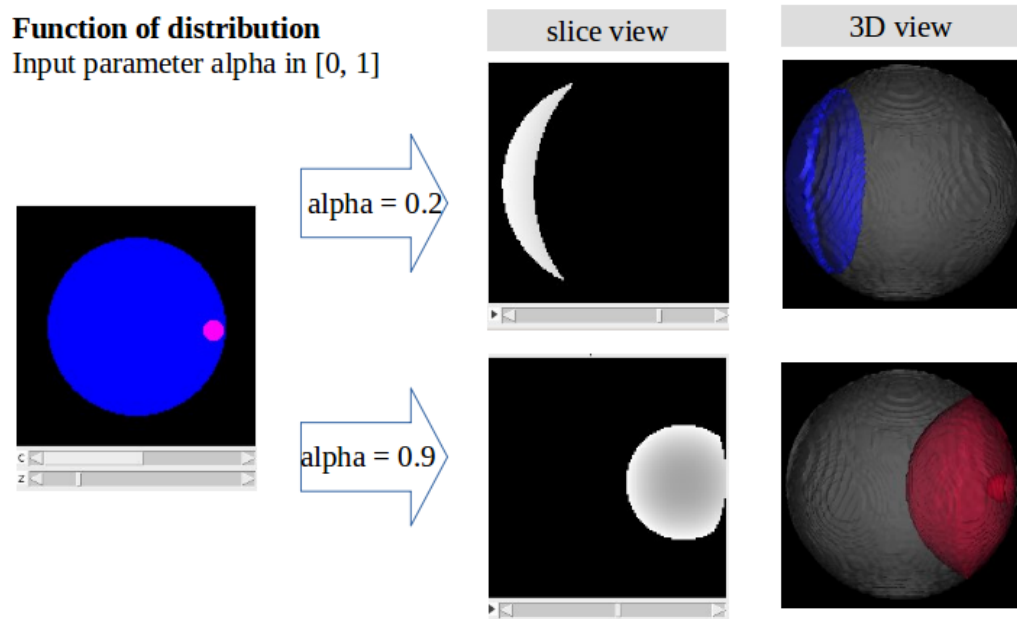


FIGURE 3.20: Distribution function using Euclidean distance map with input parameter  $\alpha$  in  $[0, 1]$ . Results shown the possible zone obtained from distribution function with  $\alpha = 0.9$  - region is close to the existing object, and  $\alpha = 0.2$  - region is far from the existing object.

#### *Distribution function*

The distribution function uses the distance map with a parameter of distribution  $\alpha \in [0, 1]$ . Let  $SK = \{S_1, S_2, \dots, S_k\}$  be the list of existing objects, we need to find the

possible zone to locate the position of object  $S_i$  so that it satisfies the relationship at local scale and at global scale. We compute the distribution function as follows:

1. First, we compute the Euclidean distance map measured from the boundary of existing objects and inside of the study area of  $P \cap SK$ .
2. We then normalize the distance map in the range of  $[0, 1]$ . The distance map includes the distance of the region which is outside of existing objects and inside of the study area.
3. We obtain the possible zone  $D_\alpha$  based on the input parameter  $\alpha$ . If  $\alpha$  is close to 1, the voxel in the possible zone is far from all existing neighbors  $SK = \{S_1, S_2, \dots, S_k\}$  as possible. On the other hand, if  $\alpha$  is close to 0, the voxels in the possible zone is close to all existing neighbors  $SK = \{S_1, S_2, \dots, S_k\}$ .

The result of the distribution function is shown in Fig. 3.20. Finally, we combine the

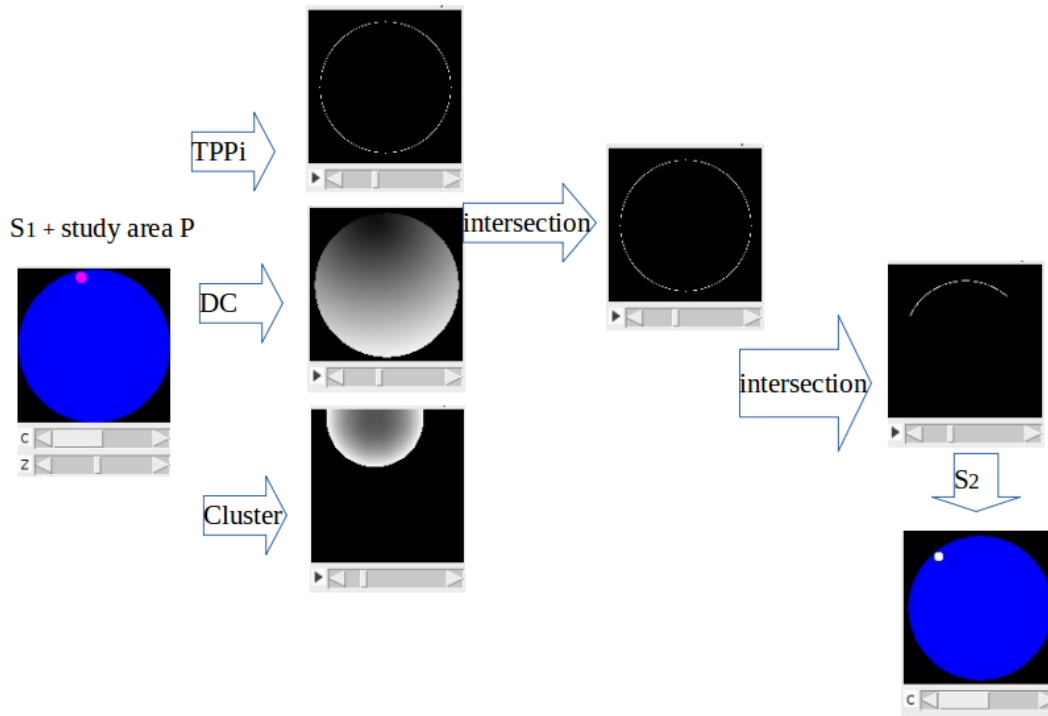


FIGURE 3.21: Algorithm to modeling the spatial organization at global scale.

two spatial relationships at the local and global scales:

1. We determine the possible zone  $E$  from modeling procedure at local scale.
2. We determine the possible zone  $D_\alpha$  from modeling procedure at global scale.



3. We intersect  $E$  and  $D_\alpha$  to get the possible zone  $D = E \cap D_\alpha$ .

The result is shown in Fig. 3.21.

## 3.7 Developing Environment

The objective is to choose a framework which can provide:

1. High throughput because of 3D thick tissue samples.
2. Support biological image processing in different formats.
3. Support 3D image analysis.
4. Easy to use across different operating systems so that it can be learned by domain experts

There are several existing platforms that we can use to develop our image analysis framework, for example, ImageJ, OpenCV, Matlab image processing toolbox, Python image processing toolbox. We choose ImageJ, a widely used image analysis software platform, to develop our application. ImageJ is an open source image processing program designed for multidimensional images. Written in Java, ImageJ is highly extensible with thousands of plugins and macros for performing a wide variety of tasks. It also has a strong and well-established user base. ImageJ with its focus on biomedical images has become a common platform for various applications in biology. With all these advantages in mind, ImageJ is the best choice for our purpose.

ImageJ has the following popular features:

1. Package 3DViewer provides a way to visualize image stacks in 3D.
2. Package bio-formats can handle many different types of images and save them all under '.tiff' format.
3. Other packages to support the analysis of 3D microscopic images, for example, mcib3d-suite plugin [37, 54].

The mcib3d-suite plugin of Boudier's group that we use to implement our framework consists of:

1. 3D Filters (mean, median, max, min, tophat, max local) and edge and symmetry filter.
2. 3D Segmentation (iterative thresholding, spots segmentation, watershed).
3. 3D Mathematical Morphology tools (fill holes, binary closing, distance map).
4. 3D RoiManager (3D display and analysis of 3D objects).
5. 3D Analysis (geometrical measurements, mesh measurements, convex hull).
6. 3D MereoTopology (relationship between objects)
7. 3D Tools (drawing ellipsoids and lines, cropping)
8. A 2D/3D spatial statistics plugin.

# Chapter 4

## Results

In the first part of this chapter, we present a detailed description of the biological datasets with various acquisition conditions which are used to test our methodologies. Then, we introduce our developed toolbox to analyze tissue spatial organization. Next, we describe our 3D segmentation algorithm and image processing process in details, followed by several comparisons between our segmentation methods and other existing frameworks for 3D cell segmentation. We then analyze and model spatial organization of cells within tissue using the segmentation outputs from the previous step. In particular, we choose to apply our set of methodologies on the dataset of islet of Langerhans. Finally, we extend our toolbox to explore and analyze the structural-functional mechanism of delta cells, one of three main cell types within the islet of Langerhans whose role is not yet determined.

### 4.1 Program Description

#### 4.1.1 Experimental datasets

##### **Immunohistochemistry of mouse and monkey pancreatic islets of Langerhans**

All animal procedures have been approved by the Institutional Animal Care and Use Committee (IACUC, protocol number 2013/SHS/816) of the SingHealth system. Experiments were carried out in accordance to experimental guidelines established by the SingHealth Experimental Medicine Center (SEMC), which is an AAALAC accredited

facility. Immunohistochemistry of mouse or monkey pancreas was performed as previously described [16]. Briefly, the pancreas from a 5-week-old wild type C57 mouse was dissected, rinsed in phosphate buffered solution (PBS) and fixed in 4 paraformaldehyde for 12 hours. Monkey pancreases were from animals used in other terminal studies. Following the fixation step, the pancreas was rinsed in PBS for 5 minutes and cryo-protected in a 30 sucrose solution for 12 hours. Next, using a razor blade, the pancreas was trimmed to smaller pieces (approx.  $2 \times 2mm$ ), embedded in freezing tissue medium FSC 22 (Leica Biosystems, Australia) and stored at  $-80C$ . Then, the pancreatic tissue was sectioned in  $30\mu m$  thick sections and placed on slides. Slides were air-dried for 1 hour on the bench top protected from light and rinsed 2x5 minutes in PBS. Next, slides were blocked and permeabilised with a blocking solution (10 fetal bovine serum and 0.3 triton X-100 in PBS) for 1 hour. Following the blocking step, samples were incubated overnight with the following primary antibodies: guinea pig anti-insulin (beta-cells, 1:200, Dako, California), rabbit anti-glucagon (alpha-cells, 1:200, Sigma Aldrich, USA) and rat anti-somatostatin (delta-cells, 1:200, Millipore, USA). After incubation, slides were rinsed 4x 10 minutes in PBS. Next, slides were incubated for 1 hour with the following secondary antibodies: donkey anti-guinea pig AlexaFluor 568 (1:400, Invitrogen, USA), donkey anti-rabbit AlexaFluor 488 (1:400, Invitrogen, USA) and donkey anti-rat AlexaFluor 647 (1:400, Invitrogen, USA). Cell Nuclei was stained with DAPI (1:400, Invitrogen, USA). After incubation, samples were rinsed 4x 10 minutes in PBS, dried and mounted for microscopy. Clear tissue samples were prepared according to Ke *et al* [51].

### Experimental datasets

In order to test the performance of our framework, we first use it to analyze 3D confocal image stacks of islets of Langerhans in whole or serial histological pancreas sections from two different species [10]:

- C57/BL6 mouse (8 weeks of age,  $n = 6$  animals and 22 stacks).
- Monkey (Macaca Fascicularis, 9-12 years of age,  $n = 6$  animals and 12 stacks).

More information about the islet used to validate our toolbox can be found in Table 4.1. In addition, we test the adaptation of our nuclei segmentation method on several raw nuclei samples from human datasets as shown in Fig. 4.2.1.

Species	Name	Size (XYZ)	Calibration (XY-Z) in nm
Mouse	Mouse1-a*	1024 x 1024 x 36	378.4 x 800.1
	Mouse1-b*	1024 x 1024 x 48	378.4 x 800.1
	Mouse1-c*	1024 x 1024 x 56	378.4 x 800.1
	Mouse2	1024 x 1024 x 83	179.7 x 500.2
	Mouse3	1024 x 1024 x 77	250.5 x 500.2
	Mouse4	528 x 391 x 104	209.2 x 839.8
	Mouse5	584 x 498 x 86	345.9 x 839.8
	Mouse6	468 x 443 x 136	283.9 x 839.8
Monkey	Monkey1	1024 x 1024 x 17	240.3 x 1007.0
	Monkey2	1248 x 1248 x 56	177.4 x 420.2
	Monkey3	1248 x 1248 x 68	177.4 x 420.2
	Monkey4	1024 x 1024 x 44	180.2 x 333.3
	Monkey5	1024 x 1024 x 40	198.3 x 999.9
	Monkey6	1024 x 1024 x 45	236.7 x 999.9
Human	Human1	1024 x 1024 x 49	319 x 999
	Human2	1024 x 1024 x 50	141 x 999.9
	Human3	1024 x 1024 x 49	207 x 999.9
	Human4	1024 x 1024 x 48	282 x 999.9
	Human5	1024 x 1024 x 62	263 x 999.9
	Human6	1024 x 1024 x 56	255 x 999.9

TABLE 4.1: Origin and characteristics of mouse, monkey, and human pancreas datasets used in this study. \* indicates serial histological sections.

The sample preparation and imaging protocols is shown above. Besides the thick tissue sections, we also analyze the whole islet, obtained from optically cleared tissues using SeeDB and imaged with 2-photon microscopy [51], from two mice, 4 and 6.

Finally, in order to explore the mechanism of delta cells [9], we also test our framework on the two different species: Mouse (n = 24 islets total, n = 7 mice) and human islets (n = 21 islets total, n = 3 donors).

#### 4.1.2 Developed toolbox to investigate tissue spatial organization

In this part, we introduce the TITSO (the Toolbox to Investigate Tissue Spatial Organization) that we have been developing (Fig. 4.2). The main goal of this framework is to analyze and model the spatial organization of cells within tissue in 3D confocal microscopy image. It could be used both to analyze tissue organization in general and to explore the structure and functions of the islet of Langerhans in particular. Also, in this thesis, we use the term ‘toolbox’ and ‘framework’ interchangeably because the former is more familiar with biologists.

TITSO is a framework that can perform various algorithms on 3D confocal microscopy images. It is implemented using the open source software platform ImageJ, a Java image processing program created by National Institute of Health and is well known among scientific community.

We applied TITSO toolbox on the captured images of the islet of Langerhans and achieved good results with efficient execution time, thus proving that our methods are both stable and robust. The main advantage of our toolbox is that it has the capability to analyze tissue microscopic images in three dimensions (3D).

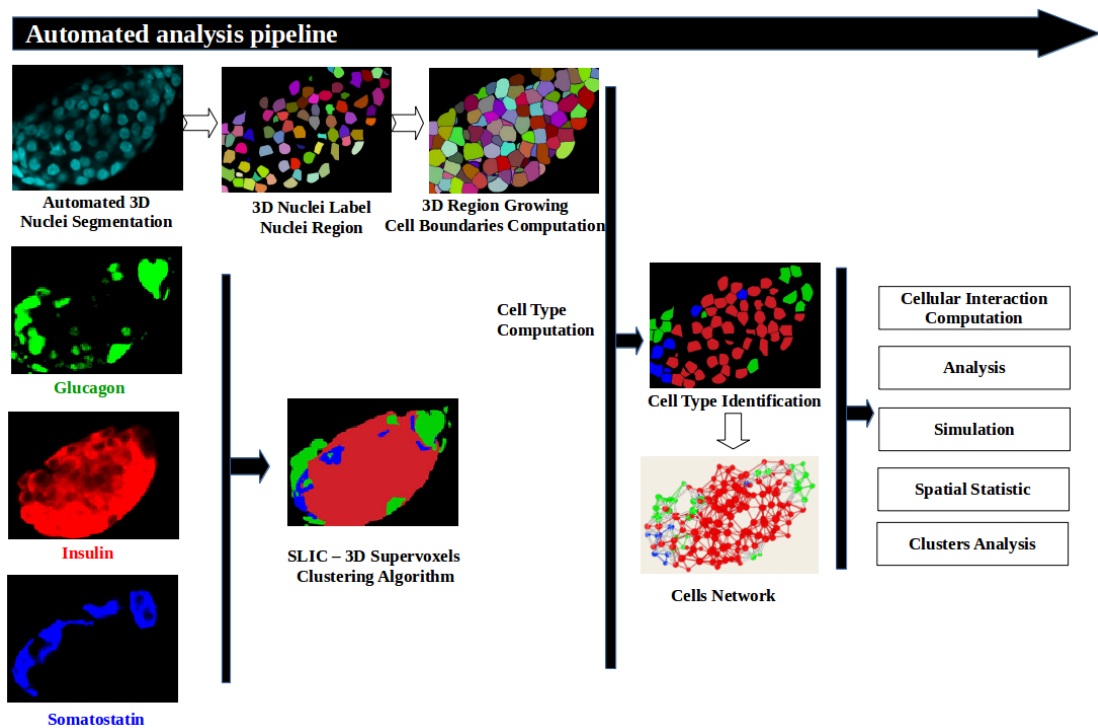


FIGURE 4.1: Workflow of the analysis. Nuclei are segmented and cell boundaries computed from the DNA-label channel. The images corresponding to the different labels of the different cell types are combined together in a SLIC image (see text for details). This combined image is used to compute each cell type. From the interactions between cells, computed using the cell boundaries information, the cells network is built.

### General workflow

The input data for TITSO consists of a 3D nuclei raw image, captured using DAPI technology with confocal microscopy, and a multi-channel set. Each channel corresponds to a fluorescent marker and associates with a different biological compartment. In the application of the islet of Langerhans, the input includes a 3D nuclei image and a three-channel set. These channels include glucagon marker - denoted alpha cells, insulin marker denoted beta cells, and somatostatin marker - denoted delta cells. The input is

encoded in ‘.tiff’ format. Other images types, which are compatible with Bio-formats plugin from Linkert et al, 2010 [60], can be opened using the same plugin under ImageJ and then saved as ‘.tiff’ images for TITSO.

Our workflow includes five main consecutive stages as shown in Fig. 4.1.

1. Nuclei are segmented and cell boundaries computed from the DNA-label channel.
2. The images corresponding to different labels of different cell types are combined together in a SLIC image.
3. This combined SLIC image and cell morphological information obtained from the first step are used to compute each cell type.
4. The cell network is then built from the interactions between cells, computed using the cell boundaries information. Using the cell morphological information and this cells network, we analyze the tissue spatial organization.
5. Finally, we model and simulate a spatial organization model.

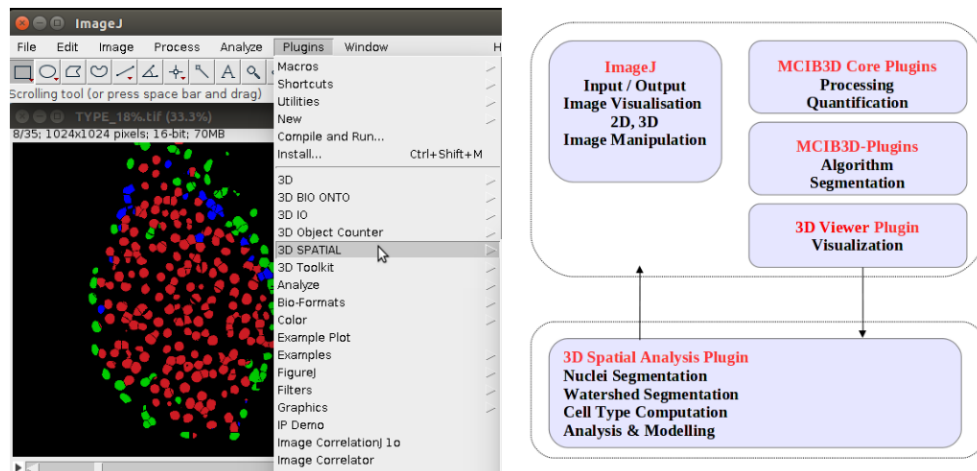


FIGURE 4.2: TITSO (Toolbox to Investigate Tissue Spatial Organization) - Plugin developed under ImageJ-image analysis platform for biomedical image. Figure show the graphical interface of TITSO under ImageJ (left side) and the dependency libraries that TITSO used.

### Availability

We implement our framework in Java as a plugin of ImageJ. The plugin has open source license and has been tested on Linux, Microsoft Windows, and MacOS X. The system specification: Intel Core i7-4800MQ, 3.07GHz, 64-bit with 8GB RAM. The source code is available on github at [https://github.com/nhuhoa/tissue\\_organization](https://github.com/nhuhoa/tissue_organization). In addition, a wiki page with instructions on how to use the toolbox together with a

sample dataset is at [http://imagejdocu.tudor.lu/doku.php?id=plugin:3d:tissue\\_organization\\_toolbox:start](http://imagejdocu.tudor.lu/doku.php?id=plugin:3d:tissue_organization_toolbox:start). The same documentation is attached at the Appendix A of this thesis.

### **Acknowledgement**

I would like to thank all people who contribute to the development of the ImageJ platform, Imagescience plugin, Convexhull3D plugin, 3dViewer plugin, and Lipid Droplet Counter plugin as I utilize parts of these libraries to develop my plugin. I would also like to thank my supervisor Dr. Thomas Boudier who provided me with mcib3d core and mcib3d plugins with many useful image processing features for three dimensional images. Finally, I wish to thank Dr. Phillippe Andrey for his statistic plugin under ImageJ which I have adapted to work with tissue application.

## **4.2 Tissue Quantification and Analysis**

The steps to investigate tissue spatial organization involve several complex image processing tasks as shown in Fig. 3.1. Due to their complexity, selected methods should be robust and have good accuracy. In our work, we have conducted a detailed study of all existing methods in the field and keep those that are most suitable for our work. So, we briefly describe how we have utilized these methods to study the spatial organization of tissue including how parameters are selected and results are achieved. This is followed by a thorough comparison of our methods with other related works in the same domain.

### **4.2.1 Nuclei Segmentation Result**

We have modified spot segmentation method to make it work with our dataset by choosing a set of parameters that could well capture the volumetric of nuclei. After experimenting and evaluating big datasets of mouse, monkey, and human, we show that spot segmentation algorithm is the most appropriate existing method for our confocal microscopy 3D input images, both in terms of accuracy and speed when comparing with other existing methods.



#### 4.2.1.1 Parameters selection

The Spot Segmentation algorithm performs segmentation of spots in 3D. The input image consists of original nuclei labelled as spots (using DAPI staining to label the cell nucleus) and an image of the seed (seeds). The local maxima of the filtered image, a Fourier domain image in our case, is used as seed for the local segmentation. This process takes into account all voxels surrounded each seed and clusters these neighboring voxels into objects (Section 3.3.1).

#### Pre-processing

We use a 3D median filter from ImageJ core package [61] with a radius of  $4 \times 4 \times 2$  in  $X \times Y \times Z$  for all datasets to reduce noise and homogenize intensities inside the nuclei. The  $z$  value is larger due to the fact that  $z$ -dimension is much thinner as shown in Table. 4.1.

#### Nuclei seed detection and comparisons

In our work, seeds are computed using the Fourier domain image instead of using the original one - grayscale image. This preprocessed nuclei image is then filtered by a 3D band pass filter [53] with a specific size interval. The performance of seed detection algorithm depends on whether the band pass filter can capture the characteristic of the image, in which frequencies between the two cut-offs remain in the output image while others are attenuated. The two cut-offs should correspond to the approximate diameter of nuclei in the image. Due to this reason, choosing the correct values is very important. In our mice dataset, we use the range  $[D_{min} - D_{max}]$  with the low value  $D_{min}$  of  $20 \pm 3$  pixels and the high value  $D_{max}$  of  $28 \pm 2$  pixels, which is approximately  $6 - 7\mu m$ . These equivalent values but with a higher degree of variance in the monkey dataset are  $D_{min}$  of  $17 \pm 3$  pixels and  $D_{max}$  of  $28 \pm 2$  pixels or  $6 - 8\mu m$ . Both the monkey and human datasets share similar values. The 3D Spot Segmentation plugin [37, 52] uses a local maxima in a radius of 4 voxels extracted from the band pass filtered image. A threshold is set to only detect local maxima corresponding to nuclei regions, which is 30,000 for both mouse and monkey datasets.

#### Comparison with different strategies of seed detection

There are existing methods which use local peaks of intensity to detect seed, however, such strategies could not work well in our case due to the high level of noise in the nuclei

image which leads to false positives. In addition, each nuclei could contain several local peaks or no peak at all.

Another strategy to the problem of seed detection is to use multi scales blob detectors based on Laplacian of Gaussian (LoG) [32] or Hessian of Gaussian [34] detectors.

In a related work, Danvek et al. (2009) [33] find peaks as seeds in the resulting image using a morphological extended maxima transformation. They also compute the seed on the transform image.

The advantage of our method is that the seed detection is computed on the transformed image instead of the original one. The transform domain is more effective than its spatial domain because it is easier to separate noise from the object in the transform domain as shown in our case. However, seed detection on transform domain still suffers from over segmentation because it is sensitive to cell texture and the occurrences of small shapes. In our work, we decide to use the band pass filter to obtain the transform image because this filter could remove both low frequencies corresponding to small objects and high frequencies corresponding to fusion of objects. Moreover, the local maxima in a radius of 4 voxels extracted from the band pass filtered image can limit the number of detected seeds. Thus, a threshold could be set to only detect the local maxima corresponding to nuclei regions, which is set to 30.000 in 16-bits local maxima image for both the mouse and monkey datasets.

In summary, the band pass filter allows us to discard any eventual intensity variations between nuclei and within the nuclei. This contributes to a higher detection efficiency observed in our dataset.

### **Impact of nuclei diameter to segmentation result**

Our nuclei segmentation method uses several parameters across all our datasets. Among these parameters, there are only two dynamical ones which are the low and high values of nuclei diameter which should be set according to the input image. Specifically, the band pass filter mentioned in the previous section could remove all objects below the predefined low value or higher than the predefined high value of the nuclei diameter.

In order to understand better the impact of nuclei diameter upon the segmentation result, we conducted some experiment on all our datasets as shown in Table 4.2. In these experiments, we fixed all others parameters and varied only the range of nuclei

Sample	Acquisition(pixel)[Dmin-Dmax]	Nb detected objects	Remark
<i>Mouse3</i>	1024 × 1024 × [14-20] 77	1277	Detect objects + splitted objects + noise
<i>Mouse3</i>	1024 × 1024 × [16-20] 77	1080	Detect objects
<i>Mouse3</i>	1024 × 1024 × [16-23] 77	1000	Detect objects + merged objects
<i>Monkey6</i>	1024 × 1024 × [20-31] 45	560	Detect objects + merged objects
<i>Monkey6</i>	1024 × 1024 × [20-28] 45	578	Detect objects
<i>Monkey6</i>	1024 × 1024 × [16-28] 45	613	Detect objects + splitted objects + noise

TABLE 4.2: Impact of nuclei diameter  $[D_{min} - D_{max}]$  to segmentation result. Testing dataset of mouse3 and monkey6 islet (Table 4.1).

diameter  $[D_{min} - D_{max}]$  before executing the Spot Segmentation algorithm. The results show that using smaller values of  $D_{min}$  we detect more objects. However, these objects contained both nuclei objects and noise (false positives). In many cases, the filter also captured the variance within nuclei and split one nuclei into two. On the other hand, when we use a higher value for  $D_{max}$ , there is a possibility that two or more objects could be fused into one, thus leading to a smaller number of detected objects.

#### 4.2.1.2 Nuclei segmentation result

The Spot Segmentation algorithm allowed us to identify and segment correctly more than 97% of the total number of nuclei in all the datasets as shown in Fig. 4.8A-C and Table 4.3). We also observed a low false-positive rate of 106 objects out of 8837 nuclei and a low false negative rate of 257 objects (Table 4.3) in our validated dataset (Table 4.3). The final F-measure is approximately 97.7%, which is comparable and even better than other state-of-the-art algorithms (Table 4.3, 4.5, Fig. 4.3). Most of the false-positive detections are due to the occasional observations of nuclei which are split into two different objects.

Dataset	True Posi- tive	False Posi- tive	False Nega- tive	Recall	Precision	F- measure
Mouse1-a*	459	2	17	96.429	99.566	97.972
Mouse1-b*	847	1	16	98.146	99.882	99.006
Mouse1-c*	915	2	29	96.928	99.782	98.334
Mouse2	568	5	20	96.599	99.127	97.847
Mouse3	993	2	19	98.123	99.799	98.954
Mouse4	194	3	20	90.654	98.477	94.404
Mouse5	568	5	22	96.271	98.594	97.594
Mouse6	494	5	19	96.296	98.997	97.268
<b>Total Mouse</b>	<b>5038</b>	<b>25</b>	<b>162</b>	<b>96.180</b>	<b>99.322</b>	<b>97.726</b>
Monkey1	695	12	9	98.722	98.303	98.512
Monkey2	754	16	21	97.29	97.922	97.605
Monkey3	605	27	21	96.645	95.728	96.184
Monkey4	561	5	18	96.891	99.117	97.991
Monkey5	580	15	9	98.47	97.479	97.971
Monkey6	604	6	17	97.262	99.016	98.131
<b>Total Monkey</b>	<b>3799</b>	<b>81</b>	<b>95</b>	<b>97.546</b>	<b>97.927</b>	<b>97.732</b>

TABLE 4.3: Nuclei segmentation results for the observed mouse and monkey datasets.  
\* indicates serial histological sections.

We also encountered a low false negative detection rate for nuclei corresponding to nuclei that were located in a cluster or below the resolution potential of the confocal microscope in the X-Y and Z-axis (Table 4.1). Finally, there was one dataset of mouse labelled ‘Mouse4’ whose accuracy was lower than others (90.6% Recall, 94.4% accuracy). The direct causes are the very high resolution of these images and the big cluster of cells as shown in Table 4.3 and Fig. 4.1).

### In-depth analysis of mis-detected cases

With the accuracy of our approach in mind, it’s worthwhile to take a closer look into all mis-detected cases to understand their natures so that we could improve the process in the future. A typical case of error is where there is a lack of strong edges between nuclei, thus, the filter only detects one big object. Similarly, two or more nuclei in close proximity could have shared a homogeneous level of intensity. The segmentation algorithm could either detect only one seed corresponding to a big object or, in some cases, two seeds with one big and one very small object. The latter happens because of the criteria in which all neighbors whose values are greater than the local threshold are examined. In contrast, when there are strong edges within nuclei, the algorithm can detect several seeds per nuclei and consequently split one nuclei into multiple objects. Finally, there is a limitation in the method in which it can only capture nuclei within the

predefined range of diameter  $[D_{min} - D_{max}]$ . In other words, all nuclei whose diameters are smaller or bigger than this range are excluded.

### Complexity

During our work, we did all the computation on a computer Intel Core i7-4800MQ, 3.07GHz, 64-bit with 8GB memory. On average, the pre-processing step took about 30 seconds while the seed detection procedure took from 1 to 2 minutes. After that, computing all thresholds took another 30 seconds to 1 minutes. Finally, the segmentation took from 1 to 2 minutes. In total, the program took approximately 3 to 6 minutes to segment one nuclei image. After getting the segmented image, we could apply a semi-automated segmentation procedure by observing the result in 3D Manager plugin [37] to correct possible errors by splitting or merging mis-detected objects. Such procedure took an addition of 1 to 5 minutes.

#### 4.2.1.3 Comparison of nuclei segmentation methods

In this part, we introduce several available softwares for 3D nuclei segmentation and followed by a comparison of the results of spot segmentation methods with these softwares.

#### Testing dataset

Our testing dataset consists of one mouse and one monkey samples as shown in Table 4.4. Both contain nucleus labeled with confocal microscopy as mentioned in Section 4.1.1. These sample images have a medium level of crowded cell density and rounded cell shapes so we consider them to be at the medium difficulty level for 3D segmentation task. Using the testing dataset, we conducted two groups of experiment, one to test the performance of segmentation methods and the other to check their accuracy. The former uses the input images as a whole while the latter uses only selected tissue areas cropped from the original images. The main reason for such setting is that some methods we want to compare against could not handle our original high resolution images.

#### Choosing segmentation methods

Existing 3D segmentation methods include 3D Iterative Thresholding [36, 37], Farsight toolkit ([32]), MINS (Modular Interactive Nuclear Segmentation) [34], 3D MLS (3D Multiple Level Sets) [35]. These methods are described in Section 2.2.4. Due to the

Testing Sample	Nb slices	Size	Resolution	Test Type
Original-mouse7	59	1248 × 1248	0.27 × 0.27 × 0.50	Complexity
Crop-mouse7	59	686 × 956	0.27 × 0.27 × 0.50	Complexity, Accuracy
Original-monkey7	45	1024 × 1024	0.23 × 0.23 × 0.99	Complexity
Crop-monkey7	45	703 × 668	0.23 × 0.23 × 0.99	Complexity, Accuracy

TABLE 4.4: Origin and characteristics of mouse7 and monkey7 nuclei labeling image using as testing data for the comparison of different nuclei segmentation methods.

amount of time required to validate the huge number of nuclei in each segmentation output, here we provide only observations and discussions about each method. MINS is written using Matlab while Farsight and 3D MLS in C++. On the other hand, our spot segmentation and 3D Iterative Thresholding are both in Java, under ImageJ platform.

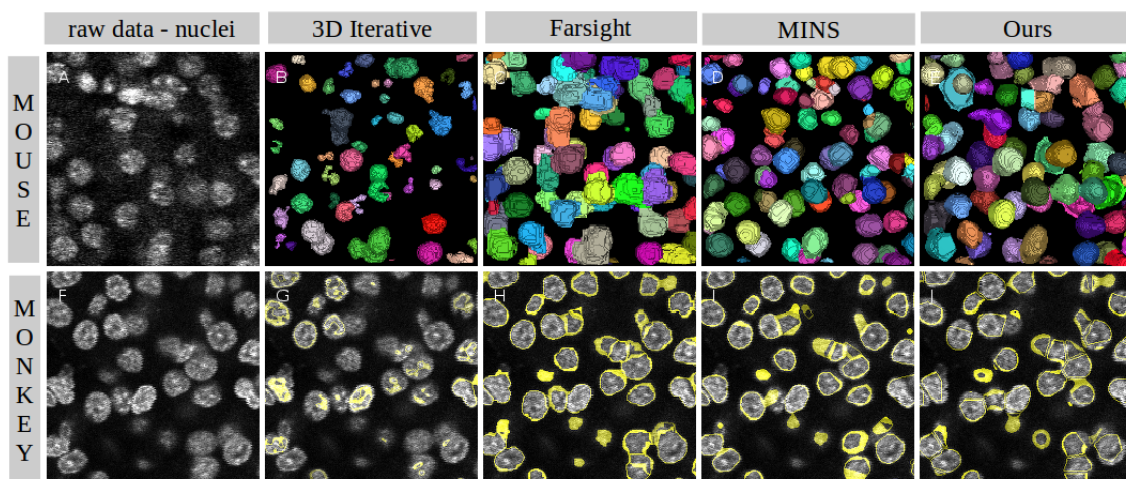


FIGURE 4.3: Comparison of 3D Nuclei Segmentation Methods. 1st column: Raw data showed one slice, 2nd: Results of 3D Iterative Thresholding method, 3rd: Segmented results of Farsight toolkit, 4th: Results of MINS platform, 5th: Results of our Spot Segmentation method. 1st row: Segmentation nuclei result of mouse sample, visualize in 3D using 3D Viewer plugin, under ImageJ platform. 2nd row: Segmentation nuclei result in monkey sample, visualize segmented contours of object using 3D ROI plugin under ImageJ platform.

With the development of new technologies, we could capture thick tissue samples in three dimensions, which, in turn, requires better and more efficient segmentation methods to analyze them. Therefore, the execution time could be considered one of the most important criteria with regard to the performance of a segmentation method. We compare the running time of our spot segmentation method with other methods on our tested server Intel Core i7-4800MQ with 8GB RAM.

The first experiment was performed on original samples of ‘mouse7’ and ‘monkey7’ islets as shown in Table 4.4. These samples contain the tissue region as well as the outside environment. MINS took 22 and 17 minutes to segment mouse and monkey nuclei images respectively. Farsight was much faster spending 4 and 3 minutes while our spot segmentation took slightly longer at 5 and 4 minutes each. 3D MLS crashed and could not perform the segmentation using our original data. Finally, 3D Iterative Thresholding took less than 3 minutes for both datasets. However, the running time of this method depends heavily on the range of image values and the number of iterations. It took order of magnitude longer at around 2 hours if there were many iterations and input images encoded in 16-bit. Therefore, this method only works with 8-bit testing samples.

In the second experiment using the selected islet areas cropped from the original images, we provided the quantitative results of 3D nuclei segmentation as shown in Table 4.5. The result sorted by the execution time is 3D Iterative Thresholding, Farsight, our spot segmentation, 3D MLS, and MINS. Because the performance of 3D Iterative depends heavily on its input parameters as mentioned above, Farsight could be considered the most stable implementation-wise with an average of 3 minutes per input. MINS took about 11 minutes for mouse image. 3D MLS required manually step-by-step optimization.

In both tests, our spot segmentation method took an acceptable amount of time to finish when comparing with others. In addition, its execution time ranged from 3 to 7 minutes per input across all our datasets with a total of 24 mouse islets, 11 monkey islets, and 21 human islets. These results prove that our spot segmentation method is stable and fast in most cases.

Although slightly slower than Farsight, the spot segmentation approach shines in term of accuracy. Both Farsight and MINS suffers from over-segmented when the intensity inside a nucleus is not homogeneous and there is less difference between the object and its background or when there are more than one local maxima inside the nucleus. Farsight and MINS also incorrectly divided a nucleus with long extensions along z-axis into several objects. In fact, MINS has the highest number of over-segmentation among all methods. Specifically, Farsight and MINS detected 705 and 847 objects in the mouse sample, much more than the more accurate number of 606 objects detected by our spot

Method	Nb detected nuclei
Mouse7 3D Iterative Thresholding	756
Mouse7 Farsight	705
Mouse7 MINS	847
Mouse7 Spot Seg	606
Monkey7 3D Iterative Thresholding	959
Monkey7 Farsight	800
Monkey7 MINS	781
Monkey7 Spot Seg	702

TABLE 4.5: The output of different methods(3D Iterative Thresholding, Farsight, MINS, our Spot Segmentation) for 3D nuclei segmentation of a nuclei labeling image of a crop-mouse7, and a nuclei labeling image of crop-monkey7 (Table 4.4). The actual number of nuclei with strong intensity level in mouse nuclei labeling image is approximately  $600 \pm 20$  nuclei, and  $700 \pm 20$  nuclei.

segmentation. To make things worse, the output of Farsight can be under-segmented when nuclei are located in a cluster.

In term of contour, MINS detected the position of nuclei correctly despite the fact that it only produced the compact segmentation and under-estimated the correct contours of the nuclei as shown in Fig. 4.3. In contrast, both Farsight and our spot segmentation slightly over-estimated the nuclei’s contours and connected some closed ones together. Finally, because the staining inside each nuclei is not homogeneous, 3D Iterative Thresholding could only detect the strong intensity regions within each nucleus as shown in Fig. 4.3.

## Conclusion

Among several existing segmentation methods, the spot segmentation algorithm [36, 37] from Boudier’s group is chosen to segment the islet of Langerhans in our work because it is relatively fast, stable, especially, proven to have equal or better accuracy than others.

## 4.2.2 Cell Zone Computation Results

### 4.2.2.1 Cell segmentation using plasma membrane labeling

In order to validate our cell zone computation method, we first tried to obtain islet cells whose plasma membranes have been labeled with F-actin. The membrane detection procedure is described in section 3.3.2.1



Fig. 4.4 shows a very poor result with many noises inside the cell area and mis-detection of the cell contours because of the low intensity at the membrane area. Such poor result inevitably leads to incorrect computation of cell boundaries and affects the performance of all consecutive stages.

The main difficulty lies in the labeling process, which introduced many gaps in the membrane staining. In addition, we could not stain the plasma membrane effectively in our study because of the limited number of possible markers that could be used in a single immuno-fluorescence experiment. We will need a different approach to solve the problem instead of relying on the difficult plasma membrane labeling process.

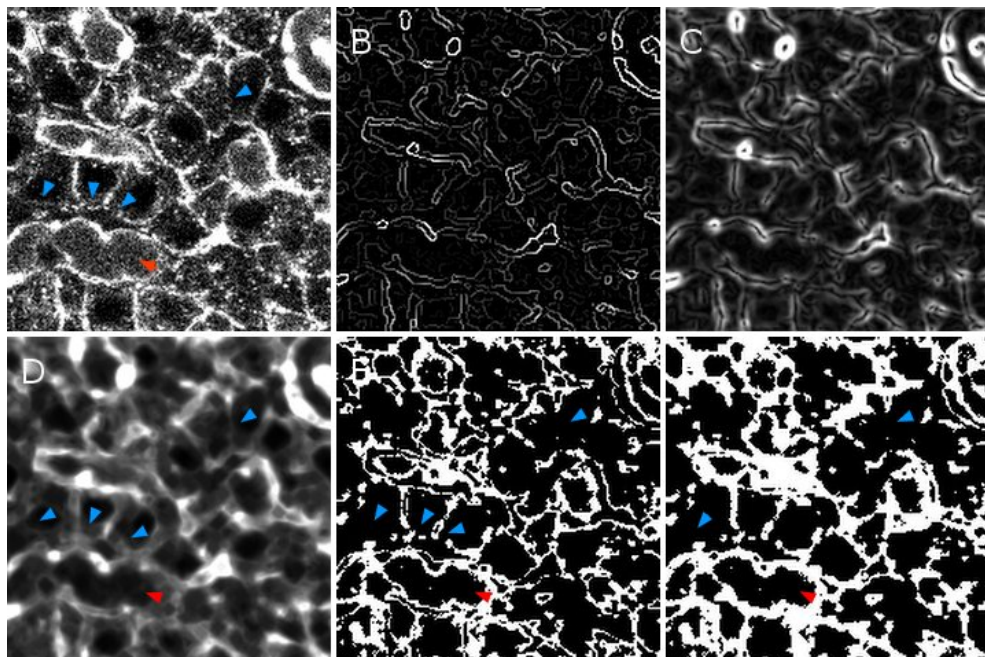


FIGURE 4.4: Results of membrane processing, slide 19 of 53 slices, using Image Edge plugin, under ImageJ platform. Red arrow: blood vessel area, blue arrow: cell contours area. (A) original membrane labeling, (B) result of membrane contours, (C) result of normalized membrane contours, (D) result of membrane filter using normalized contours + mean filter radius 2-2-1 + 3D minimum filter, (E) result of membrane segmentation with a low threshold, (F) result of membrane segmentation with a high threshold.

#### 4.2.2.2 Cell zone computation using watershed separation

##### Cell zone computation

Since we could not rely on membrane labeling for cell boundary computation, we tried to determine the area of the cell surrounding the nucleus, the cell zone, by applying the 3D watershed separation algorithm. Specifically, we considered the nucleus as the

theoretical core of the cell and predicted the maximum likelihood of the cell zone by extending the nucleus' boundaries according to the average cell size  $R_{cell}$ .

In 3D watershed separation algorithm, if two nuclei are closer than two times  $R_{cell}$ , the region will be split midway between these two cell zones as shown in Figs. 4.5C, Figs. 4.6C, and Fig. 4.7. As such, an obvious question is how to set the value of  $R_{cell}$ .

#### **Estimate the average cell size $R_{cell}$**

There are several ways to estimate the cell size  $R_{cell}$  defined as the distance from nucleus' boundaries to the cell zone contours. The first ways to do that is to use the plasma membrane and nuclei labeling obtained from the previous experiment to measure the distance between the nucleus and its cell boundaries then use this distance as the average cell size  $R_{cell}$ . The second way is to use three different markers for three cell types in order to measure the distance between the nucleus and its surrounding marker areas. This distance can then be used to estimate the average cell size  $R_{cell}$ . In our work,  $R_{cell}$  has the value of  $4\mu\text{m}$  corresponding to approximately 10 pixels in mouse dataset and 11 to 12 pixels in monkey and human datasets. The cell shapes and volumes obtained from this estimation are close to the shapes of real cells.

**Comparison with membrane segmentation.** Several tests were conducted to prove the improvement of 3D watershed approach over membrane labeling segmentation as shown in Fig. 4.5. The first test used only nuclei as seeds and predicted the cell zone by extending the nuclei boundaries according to average cell size  $R_{cell}$  of  $4\mu\text{m}$  as described in Section 4.2.2.2. The second one used intensity driven watershed with segmented nuclei as seeds on the area of labeled membrane as described in Section 4.2.2.1 and Fig. 4.4D. The output in Fig. 4.5C proved that our cell zone determination is closer to the real cell morphology. In contrast, the output of intensity driven watershed as shown in Fig. 4.5F showed poor results in which the cell area over-extended to include the surrounding environment, not to mention existing gaps in the membrane staining.

#### **Comparison of different watershed algorithms**

1. Voronoi tessellation method ([46]) as shown in Fig. 4.6A. It uses segmented nuclei images to compute the centers of nuclei as the cell cores. After this, it applies the region growing technique in Kaliman (2016) [46] to define the cell zone area. This method works fine even in the case where cells are located in close proximity and

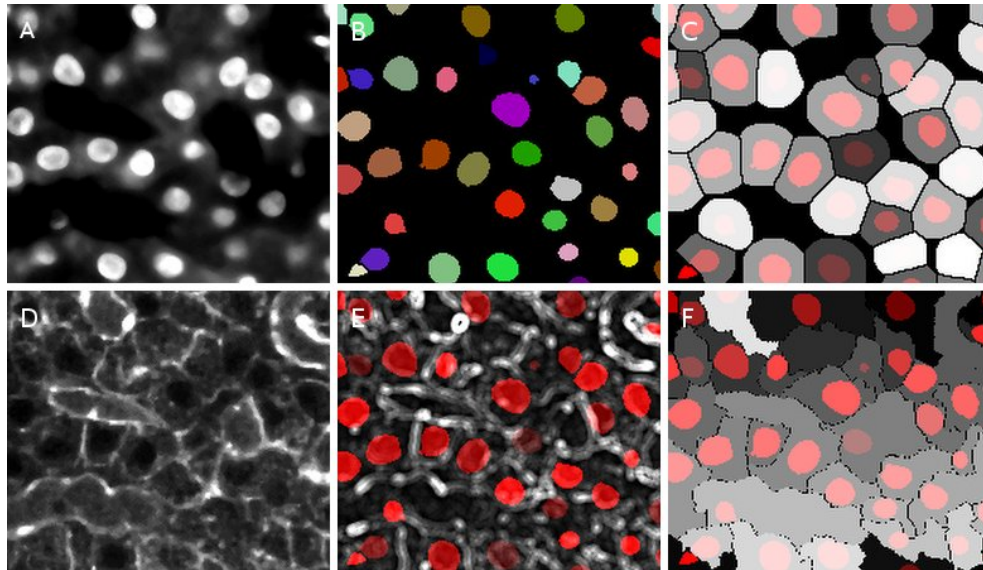


FIGURE 4.5: Comparison of cell zone obtained from membrane detection + segmented nuclei and our segmented nuclei + 3D watershed separation. Nuclei DAPI original image (A), nuclei segmentation result (B), our computation of cell zones based on watershed and maximal radius of average cell size (here 11 pixels, 4.16 microns (C)). Membrane original image(D), merge processed membrane (see Section 4.2.2.1) + nuclei segmentation result (E). Intensity driven watershed - region growing using segmented nuclei as seed based on processed membrane labeling (F).

their volumetric are similar. The main drawback of this method happens when there is a big difference in the volumetric, for example, a big cell is located in close proximity to a smaller one. In such case, the algorithm only takes into account two centers of these nuclei, thus, incorrectly separates the area into two equivalent regions. In other words, one cell zone is under-estimated while the other is over-estimated as shown in the top right corner of Fig. 4.6A. The cell zone is over-estimated along the z-axis, thus, increases the number of contacts for each cell and leads to an over-estimation of number of cell-to-cell contacts as well.

- Utilize the cell nuclei regions as seeds then extend to the entire space as shown in Fig. 4.6B. The drawback of this method is that it could not determine the exact cell zone and covers the entire space instead. Again, this leads to the over-estimation of the number of cell-to-cell contacts.

With the drawbacks of these two approaches in mind, we decide to use a constant speed of average cell size  $R_{cell}$  to enhance them as shown in Fig. 4.6C. Our approach can detect the cell zone better in terms of shape, size, and volumetric with limited errors. As a result, we can find a closer estimation of the number of cell-to-cell contacts.

In order to compare our method with the other two, we conducted an experiment in which we used mcib3d-suite plugin [37] under ImageJ platform to process the big mouse7 dataset as shown in Table 4.4. The plugin provided watershed output images in which each region had one label corresponding to a cell zone. We then used ‘3D ROI’ from mcib3d-suite plugin [37] to extract the contour of each determined cell zone for comparison.

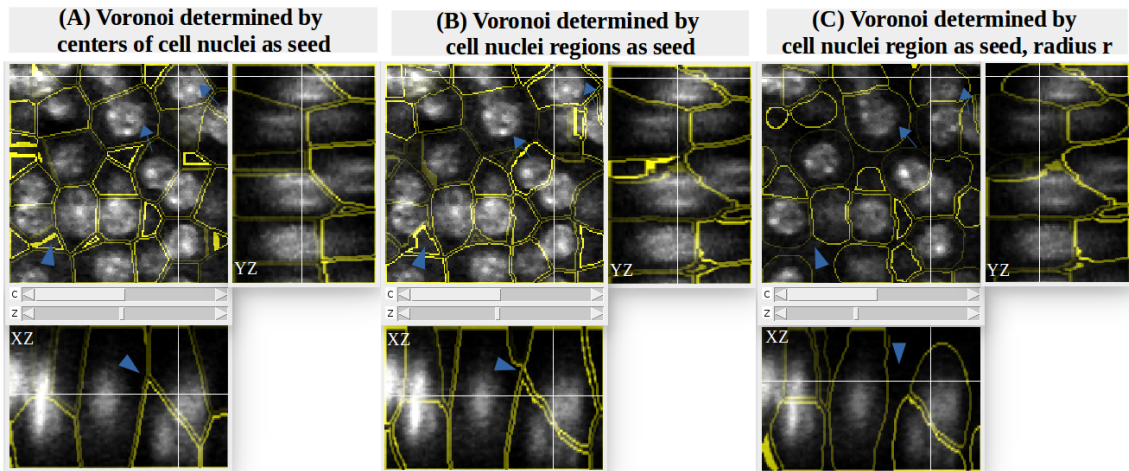


FIGURE 4.6: Result of the Voronoi with three different methods : (A) Method 1, region growing determined by centers of cell nuclei as seed, (B) Method 2, region growing determined by cell nuclei regions as seeds, (C) region growing determined by cell nuclei regions as seed, with constant speed - radius  $R_{cell}$  from boundaries of nuclei, slice 23 of 45, mouse islet sample.

### Impact of cell zone radius $R_{cell}$ to cellular interaction

The previous experiment has proved that our watershed approach can detect the cell zone better than other existing methods. However, we still need to investigate the effect that the average cell size  $R_{cell}$  has upon the cellular interaction computation. To achieve this, another experiment was done in which we applied the same watershed process on the segmented nuclei image but varied the value of the average cell size  $R_{cell}$  from 3 to 30 pixels. The obtained results is shown in Fig. 4.7. It shows that the number of contacts increases quickly when  $R_{cell}$  is in the range from 3 to 10 pixels and the rate of increase slows down when  $R_{cell}$  becomes larger from 11 to 30 pixels. In our use case of islet of Langerhans, the cell radius is around  $4\mu\text{m}$ , or 10 to 15 pixels. Therefore, we believe it's safe to use such value cause it could not impact the analysis of cell-to-cell contacts in a major way.

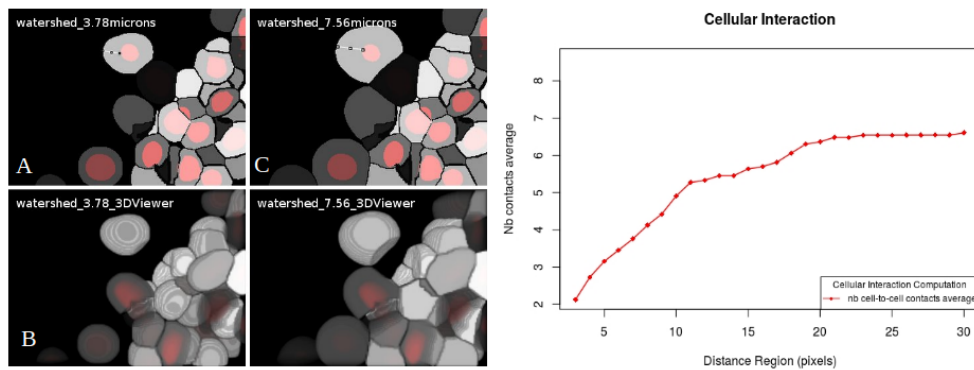


FIGURE 4.7: Result of the computation of the cell zones around each nuclei with two different values of fixed maxima radius  $R_{cell}$  around each nuclei, 3.78 microns (A, B) and 7.56 microns (C, D). Result of computation of the average number of cellular interaction based on maxima radius (radius from 3 to 30 pixels). The number of cell to cell contacts average is quite stable in the interval [10, 15] pixels corresponding to the average radius of nucleus size.

### 4.2.3 3D Cell Phenotype Computation

In this section, we go through some selected input parameters using for 3D cell phenotype computation. Followed by a small discussion about impact of choosing a study area to cell phenotype computation. Then, we provide the results of cell type identification applying on the validated mouse and monkey islets.

#### 4.2.3.1 Parameter selection

For the first step, we use 3D SLIC super-voxels clustering algorithm on three markers, glucagon, insulin, and somatostatin. In short, the input image is divided into square zones of similar size  $R$ , where  $R$  can be somewhere between 100 to 300 pixels according to the size of the nuclei or cells. The threshold of 0.1 is set to define the cell type. If the number of voxels corresponding to a marker at the study area is greater than 10% of the volume of observed cell area, we assign the cell type to the most abundant intensity inside this study area. Otherwise, the cell is marked as ‘unlabelled’.

The study area is computed using  $S1$  strategy from Section 3.3.3 in which Euclidean distance maps from nuclei boundaries to the areas inside and outside are calculated. After that, we observe the zone inside nuclei with radius  $d_1 = 2$  and the zone outside radius with radius  $d_2 = 3$ . If  $d_1$  is larger, the study area will cover all nuclei zones

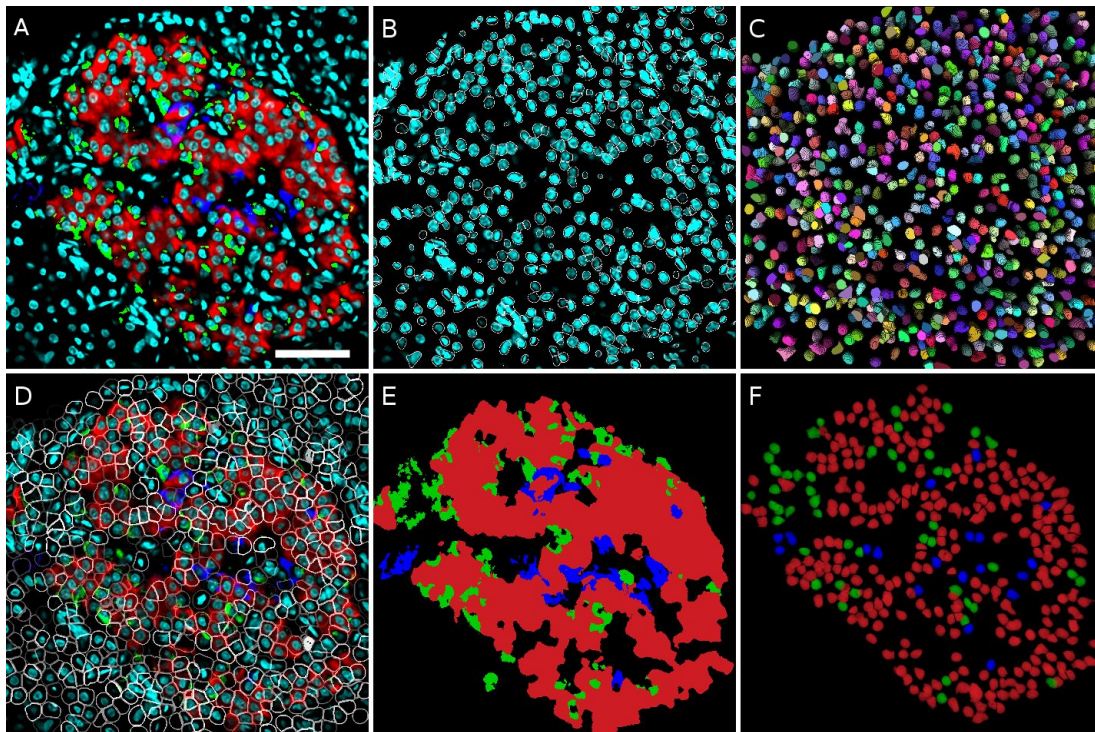


FIGURE 4.8: Results of nuclei segmentation and cell type computation for the monkey1 dataset. (A) Merged image of raw data, only one slice is displayed (red=insulin, green=glucagon, blue=somatostatin, cyan=DAPI). (B) DAPI channel with overlaid contours of detected nuclei. (C) 3D visualization of segmented nuclei. (D) Merged raw image with overlaid contours of computed cell boundaries. (E) 3D SLIC supervoxels clustering result (red=insulin, green=glucagon, blue=somatostatin) (F) 3D visualization of computed cell types (red=beta-cells, green=alpha-cells, blue=delta-cells). Scale bar,  $50\mu\text{m}$ .

including noises at the center part of the nuclei and it affects the identification of cell type.

Performance-wise, cell phenotype computation usually takes about 2 to 3 minutes for SLIC clustering and another minute for the computation.

### Impact of choosing a study area to cell phenotype computation

As described in Section 3.3.3, there are four different strategies  $S_1$ ,  $S_2$ ,  $S_3$ ,  $S_4$  to determine a study area based on nucleus zone and cell zone for detecting the cell type.

- Strategy  $S_1$ : Counting from the nucleus boundary towards the inside of the nucleus with the inner distance  $d_1$ . Counting from the nucleus boundary towards the outside of the nucleus with outer distance  $d_2$ . We call it study area  $A_1$ .
- Strategy  $S_2$ : Counting from the nucleus boundary towards the outside of the nucleus with the outer distance  $d_2$ , called study area  $A_2$ .

Average cell size(pixels)	4	5	6	7	8	9	10
Nb of $\alpha$ cells	42	42	44	41	44	43	46
Nb of $\beta$ cells	204	218	232	243	244	251	257
Nb of $\delta$ cells	28	35	37	42	45	48	48
Nb of <i>unlabelled</i> cells	305	284	266	253	246	237	228

TABLE 4.6: Effect of average cell size (for cell zone computation) to cell phenotype computation. When increasing the radius of cell size  $R_{cell}$ , and use obtained cell zone as study area (strategy  $S4$ ), number of cell phenotype is increased ('unlabeled' cell detected as a cell phenotype alpha, beta, delta) and also there are a shift of cell type between alpha-, beta-, and delta-cells (wrong identification of cell type).

- Strategy  $S3$ : Counting from the nucleus boundary towards the inside of the nucleus with the inner distance  $d_1$ , called study area  $A3$ .
- Strategy  $S4$ : Observing the entire cell zone obtained from watershed from the section of cell zone detection, called study area  $A4$ .

In our endocrine cell application, we chose the first strategy  $S1$  to compute our study area because of its stability and well-defined cytoplasmic area.

To back our decision to use  $S1$ , we did an experiment to compare it with  $S4$ , which used the cell zone determined by watershed algorithm as the study area. Under  $S4$ , the study area included all voxels in the cell zone as shown in Fig. 3.8F. The cell zone was computed using cell zone determination process with an average cell size  $R_{cell}$  as the input parameter. We used different values for  $R_{cell}$  and obtained different cell zones or, in this case, study areas respectively.

Fig. 4.9 shows the impact of  $R_{cell}$  upon cell phenotype computation. The rows are  $\alpha$ ,  $\beta$ , and  $\delta$  cells respectively. The first column is the overlay of the three markers (insulin, glucagon, somatostatin) and the nucleus contour from the segmentation process. The following columns are the result of SLIC-median labeling with different  $R_{cell}$  4, 6, 8, 10 pixels respectively. We observe that for  $\alpha$  and  $\delta$  cells, the volume of the study area has a big effect on the output of cell phenotype computation.

Table 4.6 shows the results of cell composition for different study areas corresponding to different cell zone volumetric with the average size from 4 to 10 pixels using the 'monkey5' dataset. It's clear that there is a big variance in cell composition.

The evidences, therefore, support the conclusion that strategy  $S4$  is not suitable for our application because of its large variance. In our application  $S1$  is a better choice because this method is more robust.

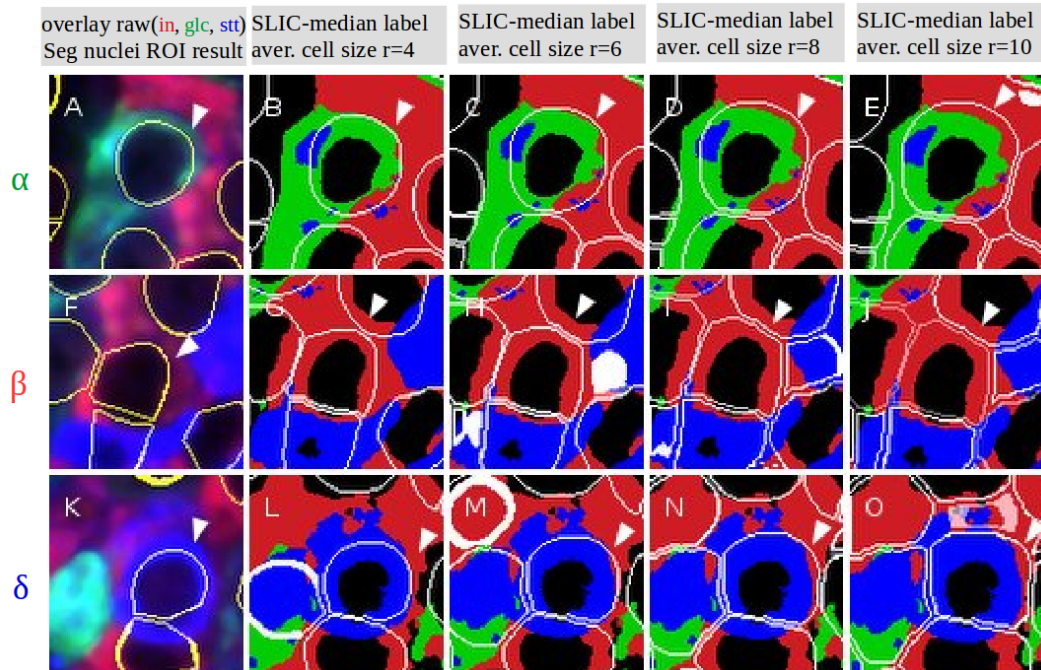


FIGURE 4.9: Result of the computation of the cell zones around each nuclei with 4 different values of fixed maximal radius around each nuclei ( $r = 4, 6, 8, 10$  pixels). A-, F-, K- overlay of 4 channels (glucagon-green, insulin-red, somatostatin-blue, and dapi-white) and nuclei ROI result. First row show experiment for  $\alpha$  cell, second row show experiment for  $\beta$  cell, third row show experiment for  $\delta$  cell which consist of overlay image, watershed ROI with radius 4, 6, 8, 10 pixels.

#### 4.2.3.2 Results of cell type identification

To check the validity of our cell type identification method, we selected 6 mice and 6 monkeys as the validated datasets to be checked by our biologist, Dr Drigo, with the help of the 3D Manager plugin. The result was then combined into a confusion matrix as shown in Table 4.7.

Each column of the matrix represents the instance in a predicted class while each row represents the instance in an actual class. Cells are classified as alpha, beta, delta, or ‘unlabelled’ cells. We achieve an accuracy of more than 97% with the mouse and monkey datasets.

#### Errors and mis-identifications

Although we had normalized each marker into the 8-bit range from 0 to 255 during the pre-processing stage, we still observed a large variance in the intensity of the three markers. In particular, insulin markers indicating beta cells had a very high level of noise in both mouse and monkey datasets. This is the direct cause that led to the wrong



<b>Mouse</b>		Prediction				
		Alpha	Beta	Delta	Unlabeled	Total
Actual	Alpha	490	3	4	2	499
	Beta	0	2146	5	16	2167
	Delta	9	3	135	2	149
	Unlabeled	20	44	11	2021	2096
	Total	519	2196	155	2041	<b>97.57%</b>
<b>Monkey</b>		Prediction				
		Alpha	Beta	Delta	Unlabeled	Total
Actual	Alpha	332	5	2	5	344
	Beta	9	1119	4	19	1151
	Delta	2	5	190	3	200
	Unlabeled	24	27	8	2123	2182
	Total	367	1156	204	2150	<b>97.08%</b>

TABLE 4.7: Actual and predicted classifications result and overall accuracy (number in bold) of cell type computation for mouse and monkey datasets.

identification of the cell type. Our method wrongly detected 44 *unlabelled* cells in the mouse datasets and 27 *unlabelled* cells in the monkey datasets as beta cells. There were also several alpha and delta cells misidentified as beta due to the high amount of insulin used in comparison with glucagon, indicating alpha cells, and somatostatin, indicating delta cells. In some rare cases, some cell area contained both glucagon and somatostatin and mis-identified these two (alpha detected as delta and vice versa).

### Endocrine cell composition

Using the results of cell type identification, we observe that the mean percentages of alpha cells ( $P_\alpha$ ) are 18.2% in mouse and 21.2% in monkey islets. While the same numbers for beta cells ( $P_\beta$ ) are 76.4% and 66.9% respectively. Finally, the remaining delta-cells ( $P_\delta$ ) stay at 5.4% and 11.9% as shown in Table 4.8. These numbers are consistent with those described in other works in the field [14, 16].

#### 4.2.4 Analysis of Cellular Interaction

##### Analysis of direct cell-to-cell interactions

Once all islet-cells had been properly segmented and identified, we computed the number of neighbors for each cell to build a network of interacting cells. This network is a graph whose vertexes are the cell centers and edges are links between two neighboring cells. We use this network to compute the spatial distance function.

Dataset	Total nuclei	Nb of unlabelled cells	Total alpha cells	Total beta cells	Total delta cells	Total alpha + beta + delta	$P_\alpha(\%)$	$P_\beta(\%)$	$P_\delta(\%)$
Mouse1-a	480	278	15	165	22	202	7.4	81.7	10.9
Mouse1-b	772	343	124	261	44	429	28.9	60.8	10.3
Mouse1-c	950	621	19	269	41	329	5.8	81.8	12.5
Mouse2	584	59	118	398	9	525	22.5	75.8	1.7
Mouse3	1012	462	146	384	20	550	26.5	69.8	3.6
Mouse4	198	21	32	138	7	177	18.1	78.0	4.0
Mouse5	425	39	23	358	5	386	6.0	92.7	1.3
Mouse6	493	218	46	222	7	275	16.7	80.7	2.5
<b>Total Mouse</b>	<b>4914</b>	<b>2041</b>	<b>523</b>	<b>2195</b>	<b>155</b>	<b>2873</b>	<b>18.2</b>	<b>76.4</b>	<b>5.4</b>
Monkey1	700	317	55	309	19	383	14.4	80.7	5.0
Monkey2	770	374	136	170	90	396	34.3	42.9	22.7
Monkey3	630	411	57	126	36	219	26.0	57.5	16.4
Monkey4	587	398	46	128	15	189	24.3	67.7	7.9
Monkey5	579	341	33	172	33	238	13.9	72.3	13.9
Monkey6	610	308	40	251	11	302	13.2	83.1	3.6
<b>Total Monkey</b>	<b>3876</b>	<b>2149</b>	<b>367</b>	<b>1156</b>	<b>204</b>	<b>1727</b>	<b>21.2</b>	<b>66.9</b>	<b>11.9</b>

TABLE 4.8: Cell composition of observed mouse and monkey islets.

### Parameters Selection

There are two important parameters that affect the cellular interaction:

1. The number of neighboring voxels  $NB_{neivox} = 26$  surrounding the observed voxel  $P(x, y, x)$  at the cell boundary. This parameter is covered in Section 3.4.1).
2. The average cell size  $R_{cell}$ , which is currently set at  $4\mu\text{m}$ . It is used to determine the cell zone for each cell. The impact of  $R_{cell}$  upon the cellular interaction can be found in Section 4.2.2.2.

### Cellular interaction of section and whole islet

Using the number of neighboring cells, we can compute the average and standard deviation (SD) of the number contacts per cell. When analyzing intact mouse islets, for example ‘mouse4’, ‘mouse5’, and ‘mouse6’ in Table 4.1, Table 4.8, Table 4.9, and Table 4.10, we observed a higher number of direct interactions per islet-cell. Specifically, the average and SD of the number of contacts per cell were 7.48, 8.98, 9.94 for ‘mouse4’, ‘mouse5’, and ‘mouse6’ respectively comparing with the same values of 3.66, 5.62, 4.54 for the three sections of ‘mouse1’ as shown in Table 4.10. These results suggest

that analyzing thin and individual histological sections may lead to an under-estimation of the number of connections in a tissue dataset.

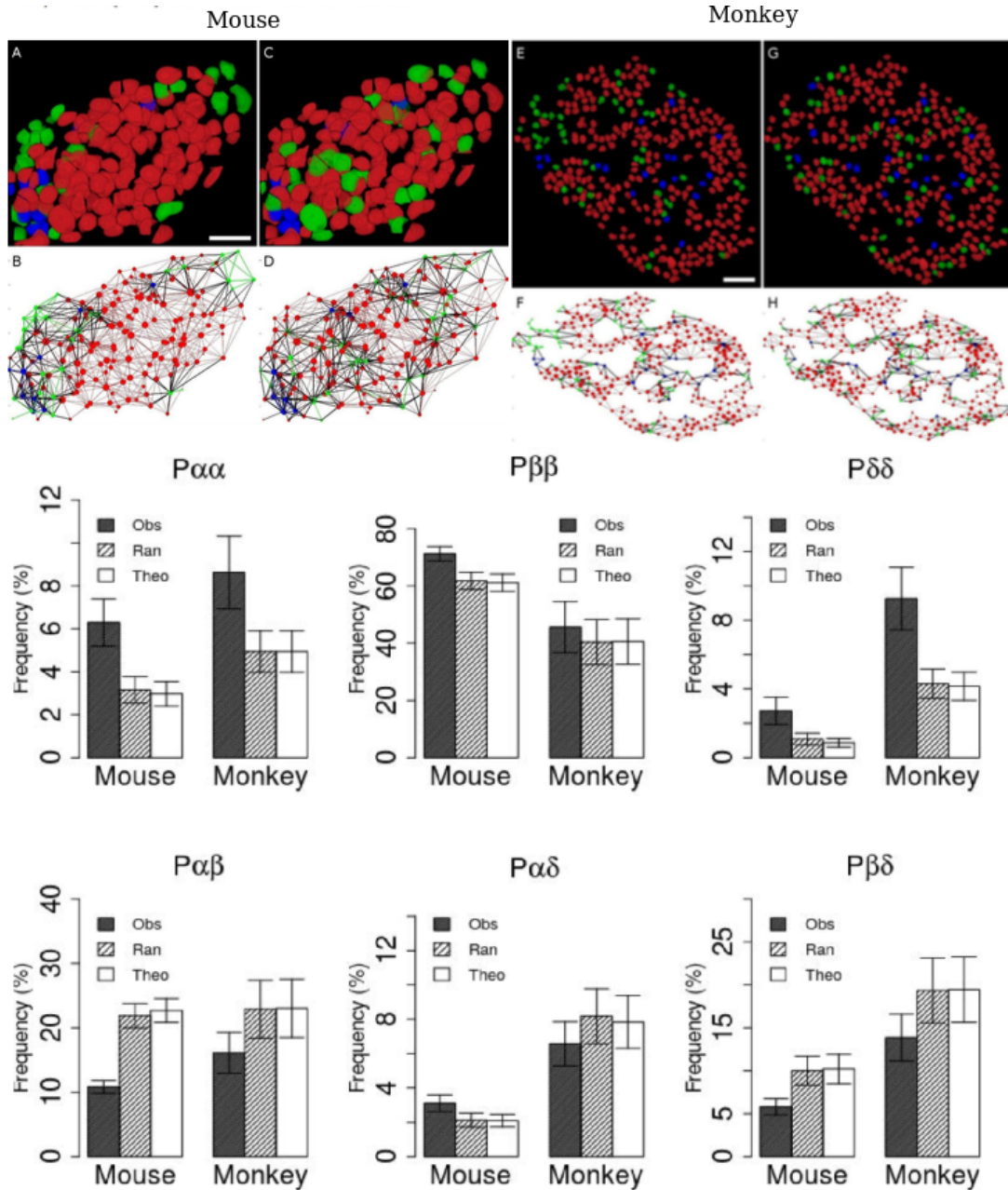


FIGURE 4.10: Cell-cell contact probabilities. Top: an example of result of cell phenotype computation (colored nuclei area, green - alpha cell, red - beta cell, blue - delta cell) and cell network (green alpha-alpha, red beta-beta, blue delta - delta, black - heterotypic interaction between alpha, beta and delta) at mouse 4 and monkey 1 dataset. Bottom: results of frequency of interaction between cells (6 mice,  $n=22$  datasets; 6 monkeys,  $n=12$  datasets) done on observed islet dataset, using random procedure or theoretical model presented in Hoang et al.(2014)[8].

### Homotypic and heterotypic contacts

After computing the relative proportions of different cell types in the tissue, we define

homotypic interactions as those between similar cells (alpha-alpha, beta-beta and delta-delta) and heterotypic interactions as those between two different ones (alpha-beta, alpha-delta, beta-delta) as shown in Table 4.10. In this, we found more homotypic beta-beta interactions in the mouse than in monkey datasets and approximately the same amount of alpha-alpha interactions between them.

### Results of the analysis a larger number of datasets

In order to continue the investigation of the different organizations of the islet of Langerhans tissues in mice and monkeys, we focused our analysis on the two major cell types, alpha and beta, as shown in Fig. 4.11. Due to the heterogeneous structure of the primate islets with a lower percentage of beta cells, we observed a significantly lower  $P_{\beta\beta}$  of only 49.4% ( $p < 0.05$  vs mouse, Fig. 4.11A). On the other hand, there was not much difference in ( $P_{\alpha\alpha}$ ), which was at 8.6% for the mouse database as shown in Fig. 4.11A. For heterotypic connections between alpha and beta cells  $P_{\alpha\beta}$ , we observed that monkey islets had a significantly higher percentage of  $P_{\alpha\beta}$  (17.1%) than mouse islets (10.8%) (Fig. 4.11B,  $p < 0.05$ ).

**Compare the observed structures with a random organization.** If cells within tissue are clustered, we should observe a higher number of homotypic and a lower number of heterotypic interactions than a randomly generated structure. Therefore, we designed a randomization process to generate cells inside a tissue and compared the cell interactions in our randomized models with the theoretical proportion presented in Section 3.4.1. We did not observe any significant differences between those randomized models and the theoretical one. So, we concluded that our randomized models could effectively capture the organization of the tissue and could be used in our work.

When using the random model as baseline, our results showed that mouse islets had significantly higher homotypic contact frequencies as shown in Fig. 4.11B, Fig. 4.11D, and Fig. 4.10. Something similar occurred with the monkey islets but did not prove to be statistically significant (Fig. 4.11C, Fig. 4.11E, Fig. 4.10). In both islets, the observed number of heterotypic contacts between alpha and beta cells was significantly lower than the values obtained in a simulated random distribution (Fig. 4.11F-G).

With these results, we can conclude that alpha and beta cells in mouse and monkey islets are not randomly distributed.

Dataset	Total number of direct interactions	Maximum number of direct interactions for one cell	Avg, sd of number of direct interactions, per cell	Avg, sd of number of direct interactions for alpha cells	Avg, sd of number of direct interactions for beta cells	Avg, sd of number of direct interactions for delta cells
Mouse1-a	370	7	3.66 ± 1.44	3.13 ± 1.09	3.76 ± 1.40	3.32 ± 1.74
Mouse1-b	1206	11	5.622 ± 1.8	4.79 ± 2.0	5.96 ± 1.5	5.98 ± 1.8
Mouse1-c	747	9	4.54 ± 1.4	3.79 ± 1.6	4.59 ± 1.3	4.56 ± 1.9
Mouse2	2064	15	7.86 ± 2.3	7.63 ± 2.5	7.9 ± 2.3	7.78 ± 2.0
Mouse3	1856	16	6.75 ± 2.3	7.19 ± 3.0	6.4 ± 1.8	9.75 ± 3.0
Mouse4	662	15	7.48 ± 3.0	6.63 ± 2.7	7.70 ± 3.0	7.00 ± 2.6
Mouse5	1734	20	8.98 ± 3.0	6.69 ± 2.8	9.13 ± 2.9	8.60 ± 1.7
Mouse6	1368	18	9.94 ± 3.2	7.43 ± 2.5	10.45 ± 3.1	10.28 ± 3.2
Monkey1	850	9	4.44 ± 1.6	3.78 ± 1.6	4.63 ± 1.5	3.32 ± 1.3
Monkey2	804	11	4.06 ± 1.7	4.06 ± 1.6	3.81 ± 1.5	4.53 ± 2.2
Monkey3	412	11	3.76 ± 2.1	4.35 ± 2.2	3.39 ± 1.9	4.11 ± 1.9
Monkey4	417	9	4.41 ± 1.6	4.19 ± 1.7	4.39 ± 1.6	5.2 ± 1.8
Monkey5	430	8	3.61 ± 1.4	3.51 ± 1.6	3.47 ± 1.4	4.45 ± 1.2
Monkey6	677	9	4.48 ± 1.6	4.57 ± 1.4	4.43 ± 1.6	5.27 ± 1.0

TABLE 4.9: Number of direct interactions, per cell, for all datasets.

Type interaction	alpha-alpha	beta-beta	delta-delta	alpha-beta	alpha-delta	beta-delta	Total
Mouse1-a	7	282	16	24	9	32	370
Mouse1-b	142	623	50	228	82	81	1206
Mouse1-c	7	566	45	32	26	71	747
Mouse2	228	1354	2	414	30	36	2064
Mouse3	332	1070	17	276	110	51	1856
Mouse4	53	468	6	98	8	29	662
Mouse5	19	1558	2	116	0	39	1734
Mouse6	61	1037	2	200	20	48	1368
Total mouse	849	6958	140	1388	285	387	10007
<b>Average (%)</b>	<b>8.48</b>	<b>69.53</b>	<b>1.4</b>	<b>13.87</b>	<b>2.85</b>	<b>3.87</b>	<b>100</b>
Monkey1	37	631	6	125	9	42	850
Monkey2	162	218	143	159	69	53	804
Monkey3	55	152	29	86	52	38	412
Monkey4	45	227	9	76	27	33	417
Monkey5	13	235	36	71	19	56	430
Monkey6	24	476	2	121	14	40	677
Total monkey	336	1939	225	638	190	262	3590
<b>Average (%)</b>	<b>9.35</b>	<b>54.01</b>	<b>6.27</b>	<b>17.77</b>	<b>5.29</b>	<b>7.31</b>	<b>100</b>

TABLE 4.10: Number of direct interactions between different islet-cell types in observed mouse and monkey islets.

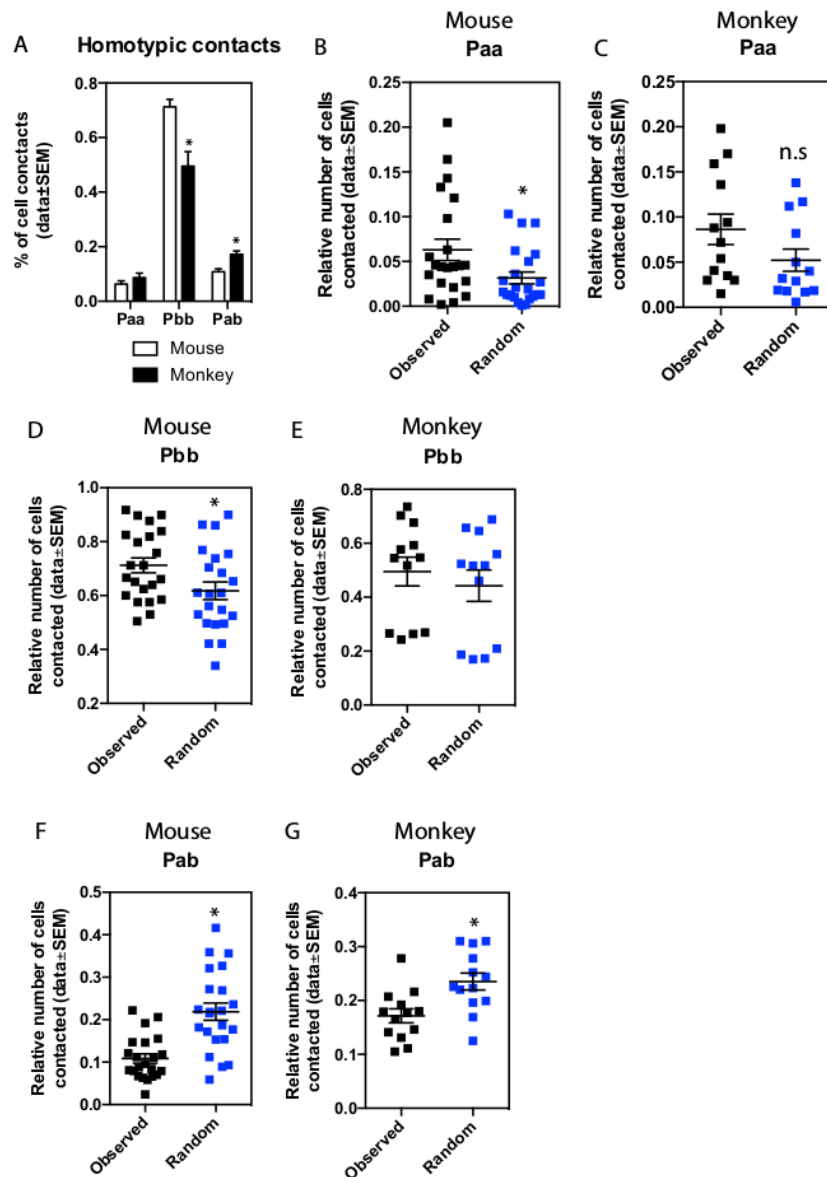


FIGURE 4.11: Relative proportions of direct cellular interactions between the two main cellular types alpha- and beta-cells for the extended datasets (6 mice,  $n=22$  datasets; 6 monkeys,  $n=12$  datasets). (A) Homotypic contacts in mouse and monkeys datasets (\*) denotes significant difference. (B-G) comparison of cellular interactions between extended mouse and monkey datasets and random models (\*) denotes significant difference. All statistics were done using ANOVA method with the Graphpad Prism V6.0 software.

### Analysis of indirect cell-to-cell interactions

The computation of indirect cell to cell interaction is described in section 3.4.1 and shown in Fig. 3.11. This logic allows the computation of relative distance maps for any cell or group of cells within our datasets. Here we compare the indirect cell-to-cell interactions in two models: observed model (raw data) and simulated random model using a randomization procedure as described in the Material and Methods section.

We first calculated the average distance between beta-cells and their closest alpha ones as shown in Table 4.11. In this table, there was a higher number of beta-cells, whose distances to their closest alpha ones were larger than or equal to 2, in the observed model than in the randomized one. It supports the hypothesis that alpha-cells are organized in a cluster-like pattern. This hypothesis is also backed up by the well-known mantle organization of alpha cells in mouse islets [14, 15]. Nevertheless, the difference between observed and randomized modules is less striking in monkey. This suggests that alpha-cells are more evenly distributed throughout the islet, thus explains the higher frequency of alpha-beta contacts as shown in Fig. 4.11 and Table 4.11.

### 4.2.5 Analysis of Spatial Organization

After checking the direct and indirect cellular interactions, we will apply spatial point pattern analysis approach, in which we compare our observed models of cells with completely random patterns, to determine whether there is a special distribution of cells within the tissue. Specifically, we first calculate the cumulative distribution function of the distance  $F$ -function,  $G$ -function. Then, we apply the Spatial Distribution Index (SDI) to determine whether the spatial distribution matches any organizational rules [4].

#### Set up the experiment

1. The study area. The experiment is performed inside the islet space where all the cells are identified based on previous segmentation and computation steps.
2. Events in the study area. Our tissue contains three different cell types alpha, beta, and delta in which one can be chosen as the observed points - events. Numerous previous works have shown that beta cells are always located in several big clusters at the shell-core, thus are not the topic of interest in this experiment. Also the number of delta cells is quite small, ranging from 5 to 44 per sample in the mouse datasets, and 11 to 90 per sample in the monkey datasets as shown in Table 4.8, so it is difficult to apply a statistical test with this type. Therefore, we end up with the alpha cell whose number is large enough for our study, from 15 to 146 in the mouse datasets, and 33 to 136 in the monkey datasets.
3. Datasets. We use the mouse and monkey islet datasets.

Cell Distance	d=1	d=2	d=3	d=4	d=5	d=6	d=7	d=8	d>=9
Observed Mouse1-a	20	28	29	31	13	13	11	11	9
Simulated Mouse1-a	40.52	42.15	32.79	20.93	11.67	6.79	4.18	2.7	3.27
Observed Mouse1-b	139	91	30	1	0	0	0	0	0
Simulated Mouse1-b	214.93	41.54	3.82	0.52	0.15	0.04	0	0	0
Observed Mouse1-c	25	32	42	43	31	36	31	16	13
Simulated Mouse1-c	60.91	72.62	60.31	39.03	20.32	8.89	3.97	1.76	1.19
Observed Mouse2	169	120	81	28	0	0	0	0	0
Simulated Mouse2	336.81	60.26	0.93	0	0	0	0	0	0
Observed Mouse3	125	86	61	43	32	21	12	4	0
Simulated Mouse3	325.57	55.83	2.5	0.1	0	0	0	0	0
Observed mouse4	50	58	20	2	0	0	0	0	0
Simulated mouse4	101.11	26.88	1.86	0.15	0	0	0	0	0
Observed Mouse5	74	85	77	69	46	5	1	0	0
Simulated Mouse5	149.56	147.25	47.99	10.36	1.64	0.2	0	0	0
Observed Mouse6	93	94	34	1	0	0	0	0	0
Simulated Mouse6	181.43	39.5	1.06	0.01	0	0	0	0	0
Observed monkey1	94	94	63	39	16	1	0	0	0
Simulated monkey1	148.23	100.48	40.21	12.43	3.69	1.07	0.49	0.23	0.17
Observed Monkey2	85	38	19	10	6	4	1	0	0
Simulated monkey2	128.6	28.82	4.62	0.81	0.12	0.02	0.01	0	0
Observed Monkey3	25	15	10	4	1	0	0	0	0
Simulated monkey3	37.93	13.73	2.62	0.57	0.13	0.02	0	0	0
Observed Monkey4	48	32	24	17	6	1	0	0	0
Simulated monkey4	87.42	33.19	6.53	0.79	0.05	0.02	0	0	0
Observed Monkey5	52	40	27	21	16	5	3	2	3
Simulated Monkey5	68.39	51.49	27.24	12.39	5.4	2.04	0.96	0.55	0.77
Observed Monkey6	89	74	45	20	13	6	1	0	0
Simulated Monkey6	114.19	81.53	35	11.94	3.56	1.1	0.38	0.16	0.1

TABLE 4.11: Histogram of cell distances between alpha and beta cells in observed and simulated random organization - the average value of 100 simulated random organization for mouse and monkey islets,  $d$  : distance unit.



Using the profile of the alpha cell (nuclei voxel coordinates and cell zone voxels coordinates), we extract the center of gravity, which is the average of all nuclei voxels coordinates, and use this as the input of our statistical test. While this approach ignores the impact of cell shape variability, it is very easy to perform.

To construct the random organization for comparison, we avoid generating new cell shapes but simply shuffles the cell types without actually moving them from their geographical positions. Specifically, we selected the first alpha cell and then randomly chose another cell in the islet structure, for example another alpha or beta cell. If the selected one was not an alpha-cell we swapped the two cell types. The process was then repeated for all alpha-cells to get the simulated random organization.

### **First experiment: using Euclidean distance function**

In this experiment, we performed all the tests using Euclidean distance as the distance function. We did the analysis on all the datasets and chose 6 mouse and 6 monkey datasets for validation.

We noted that the results of F-function (violet color) for our observed ‘mouse4’ as shown in Fig. 4.12A and ‘monkey1’ islets as shown in Fig. 4.12E were below the envelopes. This result demonstrated that our observed data are clustered.

The results of G-function (violet color) for our observed ‘mouse4’ as shown in Fig. 4.12B) and ‘monkey1’ islets as shown in Fig. 4.12F were above the envelopes. This confirmed the cluster pattern of alpha cells based on F-function test.

The SDI values of F- and G-function in mouse and monkey islets are outside the [0.05-0.95] interval (Table 4.12), thus rejects the null hypothesis of a completely random organization of all alpha-cells. This supports our previous hypothesis in Table 4.11 that alpha-cells are organized in clusters in both mouse and monkey islets.

### **Second experiment: using cell distance as the distance function**

Given the cellular architecture of the islet and the importance of paracrine signaling factors secreted by alpha cells [14], we hypothesize that the interaction between neighboring cells plays a more important role than the shortest Euclidean distance between two cells. Hence, we implement a modified version of F- and G- functions which utilize cell distance instead of Euclidean distance. Cell distance is defined in Section 3.4.1.

Dataset	SDI-F Euclidean Distance	SDI-G Euclidean Distance	SDI-F Cell Dis- tance	SDI-G Cell Dis- tance	SDI-F Cluster	SDI-G Cluster
Mouse1-a	0.99	0	0.99	0	0.31	1
Mouse1-b	1	0	1	0.805	0.08	1
Mouse1-c	0.99	0.63	1	0.02	0.97	0.98
Mouse2	1	0	1	0	0.72	0.96
Mouse3	1	0	1	0	0.97	0.82
Mouse4	1	0	1	0	1	0
Mouse5	1	0	1	0	0.96	0.81
Mouse6	1	0	1	0.01	0.62	0.98
Mouse4 randomized	0.91	0.62	0.83	0.38		
Monkey1	1	0	1	0	0.85	0.8
Monkey2	1	0	1	0	0.02	0.96
Monkey3	1	0	1	0	0.84	1
Monkey4	0.99	0.01	1	0.01	0.22	0.705
Monkey5	0.96	0.01	0.98	0.005	0.74	0.96
Monkey6	1	0.01	1	0	0.19	0.96
Monkey1 randomized	0.1	0.64	0.44	0.72		

TABLE 4.12: Spatial descriptor index (SDI) for F-function and G-function, calculated using Euclidean distance and cell distance.

Dataset	Number of clusters	Average number of cells in cluster	Maximum number of cells in cluster
Mouse1-a	8	1.88	3
Mouse1-b	39	3.18	27
Mouse1-c	13	1.46	3
Mouse2	16	7.38	34
Mouse3	20	7.3	93
Mouse4	7	4.57	10
Mouse5	7	3.3	10
Mouse6	6	7.67	36
<b>Total Mouse</b>	<b>116</b>	<b>5.43</b>	
Monkey1	25	2.2	16
Monkey2	34	4	33
Monkey3	15	3.8	22
Monkey4	13	3.53	27
Monkey5	20	1.65	4
Monkey6	21	1.9	8
<b>Total Mon- key</b>	<b>128</b>	<b>2.85</b>	

TABLE 4.13: Cluster analysis for all datasets, with number of clusters in each dataset, average number of cells in clusters and maximum number of cells in a cluster.

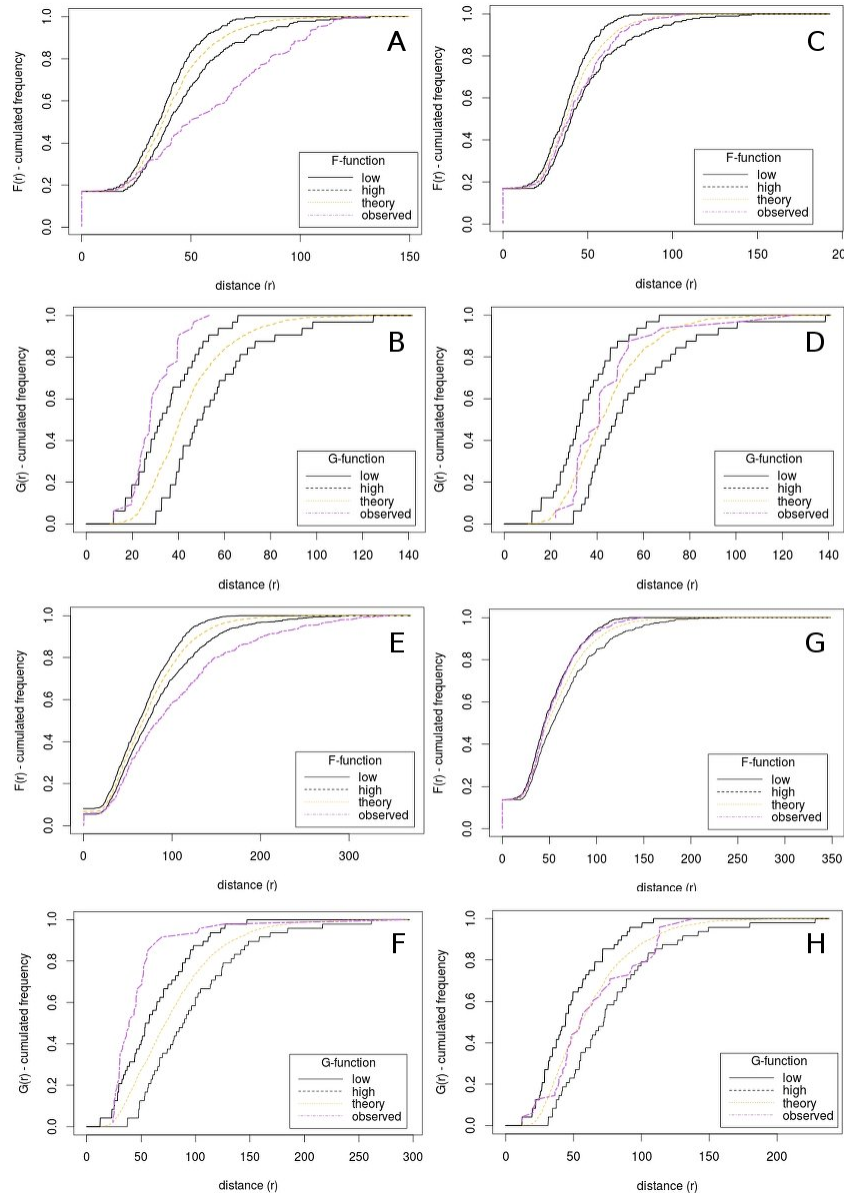


FIGURE 4.12: The corresponding F-function, G-function cumulative distribution functions of the observed data and simulated random organization for alpha-cells within mouse4 and monkey1 islets, using Euclidean distance. F-function (A) and G-function (B) of observed mouse4 islet, F-function (C) and G-function (D) of simulated random organization of mouse4 islet. F-function (E) and G-function (F) of observed monkey1 islet, F-function (G) and G-function (H) of simulated random organization of monkey1 islet. In black (low) low envelope of 5% confidence interval, in black (high) high envelope of 95% confidence interval, in orange (theory) average value of 100 simulated random organization, in violet (observed), our observed data. Note that for observed data the curve is outside interval delimited by the 5%-95% confidence interval.

Using the cell distance, we observed similar results where most SDI values were outside of the 95% range, thus rejected the null hypothesis of completely random organization (Table 4.12). Such result indicates again a non-random organization of alpha-cells in both mouse and monkey islets.

### **Validate the performance of statistical functions with R-statistic**

In our work, we extended the statistic plugin of Phillippe Andrey [4] with additional functions to work with tissue applications. These statistical functions are complex so we want to test their stability and robustness. As R-statistic is a well known tool in the field with numerous existing statistic functions, we choose it as the baseline for our implementation. Specifically, we implemented a function to compute the SDI value using R-script. In this function, we mainly used the Empirical Cumulative Distribution Function (ecdf) provided in R-core library to compute the SDI values with the result of F-function as the input data in Excel format.

The results of these implementations are similar with the error of less than 0.05. The small different is due to the fact that when the observed values are the same, R-function takes into account all the similar values. So, in terms of accuracy, they are equivalent. However, our Java implementation performs better in terms of speed in which it takes from 1 to 3 minutes to complete versus 10 minutes to 1 hours in the case of R-statistic.

### **Cluster analysis result**

The results of spatial statistic point pattern have proved that alpha cells are clustered within the islet area. Nevertheless, we do not know the degree of their cluster, a completely cluster or a partial cluster distribution. In addition, we also want to understand the variability of the structure as well as the intra cellular interactions inside an alpha cluster and those between an alpha cluster and beta or delta cells.

We first characterize the composition of alpha-cell clusters. In mouse islets, alpha-cell clusters vary in size from 1 to 93 cells with a mean of 5.43 cells per cluster as shown in Table 4.13. Monkey islets have smaller alpha-cell clusters, 1 to 27 cells with a mean of 2.85 alpha-cells per cluster. Such results are compatible with other works in the field in which beta cells of rodent islets are located in the core of islets and surrounded by a mantle of alpha and delta cells and alpha cells are located in several clusters at the mantle. Whereas, in non human primate (monkey datasets), existing observations show that alpha cells have more heterogeneous interactions with beta and delta cells.

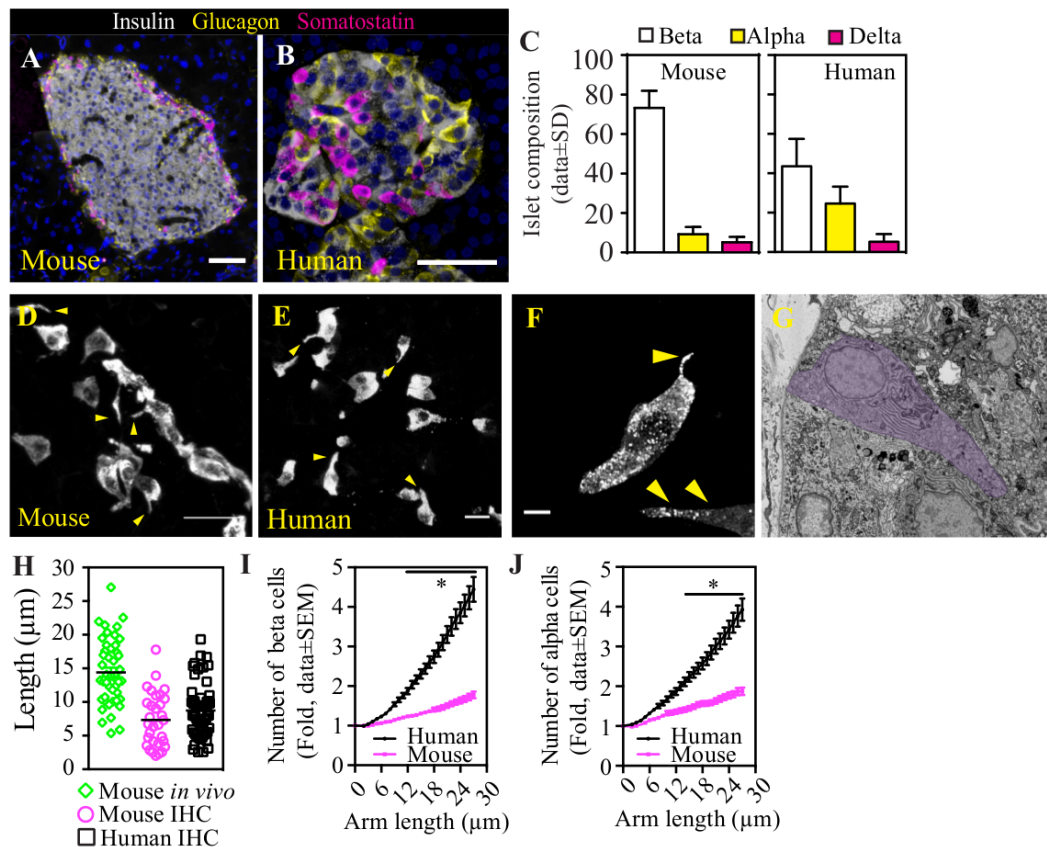


FIGURE 4.13: Cyto-architecture of delta cells in rodents and human islets. Delta cells send out cellular processes to increase their intra-islet reach. (A) and (B), maximum projection images of representative mouse (A) and a human islet (B) glucagon, insulin and somatostatin. Cell nuclei are shown in light blue. (C) Cellular composition of mouse (n=24 islets total, n=7 mice) and human islets (n=21 islets total, n=3 donors). (D) Maximum projection of a mouse islet and (E) a human islet stained with somatostatin labeling. The yellow arrowheads indicate cellular extensions from islet delta cells. (F) human delta cell image with somatostatin labeling). Yellow arrowheads indicate the delta cell extensions. (G) Scanning electron microscopy of a human islet. Somatostatin - delta cell is highlighted by a purple shade. (H) Physical length of the delta cell extension in its longest axis *in vivo* in mouse islets (green circles, n=47 cells) and in mice (black circles, n=31) and humans (purple squares, n=57) samples. (I) Relative gain in delta cell contact with islet beta cells and (J) alpha cells in mouse (pink line) and human islets (black lines). In (A) and (B), scale bar 50  $\mu\text{m}$ . In (D), scale bar 25  $\mu\text{m}$ . In (E) and (F), scale bar 10  $\mu\text{m}$ . (Figure is taken from Fig. 1 of our revision publication [9])

### 4.3 Delta Cell Analysis and Modeling

In the Section 4.2, we have presented our results in tissue quantification and analysis using alpha cells. So, we now switch our focus into the third major cell type in the endocrine pancreas, the islets of Langerhans - delta cell. In other works, it has been hypothesized that delta cells have a dynamic morphology that could reach out and contact with a large number of beta cells within the islet structure. Such hypothesis

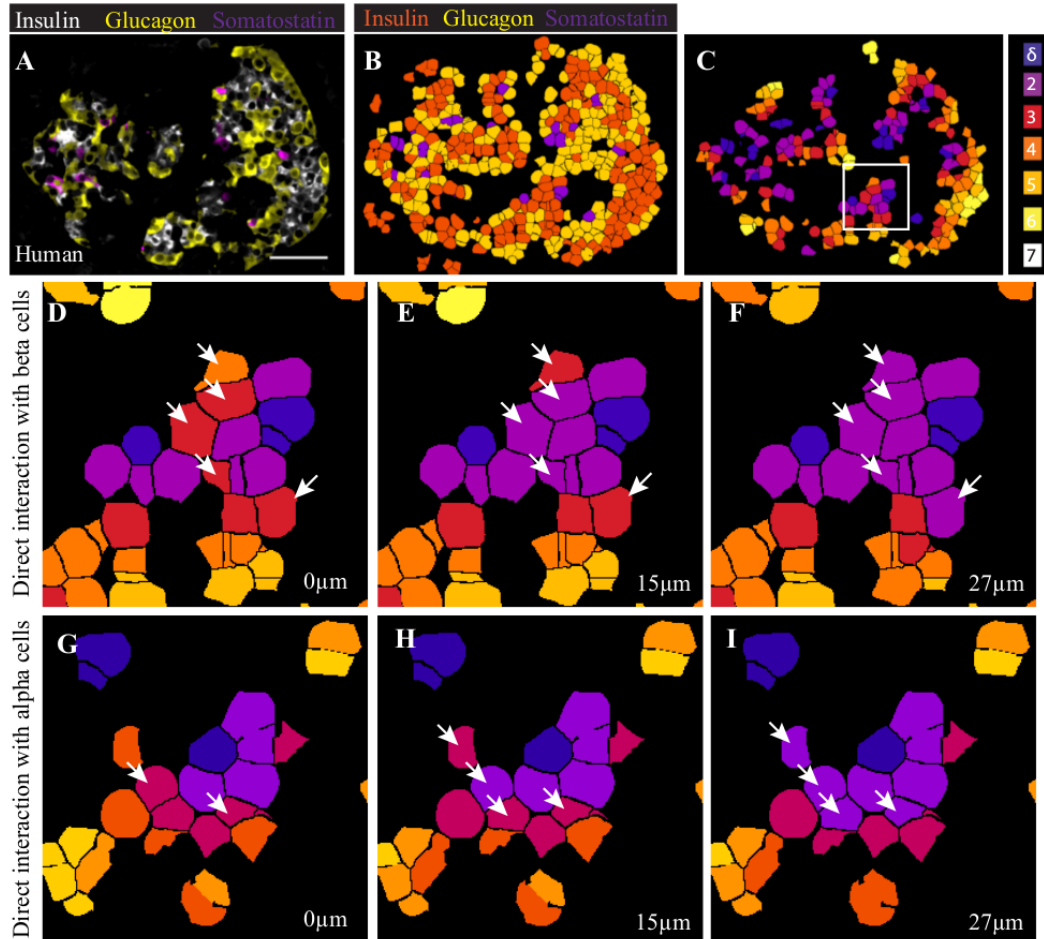


FIGURE 4.14: Mapping delta cell contacts in mouse and human islets. (A) Overlay of insulin, glucagon and somatostatin marker in a human islet. (B) 3D mapping of beta, alpha and delta cells using the algorithm described in our plugin[10]. (C) cell distance map of beta cells from the nearest delta cell (purple). (D-I) result image showing the result of simulation of increasing lengths of delta cell arms and its influence on the relative beta cell (D-F) or alpha cell (G-I) contacts in human islets. White arrows indicate cells that are not contacted directly by a target delta cell but become within reach by increasing arm length of 15 and 27  $\mu\text{m}$ . (Figure is take from from Supplementary Fig. 3 of our revision publication [9])

suggests the important role of delta cells as regulators of beta and alpha cells [14]. To achieve this goal, we use our TITSO toolbox to study the function, dynamic morphology, as well as the spatial organization of delta cells in our rodent (mouse,  $N=24$  islets total,  $n=7$  mice) and human islets (human,  $N=21$  islets total,  $n = 3$  donors).

#### 4.3.1 Quantification and analysis of delta cells

We first use the TITSO toolbox to repeat the same process that we have done with alpha cells to delta cells in the mouse and human datasets. The results show that, in

both species, the percentage of delta cells is between 1% to 5% of the total number of cells within each islet (Fig. 4.13), which is smaller than beta or alpha cells. This is one of the main difficulties in working with them.

Fig. 4.13 also shows two entire islets in mouse and human. In the first dataset, delta cells are located in the mantle area having contact with alpha cells (Fig. 4.13A). In the second one, delta cells connect more to beta cells and blood vessels.

### 4.3.2 Results of modeling and simulation of dynamic morphology of delta cell

We apply a model and simulated procedure to examine the physical advantages of dynamic morphology of delta cells to cellular interactions. The observations of delta cells over time reveal that they are elongated. In many cases, the morphology of delta cells changes and extends into some directions like an arm structure as shown in Fig. 4.13 D,E,F. In vivo, the length of the delta cell arms vary between 1 to 27  $\mu\text{m}$  and between 1 to 18  $\mu\text{m}$  in fixed section (Fig. 4.13F and H). Therefore, we choose the arm length from 1 to 27 as input parameters of our model.

In order to model the arm of delta cell, we used the 3D dilation operation (Section 2.2.5) to perform the extension of each delta cells within the radius from 1 to 27  $\mu\text{m}$ . For each input radius, we obtained a dilated region of the cell. Then we examined the number of affected beta- and alpha- cells by checking the intersection of these cells zones and the dilated region of the delta cell. The results shown that the extending of the delta cell arm length significantly increased the number of cellular interactions of delta-beta and delta-alpha contacts. Although this trend occurred in a limited way in mouse (1 to 1.8-fold of direct delta-beta, delta-alpha contacts), it has a huge impact in human islets (1 to more than 4-fold of direct delta-beta, delta-alpha contacts) as shown in Fig. 4.13 I,J). These results also confirmed the delta cells structure in mouse and human islet as presented above and demonstrate that the human islet organization is more advanced than in rodent islet.

We also examined the results of direct interactions between delta cells and the other two types. Fig. 4.14C shown the distance map of beta cells from the nearest delta cells, in which the distance from a beta to the nearest delta cell varied from 2 to 7 cell distance.

The second column showed the direct interactions of delta cells to the neighbors beta cells with different arm lengths: raw data, 15  $\mu\text{m}$  simulated arm length, and 27 $\mu\text{m}$  simulated arm length. The white arrow showed the change in the distance and demonstrated the physical advantage of having a longer arm. The longer the arm, the more number of direct interactions with beta cells there is. We observed similar results between delta and alpha cells with raw data, 15, and 27  $\mu\text{m}$  simulated arm length.

To summarize, these results give the precise quantification of delta function and suggest that delta cells morphology are dynamic. Delta cells enhance the range of somatostatin action in intra islet and have a big impact on regulating the function of beta and alpha cells.

## 4.4 Modeling Spatial Organization

In order to demonstrate the advantage of the modeling task, we present here two user cases that describe how it is used to generate a virtual model of biological components. The first use case uses a cell-like shape such as a sphere to model a biological component and generates a virtual model. The second one uses cell profiles obtained from segmentation as the materials for simulation. We simulate different scenarios of spatial organization such as random organization, cluster distribution, regular organization (repulsion, uniform) following by statistic tests to prove the correctness of our simulated models.

### Use case 1. Intra-structure of vesicles within the cell area

Vesicle is a small biological component but plays an important role in cell function. When vesicles move closely to periphery of the cell membrane, they have a tendency to break out of the periphery and connect to the outside components.

To study the vesicles spatial organization, we generate  $n$  vesicles of  $S_1, S_2, \dots, S_n$  where each vesicle  $S_i$  is represented by a sphere with the center coordinates  $C_i(x, y, z)$  and the radius of  $r_i$ . We then select a big sphere  $P$  with the center coordinate  $CP(x, y, z)$  and the radius of  $R$  as the study area. In this case, the study area is the cell area as shown in Fig. 4.15. We use Euclidean distance function to compute all the distances between different vesicles and cell area. The model is generated using the method described in Section 3.6.



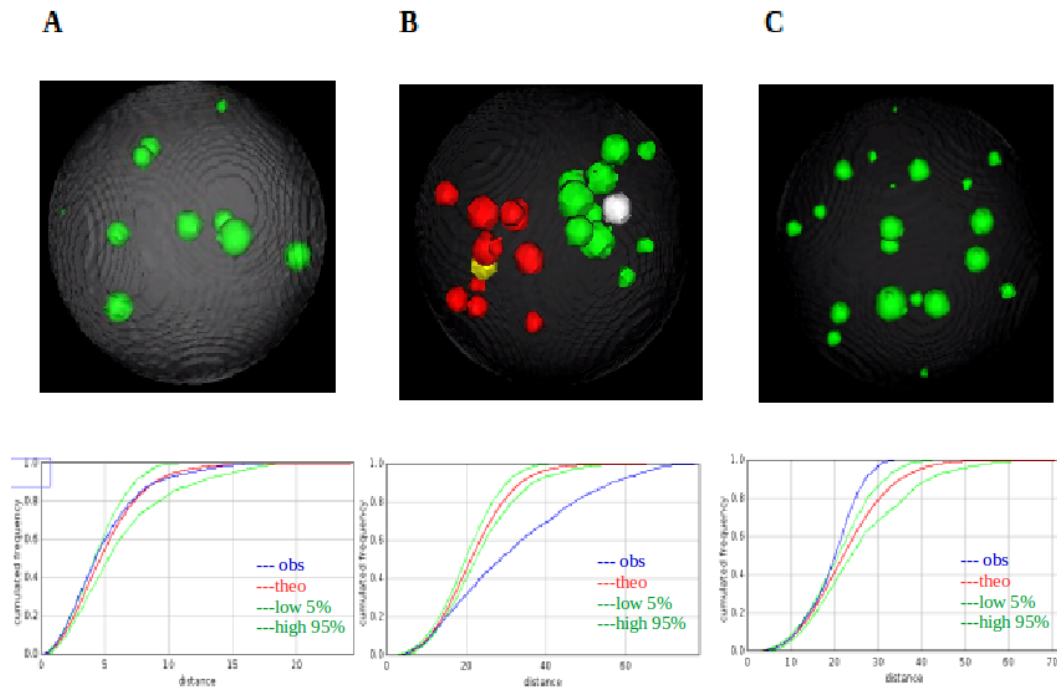


FIGURE 4.15: Results of modeling & simulation of spatial organization of vesicles within cell area, visualize in 3D (Use case 1). Each column represent one type of distribution, followed by result of F-function (blue curve: our observed model, green curves: envelopes of 5% and 95% of 100 random models, red curve: average of 100 random models). (A) random distribution  $SDI = 0.65$ , (B) cluster distribution (attraction)  $SDI=0.02$ , (C) uniform distribution (repulsion)  $SDI=1.0$ .

An example of the generated distributions is shown in Fig. 3.6 which contains a random pattern (A), a cluster pattern (B), and an uniform pattern (C). For each distribution, the result of F-function is shown at the bottom of each column.

Column A shows the result of a randomly generated pattern with 10 vesicles in the intra cell area. The F-curve of the observed pattern (green line) is within the envelopes of 5% and 95% of a completely random pattern, and  $SDI = 0.65$  is also in the same range.

Column B shows the result of a cluster pattern with two clusters. Each cluster contains a vesicle which plays the role of a centroid to attract other vesicles in its surrounding area. The F-curve of the observed pattern (green line) is below the envelopes and  $SDI$  has the value of 0.02. They show that our generated procedure works correctly.

Column C shows the result of a regular pattern with 20 vesicles in the cell area. As expected, the result of F-function is above the envelopes and  $SDI$  index is 1.0. With this, we have successfully generated three different types of distribution.

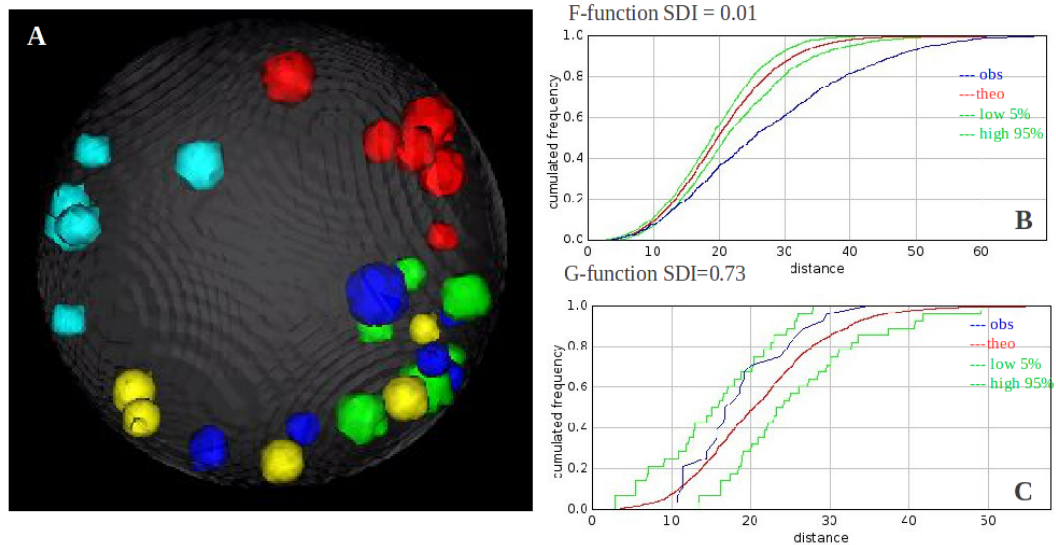


FIGURE 4.16: Results of modeling & simulation of spatial organization of vesicles within cell area at multi-scales (Use case 2). (A) visualization of generated pattern in 3D, followed by result of F-function,  $SDI=0.01$  (B), and G-function,  $SDI=0.73$  (C).

### Use case 2. Modeling spatial organization at multi-scales

Modeling a spatial organization at multi-scales were first introduced in Weston et al. 2012 [5]. The paper included many scenarios such as polar, center, or boundary distribution where biological components are located in different areas. In our case, we built the model at multi-scales using vesicles as materials and the cell area as the study area.

Clusters of vesicles sometimes move in close proximity to the cell periphery and have the tendency to break out of the membrane and connect to the outside environment, which corresponds to the spatial organization model at multi scales. At the local scale, many vesicles are located in close proximity and form clusters. At the global scales, vesicles tend to move to the periphery area.

An example of the multi-scale pattern is shown in Fig. 4.16 where there are 5 clusters. Each cluster consists of 6 vesicles in a predefined range and they all locate in close proximity to the periphery area. The result of F-function (blue line) is below the envelopes with  $SDI = 0.01$  showing that our generated pattern is clustered. However, the results of G-function (blue line) is within the envelopes with  $SDI$  index = 0.73. It means that our pattern is only partially clustered. The visualized 3D model also reaches the same conclusion that the generated multi-scales model is not completely clustered.

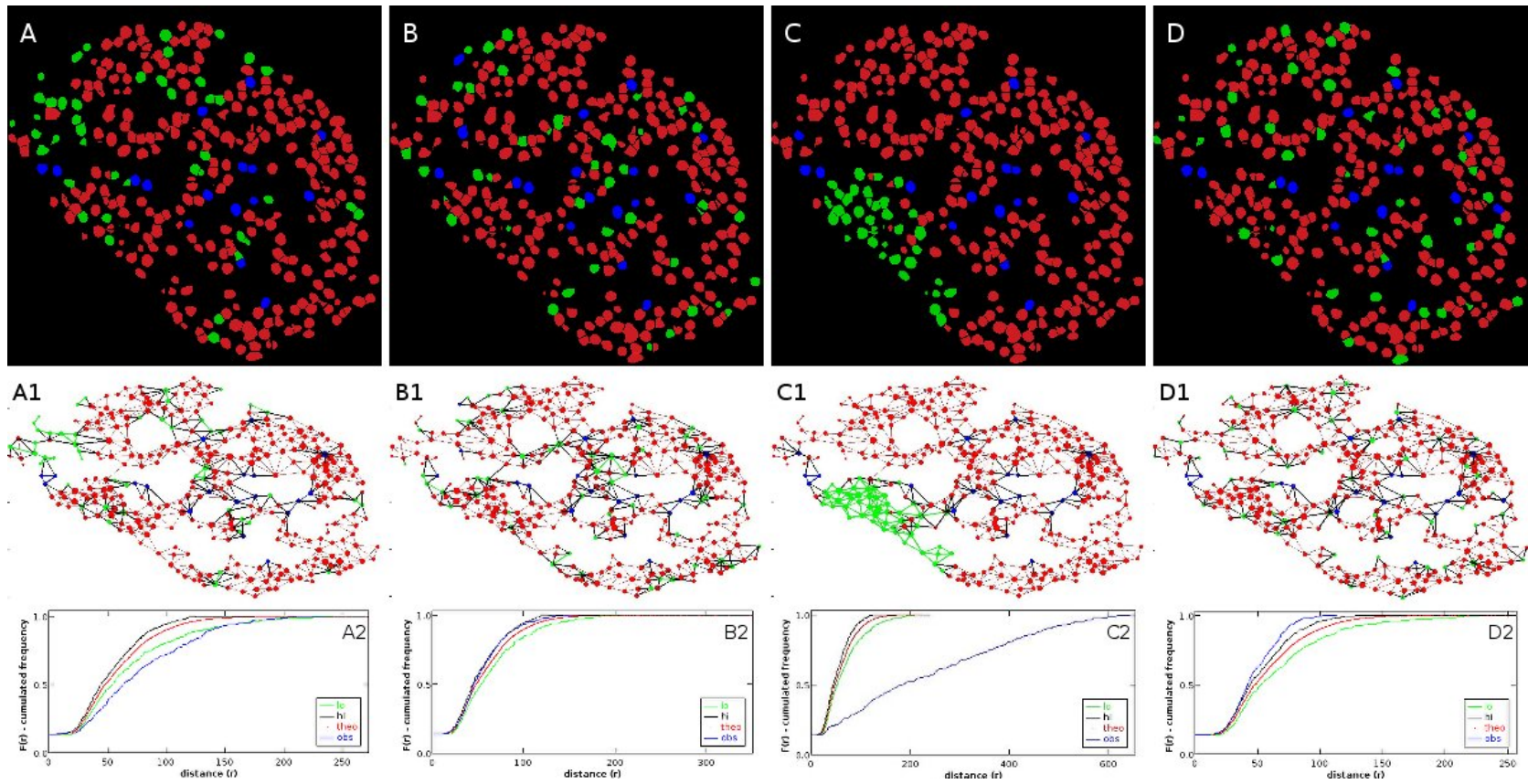


FIGURE 4.17: Results of modelling & simulation of spatial organization of cells within monkey islet (Use case 3). (A)raw data, (B) random pattern, (C) clustered pattern(attraction), (D)regular pattern (repulsion).

**Use case 3. Distributions of alpha cells within islet of Langerhans**

In the islet of Langerhans, alpha cells interact and affect the function of local beta cells. So, there is a considerable interest in understanding the impact of spatial organization of alpha cells to the function of beta cells and of islet when alpha cells are located at different positions within the islet study area. An example using the observed model and three generated ones in ‘monkey1’ tissue dataset is shown in Fig. 4.17.

The first column shows the 3D visualization of the observed pattern (cell type + nuclei zone) and its cell network. There are 48 alpha in the total of 349 cells in the sample islet. The result is below but very close to the low envelope of 5% with SDI of F-function at 1.0 and SDI of G-function at 0.0. Thus, it rejects the null hypothesis of a completely random pattern.

The second column shows the result of the random pattern of alpha cells. The observed curve (blue line) is close to the envelope of 95% with SDI of F-function at 0.99 and SDI of G-function at 0.03. The SDI values show that our generated random model is not completely random because the distance function could not capture the variability of positions well and the cells in the study area are quite dense.

The third column shows the result of the cluster pattern of alpha cells. It is clear that there is a big cluster of alpha cells at the right side of the islet area. The observed curve (blue line) is far below the envelopes of 5% with SDI of F-function at 1.0 and SDI of G-function at 0.0.

The fourth column shows the result of a regular pattern of alpha cells. The observed curve (blue line) is above the envelopes of 95% with SDI of F-function at 0.0 and SDI of G-function at 1.0. Observing this pattern, we notice that there is no direct interactions between alpha cells and they are as far from their neighbors as possible.

To summarize, our modeling procedure can generate different types of distributions. However, it becomes more difficult to generate a completely random distribution when the cells is getting denser.

**Randomizing data for a statistic test**

In all the previous use cases, we need to apply a statistical test to check whether the observed pattern follows the reference one. To do so, we generate a large number  $N$ , for example  $N = 1000$ , of reference patterns and compute their distance functions. Finally,

we compare the output of the observed distance function with those  $N$  to identify the similarity between them.

For example, in order to compare the observed model with completely random patterns, we generated  $N = 100$  random ones and measured the distance between their reference points and the events in the study area. The output of SDI index showed the rank of the observed model compared with  $N$ . The same idea can also be applied to a cluster or a regular pattern.

## Chapter 5

# Conclusion & Discussion

### 5.1 Discussion

In this part, we discuss more about the performance and the limitations of our toolbox. This will be followed by suggestions and ideas for future work.

#### **Nuclei segmentation**

One of the advantages of our algorithm is its ability to segment nuclei in 3D images captured by confocal microscopy better than other methods as shown in Section 4.2.1. Our algorithm uses a two-stage approach for segmentation in which the first stage is to detect regions with maximal intensity locally within each nuclei (the seed). The second stage then uses the seed to initialize a region growing process to achieve nucleus segmentation. However, the key point that makes our algorithm different from other segmentation methods is the seed detection stage in which we transformed the nuclei labeling input image into a Fourier domain image and then applied the band pass filter. This filter allows us to discard any eventual intensity variation in the nuclei, thus, improves the accuracy of the segmentation process. Our method is especially useful when the input data suffers from a high variance of intensity level.

Nevertheless, the disadvantage of the band pass filter is that it cuts off parts of regions with lower or higher frequency than the predefined threshold. In other words, all objects whose diameters are smaller than the minimum value  $D_{min}$  or larger than the maximum value  $D_{max}$  will not be taken into account. We only tune these values so that the number

of affected objects is small enough to be irrelevant. Such approach works in our case because the majority of cells in a tissue has similar volumes.

In terms of shape detection, our method achieves good performance in capturing convex and rounded nuclei. However, when the nuclei shape become flexible, the nuclei shape obtained by our method is slightly bigger and rounder than real boundary. In addition, when the cell is elongated through a big number of slices, our method could over-estimate and add more slices into the first and last parts of the cell in the z-axis.

We have tested our method on different cross-species datasets of mouse, monkey, and human. Although these datasets have different characteristics, we have obtained good accuracy overall. With such result, our method has proven to be useful for domain experts in their works.

One more thing that we should note is the cell division. In our work, we did not take into account mitosis cases, because these cases are very rare in mature tissue.

### **Computation of cell zones**

The computation of marked cell zones is an essential part in the analysis of the spatial organization of any given tissue. In our dataset, membranes of the islet cell are unlabeled and we have chosen the watershed approach to separate the nuclei and to predict the cell zone around each nucleus (Section 4.2.2).

As our first attempt, we used the classical watershed separation approach to compute a Voronoi diagram around the segmented nucleus without any predefined speed. However, this strategy over-estimated significantly the number of cell-to-cell contacts (Section 4.2.2). In order to bypass this limitation, we have modified the original watershed algorithm to constrain the Voronoi zone so that it does not extend further than a predefined radius from the nuclei boundaries. After applying this, the result of our cell zone determination becomes much closer to the actual cell zone.

The analysis of cellular interactions in Table 4.10 has confirmed previous results in describing cellular interactions in mouse and monkey islets. Moreover, the analysis of the whole islet made by tissue clearing has shown the largest number of cellular interactions ('mouse4' and 'mouse 6' in Table 4.9 and Table 4.10) when comparing to datasets with only single or serial sections. This result suggests that the analysis of thin tissue sections is insufficient because it underestimates the number of cellular

interactions. Consequently, to achieve a real picture of the tissue spatial organization, it is necessary to study the whole tissue image in three dimensions. Our results for cell composition and cell-to-cell interaction are also verified by other works [16], thus it validates our computational approach.

### Cell identity computation

In order to identify the cell phenotype and cellular interactions, we first capture the cytoplasmic marker corresponding to each cell type. In our case, since we were investigating three different cell types in addition to the nuclear stain (i.e. DAPI), it was a challenge to add another dye to stain the cellular membrane. Therefore, we bypass this limitation by assuming that the nucleus zone is the theoretical core of cell and the cytoplasmic volume expands only locally around the nucleus. We can then use the theoretically determined cell zone as the input for our cell phenotype identification. Finally, to process the three different markers (glucagon, insulin, somatostatin) altogether, the supervoxels clustering approach (SLIC - [55]) has proven to be the most robust and accurate clustering method for all these three channels. The proportion of each cell type is consistent with other works in this domain [6, 8, 16].

However, due to the thickness of the tissue and the variability of the staining method itself, there is a large variance of intensity between cytoplasmic markers and it affects the accuracy of our algorithm. For example, in our islet application, the insulin marker, denoting the beta cell, has a very strong level of intensity while somatostatin marker, denoting the delta cell, has a much lower level of intensity. Although we have normalized the intensity of all markers in the range of  $[0, 255]$ , the noise of insulin marker is higher than the intensity of somatostatin marker in certain regions. It leads to the wrong detection of cell phenotype (actual delta cell but detected as beta cells) as shown in Table 4.7. Our islet application also examines only three main cell types alpha-, beta-, delta-cells.

In addition, our work only looks at the region with a predefined radius surrounding each nucleus and does not take into account the special case in which the nucleus is surrounded by an external region of glucagon marker and encircled by one more region of insulin marker (beta cell wrapping alpha cell [6]). So, in order to investigate these complex cell phenotypes (containing common cell phenotypes and rare cell phenotypes),



we need a combination of different methods, for example clustering, outlier detection, and classification.

### **Analysis of spatial organization**

In order to investigate the tissue spatial organization, we need a set of robust descriptors. Classical descriptors such as SDI for F- and G- function (Table 4.12 and Fig. 4.12) are very suitable for spatial point pattern analysis. However, there are difficulties in adapting them to whole tissue analysis where the object location is quite constrained and where structural voids, for example blood vessels, may exist inside the structure. Therefore, we choose the same approach but constrain the object position to only the cell location instead of the entire study area as usual.

We also use cell distance instead of Euclidean distance (Fig. 3.11 and Supplementary Table 4.11, 4.12) and observe similar results using either of them. However, in other applications with a complex cell networks and a non-convex topology, cell distance analysis may prove to be more effective in correctly assessing the organization of cells than Euclidean distance.

In chapter 4, we have demonstrated the usefulness of spatial functions F-, G-function and spatial descriptor SDI to detect the spatial distribution of cells within tissue. However, when the cell density is either very low or very high, for example ‘monkey1’, these statistical functions could not capture correctly the type of spatial distribution as shown in Table 4.12. They fail to distinguish a random distribution and a partial cluster distribution.

### **The cyto-architecture of Islets of Langerhans of mice and monkeys**

The result of our method confirms the observation of underlying differences in the cyto-architecture of mouse and monkey islets [14, 16]. While mouse islets are rich in homotypic contacts, monkey islets display a more heterogeneous islet-cell distribution that results in more heterotypic contacts between different cell types (Fig. 4.11 and Table 4.10). This result confirms that we should not use mouse as the testing material to make assumptions about monkey and human islet and we need to do the analysis directly on monkey and human samples.

Interestingly, alpha-cells in all these species tend to be organized into a non-random cluster organization which leads to a non-random pattern of contacts between alpha-

and beta cells as shown in Fig. 4.11. This suggests that alpha-cells are in specific areas within the islet, which might be set during the developmental period [62] and highlight the potential paracrine role that alpha-cells have in islet physiology and diabetes [63–67]. Furthermore, we notice a subtle trend suggesting that alpha-cell clusters are not randomly distributed in all datasets. Alpha cells cluster more at the mantle region in mouse islets while they are near the blood vessel in monkey islets.

### **Limitation: small number of datasets for testing**

Although confocal microscopy helps capturing high throughput 3D images, it is more difficult to capture multiple tissues in one image than the low resolution 2D techniques [19]. Because of this limit, our toolbox could process only one tissue per image in the islet application. Although we have been able to extend our works with more testing datasets and achieved some results on monkey as shown in Fig. 4.11, there is still only a limited number of samples when comparing with other works in the field ([19, 22]).

Nevertheless, the main objectives of our works are to propose a set of new methodologies and to provide them in an efficient framework. We do not really focus on describing a novel biological investigation. Therefore, a small number of datasets is somewhat acceptable in our case.

### **Delta cell analysis and modeling**

Our attempts to model and to simulate the dynamic morphology of delta cells have been able to achieve good results. However, the process is still in its early phase and remains open for future development.

### **The potential of our framework**

There are several points to prove the generic and useful nature of our TITSO framework:

1. We originally develop our framework to use alpha cells as the references for the analysing of mouse and monkey datasets [10]. Since then, the framework has been extended to process delta cells in the human dataset as well.
2. We also apply our TITSO toolbox to analyze the tissue spatial organization and plan to use it to investigate the spatial organization of other biological components such as vesicles within a cell.
3. Our toolbox can analyze different types of fluorescent microscopy images.
4. Our toolbox is developed on Java as a plugin of the popular ImageJ platform.

5. The recent development of 3D bio-printed artificial pancreas in 2017 at <https://diabetes.co.uk/news/2017/apr/3d-bioprinted-artificial-pancreas-for-type-1-diabetes.html>, opens new potential use cases for our toolbox.

However, there are still some drawbacks worth mentioning here: The user interface is not intuitive for new users. They need to have an understanding of the entire work flow before being able to use it correctly. In the future, we plan to improve it further and to add more examples in its documentation.

### **Future works**

Our current model defines the interaction between two cells as direct physical contacts and does not consider the presence of important extracellular matrix proteins or blood vessels. Both of which are important factors for the islet-cell function. As such, potential future extensions of our works are to include the identification of blood vessels and to analyse the spatial relationship between islet-cell and the vasculature [68, 69]. This is further highlighted by recent studies indicating key anatomical and functional differences between mouse and human islets and their microvasculature [70, 71].

The study of different diabetic types is an active domain so domain experts could utilize our framework to explore special patterns of diabetic tissues and compare them with normal ones. Such comparison could shed light upon the underlying reason of diabetes.

## **5.2 Conclusion**

1. We have demonstrated the development of a new set of methods to visualize, analyze, and compare the different organizations of multicellular structures in mouse, monkey, and human islets.
2. Our framework provides an easy-to-use way to automate and standardize the data collection on islet morphology.
3. The fundamental differences between rodent and primate islets support the hypothesis that the physiology and paracrine signaling architectures of rodent and primate islets are different. In addition, alpha-cells are not randomly allocated within the islet ultra-structure.

4. This study highlights the differences between rodent and non primate islets and supports a shift towards islet studies using monkey and human islets directly for a more meaningful understanding of diabetes patho-physiology.

# Appendix A

## Toolbox to Investigate Tissue Spatial Organization (TITSO)

Welcome to TITSO (the Toolbox to Investigate Tissue Spatial Organization) that we have developed for microscopy image processing of tissue samples. This tutorial will present all the necessary stages to use TITSO for the study of spatial organization of tissues.

### A.1 Documentation

#### A.1.1 Introduction

TITSO is a toolbox for segmenting and analyzing cells and nuclei in 3D tissues and their spatial structure. It is originally developed by Boudier's group. The goal is to suggest a workflow in order to choose a set of optimal methodologies for quantification and modeling of tissue organization based on 3D microscopy images.

This ImageJ plugin provides easy access to biologists and has been used to analyze endocrine cells in:

- *'A novel toolbox to investigate tissue spatial organization applied to the study of the islets of Langerhans'*. . TRAN THI NHU Hoa, ARROJO E DRIGO Rafael, BERGGREN Per-Olof, and BOUDIER Thomas. Scientific Report, 7, 2017.

- *Pancreatic islet delta cells reach out to modulate beta cell function.* Arrojo e Drigo Rafael, Jacob Stefan, Zheng Xiaofeng, Garcia-Prieto Concha F., Fukuda Masahiro, *Tran Thi Nhu Hoa*, Stelmashenko Olga, Martins P. F. Leticia, Rodriguez-Diaz Rayner, Bushong Eric, Deerinck Thomas, Phan Sebastian, Ali Yusuf, Leibiger Ingo, *Boudier Thomas*, Augustine George, Ellisman Mark H., and Berggren Per-Olof. *Nature (Revision)*, 2017.

For more information, please send email to: [hoa.tran\\_thi\\_nhu@etu.upmc.fr](mailto:hoa.tran_thi_nhu@etu.upmc.fr) or [thomas.boudier@upmc.fr](mailto:thomas.boudier@upmc.fr).

### A.1.2 Data requirements

Tissue Analysis Toolbox requires 3D microscopy images as its input data:

- 3D Nuclei DAPI staining. Nuclear labelling DAPI which capture nucleus in 3D.
- Markerset. A markerset represents a set of fluorescence markers or imaging channels used in an experiment. More than one markerset may be associated with a single tissue. This allows users to perform analysis with different combinations of the available markers. In our paper, we applied our toolbox with endocrine cells which composed of three channels (three markers) named insulin staining, glucagon staining, and somatostatin staining.
- Image file format. Only 8-bit or 16-bit TIFF format file of 3D are currently supported. All images in one dataset should have the same size.
- Input. A composite image composed of all different channels and one nuclei marker image.
- We also provide an example of input images extracted from islet of Langerhans (more detail see reference paper). DAPI staining nuclei image named as C4-dapi.tif, three channels correspond to cytoplasmic (label or marker) and efficient to determine cell types: somatostatin: C1-delta.tif, insulin: C2-beta.tif, glucagon: C3-alpha.tif. In order to open these images example: launch ImageJ, open zip file and obtain the tiff format of these images.

### A.1.3 Installation

To install toolbox, we need to include these library package:

- Install ImageJ environment, see instruction in ImageJ web page. ImageJ version should be equal or higher than 1.50a and Java 1.7 or later is required JDK.
- Download the Tissue Analysis plugin in this link: [tissueanalysis.jar](#).
- To work with 3D TIFF image, some libraries are required also: [imagescience.jar](#), [imageware.jar](#).
- The ConvexHull3D plugin is required only for measurements: [quickhull3d.jar](#). Java3D is required, download our 3D Viewer version 3D Viewer plugin (version 4 or later) or go to the main website: [3D Viewer](#).
- Mcib3d core and plugin, version 3.8 or latter: [mcib3d-core-3.83.jar](#), [mcib3d-plugins-3.83.jar](#) included many functions for 3D image processing.
- Package Lipid Droplet Counter which contains 3D band pass filter, included in 3D Nuclei Segmentation of tissue organization plugin.
- Put the jar file into the plugin folder of ImageJ and refresh menu.
- More details about the required plugin you can find at [mcib3d-suite plugin \[37\]](#).

Starting TITSO:

- Launch ImageJ program.
- From ImageJ go to Plugin then go to 3D Tissue Spatial Analysis plugin.

You can find the source code of this framework at our github repository: [https://github.com/nhuhoa/tissue\\_organization](https://github.com/nhuhoa/tissue_organization)

#### A.1.4 Features

The TITSO framework offers the following features:

- Filtering.
- Nuclei segmentation.
- SLIC clustering.
- Watershed segmentation.
- Compute cell zone.
- Compute cell type.
- Analyze cellular interaction.
- Spatial statistic.
- Clusters analysis.

— Visualization of graph.

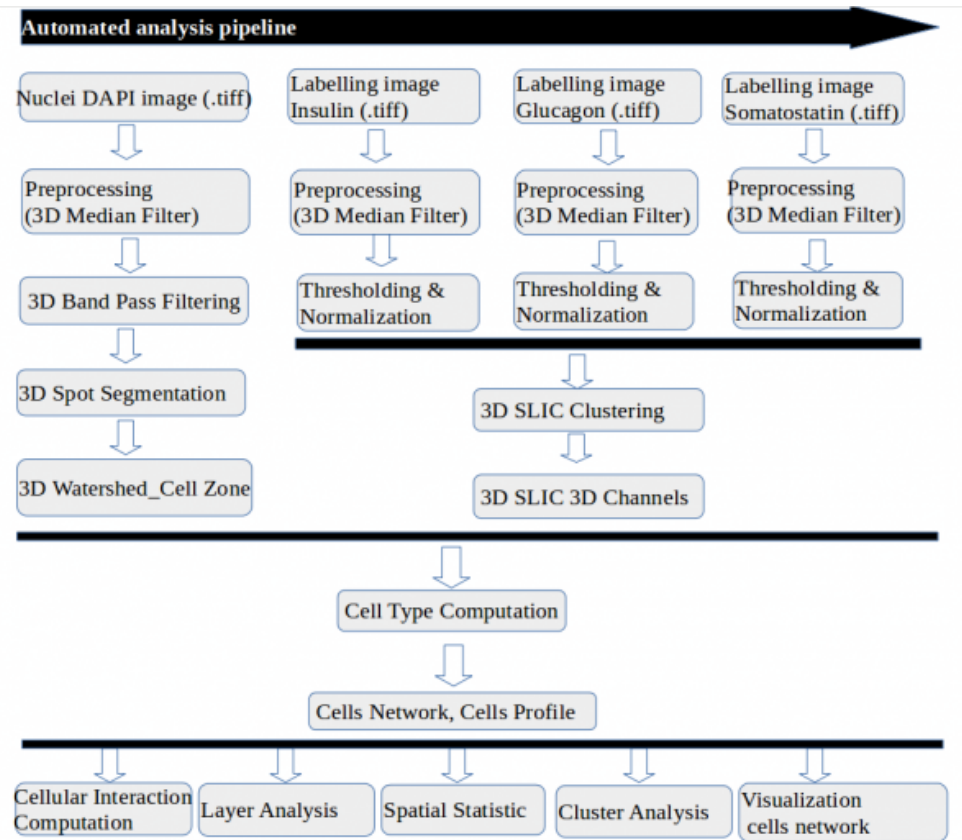


FIGURE A.1: Workflow of the TITSO (the Toolbox to Investigate Tissue Spatial Organization).

### A.1.5 Pre-filtering and normalize labelling image

Before processing the marker - labelling image, we need to filter and normalize the intensity of the input image.

- From ImageJ go to Plugin then go to 3D Tissue Spatial Analysis, sub menu Pre-filtering Input: a composite image which contains different channels and also DAPI nuclei image.
- Input parameters. Choose the label correspond to each image. It can be M1, M2, M3 (markers, cytoplasmic to determine cell type) or DAPI (nuclei), or \*NONE\* if number of channels is smaller number of options provided
- Output. Filtered images of different channels and save it into folder. The output images will be save into a folder named ‘filtered’ with the rename of images are C1.tif, C2.tif, C3.tif, C4-dapi.tif.



- During the normalization phase, select image and its input threshold value. The framework will remove all voxels whose intensity values are below this threshold and normalize the input image within the range [0, 255]. This step removes noise in 3D microscopy image.

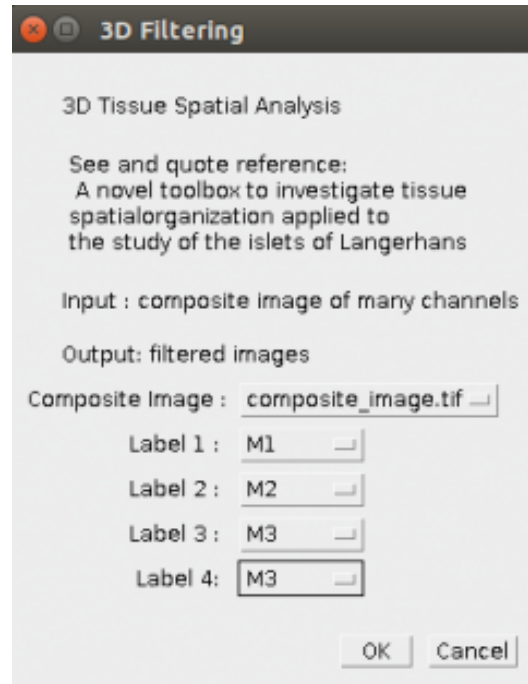


FIGURE A.2: Pre-filtering and normalize labelling image sub menu.

### A.1.6 Nuclei segmentation

The next step is to detect nuclei in the image by applying the appropriate segmentation algorithm. Images corresponding to nuclear labeling (DAPI) will first be processed by 3D median filter to reduce noise and homogenized intensities inside nuclei. After that, these images will be filtered by a 3D band pass filter which uses the approximated diameter of the nuclei region. The output are 32-bit images, which we then scale down to 16-bit. Our 3D Spot Segmentation plugin is used to segment the nuclei, where local maxima from the band pass filtering is used as seed.

A set of default parameters for segmentation is supplied but quality of nuclei image, volume of nucleus vary depending on different species and different tissue. Therefore, the toolbox allows you to adjust these input parameters and observe changes in the output.

From ImageJ go to Plugin and then to 3D Tissue Spatial Analysis, choose sub menu NUCLEI SEGMENTATION to adjust segmentation parameters.

- Input images. A DAPI nuclei image.
- Pre-processing. Choose pre-filtering (3D Median 4-4-2) if the image has not yet been filtered.
- Maximal nucleus size, Minimal nucleus size. You need to provide the range [min-size, max-size] of nucleus radius size in the image.
- Seed threshold. Threshold used to detect seed in the Band Pass Filtered Image. Normally, we use the threshold of 30000 to extract the brightest voxel correspond to seed in the 16-bit image. If image is at a lower intensity, we can reduce this parameter accordingly.
- Output. The output is a 16-bit image of nuclei segmentation which contains different regions. Each region is indexed by one value of intensity and has a different color by applying LUT 3-3-2 RGB.

To observe the segmentation result:

- Open segmented image in 3D Manager and 3D Viewer.
- A window of 3D Manager will appear and show the list of objects, 3D Viewer show the visualization of objects in 3D.
- From 3D Manager, you can check the 3D nucleus object profile and have manual validation: detect merge objects and split them. See more of 3D Manager at the following link:[http://imagejdocu.tudor.lu/doku.php?id=plugin:stacks:3d\\_roi\\_manager:start](http://imagejdocu.tudor.lu/doku.php?id=plugin:stacks:3d_roi_manager:start).

### A.1.7 Cell zone detection

The normal way to detect a cell is by using segmented nuclei as seeds inside the membrane region. However, for 3D microscopy image, 3D membrane label is generally unavailable. In our work, we apply Watershed - region growing technique to obtain the cell zone.

From ImageJ menu go to Plugin then go to 3D Tissue Spatial Analysis plugin and sub menu CELL ZONE.

- Input. A nuclei segmented image.

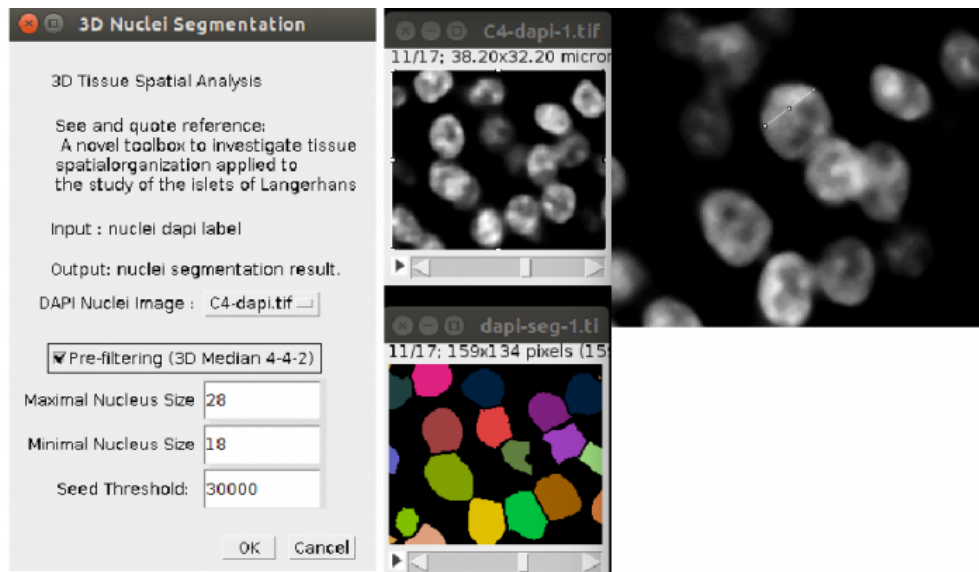


FIGURE A.3: Nuclei segmentation sub menu.

- Distance. The count from border of segmented nucleus to outside. User can define this parameter based on observed cells. If cells are close, the boundary is counted half way from maximum distance between the border of two cells.
- Output. A Watershed image which defines cell zone around each nucleus as seed.

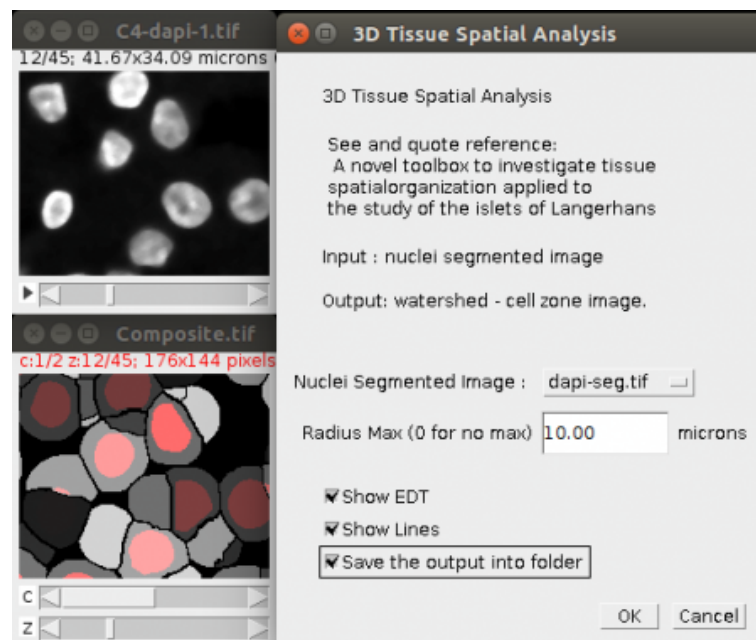


FIGURE A.4: Cell zone detection sub menu.

### A.1.8 3D SLIC clustering

In order to combine information of different markers, we use the 3D Multi-Voxels Simple Linear Iterative Clustering (SLIC) algorithm. This algorithm performs an unsupervised learning of different channels. From ImageJ menu go to Plugin then go to 3D Tissue Spatial Analysis plugin and sub menu SLIC 3D CLUSTERING.

- Input. Different channels - markerset. Choose a number of input images.
- Min size, Max sizes. Use [min\_size, max\_size] to provide the range, which is the possible volume of each clustering region. The default volume is [100, 300] which corresponds to the max size and min size of nucleus in dataset.
- Nb iterations. The number of iterations that the framework will take to perform clustering and refine the result of clustering. The default value is 10. The complexity of SLIC algorithm depends on the size XYZ of 3D image.
- Output. An image which contains many regions, each region corresponds to one cluster of voxels, named label.tiff.

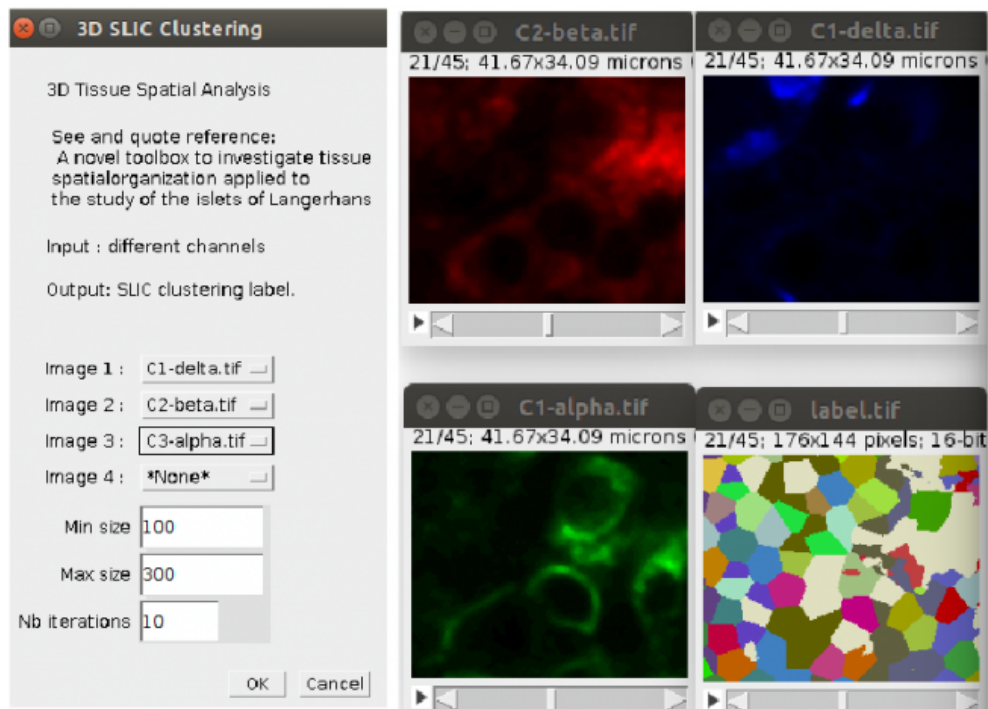


FIGURE A.5: 3D SLIC Clustering sub menu.

### A.1.9 3D SLIC 3 channels

We finally assign each zone to the channel having the highest median value inside the zone. If all the median values inside the zone are lower than predefined thresholds, the zone is assigned to the unlabelled type. An image is then created with the voxels having the value of their corresponding zone. From ImageJ menu go to Plugin then go to 3D Tissue Spatial Analysis plugin and sub menu SLIC 3D 3 CHANNELS.

- Input. Choose image correspond to different labels and SLIC label as the result of SLIC 3D CLUSTERING.
- Threshold. Input threshold value. If the median value of a SLIC region is below this value, this region won't be taken into account.
- Output. SLIC with intensity value of region correspond to only one channel (one label) in the list of existing label (M1=1, M2=2, M3=3).

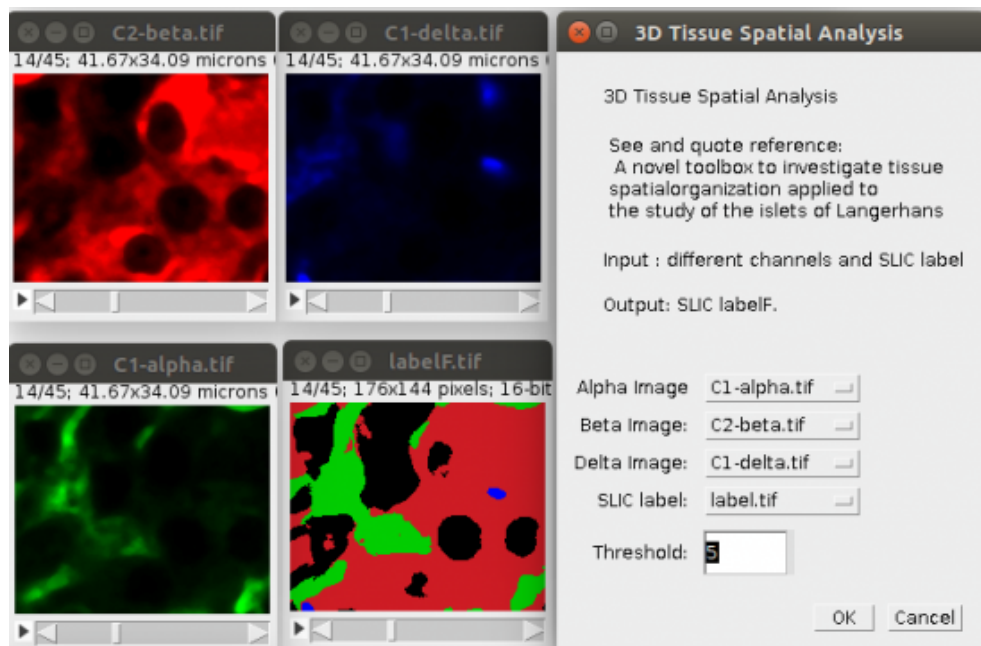


FIGURE A.6: 3D SLIC 3 Channels sub menu.

### A.1.10 Cell type detection

Take into account different regions and use intensity value of three channels and cell profile that we obtain from the previous step, we compute the cell type of each cell by quantifying the marker inside each cell zone. From ImageJ menu go to Plugin then go to 3D Tissue Spatial Analysis plugin and sub menu CELL TYPE DETECTION.

- Inputs. A watershed - cell zone image, segmented nuclei image, and SLIC labelF image.
- Region observed: INSIDE\_OUTSIDE NUCLEUS, OUTSIDE NUCLEUS, INSIDE NUCLEUS, WATERSHED REGION.
- Ratio of histogram of label.
- Min distance.
- Maximum distance inside of nuclei region take into account.
- Maximum distance outside of nuclei region take into account.
- Output. Cell type image.
- Cell profile. Nuclei.zip and Region.zip.

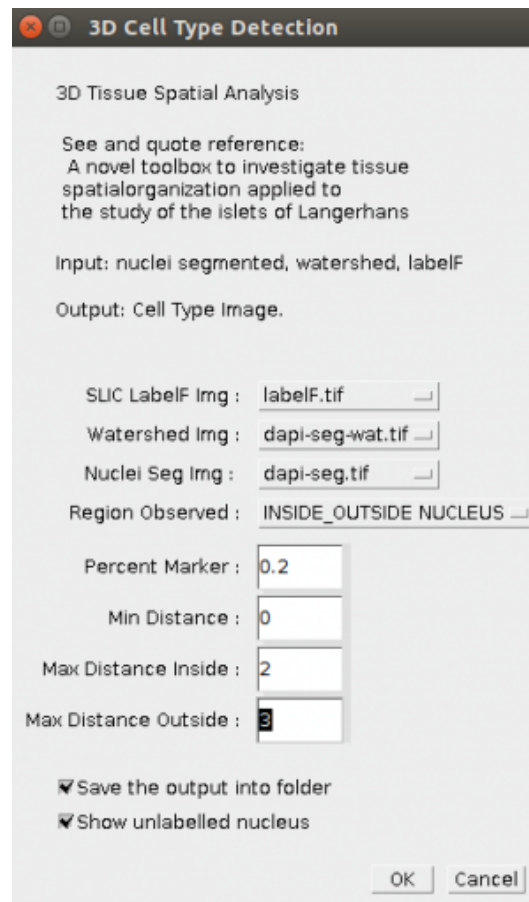


FIGURE A.7: Cell Type Detection sub menu.

### A.1.11 Cellular interaction computation

We define cell-cell contact based on nucleus zone as follows: Two cells are neighbor if their nucleus zones are touching, which means that they have the voxel contact. Users

can handle the number of voxels contact as input parameter in the next version.

- Input. A watershed image, 3D ROI file Nuclei.zip, Regions.zip. At the current version our plugin will read 2 files: Nuclei.zip and Regions.zip are located in the same folder with watershed image.
- Choose options. Provide different options for computation: cell to cell contact, layer contact, histogram contact between beta cells and alpha cells, cells network visualization.
- Output. Excel files: AllContact.csv, LayerContact.csv, histogram\_beta\_alpha.csv, Image of cells network visualization.

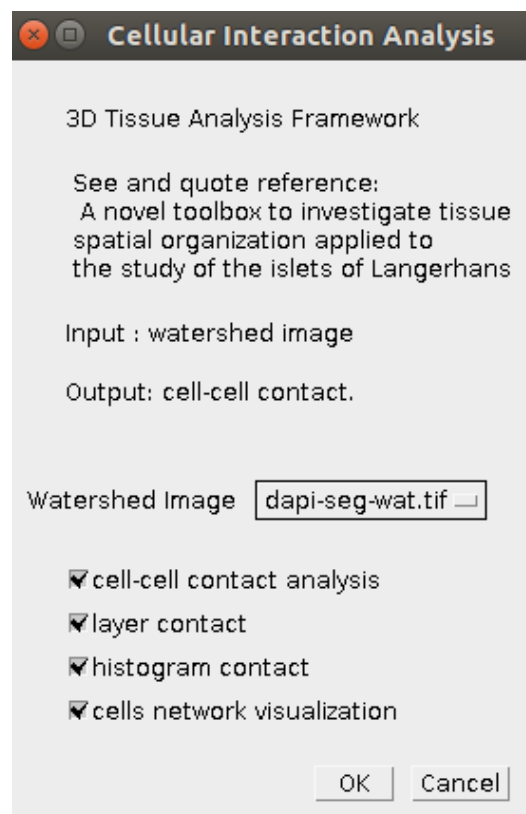


FIGURE A.8: Cellular interaction computation sub menu.

### A.1.12 Random organization

The program will compare the cellular interaction between the observed raw data and the average of 100 simulated random organization models. From ImageJ menu go to Plugin then go to 3D Tissue Spatial Analysis plugin and sub menu RANDOM ORGANIZATION.

Input:

- Input. A watershed image.
- Choose one of cell type observed: TYPE1, TYPE2, TYPE3.
- Choose source cell type observed: TYPE1, TYPE2, TYPE3.
- Number of random organization. By default it is 100 times simulated random organization.

Output:

- The number of interactions of observed raw data.
- The number of interactions of 100 simulated models and average.

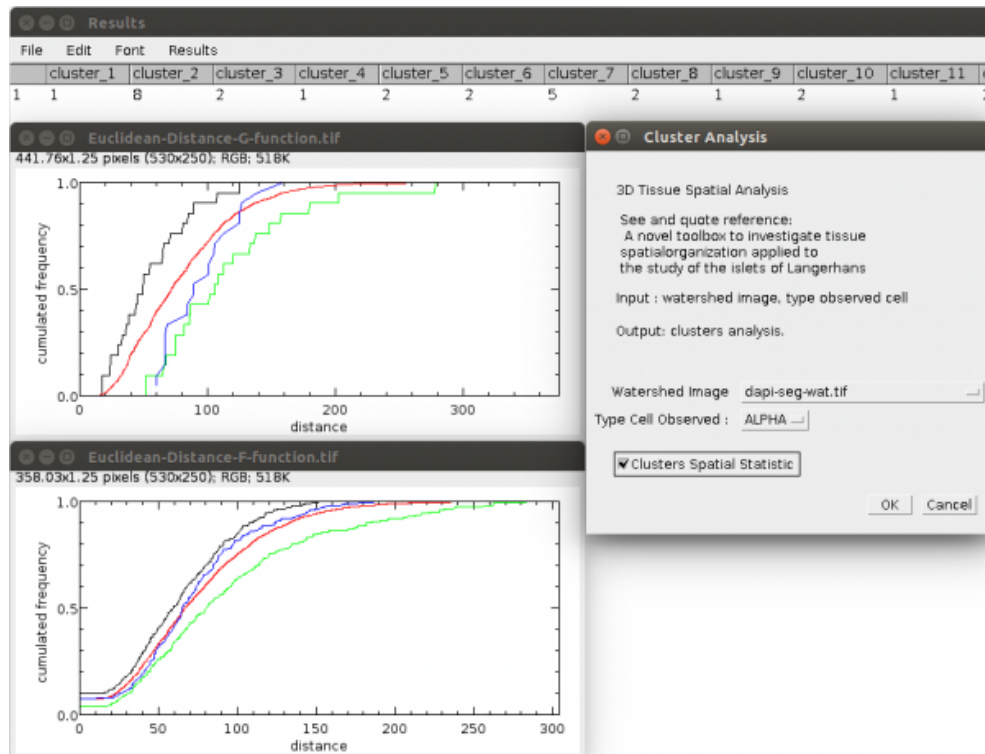


FIGURE A.9: Random Organization sub menu.

### A.1.13 Cluster analysis

From ImageJ menu go to Plugin then go to 3D Tissue Spatial Analysis plugin and sub menu CLUSTER ANALYSIS.

Input:

- Input. A watershed image.
- Choose the type cell observed: TYPE1, TYPE2, TYPE3.
- Choose whether to compute Clusters Spatial Statistic.



The output is an Excel file Cluster.csv which includes the following computation results: The index cluster, the number of observed cells in each clusters, the number of clusters, the average cells in total clusters, the max number of cells in the biggest cluster, the graphs of spatial statistic F-function and G-function of clusters and SDI values.

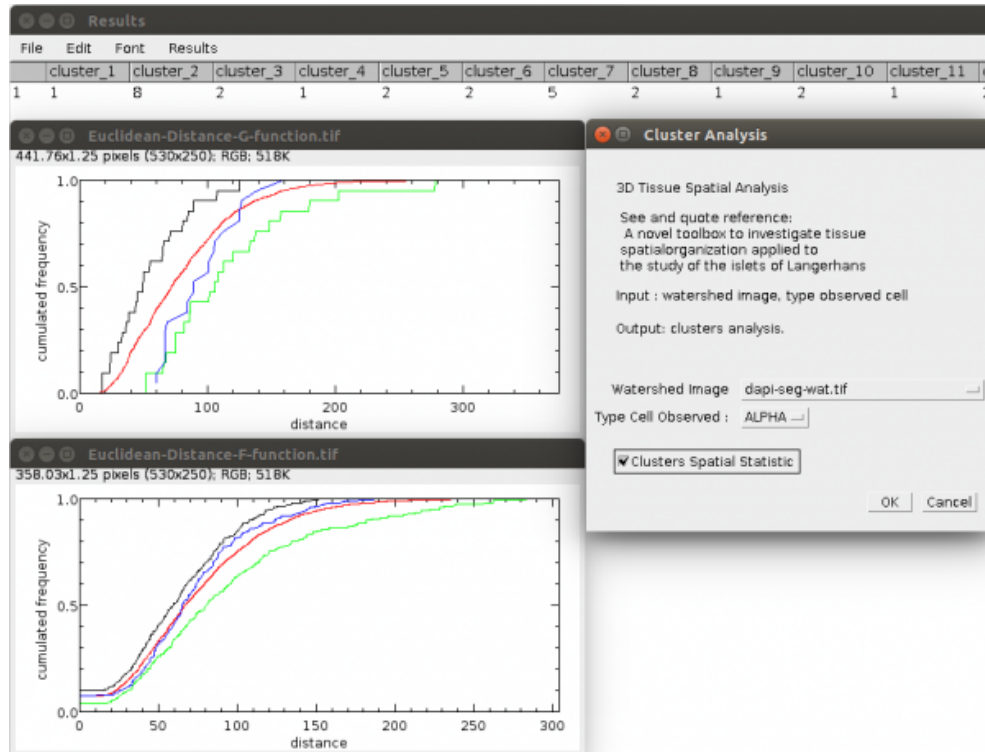


FIGURE A.10: Random Organization sub menu.

#### A.1.14 Spatial statistic analysis

In order to compute the spatial statistic of organization model, from ImageJ menu go to Plugin then go to 3D Tissue Spatial Analysis plugin and sub menu SPATIAL STATISTIC.

Input:

- Input. A watershed image.
- Cell Profiles: Nuclei.zip, Region.zip.
- Choose cell type observed: TYPE1, TYPE2, TYPE3.
- Choose cell type source: TYPE1, TYPE2, TYPE3.
- Choose type distance: Euclidean Distance, Cell Distance.
- Choose model for compute spatial statistic: Raw Data, Random Organization, Clusters.

- Choose spatial statistic function to compute: F-function, G-function.

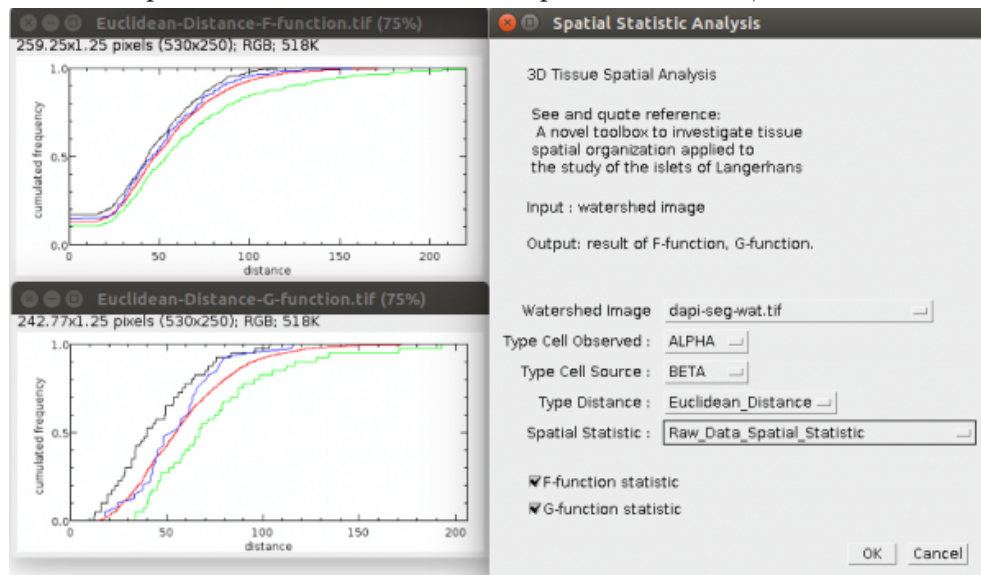


FIGURE A.11: Spatial Statistic Analysis sub menu.

#### A.1.15 Modeling and simulation of dynamic morphology of a cell

Observing the confocal microscopic images of cells within tissue reveals that there are some specific cells which are elongated and, in most cases, have a well-defined filopodia-like. Because of the elongated shape, these specific cells can reach out and contact numerous other cells in the tissue. We use our framework [10] to identify these observed cells and verify the physical advantage provided by the filopodia in a crowded islet environment.

In case the observed cell reach out in all directions, we define the dynamic morphology of the observed cell by an affected range with radius  $r$  - simulated physical length of filopodia area. The affected range is computed by dilating the observed cell zone with a predefined radius  $r$ , starting from the boundary of delta cell zone.

In order to model and simulate a dynamic morphology of a cell, from ImageJ menu go to Plugin then go to 3D Tissue Spatial Analysis plugin and sub menu CELL PROTRUSION.

Input:

- Input. A watershed image, original nuclei labeling image.
- Cell Profiles: Nuclei.zip, Region.zip.

- Choose cell type as target: TYPE1, TYPE2, TYPE3.
- Choose cell type as source: TYPE1, TYPE2, TYPE3. Source cell (or protrusion cell) is the cell have dynamic morphology.
- Min distance: minimum length of filopodia.
- Max distance: maximum length of filopodia.
- Choose observed cells, the cells have dynamic morphology : ALL CELL PROTRUSION, EACH CELL PROTRUSION, SPECIFIC CELL PROTRUSION.
- Choose the observed region: CELL REGION, NUCLEUS REGION. The length of extension (filopodia) is counted from cell boundary or nuclei boundary.

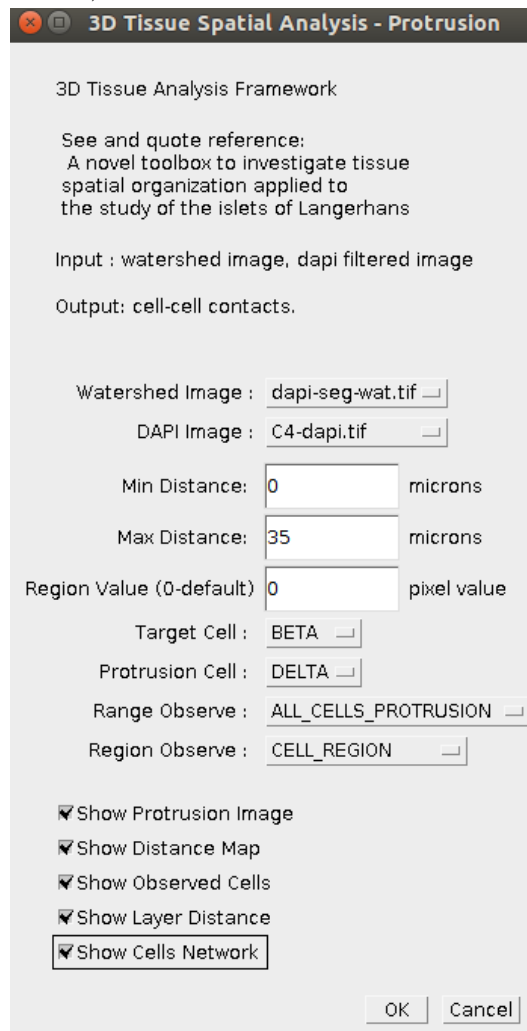


FIGURE A.12: Modeling and simulation of dynamic morphology of a cell sub menu.

## Appendix B

# Tissue Quantification and Analysis Results

Series	na	nb	nd	nabd	pa	pb	pd	aa	bb	dd	ab	ad	bd	all-contacts	paa	pbb	pdd	pab	pad	pbd	P-homo typic	P-hetero typic	delta-hetero typic
Mouse1-A	15	165	22	202	0.074	0.817	0.109	7	282	16	24	9	32	370	0.019	0.762	0.043	0.065	0.024	0.086	0.824	0.175	0.11
Mouse1-B	124	261	44	429	0.289	0.608	0.103	142	623	50	228	82	81	1206	0.118	0.517	0.041	0.189	0.068	0.067	0.676	0.324	0.135
Mouse1-C	19	269	41	329	0.058	0.818	0.125	7	566	45	32	26	71	747	0.009	0.758	0.06	0.043	0.035	0.095	0.827	0.173	0.13
201564-Mouse2-S6	118	398	9	525	0.225	0.758	0.017	228	1354	2	414	30	36	2064	0.11	0.656	9.690E-4	0.201	0.015	0.017	0.766	0.233	0.032
201564-Mouse3-S8	146	384	20	550	0.265	0.698	0.036	332	1070	17	276	110	51	1856	0.179	0.577	0.009	0.149	0.059	0.027	0.765	0.235	0.086
201430120-SeeDB-Mouse4	32	138	7	177	0.181	0.78	0.04	53	468	6	98	8	29	662	0.08	0.707	0.009	0.148	0.012	0.044	0.796	0.204	0.056
2015225-SeeDB-Mouse5-S1	23	358	5	386	0.06	0.927	0.013	19	1558	2	116	0	39	1734	0.011	0.899	0.001	0.067	0	0.022	0.911	0.089	0.022
2015225-SeeDB-Mouse6-S2	46	222	7	275	0.167	0.807	0.025	61	1037	2	200	20	48	1368	0.045	0.758	0.001	0.146	0.015	0.035	0.804	0.196	0.05
2014-B6-5-S25	41	248	13	302	0.136	0.821	0.043	33	482	6	53	6	24	604	0.055	0.798	0.01	0.088	0.01	0.04	0.863	0.137	0.05
2014-B6-4-S8	34	85	14	133	0.256	0.639	0.105	32	141	9	27	14	18	241	0.133	0.585	0.037	0.112	0.058	0.075	0.755	0.245	0.133
2014-B6-4-S11	16	137	5	158	0.101	0.867	0.032	6	231	1	27	5	10	280	0.021	0.825	0.004	0.096	0.018	0.036	0.85	0.15	0.054
2014-B6-4-S14	12	73	10	95	0.126	0.768	0.105	6	89	5	11	6	22	139	0.043	0.64	0.036	0.079	0.043	0.158	0.719	0.281	0.201
Mouse-19-S11	35	174	24	233	0.15	0.747	0.103	30	317	21	30	30	48	476	0.063	0.666	0.044	0.063	0.063	0.101	0.773	0.227	0.164
Mouse-19-S19	10	46	10	66	0.152	0.697	0.152	3	49	4	10	7	12	85	0.035	0.576	0.047	0.118	0.082	0.141	0.659	0.341	0.223
Mouse-8-S25	13	91	2	106	0.123	0.858	0.019	6	185	0	14	1	5	211	0.028	0.877	0	0.066	0.005	0.024	0.905	0.095	0.029
Mouse-8-S28	13	77	2	92	0.141	0.837	0.022	8	149	0	22	2	1	182	0.044	0.819	0	0.121	0.011	0.005	0.863	0.137	0.016

TABLE B.1: Results of cell composition, cellular interaction, homotypic, and heterotypic contact on C57/BL6 mouse islets (8 weeks of age,  $n = 6$  animals and 8 stacks)

S38 2016617																	
Observed						Theoretical						Random					
paa	pbb	pdd	pab	pad	pbd	paa	pbb	pdd	pab	pad	pbd	paa	pbb	pdd	pab	pad	pbd
0.041	0.593	0.023	0.177	0.023	0.142	0.029	0.527	0.011	0.245	0.035	0.152	0.029	0.524	0.009	0.244	0.037	0.157
S45 2016617																	
Observed						Theoretical						Random					
paa	pbb	pdd	pab	pad	pbd	paa	pbb	pdd	pab	pad	pbd	paa	pbb	pdd	pab	pad	pbd
0.088	0.545	0.015	0.278	0.031	0.044	0.041	0.558	0.002	0.304	0.02	0.074	0.04	0.56	0.003	0.306	0.02	0.072
S49 2016617																	
Observed						Theoretical						Random					
paa	pbb	pdd	pab	pad	pbd	paa	pbb	pdd	pab	pad	pbd	paa	pbb	pdd	pab	pad	pbd
0.072	0.269	0.215	0.105	0.13	0.21	0.033	0.212	0.129	0.167	0.13	0.33	0.032	0.21	0.127	0.169	0.13	0.332
S63 2016617																	
Observed						Theoretical						Random					
paa	pbb	pdd	pab	pad	pbd	paa	pbb	pdd	pab	pad	pbd	paa	pbb	pdd	pab	pad	pbd
0.015	0.677	0.054	0.111	0.016	0.127	0.006	0.654	0.013	0.125	0.018	0.185	0.006	0.658	0.017	0.125	0.018	0.177
S22 2014101																	
Observed						Theoretical						Random					
paa	pbb	pdd	pab	pad	pbd	paa	pbb	pdd	pab	pad	pbd	paa	pbb	pdd	pab	pad	pbd
0.198	0.264	0.095	0.18	0.133	0.13	0.135	0.179	0.044	0.311	0.154	0.177	0.138	0.17	0.046	0.31	0.16	0.176
S37 2014101																	
Observed						Theoretical						Random					
paa	pbb	pdd	pab	pad	pbd	paa	pbb	pdd	pab	pad	pbd	paa	pbb	pdd	pab	pad	pbd
0.054	0.577	0.088	0.131	0.035	0.115	0.021	0.508	0.021	0.205	0.041	0.205	0.019	0.518	0.025	0.196	0.041	0.202
S40 2014101																	
Observed						Theoretical						Random					
paa	pbb	pdd	pab	pad	pbd	paa	pbb	pdd	pab	pad	pbd	paa	pbb	pdd	pab	pad	pbd
0.136	0.266	0.158	0.146	0.092	0.203	0.081	0.202	0.071	0.255	0.151	0.239	0.082	0.187	0.074	0.252	0.166	0.239

TABLE B.2: Result of cellular interaction of observed models, theoretical models and simulated random models on 7 monkey islets.

Series	na	nb	nd	nabd	pa	pb	pd	aa	bb	dd	ab	ad	bd	all-contacts	paa	pbb	pdd	pab	pad	pbd	P-homo typic	P-hetero typic
<b>HP1-18</b>	85	52	3	140	0.607	0.371	0.021	162	90	2	88	11	0	353	0.459	0.255	0.006	0.249	0.031	0	0.72	0.28
<b>HP1-28</b>	73	77	56	206	0.354	0.374	0.272	112	155	90	106	113	85	661	0.169	0.234	0.136	0.16	0.171	0.129	0.539	0.46
<b>HP1-30</b>	19	88	17	124	0.153	0.71	0.137	10	151	17	40	3	25	246	0.041	0.614	0.069	0.163	0.012	0.102	0.724	0.277
<b>HP1-38</b>	24	77	7	108	0.222	0.713	0.065	29	159	5	45	3	15	256	0.113	0.621	0.02	0.176	0.012	0.059	0.754	0.247
<b>HP1-44</b>	85	72	37	194	0.438	0.371	0.191	109	118	50	181	74	66	598	0.182	0.197	0.084	0.303	0.124	0.11	0.463	0.537
<b>HP2-5</b>	278	256	34	568	0.489	0.451	0.06	456	426	18	536	85	67	1588	0.287	0.268	0.011	0.338	0.054	0.042	0.566	0.434
<b>HP2-9</b>	99	63	16	178	0.556	0.354	0.09	205	113	14	165	71	30	598	0.343	0.189	0.023	0.276	0.119	0.05	0.555	0.445
<b>HP2-13</b>	164	258	17	439	0.374	0.588	0.039	210	404	7	253	32	46	952	0.221	0.424	0.007	0.266	0.034	0.048	0.652	0.348
<b>HP2-18</b>	70	279	20	369	0.19	0.756	0.054	59	649	8	241	16	65	1038	0.057	0.625	0.008	0.232	0.015	0.063	0.69	0.31
<b>HP2-22</b>	32	52	8	92	0.348	0.565	0.087	18	61	2	23	5	19	128	0.141	0.477	0.016	0.18	0.039	0.148	0.634	0.367
<b>HP2-25</b>	24	118	2	144	0.167	0.819	0.014	11	181	0	51	1	6	250	0.044	0.724	0	0.204	0.004	0.024	0.768	0.232
<b>HP2-28</b>	100	142	16	258	0.388	0.55	0.062	80	229	16	140	10	36	511	0.157	0.448	0.031	0.274	0.02	0.07	0.636	0.364
<b>HP3-16</b>	81	298	11	390	0.208	0.764	0.028	68	445	8	228	11	28	788	0.086	0.565	0.01	0.289	0.014	0.036	0.661	0.339
<b>HP3-22</b>	27	211	7	245	0.11	0.861	0.029	15	456	5	78	7	15	576	0.026	0.792	0.009	0.135	0.012	0.026	0.827	0.173
<b>HP3-29</b>	124	199	14	337	0.368	0.591	0.042	157	453	13	290	12	23	948	0.166	0.478	0.014	0.306	0.013	0.024	0.658	0.343
<b>HP3-35</b>	182	124	7	313	0.581	0.396	0.022	298	222	2	264	19	11	816	0.365	0.272	0.002	0.324	0.023	0.013	0.639	0.36
<b>HP3-40</b>	39	81	2	122	0.32	0.664	0.016	24	96	0	68	2	5	195	0.123	0.492	0	0.349	0.01	0.026	0.615	0.385
<b>HP3-45</b>	184	190	43	417	0.441	0.456	0.103	288	363	40	372	78	107	1248	0.231	0.291	0.032	0.298	0.062	0.086	0.554	0.446
<b>HP3-49</b>	170	162	86	418	0.407	0.388	0.206	242	293	134	331	184	164	1348	0.18	0.217	0.099	0.246	0.136	0.122	0.496	0.504
<b>HP3-53</b>	28	126	33	187	0.15	0.674	0.176	19	204	46	71	21	69	430	0.044	0.474	0.107	0.165	0.049	0.16	0.625	0.374
<b>HP3-57</b>	24	99	10	133	0.18	0.744	0.075	11	133	7	33	9	7	200	0.055	0.665	0.035	0.165	0.045	0.035	0.755	0.245

TABLE B.3: Results of cell composition, cellular interaction, homotypic, and heterotypic contact on human islets ( $n = 21$  islets total,  $n = 3$  donors)

# Bibliography

- [1] The Editors of Encyclopedia Britannica. Islets of langerhans. *Journal of microscopy*, 2009. URL <https://www.britannica.com/science/islets-of-Langerhans>.
- [2] Ben T Grys, Dara S Lo, Nil Sahin, Oren Z Kraus, Quaid Morris, Charles Boone, and Brenda J Andrews. Machine learning and computer vision approaches for phenotypic profiling. *J Cell Biol*, pages jcb-201610026, 2016.
- [3] David A Randell, Gabriel Landini, and Antony Galton. Discrete mereotopology for spatial reasoning in automated histological image analysis. *IEEE transactions on pattern analysis and machine intelligence*, 35(3):568–581, 2013.
- [4] Philippe Andrey, Kiên Kiêu, Clémence Kress, Gaëtan Lehmann, Leïla Tirichine, Zichuan Liu, Eric Biot, Pierre-Gaël Adenot, Cathy Hue-Beauvais, Nicole Houba-Hérin, Véronique Duranthon, Eve Devino, Nathalie Beaujean, Valérie Gaudin, Yves Maurin, and Pascale Debey. Statistical analysis of 3d images detects regular spatial distributions of centromeres and chromocenters in animal and plant nuclei. *PLoS Comput Biol*, 6(7):1–15, 07 2010. doi: 10.1371/journal.pcbi.1000853. URL <http://dx.doi.org/10.1371%2Fjournal.pcbi.1000853>.
- [5] David J Weston, Niall M Adams, Richard A Russell, David A Stephens, and Paul S Freemont. Analysis of spatial point patterns in nuclear biology. *PloS one*, 7(5):e36841, 2012.
- [6] Domenico Bosco, Mathieu Armanet, Philippe Morel, Nadja Niclauss, Antonino Sgroi, Yannick D Muller, Laurianne Giovannoni, Géraldine Parnaud, and Thierry Berney. Unique arrangement of  $\alpha$ -and  $\beta$ -cells in human islets of langerhans. *Diabetes*, 59(5):1202–1210, 2010.
- [7] Benjamin Schmid, Johannes Schindelin, Albert Cardona, Mark Longair, and Martin Heisenberg. A high-level 3d visualization api for java and imagej. *BMC bioinformatics*, 11(1):1, 2010.
- [8] Danh-Tai Hoang, Hitomi Matsunari, Masaki Nagaya, Hiroshi Nagashima, J Michael Millis, Piotr Witkowski, Vipul Periwal, Manami Hara, and Junghyo Jo. A conserved rule for pancreatic islet organization. *PloS one*, 9(10):e110384, 2014.
- [9] Arrojo e Drigo Rafael, Jacob Stefan, Zheng Xiaofeng, Garcia-Prieto Concha F., Fukuda Masahiro, Tran Thi Nhu Hoa, Stelmashenko Olga, Martins Peçanha Flavia Leticia, Rodriguez-Diaz Rayner,



- Bushong Eric, Deerinck Thomas, Phan Sebastian, Ali Yusuf, Leibiger Ingo, Boudier Thomas, Augustine George, Ellisman Mark H., and Berggren Per-Olof. Pancreatic islet delta cells reach out to modulate beta cell function. *Nature (Revision)*, 2017.
- [10] Tran Thi Nhu Hoa, Rafael Arrojo E Drigo, Per-Olof Berggren, and Thomas Boudier. A novel toolbox to investigate tissue spatial organization applied to the study of the islets of langerhans. *Scientific Reports*, 7, 2017.
- [11] Bruce Alberts, Dennis Bray, Karen Hopkin, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Essential cell biology*. Garland Science, 2013.
- [12] Mark J. Hubley. *Human Anatomy and Physiology I, Tissue Organization*. BIO 2050, 2016. URL <http://academic.pgcc.edu/~mhubley/a&p/205/notes.htm>.
- [13] R. Abu-Eid and G. Landini. Morphometrical differences between pseudo-epitheliomatous hyperplasia in granular cell tumours and squamous cell carcinomas. *Histopathology*, 48(4):407–416, 2006. ISSN 1365-2559. doi: 10.1111/j.1365-2559.2006.02350.x. URL <http://dx.doi.org/10.1111/j.1365-2559.2006.02350.x>.
- [14] Rafael Arrojo e Drigo, Yusuf Ali, Juan Diez, Dinesh Kumar Srinivasan, Per-Olof Berggren, and Bernhard O. Boehm. New insights into the architecture of the islet of langerhans: a focused cross-species assessment. *Diabetologia*, 58(10):2218–2228, 2015. ISSN 1432-0428. doi: 10.1007/s00125-015-3699-0. URL <http://dx.doi.org/10.1007/s00125-015-3699-0>.
- [15] Christopher S. Bjornsson, Gang Lin, Yousef Al-Kofahi, Arunachalam Narayanaswamy, Karen L. Smith, William Shain, and Badrinath Roysam. Associative image analysis: A method for automated quantification of 3d multi-parameter images of brain tissue. *Journal of Neuroscience Methods*, 170(1):165 – 178, 2008. ISSN 0165-0270. doi: <http://dx.doi.org/10.1016/j.jneumeth.2007.12.024>. URL <http://www.sciencedirect.com/science/article/pii/S0165027008000034>.
- [16] Over Cabrera, Dora M. Berman, Norma S. Kenyon, Camillo Ricordi, Per-Olof Berggren, and Alejandro Caicedo. The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7):2334–2339, 2006. doi: 10.1073/pnas.0510790103. URL <http://www.pnas.org/content/103/7/2334.abstract>.
- [17] Constantin Ionescu-Tirgoviste, Paul A Gagniuc, Elvira Gubceac, Liliana Mardare, Irinel Popescu, Simona Dima, and Manuella Militaru. A 3d map of the islet routes throughout the healthy human pancreas. *Scientific reports*, 5, 2015.
- [18] EJP De Koning, NL Bodkin, BC Hansen, and A Clark. Diabetes mellitus in macaca mulatta monkeys is characterised by islet amyloidosis and reduction in beta-cell population. *Diabetologia*, 36(5):378–384, 1993.
- [19] German Kilimnik, Junghyo Jo, Vipul Periwal, Mark C Zielinski, and Manami Hara. Quantification of islet size and architecture. *Islets*, 4(2):167–172, 2012.

- [20] Abraham Kim, Kevin Miller, Junghyo Jo, German Kilimnik, Pawel Wojcik, and Manami Hara. Islet architecture: a comparative study. *Islets*, 1(2):129–136, 2009.
- [21] Ananta Poudel, Jonas L Fowler, Mark C Zielinski, German Kilimnik, and Manami Hara. Stereological analyses of the whole human pancreas. *Scientific reports*, 6, 2016.
- [22] Deborah A Striegel, Manami Hara, and Vipul Periwal. The beta cell in its cluster: Stochastic graphs of beta cell connectivity in the islets of langerhans. *PLoS Comput Biol*, 11(8):e1004423, 2015.
- [23] German Kilimnik, Billy Zhao, Junghyo Jo, Vipul Periwal, Piotr Witkowski, Ryosuke Misawa, and Manami Hara. Altered Islet Composition and Disproportionate Loss of Large Islets in Patients with Type 2 Diabetes. *PLOS ONE*, 6(11):e27445, November 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0027445. URL <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0027445>.
- [24] Anne E. Carpenter, Thouis R. Jones, Michael R. Lamprecht, Colin Clarke, In Han Kang, Ola Friman, David A. Guertin, Joo Han Chang, Robert A. Lindquist, Jason Moffat, Polina Golland, and David M. Sabatini. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, 7:R100, 2006. ISSN 1474-760X. doi: 10.1186/gb-2006-7-10-r100. URL <http://dx.doi.org/10.1186/gb-2006-7-10-r100>.
- [25] Pasi Kankaanpää, Lassi Paavolainen, Silja Tiitta, Mikko Karjalainen, Joacim Päivärinne, Jonna Nieminen, Varpu Marjomäki, Jyrki Heino, and Daniel J White. Bioimagexd: an open, general-purpose and high-throughput image-processing platform. *Nature methods*, 9(7):683–689, 2012.
- [26] Dirk Drasdo, Stefan Hoehme, and Jan G Hengstler. How predictive quantitative modelling of tissue organisation can inform liver disease pathogenesis. *Journal of hepatology*, 61(4):951–956, 2014.
- [27] Hernán Morales-Navarrete, Fabián Segovia-Miranda, Piotr Klukowski, Kirstin Meyer, Hidenori Nonaka, Giovanni Marsico, Mikhail Chernykh, Alexander Kalaidzidis, Marino Zerial, and Yannis Kalaidzidis. A versatile pipeline for the multi-scale digital reconstruction and quantitative analysis of 3d tissue architecture. *Elife*, 4:e11214, 2015.
- [28] Andreas Kårsnäs, Robin Strand, Johan Doré, Thomas Ebstrup, Michael Lippert, and Kim Bjerrum. A histopathological tool for quantification of biomarkers with sub-cellular resolution. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 3(1):25–46, 2015.
- [29] Serge Beucher et al. The watershed transformation applied to image segmentation. *Scanning Microscopy - Supplement*, pages 299–299, 1992.
- [30] Jos BTM Roerdink and Arnold Meijster. The watershed transform: Definitions, algorithms and parallelization strategies. *Fundamenta informaticae*, 41(1, 2):187–228, 2000.
- [31] Carlos Arteta, Victor Lempitsky, J Alison Noble, and Andrew Zisserman. Learning to detect cells using non-overlapping extremal regions. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 348–356. Springer, 2012.

- [32] Yousef Al-Kofahi, Wiem Lassoued, William Lee, and Badrinath Roysam. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Transactions on Biomedical Engineering*, 57(4):841–852, 2010.
- [33] Ondřej Daněk, Pavel Matula, Carlos Ortiz-de Solórzano, Arrate Muñoz-Barrutia, Martin Maška, and Michal Kozubek. Segmentation of touching cell nuclei using a two-stage graph cut model. *Image Analysis*, pages 410–419, 2009.
- [34] Xinghua Lou, Minjung Kang, Panagiotis Xenopoulos, Silvia Munoz-Descalzo, and Anna-Katerina Hadjantonakis. A rapid and efficient 2d/3d nuclear segmentation method for analysis of early mouse embryo and stem cell image data. *Stem cell reports*, 2(3):382–397, 2014.
- [35] Rambabu Chinta and Martin Wasser. Three-dimensional segmentation of nuclei and mitotic chromosomes for the study of cell divisions in live drosophila embryos. *Cytometry Part A*, 81A(1):52–64, 2012. ISSN 1552-4930. doi: 10.1002/cyto.a.21164. URL <http://dx.doi.org/10.1002/cyto.a.21164>.
- [36] Jaza Gul-Mohammed, Ignacio Arganda-Carreras, Philippe Andrey, Vincent Galy, and Thomas Boudier. A generic classification-based method for segmentation of nuclei in 3d images of early embryos. *BMC bioinformatics*, 15(1):1, 2014.
- [37] Thomas Boudier and Jean Ollion. 3d imagej suite. *ImageJ Documentation Wiki*, 2012. URL [http://imagejdocu.tudor.lu/doku.php?id=plugin:stacks:3d\\_ij\\_suite:start](http://imagejdocu.tudor.lu/doku.php?id=plugin:stacks:3d_ij_suite:start).
- [38] Saad Ullah Akram, Juho Kannala, Mika Kaakinen, Lauri Eklund, and Janne Heikkilä. Segmentation of cells from spinning disk confocal images using a multi-stage approach. In *Asian Conference on Computer Vision*, pages 300–314. Springer, 2014.
- [39] Benjamin Schmid Johannes Schindelin, Francois Kuzstos. Segmentation editor. *ImageJ Documentation Wiki*, 2007. URL [https://imagej.net/Segmentation\\_Editor](https://imagej.net/Segmentation_Editor).
- [40] Theo E Schouten, Harco C Kuppens, and Egon L van den Broek. Three-dimensional fast exact euclidean distance (3d-feed) maps. In *Vision Geometry XIV*, volume 6066, page 60660F. International Society for Optics and Photonics, 2006.
- [41] Norberto Malpica, Carlos Ortiz de Solórzano, Juan José Vaquero, Andrés Santos, Isabel Vallcorba, José Miguel Garcia-Sagredo, and Francisco del Pozo. Applying watershed algorithms to the segmentation of clustered nuclei. 1997.
- [42] K Keraudren, M Spitaler, V Braga, D Rueckert, and L Pizarro. Two-step watershed segmentation of epithelial cells. In *Proceedings of the 6th International Workshop on Microscopic Image Analysis with Applications in Biology*, 2011.
- [43] Johannes Stegmaier, Fernando Amat, William C Lemon, Katie McDole, Yinan Wan, George Teodoro, Ralf Mikut, and Philipp J Keller. Real-time three-dimensional cell segmentation in large-scale microscopy data of developing embryos. *Developmental cell*, 36(2):225–240, 2016.

- [44] Thouis Jones, Anne Carpenter, and Polina Golland. Voronoi-based segmentation of cells on image manifolds. *Computer Vision for Biomedical Image Applications*, pages 535–543, 2005.
- [45] Andres Kriete, Elisabeth Papazoglou, Bahar Edrissi, Harriet Pais, and Kambiz Pourrezaei. Automated quantification of quantum-dot-labelled epidermal growth factor receptor internalization via multiscale image segmentation. *Journal of microscopy*, 222(1):22–27, 2006.
- [46] Sara Kaliman, Christina Jayachandran, Florian Rehfeldt, and Ana-Sunčana Smith. Limits of applicability of the voronoi tessellation determined by centers of cell nuclei to epithelium morphology. *Frontiers in physiology*, 7, 2016.
- [47] Danai Laksameethanasan, Rui Zhen Tan, Geraldine Wei-Ling Toh, and Lit-Hsin Loo. cellxpress: a fast and user-friendly software platform for profiling cellular phenotypes. *BMC bioinformatics*, 14(16):S4, 2013.
- [48] Nicola Bougen-Zhukov, Sheng Yang Loh, Hwee Kuan Lee, and Lit-Hsin Loo. Large-scale image-based screening and profiling of cellular phenotypes. *Cytometry Part A*, 91(2):115–125, 2017.
- [49] Oliver Dürr and Beate Sick. Single-cell phenotype classification using deep convolutional neural networks. *Journal of biomolecular screening*, 21(9):998–1003, 2016.
- [50] Céline Hudelot, Jamal Atif, and Isabelle Bloch. Fuzzy spatial relation ontology for image interpretation. *Fuzzy Sets and Systems*, 159(15):1929–1951, 2008.
- [51] Meng-Tsen Ke, Satoshi Fujimoto, and Takeshi Imai. Seedb: a simple and morphology-preserving optical clearing agent for neuronal circuit reconstruction. *Nature neuroscience*, 16(8):1154–1161, 2013.
- [52] Nicolas Heck, Marc Dos Santos, Brahim Amairi, Marine Salery, Antoine Besnard, Etienne Herzog, Thomas Boudier, Peter Vanhoutte, and Jocelyne Caboche. A new automated 3d detection of synaptic contacts reveals the formation of cortico-striatal synapses upon cocaine treatment in vivo. *Brain Structure and Function*, 220(5):2953–2966, 2015.
- [53] Moll Samuel. Lipid droplet counter. *ImageJ Documentation Wiki*, 2015. URL [http://imagejdocu.tudor.lu/doku.php?id=plugin:analysis:droplet\\_counter:start](http://imagejdocu.tudor.lu/doku.php?id=plugin:analysis:droplet_counter:start).
- [54] Jean Ollion, Julien Cochenec, Francois Loll, Christophe Escude, and Thomas Boudier. Tango: a generic tool for high-throughput 3d image analysis for studying nuclear organization. *Bioinformatics*, page btt276, 2013.
- [55] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.
- [56] Donald B. Johnson. A note on dijkstra’s shortest path algorithm. *J. ACM*, 20(3):385–388, July 1973. ISSN 0004-5411. doi: 10.1145/321765.321768. URL <http://doi.acm.org/10.1145/321765.321768>.

- [57] Michèle Ernoul-Lange, Sonia Baconnais, Maryannick Harper, Nicola Minshall, Sylvie Souquere, Thomas Boudier, Marianne Bénard, Philippe Andrey, Gérard Pierron, Michel Kress, et al. Multiple binding of repressed mrnas by the p-body protein rck/p54. *Rna*, 18(9):1702–1715, 2012.
- [58] Sally Galbraith, James A Daniel, and Bryce Vissel. A study of clustered data and approaches to its analysis. *Journal of Neuroscience*, 30(32):10601–10608, 2010.
- [59] Robert F Murphy. Building cell models and simulations from microscope images. *Methods*, 96:33–39, 2016.
- [60] Melissa Linkert, Curtis T Rueden, Chris Allan, Jean-Marie Burel, Will Moore, Andrew Patterson, Brian Loranger, Josh Moore, Carlos Neves, Donald MacDonald, et al. Metadata matters: access to image data in the real world. *The Journal of cell biology*, 189(5):777–782, 2010.
- [61] Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. Nih image to imagej: 25 years of image analysis. *Nature methods*, 9(7):671, 2012.
- [62] L. Charles Murtaugh. Pancreas and beta-cell development: from the actual to the possible. *Development*, 134(3):427–438, 2007.
- [63] Roger H Unger and Lelio Orci. Paracrinology of islets and the paracrinopathy of diabetes. *Proceedings of the National Academy of Sciences*, 107(37):16009–16012, 2010.
- [64] Over Cabrera, M Caroline Jacques-Silva, Stephan Speier, Shao-Nian Yang, Martin Köhler, Alberto Fachado, Elaine Vieira, Juleen R Zierath, Richard Kibbey, Dora M Berman, et al. Glutamate is a positive autocrine signal for glucagon release. *Cell metabolism*, 7(6):545–554, 2008.
- [65] Rayner Rodriguez-Diaz, Robin Dando, M Caroline Jacques-Silva, Alberto Fachado, Judith Molina, Midhat H Abdulreda, Camillo Ricordi, Stephen D Roper, Per-Olof Berggren, and Alejandro Caicedo. Alpha cells secrete acetylcholine as a non-neuronal paracrine signal priming beta cell function in humans. *Nature medicine*, 17(7):888–892, 2011.
- [66] C. J. Ramnanan, D. S. Edgerton, G. Kraft, and A. D. Cherrington. Physiologic action of glucagon on liver glucose metabolism. *Diabetes, Obesity and Metabolism*, 13(s1):118–125, 2011.
- [67] Barbara Leibiger, Tilo Moede, Thusitha P Muhandiramlage, Daniel Kaiser, Pilar Vaca Sanchez, Ingo B Leibiger, and Per-Olof Berggren. Glucagon regulates its own synthesis by autocrine signaling. *Proceedings of the National Academy of Sciences*, 109(51):20925–20930, 2012.
- [68] Eckhard Lammert, Ondine Cleaver, and Douobalas Melton. Induction of pancreatic differentiation by signals from blood vessels. *Science*, 294(5542):564–567, 2001.
- [69] Marcela Brissova, Alena Shostak, Masakazu Shiota, Peter O Wiebe, Greg Poffenberger, Jeannelle Kantz, Zhongyi Chen, Chad Carr, W Gray Jerome, Jin Chen, et al. Pancreatic islet production of vascular endothelial growth factor-a is essential for islet vascularization, revascularization, and function. *Diabetes*, 55(11):2974–2985, 2006.

- 
- [70] Marcela Brissova, Alena Shostak, Corinne L Fligner, Frank L Revetta, Mary K Washington, Alvin C Powers, and Rebecca L Hull. Human islets have fewer blood vessels than mouse islets and the density of islet vascular structures is increased in type 2 diabetes. *Journal of Histochemistry & Cytochemistry*, 63(8):637–645, 2015.
- [71] Rayner Rodriguez-Diaz, Midhat H Abdulreda, Alexander L Formoso, Itai Gans, Camillo Ricordi, Per-Olof Berggren, and Alejandro Caicedo. Innervation patterns of autonomic axons in the human endocrine pancreas. *Cell metabolism*, 14(1):45–54, 2011.