Thesis Defense

Graph-based Mathematical Morphology for the Characterization of the Spatial Organization of Histological Structures in High-Content Images: Application to Tumor Microenvironment in Breast Cancer

Bassem Ben Cheikh

Ph.D Supervisor:

Daniel RACOCEANU

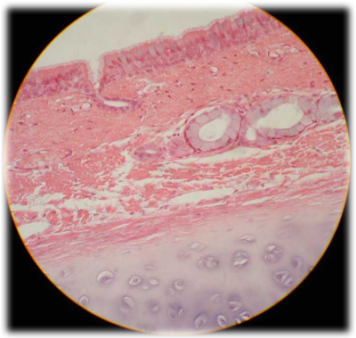
September 26th, 2017

bassem.bencheikh@lib.upmc.fr

Context

Glass slide

Classical Histopathology

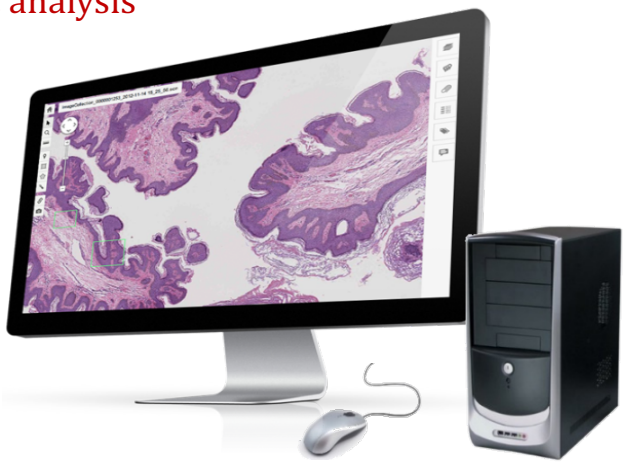


- Diagnosis
- Disease detection
- Interpretation

Digital Histopathology

scanner

- Annotation
- Automatic analysis
- Storage
- Sharing



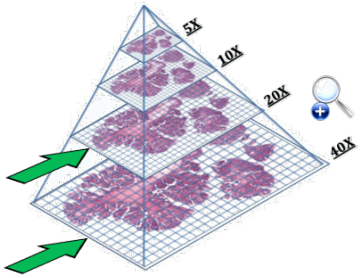
➔ **Digital Pathology** approved by the **FDA** for primary diagnosis only in April 2017 [1].

[1] FDA Press Announcements. "FDA allows marketing of first whole slide imaging system for digital pathology April 12", (2017)

Context

Whole Slide Imaging (Virtual Slides): Tiled pyramidal representation

➔ View at s

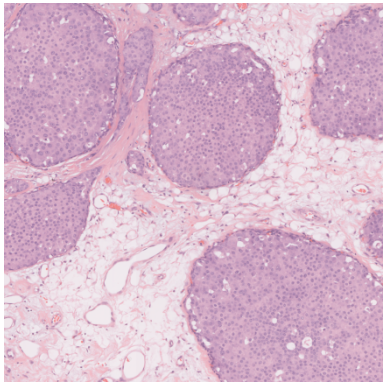


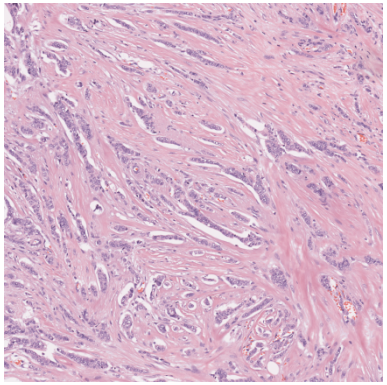
Virtual Slide 44200 X 58600 pixels

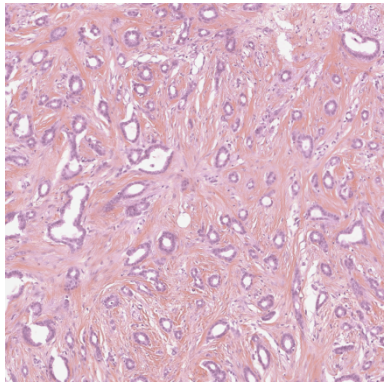
Breast tissue - Haematoxylin & Eosin stain

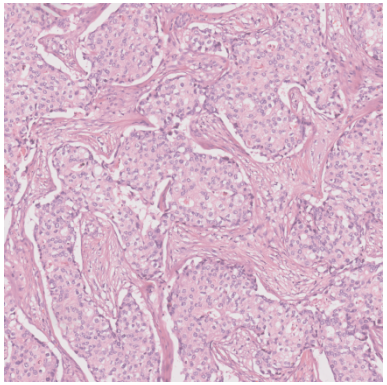
Context

Breast Cancer Histopathology


Ductal carcinoma in situ - 13%


Invasive lobular carcinoma - 5%


Tubular carcinoma - 4%


Medullary carcinoma - 3%

more subtypes: Cribriform carcinoma, Mucinous carcinoma, Papillary carcinoma, Micropapillary carcinoma, Apocrine carcinoma, Metaplastic carcinomas, Inflammatory carcinoma, etc…

55% of breast tumors do not have the "specific differentiating features"
a.k.a. Invasive carcinoma of no special type (NST)

High architectural heterogeneity

- Quantitative evaluation of tissue
- New tissue-derived parameters

Tumor patterns: Density, Size, Shape, Morphology, Sparsity, Relative distance

[2] Sinn HP et al. "**A brief overview of the WHO classification of breast tumors**". Breast Care 8 (2013)

Context

Tumor Micro-Environment (TME)

Immune cell

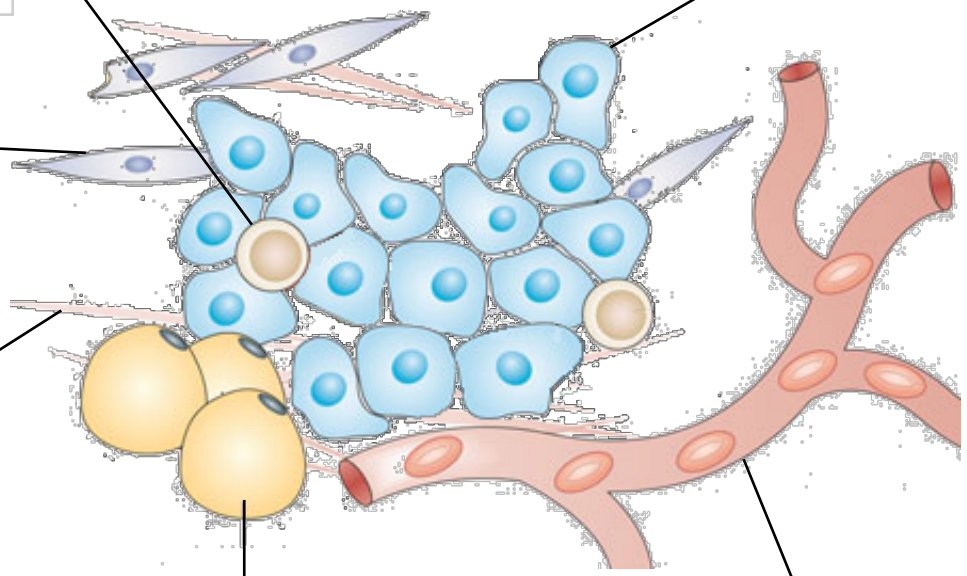
may potentially **attack** cancer cells [4]

Fibroblast

may **drive** the progression of cancer metastasis [5]

Collagen

cancer cell invasion **oriented** along aligned collagen fibres [7]



Tumor Cell

Adipocyte

fuel the growth of malignant cells [6]

Blood Vessel

supply tumor cells with oxygen and nutrients [8]

[8] DeLisser H.M. and al. "Involvement of endothelial PECAM-1/CD31 in angiogenesis". The American journal of pathology (1997)

[7] P. Provenzano et al, "*Collagen reorganization at the tumor-stromal interface facilitates local invasion"*, BMC Medicine 2006

[6] M. Wagner et al, "*Inflamed tumor-associated adipose tissue is a depot for macrophages that stimulate tumor growth and angiogenesis*", Angiogenesis, 2012.

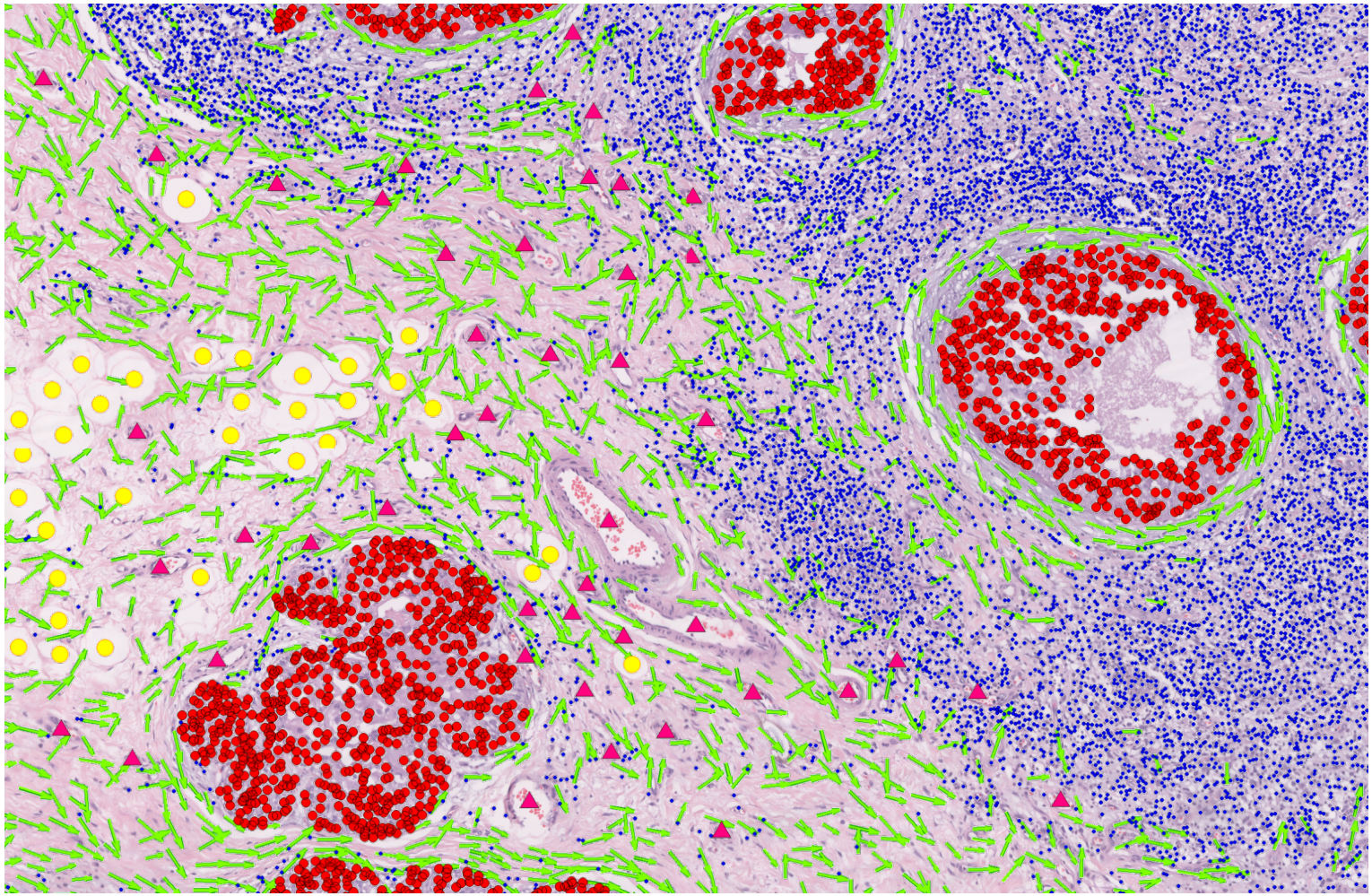
[5] G.S.K. et al, "*Cancer-Associated Fibroblasts Drive the Progression of Metastasis through both Paracrine and Mechanical Pressure on Cancer Tissue*", Mol Cr Res, 2012

[4] W. H. Fridman et al, "*The immune contexture in human tumours: impact on clinical outcome*", Nature Reviews Cancer, 2012

[3] Pattabiraman DR. and Weinberg RA. "Tackling the cancer stem cells - what challenges do they pose?" Nature reviews Drug disco (2014)

Context

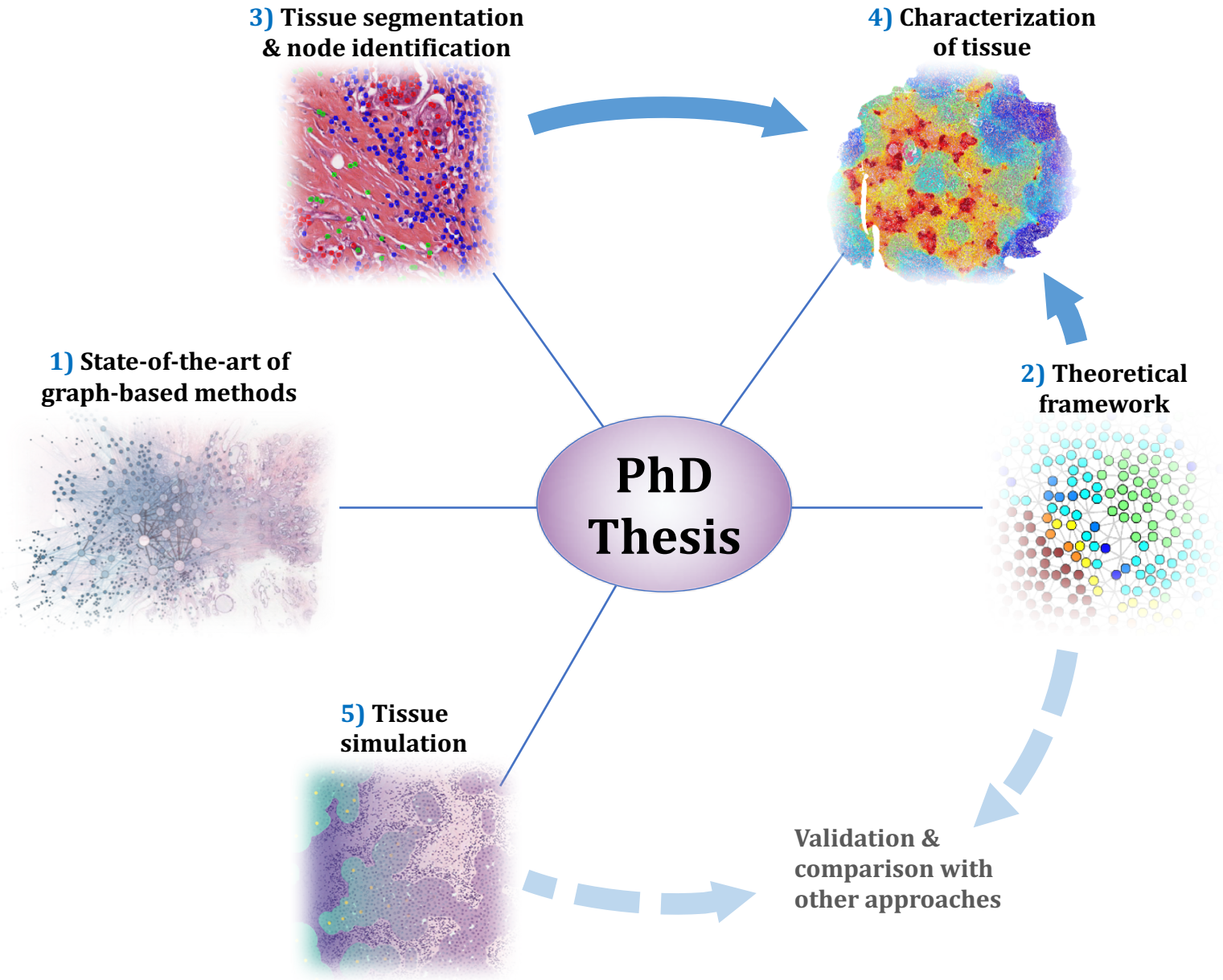
- Cancer cells
- Immune cells
- Collagen
- Adipocytes
- Blood vessels



Spatial heterogeneity in cancer ecosystem

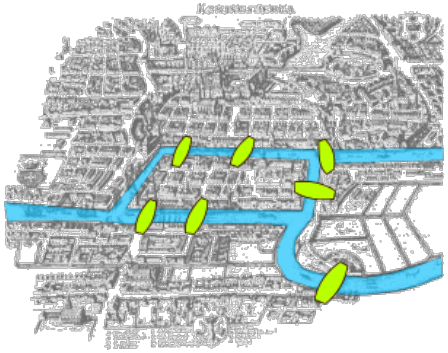
Quantitative analysis of spatial heterotypic interactions in TME: Proximity/Adjacency, Surroundedness/Encloseness, Betweenness

Content

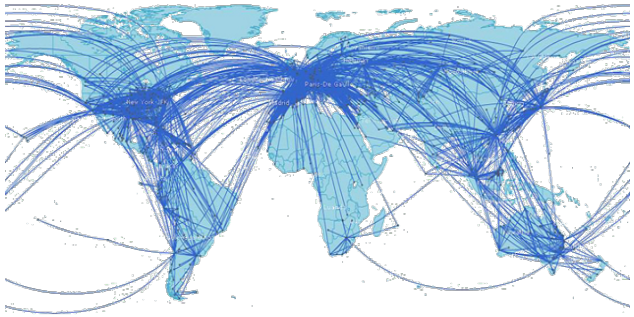

3) Tissue segmentation & node identification
4) Characterization of tissue
1) State-of-the-art of graph-based methods
2) Theoretical framework
PhD Thesis
5) Tissue simulation
Validation & comparison with other approaches

1) Graph-based methods in histopathology

Graph Theory:

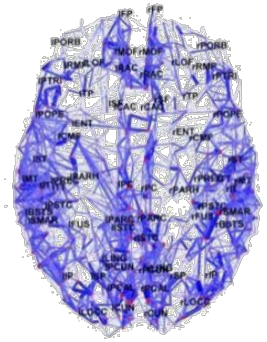


Seven Bridges of Königsberg
(L. Euler in 1736)



Air travel

Social network



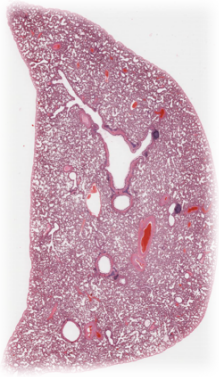
Brain function

Celestial phenomena

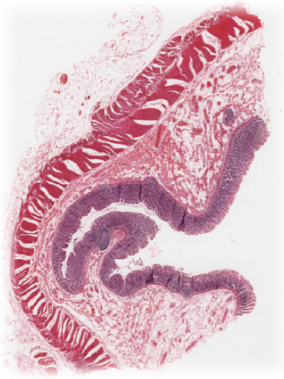
Histopathology

1) Graph-based methods in histopathology

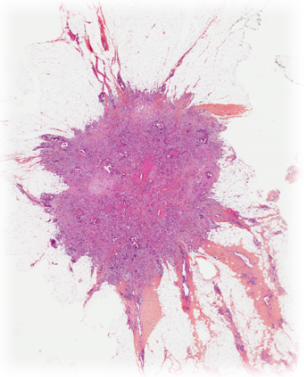
Histological Sections:



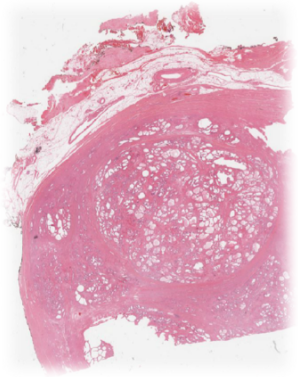
Lung



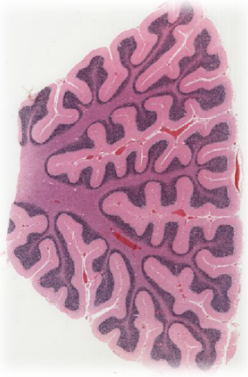
Colon



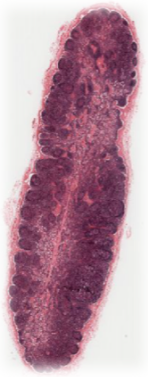
Breast



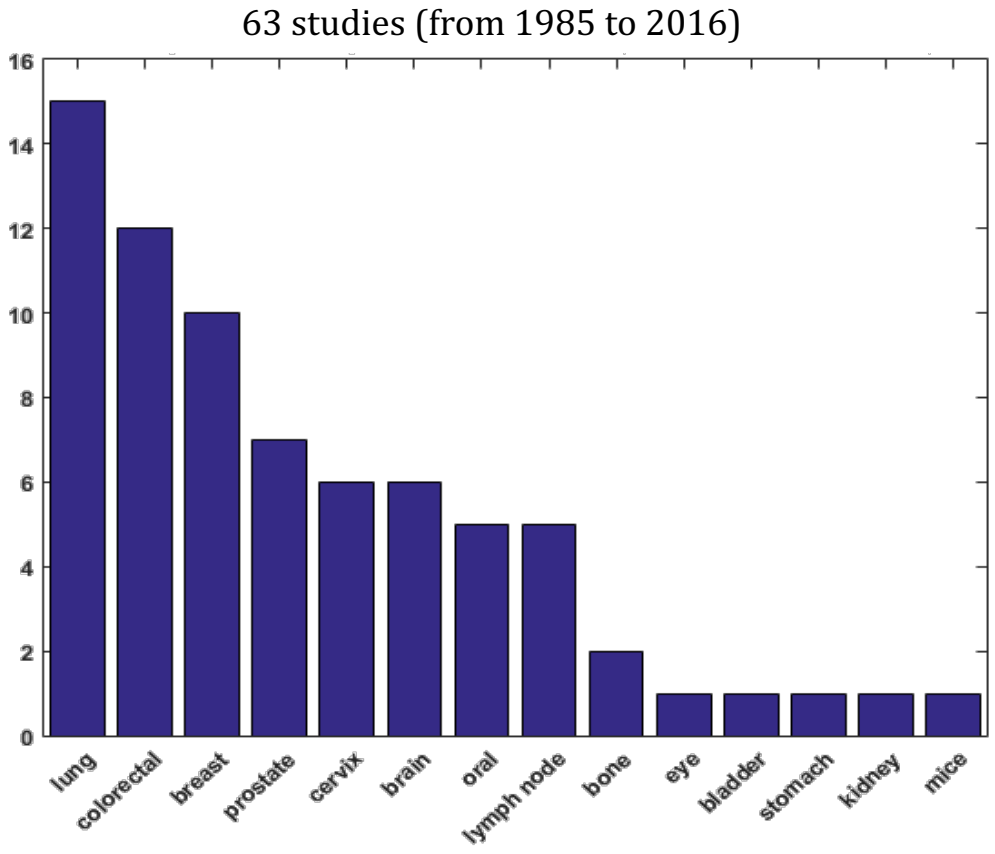
Prostate



Brain



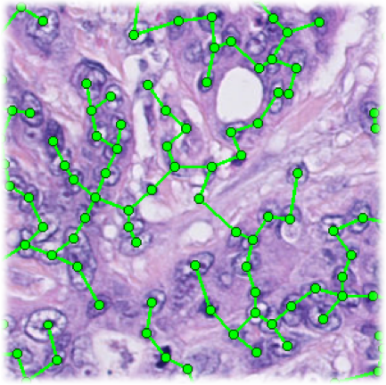
Lymph node



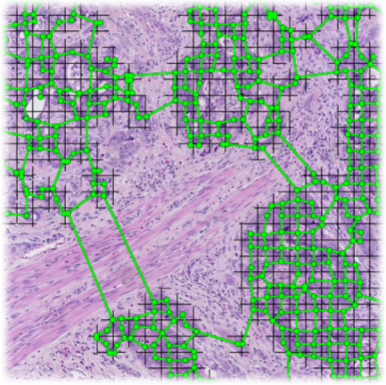
[9] Images from: University of British Columbia - Slide Box

1) Graph-based methods in histopathology

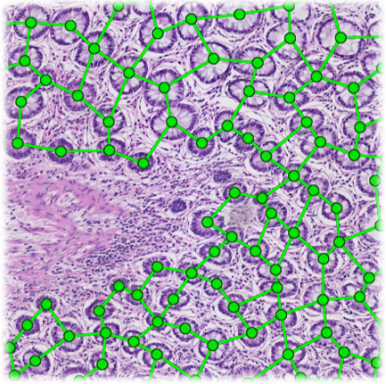
Histological Objects:



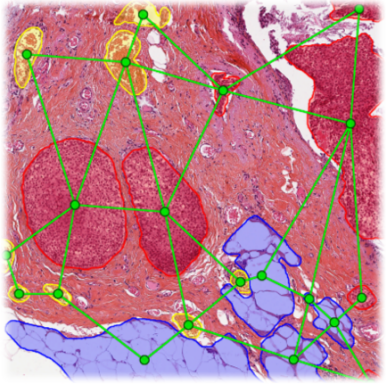
Cell Nuclei



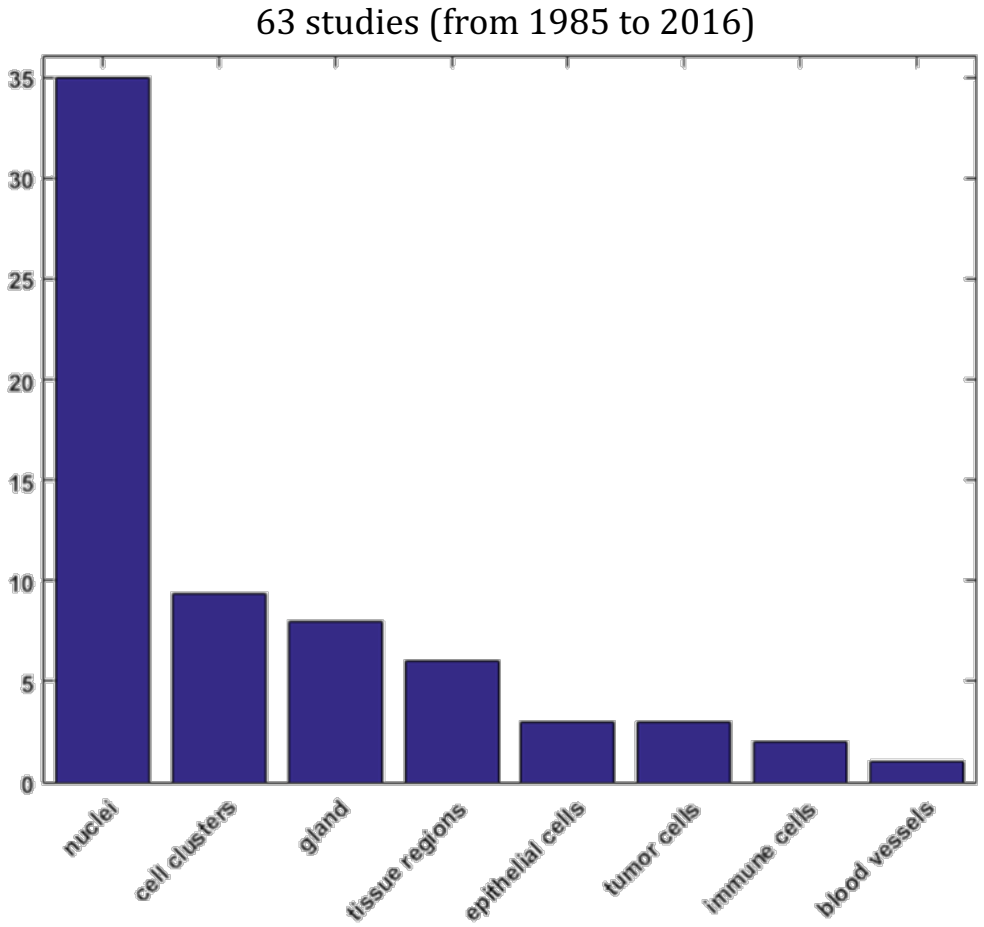
Cell clusters



Glands

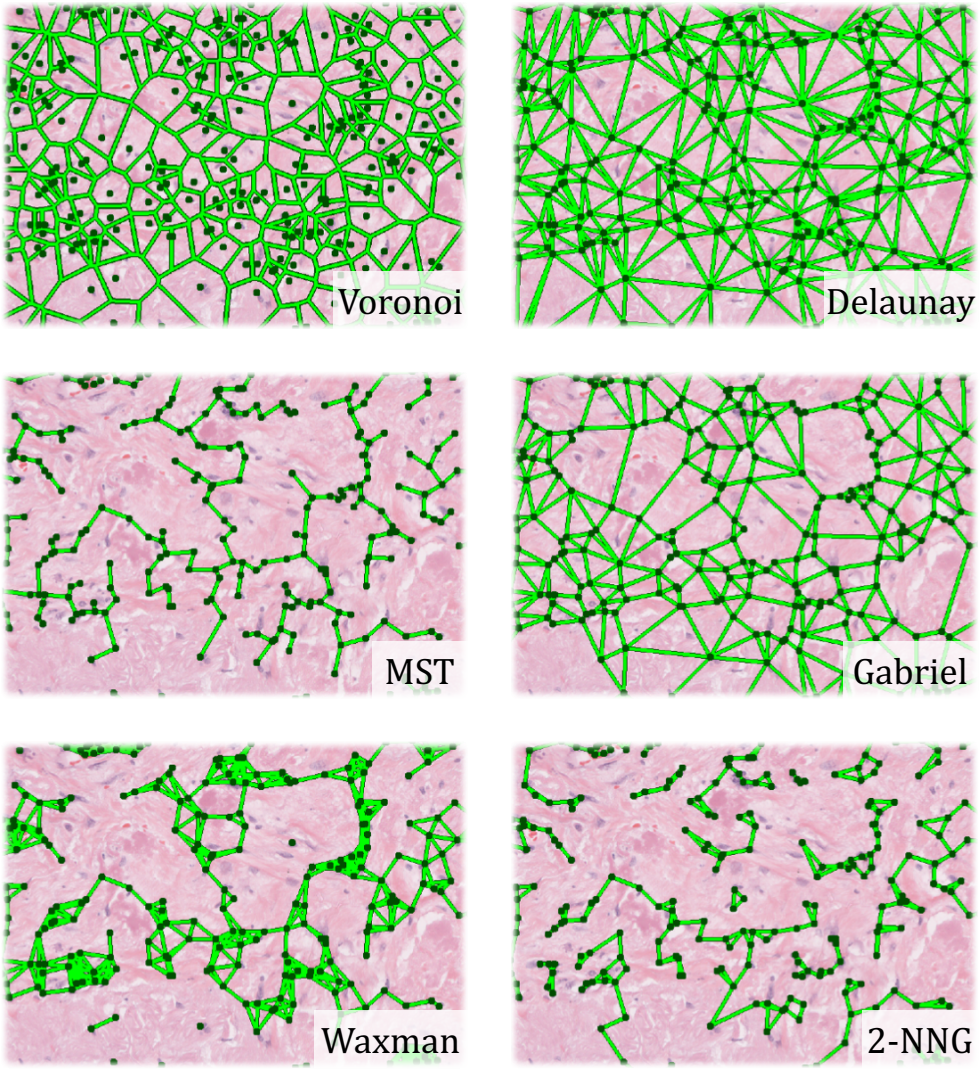


Tissue regions



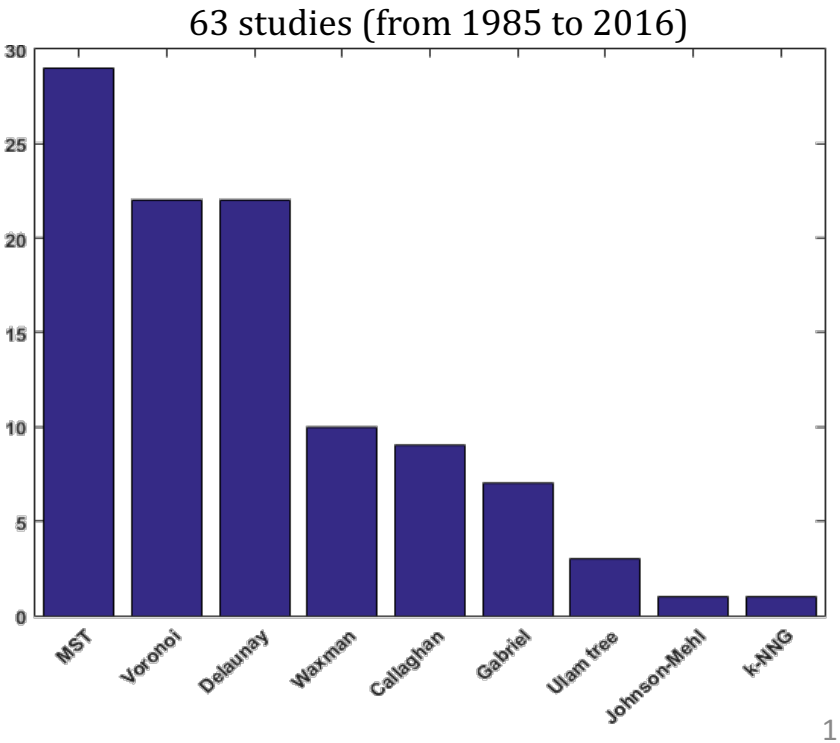
1) Graph-based methods in histopathology

Graph types:



- Delaunay = dual(Voronoi)
- MST ⊆ Gabriel ⊆ Delaunay

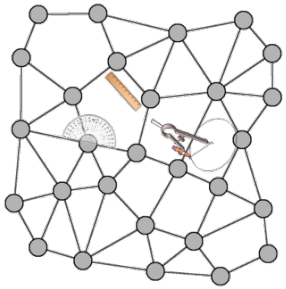
➔ Different degrees of connectedness based on geometric rules



1) Graph-based methods in histopathology

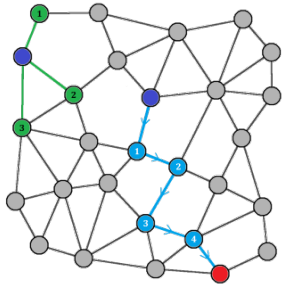
Feature extraction:

1) Syntactic Structure Analysis



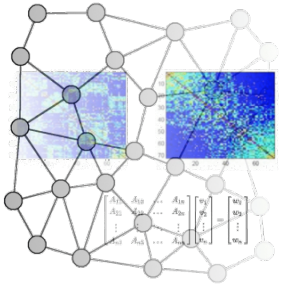
- Polygon area
- Polygon perimeter
- Polygonal size
- Polygon shape
- Polygon roundness
- Polygon compacity
- Polygon density
- Minimal to maximal side length
- Edge length
- chord length
- Minimum angle
- Maximum angle
- Nearest neighbor distance
- Distance to the k-nearest neighbors
- Nbr. of neighbors within a disk of radius r
- Weighted compacity
- Divergence from the regular number tree

2) Network Analysis



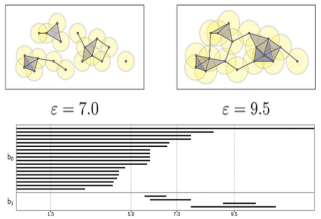
- Order
- Size (nbr. of edges)
- Average degree
- Average eccentricity
- Radius (min. ecc.)
- Diameter (max. ecc.)
- Clustering coefficient
- Closeness
- Nbr. of conn. comp.
- Giant conn. comp.
- Nbr. of isolated nodes
- Nbr. of nodes with k neighbors
- Nbr. of central vertices
- Nbr. of spanning trees
- Betweeness
- Rich club coefficient
- Cyclomatic number
- Persitence
- Girth
- Expansion

3) Spectral Analysis



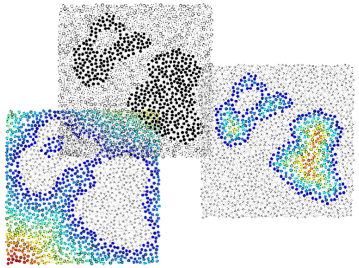
- Spectral radius
- Fiedler vector
- Cheeger constant
- Algebraic connectivity
- Principal eigenvector
- Nbr. of eigenvalues
- Second largest absolute value of eigenvalues
- Eigen exponent
- Energy
- Lower slope
- Nbr. of eignervalues equal to 1
- Nbr. of eignervalues equal to 2
- Upper-slope
- Trace
- Eigenmode perimeter
- Eigernmode volume

4) Persistent Homology

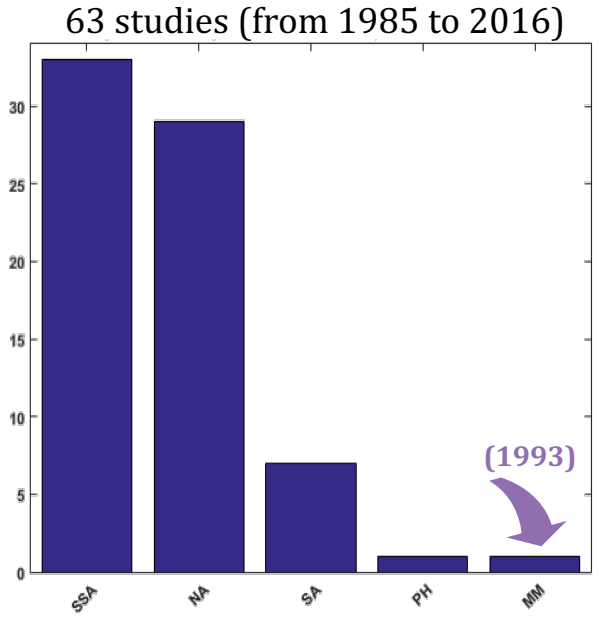


- Betti numbers

5) Mathematical Morphology



- Distance transform
- Granulometry



[10] Raymond E. et al. "Germinal center analysis with the tools of mathematical morphology on graphs". Cytometry (1993)

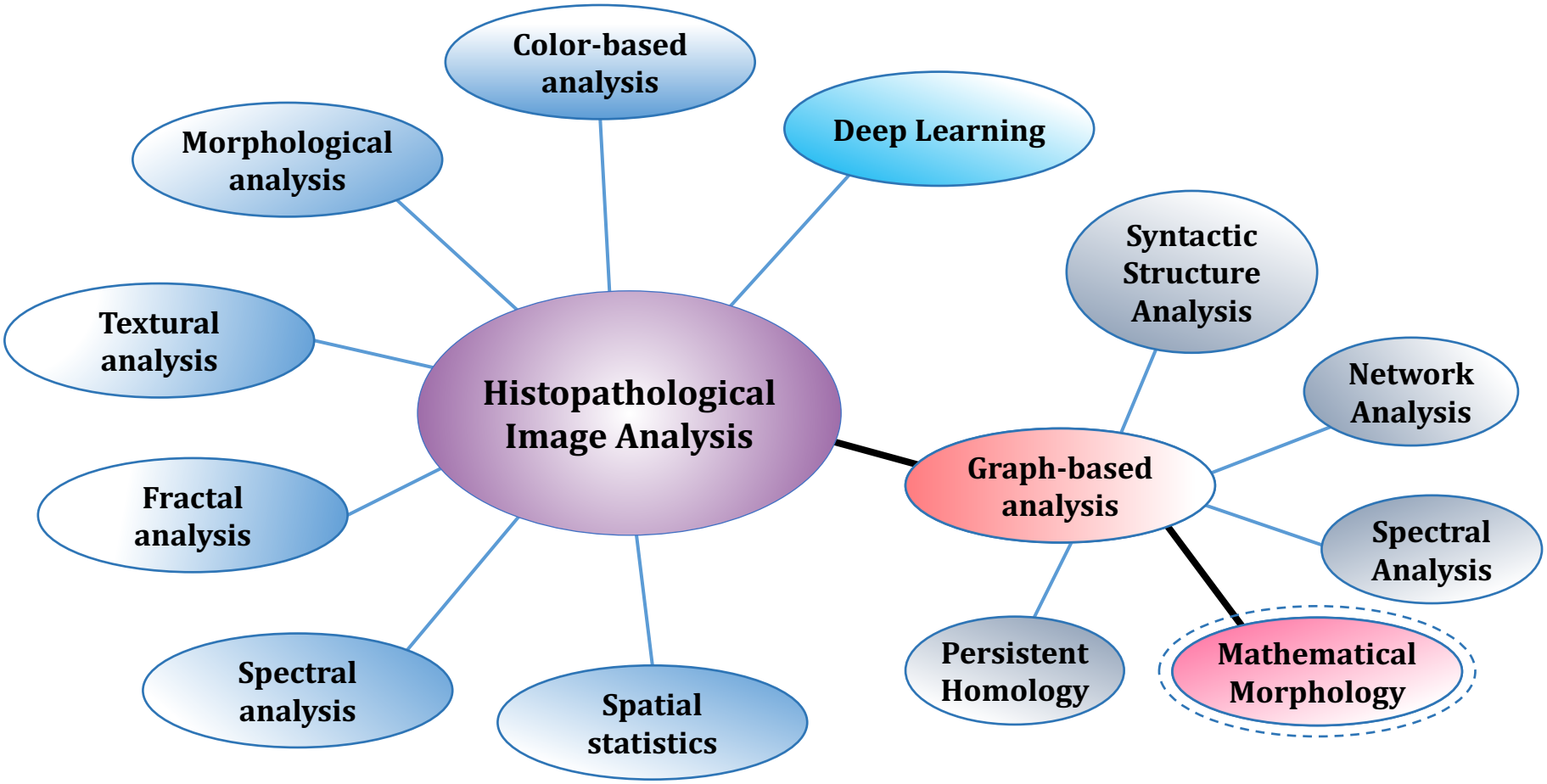
1) Graph-based methods in histopathology

Limitations:

- Analysis limited to small ROI/image frames
- Imprecision in nuclei detection/manual annotation
- Cells are treated indiscriminately
- High computational cost of some features

Requirements:

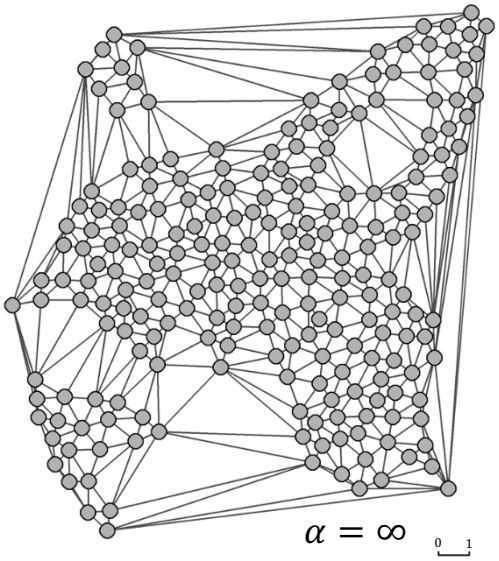
- Global analysis: Whole Slide Image
- Heterotypic spatial interactions (TME)
- Nuclei detection & classification
- More holistic approaches for feature extraction

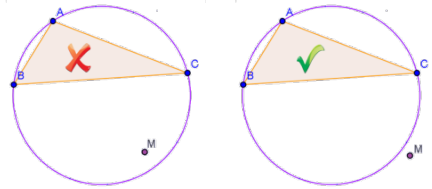

Color-based analysis
Morphological analysis
Deep Learning
Textural analysis
Histopathological Image Analysis
Syntactic Structure Analysis
Network Analysis
Graph-based analysis
Fractal analysis
Spectral Analysis
Spectral analysis
Spatial statistics
Persistent Homology
Mathematical Morphology

2) Spatial Point Analysis using Mathematical Morphology

Graph reconstruction:

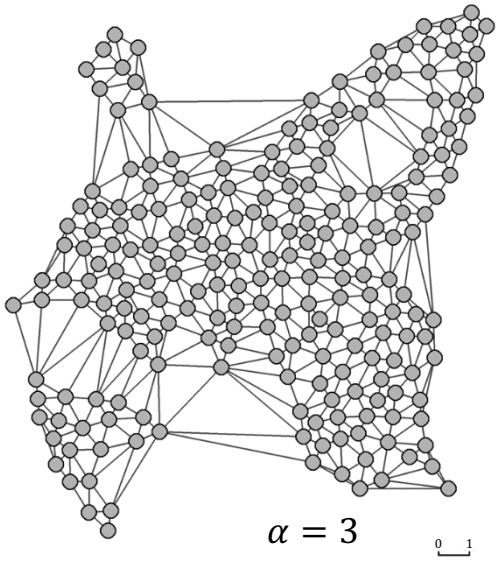
Delaunay graph:

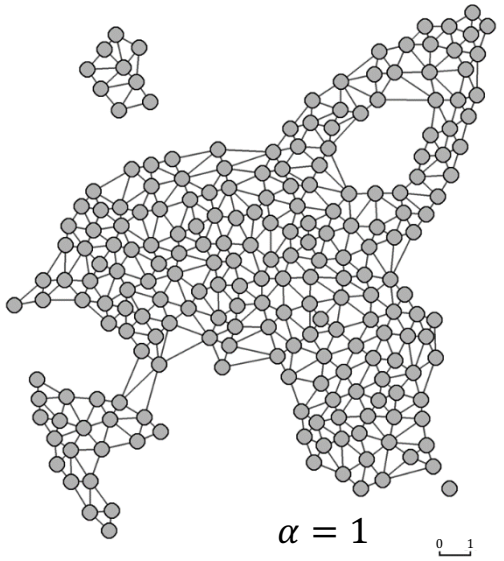


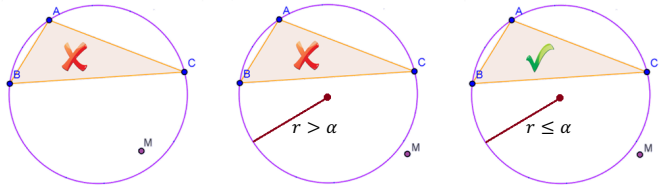


➔High degree of connectedness
➔Distant points may be connected

α-Shape filtering: sculpting Delaunay graph



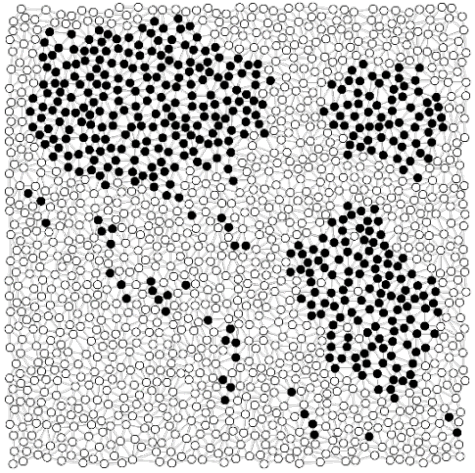




➔High degree of connectedness
➔~~Distant points may be related~~

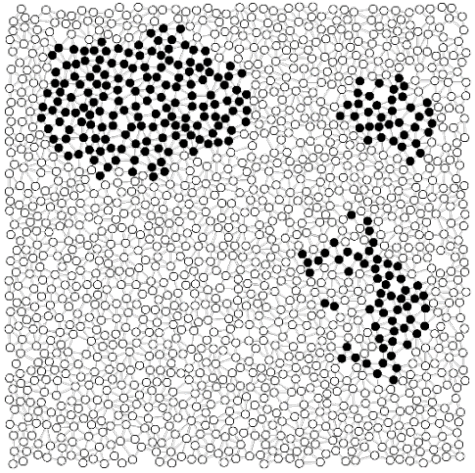
2) Spatial Point Analysis using Mathematical Morphology

Mathematical Morphology on Graphs:

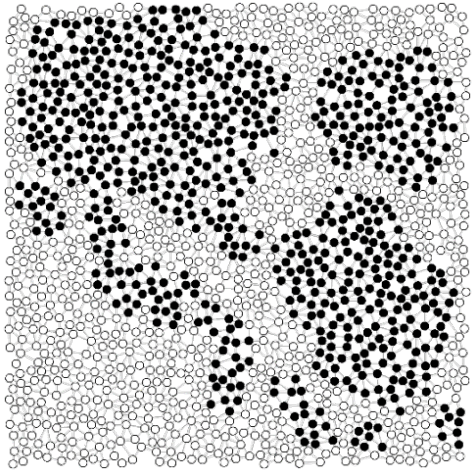


Binary graph: $G(V, E)$

$\circ = 0$, $\bullet = 1$



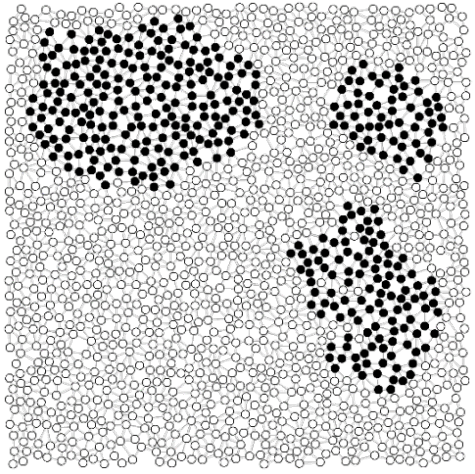
Erosion $\varepsilon_1(G)$



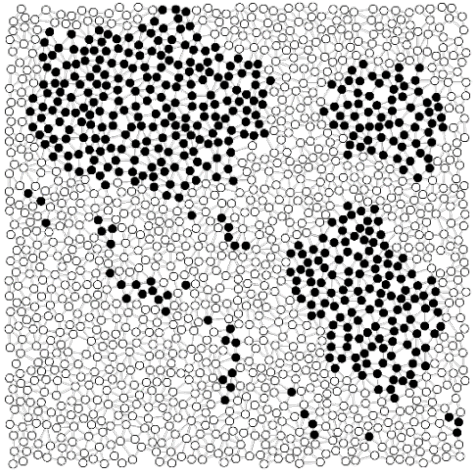
Dilation $\delta_1(G)$

$\forall\, v \in V$,

- $\varepsilon_n(G)(v) = \min\{G(v), d_E(u, v) \leq n\}$
- $\delta_n(G)(v) = \max\{G(v), d_E(u, v) \leq n\}$
- $\gamma_n(G) = \delta_n \circ \varepsilon_n(G)$
- $\varphi_n(G) = \varepsilon_n \circ \delta_n(G)$



Opening $\gamma_1(G)$

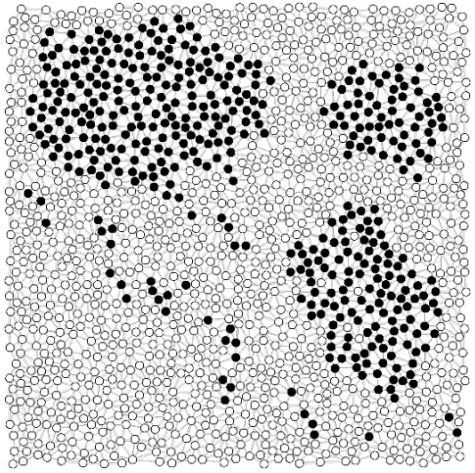


Closing $\varphi_1(G)$

[11] Luc Vincent, "**Graphs and Mathematical Morphology**", Signal Processing, 1989.

2) Spatial Point Analysis using Mathematical Morphology

Mathematical Morphology on Graphs:



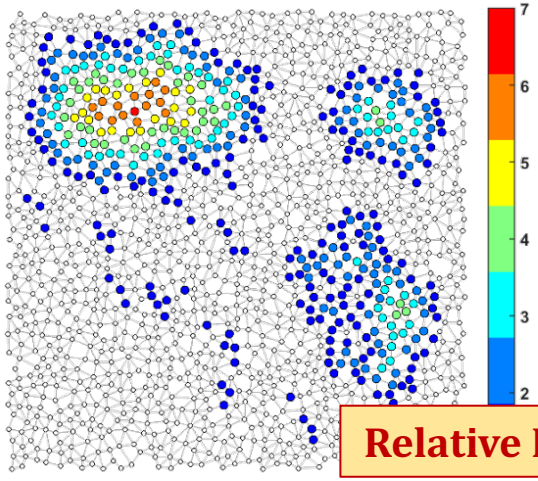
Binary graph: $G(V,E)$

$\circ = 0\,,\ \bullet = 1$

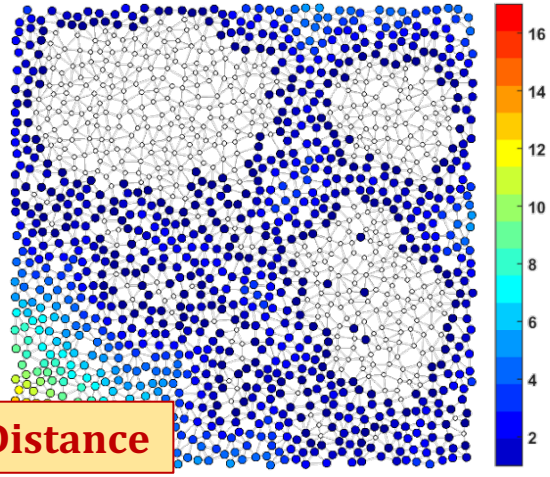
$$D(G) = \sum_{n=0}^{k} \varepsilon_n(G) \quad with \quad \varepsilon_k(G) = \varepsilon_{k'}(G), \qquad \forall k' \geq k$$

$$\Omega(G) = \sum_{n=0}^{k} \gamma_n(G) \quad with \quad \gamma_k(G) = \gamma_{k'}(G), \qquad \forall k' \geq k$$

$$\zeta(G) = \sum_{n=0}^{k} \varphi_n(G) \quad with \quad \varphi_k(G) = \varphi_{k'}(G), \qquad \forall k' \geq k$$

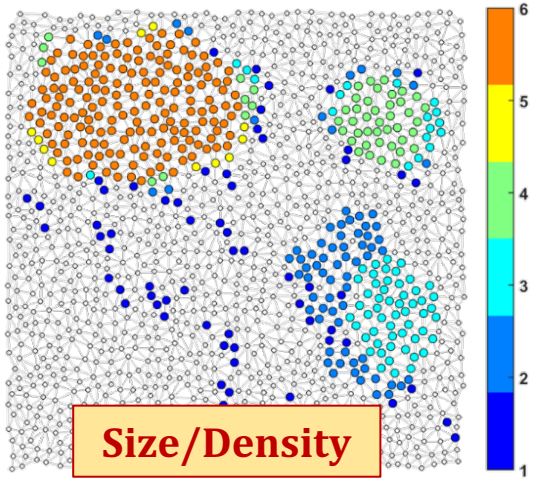


Distance map $D(G)$



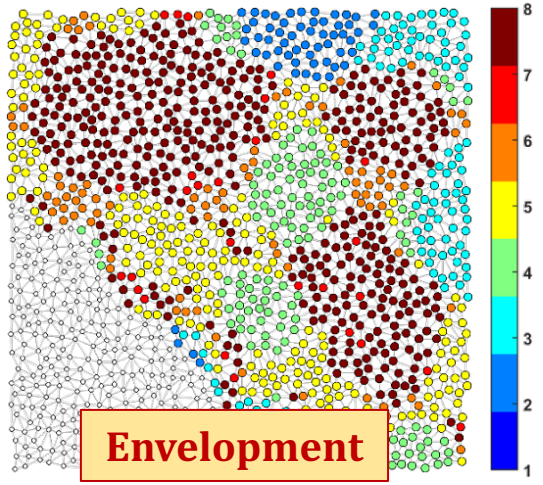
Distance map $D(\bar{G})$

Relative Distance



Openness $\Omega(G)$

Size/Density



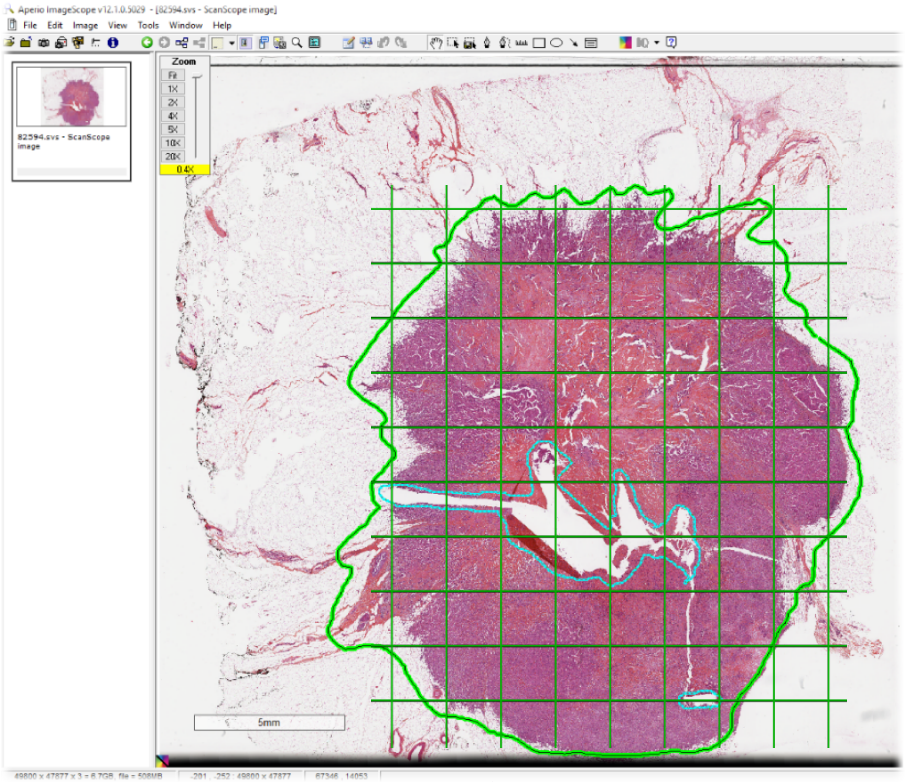
Closeness $\zeta(G)$

Envelopment

3) Node Identification

Data:

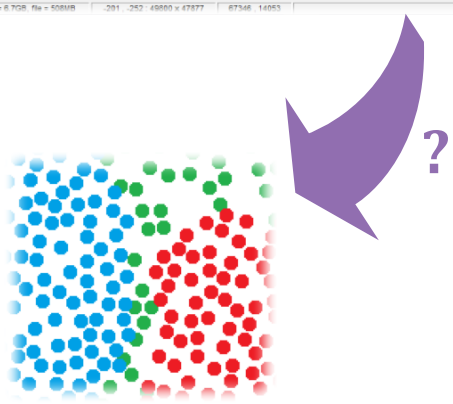
- 55 WSIs (breast cancer)
- Hematoxylin-Eosin-Saffron (HES),
- Aperio ScanScope CS (20X objective)
- Resolution: 0.5 $\mu m/pixel$.



Framework:

HES image tile

(1024×1024)

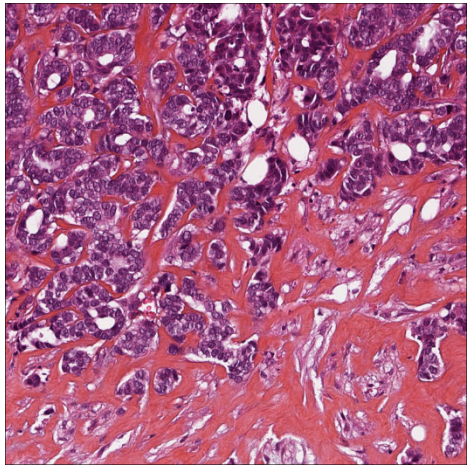


3) Node Identification

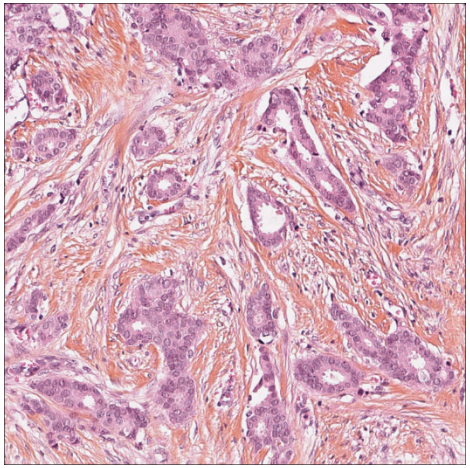
Preprocessing: color normalization

Reinhard:

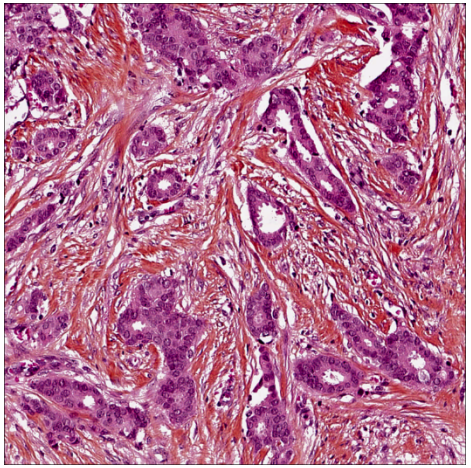
$$I_{norm}^{(i)} = (I_{source}^{(i)} - \mu_{source}^{(i)}) \frac{\sigma_{target}^{(i)}}{\sigma_{source}^{(i)}} + \mu_{target}^{(i)}, \quad i \in \{L, a, b\}$$



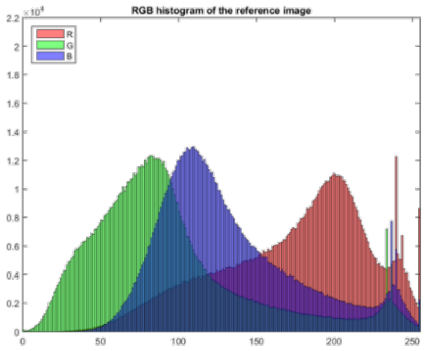
Reference image

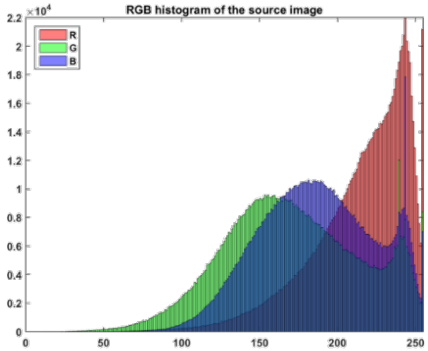


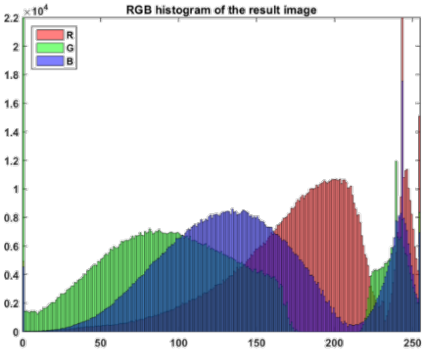
Under-stained image



Normalized image







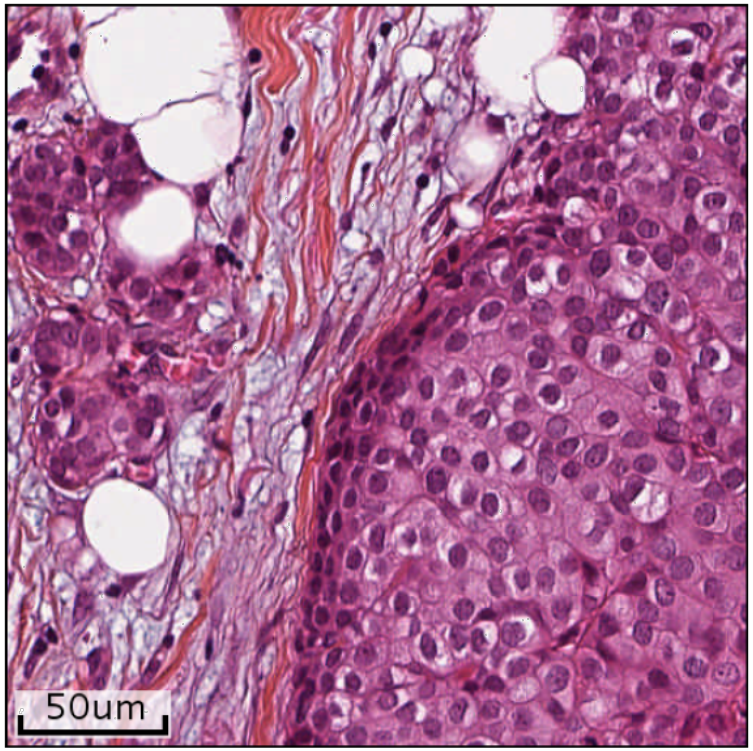
[12] Reinhard E. et al. "**Color transfer between images**". IEEE Computer graphics and applications (2001)

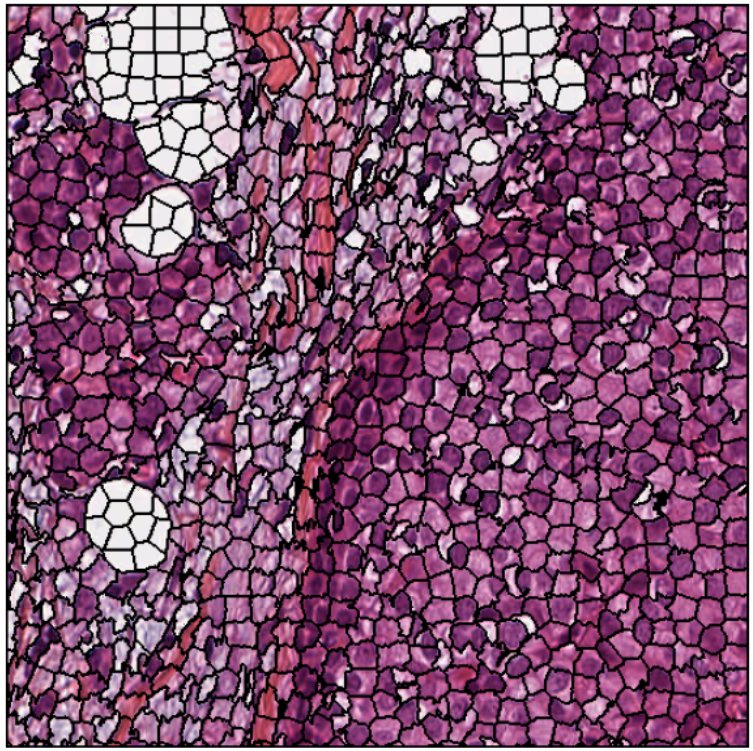
3) Node Identification

Preprocessing: superpixel segmentation

Simple Linear Iterative Clustering (SLIC):

- number of superpixels: 3500
- compact factor: 35

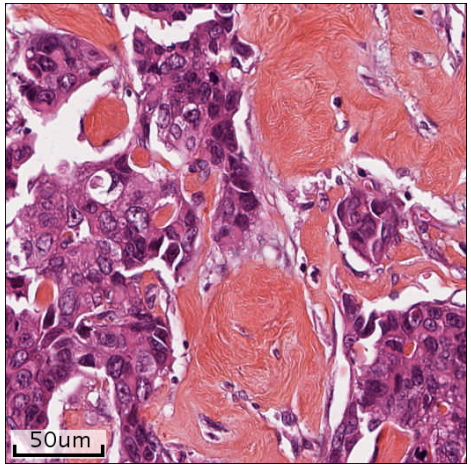




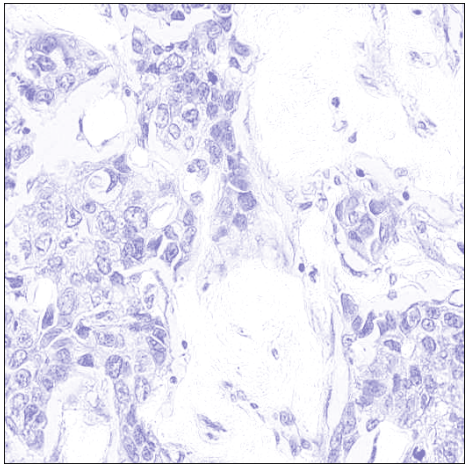
[13] Achanta R. et al. "**SLIC Superpixels**". EPFL Technical Report no. 149300 (2010)

3) Node Identification

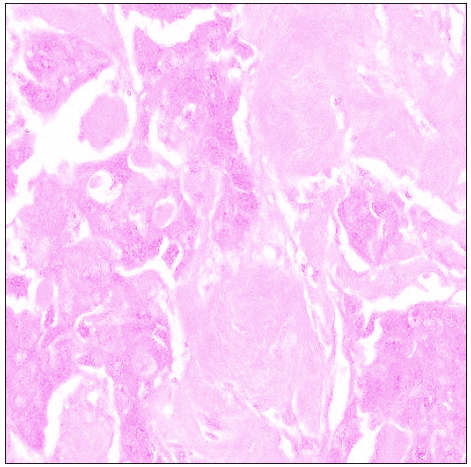
Nuclei detection: color deconvolution



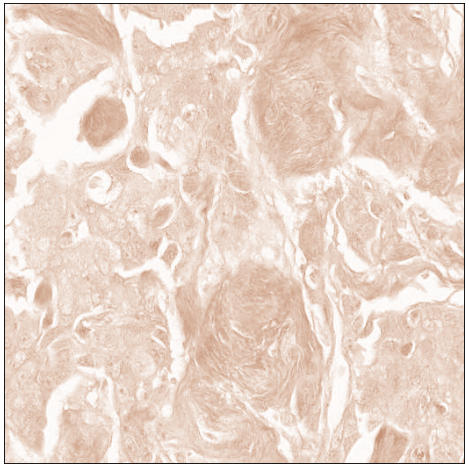
HES



Haematoxylin (H)



Eosin (E)



DAB (~S)

- Separation of histochemical staining

$$\begin{bmatrix} H_x \\ E_x \\ DAB_x \end{bmatrix} = OD^{-1} \times \begin{bmatrix} -log(\frac{R_x+1}{256}) \\ -log(\frac{G_x+1}{256}) \\ -log(\frac{B_x+1}{256}) \end{bmatrix}$$

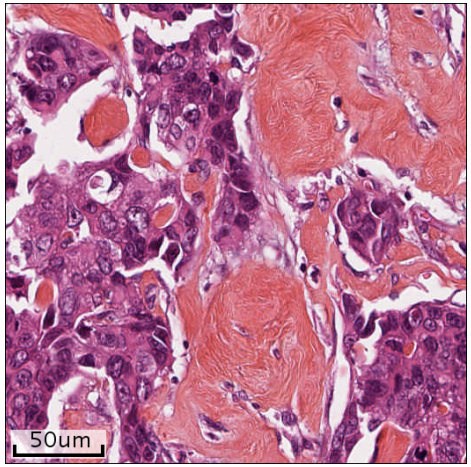
$$OD = \begin{bmatrix} 0.65 & 0.70 & 0.29 \\ 0.07 & 0.99 & 0.11 \\ 0.27 & 0.57 & 0.78 \end{bmatrix} \begin{matrix} \leftarrow \text{H} \\ \leftarrow \text{E} \\ \leftarrow \text{DAB} \end{matrix}$$

Optical density matrix

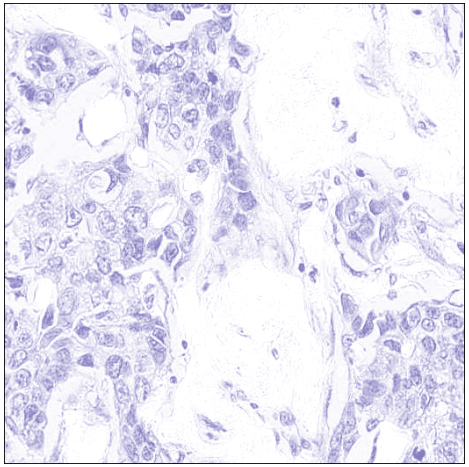
[14] Ruifrok A.C. et al. "**Quantification of histochemical staining by color deconvolution**". In: Analytical and quantitative cytology and histology (2001)

3) Node Identification

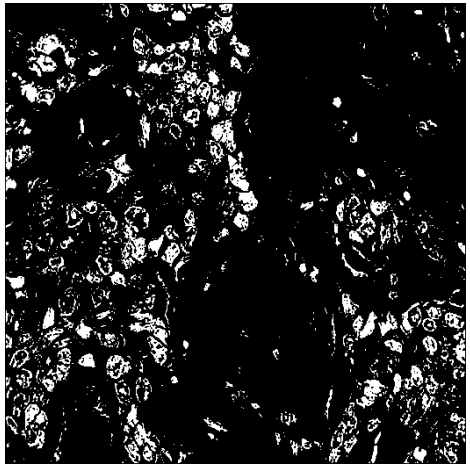
Nuclei detection: superpixel selection



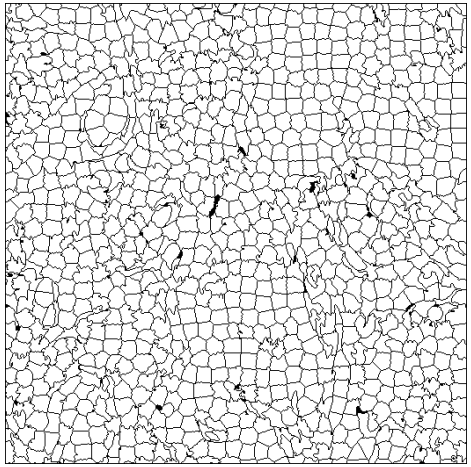
HES



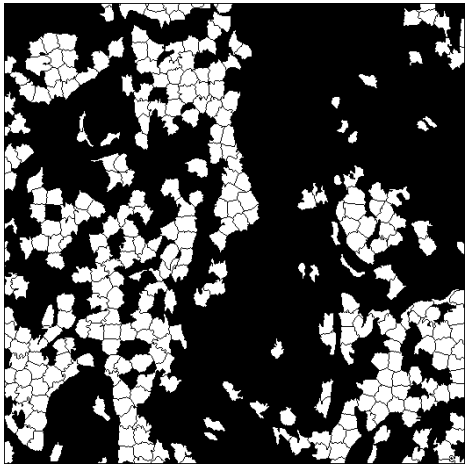
Haematoxylin (H)



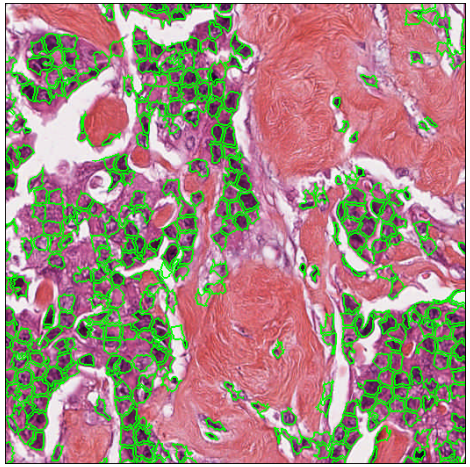
$Open_{R_d}(H > T_h)$ ← $R_d?\,;T_h?$



SLIC

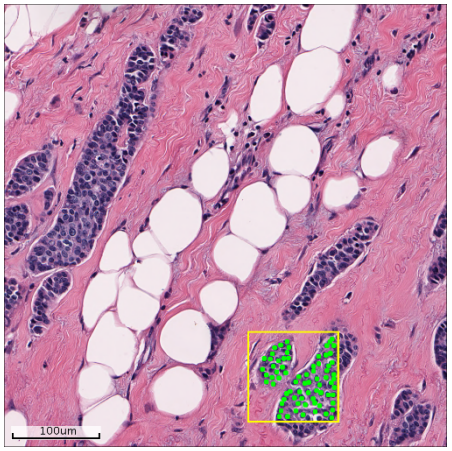


$Rec\left(Open_{R_d}(H > T_h), SLIC\right)$

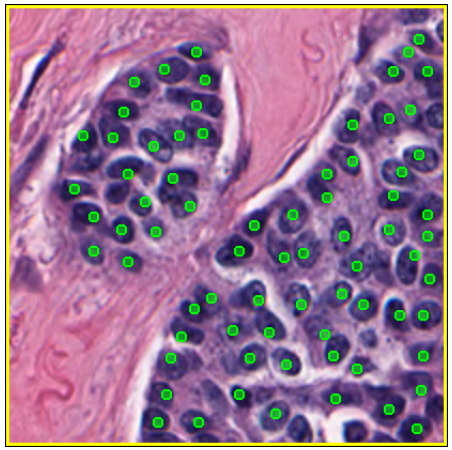


3) Node Identification

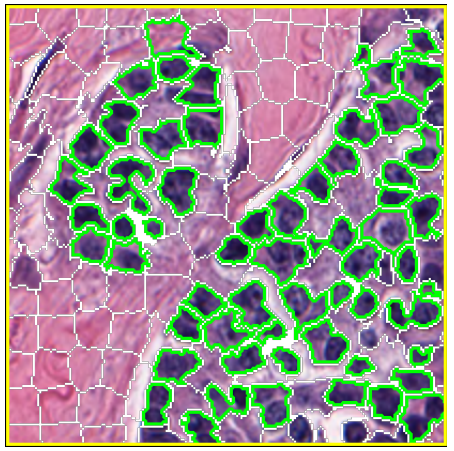
Nuclei detection: superpixel selection



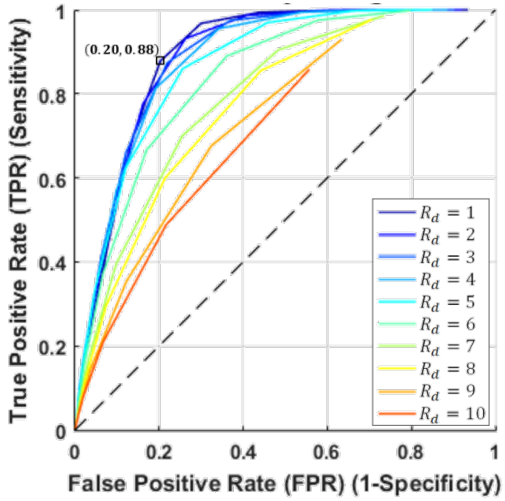
Dataset [14]



Ground Truth



Superpixel Ground Truth



- 512 images (color normalization)
- 17 294 nuclei ⇒ 14 830 (+) superpixels
  71 191 (−) superpixels
  } within the validation ROIs

➔ $R_d = 1$ and $T_h = 0.6$

F-score	Precision	Recall	Specificity	Accuracy
0.60729	0.46406	0.87843	0.7971	0.81065

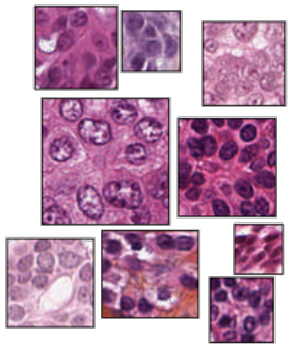
(50% training | 50% testing)

0.8449 [14]

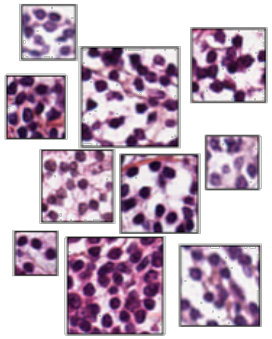
[15] Xu J. et al. "**Stacked sparse auto-encoder (SSAE) for nuclei detection on breast cancer histopathology images**". IEEE trans. on medical imaging (2016)

3) Node Identification

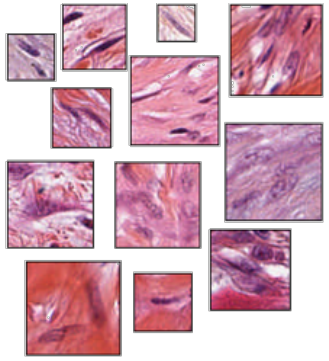
Nuclei classification:



Cancer cells



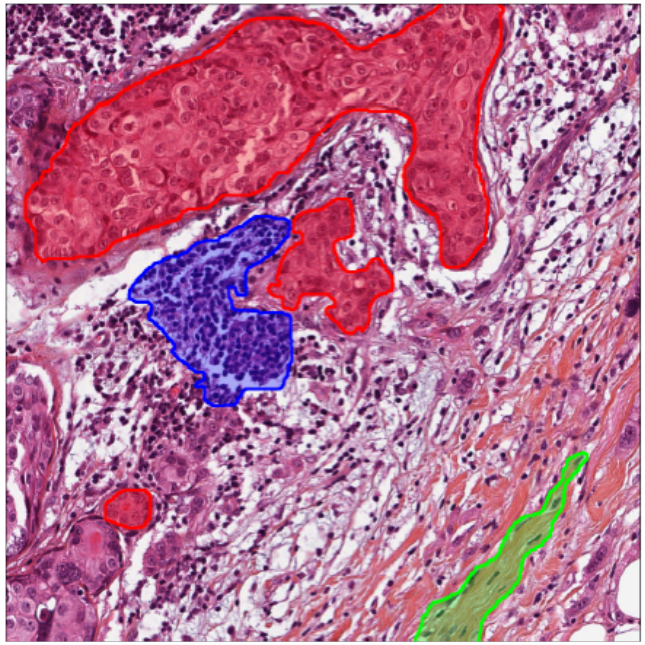
Immune cells



Fibroblasts

- 17 WSI
- 2533 regions
  ⇒ 1005 (1024×1024) images

Cell type	Nbr. of detected nuclei within the annotated areas
Cancerous	77676
Immune cell	31037
Fibroblast	3412
Total	**112125**



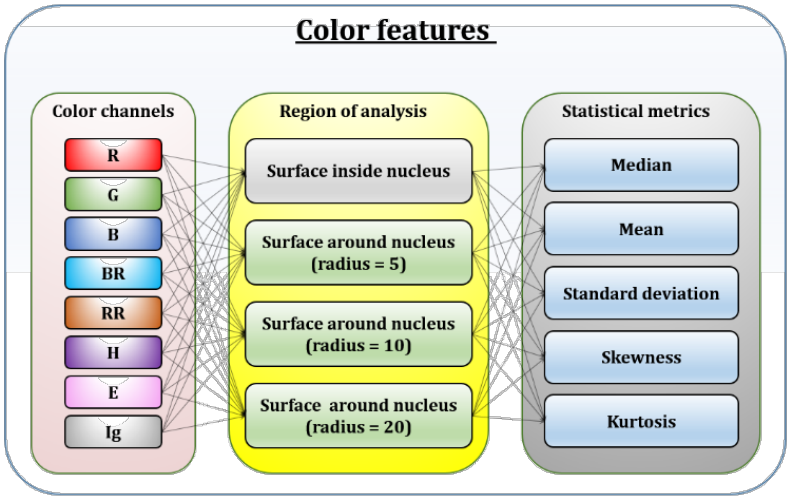
Cancer cells

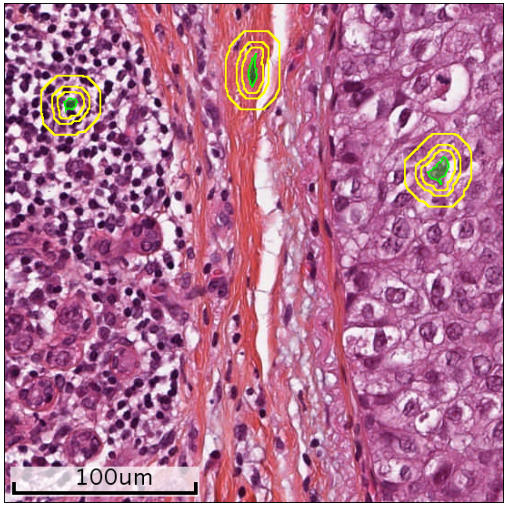
Immune cells

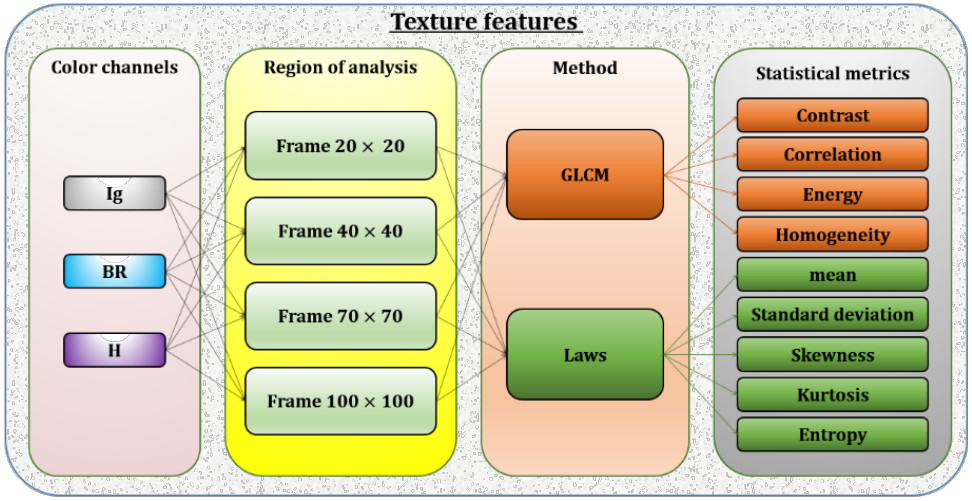
Fibroblasts

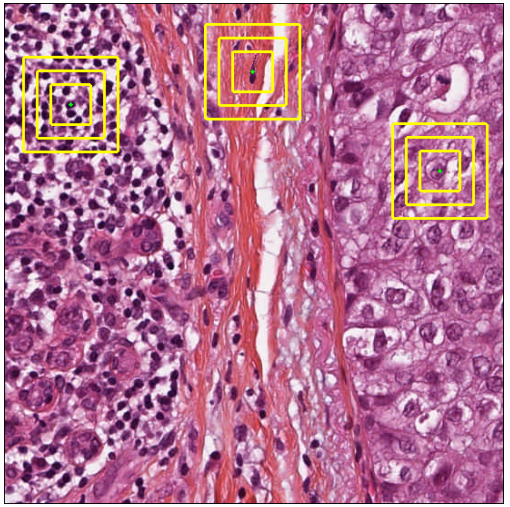
3) Node Identification

Nuclei classification: feature extraction









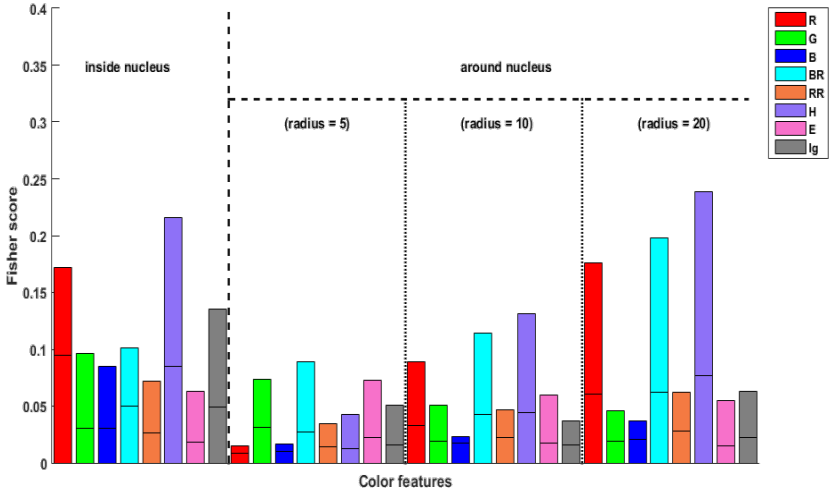
3) Node Identification

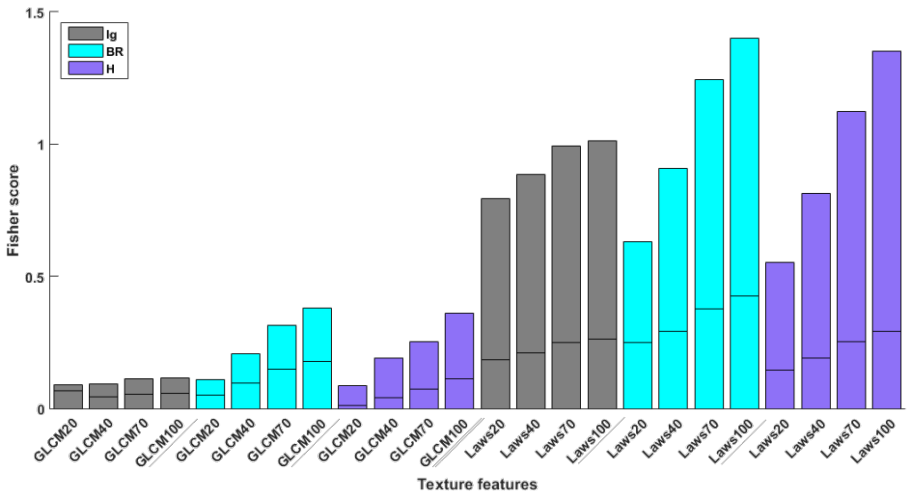
Nuclei classification: feature selection

- 160 color features per nucleus
- 1092 texture features
  - 900 Laws features
  - 192 GLCM features

➔ **1252 features per nucleus**

Fisher method: $$F_s = \frac{\sum_{k=1}^{K} n_k(\mu - \mu_k)^2}{\sum_{k=1}^{K} n_k \sigma_k^2}$$





➔ **337 features per nucleus**

3) Node Identification

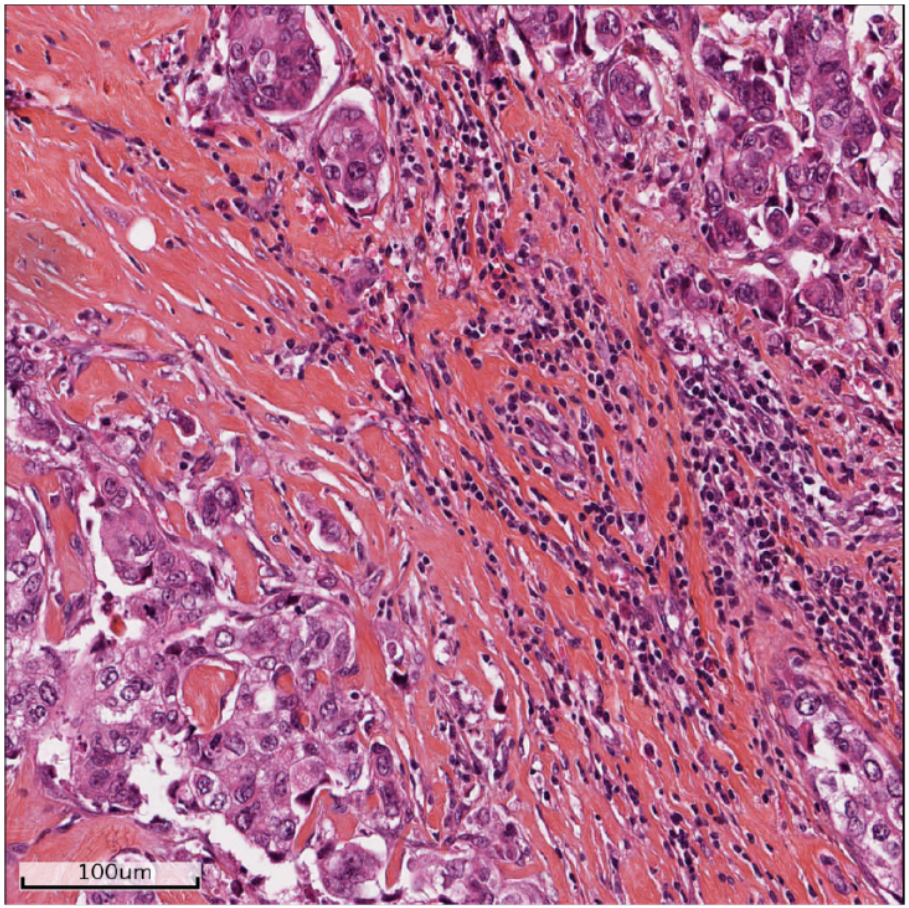
Nuclei classification: Random Forest

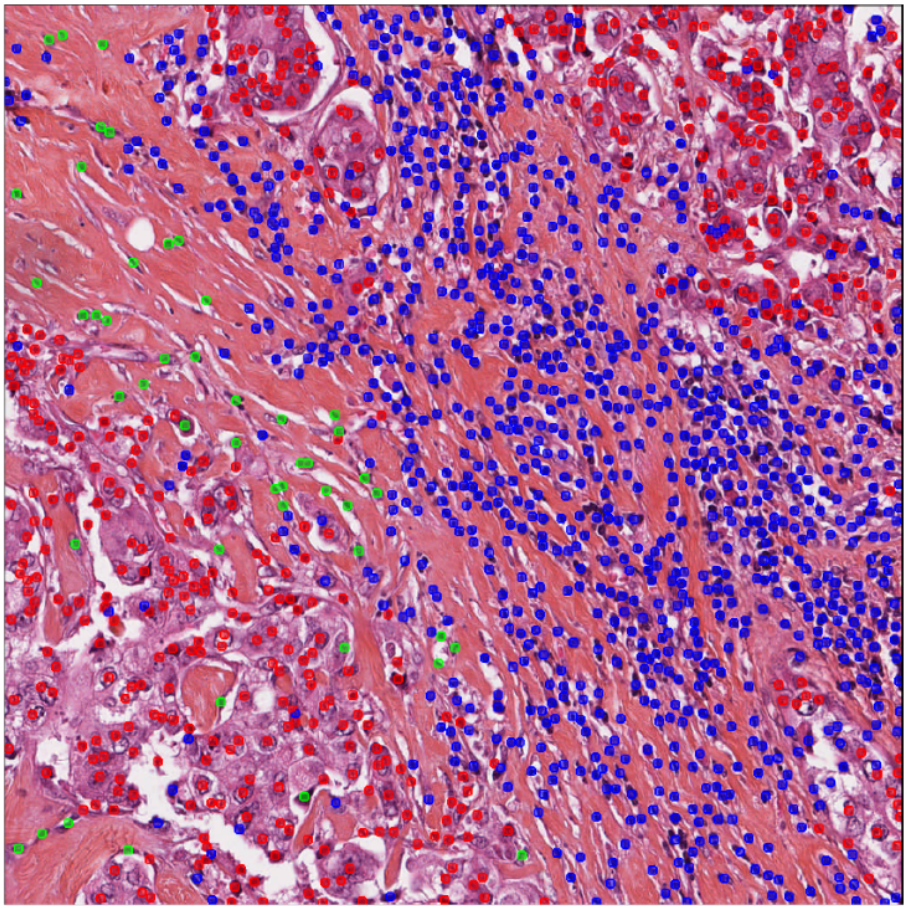
Average over 10-fold cross validation

Accuracy_M	Precision_M	Recall_M	Specificity_M	$F_1\text{-score}_M$
0.9784	0.9383	0.8745	0.9737	0.9053

Confusion Matrix

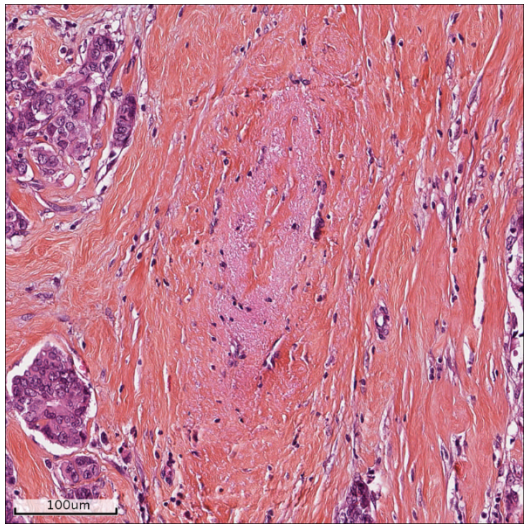
Actual class \ Predicted class	Cancer cells	Immune cells	Fibroblasts
Cancer cells	**68858** (98.50%)	901 (1.29%)	149 (0.21%)
Immune cells	1087 (3.89%)	**26705** (95.60%)	142 (0.51%)
Fibroblasts	771 (25.11%)	187 (6.09%)	**2113** (68.80%)



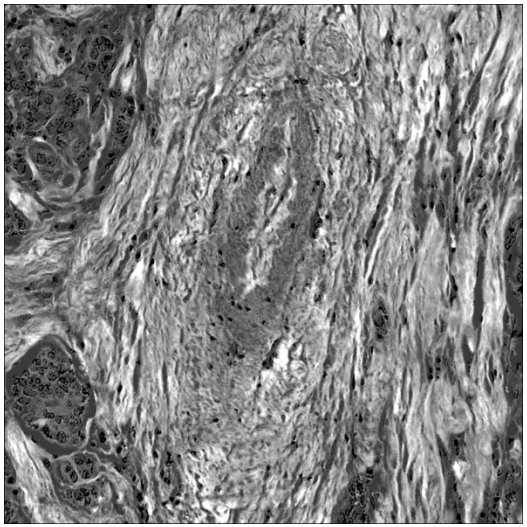


3) Node Identification

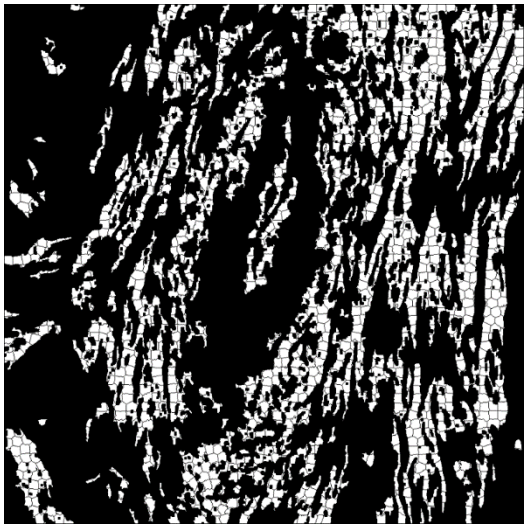
Collagen segmentation:



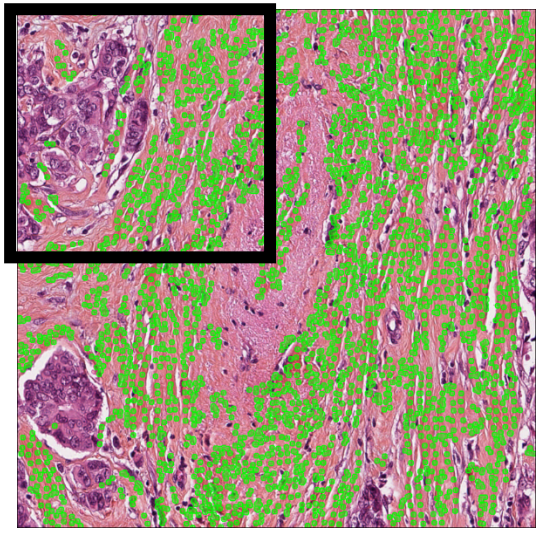
HES

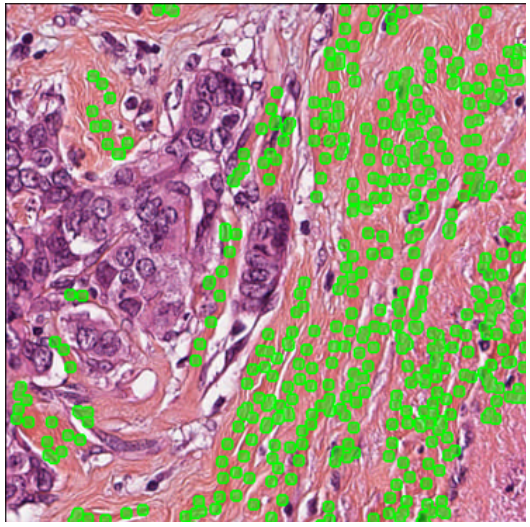


I_{RR}



$Rec(Open_2(I_{RR} > 0.5), SLIC)$



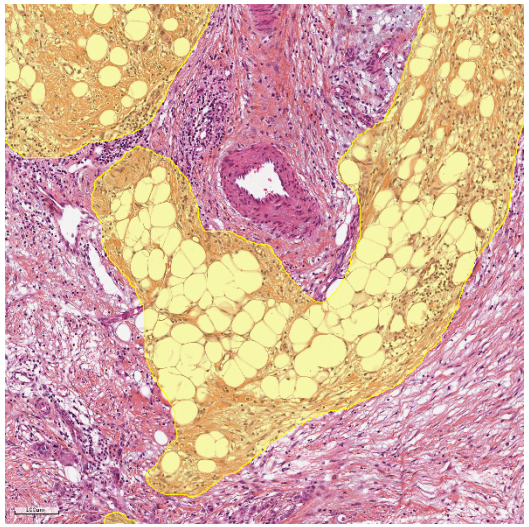


$$I_{RR} = \frac{R}{1 + G + B + \frac{(G-B)^2}{1+G+B} + R}$$

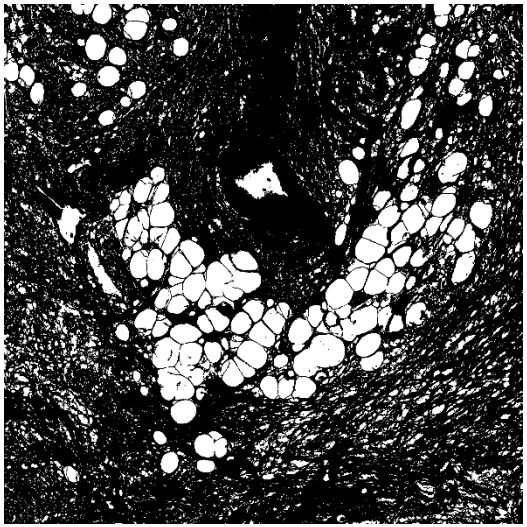
Qualitative Evaluation

3) Node Identification

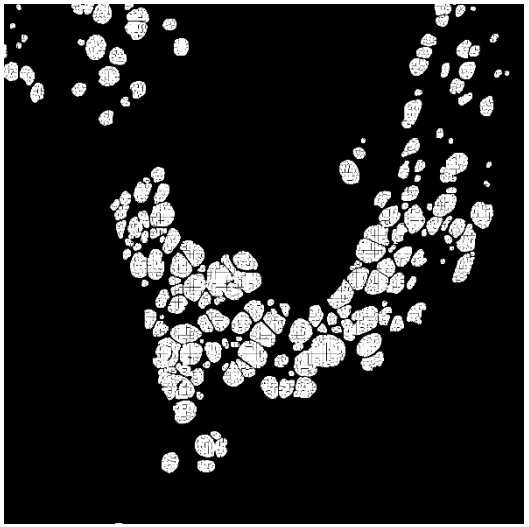
Adipose tissue segmentation:



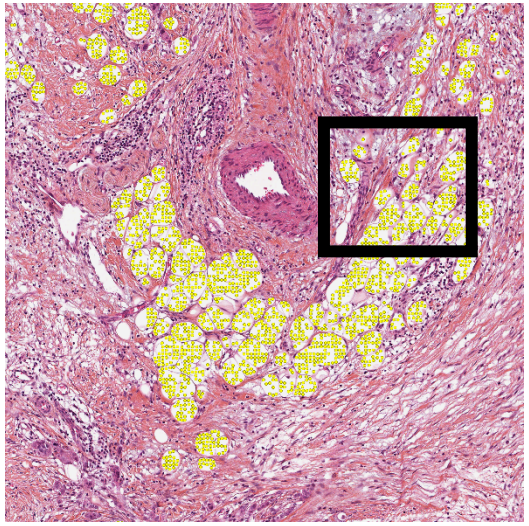
HES

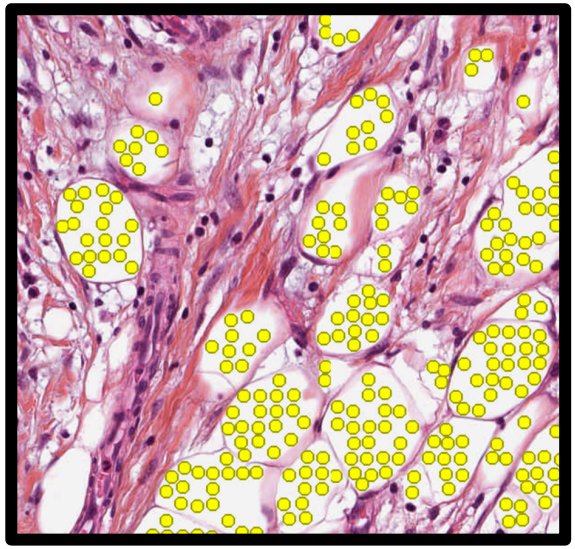


HES>220



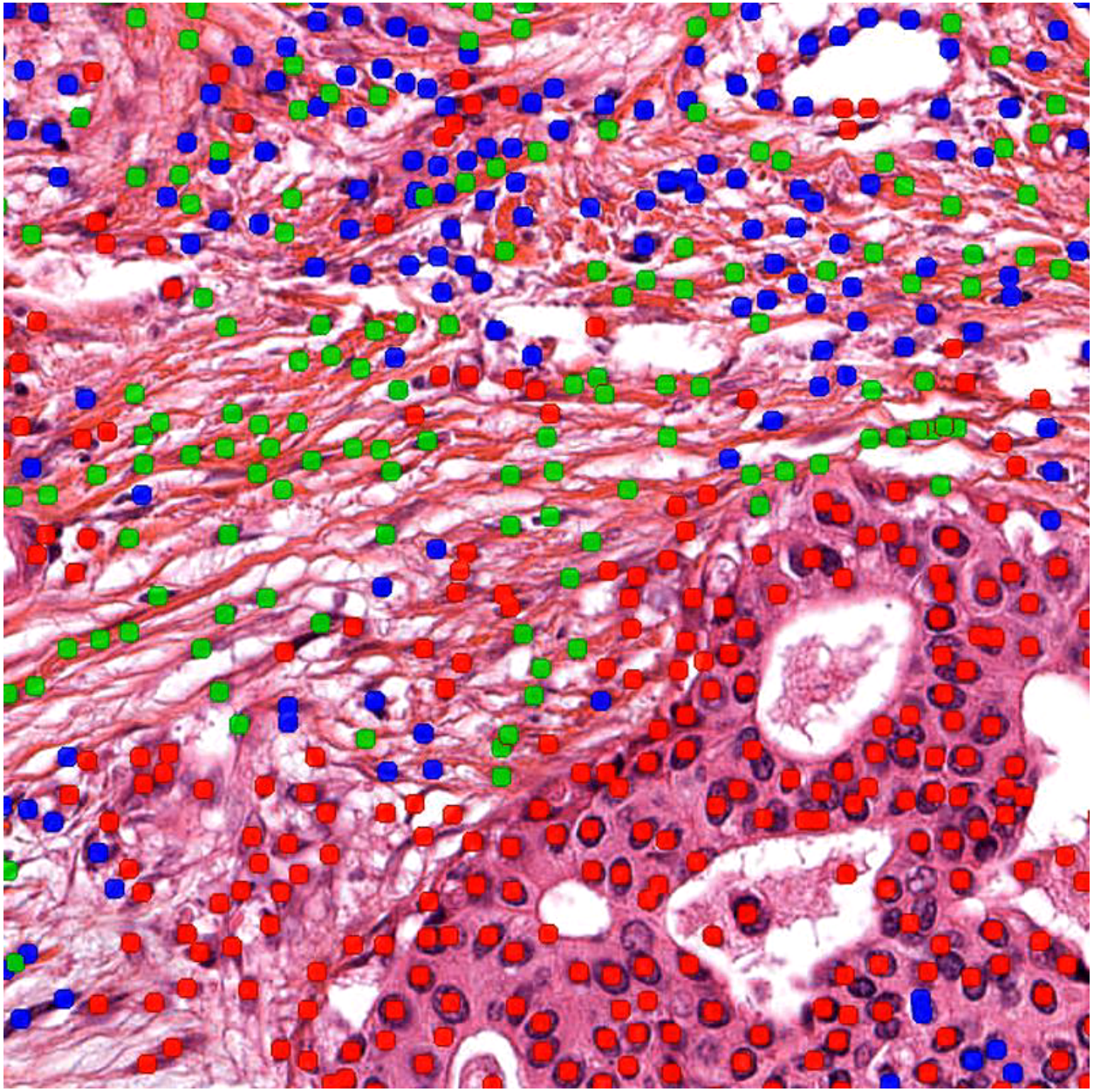
$Rec(Open_{10}(\text{HES>220}), SLIC)$





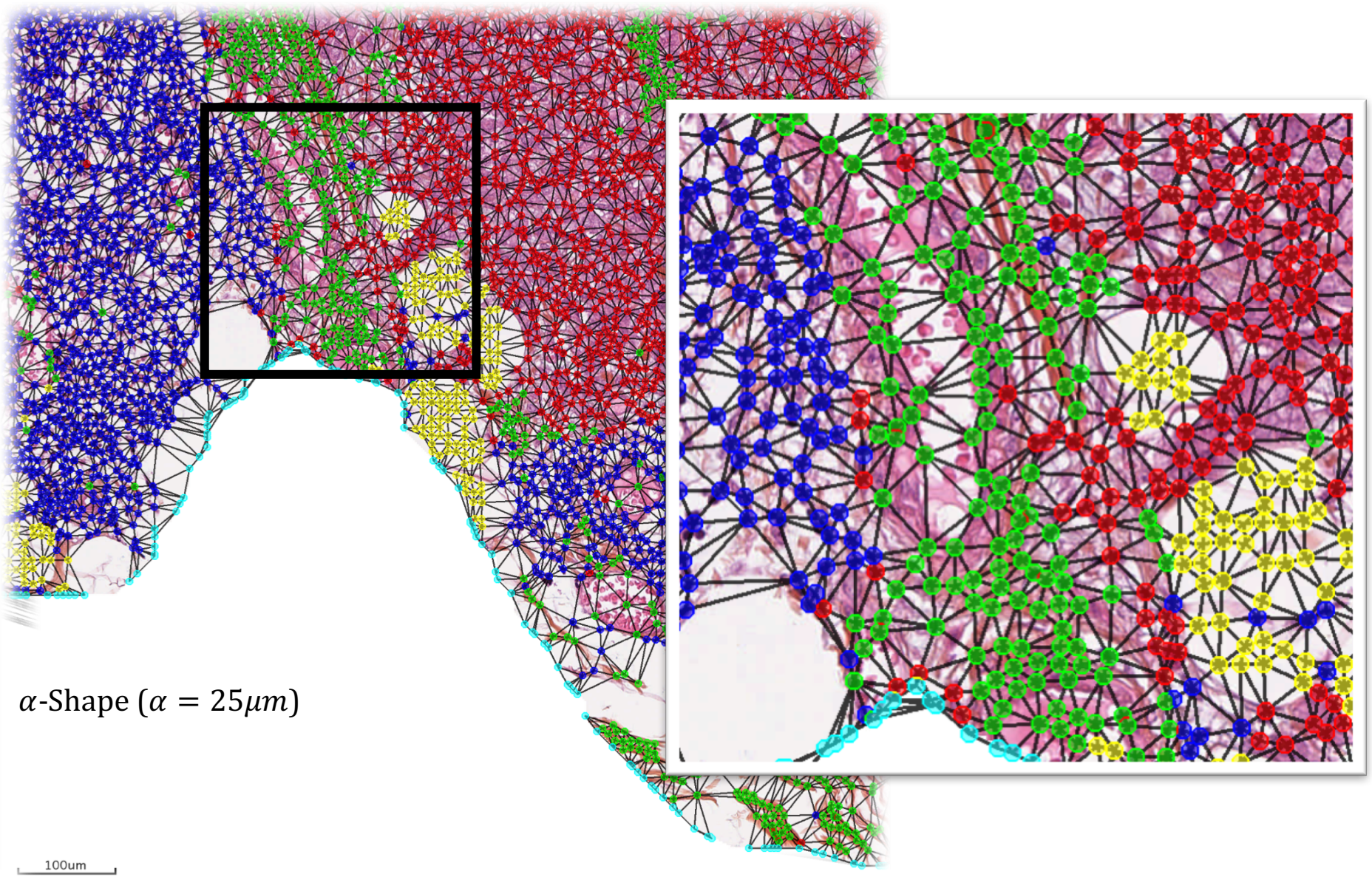
Qualitative Evaluation

3) Node Identification



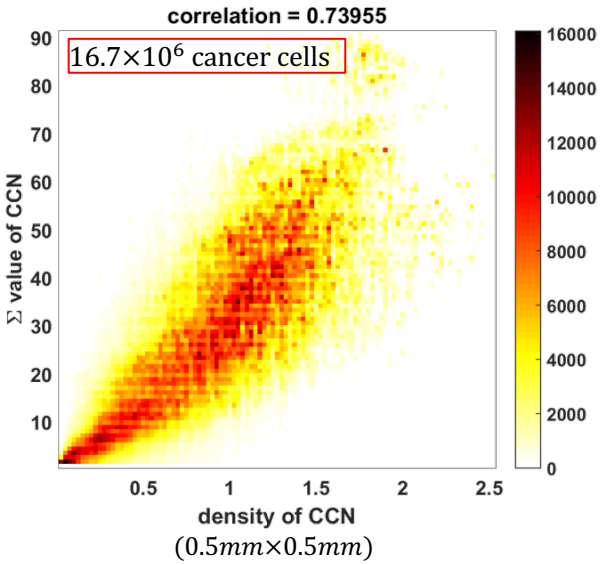
- Cancer cells
- Immune cells
- Collagen + Fibroblasts
- Adipocytes

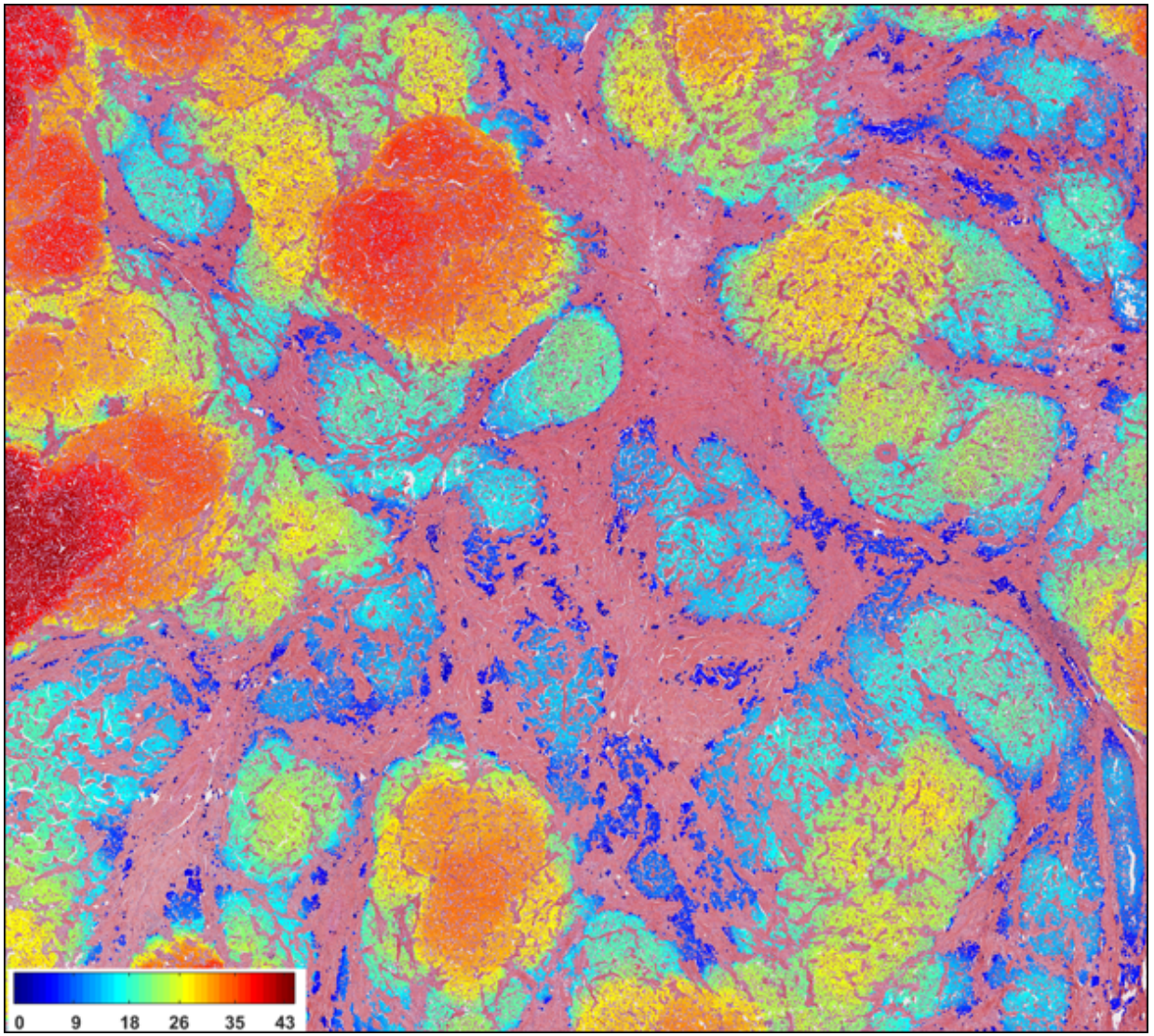
4) Spatial organization of histological structures



4) Spatial organization of histological structures

Expanse:





Expanse function:

$$\Sigma(G_X) = \frac{1}{N+1}\sum_{n=0}^{N} \Omega \circ \varphi_n(G_X)$$

$N = 5$

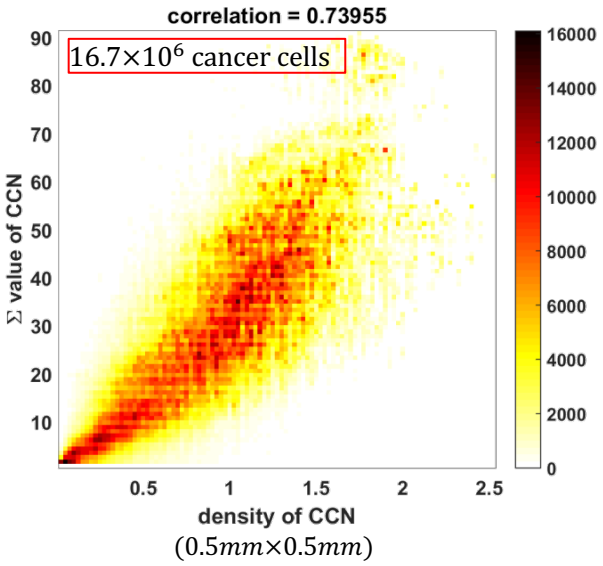
$\Omega \circ \varphi_n(G_X)$

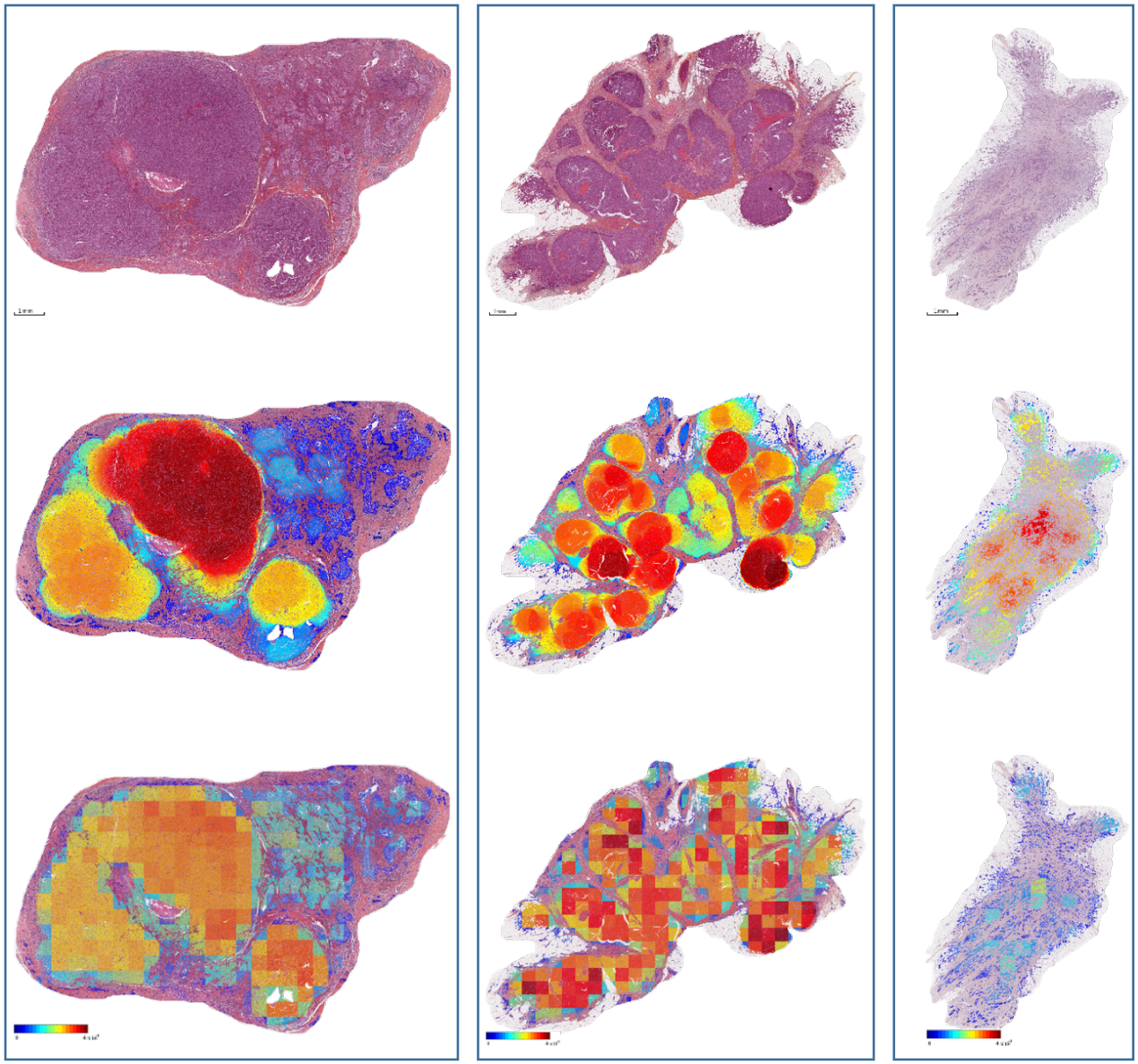
- $n = 4$

- X = cancer cells
- $\Omega(G) = \sum_{n=0}^{k} \gamma_n(G) \quad with \quad \gamma_k(G) = \gamma_{k'}(G),\ \forall k' \geq k$

4) Spatial organization of histological structures

Expanse:





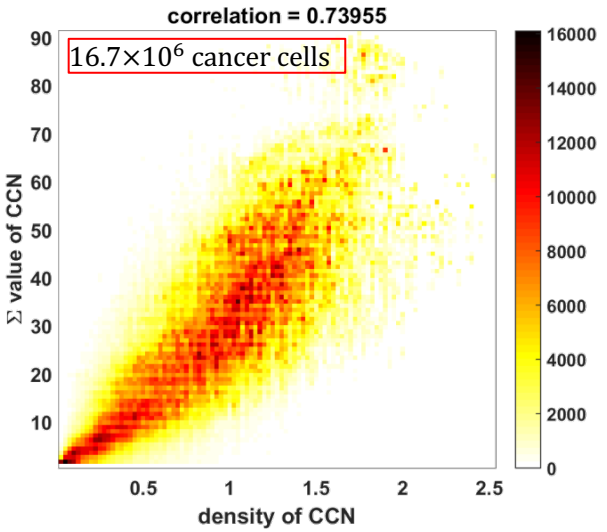
Expanse function:

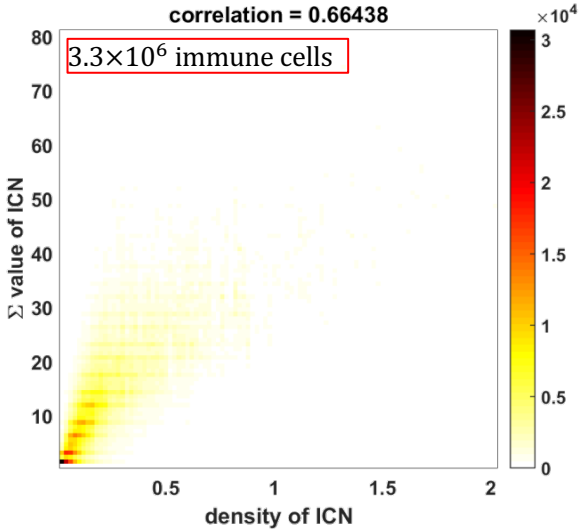
$$\Sigma(G_X) = \frac{1}{N+1}\sum_{n=0}^{N} \Omega \circ \varphi_n(G_X)$$

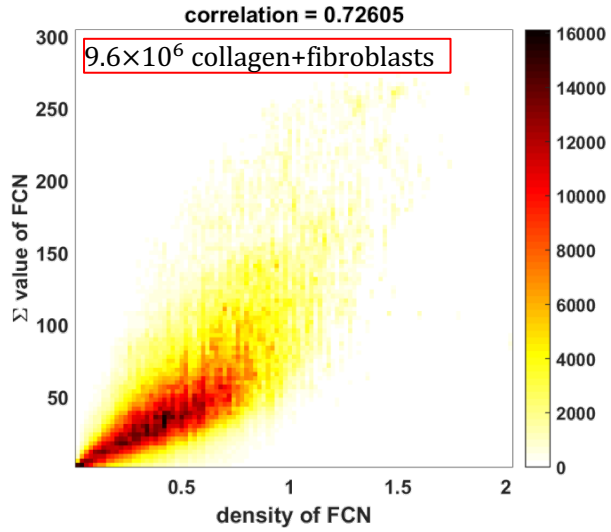
$N = 5$

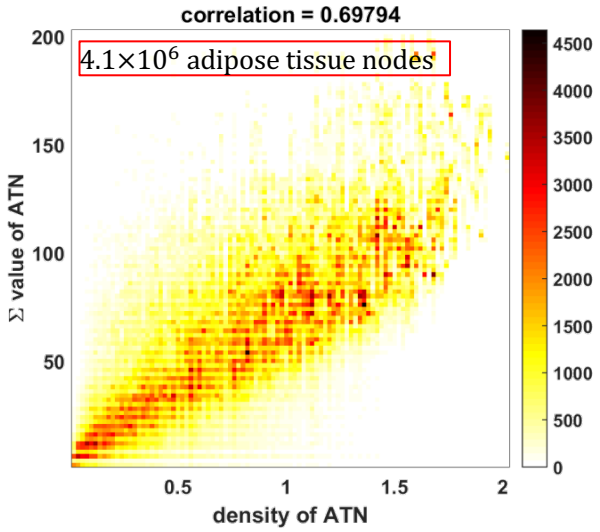
4) Spatial organization of histological structures

Expanse:









4) Spatial organization of histological structures

Sparsity:

Immune cell aggregates

Envelope immune cell aggregates

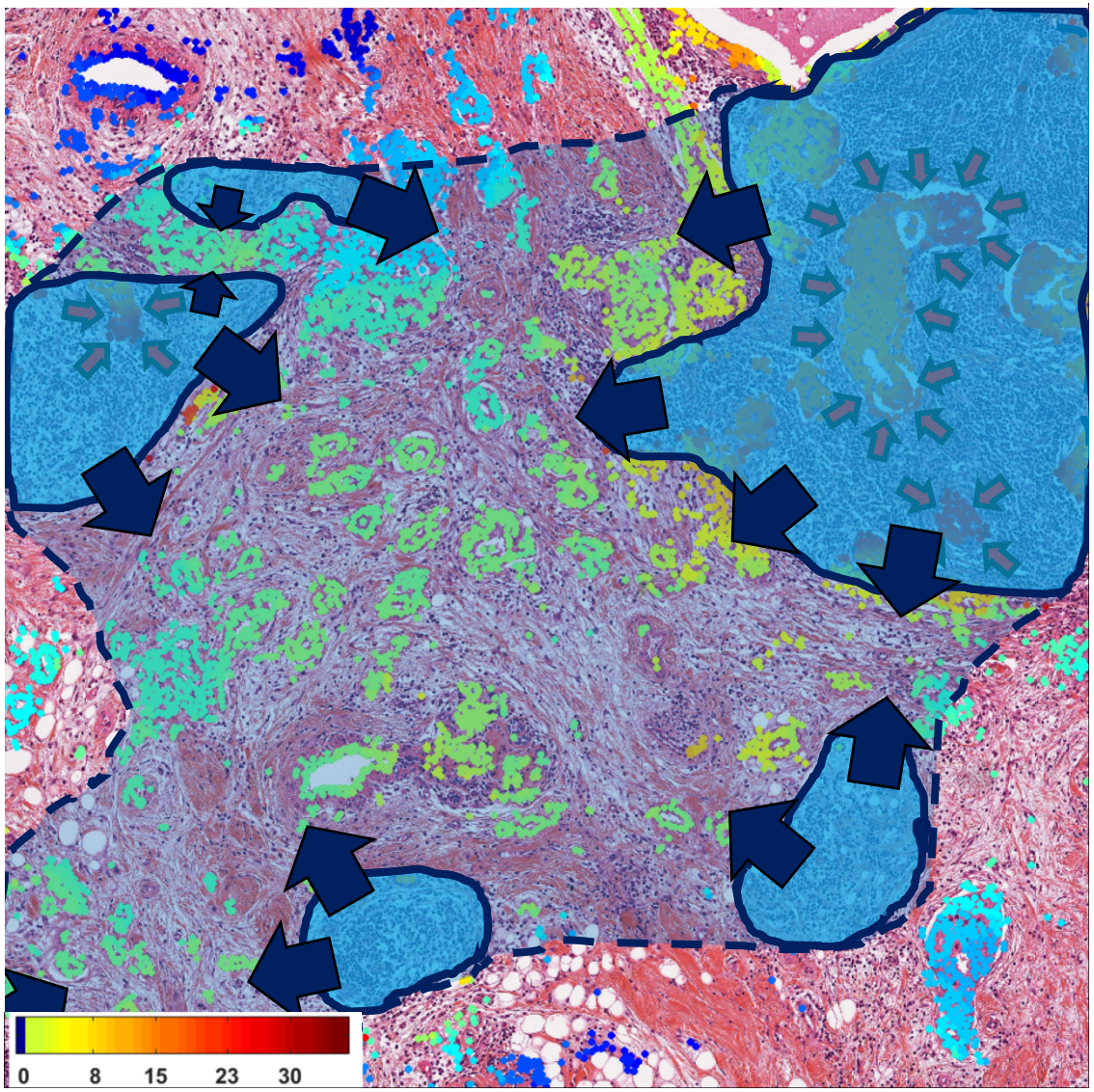
Cancer cells

- highly enclosed by immune cells
- surrounded by immune cells
- between distant immune cell aggregates
- free from immune cells

Sparsity function:

$$\Gamma(G_X) = \frac{1}{N+1}\sum_{n=0}^{N} \zeta \circ \gamma_n(G_X)$$

$N = 5$



$\Gamma(G_X)(G)$

$\zeta(G_X)(G)$

- X = immune cells
- Y = cancer cells
- $n = 4$
- $\zeta(G) = \sum_{n=0}^{k} \varphi_n(G)$ *with* $\varphi_k(G) = \varphi_{k'}(G),\ \forall k' \geq k$

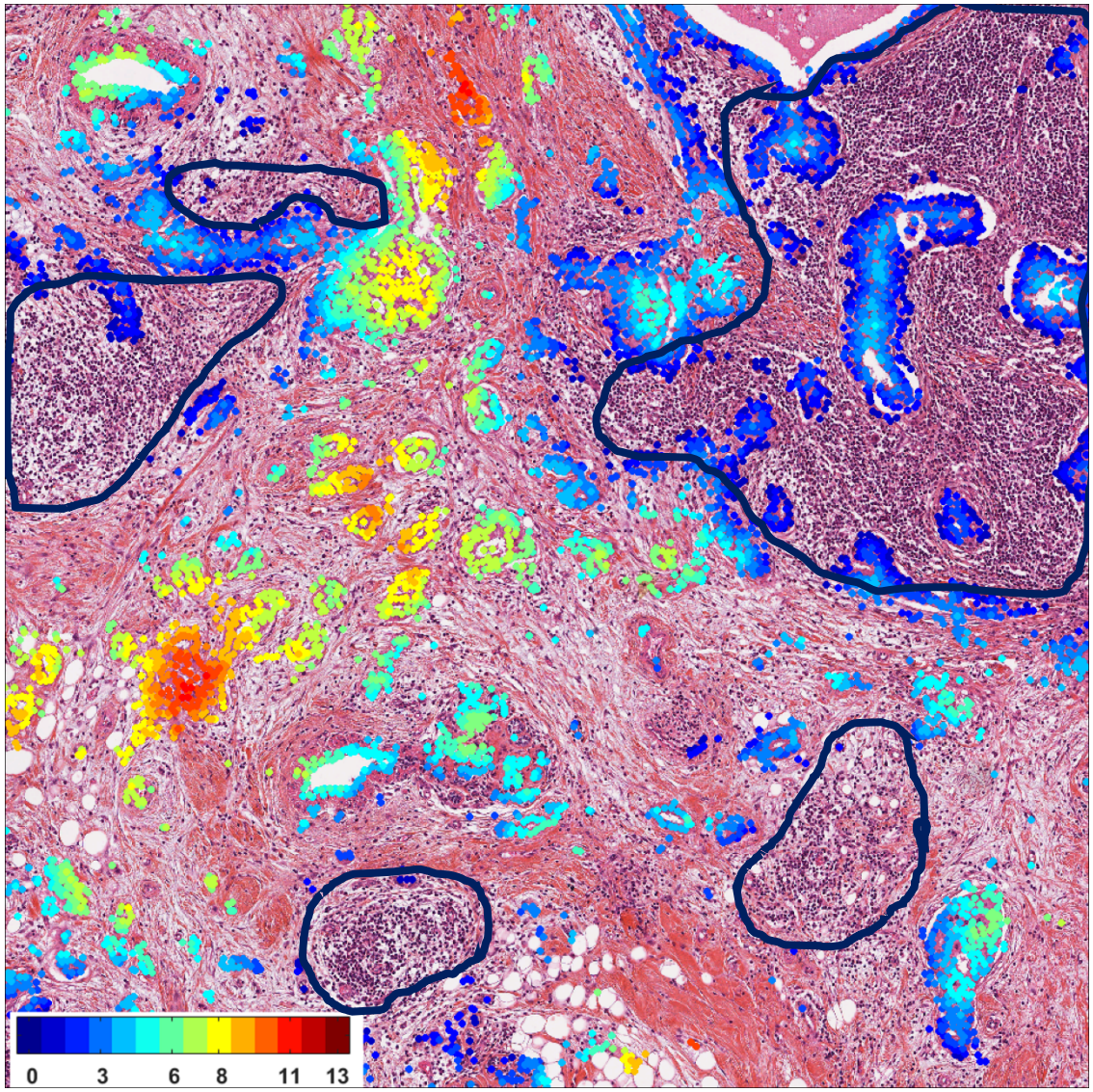
4) Spatial organization of histological structures

Relative distance:

Immune cell aggregates

- Cancer cells distant from immune cells

- Cancer cells close to immune cells



Average distance function:

$$\Psi(G_X) = \frac{1}{N+1}\sum_{n=0}^{N} D \circ \gamma_n(G_X)$$

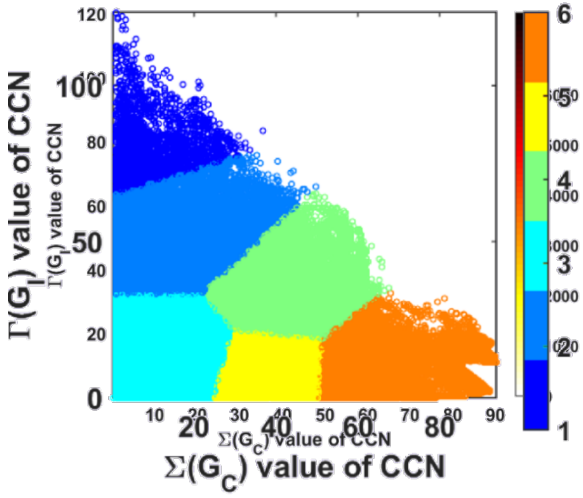
$N = 5$

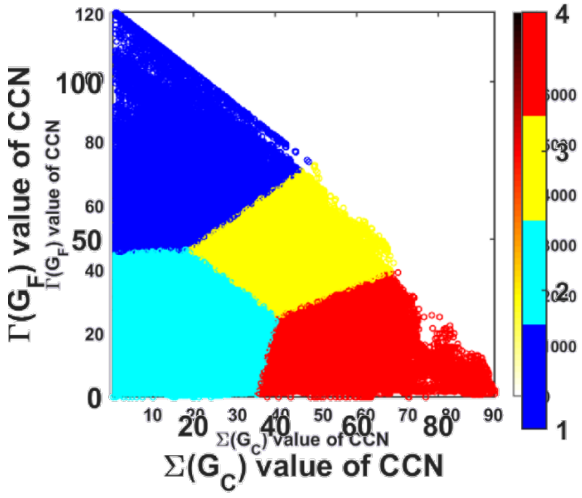
$D(G_X)(G_Y)$

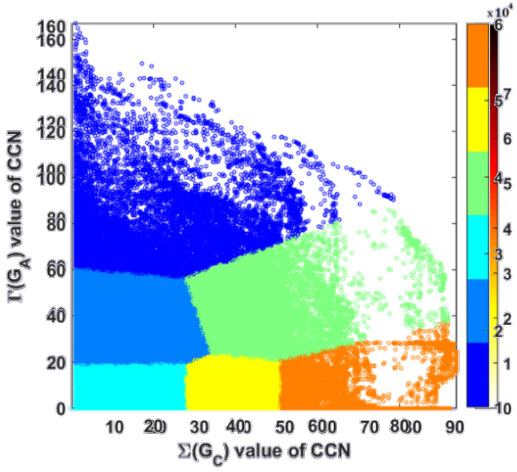
- X = immune cells
- Y = cancer cells
- $n = 4$
- $D(G) = \sum_{n=0}^{k} \varepsilon_n(G)$ $with$ $\varepsilon_k(G) = \varepsilon_{k'}(G),\ \forall k' \geq k$

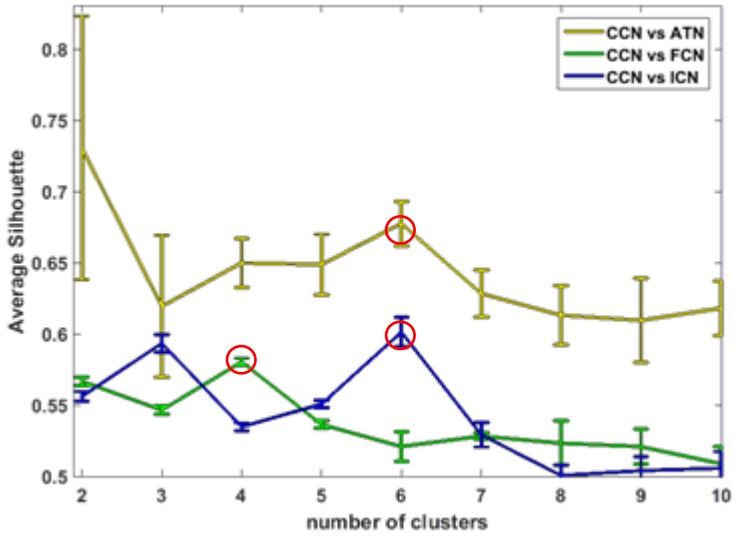
4) Spatial organization of histological structures

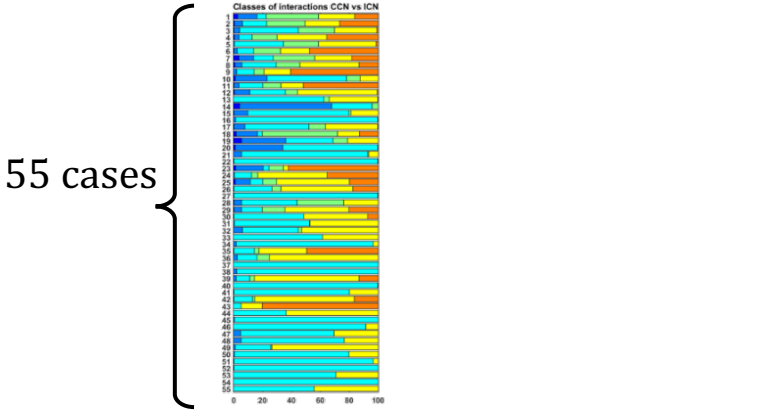
Interactions:





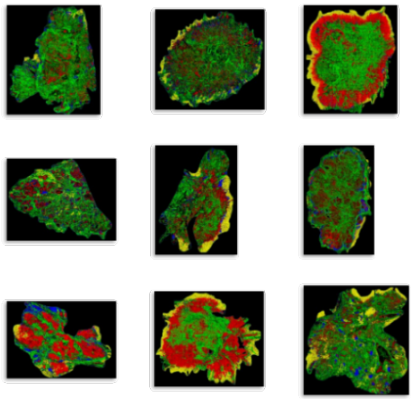


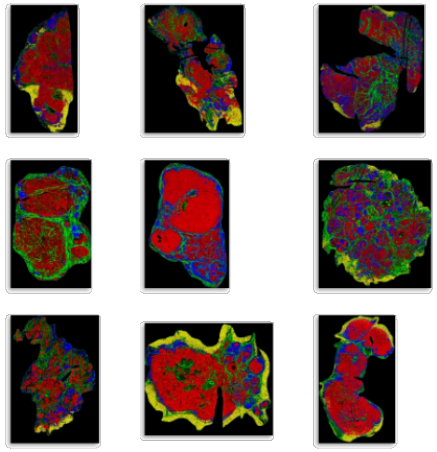




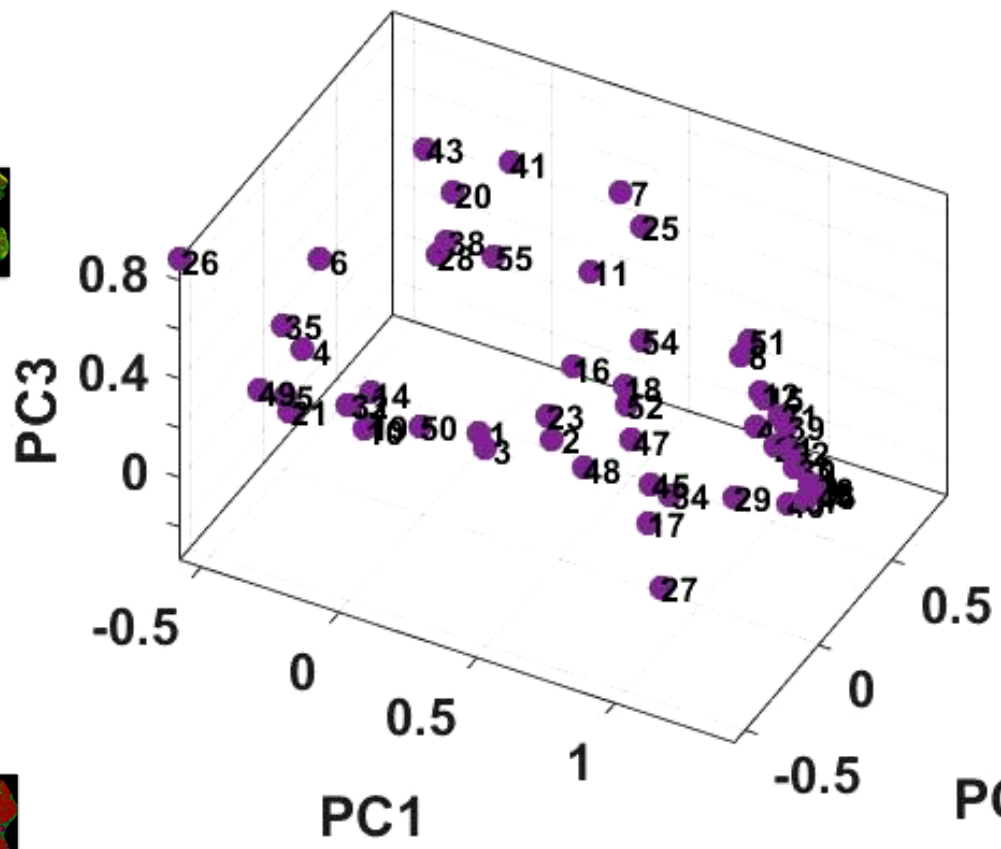
4) Spatial organization of histological structures

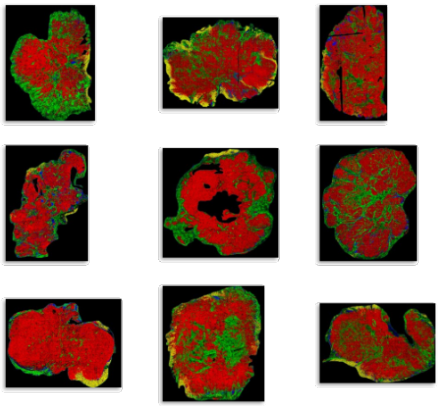
Interactions:

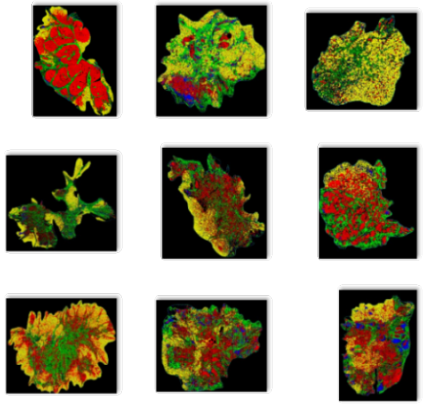




Principal Component Analysis







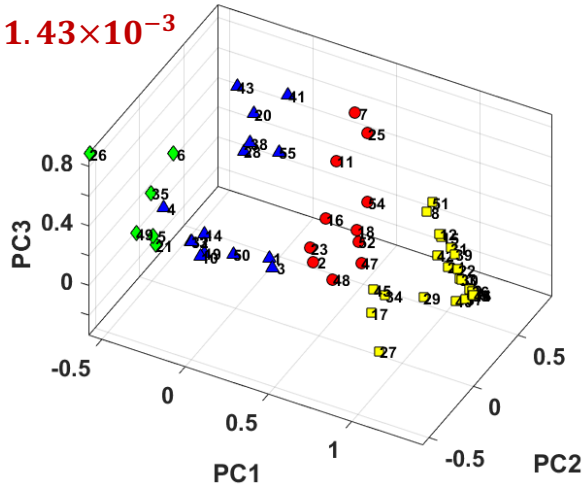
K-means clustering (k=4)

4) Spatial organization of histological structures

Interactions:

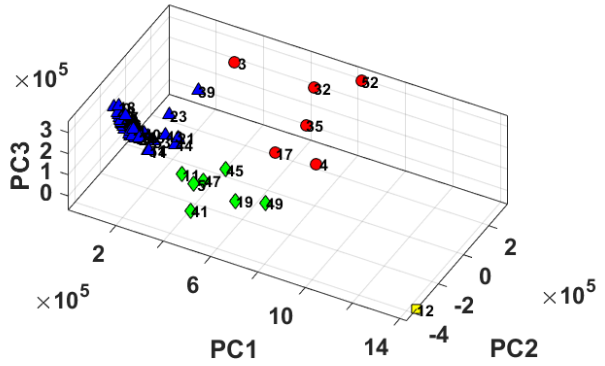
Spatial interactions parameters:

$PRESS = 1.43\times10^{-3}$



Node counting parameters:

$PRESS = 5.67\times10^{-2}$



Overall quality of the PCA : Predicted Residual Sum of Squares:

$$\mathrm{PRESS}_{M=3} = \left\| X - \tilde{X}^{[M]} \right\|$$

X : principal component coefficients
M : nbr. of components

➢ Leave-one-out cross validation

4) Spatial organization of histological structures

Outcome:

Tumor cell
- Density of tumor pattern
- Encloseness by TME_X
- Relative distance to TME_X

$X \in \{immune\ cells, collagen, adipose\ tissue\}$

Advantages:

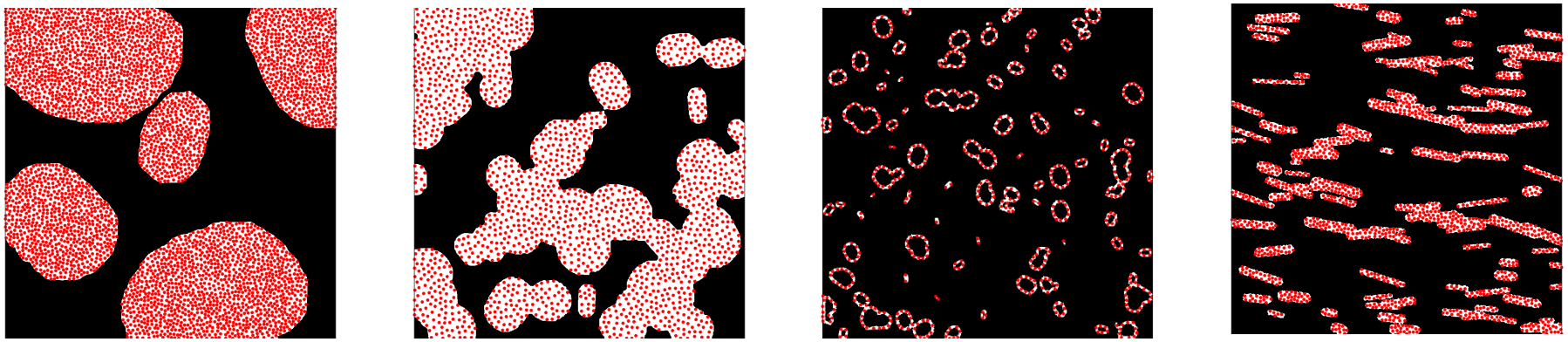
- Objective estimation of tumor pattern density (~~frame size~~)
- Estimation of tumor pattern size (~~connected component labeling~~)
- Robustness to noise (node identification errors)
- Low computational cost (WSI)

Limitations:

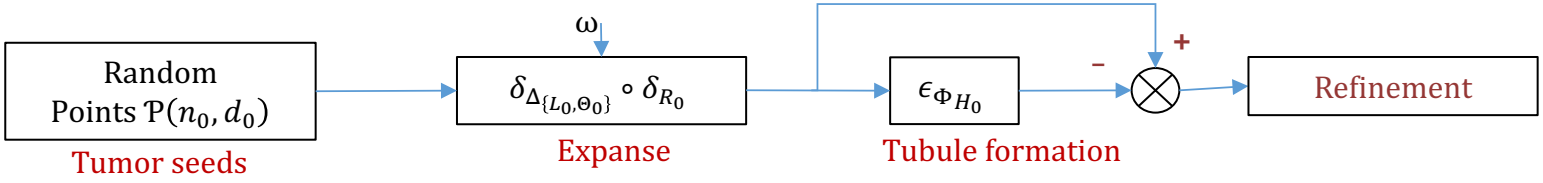
- Unsupervised learning (unlabeled data)
- Qualitative evaluation / prospective analysis

5) Simulation

Tumor Patterns:



(a) (b) (c) (d)



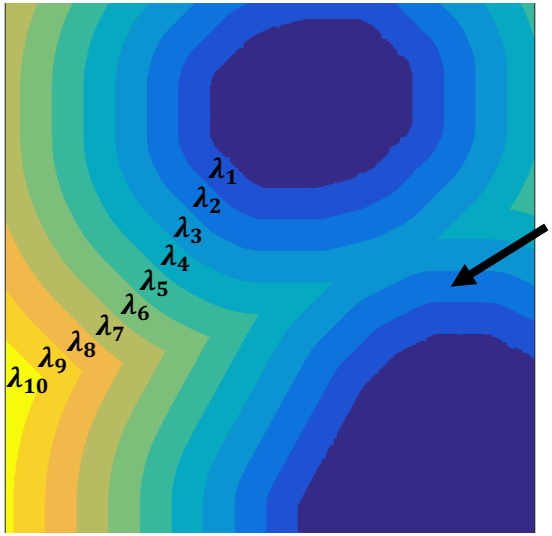
	n_0	d_0
(a)	$\mathcal{N}_1(7,2)$	800
(b)	$\mathcal{N}_1(40,8)$	100
(c)	$\mathcal{N}_1(85,5)$	40
(d)	$\mathcal{N}_1(100,5)$	40

$\mathcal{N}_n(\mu,\sigma)$: n random values generated from a normal distribution

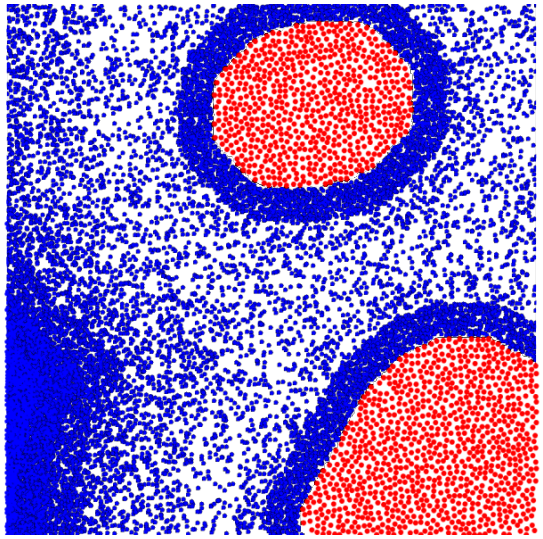
$\mathcal{U}_n(a,b)$: n random values generated from a discrete uniform distribution

5) Simulation

Tumor Micro-Environment:



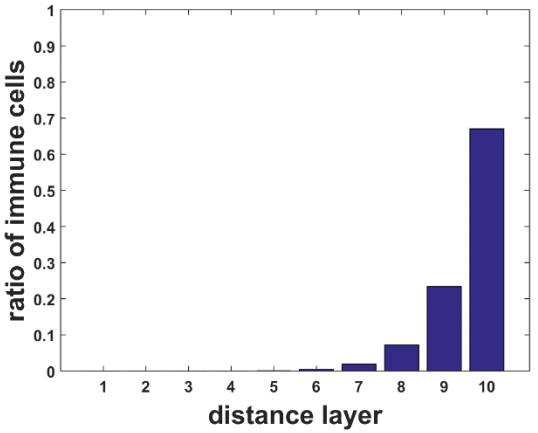
$n_L \times \lambda_i$ immune cells



- Distance map $(k = 10)$

$\alpha = 0, 1, 2, 10$

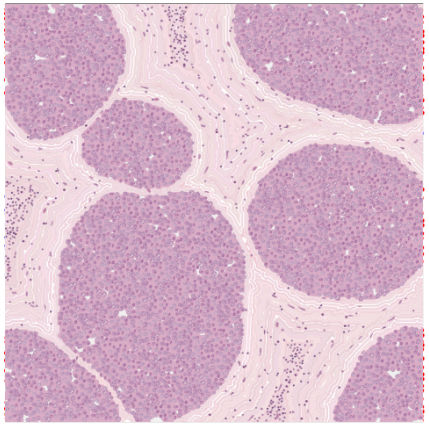
- k: number of layers
- Weight coefficients: $\lambda_i = \frac{i^{-\alpha}}{\sum_{i=1}^{k} i^{-\alpha}}$
- n_L: desired number of immune cells



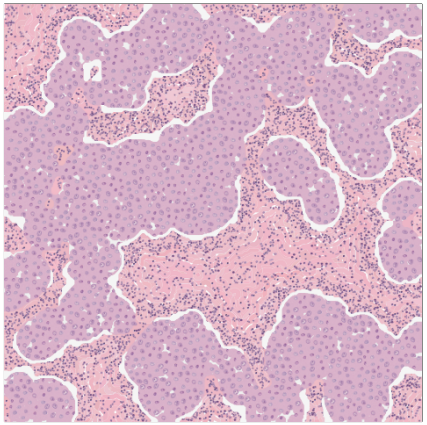
- $\alpha > 0$: immune cells close to tumor

5) Simulation

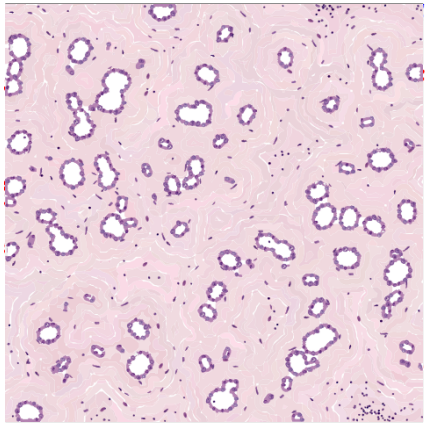
Synthetic data



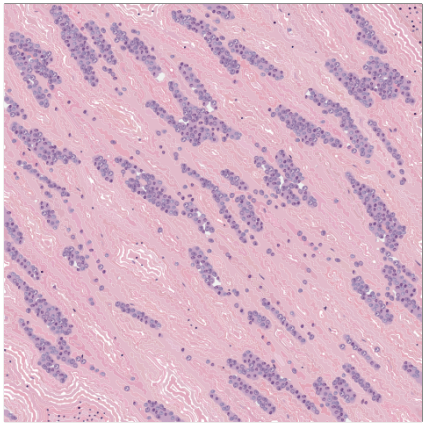
(a)



(b)

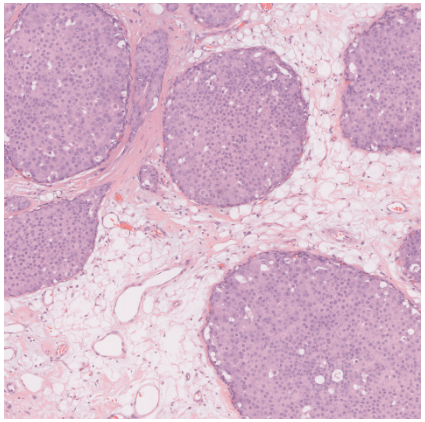


(c)

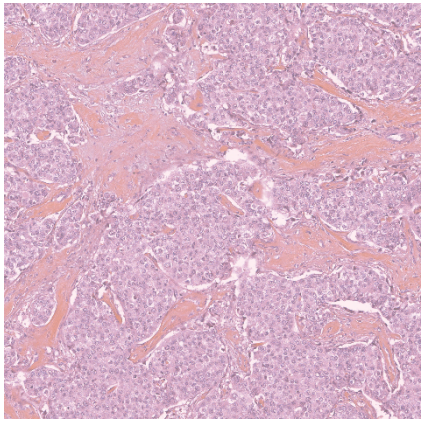


(d)

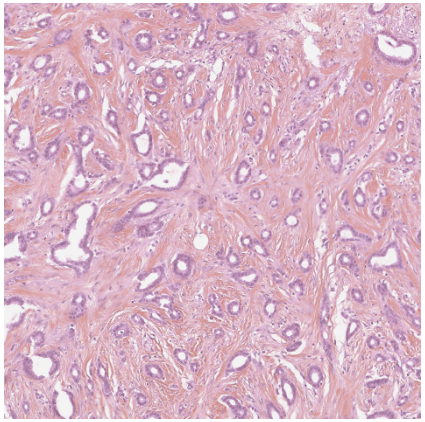
Real data



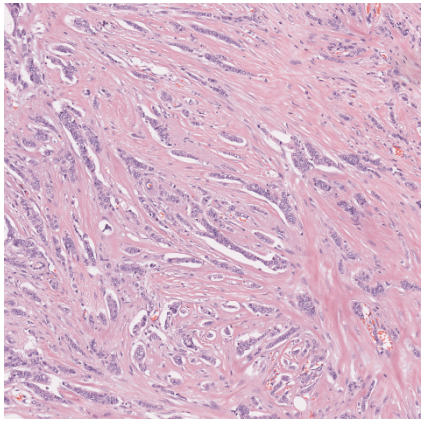
DCIS



IDC



TC



ILC

6) Perspectives

Graph-based Mathematical Morphology:

- **Significance:** different spatial point-set configurations
- **Comparison:** SSA / network analysis / spectral analysis / spatial statistics
- **Effectiveness:** stability of morphological functions *vs* noise
- **Variation:** different graphs / linking rules / parameters
- **Exploration:** new morphological functions / spatial aspects
- **Adaptation:** Spatial Point Pattern Analysis (GIS, astronomy, histopathology,…)

Tumor Heterogeneity:

- **Entities:** TME, IHC (+/-), FISH
- **Implication:** clinical parameters, cancer subtypes, gene expression
- **Comparison:** medical imaging parameters (US, Elastography,…)
- **Evolution:** tumor growth (culture / xenograft), 3D/4D
- **Anisotropy:** Collagen orientation (TACS)

Thanks

For Your Attention

September 26th, 2017

bassem.bencheikh@lib.upmc.fr

References

[1] FDA Press Announcements. “FDA allows marketing of first whole slide imaging system for digital pathology April 12”, (2017)

[2] Sinn HP et al. “A brief overview of the WHO classification of breast tumors”. Breast Care 8 (2013)

[3] Pattabiraman DR. and Weinberg RA. “Tackling the cancer stem cells - what challenges do they pose?” Nature reviews Drug disco (2014)

[4] W. H. Fridman et al, “*The immune contexture in human tumours: impact on clinical outcome*”, Nature Reviews Cancer, 2012

[5] G.S.K. et al, “*Cancer-Associated Fibroblasts Drive the Progression of Metastasis through both Paracrine and Mechanical Pressure on Cancer Tissue*”, Mol Cr Res, 2012

[6] M. Wagner et al, “*Inflamed tumor-associated adipose tissue is a depot for macrophages that stimulate tumor growth and angiogenesis*”, Angiogenesis, 2012.

[7] P. Provenzano et al, “*Collagen reorganization at the tumor-stromal interface facilitates local invasion”*, BMC Medicine 2006

[8] DeLisser H.M. and al. “Involvement of endothelial PECAM-1/CD31 in angiogenesis”. The American journal of pathology (1997)

[9] Images from: University of British Columbia - Slide Box

[10] Raymond E. et al. “Germinal center analysis with the tools of mathematical morphology on graphs”. Cytometry (1993)

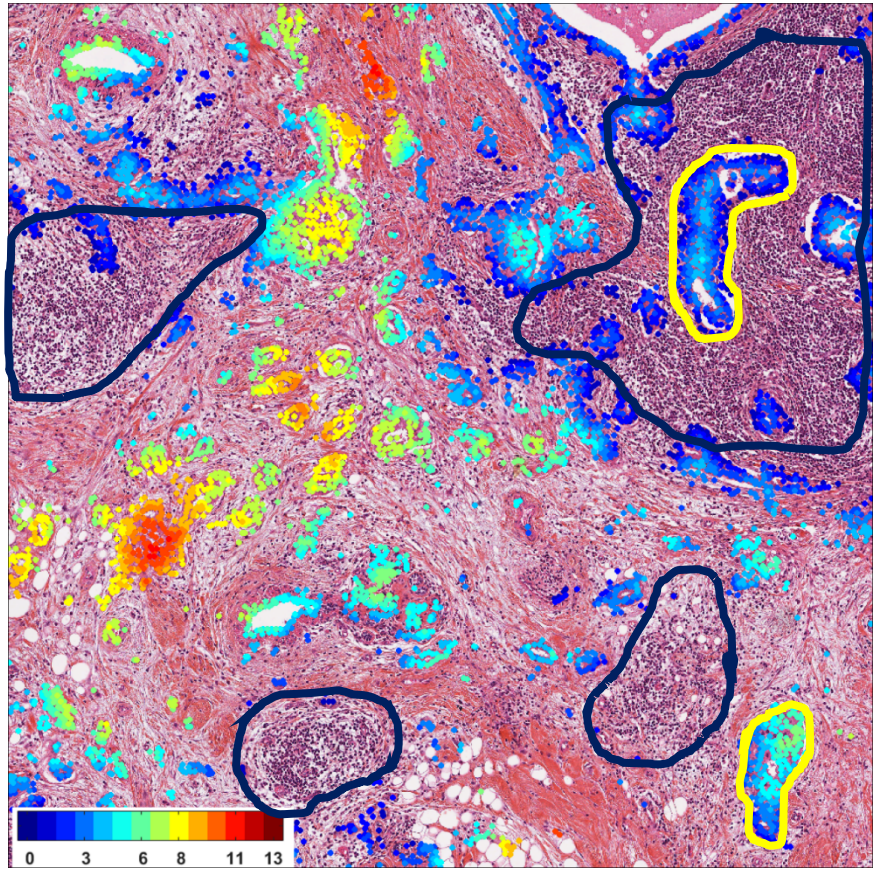
[11] Luc Vincent, "Graphs and Mathematical Morphology", Signal Processing, 1989.

[12] Reinhard E. et al. “Color transfer between images”. IEEE Computer graphics and applications (2001)

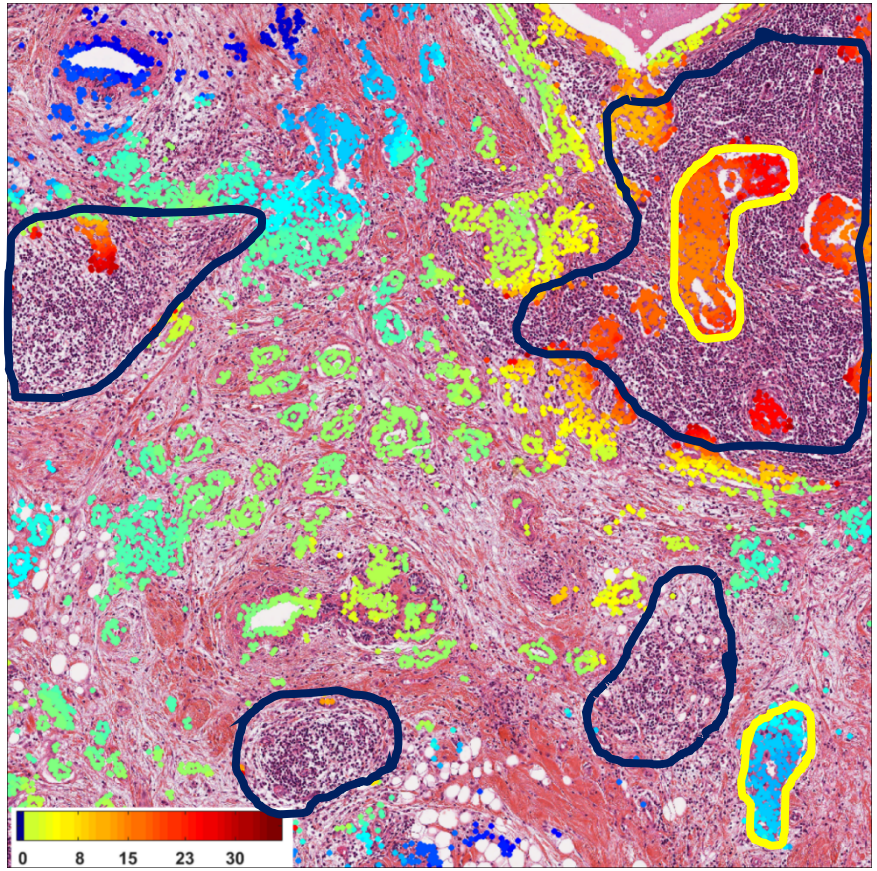
[13] Achanta R. et al. “SLIC Superpixels”. EPFL Technical Report no. 149300 (2010)

[14] Ruifrok A.C. et al. “Quantification of histochemical staining by color deconvolution”. In: Analytical and quantitative cytology and histology (2001)

[15] Xu J. et al. “Stacked sparse auto-encoder (SSAE) for nuclei detection on breast cancer histopathology images”. IEEE trans. on medical imaging (2016)

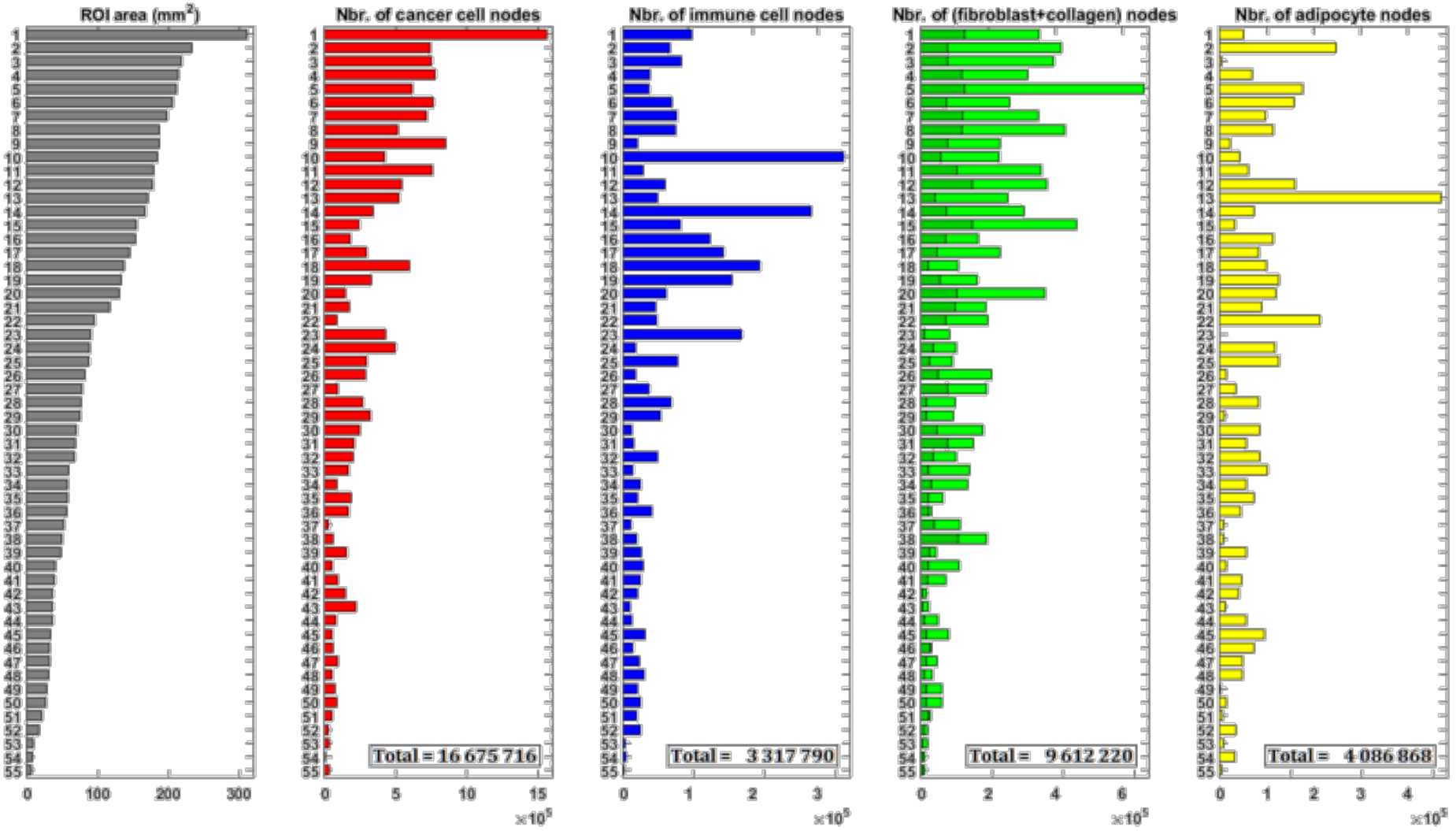


$\mathrm{D}(G_X) * G_Y$

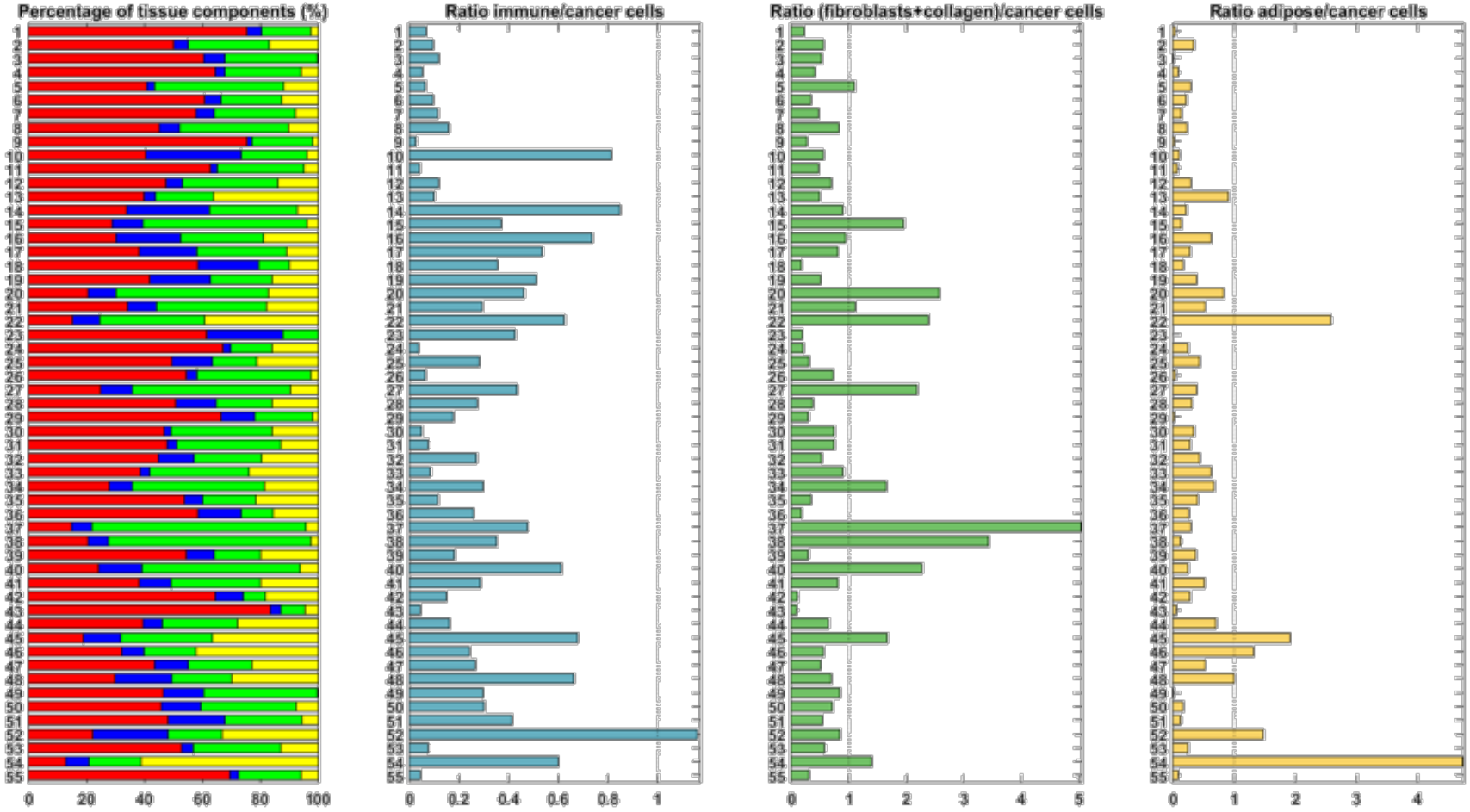


$\Gamma(G_X) * G_Y$

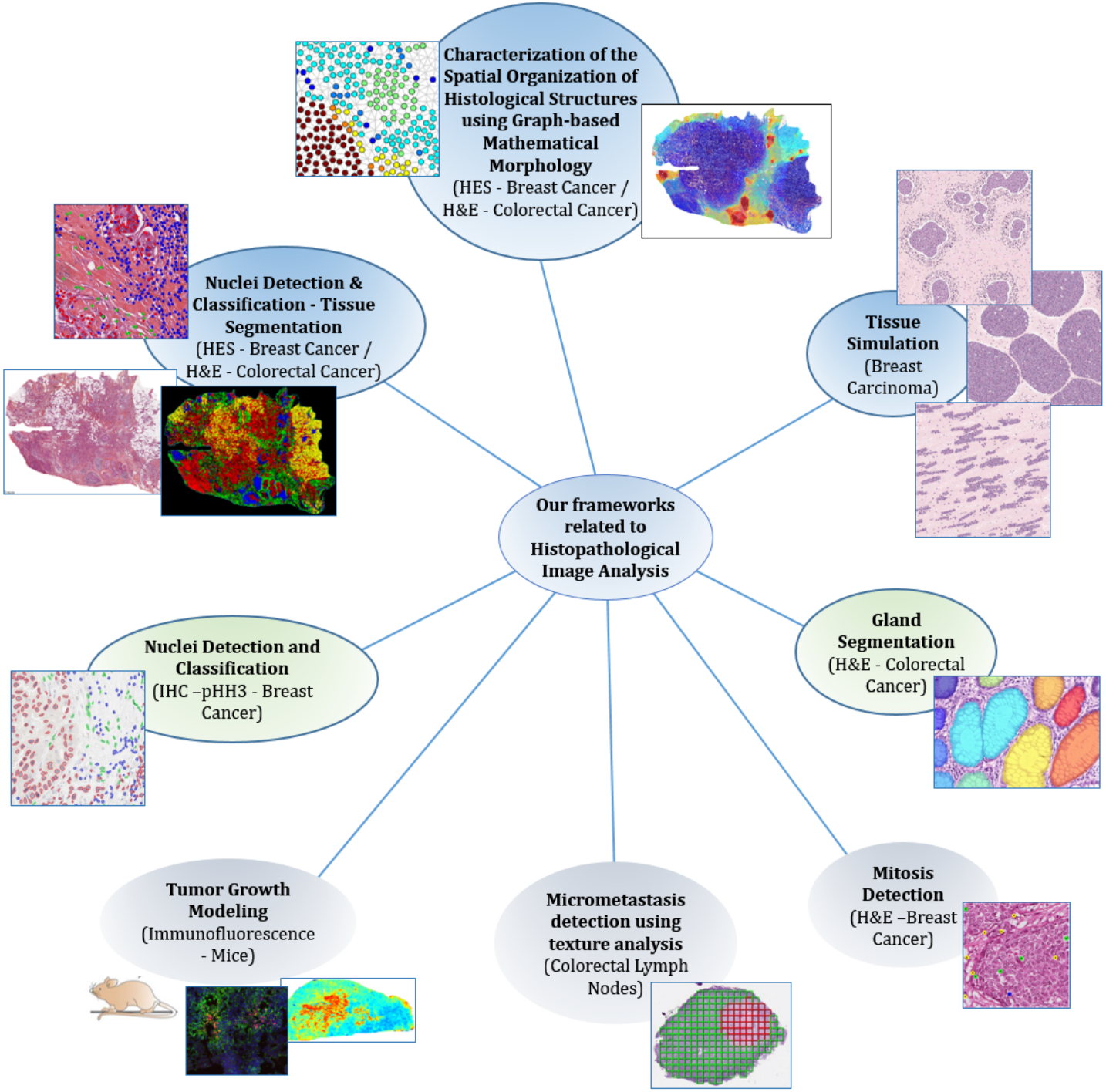
- X = immune cells
- Y = cancer cells

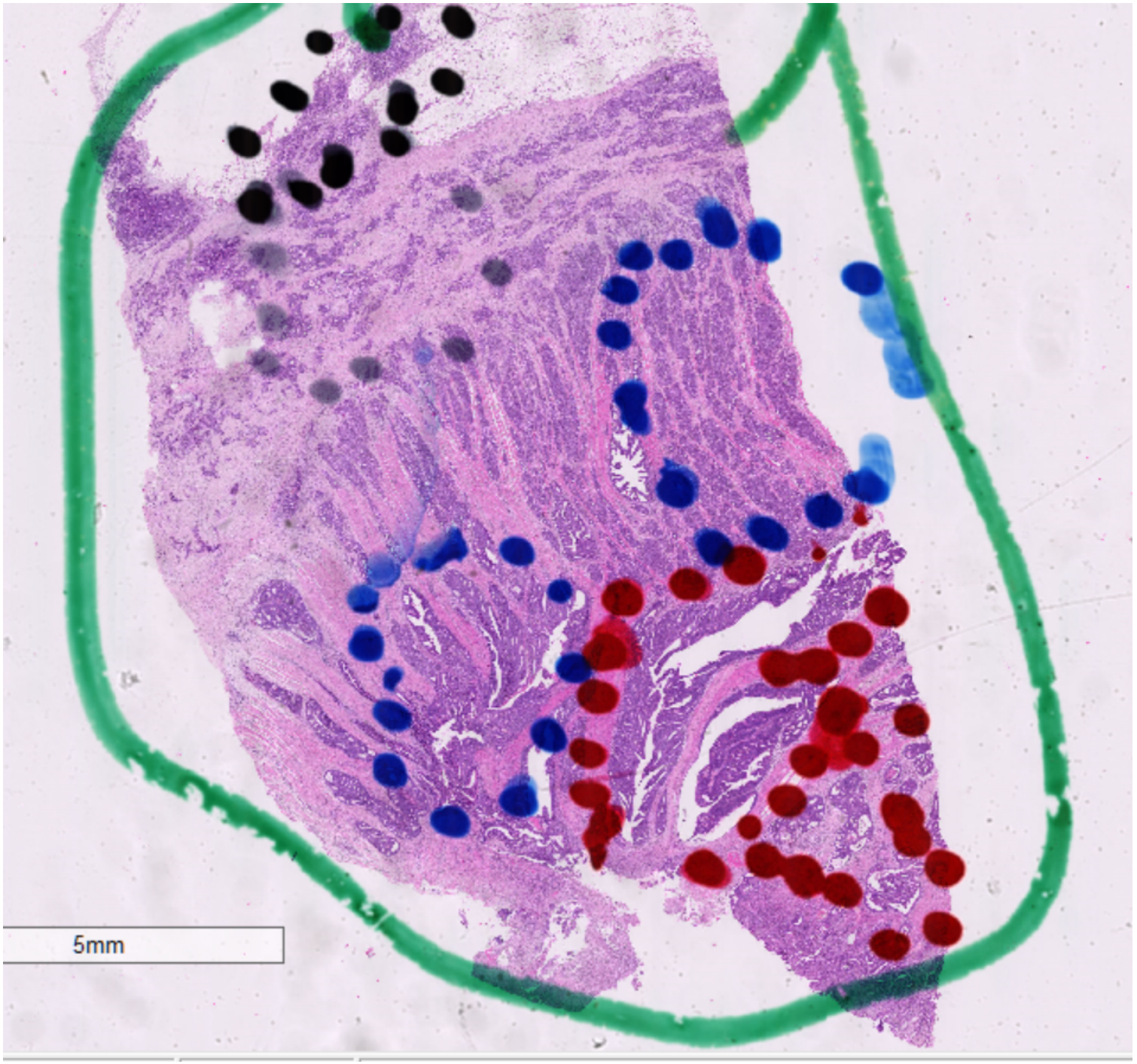


	ROI area	CCN	ICN	FCN	ATN
ROI area	1.0000	0.8893	0.4896	0.7719	0.4171
CCN	0.8893	1.0000	0.3116	0.5726	0.2684
ICN	0.4896	0.3116	1.0000	0.1966	0.1047
FCN	0.7719	0.5726	0.1966	1.0000	0.2487
ATN	0.4171	0.2684	0.1047	0.2487	1.0000

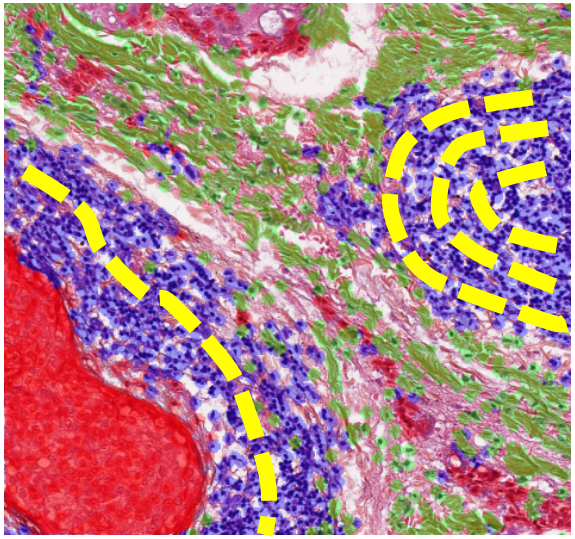


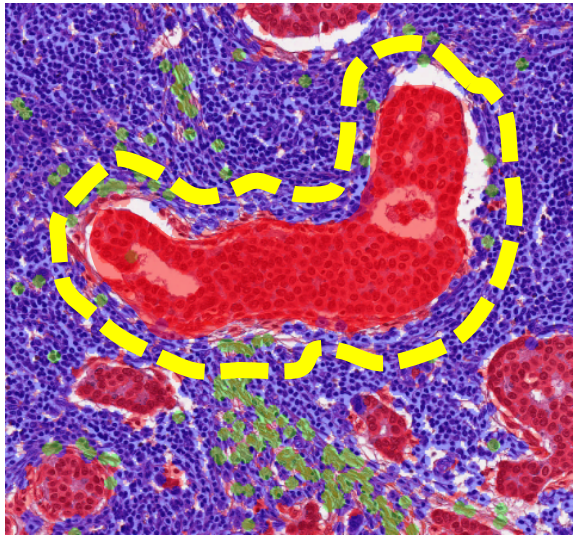
	ROI area	% CCN	% ICN	% FCN	% ATN	ICN/CCN	FCN/CCN	ATN/CCN
ROI area	1.0000	0.4815	-0.1443	-0.3043	-0.2178	-0.3024	-0.3127	-0.2703
% CCN	0.4815	1.0000	-0.2767	-0.6003	-0.5036	-0.6746	-0.7359	-0.5728
% ICN	-0.1443	-0.2767	1.0000	-0.1699	-0.0258	0.8135	-0.0442	0.0166
% FCN	-0.3043	-0.6003	-0.1699	1.0000	-0.2466	0.1303	0.8654	-0.0750
% ATN	-0.2178	-0.5036	-0.0258	-0.2466	1.0000	0.2825	0.0225	0.8676
ICN/CCN	-0.3024	-0.6746	0.8135	0.1303	0.2825	1.0000	0.3577	0.4115
FCN/CCN	-0.3127	-0.7359	-0.0442	0.8654	0.0225	0.3577	1.0000	0.2222
ATN/CCN	-0.2703	-0.5728	0.0166	-0.0750	0.8676	0.4115	0.2222	1.0000

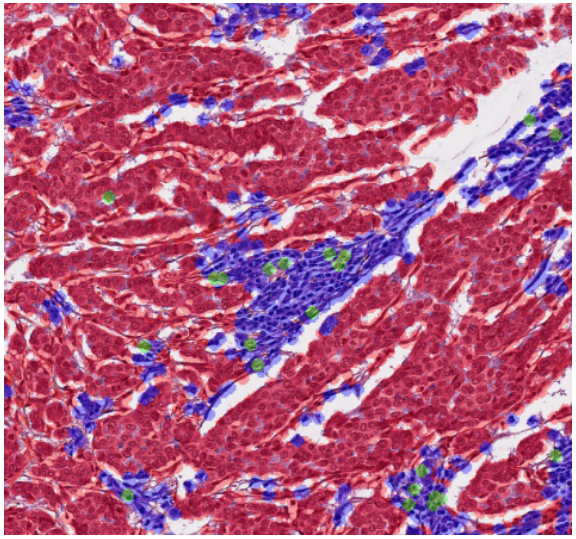

Our frameworks related to Histopathological Image Analysis
Characterization of the Spatial Organization of Histological Structures using Graph-based Mathematical Morphology
(HES - Breast Cancer / H&E - Colorectal Cancer)
Nuclei Detection & Classification - Tissue Segmentation
(HES - Breast Cancer / H&E - Colorectal Cancer)
Tissue Simulation
(Breast Carcinoma)
Nuclei Detection and Classification
(IHC –pHH3 - Breast Cancer)
Gland Segmentation
(H&E - Colorectal Cancer)
Tumor Growth Modeling
(Immunofluorescence - Mice)
Micrometastasis detection using texture analysis
(Colorectal Lymph Nodes)
Mitosis Detection
(H&E –Breast Cancer)


5mm

❑ Spatial heterogeneity of immune cells:

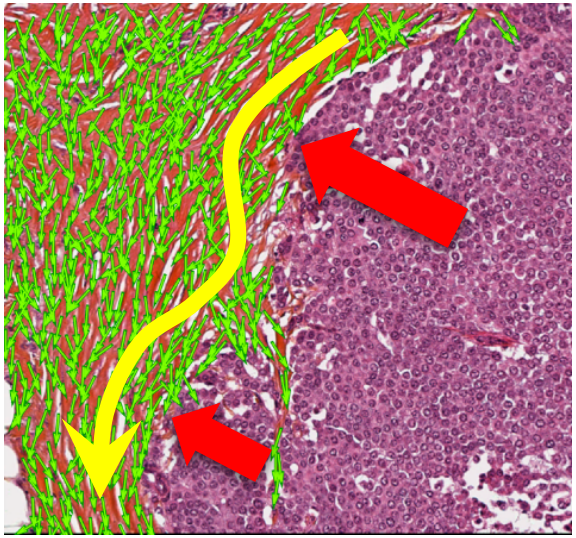


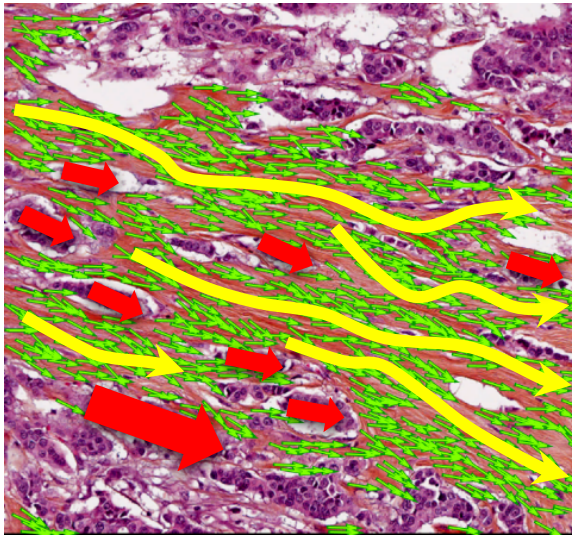


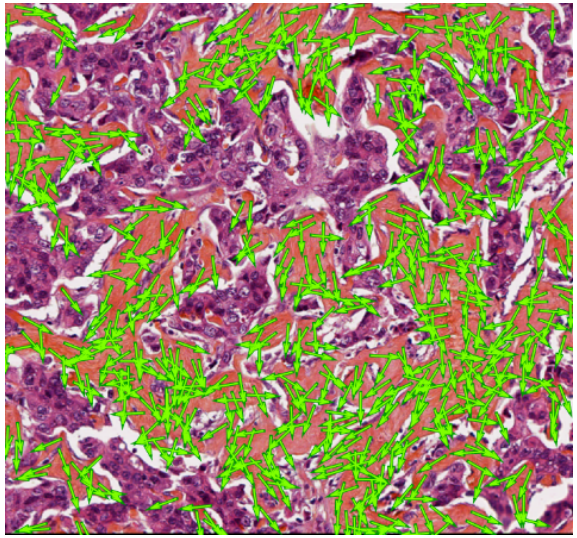


- **Relative abundance**
- **Spatial proximity**
- **Surroundedness**

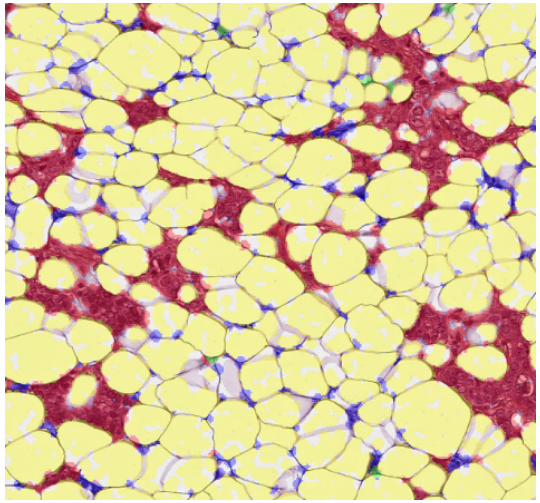
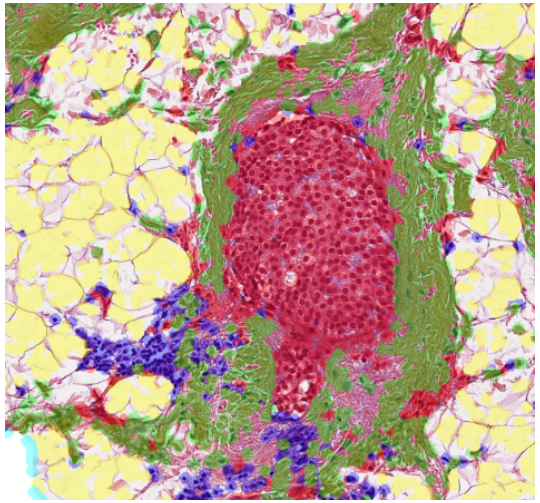
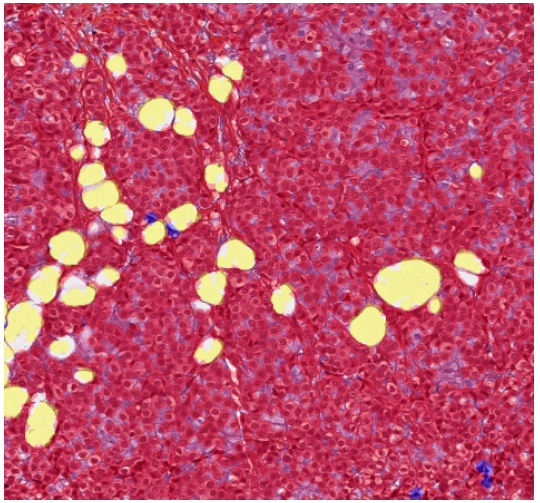
❑ Heterogeneity of the spatial organization of collagen fibers:







- **Orientation**
- **Density**
- **Spatial proximity**
- **Surroundedness**
- **Stretchiness**

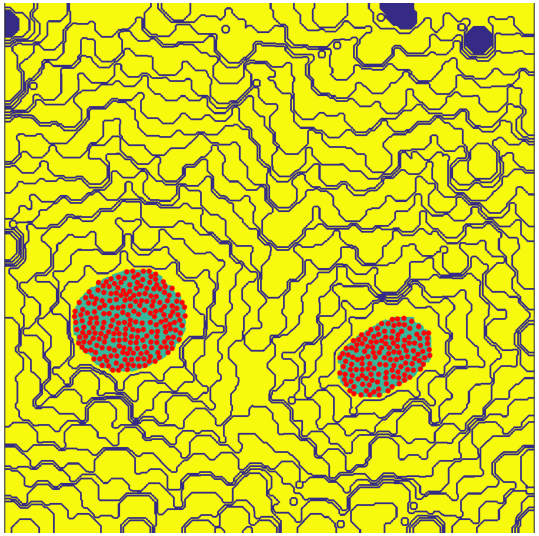
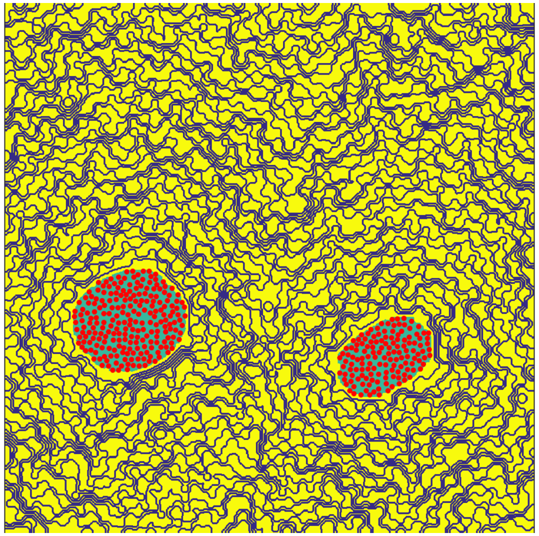
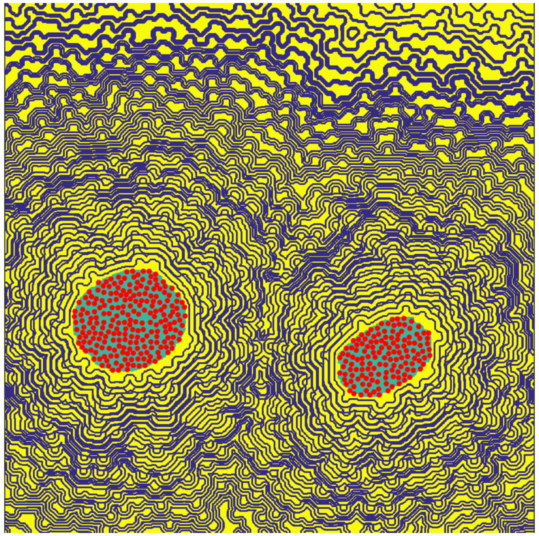




Example of spatial heterogeneity of adipose tissue

❑ A model of collagen spatial arrangement:

- Reshaping (parameter $\boldsymbol{\gamma}$)
- Distance map
- Quantization (parameter $\boldsymbol{k}$)
- Dilation (parameter $\boldsymbol{R}$)

➢ Morphological of deformations of tumor patterns define various distance maps, which are quantized.

stretchiness	**thickness**	**density and spatial proximity**
		
$\{k, R, \gamma\}=$ {1, 40, 60}	{1, 20, 30}	[illegible]

❑ A model of collagen spatial arrangement:

- Reshaping (parameter $\boldsymbol{\gamma}$)
- Distance map
- Quantization (parameter $\boldsymbol{k}$)
- Dilation (parameter $\boldsymbol{R}$)

➢ Morphological of deformations of tumor patterns define various distance maps, which are quantized.

Orientation relative to tumor