Statistical analysis of spatial patterns in tumor microenvironment images: Supplementary Information

Mohamed M. Benimam[1,*] Vannary Meas-Yedid [1,*] Suvadip Mukherjee[1,2,*]
Astri Frafjord[3] Alexandre Corthay[3,4]
Thibault Lagache[1,†,‡] Jean-Christophe Olivo-Marin[1,†,‡]

[1] *Institut Pasteur, Université de Paris Cité, CNRS UMR 3691, BioImage Analysis Unit 75015 Paris, France*
[2] *Current affiliation: KLA Corporation, Ann Arbor, Michigan, USA*
[3] *Tumor Immunology Lab, Department of Pathology, Rikshospitalet, Oslo University Hospital and University of Oslo, Oslo, Norway*
[4] *Hybrid Technology Hub - Centre of Excellence, Institute of Basic Medical Sciences. University of Oslo, Oslo, Norway*
[*] *These authors contributed equally*, [†] *These authors jointly supervised this work*
[‡] *Corresponding authors: thibault.lagache@pasteur.fr, jcolivo@pasteur.fr*

Zone Width (μm)	RMSE ACI	RMSE SAI	RMSE ASI
5	6.69	0.66	0.31
10	6.69	0.64	0.39
20	6.81	1.61	0.12
50	6.69	2.22	0.13

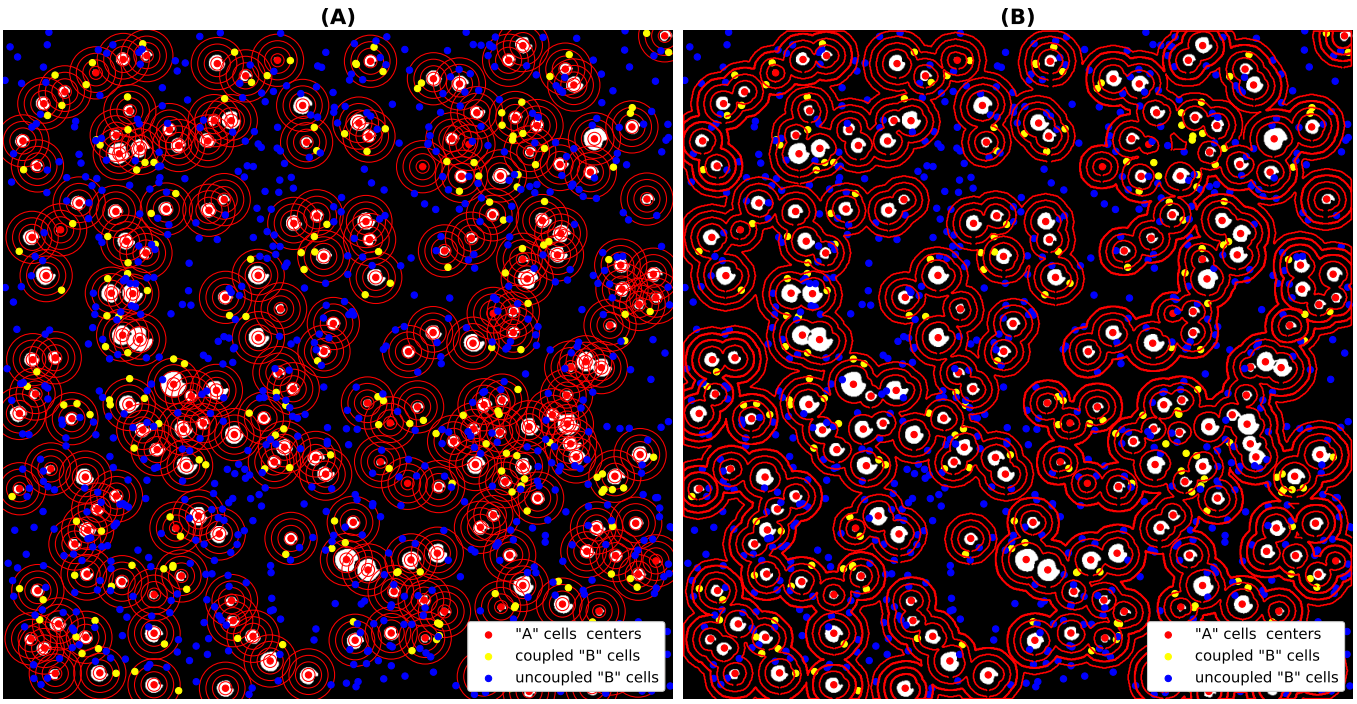
Supplementary Table I: Root mean square error (RMSE) between simulated spatial association (α) and estimation indices (ACI, SAI and ASI) for different zone widths.

CellType	Total Cells	Cells/Patient				
		Mean	Std	Median	Min	Max
Mast cells	5919	739.875	406.715	788.0	97	1241
T cells	100369	12546.125	11845.931	10562.0	2202	37514

Supplementary Table II: Distribution of the number of T cells and mast cells in all NSCLC patients ($n = 8$).

	ASI					delta					log(pvalue)				
configuration	median	min	25th	75th	max	median	min	25th	75th	max	median	min	25th	75th	max
T cells to mast cells	0.426	0.196	0.308	0.529	0.802	9.948	5.845	7.798	10.539	11.698	-17.600	-33666.447	-439.376	-10.259	-7.023
Mast cells to T cells	0.221	0.056	0.185	0.321	0.513	31.863	20.759	27.981	34.653	48.395	-79.020	-259348.131	-21273.046	-29.533	-13.294
TER to T cells	0.188	0.004	0.042	0.232	0.297	76.114	54.424	70.546	84.006	100.531	-10.362	-95.075	-40.606	-4.104	-1.436
TER to mast cells	0.115	0.000	0.052	0.243	0.460	12.019	-1.000	6.149	15.868	39.031	-3.637	-46459.356	-6.952	-1.605	0.000

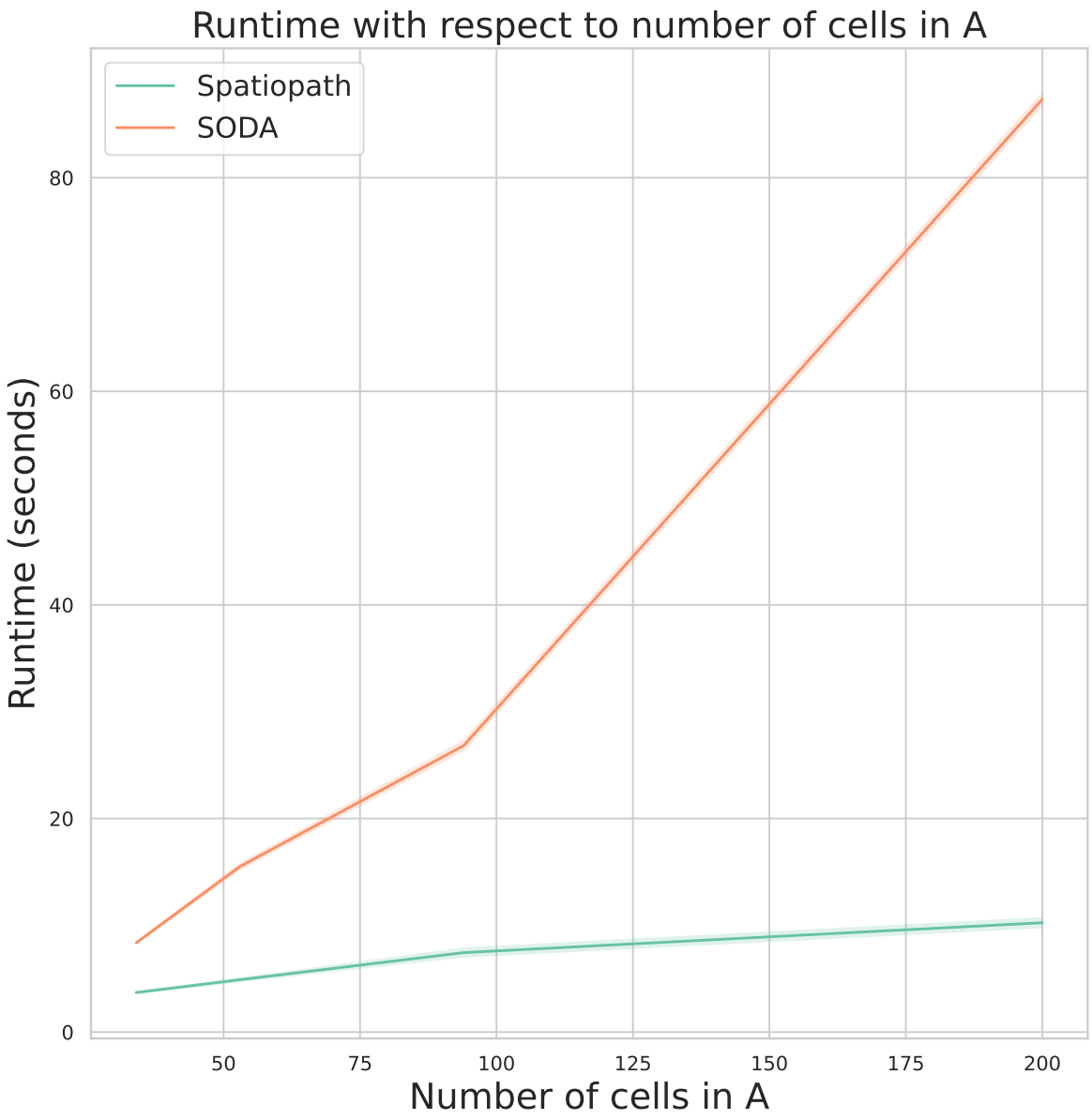
Supplementary Table III: Table of ASI, delta, and p-value statistics



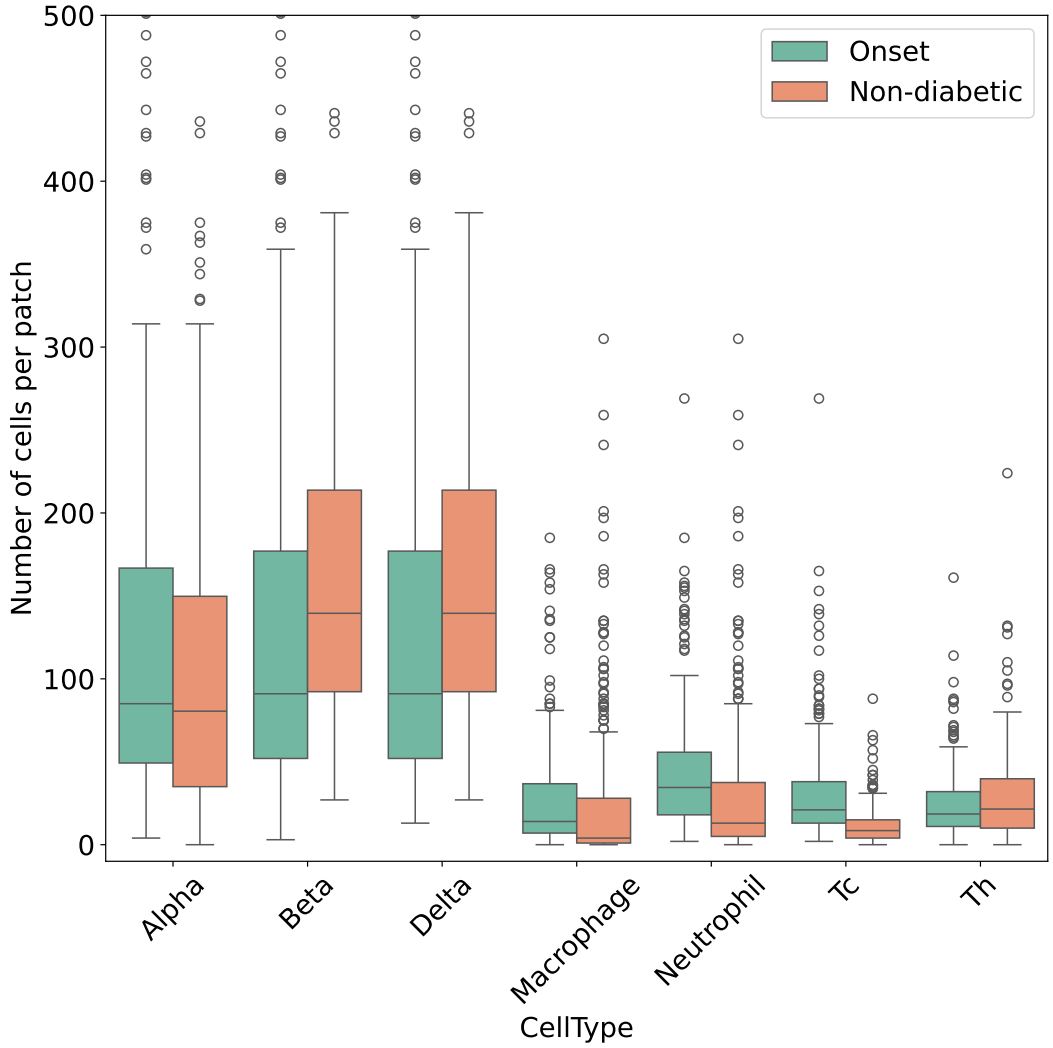
Supplementary Figure 1: **Methodology differences between SPATIOPATH and SODA for quantifying spatial associations. (A)** SODA reduces A cells (white masks) to their positional center-of-mass (red dots) and uses Ripley's K function to calculate the spatial accumulation of points (set B) around other points (set A) in expanding concentric rings (red circles around A cells). A subset of B points are randomly distributed in Ω (blue dots), the other B points being coupled to A cells (yellow dots). After statistical thresholding of the K function, the number of associated (coupled) B points is determined. An unmixing correction is applied when the coupling distance is within the typical inter-point spacing of set A, preventing overestimation of coupling due to closely spaced A points sharing B points. **(B)** In contrast, SPATIOPATH represents all cells using a unified level-set function (red isolines), enabling it to manage large objects and intricate cellular architectures. This approach also enhances computational efficiency and precision when analyzing association patterns across varying distances.

CellType	Patient	Cells/Patient	Cells/Patch				
			Mean	Std	Median	Min	Max
Mast cells	A01	854	50.24	29.66	41.00	4	115
	A02	97	5.39	5.95	2.50	0	19
	A03	1240	40.00	26.57	34.00	5	128
	A04	325	46.43	13.50	52.00	25	60
	A05	722	65.64	35.33	73.00	8	116
	A06	551	27.55	18.77	30.50	2	74
	A07	889	42.33	29.56	44.00	3	92
	A08	1241	155.13	128.02	143.00	31	399
T Cells	A01	16572	974.82	505.63	899.00	226	1740
	A02	2202	122.33	94.10	98.00	20	357
	A03	16646	536.97	244.51	510.00	120	1121
	A04	3180	454.29	450.06	329.00	76	1441
	A05	6401	581.91	372.19	410.00	105	1208
	A06	37514	1875.70	798.89	1766.00	1020	4404
	A07	14723	701.10	419.60	616.00	162	1714
	A08	3131	391.38	116.73	354.00	218	563

Supplementary Table IV: **Patch Summary**. The column Cells/Patient refers to the total number of cells per patient and the other Cells/Patch columns (Mean, Std, Median, Min, Max) to these values per selected patch within the tumor region.



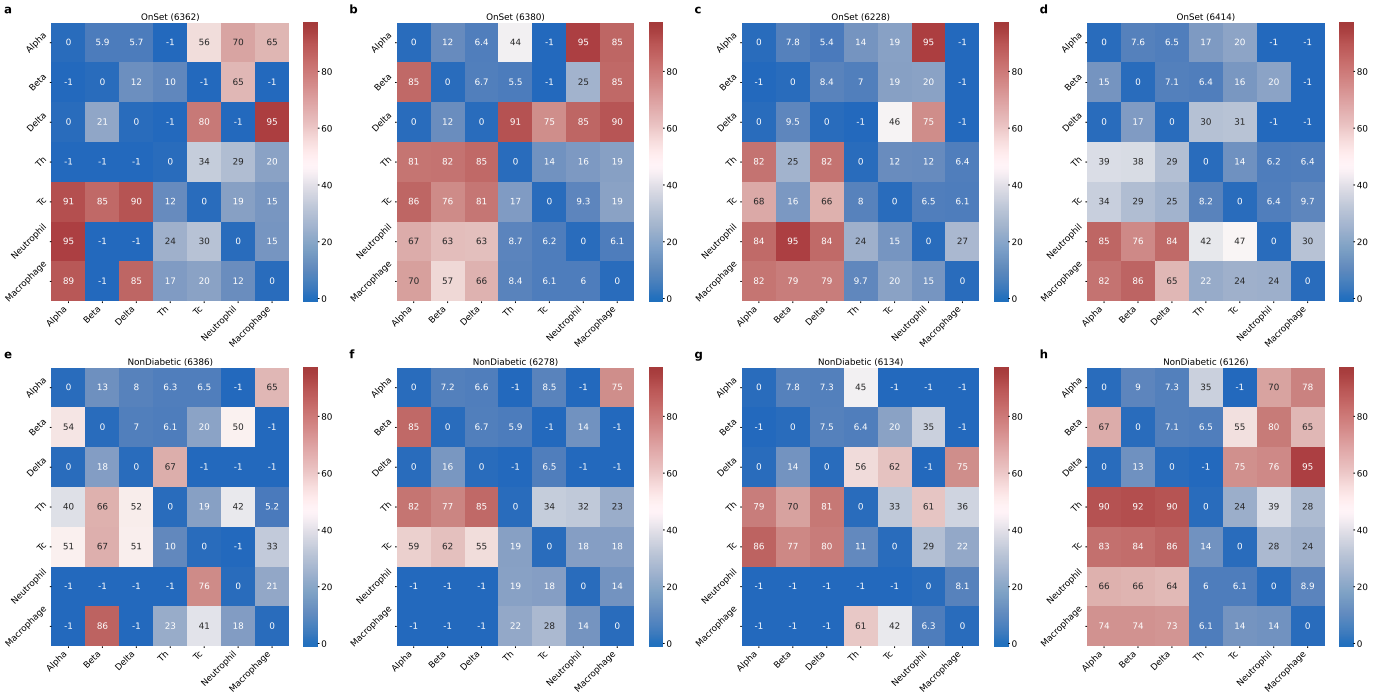
Supplementary Figure 2: **Computational analysis** The figure compares the computational time (in seconds) for SPATIOPATH in green, *versus* SODA in red with respect to the number of A cells. Solid lines are the mean values and shaded envelopes correspond to $\pm$ standard error ($n = 10$ simulations per number of A cells). Source data are provided as a Source Data file.



Supplementary Figure 3: **Numbers of the different cell types in non-diabetic patients and those with onset diabetes**. Boxplots of the different cell numbers in IMC images of non-diabetic patients (n=274 patches, red) and patients with onset diabetes (n=290 patches, green). Source data are provided as a Source Data file.

Specificity of primary antibody	**Antibody clone**	**Species/ Isotype**	**Final dilution (conc. µg/ml)**	**Supplier**	**Cat. No.**	**Incubation time**	**Isotype-matched control - Supplier(Cat. No.)**
CD3	2GV6	Rabbit IgG	ready to use (0.40)	Roche	790-4341	40 min	Vector Laboratories (I-1000)
Tryptase	EPR8476	Rabbit IgG	1:500 (3.0)	Abcam	Ab134932	60 min	Vector Laboratories (I-1000)
Cytokeratins	AE1/AE3/PCK26	Mouse IgG1	ready to use (48.0)	Roche	760-2135	20 min	Dako (X0931)

Supplementary Table V: List of primary antibodies used to detect immune cells and cytokeratins by IHC.

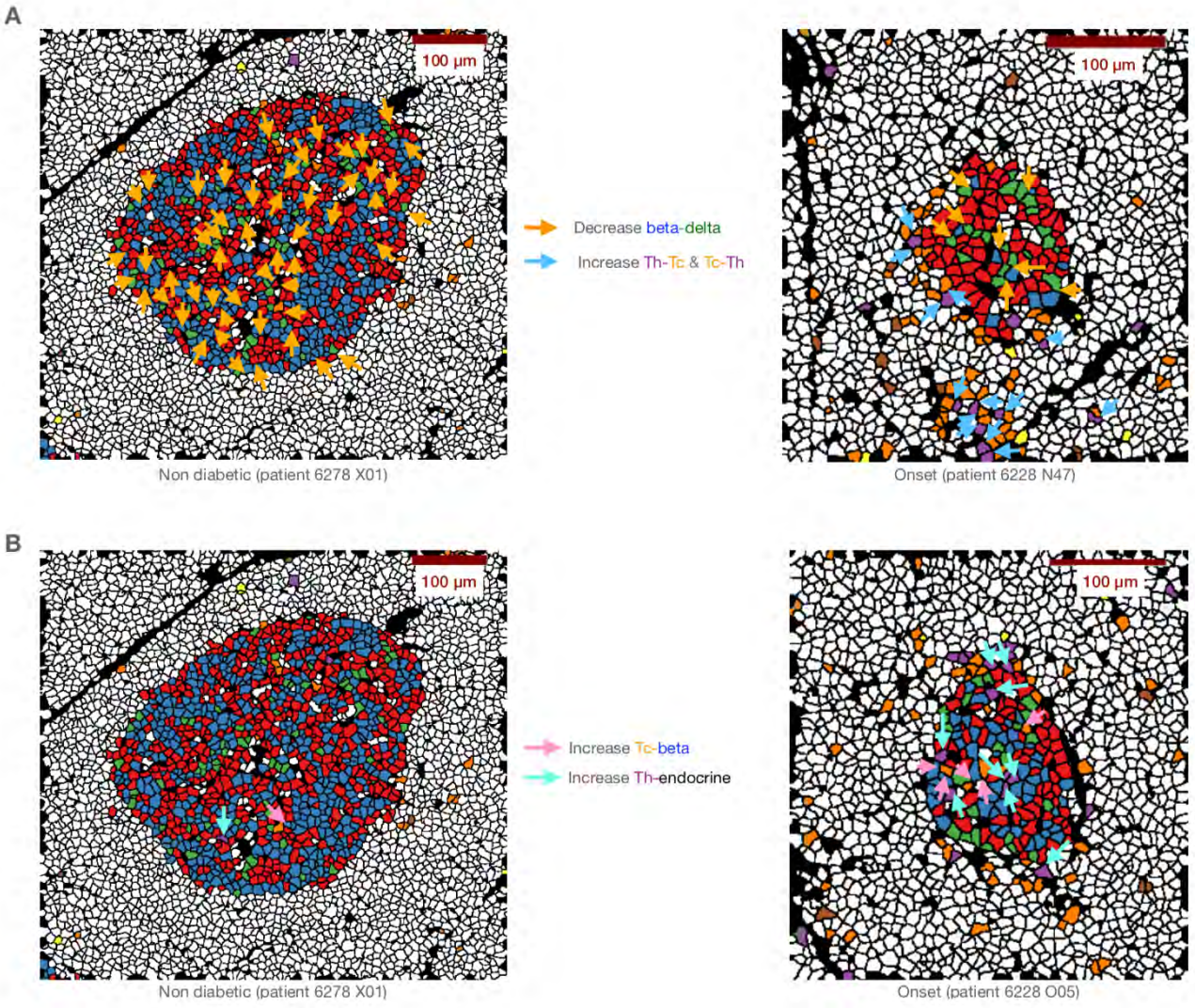


Supplementary Figure 4: **Mean association distance (δ_a in μm) between endocrine and immune cells in non-diabetic patients and those with onset diabetes**. For each patient of each diabetes stage (Onset diabetes (first row) and non-diabetic (second row)) we measured the mean distance of spatial association (δ_a) between the different endocrine cell subsets (alpha, beta, and delta) and the immune cell subsets (T helper (Th), T cytotoxic (Tc), neutrophils and macrophages). The element (i,j) of each table is equal to the association distance of cell type j to cell type i. Source data are provided as a Source Data file.

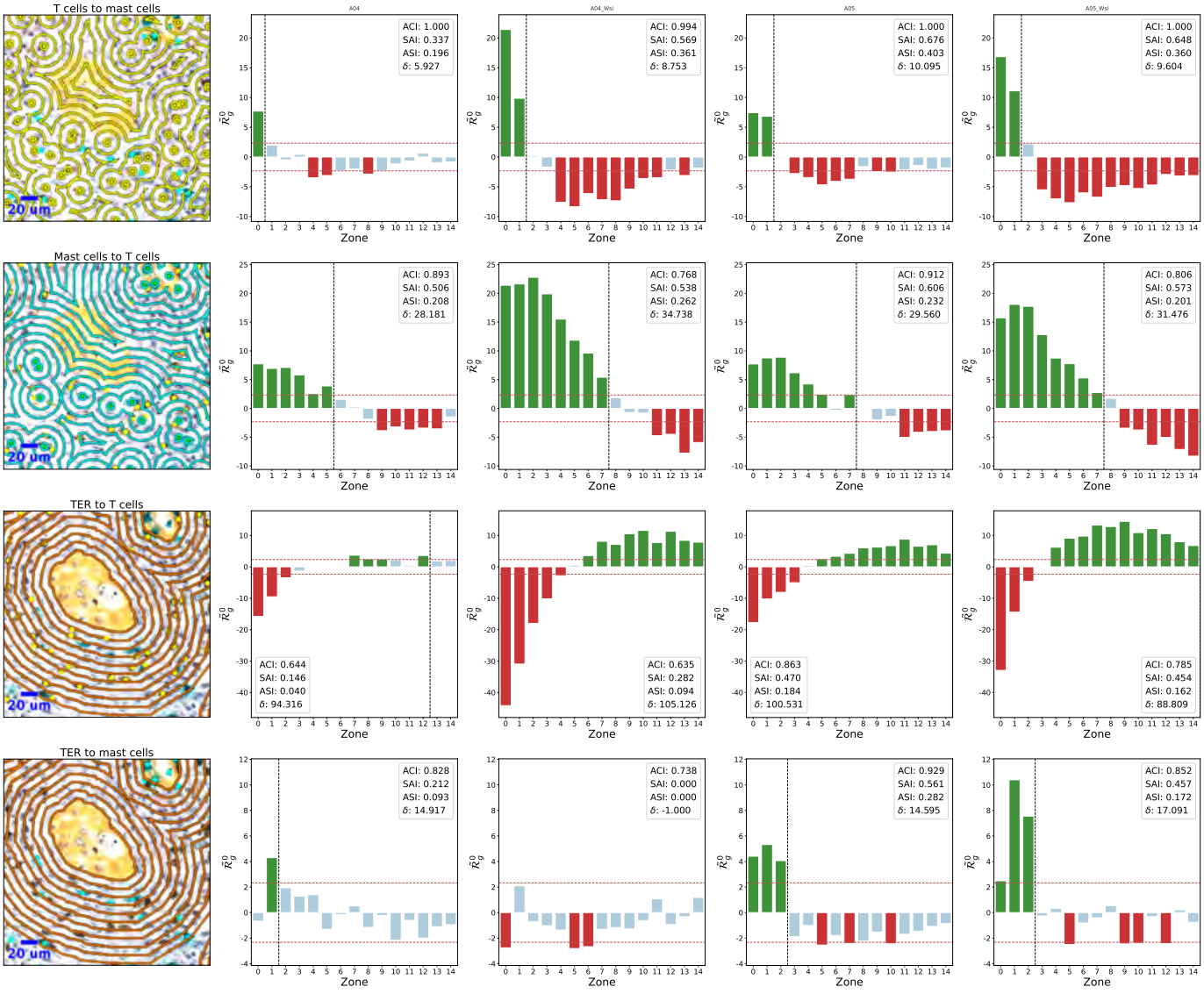
Abbreviations: AP, alkaline phosphatase; HRP, horseradish peroxidase; NP, hapten nitropyrazole.

Primary antibody	**Secondary antibodies**	**Dilution**	**Supplier**	**Cat. No.**	**Incubation time**
Anti-CD3 rabbit IgG	DISCOVERY OmniMap IgG HRP anti-rabbit	Ready to use	Roche	760-4311	16 min
Anti-Tryptase rabbit IgG	DISCOVERY OmniMap IgG HRP anti-rabbit	Ready to use	Roche	760-4311	16 min
Anti-Cytokeratins Mouse IgG1	DISCOVERY UltraMap IgG AP anti-mouse	Ready to use	Roche	760-4312	16 min

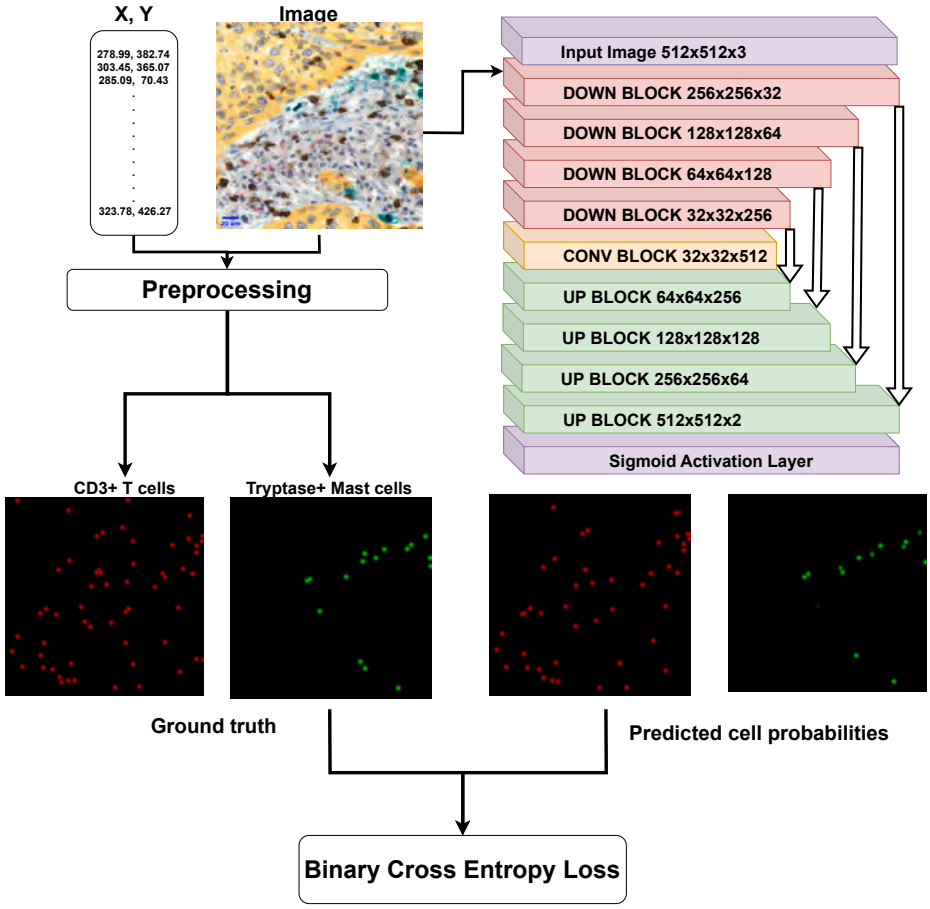
Supplementary Table VI: List of secondary antibodies used for multiplex IHC.



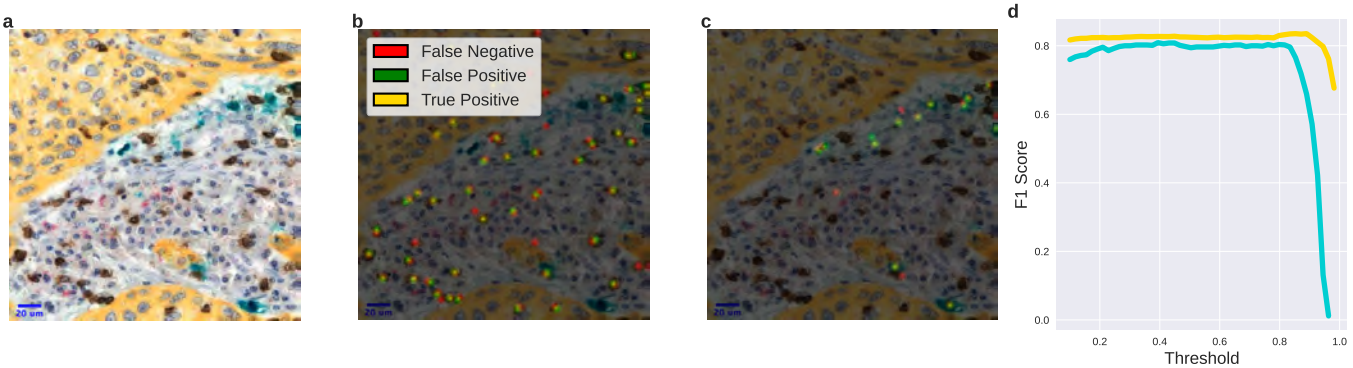
Supplementary Figure 5: **Main changes in cell-to-cell spatial associations between non-diabetic patients and patients with onset diabetes (example illustrations)**. **A-** Decrease of the association of delta (green) to beta (blue) cells in patients with onset diabetes (right) compared to non diabetic (left) patients (orange arrows), and increase of Th (violet) - Tc (orange) apposition (blue arrows) **B-** Increase of the beta (blue) to Tc (orange) apposition in patients with onset diabetes (pink arrows), and increase of the infiltration of Th cells (violet) in endocrine islest (cyan arrows).



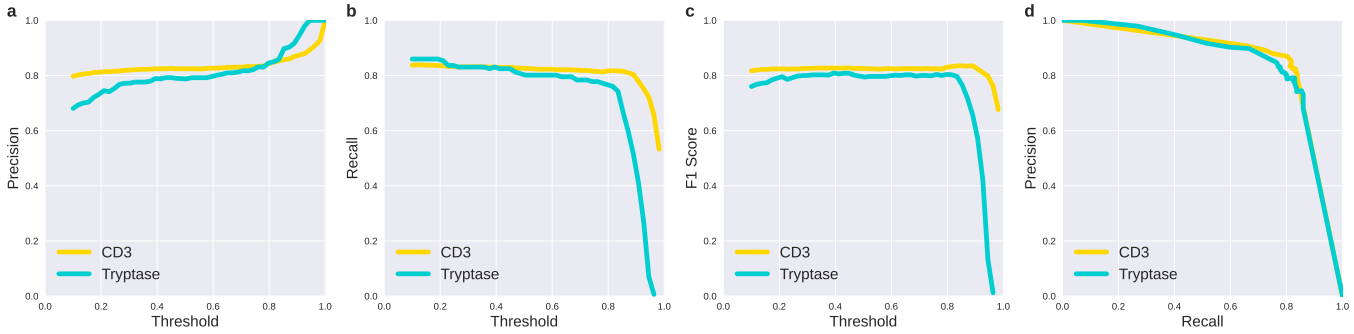
Supplementary Figure 6: Comparison of statistical analysis ($\tilde{R}_g^0$) performed over selected patches representing $\sim 10\%$ of the tumor region (columns 2 and 4 for patients A04 and A05) or the whole tumor region (columns 3 and 5 for patients A04 and A05). Source data are provided as a Source Data file.



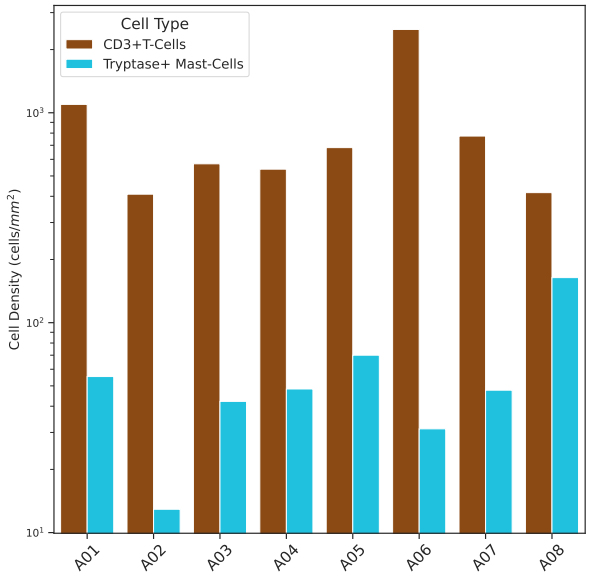
Supplementary Figure 7: **U-Net architecture and preprocessing pipeline for cell detection**. The detection process starts with an image and associated X, Y coordinates (ground truth). Coordinates are preprocessed (maximum between Gaussian convolutions), before being used as training data for the Convolutional Neural Network (U-Net). The U-Net model consists of 4 down-sampling (DOWN BLOCK) and up-sampling (UP BLOCK) layers with intermediate convolutional (CONV BLOCK) layers. The network outputs cell probabilities, which are compared to the ground truth masks generated from the X, Y coordinates. The Binary Cross-Entropy (BCE) loss function is used to compare the predicted cell probabilities with the ground truth and train the model.



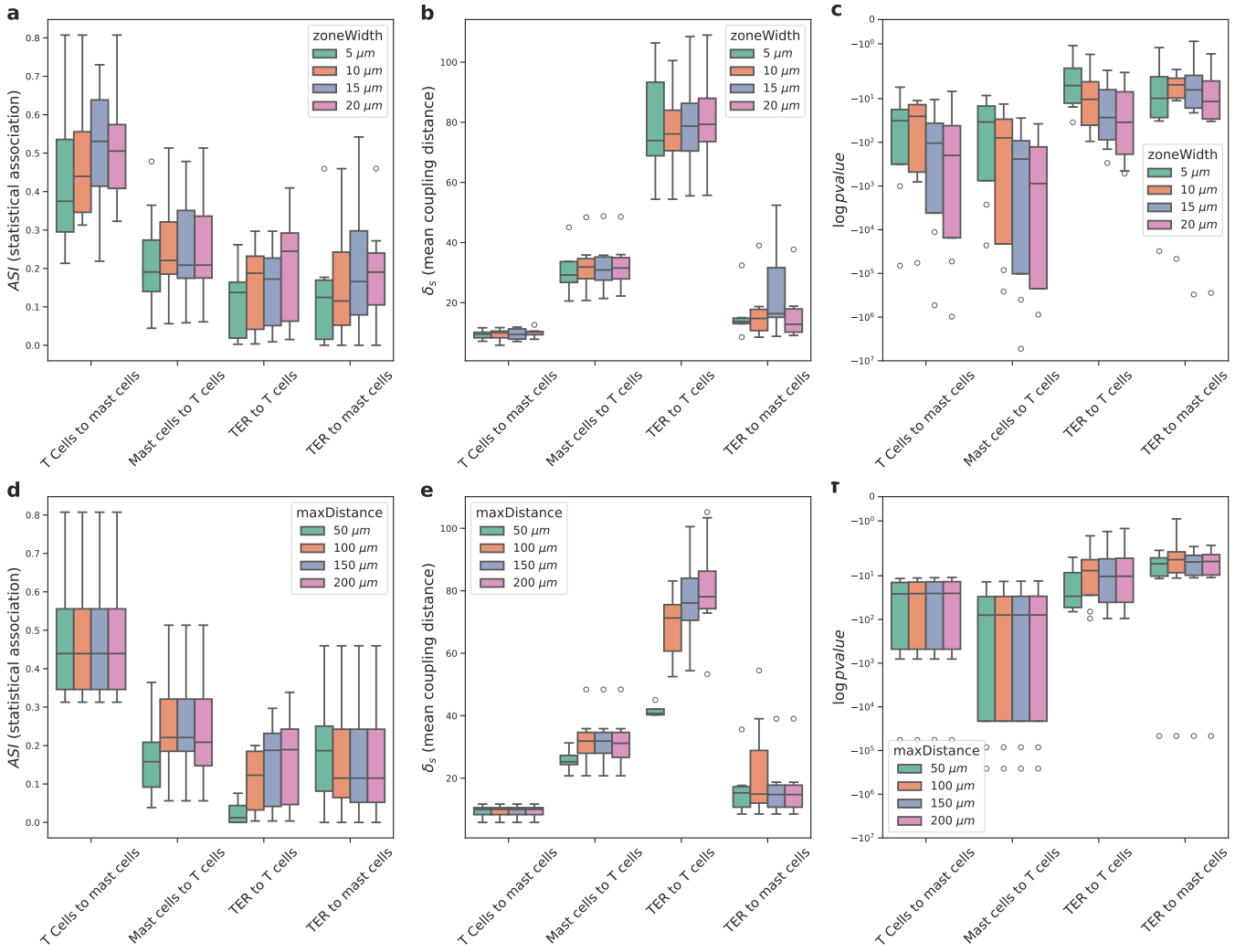
Supplementary Figure 8: **Cell detection with convolutional U-Net**. **(a)** Original histological image with CD3+ T cells (brown) and tryptase+ mast cells (cyan). **(b)** Detections of CD3+ T cells *vs.* ground truth. Red spots indicate false negatives (*i.e.* missed detections), green spots indicate false positives (*i.e.* false detections), and yellow spots indicate true positives. **(c)** Detections of tryptase+ mast cells *vs.* ground truth. The same color coding as in (b) is used. **(d)** F1-score for CD3+ T cells (yellow curve) and tryptase+ mast cells (cyan curve) is plotted for increasing detection threshold T_p.



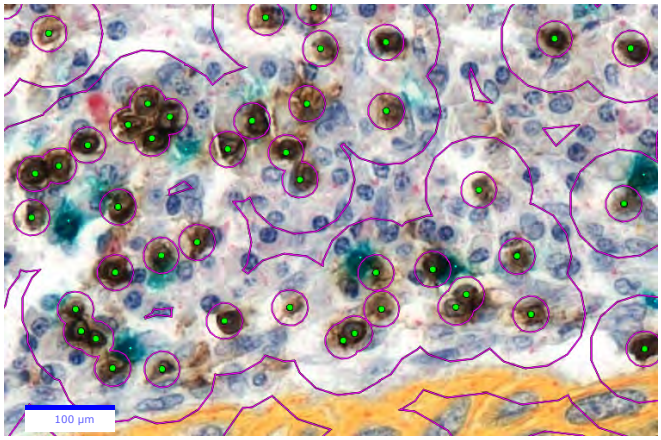
Supplementary Figure 9: **Performance metrics for deep-learning-based detection of CD3+ T cells (yellow) and tryptase+ mast cells (cyan)**. **(a)** Precision with respect to threshold T_p.**(b)** Recall with respect to threshold T_p. **(c)** F1-score with respect to threshold T_p. **(d)** Precision-Recall curve. Source data are provided as a Source Data file.



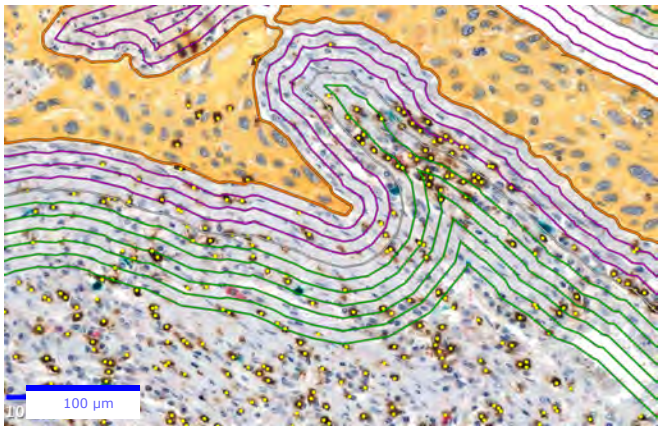
Supplementary Figure 10: **Cell densities for the $n = 8$ patients investigated**. Densities of CD3+ T cells (brown) and tryptase+mast cells (cyan) are displayed in cells/mm² for each patient. The y-axis is on a logarithmic scale to accommodate the range of observed densities. Source data are provided as a Source Data file.



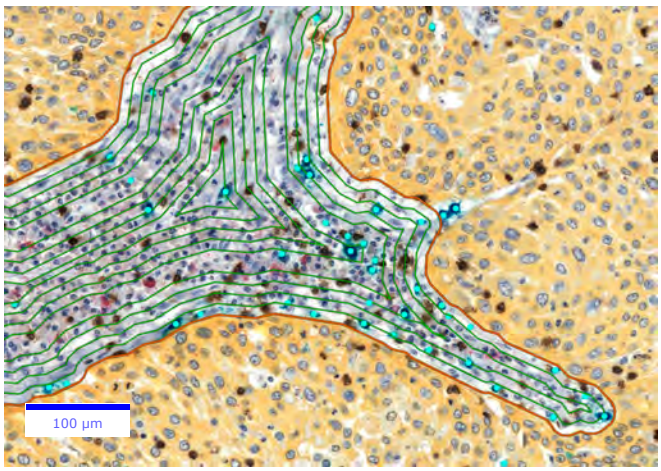
Supplementary Figure 11: **Sensitivity analysis of spatial association between immune cells and tumor epithelium regions in cancer patients**. **(a)** Impact of *zone width* (5 (green), 10 (orange), 15 (blue), and 20 (purple) μm) on estimated association index ASI for the different configurations (T cells to mast cells, mast cells to T cells, TER to T cells and TER to mast cells).**(b)** Impact of *zone width* on estimated distance δ_a. **(c)** Impact of *zone width* on statistical significance $\log(p - value)$. **(d)** Impact of the maximum distance of analysis l_N (50 (green), 100 (orange), 150 (blue), and 200 μm (purple)) on estimated association index ASI. **(e)** of the maximum distance of analysis l_N on estimated distance δ_a. **(f)** Impact of the maximum distance of analysis l_N on statistical significance $\log(p - value)$. Source data are provided as a Source Data file.



(a) Close contact between T cells and mast cells

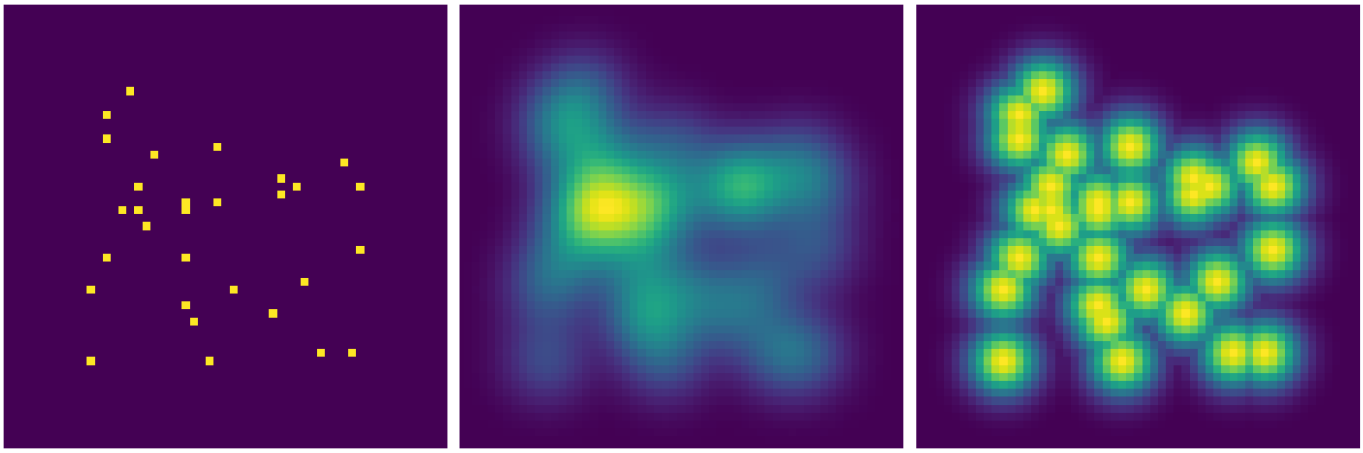


(b) Depletion of T cells along the TER



(a) Accumulation of mast cells along the TER

Supplementary Figure 12: **Illustration of detected spatial patterns (a)** Spatial proximity of mast cells (cyan) to T cells (brown with green dots) showing direct physical contacts (red arrows), **(b)** Depletion of T cells (brown with yellow dots) in the vicinity of of the tumor epithelium region (TER, yellow area). The depletion zone is $\sim 20-40\mu m$, corresponding to $\sim 2-4$ levelsets (magenta lines) and **(c)**) Mast cells (cyan dots) accumulate close to the border of the TER (orange line).



Supplementary Figure 13: **Preprocessing ground-truth coordinates**. Left panel shows the original cell positions with each cell represented as a single pixel set to 1. Middle panel displays the cell density computed using a Gaussian convolution of ground-truth pixels. Size of the Gaussian kernel is determined by the formula: kernel size $= 2 \cdot \sigma + 1$, with $\sigma = \frac{\text{size of a cell}}{\text{pixel size}}$. Right panel shows the density computed by keeping the maximum value from the overlap of individual Gaussian functions for each cell, given by the formula: $\max(G_i)$, where G_i is the Gaussian function for each cell i. The image is normalized to ensure the maximum value is 1. Such preprocessing keeps distinct intensity peaks for each cell.