

Supplementary Material

Liu and Bosse et al., Reproducible, high-dimensional imaging in archival human tissue
by Multiplexed Ion Beam Imaging by Time-of-Flight (MIBI-TOF)

Supplementary Figure 1: MIBI-TOF instrument stability

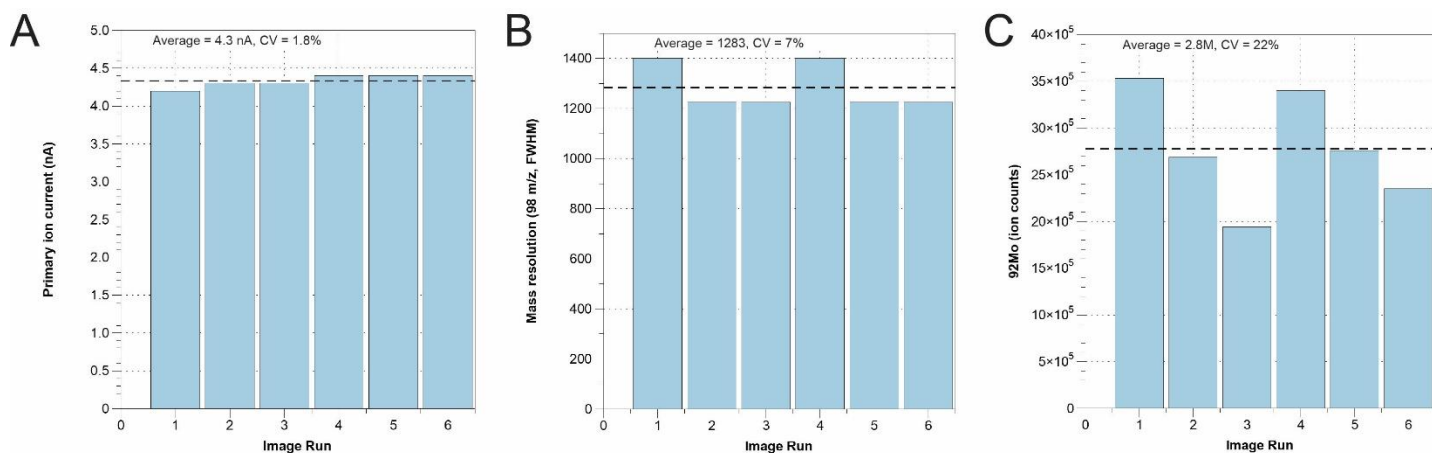


Figure S1. MIBI-TOF instrument stability was assessed prior to data acquisition using a molybdenum foil standard. (A) Primary ion current, (B) mass resolution, and (C) ion detector sensitivity were assessed and did not vary significantly over the course of the study.

Supplementary Figure 2: Effect of normalization coefficient

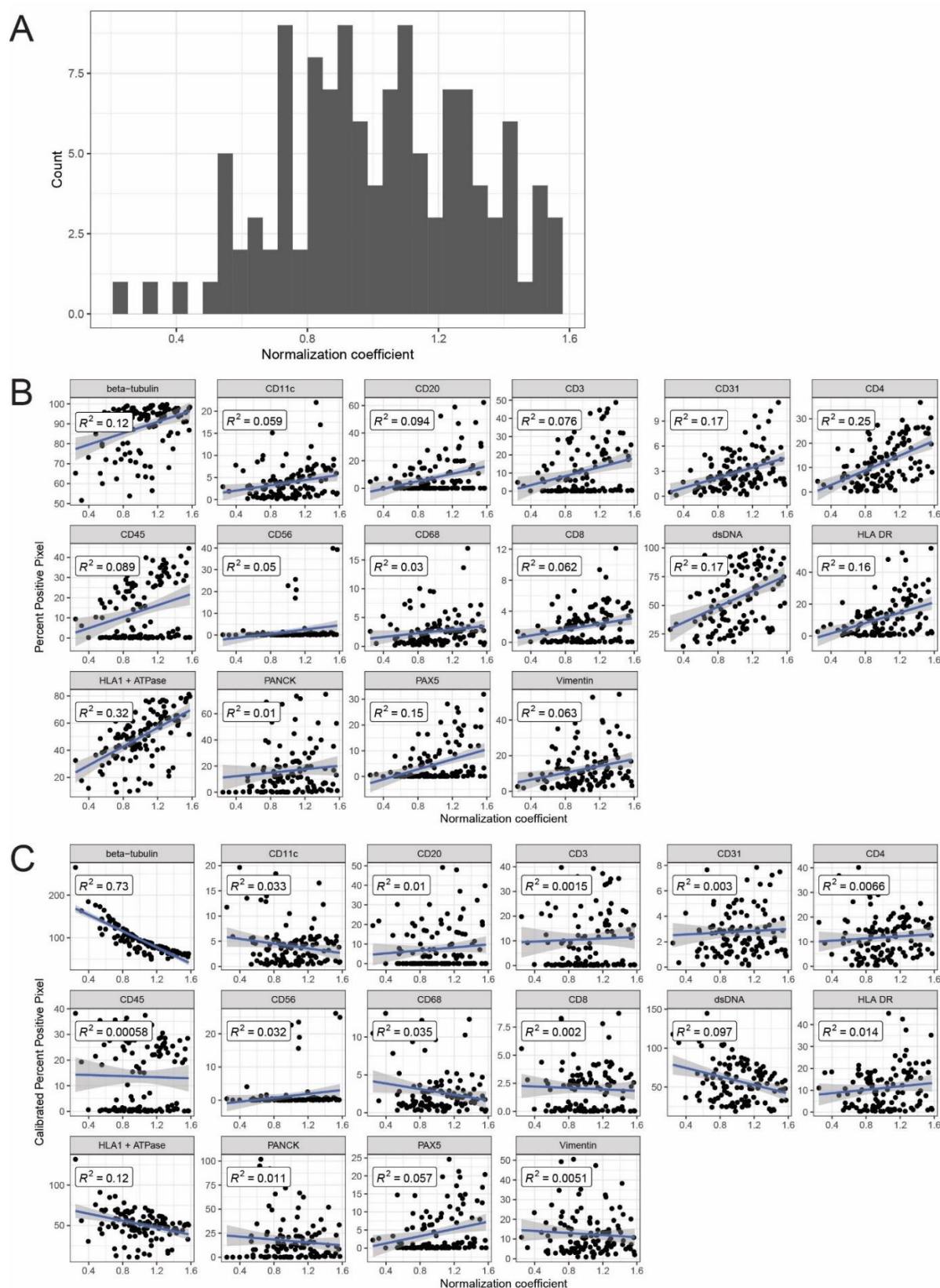


Figure S2. (A) Histogram of normalization coefficients for all fields-of-view (FOVs) in this study. (B) Correlation of the normalization coefficient vs. uncalibrated percent positive pixel for each FOV. (C) Correlation of the normalization coefficient vs. calibrated percent positive pixel for each FOV.

Supplementary Figure 3: Co-registration with brightfield microscopy

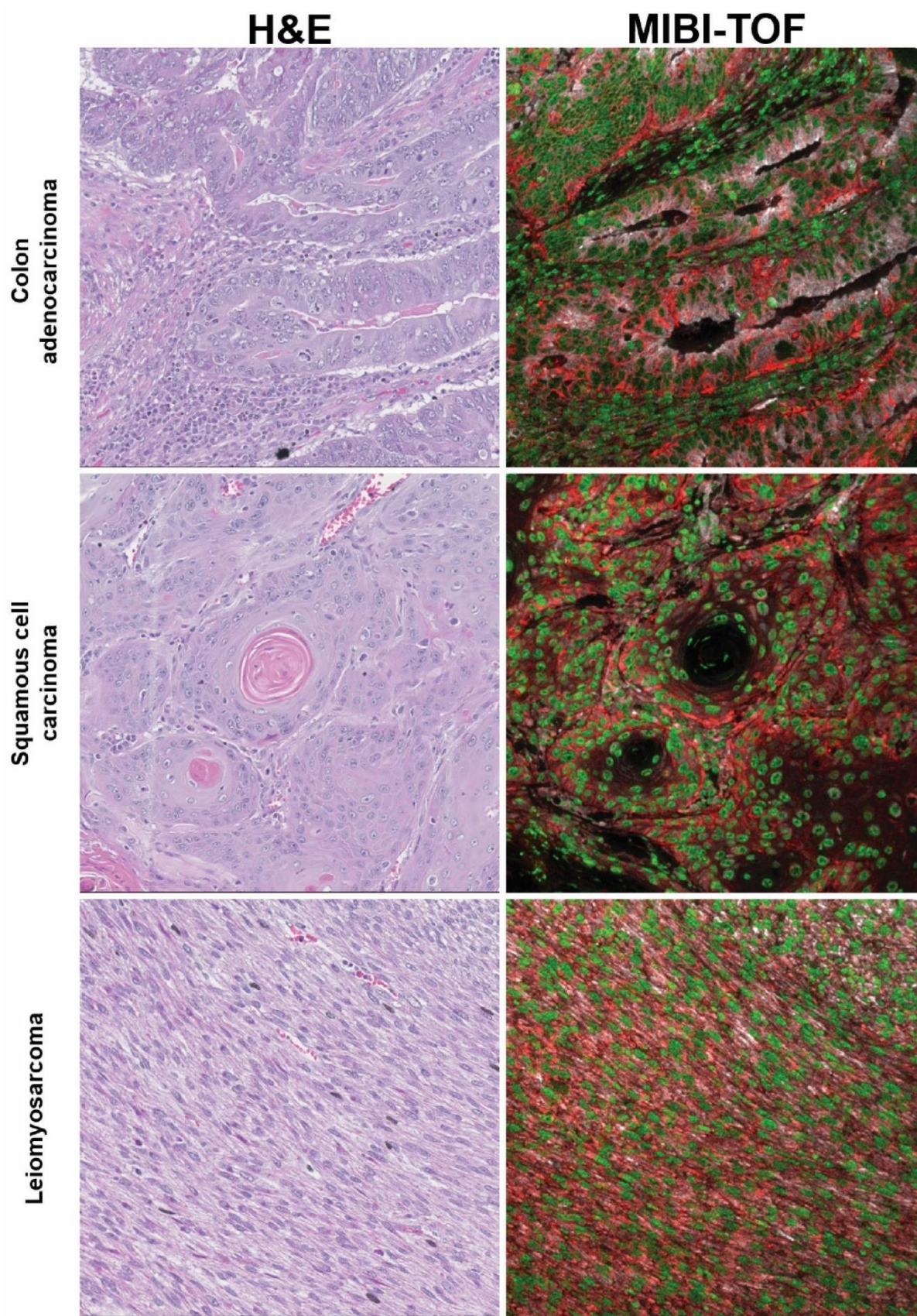


Figure S3. Representative co-registered H&E and MIBI-TOF images of 3 tissue cores. Green: dsDNA, red: HLA1 + ATPase.

Supplementary Figure 4: Concordance of Percent Positive Pixels by MIBI-TOF

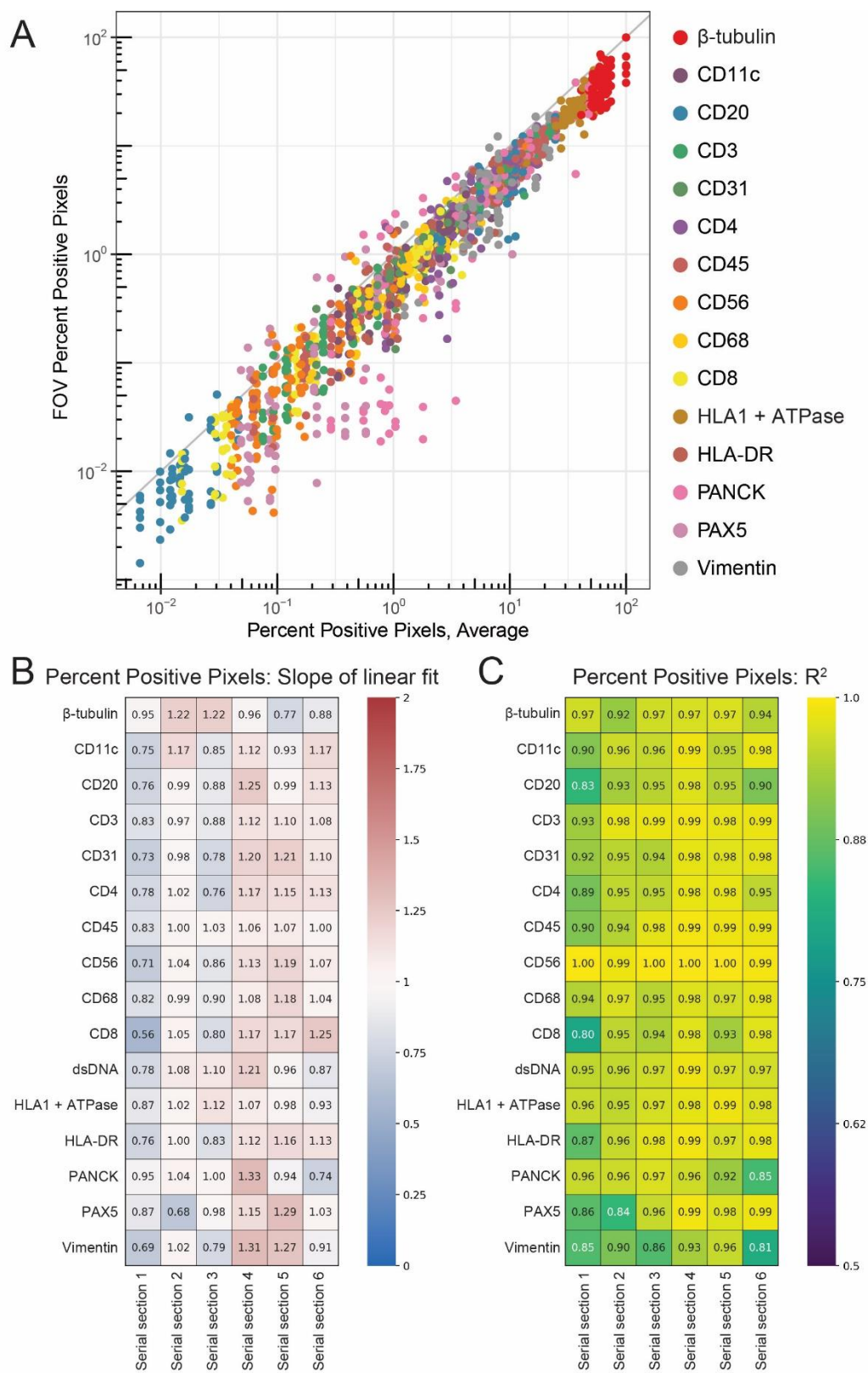


Figure S4. (A) Plot of the average Percent Positive Pixels (PPP) of all FOVs of the same TMA vs. the PPP of each FOV. Each color represents a different marker. The gray line is a reference line with slope 1 and intercept 0. We performed least squares linear regression of the average PPP of the core vs the PPP of all six serial sections and found the slope m (B) and coefficient of determination R^2 (C) for all 16 markers.

Supplementary Figure 5: Variation of serial sections within each tissue core

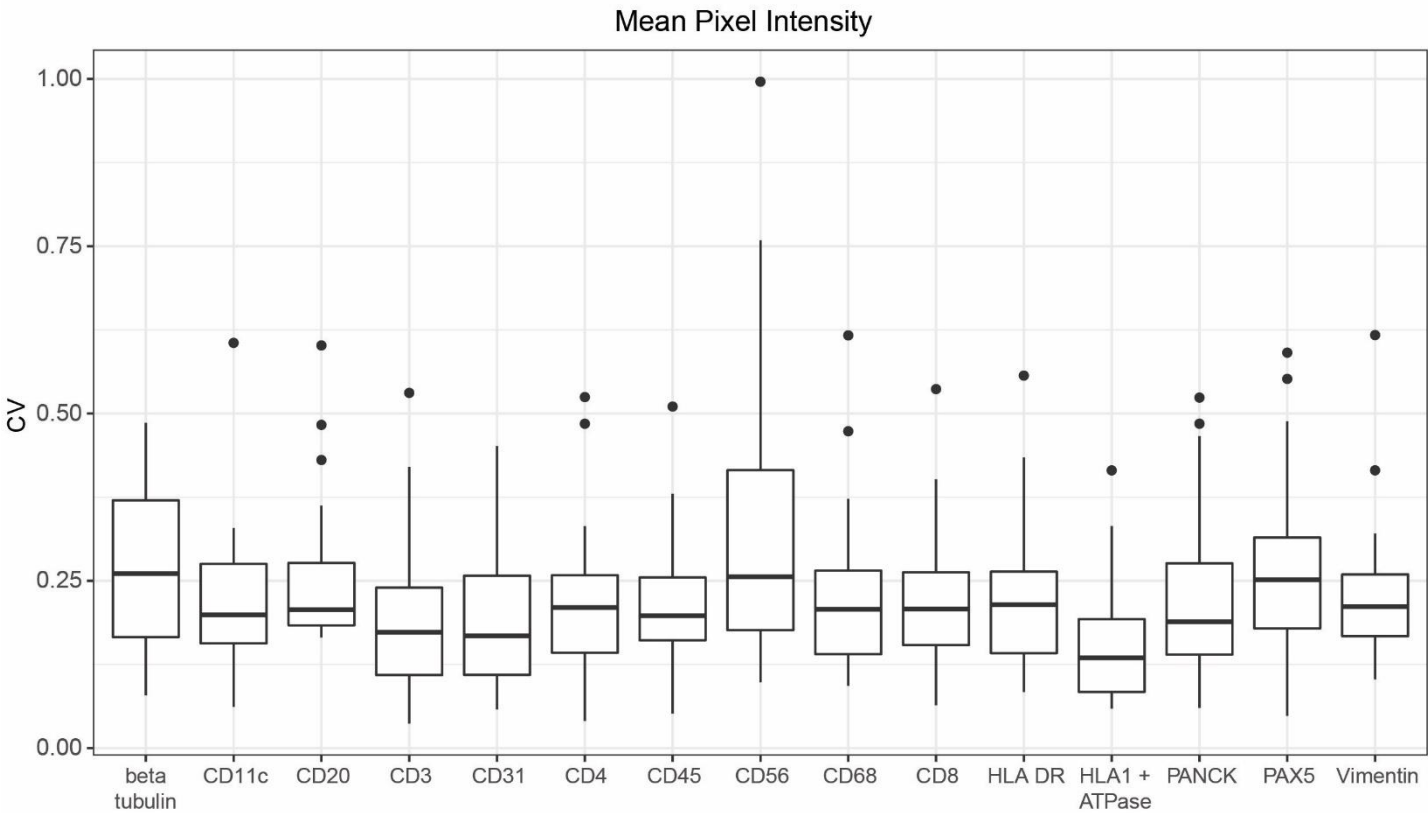


Figure S5. Boxplot of the CV of mean pixel intensity within each tissue core for each marker. Each data point in the boxplots represents one tissue core.

Supplementary Figure 6: Reproducibility of cell segmentation by MIBI-TOF

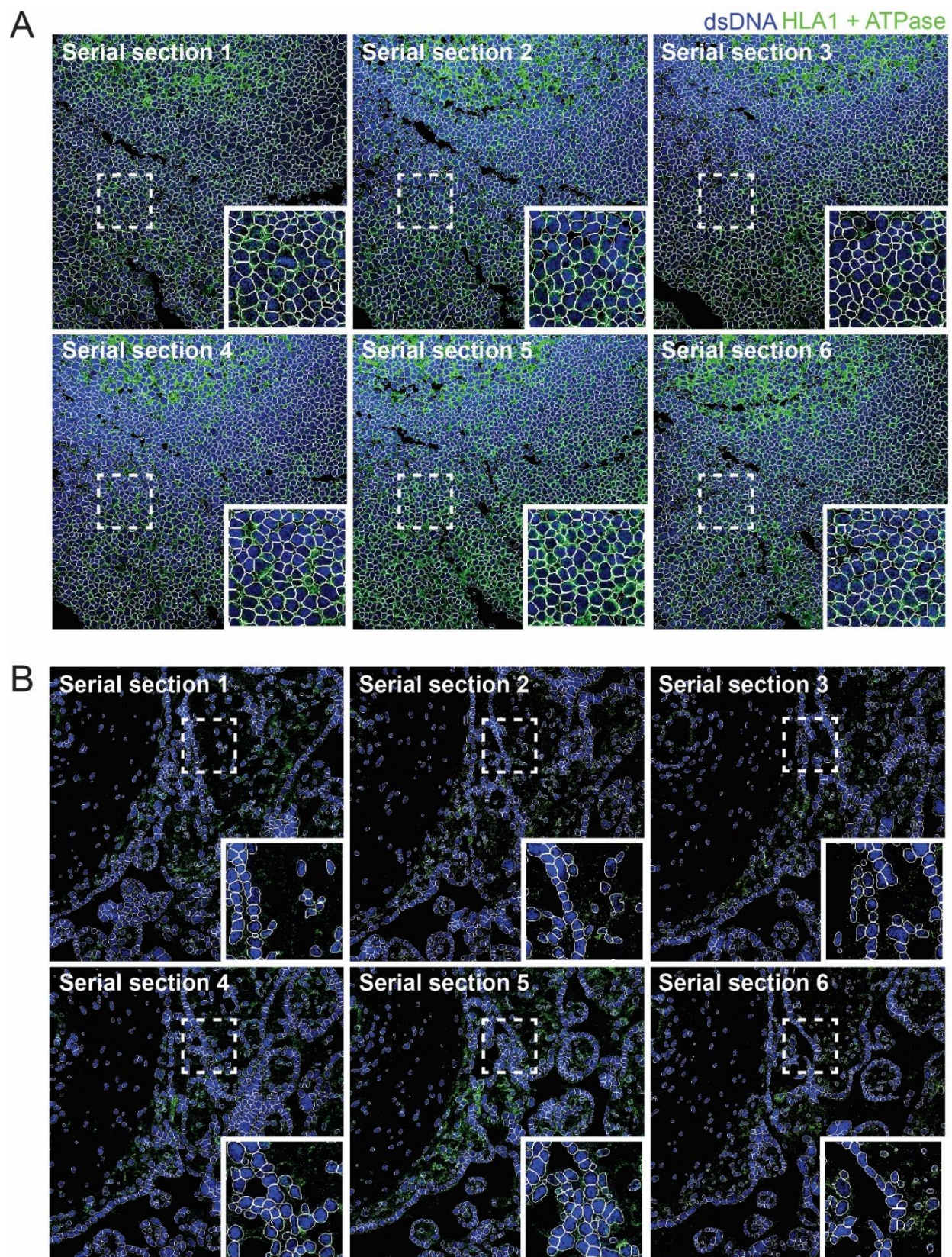


Figure S6. Representative images showing cell segmentation across six serial sections of the same TMA core of tonsil tissue (A) and placenta (B). We used dsDNA (blue) as the nuclear marker and HLA class 1 + NaKATPase (green) as the membrane marker as input to the Mesmer segmentation algorithm.

Supplementary Figure 7: Reproducibility of cell phenotyping by MIBI-TOF

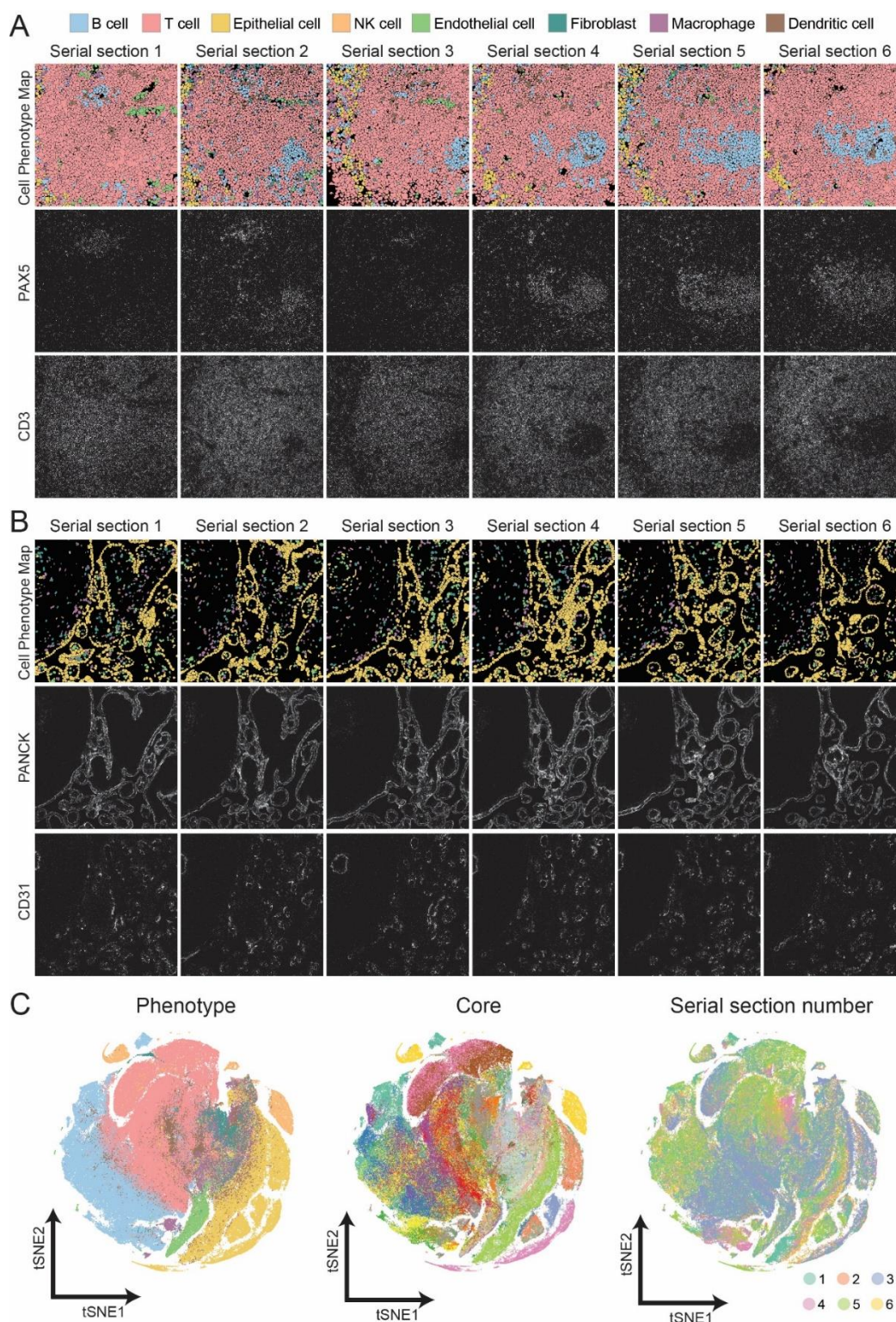


Figure S7. Additional examples showing cell type assignment reproducibility across six serial sections of the same TMA core of tonsil tissue (A) and placenta (B). In both (A) and (B), top row: cell phenotype map colored according to the eight cell types; middle and bottom rows: single-marker MIBI-TOF images. (C) tSNE plots of all cells in all images, where each point represents a single cell colored according to cell phenotype (left), tissue core (middle), or serial section number (right).

Supplementary Figure 8: Cell phenotype correlation as a function of slide distance

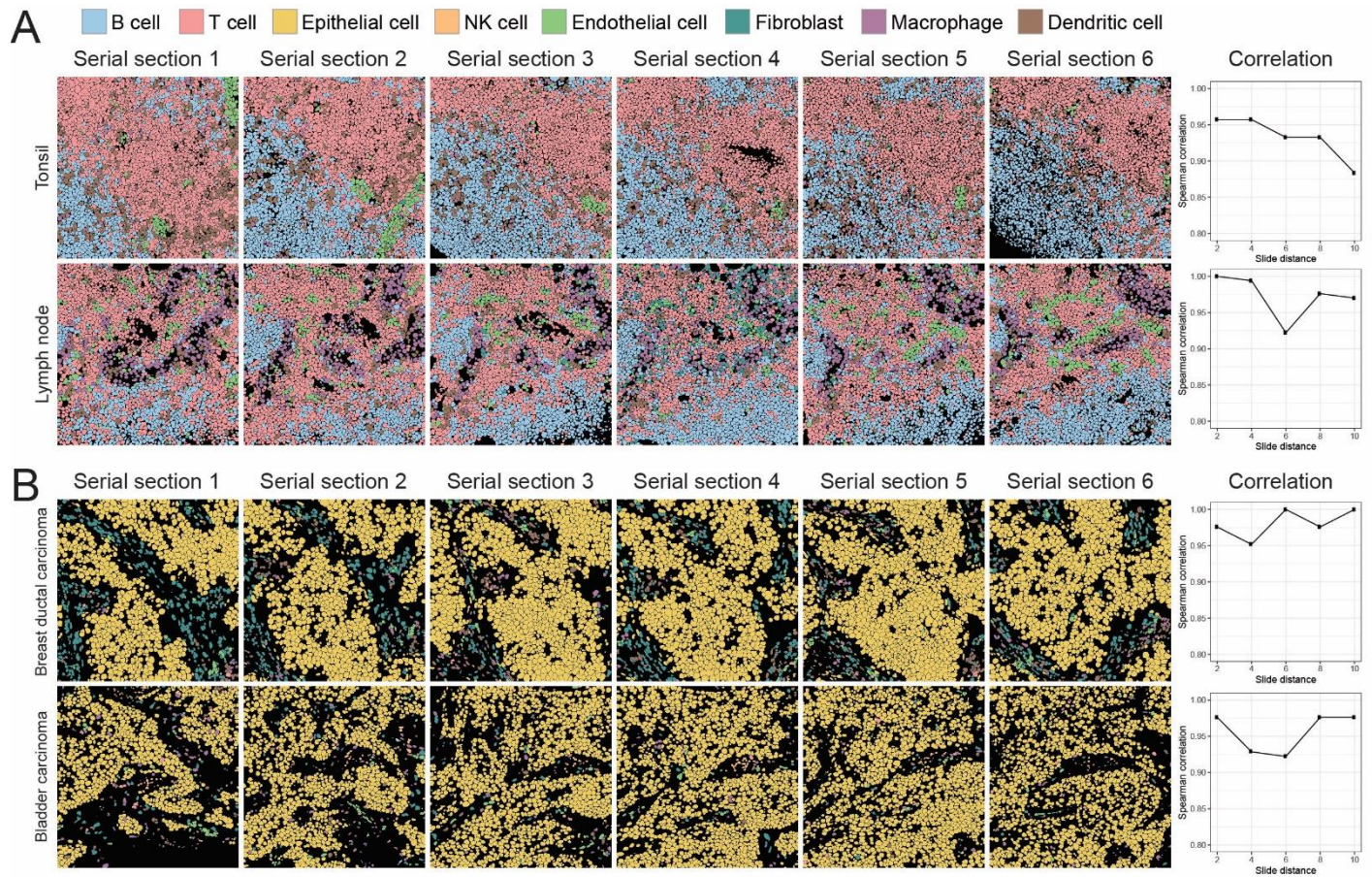


Figure S8. Representative images showing the comparison of serial sections as a function of slide distance. For each tissue core, Spearman correlations of cell phenotype frequency between the first serial section and the subsequent serial sections were determined. (A) Examples of tissues with features such as germinal centers that visibly change as we cut through the tissue core. (B) Examples of tissues that are largely dominated by one or two cell types. Therefore, changes of correlation across sections largely capture small differences.

Supplementary Figure 9: Concordance of MIBI-TOF with single-plex chromogenic IHC

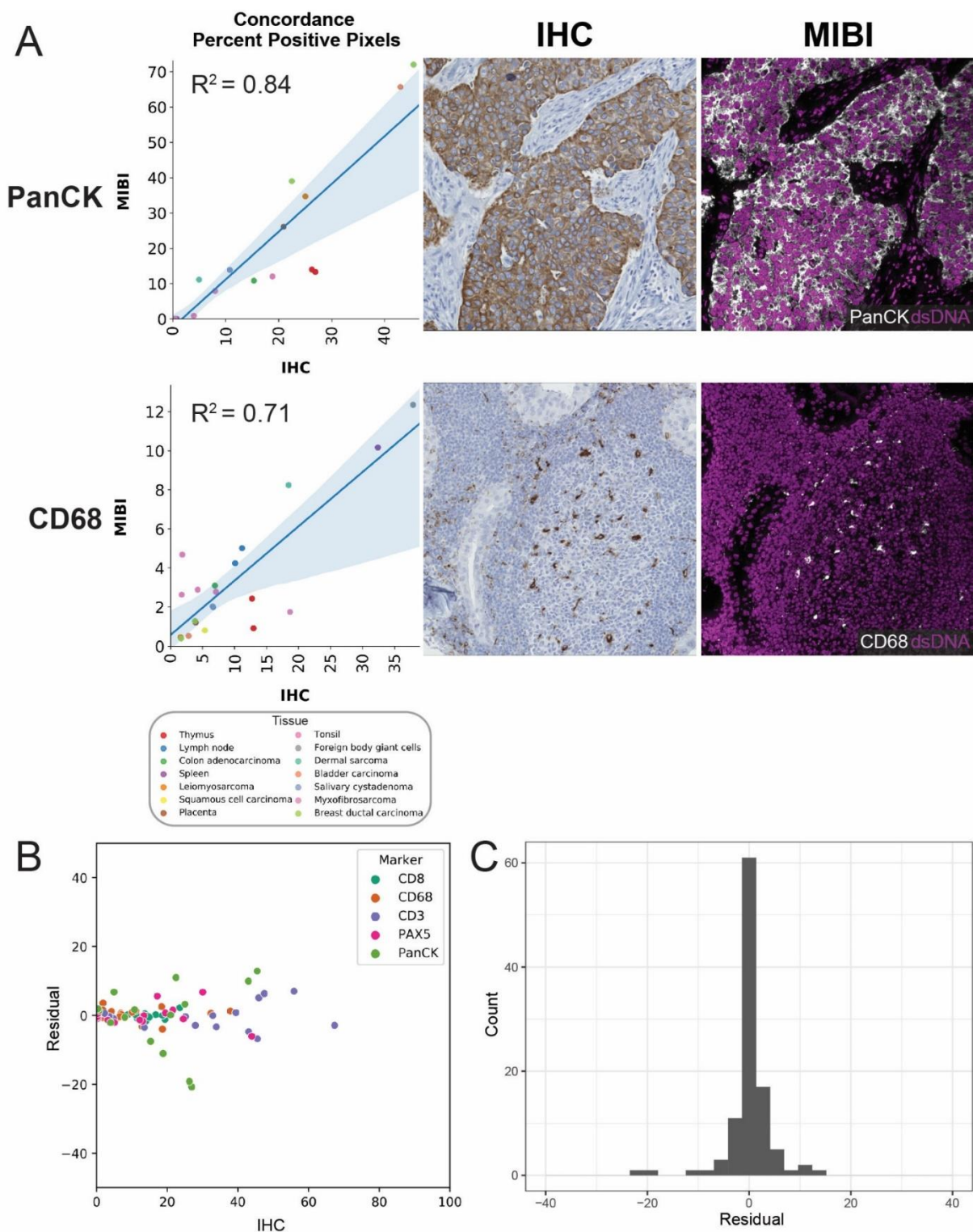


Figure S9. (A) Representative images (PanCK: breast ductal carcinoma, CD68: salivary cystadenoma) of the comparison of MIBI-TOF images co-registered with single-plex chromogenic IHC stains. Each data point represents the PPP by MIBI-TOF and IHC for a single tissue core. Colors of data points indicate tissue type. Shaded area represents 95% confidence interval. (B) Plot of the residuals of the linear regressions shown in Figure 5 and Supplementary Figure 9A, as a function of the IHC value. (C) Histogram of all residuals.

Supplementary Table 1: MIBI-TOF background correction parameters

Channel	Mass Window		Slide Background			Isobaric Correction					
	Start [m _a]	Stop [m _a]	¹³⁸ Ba [counts]	¹⁸¹ Ta [counts]	¹⁹⁷ Au [counts]	dsDNA (⁸⁹ Y)	CD4 (¹⁴³ Nd)	CD56 (¹⁵¹ Eu)	CD68 (¹⁵⁶ Gd)	CD3 (¹⁵⁹ Tb)	HLA DR (¹⁷² Yb)
dsDNA (⁸⁹ Y)	88.70	89.00		20	50						
beta-tubulin (¹¹³ In)	112.70	113.00		20	50	3.30%					
CD4 (¹⁴³ Nd)	142.70	143.00		20	50						
CD11c (¹⁴⁴ Nd)	143.70	144.00	15	20	50		2.00%				
CD56 (¹⁵¹ Eu)	150.70	151.00	15	20	50						
CD31 (¹⁵² Sm)	151.70	152.00	15	20	50						
CD68 (¹⁵⁶ Gd)	155.70	156.00	15	20	50						
CD8 (¹⁵⁸ Gd)	157.70	158.00		20	50				0.96%		
CD3 (¹⁵⁹ Tb)	158.70	159.00		20	50						
Vimentin (¹⁶³ Dy)	162.70	163.00	15	20	50						
PAX5 (¹⁶⁵ Ho)	164.70	165.00	15	20	50						
CD20 (¹⁶⁷ Er)	166.70	167.00		20	50		1.50%				
HLA DR (¹⁷² Yb)	171.70	172.00		20	50						
PANCK (¹⁷³ Yb)	172.70	173.00		20	50				5.00%	1.00%	5.20%
CD45 (¹⁷⁵ Lu)	174.70	175.00		20	50			1.00%			
HLA class 1 A, B, and C Na-K-ATPase alpha (¹⁷⁶ Yb)	175.70	176.00		20	50						

Supplementary Table 2: MIBI-TOF noise filtering parameters

Channel	Aggregates diameter [um]	Two intensity levels	Initial noise estimation	Reachability cut-off		Reachability mask
				Slope	Intercept	
dsDNA (⁸⁹ Y)			Reachability	-0.16	1.16	
beta-tubulin (¹¹³ In)		TRUE	Conjunction	0.00	1.25	
CD4 (¹⁴³ Nd)	2.5		Max	-0.44	1.57	TRUE
CD11c (¹⁴⁴ Nd)	2.5		Max	-0.29	1.43	TRUE
CD56 (¹⁵¹ Eu)	2.5		Reachability	-0.25	1.25	TRUE
CD31 (¹⁵² Sm)			Reachability	-0.52	0.90	TRUE
CD68 (¹⁵⁶ Gd)			Max	-0.37	1.21	TRUE
CD8 (¹⁵⁸ Gd)			Reachability	-0.34	1.41	TRUE
CD3 (¹⁵⁹ Tb)	2.5		Reachability	-0.32	1.50	TRUE
Vimentin (¹⁶³ Dy)			Reachability	-0.24	1.37	
PAX5 (¹⁶⁵ Ho)	2.5		Max	-0.27	1.41	TRUE
CD20 (¹⁶⁷ Er)	2.5		Connectivity	-0.33	1.39	TRUE
HLA DR (¹⁷² Yb)			Reachability	-0.41	1.27	TRUE
PANCK (¹⁷³ Yb)		TRUE	Reachability	-0.30	1.60	
CD45 (¹⁷⁵ Lu)	2.5		Reachability	-0.22	1.53	TRUE
HLA class 1 A, B, and C Na-K-ATPase alpha (¹⁷⁶ Yb)		TRUE	Conjunction	-0.26	1.78	

Supplementary Table 3: IHC thresholds

Channel	DAB Threshold	Hematoxylin Threshold
CD68	0.130	0.210
CD8	0.126	0.210
CD3	0.177	0.168
PAX5	0.169	0.207
PANCK	0.110	0.262

Image Processing Summary

Introduction

Multiplexed ion beam imaging (MIBI) analyzes samples by staining tissues with metal-tagged antibodies and then sputtering with a focused ion beam. In the picoseconds following a primary ion impact at the surface of the tissue, a plume of neutral and charged atoms, molecules, and molecular fragments is formed. The positively charged ions from the plume are then extracted into the time-of-flight mass analyzer.

The MIBI experiment relies on detecting the metal tags liberated from conjugated antibodies during the sputtering process. However, there is a certain probability that either the metal atom may not cleanly separate from the conjugation molecule, that it may recombine with another fragment from the plume, or that other ions of similar masses are liberated. While the MIBIScope instrument is designed and tuned to preferentially allow monatomic species rather than larger polyatomic species into the mass analyzer, thus specifically selecting for the metal tags, it cannot completely prevent diatomic or triatomic species from passing through. As such, when interpreting the mass spectrum, it cannot be assumed that all measurements are of monatomic metal tags.

Mass Windows

In addition to the monatomic isotopic labels, polyatomic ions containing hydrogen are liberated from the organic tissue. These can be recognized in the mass spectrum as mass peaks above the integer atomic masses. On the contrary, the mass deficient monoatomic peaks are located below the integer atomic masses. This is due to the proton mass being considerably higher in hydrogen than when bound in heavier atoms.

Provided that mass resolution of the MIBI analyzer allows the separation of the polyatomic organic peak, the monoatomic peak can be selected by setting a narrow mass window around the metal atom tag's mass. In our case, we integrated all peaks within a $(-0.3, 0)$ mass window for each integer mass.

Slide Background

MIBI experiments are performed on slides coated with Ta and Au metal layers to ensure the conductivity needed for ion sputtering. In areas where the tissue has been sputtered through or is not present, there is elevated intensity of Ta and Au peaks, but also a substantial increase in background across the entire mass range. The time-dependent presence of Au and Ta within each pixel can therefore be used as an indicator of when tissue is not present. This is represented in Figure 1.

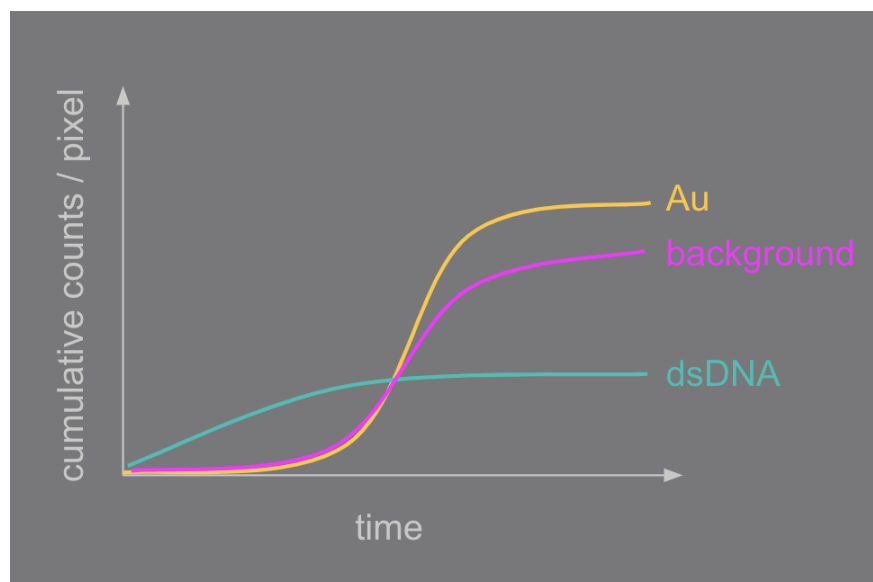


Figure 1. Schematic representation of the time dependence of the cumulative intensity of tag metal (dsDNA) and slide background when sputtering through tissue and then into the gold slide coating. It can be seen that the tag metal intensity grows linearly and then saturates, after which the background and gold intensities exhibit a steep growth.

The correlated temporal component of the slide metals and slide background intensities can be used to exclude background. This can be done by stopping ion count accumulation when the cumulative counts of the slide metals reach certain predefined thresholds. Stopping accumulation when a certain threshold is reached ensures that there is a minimal slide background bleed into other channels.

Slide background can also be thresholded on other ions. In this assay slides exhibited elevated Ba contamination that was significantly affecting some tag metal counts, therefore Ba was included in the species used to indicate and exclude slide background.

Isobaric Correction

As mentioned above, during the sputtering process, the heavy metal atoms used as tags may be chemically bound to other atoms or fragments of molecules, in which case the mass of the detected species is no longer that of the atom itself but rather the molecule it has now formed. Some of the metal atoms used as labels in MIBI have been seen to react and bond with small hydrogen, oxygen, nitrogen, and carbon species. These new species cannot be completely resolved by the mass analyzer from another metal tag that has the same integer mass, resulting in an isobaric interference. Fortunately, this chemical process is dependent on the metal atom and interfering species, not the particular isotope of the metal element or other sample-specific variables, and thus is consistent across images. This enables it to be reliably characterized by comparing the signal from a metal tag channel to that of an unlabeled channel that is susceptible to an isobaric interference.

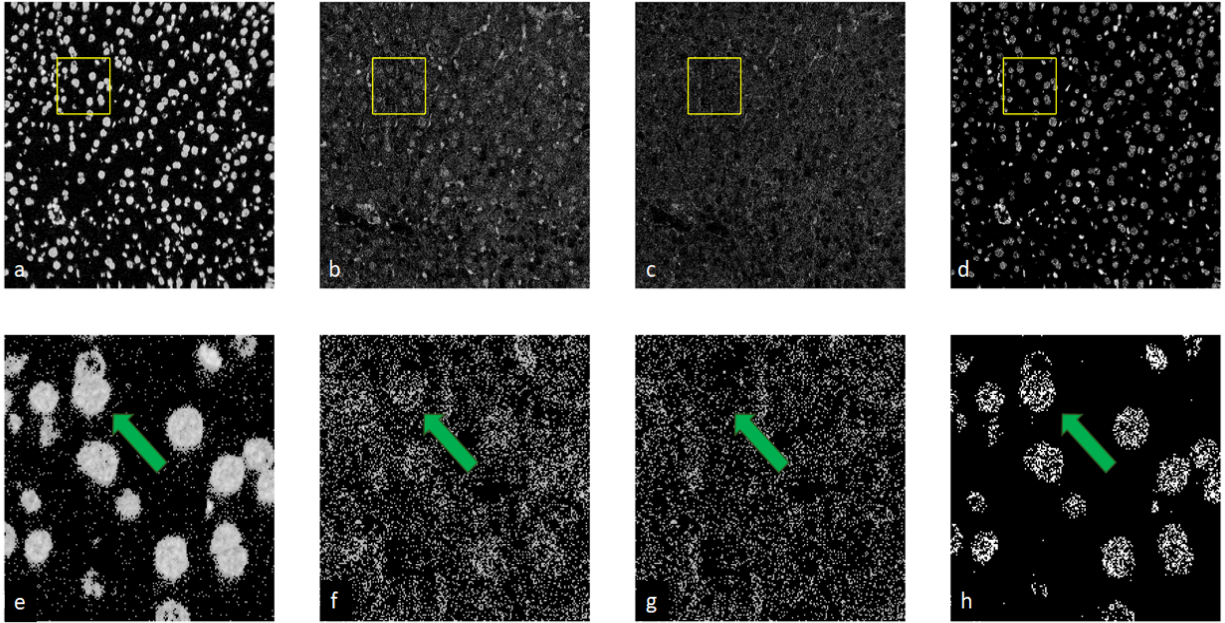


Figure 2. Example of an observed isobaric interference and subsequent correction using the method described here. In this case, dsDNA (nuclear marker) on ^{89}Y interferes into beta-tubulin (cytoplasmic marker) on ^{113}Y by an expected fraction of 0.033. a) Original dsDNA image which is the donor of the interference, b) original beta-tubulin image which is the recipient of the interference, c) beta-tubulin image after performing the detailed isobaric correction method, d) difference image between (b) and (c) showing signal that was removed during the correction process, e-h) zoomed in ROI corresponding to the yellow box in (a) - (d), including a green arrow which highlights an example of a nuclear feature from the dsDNA channel (e) being observed in the cytoplasmic channel beta-tubulin (f) before being corrected for in (g) and the difference image (h).

To correct for the above described isobaric interferences, we utilize the spatial distribution of the donor signal as a source of information as to where the interfering signal in a recipient channel is expected to manifest. For correcting a suspected interference, the donor channel (source of the interfering metal tag) and recipient channel (the channel which is susceptible to the interference) are specified, as well as the fraction of the donor signal that is expected to interfere with the recipient channel. The expected fraction was determined empirically by IONpath and has been shown to be consistent across samples and tissue types

Given the mass images of the donor channel and recipient channel, as well as the expected fraction of interference, the isobaric correction algorithm begins by calculating the expected number of counts in the recipient channel caused by the isobaric interference. Then, the process selects pixels ordered by descending intensity in the donor image, and at each iteration removes at most one count from the recipient image from this same pixel. A count is only removed if there is sufficient signal in the recipient image, and if the expected contribution of the isobaric interferences has not yet been reached. This process is repeated until either the maximum number of counts to remove (total donor counts multiplied by expected fraction of interference) has been reached or the copy of the donor image has run out of counts in affected

pixels. An example of performing this procedure is shown in Figure 2 where a 0.033 fraction of dsDNA on ^{89}Y interferes into beta-tubulin on ^{113}Y .

Noise Filtering

Apart from a background that can exhibit some kind of spatial structure, there also exists some inherent noise in multiplexed ion beam imaging (MIBI) data. The instrumental noise is Poisson in nature^[1] and is distributed across not only the x- and y- spatial dimensions, but also across the mass spectral dimension. Non-specific antibody binding may contribute additional background. As a result, the noise within each image is distributed differently across each mass channel.

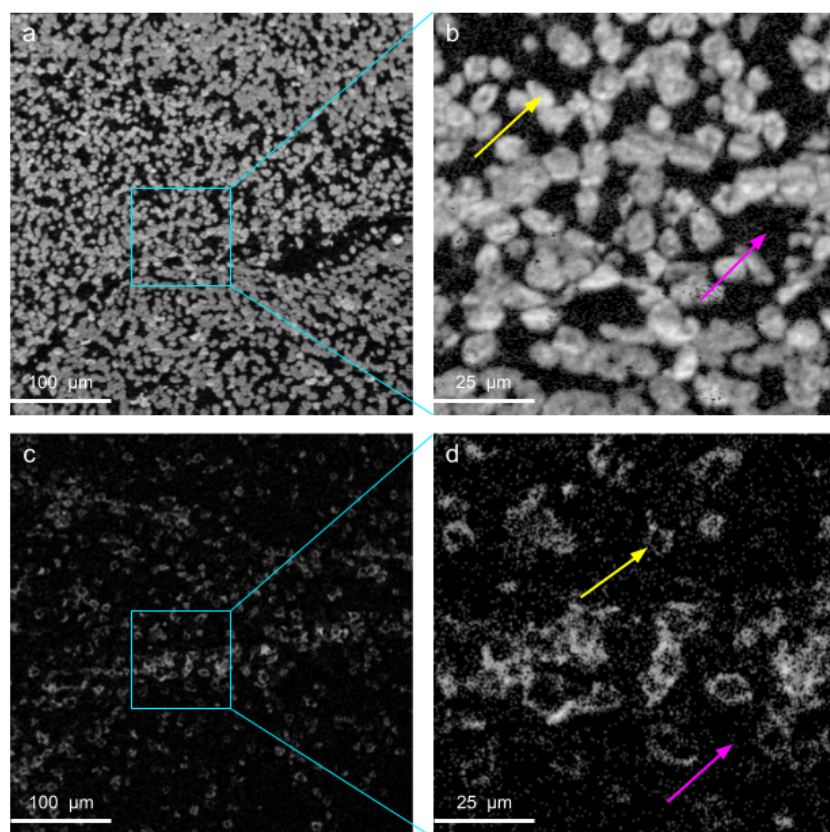


Figure 3. Two channels of a MIBI image containing both biological signal and spatially random noise. (a) dsDNA channel image. (b) Zoomed area corresponding to cyan box in (a) to highlight signal vs. noise regions. The yellow arrow indicates a nuclear feature corresponding to the expected distribution of dsDNA in the sample, and the magenta arrow points to a region of presumed noise. (c) CD3 channel image. (d) Zoomed area corresponding to cyan box in (c) to highlight signal vs. noise regions. The yellow arrow indicates a membrane feature corresponding to the expected distribution of CD3 in the sample, and the magenta arrow points to a region of presumed noise. Note that the CD3 features are more subtle than those of dsDNA, but in both cases the biological signal has spatial structure whereas the noise is randomly distributed.

The amount of noise present in an image is relatively low compared to that of the expected biological signal and rarely exceeds one count per pixel within a single channel. For images of highly expressed markers, the signal to noise ratio is often large enough that the intensity level of noise is well separated from that of the true signal. However, for low-abundant markers, pixels

with only one or two noise counts have intensities on the same order of magnitude as pixels with true biological information. The overlap of the per-pixel signal and noise intensity distributions at low counts per pixel makes it challenging to filter noise based solely on thresholding without losing true signal.

Instead of thresholding the original ion image, it generally yields better results to employ 2D density estimates. We use a combination of several such metrics, including reachability density, connectivity, and Voronoi tessellation to estimate and remove noise from each channel. This approach, described in more detail below, leverages both spatial and intensity information to assign a floating-point density value to each pixel, which in turn allows a more refined noise threshold selection. However, the density estimation parameters must be chosen with caution to avoid either diminishing signal features or enhancing background artifacts due to small variations in density.

Reachability Density

The reachability distance from one point to another is a variation on k-nearest-neighbor distance that better accounts for variations in local density. The reachability distance from A to B is the maximum of the distance between A and B, and the distance from B to its kth-nearest-neighbor. The reachability density at A is the inverse of the average reachability distance of the object A from its neighbors. In practice, this means that if B is one of A's nearest neighbors, A is assigned a lower density if it is not also one of B's nearest neighbors. This helps amplify differences in density between regions of positive and negative expression

To generate the reachability density of an ion image, each ion count is considered to be a point at its pixel coordinates. Multiple counts on the same pixel have zero distance to each other. The reachability density can thus be calculated for each pixel that has nonzero counts. For image channels where there is strong positive expression, the reachability density distribution is bimodal, with distinct peaks corresponding to signal and to noise. This is useful for estimating the threshold between signal and noise, but when applied alone can create artifacts by removing sparse counts and thus highlighting small areas of slightly denser noise left behind.

Connectivity

The connectivity is a simple measure of how many of a nonzero pixel's adjacent pixels are also nonzero. Since cellular features tend to span multiple pixels, this can be a helpful metric when assessing likelihood that a count is signal or noise, but its discrete nature makes insufficient on its own.

Voronoi Tessellation

A Voronoi tessellation provides a parameter-free method of estimating the density. By definition, Voronoi diagrams partition a space containing seed points into Voronoi cells, where each cell consists of all points in the plane closer to that seed than to any other seed. When applied to a MIBI image, the seed of each cell corresponds to a pixel that contains one or more ion counts. The sparsely distributed noise gets assigned to bigger cells, and any dense biological signal is assigned to smaller cells.

In order to produce a density estimate, the ion counts from the seed pixels are distributed evenly across a Voronoi cell area and scaled intensity is assigned to pixels within a Voronoi cell. This step essentially increases the dynamic range of the image by mapping the sparse and discrete pixel counts to a more continuous distribution. In the resulting Voronoi tessellation, larger Voronoi cells contain pixels with low fractional values which are more likely to represent noise, while higher pixel values indicative of true signal correspond to smaller cells.

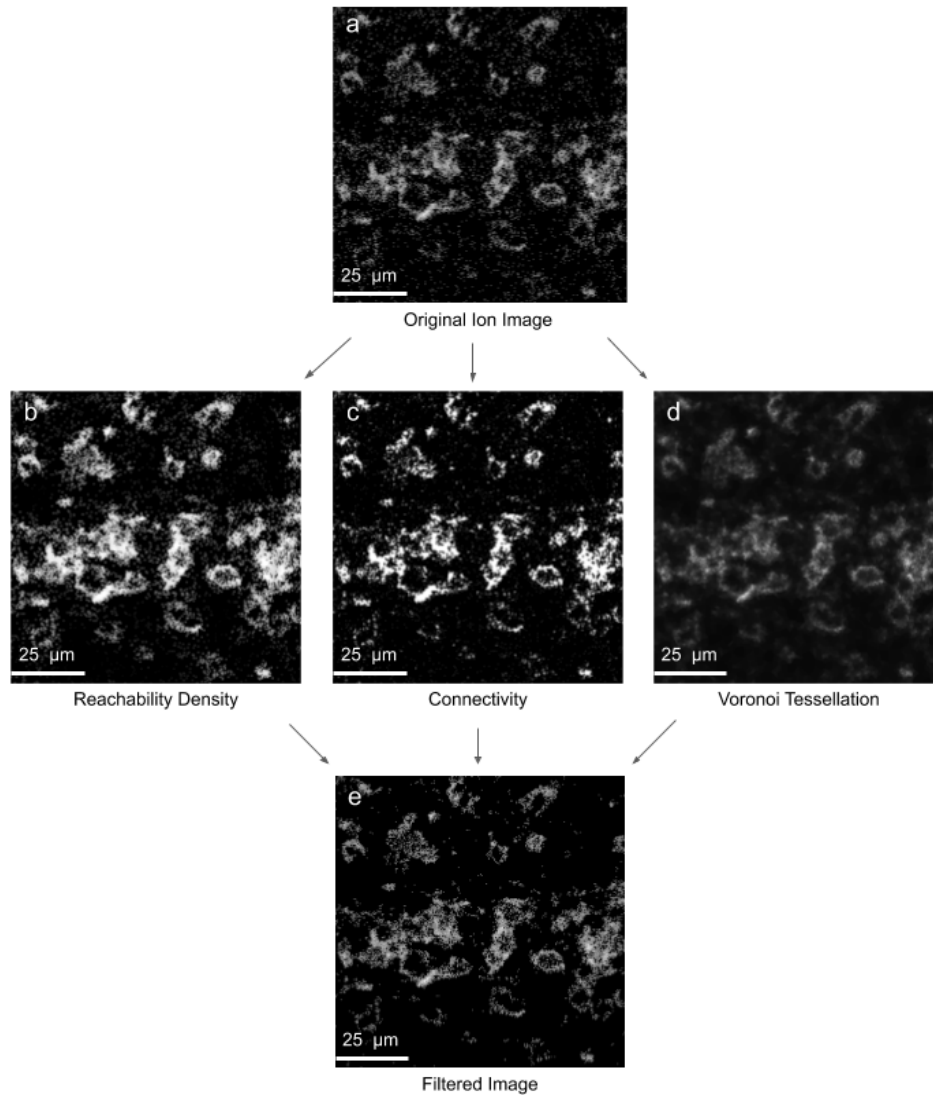


Figure 4. Filtering using combined density estimations. (a) The original ion image is the input to the reachability density, connectivity calculation, and Voronoi tessellation in (b), (c) and (d), respectively. The connectivity and reachability are used to determine a threshold on the Voronoi tessellation below which counts are considered to be noise. (e) The filtered image is output after masking with the thresholded Voronoi tessellation.

The Voronoi tessellation has several beneficial properties when compared to the original ion count image:

- Every pixel in the Voronoi tessellation has non-zero intensity.

- Voronoi cells are equivalent to pixels in the dense regions, where every pixel has at least one count.
- The summed intensity of the Voronoi tessellated image is very close to the total original ion counts (the deviation in the total count is a result of the Voronoi cell area not strictly corresponding to the number of pixels in a Voronoi cell).
- Creating the Voronoi tessellation is parameter-free (as opposed to choosing convolution kernels or number of nearest neighbors).

However, by itself, a Voronoi tessellation does not determine a density threshold for signal-noise separation. Thus, we use it in combination with the other methods that better amplify the difference between signal and noise. Provided an initial estimate of noise levels, the Voronoi tessellation can then be used as an effective regularization step.

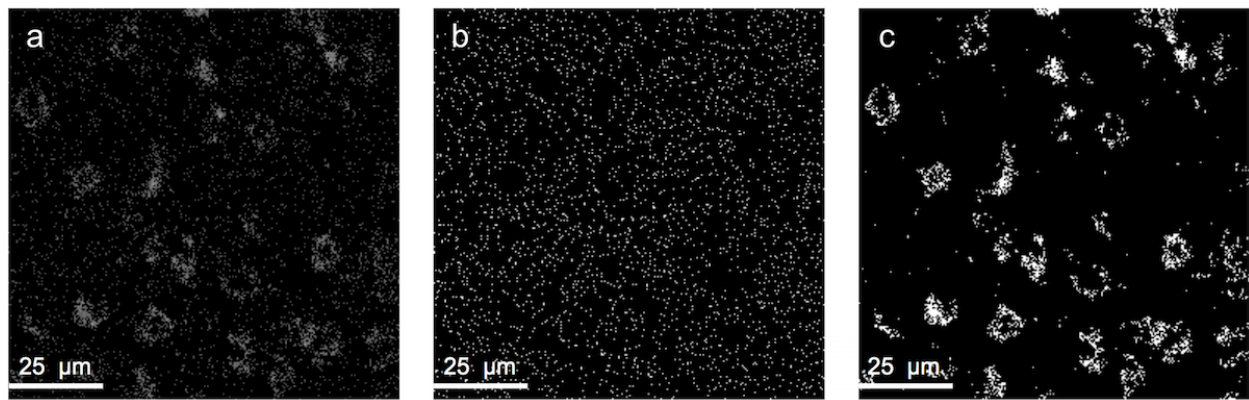


Figure 5. Filtering check comparing original image, estimated noise, and resulting signal. (a) An original CD3 image channel. (b) A binary image of the counts determined to be noise and thus removed from the image. (c) A binary image of the counts remaining in the image channel after filtering.

Determining and QCing Filtering Parameters

The following guidelines are applied when determining the filtering parameters for a dataset:

- Each channel should be filtered consistently across the dataset. In other words, conditions are chosen for a channel, not for each image. This preserves any comparisons of relative expression between samples.
- Assumptions about where signal is expected to be present should be applied only to controls or images where there is sufficient prior knowledge, and even then mostly as a quality assurance measure rather than to set parameters a priori.
- It is preferable to retain some spatially random noise rather than potentially remove biological signal.

To accomplish this, a subset of images are selected for each channel that demonstrates the range of expression observed in that dataset. Parameters are determined from the density distributions that best separate signal from noise in these images. A fit to these parameters is determined from the mean image channels' ion counts per unit area. This fit is then used to set parameters and filter images for that channel across the entire data set.

The resulting parameters are QCed by three trained users on controls plus additional images for validation, and adjusted until consensus is reached. A comparison leveraged in this

process is to examine not only the original and filtered images, but also their difference, as the spatial randomness of the purported noise is an important indicator of the method's accuracy.

IHC images

Hematoxylin and DAB channels were extracted from the original IHC image via spectral unmixing.^[3] Marker positivity was established by thresholding of the deconvolved images. The threshold for hematoxylin and each DAB channel was set collectively for all images in the staining batch. An initial threshold approximation was set with automatic image histogram thresholding (Li thresholding).^[2] The thresholds were then QCed by three trained users on controls and additional images for validation, and adjusted until consensus was reached.

References

1. Keenan, M.R., and Kotula, P.G. (2004). Accounting for Poisson noise in the multivariate analysis of ToF-SIMS spectrum images. *Surface and Interface Analysis* 36, 203–212.
2. Li, C.H., and Lee, C.K. (1993). Minimum cross entropy thresholding. *Pattern Recognit.* 26, 617–625.
3. Ruifrok, A.C., and Johnston, D.A. (2001). Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol.* 23, 291–299.