Reproducible, high-dimensional imaging in archival human tissue by Multiplexed Ion Beam Imaging by Time-of-Flight (MIBI-TOF) — Source link 🔗

Candace C. Liu, Marc Bosse, Alex Kong, Adam Kagel ...+4 more authors

Institutions: Stanford University, National Institutes of Health

Published on: 16 Oct 2021 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Mass spectrometry imaging

Related papers:

- An unsupervised MVA method to compare specific regions in human breast tumor tissue samples using ToF-SIMS

- Multiplex immunohistochemistry image cytometry

- Digital immunohistochemistry platform for the staining variation monitoring based on integration of image and statistical analyses with laboratory information system

- Automatic single cell segmentation on highly multiplexed tissue images

- Computerized methods for tissue analysis using n-gene profiling

Reproducible, high-dimensional imaging in archival human tissue by Multiplexed Ion Beam Imaging by Time-of-Flight (MIBI-TOF)

Candace C. Liu[1][†], Marc Bosse[1][†], Alex Kong[1], Adam Kagel[1], Robert Kinders[2], Stephen M. Hewitt[3], Sean C. Bendall[1][*], Michael Angelo[1][*]

[1]Department of Pathology, Stanford University, Stanford, California 94304
[2]Office of the Director, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD 20892
[3]Experimental Pathology Laboratory, Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892

[†]Authors contributed equally to this work
[*]Corresponding Author

Abstract

Multiplexed ion beam imaging by time-of-flight (MIBI-TOF) is a form of mass spectrometry imaging that uses metal labeled antibodies and secondary ion mass spectrometry to image dozens of proteins simultaneously in the same tissue section. Working with the National Cancer Institute's (NCI) Cancer Immune Monitoring and Analysis Centers (CIMAC), we undertook a validation study, assessing concordance across a dozen serial sections of a tissue microarray of 21 samples that were independently processed and imaged by MIBI-TOF or single-plex immunohistochemistry (IHC) over 12 days. Pixel-level features were highly concordant across all 16 targets assessed in both staining intensity ($R^2 = 0.94 \pm 0.04$) and frequency ($R^2 = 0.95 \pm 0.04$). Comparison to digitized, single-plex IHC on adjacent serial sections revealed similar concordance ($R^2 = 0.85 \pm 0.08$) as well. Lastly, automated segmentation and clustering of eight cell populations found that cell frequencies between replicates yielded an average correlation of $R^2 = 0.92 \pm 0.06$. Taken together, we demonstrate that MIBI-TOF, with well-vetted reagents and automated analysis, can generate consistent and quantitative annotations of clinically relevant cell states in archival human tissue, and more broadly, present a scalable framework for benchmarking multiplexed IHC approaches.

Introduction

34

35   Immunohistochemistry (IHC) is commonly used in clinical diagnostics and basic research

36   to visualize proteins in intact tissue using chromogenic or fluorescent reporters.[1–3] IHC staining is

37   used routinely to guide diagnoses and therapeutic selection in the vast majority of solid tissue

38   malignancies.[4–6] Although it remains an indispensable tool in anatomic pathology, chromogenic

39   IHC has inherent limitations that hinder quantitative interpretation and prevent routine multiplexed

40   staining.[1,7 8–10] These drawbacks are particularly limiting in the field of cancer immunotherapy,

41   where accurate evaluation of the tumor immune microenvironment requires the simultaneous

42   mapping of dozens of proteins.[11,12] The emerging field of spatial-omics with a multitude of

43   analytical solutions is working towards replacement of conventional IHC-driven decision

44   making.[11,13] While it is clear that new technologies and assays have the potential to generate new

45   types of data, it is unclear whether they can reliably replicate the decision-making information in

46   current 'gold standard' assays.

47   Our lab has developed multiplexed ion beam imaging by time-of-flight (MIBI-TOF), which

48   avoids the limitations of optical imaging by using secondary mass spectrometry to image dozens

49   of proteins on the same tissue section.[14,15] In the place of chromogenic or fluorescent reporters,

50   MIBI-TOF uses primary antibodies that are labeled with isotopically enriched metal reporters that

51   can be cleanly delineated and quantified using time-of-flight (TOF) mass spectrometry. Tissue

52   sections are treated with all metal labeled primary antibodies simultaneously using a simple

53   protocol that does not include secondary antibodies, enzymatic amplification, or cyclical staining.

54   During MIBI-TOF analysis, the tissue is sputtered by a primary ion beam in a pixel-by-pixel fashion

55   that liberates the metal tags as secondary ions that are subsequently quantified by TOF. Our lab

56   routinely quantifies 40 targets simultaneously and are currently working towards increasing this

57   capability to 60 or more.[16–18] Notably, MIBI-TOF is compatible with formalin-fixed, paraffin-

58   embedded (FFPE) samples and can detect both low and high abundant targets with a dynamic

59   range that spans six orders of magnitude.[15 19–21]

60   For technologies like MIBI-TOF to be used in large translational studies and ultimately for

61   routine clinical diagnostics, robustness and reproducibility studies are needed, as are

62   standardized workflows for interpreting these complex data. Furthermore, it is important to show

63   that new technologies are concordant with existing, established clinical assays. In collaboration

64   with the National Cancer Institute (NCI) as part of the Cancer Immune Monitoring and Analysis

65   Centers-Cancer Immunologic Data Commons (CIMAC-CIDC) network[22], we benchmarked the

66   reproducibility of multiplexed antibody staining using MIBI-TOF. To achieve this, we compared

67   MIBI-TOF imaging data across six independent replicates of adjacent tissue microarray (TMA)

68   serial sections. Antibody staining and MIBI-TOF imaging of each serial section were carried out

69   independently and randomized with respect to all experimental parameters. Additionally, we

70   assessed MIBI-TOF concordance with single-plex IHC chromogenic staining. In keeping with the

71   goal of CIMAC-CIDC to develop comprehensive, standardized immune monitoring analysis for

72   biomarker discovery, we show that MIBI-TOF is a highly reproducible assay and concordant with

73   single-plex IHC.

74

75   **Materials and Methods**

76   **Tissue microarray construction and sectioning**

77   A tissue microarray (TMA) was constructed using human FFPE blocks from Stanford

78   Pathology. The TMA consisted of disease-free controls as well as multiple types of carcinomas,

79   sarcomas, and central nervous system lesions. A table of the cores included in this study can be

80   found in Table 1 (1 mm cores). For each TMA tissue block, 13 consecutive serial sections (4 µm

81   section thickness) were acquired. The IHC and MIBI-TOF recuts were alternated so that for the

82   IHC concordance analysis, MIBI-TOF and IHC could be compared in adjacent sections. The

83   assays performed on each section were as follows:

84   Section 1: H&E

85   Sections 2,4,6,8,10: Chromogenic IHC

3

86   Sections 3,5,7,9,11,13: MIBI-TOF full panel staining

87   Section 12: MIBI-TOF unstained control

Slide Number	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Replicate 5	Replicate 6
Stain Order	1	2	6	5	3	4
Day of Analysis	5	2	6	3	4	1

FOV	Order of Acquisition					
CONTROL TONSIL	14	4	17	1	19	1
CONTROL TONSIL	18	20	21	7	20	2
CONTROL TONSIL	19	10	6	5	13	3
CONTROL TONSIL	11	1	2	18	5	4
CONTROL LYMPH NODE	9	3	18	2	11	5
CONTROL LYMPH NODE	15	6	11	12	10	6
CONTROL LYMPH NODE	3	7	15	9	4	7
CONTROL THYMUS	7	11	14	20	3	8
CONTROL THYMUS	5	12	5	13	8	9
CONTROL SPLEEN	17	19	16	16	21	10
SALIVARY WARTHIN SALIVARY	2	17	4	4	9	11
HIGH GRADE, CARCINOMA BLADDER	16	15	12	19	18	12
CONTROL PLACENTA	8	9	20	14	7	13
DUCTAL CARCINOMA BREAST	20	18	3	21	1	14
DUCTAL CARCINOMA BREAST	1	2	13	15	12	15
LEIOMYOSARCOMA, MET SOFT TISSUE	12	13	7	8	6	16
HIGH-GRADE MYXOFIBROSARCOMA SOFT TISSUE	6	16	19	10	15	17
FOREIGN BODY GIANT CELLS SOFT TISSUE	21	21	9	6	14	18
ADENOCARCINOMA COLON	10	8	1	3	2	19
PLEOMORPHIC DERMAL SARCOMA SKIN	13	14	10	11	16	20
SQUAMOUS CELL CARCINOMA SKIN	4	5	8	17	17	21

Table 1: Description of tissue cores and parameter randomization

88

89   **Antibody preparation**

90   The full MIBI-TOF panel containing 16 antibodies can be found in Table 2. The lyophilized

91   antibody panel was obtained from Ionpath ("Cell Classification" human panel). Individual metal

92   labeled antibodies for IHC were also obtained from Ionpath: CD3 (D7A6E), CD8 (C8/144B), CD68

93   (D4B9C), Pax5 (D7H5X), PanCK (AE1/AE3). Prior to staining, the lyophilized antibody panel was

94   suspended in antibody diluent buffer (TBS-IHC tween, Donkey serum 3%) and filtered with a 0.1

95   μm filter (Millipore).

Target	Clone
dsDNA	35I9 DNA
beta-tubulin	D3U1W
CD4	EPR6855
CD11c	EP1347Y
CD56	MRQ-42
CD31	EP3095
CD68	D4B9C
CD8	C8/144B
CD3	D7A6E
Vimentin	D21H3
CD20	L26
HLA DR	EPR3692
CD45	2B11 & PD7/26
Na-K-ATPase alpha 1	D4Y7E
Pax5	D7H5X
PanCK	AE1/AE3

Table 2: MIBI-TOF staining panel. Targets used for IHC concordance analysis are highlighted in green.

Tissue staining

96

97   Tissue staining for the six MIBI-TOF recuts was carried out over six separate days,

98 randomized with respect to serial section number (Table 1). Detailed protocols can be found here:

99 dx.doi.org/10.17504/protocols.io.byzrpx56,   dx.doi.org/10.17504/protocols.io.bf6ajrae,

100 dx.doi.org/10.17504/protocols.io.bhmej43e. We used the Sequenza manual staining system for

101 both MIBI-TOF and IHC staining (dx.doi.org/10.17504/protocols.io.bmc6k2ze).

102   Briefly, slides were baked at 70°C overnight followed by deparaffinization and rehydration

103 with successive washes of 30 seconds and 3 dips each in xylene (3 washes), 100% ethanol (2

104 washes), 95% ethanol (2 washes), 80% ethanol (1 wash), 70% ethanol (1 wash), and ddH2O (2

105 washes) with a Leica ST4020 Linear Stainer (Leica Biosystems). Tissues next underwent antigen

106 retrieval by submerging slides in 3-in-1 Target Retrieval Solution (pH 9, DAKO Agilent) and

107 incubating at 97°C for 40 minutes in a Lab Vision PT Module (Thermo Fisher Scientific). After

108 cooling to room temperature, slides were assembled with a cover plate on a Sequenza rack and

109 washed with PBS wash buffer. The tissue was blocked for 1 hour at room temperature with 1x

110 TBS IHC Wash Buffer with Tween 20 with 3% (v/v) normal donkey serum (Sigma-Aldrich), 0.1%

5

111   (v/v) cold fish skin gelatin (Sigma Aldrich), 0.1% (v/v) Triton X-100, and 0.05% (v/v) sodium azide.

112   The blocking buffer was washed by adding 200 µL of antibody diluent buffer in the Sequenza

113   upper chamber. After the diluent buffer flow through, 120 µL of the suspended panel of antibodies

114   was added to the slides. The Sequenza rack was then placed at 4°C overnight (16 hr).

115   Following the overnight incubation, slides were washed twice with 1 mL of PBS wash

116   buffer and fixed in a solution of 2% glutaraldehyde (Electron Microscopy Sciences) solution in

117   low-barium PBS for 5 minutes. Slides were successively washed with 30 seconds and 3 dips per

118   wash in PBS (1 wash), 0.1 M Tris at pH 8.5 (3 washes), ddH2O (2 washes), and then dehydrated

119   by washing in 70% ethanol (1 wash), 80% ethanol (1 wash), 95% ethanol (2 washes), and 100%

120   ethanol (2 washes). Slides were dried under vacuum prior to imaging.

121   Single-plex chromogenic IHC was performed for five targets using the same antibody

122   clone as used in the MIBI-TOF staining panel (Table 2). Dewaxing, epitope retrieval, blocking,

123   hybridization, and washing for IHC tissue sections were identical to that of MIBI, with the addition

124   of blocking endogenous peroxidase activity with 3% H_2O_2 (Sigma Aldrich) in ddH2O after epitope

125   retrieval. After overnight primary antibody staining, tissues were washed twice with 1 mL wash

126   buffer and the antigen:antibody complex detected with the ImmPRESS universal (Anti-

127   Mouse/Anti-Rabbit) kit (Vector labs). Chromogenic solution (DAB) was incubated for 40 s and the

128   reaction was stopped with PBS. The slides were counterstained with hematoxylin.

129   H&E tissue sections were reviewed to identify missing tissue cores, cores with large areas

130   of necrosis, tissue folding, and any other macroscopic defect. All IHC and MIBI-TOF data were

131   manually reviewed for excessive background staining.

132   **MIBI-TOF imaging**

133   The order of image acquisition for each recut was randomized with respect to section

134   number and carried out on six separate days (Table 1). Within each run for each recut, the order

135   of acquisition for the TMA cores was randomized as well (Table 1). Area normalized Xe+ primary

136   ion dose of 9nA*hr*mm^{-2} was used for all MIBI-TOF image acquisitions. At the end of each

6

137   imaging run, sensitivity and mass resolution were quantified using a molybdenum foil standard

138   (Supplementary Figure 1).

139   **Image processing**

140   MIBI-TOF data processing was performed using a standardized pipeline from Ionpath.

141   During data extraction, to only capture the monoatomic peaks that correspond to the monoatomic

142   isotopic labels and avoid the polyatomic ions containing hydrogen, we integrated all peaks using

143   a mass window of (X-0.3, X) where X is the nominal mass-to-charge (m/z) for each metal reporter

144   tag.

145   Background signal arising from bare slide was removed as described previously using

146   predefined thresholds for Ba, Ta, and Au.[16,23] Isobaric interferences arising from diatomic reporter

147   adducts and polyatomic hydrocarbons were removed iteratively using empirically defined

148   compensation coefficients (Supplementary Table 1). Lastly, we used a combination of several

149   density metrics, including reachability density, connectivity, and Voronoi tessellation to estimate

150   and remove noise from each channel (Supplementary Table 2). These density metrics leverage

151   both spatial and intensity information.

152   Hematoxylin counterstained IHC slides were digitized with the NanoZoomer slide scanner

153   at 40x magnification (Hamamatsu). The threshold for hematoxylin and each DAB channel was

154   set collectively for all images in the staining batch using a DAB channel extracted from the original

155   counterstained image via spectral unmixing. An initial threshold approximation was set with

156   automatic image histogram thresholding. The thresholds were then confirmed by three trained

157   users on controls and additional images for validation and adjusted until consensus was reached

158   (Supplementary Table 3).

159   **Marker quantitation**

160   In all MIBI-TOF images, for each marker, the frequency and mean intensity of positive

161   pixels in each tissue core was quantified. The resultant value for percent positive pixels (PPP)

162   and mean pixel intensity (MPI) for these images was then normalized with respect to the signal

163   intensity of endogenous carbon-containing mass peaks originating from the tissue itself. Peaks at

164   36, 37 and 38 m/z corresponding to 3C, 3C1H, and 3C2H were integrated for each image (referred

165   to here as C_i) and the average of these values across all images in all six TMAs was calculated

166   (referred here to C_{avg}). To calculate the normalization coefficient (C_N) for a given tissue core, C_i

167   for that tissue core was divided by C_{avg} ($C_N = C_i / C_{avg}$). Median and standard deviation for C_N

168   across all tissue cores was 1.03 ± 0.3. For IHC images, the frequency of positive pixels was

169   quantified using deconvolved and thresholded DAB images (PPP).

170   **Reproducibility and concordance**

171   To assess replicate reproducibility, least squares linear regression was used to compare

172   PPP for tissue cores in each TMA to the average PPP across all six TMAs. Similarly, least squares

173   linear regression was used to compare MPI for tissue cores in each TMA to the average MPI

174   across all six TMAs. For each comparison, we calculated the slope (m) and coefficient of

175   determination (R^2). To assess concordance of MIBI-TOF with single-plex IHC, we used least

176   squares linear regression to compare values for PPP attained by MIBI-TOF and IHC in adjacent

177   serial tissue sections.

178   **Cell segmentation and phenotyping**

179   To delineate the location of single cells in MIBI-TOF images, we performed cell

180   segmentation using the pre-trained Mesmer convolutional neural network architecture.[24] We used

181   dsDNA as the nuclear marker and HLA class I and Na-K-ATPase as the membrane markers as

182   input to the network. The output of Mesmer is the location of each cell in the image.

183   After cell segmentation, the next step of the analysis pipeline was to determine the

184   phenotype of each individual cell. Pre-processed MIBI-TOF images were first Gaussian blurred

185   using a standard deviation of 2 for the Gaussian kernel. Pixels were normalized by their total

186   expression, such that the total expression of each pixel was equal to 1. A 99.9% normalization

187   was applied for each marker. Pixels were clustered into 100 clusters using FlowSOM[25] based on

188   the expression of 12 phenotypic markers: CD3, CD4, CD8, CD11c, CD20, CD31, CD45, CD56,

8

189 CD68, PanCK, Pax5, and Vimentin. The average expression of each of the 100 pixel clusters was

190 found and the z-score for each marker across the 100 pixel clusters was computed. All z-scores

191 were capped at 3, such that the maximum z-score was 3. Using these z-scored expression values,

192 the 100 pixel clusters were meta-clustered using consensus hierarchical clustering into 12 meta-

193 clusters. Next, by applying the segmentation masks that delineate the boundaries of all cells in

194 the images, we counted the number of each of the 12 pixel clusters in each cell. This resulted in

195 a pixel cluster by cell count table. These counts were then normalized by cell size. Using these

196 frequency measurements as the feature vector, the cells were clustered using FlowSOM into 100

197 cell clusters. Similarly to the pixel clusters, the average expression of each of the 100 cell clusters

198 was found and the z-score was computed. All z-scores were capped at 3, such that the maximum

199 z-score was 3. Using these z-scored values, the 100 cell clusters were meta-clustered using

200 consensus hierarchical clustering into 10 cell meta-clusters. Each of the cell meta-clusters was

201 then manually annotated with its cell phenotype by assessing marker expression, resulting in a

202 total of eight cell types: B cells, T cells, dendritic cells, macrophages, NK cells, fibroblasts,

203 endothelial cells, and epithelial cells. To assess the concordance of cell phenotyping between

204 serial sections of the same tissue core, we quantified the number of each cell type in each image,

205 then calculated the average Spearman correlation of these frequencies between replicate images

206 of each core.

207

208 **Results**

209 **Experimental design for MIBI-TOF replicate and IHC concordance**

210 To assess the reproducibility of MIBI as well as benchmark against single-plex

211 chromogenic IHC, we took 13 serial sections from a TMA constructed for this study (see Methods)

212 that included disease-free controls and multiple types of carcinomas, sarcomas, and central

213 nervous system lesions (Figure 1). Here, every other slide was stained and processed for either

214 MIBI-TOF analysis or single-plex chromogenic IHC for the indicated target, in addition to slides

9

215   for H&E and unstained tissue control. To control for confounding factors due to batch effects

216   resulting from day of tissue staining and imaging, the order that each TMA was stained and

217   imaged were randomized with respect to serial sectioning order. Additionally, the order that each

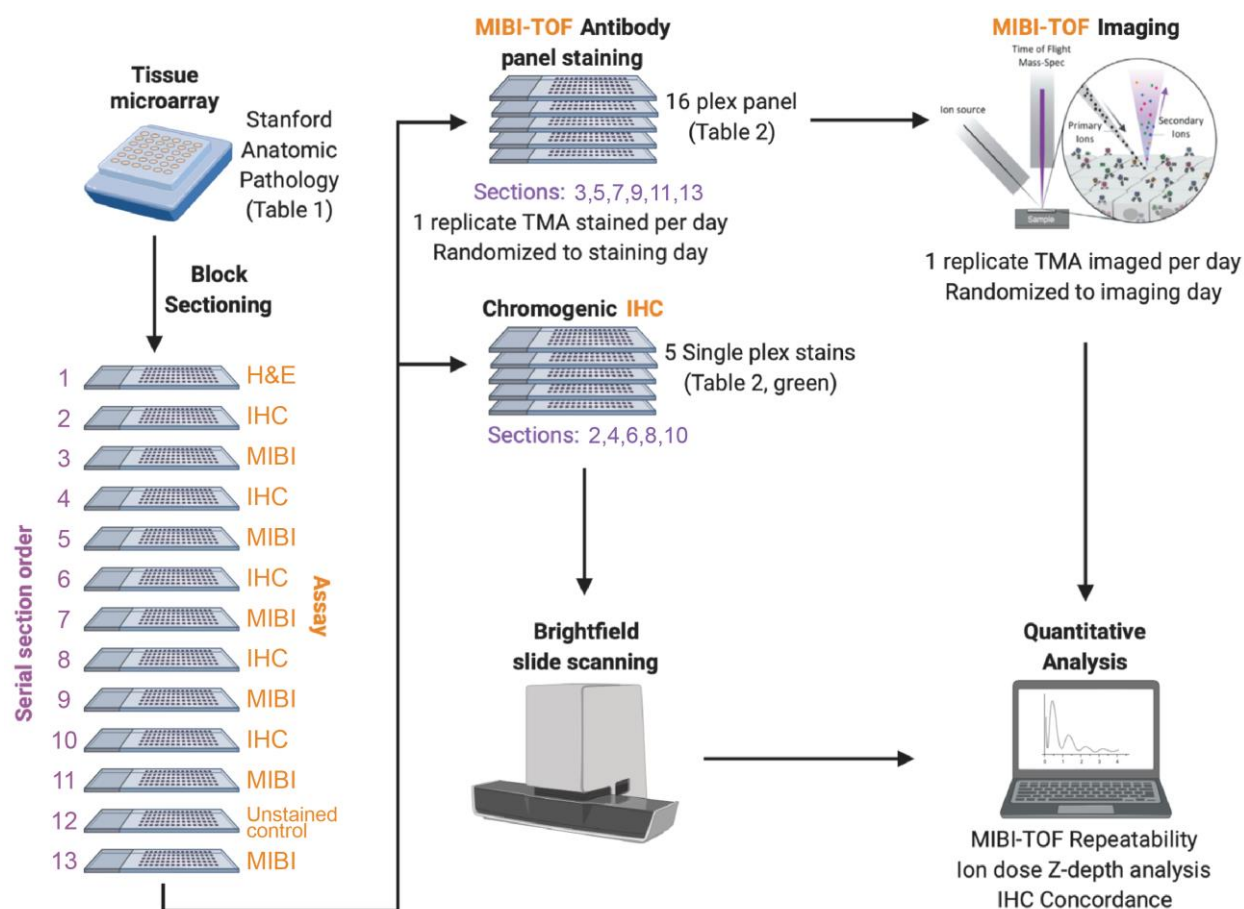218   tissue core was acquired for each TMA was randomized as well (Table 1).

Figure 1: MIBI-TOF validation overview. A tissue microarray (TMA) was constructed using human FFPE tissue blocks, and the TMA was serially sectioned for analysis by single-plex chromogenic IHC and MIBI-TOF. The order that each serial section was stained and imaged was randomized. We then analyzed the concordance between IHC and MIBI-TOF, as well as assessed reproducibility of MIBI-TOF between serial sections.

219

MIBI-TOF replicate concordance

221   For MIBI-TOF, 21 tissue cores in six replicate TMAs were stained with a 16-plex antibody

222   panel and analyzed for a total of 126 MIBI-TOF images (Table 2). Instrument stability was

223   assessed prior to data acquisition using a molybdenum foil standard, for which we know the

10

224 expected range of ion counts. Primary ion current and mass resolution varied insignificantly over

225 the course of the study (SD = 7% and 1.8%, respectively, Supplementary Figure 1A,B). Minor

226 variations in ion detector sensitivity (SD = 22%) were within historical norms and were not found

227 to impact subsequent quantitative comparisons (Supplementary Figure 1C). Fields in each tissue

228 core were manually co-registered with respect to a slide scan of a reference H&E serial section,

229 which showed high concordance of histological features (Supplementary Figure 2). High quality

230 MIBI-TOF imaging data was obtained for all targets in the antibody staining panel in all six TMA

231 slides. Manual evaluation of each marker demonstrated robust and specific staining to the target

232 cells of interest in both normal and disease tissue cores (Figure 2).

233 Assay reproducibility was assessed by comparing marker intensity and abundance in each

234 TMA replicate. An example of co-registered fields-of-view (FOVs) from 6 serial sections of a

235 representative tissue core is shown in Figure 3A, illustrating the high degree of reproducibility

236 between the replicate sections. PPP values for CD3, a marker for T cells, and CD20, a marker for

237 B cells, in the fields for this tissue core were 22.6% ± 2.5 and 20.6% ± 8.1 (mean ± SD),

238 respectively. The significantly higher standard deviation observed with CD20 is attributable to

239 section-to-section changes in the area occupied by two germinal centers (marked by expression

240 of the B cell marker CD20) as sections were taken from the tissue core.

241 Linear regression of normalized values for marker intensity (MPI) and marker frequency

242 (PPP) for each tissue core in each TMA replicate (see Methods) were used to determine the

243 degree of concordance with respect to the average across all six replicates. For MPI, which

244 quantifies the average marker intensity in positive pixels, the resultant slope (m) and coefficient

245 of determination (R^2) values for all markers in all six replicates were close to 1, with a m of 0.96 ±

246 0.14 and a R^2 of 0.94 ± 0.04 (mean ± SD, Figure 3B-D). MPI captures marker intensity, thus m

247 and R^2 values close to 1 indicate that when a marker is present, the brightness of the marker is

248 highly concordant across replicates. Similarly, for PPP, which quantifies the total number of pixels

249 that are positive for a given marker, m and R^2 in all replicates and markers were also close to 1,
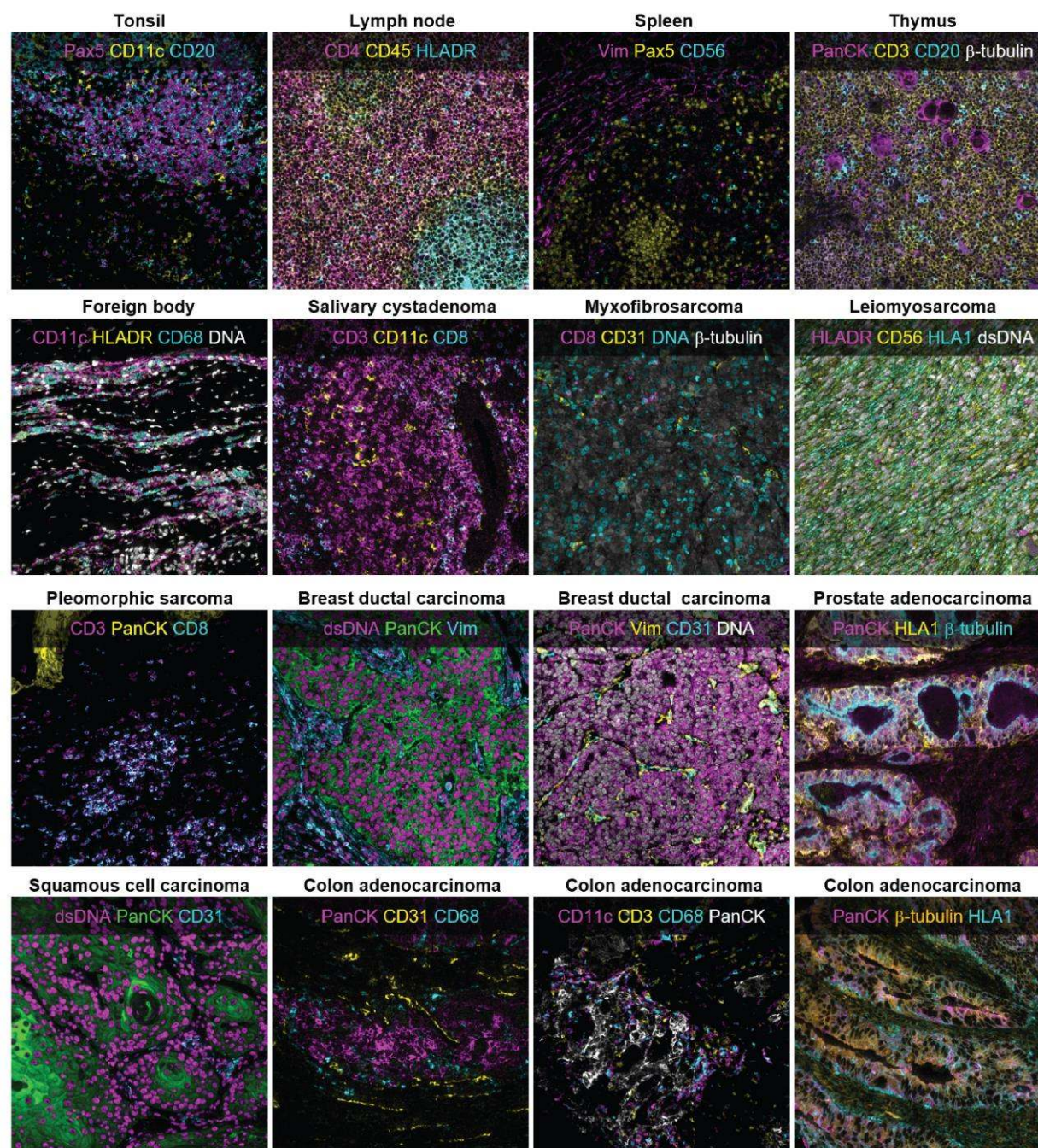
Figure 2: Representative MIBI-TOF imaging data. High-quality MIBI-TOF imaging data was attained for 16 markers for all replicate TMA slides. Representative images of the various tissue types in the TMA are shown here.

250   with a m of 0.98 ± 0.15 and a R^2 of 0.95 ± 0.04 (mean $\pm$ SD, Supplementary Figure 3). PPP m

251   and R^2 values close to 1 indicate that the number of positive pixels across replicates of the same

252   core is consistent. Taken together, these two metrics demonstrate that MIBI-TOF images of serial

12

253 sections of the same tissue core that were randomized for staining and imaging day were highly

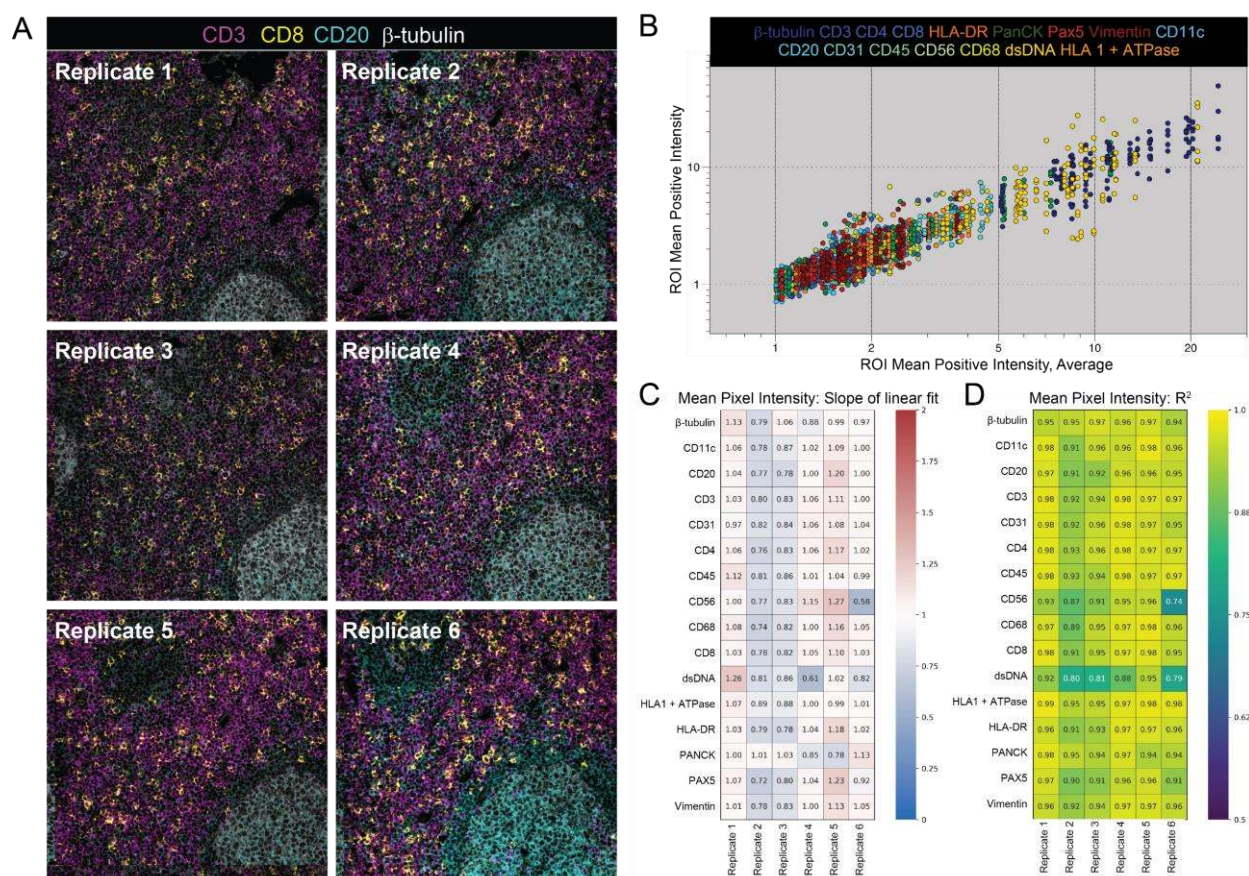254 concordant, lending evidence for the reproducible nature of MIBI-TOF.

Figure 3: Concordance of Mean Pixel Intensity by MIBI-TOF. (A) Representative MIBI-TOF replicate images from serial sections of the same TMA core of lymph node tissue. (B) Plot of Mean Positive Intensity (MPI) of each FOV vs. the average MPI of all FOVs of the same TMA core. Each color represents a different marker. We performed least squares linear regression of the MPI of all six replicates of each TMA core vs the average MPI of the core and found the slope m (C) and coefficient of determination R^2 (D) for all 16 markers.

Reproducibility of cell immunophenotypic assignments by MIBI-TOF

256 An important step in the analysis of high-dimensional imaging data is the identification of

257 single cell lineages by immunophenotype in the images. Thus, we next assessed the

258 reproducibility of single-cell annotations across the replicate images of each tissue core. Our lab

259 has recently developed a cell segmentation algorithm that can accurately identify the location of

260 cells in solid tissue without the need for manual fine-tuning or user adjustment.[24] We used this

13

261 method to delineate the location of single cells in the images, then used a pixel-clustering based

262 approach to classify the single cells into eight cell lineages: B cells, T cells, dendritic cells,

263 macrophages, NK cells, fibroblasts, endothelial cells, and epithelial cells (Figure 4A, see

264 Methods).[17] After classifying the cells in each of the images, we quantified the number of each

265 cell type in each image and assessed concordance between replicates by calculating the average
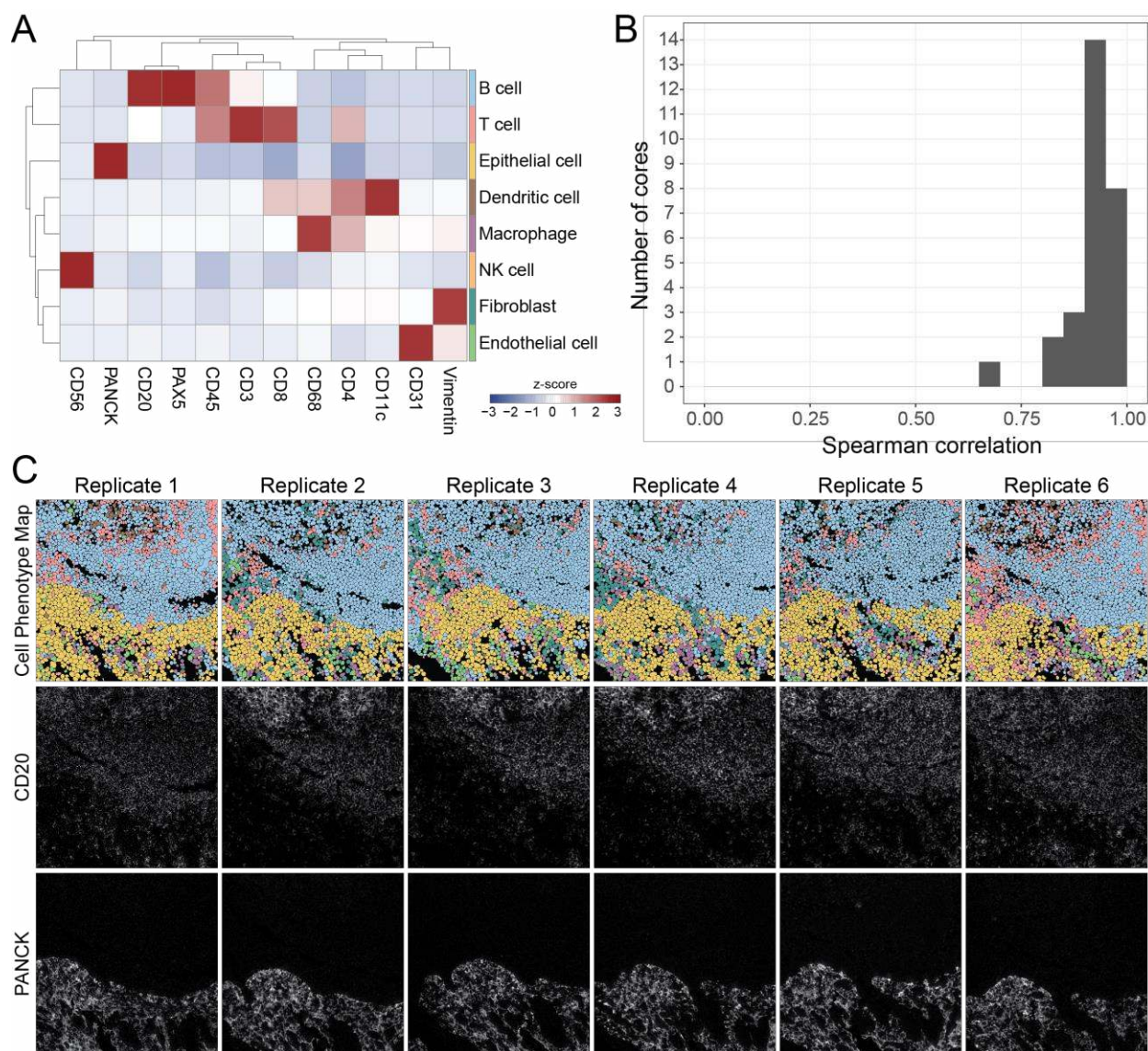

Figure 4: Reproducibility of cell phenotyping by MIBI-TOF. (A) Cells were assigned to a phenotype using an unsupervised clustering approach and manually annotated into eight major cell types. Expression values were z-scored for each marker. High z-scores are an indicator of high marker specificity. (B) The Spearman correlation between all serial sections of each TMA core using the frequency of cell types in each FOV. (C) Representative images of six serial sections of the same TMA core of ttonsil tissue. Top row: Cell phenotype map colored according to the eight cell types shown in (A). Middle and bottom rows: MIBI-TOF images showing CD20 and PanCK expression.

14

266   Spearman correlation of these frequencies between the replicate images of each tissue core

267   (Figure 4B). We expected variation in these Spearman correlation values, since the replicate

268   sections of the same tissue core are serial sections of 4 μm each (with additional serial sections

269   taken for IHC concordance analysis). In total, 13 serial sections were obtained for each tissue for

270   MIBI-TOF or IHC (see Methods). Therefore, we would expect true biological differences when

271   comparing serial sections up to 48 μm apart. When comparing these replicate MIBI-TOF images,

272   we observed that broad histological features were conserved, such as germinal centers or vessels

273   (Figure 3A, 4C, Supplementary Figure 4). The average Spearman correlation was high, 0.92 ±

274   0.06 (mean ± SD), demonstrating that cell phenotyping, a necessary step in high-dimensional

275   image analysis, was reproducible between MIBI-TOF images of the same tissue core.

276

Concordance of MIBI-TOF with single-plex chromogenic IHC

278   Chromogenic IHC is the standard modality for disease prognosis and therapeutic selection

279   in the vast majority of solid tumors. Therefore, we assessed concordance to determine if

280   multiplexed imaging by MIBI-TOF can quantitatively recapitulate single-plex immunoassays.

281   Similar to the analyses described above that were used to compare MIBI-TOF data across

282   replicates, the frequency of marker positive pixels in MIBI-TOF images co-registered with single-

283   plex chromogenic IHC stains were compared by linear regression across all 21 tissue cores.

284   Because we alternated recuts for MIBI-TOF and IHC, we could compare MIBI-TOF and IHC on

285   adjacent serial sections. Single-plex chromogenic IHC images were acquired for five targets:

286   CD3, CD8, Pax5, PanCK, and CD68. Marker positivity by single-plex IHC for CD3, CD8, Pax5,

287   PanCK, and CD68 in adjacent serial sections were compared with the respective values found by

288   MIBI-TOF using linear regression (Figure 5, Supplementary Figure 5). For all five targets,

289   concordance of MIBI-TOF and IHC was high, with an R^2 of 0.85 ± 0.08 (mean ± SD, all R^2 > 0.7),

290   demonstrating that MIBI-TOF data is consistent with single-plex chromogenic IHC data.
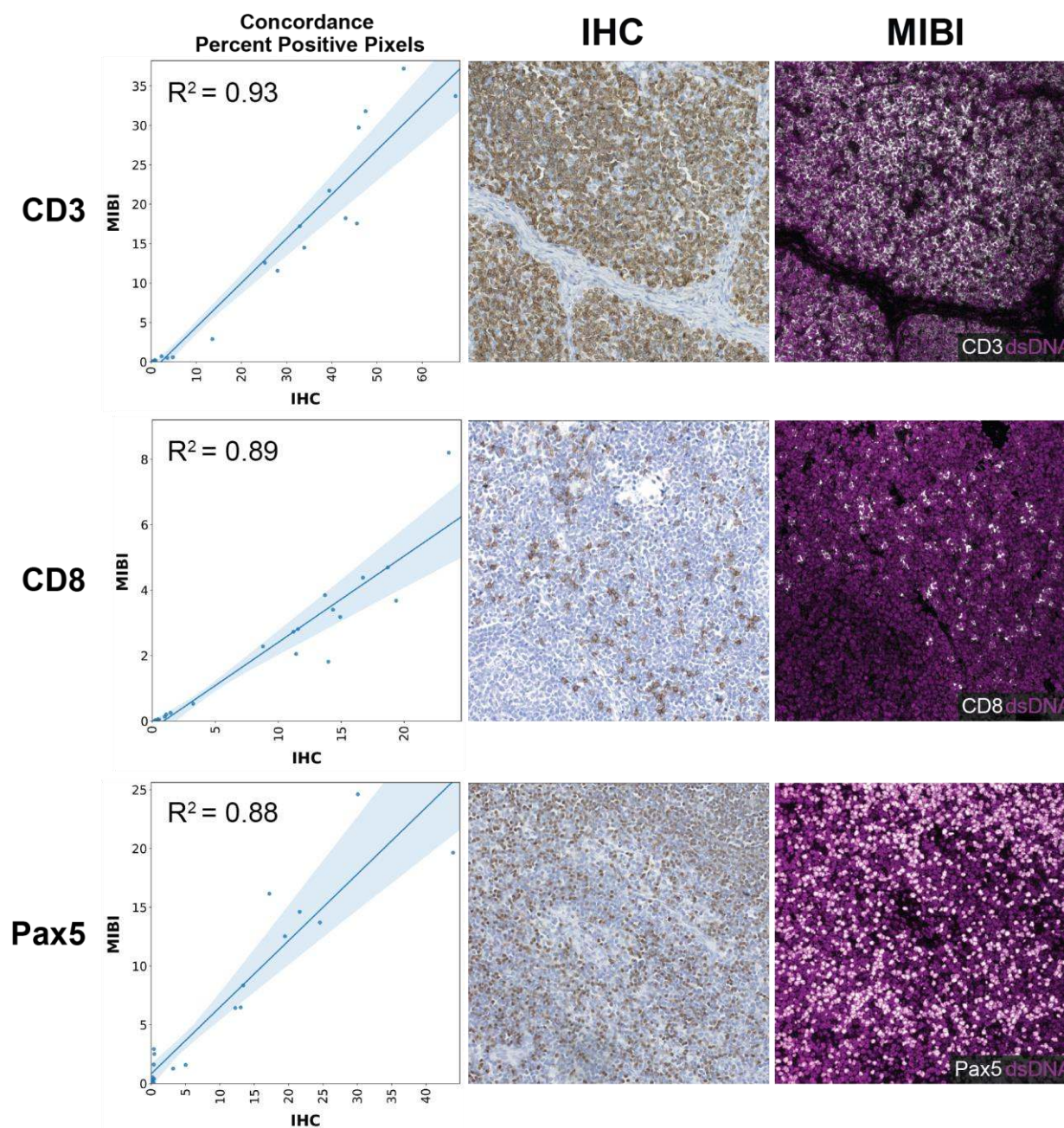
Figure 5: Concordance of MIBI-TOF with single-plex chromogenic IHC. Representative images of the comparison of MIBI-TOF images co-registered with single-plex chromogenic IHC stains. Each data point represents the PPP by MIBI-TOF and IHC for a single tissue core. Shaded area represents 95% confidence interval. Additional markers are shown in Supplementary Figure 5.

291

Discussion

293   In this study, we assessed the reproducibility of MIBI-TOF by evaluating assay

294 concordance on six serial sections from a TMA, including disease-free tonsil, lymph node, thymus,

16

295   and spleen, in addition to multiple types of carcinomas, sarcomas, and central nervous system

296   lesions. Importantly, each serial section was stained independently on six separate days with a

297   16-marker staining panel and subsequently analyzed on six different days. 21 tissue cores were

298   analyzed in each serial section for a total of 126 MIBI-TOF images. The staining intensity and

299   frequency of positive pixels for each serial section (MPI and PPP, respectively) were compared

300   against average values across all runs using least squares linear regression. In all comparisons,

301   concordance between serial sections of the same tissue core was high, with m of 0.96 ± 0.14 and

302   R^2 of 0.94 ± 0.04 when assessing MPI and m of 0.98 ± 0.15 and R^2 of 0.95 ± 0.04 when assessing

303   PPP (mean ± SD). The staining pattern of each marker was manually inspected to confirm, where

304   relevant, appropriate subcellular localization and histologic distribution. For example, CD20

305   staining in positive regions of tonsil and lymph node in all replicates was verified to be membrane-

306   localized and enriched in follicles. Furthermore, we assessed the reproducibility of cell

307   phenotyping across replicates and found that across replicate sections, the location and lineage

308   of cells identified using our data analysis pipeline were highly concordant, with an average

309   Spearman correlation of cell type frequencies between replicates of 0.92 ± 0.06 (mean ± SD).

310   Importantly, the data analysis pipeline used to go from MIBI-TOF images to cell phenotypes was

311   nearly fully automated. Cell segmentation and clustering were fully automated with no user input

312   and only minimal user intervention was needed when labelling cell clusters with cell annotations,

313   demonstrating that the biological annotation of MIBI-TOF images is robust without the need for

314   significant researcher intervention. In addition, we assessed concordance of PPP by MIBI-TOF

315   for five markers with single-plex IHC and found that MIBI-TOF and single-plex IHC were highly

316   concordant with an average R^2 of 0.85. Taken together, the combination of consistent staining,

317   imaging, and feature extraction illustrates the highly reproducible nature of MIBI-TOF.

318   One limitation of validation studies involving serial tissue sections is that exact biological

319   replicates cannot be obtained since each tissue section is somewhat different and is not a

320   homogenized bulk sample. In this study, since the tissue sections are 4 µm each and a total of

17

321 13 serial sections were acquired, the first and last tissue section were separated axially by 48 μm.

322 Also, when balancing assay replication versus concordance with an orthogonal assay, like single-

323 plex IHC here, one needs to prioritize what physically adjacent slides are used for. In this case,

324 we selected physically adjacent slides for benchmarking single-plex IHC since any one marker's

325 distribution would be heavily dependent on cell composition. Lymphocytes that are approximately

326 ~10 μm would be best preserved in adjacent sections. Although there will be true biological

327 differences in this volume of tissue, this study represents the next best situation, in which we

328 compared adjacent serial sections as a proxy for true replicates. Furthermore, this was a single-

329 site study, in which all staining and imaging was performed in our laboratory. We are currently

330 planning a multi-site study to assess inter-institutional differences in instrument performance and

331 tissue staining. Still, we believe the overall metrics for benchmarking outlined here will be

332 sufficient to assess these future efforts.

333 This study was performed as a part of the Cancer Immune Monitoring and Analysis

334 Centers Cancer Immunology Data Commons (CIMAC-CIDC) network[22], which is a NCI Cancer

335 Moonshots initiative to provide the technology and expertise for immunotherapy clinical trials.

336 Multiplexed tissue imaging is an integral assay for fully characterizing the tumor immune

337 microenvironment, which requires the simultaneous profiling of multiple tumor and immune cell

338 types. MIBI-TOF can routinely provide quantitative, multiplexed imaging data that is back

339 compatible with archival FFPE tissue and conventional anatomic pathology workflows. Technical

340 innovations in reagents and instrumentation will further increase the throughput and multiplexing

341 ability of MIBI-TOF. In the future, we envision that MIBI-TOF will not only be used for reproducible

342 basic science research, but also enable the adoption of quantitative spatial signatures in the clinic

343 for more accurate diagnosis and therapeutic selection. At the same time, as efforts to develop

344 other IHC-centric spatial imaging technologies evolve, the analytical framework and approach laid

345 out in this study should serve as a guide for their assessment.

346

347 **Data Availability**

348 The datasets generated and/or analyzed during the current study are available in Zenodo:

349 10.5281/zenodo.5542727.

350

351 **Acknowledgements**

352 The authors thank Sushama Varma and Matt van de Rijn for creating the tissue microarray for

353 this study.

354

355 **Conflict of Interest**

356 M.A. and S.C.B. are inventors on patents related to MIBI technology. M.A. and S.C.B. are

357 consultants, board members and shareholders in Ionpath Inc.

358

359 **Author Contributions**

360 M.B. performed experiments and imaging. A.K., A.K., and C.C.L. performed data analysis. C.C.L.

361 and M.A. wrote the manuscript. S.C.B., R.K., and S.M.H. reviewed the manuscript. S.C.B. and

362 M.A. supervised the work.

363

364 **Funding**

365 This work was supported by DOD EOH W81XWH2110143, 1-DP5-864 OD019822,

366 1R01AG056287, 1R01AG057915, R01AG068279, 1UH3CA246633, 1U24CA224309, and the

367 Bill and Melinda Gates Foundation.

References

1. Levenson, R. M., Borowsky, A. D. & Angelo, M. Immunohistochemistry and mass spectrometry for highly multiplexed cellular molecular imaging. *Lab Invest* **95**, 397–405 (2015).

2. Matos, L. L. de, Trufelli, D. C., de Matos, M. G. L. & da Silva Pinhal, M. A. Immunohistochemistry as an important tool in biomarkers detection and clinical practice. *Biomark Insights* **5**, 9–20 (2010).

3. Chlipala, E. A. *et al.* An Image Analysis Solution For Quantification and Determination of Immunohistochemistry Staining Reproducibility. *Appl Immunohistochem Mol Morphol* **28**, 428–436 (2020).

4. Thunnissen, E. *et al.* The Use of Immunohistochemistry Improves the Diagnosis of Small Cell Lung Cancer and Its Differential Diagnosis. An International Reproducibility Study in a Demanding Set of Cases. *Journal of Thoracic Oncology* **12**, 334–346 (2017).

5. Yatabe, Y. *et al.* Best Practices Recommendations for Diagnostic Immunohistochemistry in Lung Cancer. *Journal of Thoracic Oncology* **14**, 377–407 (2019).

6. Zaha, D. C. Significance of immunohistochemistry in breast cancer. *World J Clin Oncol* **5**, 382–392 (2014).

7. McCabe, A., Dolled-Filhart, M., Camp, R. L. & Rimm, D. L. Automated quantitative analysis (AQUA) of in situ protein expression, antibody concentration, and prognosis. *J Natl Cancer Inst* **97**, 1808–1815 (2005).

8. Taube, J. M. *et al.* The Society for Immunotherapy of Cancer statement on best practices for multiplex immunohistochemistry (IHC) and immunofluorescence (IF) staining and validation. *J Immunother Cancer* **8**, e000155 (2020).

9. Zimak, J., Schweller, R. M., Duose, D. Y., Hittelman, W. N. & Diehl, M. R. Programming in situ immunofluorescence intensities through interchangeable reactions of dynamic DNA complexes. *Chembiochem* **13**, 2722–2728 (2012).

394   10.   Neher, R. & Neher, E. Optimizing imaging parameters for the separation of multiple labels

395   in a fluorescence image. *J Microsc* **213**, 46–62 (2004).

396   11.   Patel, S. S. & Rodig, S. J. Overview of Tissue Imaging Methods. in *Biomarkers for*

397   *Immunotherapy of Cancer: Methods and Protocols* (eds. Thurin, M., Cesano, A. & Marincola,

398   F. M.) 455–465 (Springer New York, 2020). doi:10.1007/978-1-4939-9773-2_21.

399   12.   Binnewies, M. *et al.* Understanding the tumor immune microenvironment (TIME) for

400   effective therapy. *Nat Med* **24**, 541–550 (2018).

401   13.   Bodenmiller, B. Multiplexed Epitope-Based Tissue Imaging for Discovery and Healthcare

402   Applications. *Cell Systems* **2**, 225–238 (2016).

403   14.   Angelo, M. *et al.* Multiplexed ion beam imaging of human breast tumors. *Nat Med* **20**, 436–

404   442 (2014).

405   15.   Keren, L. *et al.* MIBI-TOF: A multiplexed imaging platform relates cellular phenotypes and

406   tissue structure. *Sci Adv* **5**, eaax5851 (2019).

407   16.   Keren, L. *et al.* A Structured Tumor-Immune Microenvironment in Triple Negative Breast

408   Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell* **174**, 1373-1387.e19 (2018).

409   17.   Risom, T. *et al.* Transition to invasive breast cancer is associated with progressive

410   changes in the structure and composition of tumor stroma. *bioRxiv* 2021.01.05.425362 (2021)

411   doi:10.1101/2021.01.05.425362.

412   18.   McCaffrey, E. F. *et al.* Multiplexed imaging of human tuberculosis granulomas uncovers

413   immunoregulatory features conserved across tissue and blood. *bioRxiv* 2020.06.08.140426

414   (2020) doi:10.1101/2020.06.08.140426.

415   19.   Goltsev, Y. *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed

416   Imaging. *Cell* **174**, 968-981.e15 (2018).

417   20.   Schulz, D. *et al.* Simultaneous Multiplexed Imaging of mRNA and Proteins with Subcellular

418   Resolution in Breast Cancer Tissue Samples by Mass Cytometry. *Cell Syst* **6**, 25-36.e5 (2018).

419   21.   Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by

420   mass cytometry. *Nat Methods* **11**, 417–422 (2014).

421   22.   Chen, H. X. *et al.* Network for Biomarker Immunoprofiling for Cancer Immunotherapy:

422   Cancer Immune Monitoring and Analysis Centers and Cancer Immunologic Data Commons

423   (CIMAC-CIDC). *Clin Cancer Res* **27**, 5038–5048 (2021).

424   23.   Baranski, A. *et al.* MAUI (MBI Analysis User Interface)—An image processing pipeline for

425   Multiplexed Mass Based Imaging. *PLOS Computational Biology* **17**, e1008887 (2021).

426   24.   Greenwald, N. F. *et al.* Whole-cell segmentation of tissue images with human-level

427   performance using large-scale data annotation and deep learning. *bioRxiv* 2021.03.01.431313

428   (2021) doi:10.1101/2021.03.01.431313.

429   25.   Van Gassen, S. *et al.* FlowSOM: Using self-organizing maps for visualization and

430   interpretation of cytometry data. *Cytometry A* **87**, 636–645 (2015).

431