1 **Systematic benchmarking of imaging spatial transcriptomics platforms in FFPE tissues**

2 Huan Wang[1,*], Ruixu Huang[2,*], Jack Nelson[1,*], Ce Gao[3], Miles Trans[3], Anna Yeaton[4], Kristen

3 Felt[5], Kathleen L. Pfaff[6], Teri Bowman[7], Scott J. Rodig[6,7], Kevin Wei[,3,7], Brittany A. Goods[2,**],

4 Samouil L. Farhi[1,**]

5

6 **Affiliations**

7 [1]Spatial Technology Platform, Broad Institute of MIT and Harvard, Cambridge, MA 02142

8 USA.

9 [2]Thayer School of Engineering, Molecular and Systems Biology, and Program in Quantitative

10 Biomedical Sciences at Dartmouth College, Hanover, NH 03755, USA.

11 [3]Division of Rheumatology, Inflammation, and Immunity, Brigham and Women's Hospital and

12 Harvard Medical School, Boston, MA 02215, USA.

13 [4]Present affiliation: Immunai, New York, NY 10016, USA.

14 [5]ImmunoProfile, Brigham & Women's Hospital and Dana-Farber Cancer Institute, Boston, MA

15 02215, USA.

16 [6]Center for Immuno-Oncology, Tissue Biomarker Laboratory, Dana-Farber Cancer Institute,

17 Boston, MA 02215, USA.

18 [7]Department of Pathology, Brigham and Women's Hospital, Boston, MA 02215, USA.

19

20 [*] These authors contributed equally to this work

21 [**] These authors jointly supervised this work

22

23 Corresponding author: Samouil L. Farhi: sfarhi@broadinstitute.org

Abstract

Emerging imaging spatial transcriptomics (iST) platforms and coupled analytical methods can recover cell-to-cell interactions, groups of spatially covarying genes, and gene signatures associated with pathological features, and are thus particularly well-suited for applications in formalin fixed paraffin embedded (FFPE) tissues. Here, we benchmarked the performance of three commercial iST platforms on serial sections from tissue microarrays (TMAs) containing 23 tumor and normal tissue types for both relative technical and biological performance. On matched genes, we found that 10X Xenium shows higher transcript counts per gene without sacrificing specificity, but that all three platforms concord to orthogonal RNA-seq datasets and can perform spatially resolved cell typing, albeit with different false discovery rates, cell segmentation error frequencies, and with varying degrees of sub-clustering for downstream biological analyses. Taken together, our analyses provide a comprehensive benchmark to guide the choice of iST method as researchers design studies with precious samples in this rapidly evolving field.

MAIN

Spatial transcriptomics (ST) tools measure the gene expression profiles of tissues or cells *in situ*. These approaches overcome the limitations of single-cell RNA-sequencing (scRNA-seq) methods by negating the need for cellularization and maintaining both local and global spatial relationships between cells within a tissue. ST can thus recover cell-cell interactions with high confidence, groups of spatially covarying genes, groups of cells predictive of cancer survival, and gene signatures associated with pathological features [1, 2]. These advantages, coupled with rapidly emerging computational and analytical methods, have led to substantial excitement about deploying these platforms in fundamental biology studies, and in the clinic for research and diagnostic purposes [3, 4, 5].

ST tools can be split into two broad categories: sequencing (sST) and imaging (iST) based modalities. sST methods tag transcripts with an oligonucleotide address indicating spatial location, most commonly by placing tissue slices on a barcoded substrate; isolating tagged mRNA for next-generation sequencing; and computationally mapping transcript identities to locations [6]. In contrast, iST methods most commonly use variations of fluorescence *in situ* hybridization (FISH) where mRNA molecules are tagged with hybridization probes which are detected in a combinatorial manner over multiple rounds of staining with fluorescent reporters, imaging, and de-staining **(Fig. 1a)** [7]**.** Computational reconstruction then yields maps of transcript identity with single-molecule resolution. Compared to sST methods, iST methods are targeted to subsets of the transcriptome due to their reliance on pre-defined gene panels and they adopt the higher spatial resolution and sensitivity of FISH, yielding single-cell resolution data [8].

While the iST methods share some similarities, significant differences arise in primary signal detection and amplification, sample processing, and the subsequent fluorescent cycling

61   chemistry (**Fig. 1b**) [9,10,11]. The need for amplification of signal is coupled to the sample processing,

62   namely whether the sample is cleared, gel-embedded, or photobleached to quench

63   autofluorescence. There are tradeoffs due to differences in sample processing for each iST method.

64   For example, clearing of the sample increases signal quality but can prevent follow-up H&E

65   staining and complicate immunostaining, which, in turn, can make cell segmentation more

66   challenging. Finally, there are tradeoffs between imaging time, molecular plex, and imaging area

67   covered, which result from the particular combination of the molecular protocol and the imaging

68   hardware implementation [12].

69   A key historic limitation in the widespread use of iST methods with human clinical samples

70   was the incompatibility of most methods with formalin-fixed, paraffin-embedded (FFPE) tissue

71   samples [13, 14]. FFPE is the standard format for clinical sample preservation for pathology due to

72   its ability to maintain tissue morphology and sample stability at room temperature for decades [15].

73   The ability to process FFPE samples with iST would enable the use of archival tissue banks for

74   studies and obviate the need for specialized sample harvesting workflows. However, FFPE

75   samples tend to suffer from decreased RNA integrity, particularly after having been stored in

76   archives for extended periods of time [16].

77   Three companies recently released the first FFPE compatible commercial iST platforms:

78   10X's Xenium, Nanostring's CosMx, and Vizgen's MERSCOPE [9,10,11,17]. These three platforms

79   each use different protocols, probe designs, signal amplification strategies, and computational

80   processing methods, and therefore may potentially yield different sensitivities and downstream

81   results. The main chemistry difference lies in transcript amplification: 10X Xenium uses a small

82   number of padlock probes with rolling circle amplification; CosMx uses a low number of probes

83   amplified with branch chain hybridization; and MERSCOPE uses direct probe hybridization but

84   amplifies by tiling the transcript with many probes (**Fig. 1b**). However, no head-to-head

85   performance comparisons on matched samples have been published. Understanding the key

86   differences across platforms will allow users to make better-informed decisions regarding panel

87   design, method choice, and sample selection as they design costly experiments, often on precious

88   samples that have been bio-banked for years [18].

89   In this study, we compared currently available FFPE-compatible iST platforms on matched

90   tissue samples. We prepared a set of samples representative of typical archival FFPE tissues,

91   comprised of 23 different tissue types, and acquired matched data from sequential sections

92   according to the manufacturer's best practices at the time of writing, generating a dataset of >3.3M

93   cells. We analyzed the relative sensitivity and specificity of each method on shared transcripts,

94   and further quantified the concordance of the iST data across each platform with orthogonal data

95   sets from The Cancer Genome Atlas (TCGA) program and Genotype-Tissue Expression (GTEx)

96   databases [19,20]. Then we focused on cell-level comparisons, evaluating the out-of-the-box

97   segmentation for each platform based on detected genes and transcripts and coexpression patterns

98   of known disjoint markers. Finally, we cross-compared the ability of each platform to identify cell

99   type clusters with breast and breast cancer tissues as an example use case. Taken together, our

100   work provides the first head-to-head comparison of these platforms across multiple archival

101   healthy and cancerous FFPE tissue types.

102   RESULTS

103   **Collection of matched iST data across 23 FFPE tissue types reveals high transcript counts**

104   **obtained by Xenium and CosMx.**

105   To test the performance of the latest generation of FFPE-compatible iST tools, we

106   measured the spatial expression of the same genes on the same samples as much as possible given

107   current panel configurations. To accomplish this, we used two previously generated multi-tissue

108   tissue microarrays (TMAs) from clinical discarded tissue (**see Methods**). We focused on FFPE

109   tissues as the standard method for sample processing and archival in pathology. One TMA

110   consisted of one hundred and seventy-three 0.6 mm diameter cores (i.e. sampled regions) from

111   seven different cancer types, with 3-6 patients per cancer type, and 3-6 cores per patient (**Fig. 1c,d,**

112   **Supplementary Table 1**). A separate TMA consisted of forty-eight 1.2 mm diameter cores

113   spanning sixteen normal tissue types isolated with each tissue type coming from one patient and

114   represented in 2-3 cores (**Fig. 1c,d, Supplementary Table 2**). CosMx and Xenium suggest pre-

115   screening samples based on H&E, while MERSCOPE recommends a $DV_{200} > 60\%$. Since our goal

116   was to determine the compatibility of iST platforms under typical workflows for biobanked FFPE

117   tissues, and since TMAs are challenging to assay by DV_{200}, samples were not prescreened based

118   on RNA integrity. Samples were screened by H&E in the process of TMA assembly. Both TMAs

119   were sliced into serial sections for processing by 10X Xenium, Vizgen MERSCOPE, and

120   Nanostring CosMx, following manufacturer instructions (**see Methods**).

121   The three different iST platforms offer different degrees of customizability and panel

122   compositions. In terms of panel design, MERSCOPE and Xenium offer either fully customizable

123   panels or standard panels with optional add-on genes, while CosMx offers a standard 1K

124   (substantially larger plex) panel with optional add-on genes. We opted to run the CosMx 1K panel

125   as available commercially, as well as the Xenium human breast, lung, and multi-tissue off-the-

126   shelf panels. We then designed two MERSCOPE panels to match the pre-made Xenium breast and

127   lung panels, by filtering out any genes which could potentially lead to high expression flags in any

128   tissue in the Vizgen online portal. This resulted in a total of six panels, with each panel overlapping

129   the others on >94 genes (**Fig. 1e, Supplementary Table 3**). Samples were run following

130   manufacturer instructions over the course of 74 days after slicing (**Fig. 1f, Supplementary Table**

131   **4**), with efforts made to ensure that head-to-head comparisons were available at similar time points

132   for each pair of platforms. In data review, we noticed that MERSCOPE breast and lung panel were

133   originally acquired with a 5 µm imaging depth, which was unintentionally thinner than the

134   manufacturer recommendation of 10 µm, and could thus lead to aberrantly low counts. Thus, a

135   second round of breast panel acquisition was performed with a 10 µm imaging depth

136   (**Supplementary Table 1a**), resulting in a median 3.0-fold increase in expression across all

137   transcripts. We excluded the 5 µm MERSCOPE breast panel data from all further comparisons but

138   left the lung panel data in as an illustrative example of an unsuccessful run. However, we

139   emphasize that MERSCOPE performance should be judged based on the rerun breast panel.

140   Each data set was processed according to the standard base-calling and segmentation

141   pipeline provided by each manufacturer. The resulting count matrices and detected transcripts were

142   then subsampled and aggregated to individual cores of the TMA (**Methods**). Across all datasets

143   we generated >190 million transcripts, >3.3 million cells, across 7 tumor types, and 16 normal

144   tissue types. Overall, we found that the cores from each TMA were generally well adhered to the

145   tissue and detected transcripts, and we were able to collect data from all three modalities for 217

146   cores (**Supplementary Table 4**). The total number of transcripts recovered for each run was

147   highest for Xenium, followed closely by CosMx, and then MERSCOPE (**Supplementary Table**

148   **4**). The total number of cells initially reported was highest for Xenium followed by MERSCOPE

149   and CosMx (**Supplementary Table 4**). Based on the initially reported number of transcripts, the

150   tumor TMA appeared to provide more counts than the normal tissue TMA, which we ascribed to

151   a higher tissue quality in the tumor samples (**Supplementary Table 4**). We note that the total

152   number of transcripts from the MERSCOPE normal TMA run was below what would be typically

153   thought of as a successful run, even when rerunning with the breast panel at 10 μm imaging depth.

154   Such a sample would normally be excluded from analysis, but we continued the data through to

155   illustrate how low transcript capture affects downstream results.
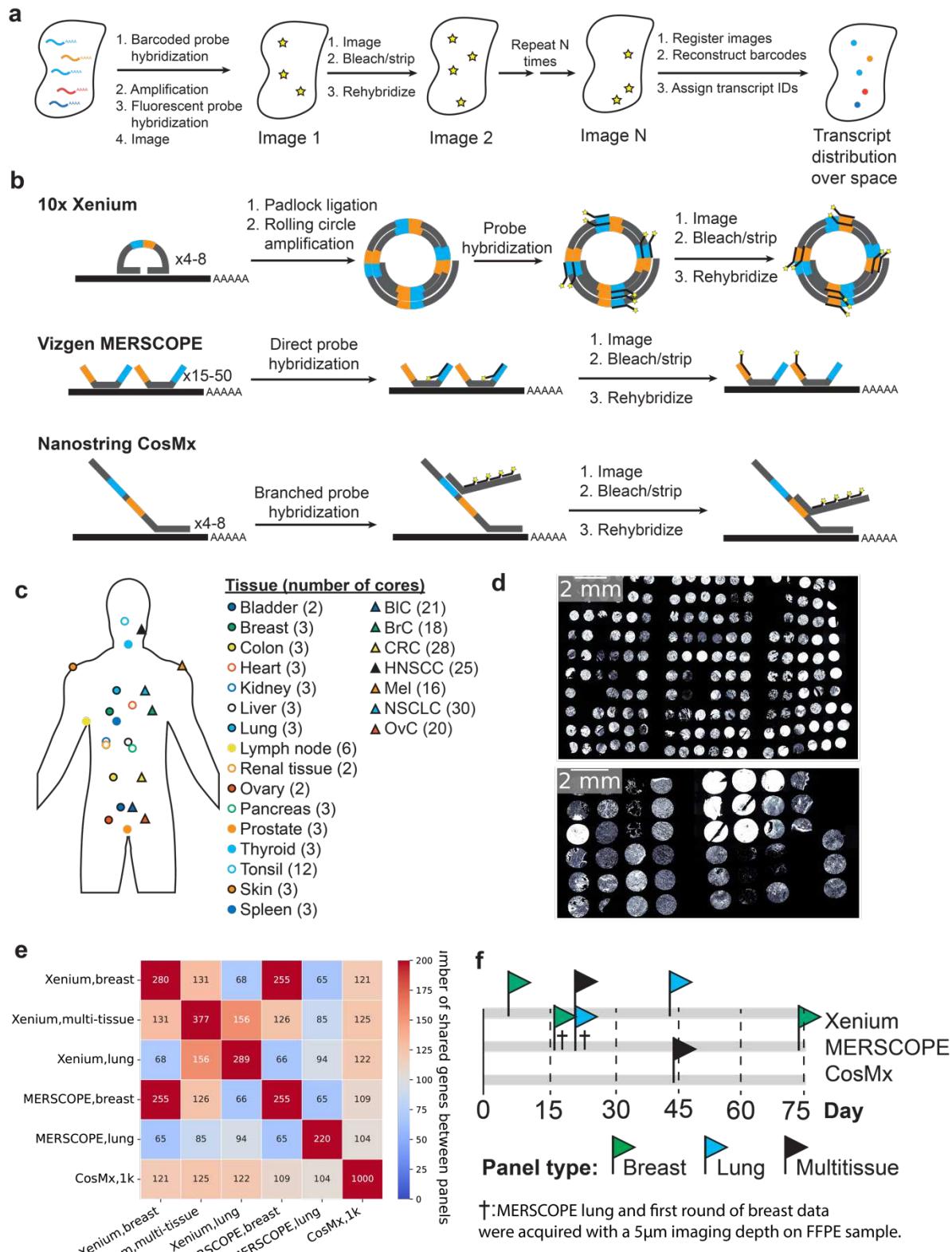
Figure 1: Experimental design and iST platforms. (a) Overall approach for generating iST data. (b) Different amplification approaches for Xenium, MERSCOPE, and CosMx. (c) Overview of

159   the tissue types and numbers of cores used in this study. BlC = bladder cancer, BrC = breast cancer,

160   CRC = colorectal cancer, HNSCC = head and neck squamous cell carcinoma, Mel = Melanoma,

161   NSCLC = non-small cell lung cancer, OvC = ovarian cancer. (d) DAPI images from the Xenium

162   run of each TMA, including tumors (top) and normal tissues (bottom) (e) The number of common

163   target genes in each panel used in this study. (f) Overall timeline of the imaging days for each

164   study. Day = 0 corresponds to the day of slicing. † denotes the MERSCOPE breast and lung panels

165   acquired with a 5 μm imaging thickness, thinner than manufacturer instructions.

166

167   10X Xenium shows higher transcript counts per gene without sacrificing specificity

168   We next sought to directly compare the performance of each iST platform on matched

169   genes. We began with a pseudo-bulk-based approach at the core level since this would not depend

170   on differences in cell segmentation performance (see **Methods**) [21].

171   First, we examined the run-to-run reproducibility within a single platform for Xenium and

172   MERSCOPE, finding that the total transcript count of all shared genes was highly correlated across

173   data sets acquired with different panels, regardless of the tissue of origin (**Supplementary Fig.**

174   **1a**). We also examined the pseudo-bulk gene expression correlation for cores from the same patient

175   in the same dataset and found that correlation was high (Pearson's r => 0.7) in almost all cases

176   (**Supplementary Fig. 1b-c**), indicating good sample-to-sample reproducibility within a given

177   platform.

178   To evaluate the relative sensitivity of each platform, we plotted the total transcript counts

179   of every shared gene between all combinations of platform and panel, summed across all matched

180   cores. We found generally linear relationships between all pairs of platforms (**Fig. 2a-c,**

181   **Supplementary Fig. 2**). Xenium consistently showed higher expression levels on the same genes

182   than CosMx in the tumor TMA, with the Xenium breast having 14.6-fold more counts than the

183   CosMx multi-tissue data sets (**Fig. 2a**). The Xenium multi-tissue panel data showed a slightly

184   smaller difference, with 12.3-fold higher expression on the same genes (**Fig. 2a**), while the lung

185   panel, which was acquired closest in time following slicing, also displayed a median of 14.0-fold

186   higher expression **(Fig. 2a)**. MERSCOPE showed higher expression levels than CosMx when

187   using the breast (10 μm) panel (median of 5.4-fold higher), and comparable expression levels even

188   when using the lung (5 μm) panel (median of 1.1-fold) **(Fig. 2b)**. Finally, Xenium showed 2.6-

189   fold higher median expression with the breast panels (10 μm) than MERSCOPE, and 13.6-fold

190   higher median expression with the lung panels (5 μm) **(Fig. 2c)**. In the normal tissue TMA, we

191   found that results were generally consistent, except that the MERSCOPE breast panel showed

192   decreased transcript counts relative to the same panel in the tumor TMA (**Supplementary Fig. 2b-**

193   **c**), which is consistent with this TMA being unsuccessful for MERSCOPE. Considering the overall

194   higher transcripts per cell across platforms for the tumor TMA **(Supplementary Table 4)**, this

195   suggests that the ability to detect transcripts falls off more strongly with sample quality with

196   MERSCOPE, altering the performance relative to CosMx but not Xenium. Examining the CosMx

197   as compared to Xenium data also revealed an upward curve in the lower expression regime

198   indicative of higher-than-expected calls associated with the low expression regime by CosMx

199   (**Supplementary Fig. 2a**).

200   We next wanted to assess the specificity of each platform. Each of the three platforms

201   includes negative controls which are used to evaluate sample quality [22, 23]. Xenium and CosMx

202   include both negative probes (e.g. real probes targeting nucleic acids that are not present in human

203   tissue) and negative barcodes (e.g. algorithmically allowable barcodes that are not associated with

204   any probe in the experimental panel). MERSCOPE includes only negative barcodes by default. To

205   determine specificity, we first calculated the fraction of negative barcodes and probes relative to

206   the number of transcripts for each tissue type **(Fig. 2e).** We found that MERSCOPE and Xenium

207   consistently showed the highest on-target fraction, while CosMx was lower across each tissue type

208   (**Fig. 2e**). However, this measurement is biased because of the relative numbers of controls and

209   target barcodes. We therefore also adopted a false discovery rate (FDR) calculation which

210   normalizes for these differences and is calculated against both the negative probes and negative

211   barcodes (see **Methods, Fig. 2f-g**). We found that Xenium consistently showed the lowest FDR

212   while CosMx showed the highest FDR regardless of whether we standardized to negative control

213   barcodes or probes. This finding is consistent with the upswing in the gene-gene expression plots

214   in **Fig 2a** and **Supplementary Fig. 2a**—as both indicate a higher FDR at the low end of the gene

215   expression range. These results are consistent when visualized across panel **(Fig. 2e-g)**.

216   Finally, we used the negative control barcodes to evaluate the number of genes reliably

217   detected by each platform in each tissue type. For each core, we calculated the number of genes

218   that were detected two standard deviations above the average expression of the negative control

219   probes. These numbers were then averaged for cores of the same tissue type. Because the CosMx

220   panel was almost three times larger, it yielded a larger absolute number of detected genes in 14

221   out of 20 tissue types while the Xenium breast panel was higher in the remaining 6 tissue types

222   (**Fig. 2h**, **Supplementary Table 5**). However, Xenium consistently detected the highest fraction

223   of genes in a panel, followed by MERSCOPE and CosMx (**Fig. 2i**).


224
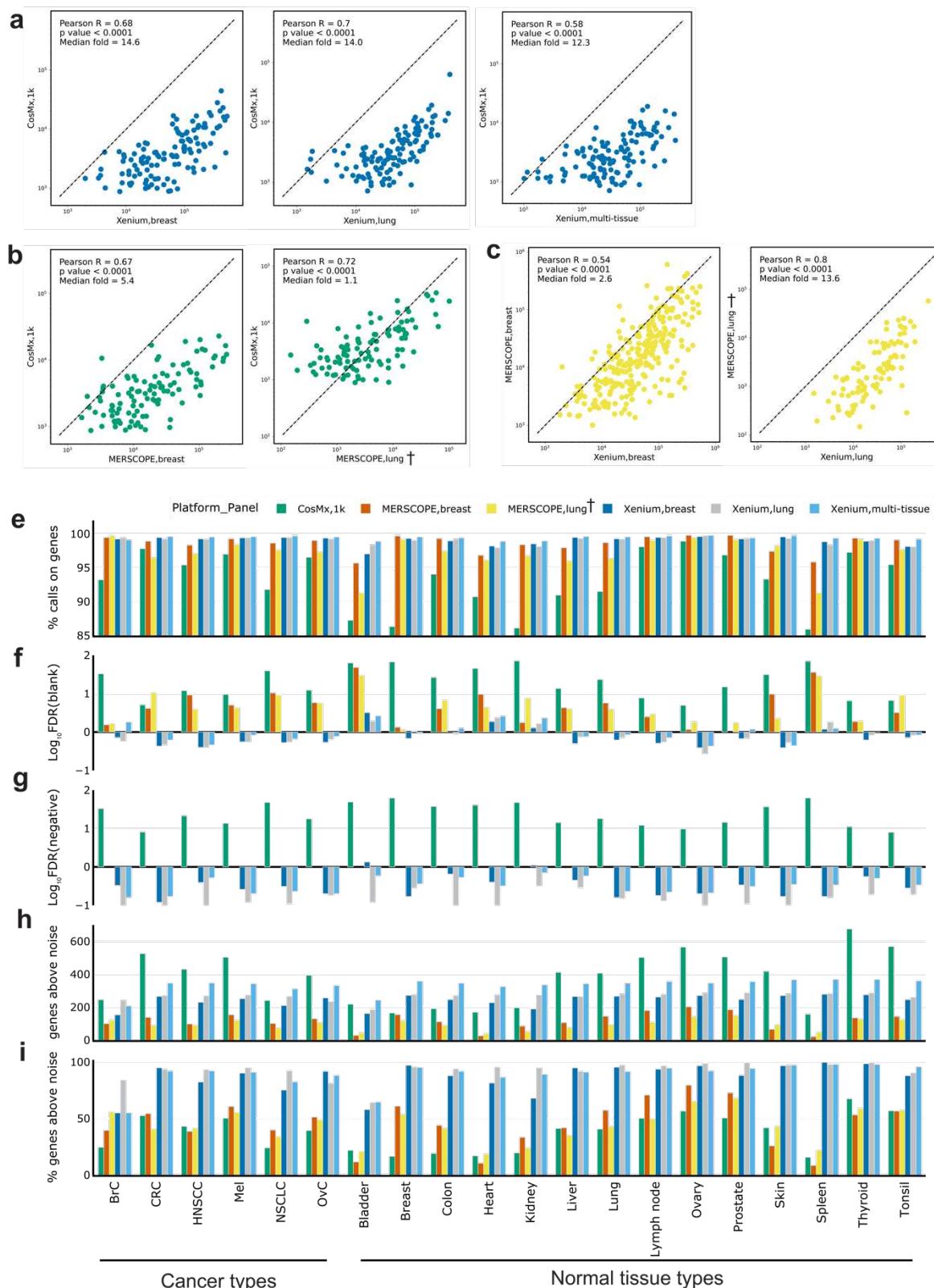
Figure 2: Technical performance comparison of iST platforms grouped by tissue types. (a)
Scatter plots of summed gene expression levels (natural log transformed) of every shared gene
between Xenium (breast/lung) and CosMx (1k) data, captured from matched tumor TMA cores.
Each data point corresponds to a gene. **(b)** Same as (a) but between MERSCOPE (breast/lung) and
CosMx(1k). **(c)** Same as (a) but between Xenium(breast/lung) and MERSCOPE(breast/lung). **(d)**
Same as (a) but between Xenium(multi-tissue) and CosMx(1k). **(e)** Bar plot of percentage of all
transcripts corresponding to genes relative to the total number of calls (including negative control
probes and unused barcodes) averaged across cores of the same tissue type. Results are presented
by panel including breast, lung, and multi-tissue panels from Xenium; breast and lung panels from
MERSCOPE; and multi-tissue 1k panel from CosMx. **(f)** Bar plot of false discovery rate
(FDR) where FDR(%) = (blank barcode calls / total transcript calls) x (Number of panel genes
/Number of blank barcode) x 100. FDR values were $\log_{10}$ transformed to better show the
differences between panels. **(g)** Same as (f) but using negative control probes to replace blank
barcodes. MERSCOPE is missing in this bar plot as it does not have negative control probe by
design. **(h)** Bar plot of number of genes detected above noise, estimated as two standard deviations
above average of the negative control probes. **(i)** Same as (h) but normalized to the number of
genes in a panel. † denotes the MERSCOPE lung panel acquired with a 5 μm imaging thickness.

iST platforms are all concordant with orthogonal RNA-seq data sets

In the absence of ground truth, it is difficult to evaluate whether a higher number of

expressed genes is representative of increased sensitivity to real biology or increased false positive

rates. We thus evaluated the correlation of iST data to reference RNA-seq data. We first aggregated

pseudo-bulk normal tissue TMA results from all panels of the three platforms and compared them

to data from the TCGA program **(see Methods)** [11]. We observed similar correlation coefficients

across all gene panels relative to pseudo-bulk RNA-seq expression data (**Fig. 3a-b,**

Supplementary Table 8). However, notably, the CosMx data showed a characteristic upswing in

the low expression regime, similar to that observed when plotting gene-by-gene expression against

MERSCOPE and Xenium (**Fig. 2, Supplementary Fig. 2**).

We also compared the pseudo-bulk results from the normal tissue TMA with bulk RNA-

seq data obtained from GTEx[12] The Xenium breast, Xenium multi-tissue, and CosMx data sets

showed similar correlations to breast data obtained from GTEx, while the MERSCOPE had

256   significantly lower correlation, consistent with a run which doesn't pass QC (Pearson's r of 0.36

257   vs 0.71, 0.68, and 0.69, respectively, **Fig. 3c**). Similar trends were observed in the lung data, with

258   MERSCOPE lung (5 μm) showing the lowest correlation while the other three data sets showed

259   higher correlations to GTEx data (**Fig. 3d**). These relative trends remained true across most normal

260   tissue types, though we found that thyroid, pancreas, and lymph nodes showed the lowest

261   correlations across all panels while prostate, tonsil, and liver showed the highest correlations

262   (**Supplementary Table 9**). Overall, our comparison to TCGA and GTEx data suggests that while

263   some platforms may be more highly correlated to reference datasets in some cases, all are within

264   a similar correlation regime regardless of tissue type.

265   We next wanted to determine how the expression of tissue-specific transcript markers

266   varied across each platform. To accomplish this, we curated tissue markers that are unique to each

267   tissue type by selecting genes whose expression in a single tissue exceeds 20 times the sum of

268   other tissues from the GTEx database (see **Methods**). We found tissue-specific expression patterns

269   of several of these markers across all selected panels when visualized across each healthy tissue

270   type (**Fig. 3e**). MERSCOPE consistently showed expression of tissue-specific markers in multiple

271   tissue types, consistent with unsuccessful normal TMA data acquisition. However, some canonical

272   markers were not enriched in certain tissues. For example, we found that ABCC11 was highly

273   expressed in the breast tissue in the Xenium and MERSCOPE datasets, but not in CosMx. In fact,

274   although CosMx showed satisfying expression patterns for some tissue markers, many canonical

275   markers are not enriched in the expected tissues, possibly due to the high false discovery rate (**Fig.**

276   **3f-g**). Across marker genes, Xenium data had a distinct expression pattern in all tissues, whereas

277   CosMx and MERSCOPE showed a less distinct pattern in many tissue types.
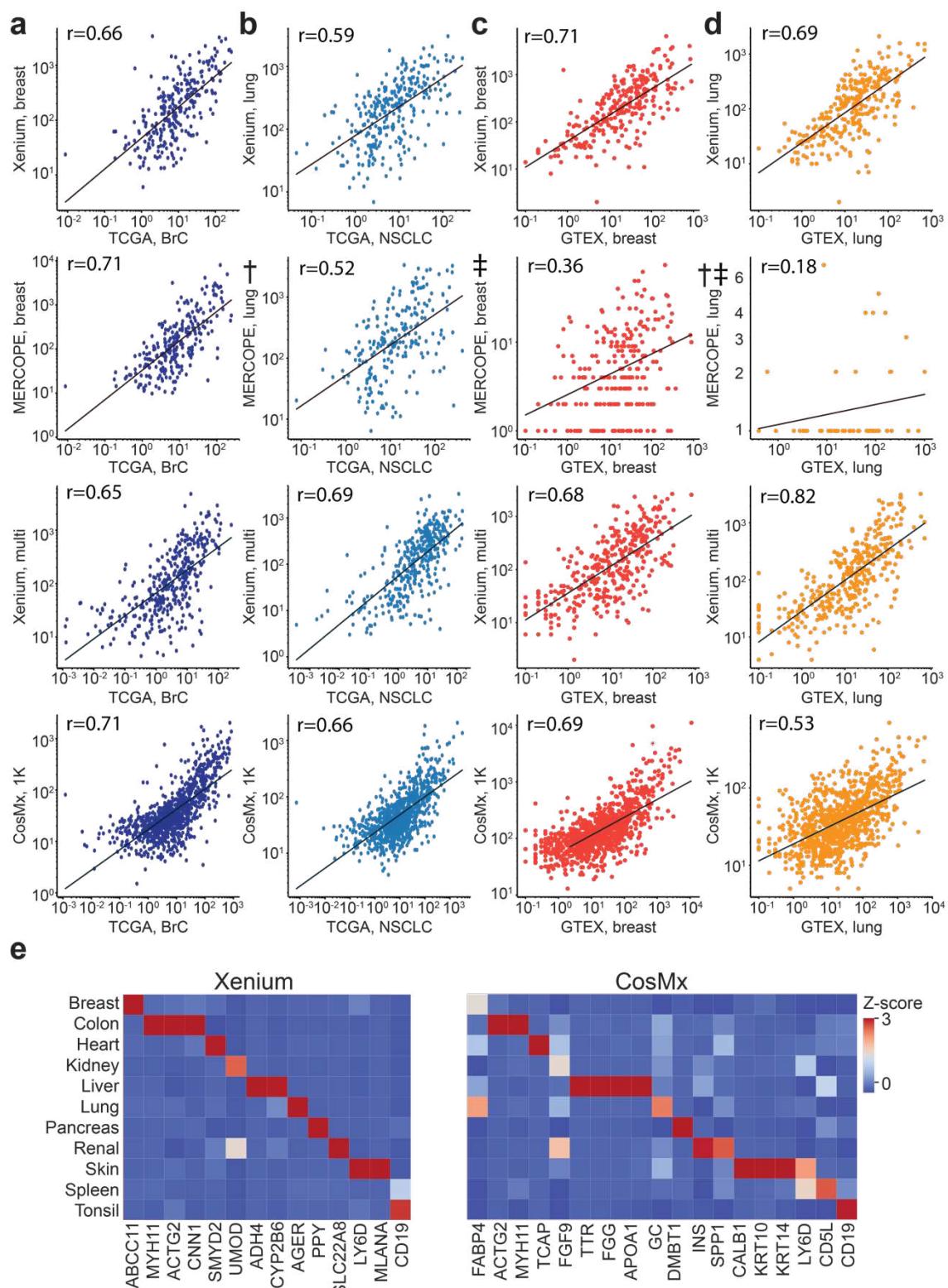
Figure 3: Concordance of iST data with reference RNA-seq datasets. (a) Scatter plots of overlapping genes, showing the averaged expression of a gene across breast cancer cores profiled by the indicated panel, normalized to 100,000 vs the average FPKM from TCGA for all samples

282 of a matched tissue type (BRCA). **(b)** Same as (a) but for lung cancer cores plotted vs averaged
283 LUAD and LUSC samples from TCGA. **(c)** Same as (a) but showing breast cores vs averaged
284 nTPM values from GTEx breast samples. **(d)** Same as (a) but for lung cores and samples. † denotes
285 the MERSCOPE lung data acquired with a 5-µm imaging depth on FFPE sample. ‡ denotes the
286 normal tissue TMA data of MERSCOPE which failed initial QC. **(e)** Heatmap of Z-scored average
287 gene expression for several canonical marker genes in the indicated tissue cores for the Xenium
288 multi-tissue panel (left) and CosMx 1K panel (right).
289
290 **Out of the box segmentation and filtration can yield cells with comparable numbers of**

291 **detected transcripts and genes from each platform.**

292   Next, we compared the performance of each iST method on a single-cell level. The three

293 platforms generate cell boundaries based on a DAPI image alone (Xenium) or a DAPI image

294 combined with a membrane marker (CosMx and MERSCOPE). When we visually examined the

295 segmentation outputs, Xenium data showed cell boundaries that appeared to include large regions

296 of non-cellular space, in contrast to MERSCOPE and CosMx which tightly followed the visualized

297 cell nucleus (**Fig. 4a**). When transcripts were overlaid with these segmentation boundaries,

298 Xenium cell boundaries fell between regions of transcripts and thus most transcripts were assigned

299 to cells. MERSCOPE and CosMx's tighter nuclei removed more transcripts, though those that

300 remain appeared more confidently assigned to cells. Overall, when normalized to the imaged tissue

301 area, Xenium and CosMx identified the most putative cells, followed by MERSCOPE (**Fig. 4b,**

302 **Supplementary Table 6**). In line with the visual inspection, Xenium cells were consistently larger,

303 regardless of data set or panel, followed by CosMx and finally MERSCOPE (**Fig. 4c,**

304 **Supplementary Table 7**).

305   We filtered out empty regions of space and cells without any transcripts for downstream

306 examination and quantified the fraction of cells containing differing numbers of transcripts per cell

307 **(Fig. 4d).** We chose a permissive threshold of removing cells with fewer than 10 transcripts for

308 Xenium and MERSCOPE, and 20 transcripts for CosMx from downstream analysis. [11, 24, 25]. The

309 tumor TMA consistently had a greater fraction of cells passing filtration, with Xenium retaining

310 the most cells (97.43% breast, 97.10% multi-tissue, 95.08% lung) followed by CosMx (83.41%)

311 and MERSCOPE (68.46% breast (10 μm), 25.77% lung (5 μm) (**Supplementary Table 3**). The

312 normal tissue TMA had overall lower cell retention performance, but the relative performance of

313 the platforms based on the fractions of cells remained the same. Notably, while CosMx and

314 Xenium still retained > 77% of the cells, MERSCOPE data of the normal tissue TMA had <3% of

315 cells retained and was thus not used in downstream analysis. Unsurprisingly, filtration decreased

316 the number of retained cells per unit area for all platforms, with the smallest decrease coming for

317 CosMx. The cells retained from CosMx had similar areas, while filtration of the Xenium and

318 MERSCOPE data sets resulted in a higher average cell area (**Fig. 4c**).

319 After filtration, we compared the number of transcripts and the number of unique genes

320 per retained cell across all tissues and all panels, focusing only on cores that were sampled by all

321 three platforms **(Fig. 4e-f)**. Xenium breast panel gave the highest numbers of transcripts per cell

322 in most tissue types, 17 out of 22. The CosMx data showed the highest numbers of transcripts in

323 heart, lymph node, spleen, thyroid, and tonsil; and comparable transcript counts in breast cancer,

324 ovary, and ovarian cancer to the Xenium breast panel. The MERSCOPE data generally had the

325 lowest number of transcripts per cell, though bladder cancer and breast cancer measured with the

326 MERSCOPE breast panels approached the results from Xenium, and the bladder cancer and skin

327 data sets had higher transcripts per cell than CosMx. As expected given its larger panel size,

328 CosMx found many unique genes per cell, showing the largest numbers in 9 tissue types: breast

329 cancer, colon, heart, lymph node, ovary, ovarian cancer, spleen, thyroid, and tonsil; while Xenium

330 breast panel found the most unique genes per cell in 12 tissue types: bladder cancer, bladder, breast,

331 CRC, HNSCC, kidney, lung, melanoma, NSCLC, pancreas, prostate, and skin (**Fig. 4f**). If these

332   analyses were restricted to only the shared genes across all panels, numbers were much lower, but

333   Xenium showed higher expression levels and unique numbers of genes than either CosMx or

334   MERSCOPE (**Supplementary Fig 3c-d**).

335   We then wanted to determine how effectively different iST platforms' segmentation

336   algorithms perform. We examined the co-expression of *CD19*, a canonical B-cell marker, and

337   *CD3e*, a canonical T-cell marker across all filtered cells; the co-expression of *CD8* and *CD4*,

338   markers of T-cell subsets; and the co-expression of *CD3e* and *EPCAM*, a marker for epithelial

339   cancer cells [26, 27]. All these marker gene pairs are disjointly expressed, and a well-performing

340   segmentation algorithm should yield few cells expressing both markers. We pooled all the filtered

341   cells from matched cores and all available panels of each platform and plotted the expression of

342   one gene against the other and converted the scatter plot to a heatmap to show cell fractions. We

343   found that Xenium—despite its less visually accurate cell boundaries—and MERSCOPE, showed

344   clear patterns of disjoint expression, separating cells from different lineages, while CosMx showed

345   such a pattern for *EPCAM* vs *CD3e* but not for the other two pairs (**Fig. 4g**). Given the low counts

346   of the immune genes, it was difficult to determine if these were false positive calls or segmentation

347   errors. Nevertheless, since the CosMx panel is much higher plex, and retained similar numbers of

348   transcripts and genes to Xenium, we wondered how these two methods performed in terms of cell

349   type recovery.
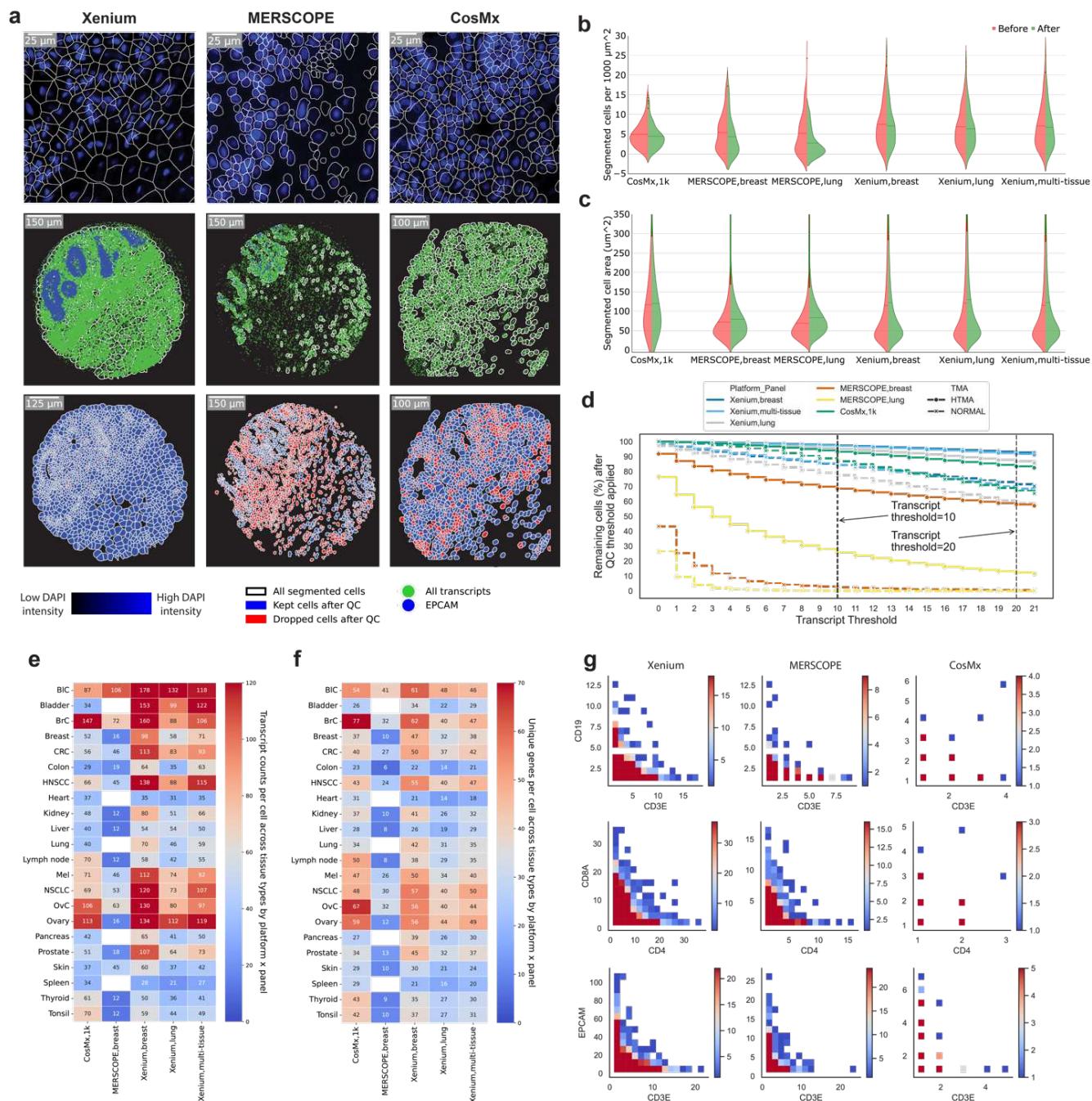
Figure 4: Comparison of cell segmentation results from each iST platform. (a) Top row: DAPI image overlaid with cell segmentation boundaries (subset). Middle row: all the transcripts in green dots, white lines for the cell boundaries, and *EPCAM* in blue dots. Bottom row: segmented cell boundaries before and after filtration. **(b)** Violin plot of segmented cells per unit area before (left half) and after filtration (right half) grouped by panel with tumor and normal TMA data combined. **(c)** Same as (b) but showing cell areas before and after filtration. **(d)** Line plot showing remaining cells in percentage after filtering with various thresholds (transcripts per cell). Dotted lines indicate selected thresholds: 10 transcripts or above for Xenium and MERSCOPE and 20 for CosMx. **(e)** Heatmap of transcripts per cell after filtration. All available genes are considered here for each

360 panel. MERSCOPE lung panel (5 μm) excluded from this heatmap. **(f)** Same as (e) but showing
361 unique genes per cell. **(g)** Co-expression density map for three pairs of disjoint genes (rows) from
362 all three platforms (columns). All cells across all tissues which include at least one detected
363 transcript of either of the indicated genes are plotted together, with color indicating the number of
364 cells at the indicated expression levels of each gene.
365

366

367 **Clustering analyses reveal differences in cell type recovery across platforms**

368 In a typical iST workflow, a key step is reducing the dimensionality of the data by

369 identifying cell types, their unique states, and their expression patterns for further analysis

370 leveraging spatial information[28]. To compare across platforms, we clustered the data from the

371 filtered cells from all the cores for each TMA with a focus on breast tissues. The initial clustering

372 of TMAs from datasets (except MERSCOPE normal tissue) showed expected batch effects caused

373 by patients and tissue types with broadly similar cluster arrangements around morphological tissue

374 features (**Supplementary Fig. 4a-d**). We removed batch effects (**see Methods**) and then

375 performed targeted clustering and cell type annotation for breast samples from the CosMx and

376 Xenium breast datasets; lung samples from the CosMx and Xenium lung datasets; and breast

377 cancer from the CosMx, MERSCOPE breast, and Xenium breast datasets.

378 In breast samples, we were able to identify nine cell types, including all known major cell

379 types, (adipocytes, alveolar cells, B cells, basal cells, fibroblast cells, hormone-sensing cells,

380 myeloid, T cells, vascular & lymphatic cells) from the Xenium data, using previously established

381 markers (**Fig. 5a, Supplementary Fig. 5**) [29,30,31]. In the CosMx data, we were only able to identify

382 six cell types, including several major cell types, but failed to recognize cell subtypes (B cells,

383 basal cells, fibroblast, hormone-sensing cells, immune cells, and vascular & lymphatic cells) (**Fig.**

384 **5a, Supplementary Fig. 5**). A high gene-to-gene correlation was found between all overlapping

385 cell types between Xenium and CosMx (**Fig. 5a**). Similarly, in the lung samples, we were able to

386   identify nine cell types (alveolar epithelial type 1 cell, alveolar epithelial type 2 cell, endothelial

387   capillary cells, endothelial cells, fibroblasts, immune cells, macrophages, mast cells, and stroma

388   cells) in the Xenium lung panel, successfully covering all known major cell types (**Fig. 5b,**

389   **Supplementary Fig. 5**) [32,33]. Four of the six major clusters were identified and annotated in the

390   lung samples from the CosMx data (endothelial cells, epithelial cells, immune cells, and

391   macrophages), while the other two clusters remained difficult to annotate due to the non-traditional

392   enriched gene markers (**Fig. 5b, Supplementary Fig. 5**). Correlation heatmaps show a strong

393   correlation between the two macrophage clusters identified in Xenium and CosMx (**Fig. 5b**). The

394   epithelial cell cluster from CosMx correlates strongly with alveolar epithelial type 1 cell from

395   Xenium and the endothelial cell cluster from CosMx correlates with endothelial capillary cell

396   cluster from Xenium (**Fig. 5b**).

397   Finally, in breast cancer, after batch effect removal (**Supplementary Fig. 5d-f**), Xenium

398   resulted in nine cell types (alveolar cells, B cells, basal cells, fibroblast, hormone-sensing cells,

399   immune cells, myeloid, T cells, and vascular & lymphatic cells) (**Fig. 5c, Supplementary Fig. 5**)

400   [34,35,36]. On the other hand, CosMx resulted in eight cell types (alveolar cells, basal cells, epithelial

401   cells, fibroblast cells, hormone-sensing cells, immune cells, myeloid, and vascular & lymphatic

402   cells). MERSCOPE resulted in six cell types, including alveolar cells, fibroblast cells, hormone-

403   sensing cells, myeloid cells, T cells, and vascular & lymphatic cells. The cell type annotation of

404   Xenium and CosMx is comparable in terms of both transcriptomic profile and subtype depth, with

405   CosMx only unable to annotate immune cell subtypes (B cell and T cell). Gene expression of the

406   same cell type from both platforms correlated well (**Fig. 5c, Supplementary 5**). The cell type

407   annotation of CosMx, however, was especially difficult compared to Xenium because of its

408   atypical gene markers shown for each cluster in the heatmaps (**Supplementary Fig. 5**) and low

409   expression of transcripts from canonical markers (**Supplementary Fig. 5g-h**). MERSCOPE, on

410   the other hand, identified most, but not all, the cell types recognized by Xenium and CosMx,

411   including alveolar cells, fibroblast cells, hormone-sensing cells, myeloid, T cells, and vascular &

412   lymphatic cells. MERSCOPE and Xenium showed a high correlation for almost all matching

413   clusters. The correlation map shows a clearer one-to-one mapping between MERSCOPE and

414   Xenium clusters than Xenium and CosMx clusters.


415
416
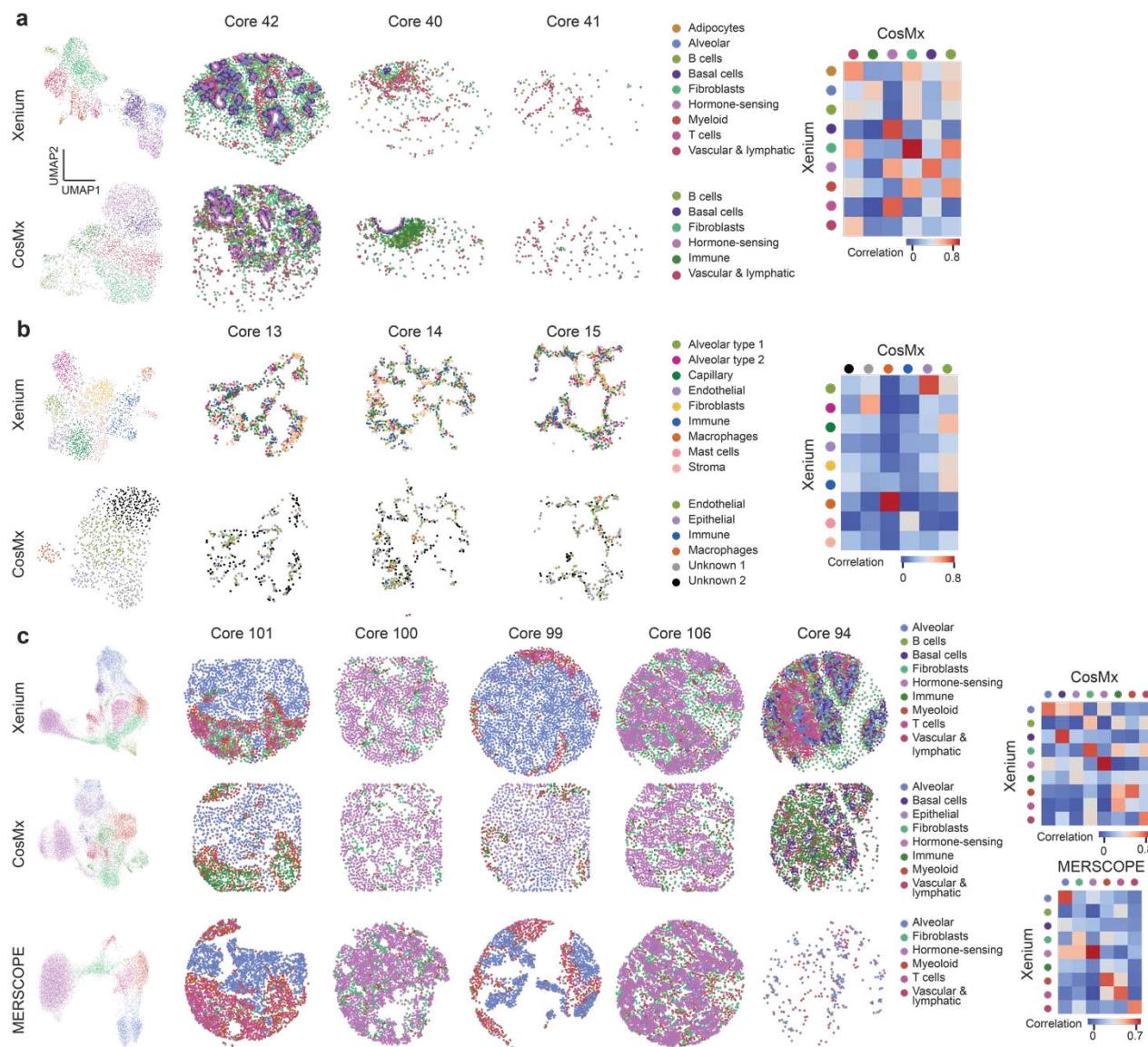
417 **Figure 5: Cell type recovery performance across technology. (a)** Clustering results of breast
418 samples in normal TMA from Xenium breast (bottom left) and CosMx multi-tissue 1k (top).
419 Correlation plot showing the correlation between cell types identified (bottom right). **(b)**
420 Clustering results of lung samples in normal TMA from Xenium lung (bottom left) and CosMx
421 multi-tissue 1k (top). Correlation plot showing the correlation between cell types identified
422 (bottom right). **(c)** Clustering results of breast cancer samples in tumor TMA from Xenium breast
423 (bottom left), MERFISH (bottom right) and CosMx multi-tissue 1k (top). Correlation plot showing
424 the correlation between cell types identified in CosMx and Xenium (left), MERFISH and Xenium
425 (right).
426

DISCUSSION

In this study, we compared data obtained with three commercially available iST platforms with archival FFPE tissues to assess overall technical performance and help guide experimental design with human samples that represent an important use case of these platforms. We focused our analyses on technical performance as a function of tissue type, including 7 different tumor types and 16 normal tissue types. Overall, we found that each iST platform presented various tradeoffs in terms of implementation, panel design and panel options, and resulting total transcript quantification and downstream analyses, including cell segmentation, cell quality, and biological interpretation. All these factors must be considered when designing iST experiments.

There are significant workflow differences between the different platforms which factor into the choice of method. Cutting samples onto MERSCOPE coverslip is more difficult than on standard microscope slides. The total hands-on time for running a slide on Xenium is 2-3 days compared to 5-7 days for MERSCOPE and 2 days on CosMx. We found that MERSCOPE and CosMx are well set up for batch processing in the wet lab, either due to built-in pause points or the instrument's ability to run multiple samples. Xenium is limited for batch processing by a need for a separate thermocycler for each slide pair processed in parallel. After staining, selecting regions of interest (ROIs) presented a surprising challenge for some systems: the Xenium platform could readily image the entire slide as a single ROI which easily covered entire TMAs, but the MERSCOPE ran into a $1cm^2$ imaging area limit which meant cores in the addressable region were left unimaged, while the CosMx workflow required a demanding manual selection of ROIs for each core. These factors are likely to change as each company updates its protocol, but currently, Xenium offers the shortest and least hands-on workflow.

449   From a technical perspective, we analyzed each resulting dataset with a combination of

450   manufacturer recommended processes for each platform and computational tools that can be

451   implemented by the user downstream. These pipelines each result in count matrices and detected

452   transcripts that can be analyzed using a whole suite of emerging tools. For our purposes, when

453   analyzed at a core level to abrogate the effects of individual cell-segmentation performance, we

454   found that the total number of transcripts varied substantially across iST platform, with Xenium

455   yielding the highest number of transcripts captured followed by CosMx. Indeed, this trend held

456   when normalized for the number of cores imaged and on a per cell basis per area. When this

457   analysis was also restricted to shared genes, we also found that Xenium consistently had higher

458   expression levels across each tissue type, with no clear differences between performance on either

459   tumor or normal tissue.

460   Using a pseudo-bulk approach, again at the core level, we assessed overall correlation,

461   reproducibility, and sensitivity of each platform. We found high correlation between replicates of

462   the same patient, suggesting that there is high reproducibility across technical replicates on each

463   platform. This is important to consider since cost or input material availability can be prohibitive

464   to implementing experimental designs that leverage technical replicates—though additional tissue

465   may still be valuable for powering cell-cell interaction analysis. We additionally found high

466   correlation on a gene-by-gene basis between MERSCOPE and Xenium platforms. Xenium and

467   MERSCOPE also showed consistently high specificity across tissue types. CosMx displayed a

468   characteristic upward curve when compared to MERSCOPE or Xenium on a gene-by-gene basis,

469   indicating more frequent calls in the lower expression regime. This, coupled with the lower

470   specificity across several tissues for CosMx and the high false discovery rate, suggest that CosMx

471   is prone to errors in calling lowly expressed genes. Finally, Xenium had the highest sensitivity

472   across tissues. CosMx and MERSCOPE both detected fewer transcripts than Xenium.

473   MERSCOPE outperformed CosMx in the higher quality (as judged by relative performance across

474   all platforms) tumor TMA but underperformed it in lower quality normal tissue TMA. In general,

475   our analyses also suggest similar performance within a given platform across a vast array of tissue

476   types assayed here. We note that given the small number of replicates from each tissue, particularly

477   in the normal tissue, we stop short of making blanket statements about relative performance across

478   a particular tissue type. The data suggests, instead, that Xenium and MERSCOPE provide more

479   reliable true-positive signals of lowly expressed genes and that Xenium's overall performance is

480   less dependent on sample input quality than the other two platforms. MERSCOPE, especially,

481   appears to be particularly sensitive to sample input, highlighting the importance of prescreening

482   RNA integrity according to manufacturer instructions.

483   When we compared each dataset to existing RNA-seq datasets, we found comparable

484   correlation of pseudo-bulk data to RNA-seq data from GTEx or the TCGA across each panel and

485   platform. However, the presence of a characteristic upswing for CosMx, even when comparing to

486   orthogonal data, further shows that there is a higher false positive rate for lower expression level

487   genes in CosMx data. This upswing could be explained by the absence of probing genes in a

488   particular tissue in a larger panel. However, the Xenium multi tissue panel also includes genes not

489   expressed in breast and lung but does not show a similar upswing. Thus, a more likely

490   interpretation is that the CosMx is prone to a high FDR at the low expression regime. This could

491   also suggest that the CosMx transcript counts and detected gene numbers may be slightly inflated

492   by false discoveries.

493   From a tissue-specific expression perspective, Xenium showed a distinct expression

494   pattern of key tissue-markers, whereas CosMx and MERSCOPE did not. Additionally,

495   MERSCOPE and CosMx consistently showed expression of known tissue markers in unexpected

496   tissue types. This could be partly explained by an overall low performance on this particular normal

497   tissue TMA for MERSCOPE due to RNA quality. This performance difference could be

498   problematic for studies that are designed to compare tissue-specific factors. For studies whose

499   main biological variable of interest are within the same tissue, factors like sensitivity, specificity,

500   and panel availability may be a more important guide for iST experimental design.

501   A significant advantage of spatial transcriptomics data is the ability to map expression in

502   single cells. We compared each platform on a cell-level basis by assessing cell identification and

503   cell clustering. Overall, it appears that the out of the box segmentation from Xenium performs

504   poorly in terms of drawing cell boundaries specific to a single cell, while MERSCOPE and Xenium

505   much more closely match cell boundaries. This did not appear to differ on a tissue-by-tissue basis,

506   thus, is likely inherent to the overall approach used by each platform. After applying an expression

507   level filter, Xenium overall retained the highest number of cells across various filtering

508   stringencies. Despite Xenium's cell boundaries not clearly matching nuclei, both it and

509   MERSCOPE were able to effectively separate cells from different lineage markers, as judged by

510   finding minimal coexpression of disjoint markers, while CosMx showed more double positive

511   cells (out of, it should be noted, fewer cells expressing the target genes overall).

512   To determine whether clearer identification of lineage markers resulted in improved ability

513   to identify cell types, we performed clustering analyses specifically in the breast tissue and breast

514   cancer samples. We note that we used the full panel, not only the shared genes, when performing

515   these clustering analyses. Xenium allowed for identification of all major cell lineages in the breast

516   when compared to several reference breast atlases. Both the global and tissue clustering results

517   show that CosMx is also able to recognize the major cell types, but cannot identify cell subtypes.

518 Additionally, since the cluster-enriched genes do not correspond to well-known markers, probably

519 due to the low expression caused by low sensitivity and specificity, cell type annotation was

520 particularly difficult. Lastly, despite lower transcript counts and fewer cells, MERFISH still

521 successfully identified cell groups, capturing the patterns seen in other platforms. These

522 differences in cell typing, can also be attributed to the differential performance of cell segmentation

523 pipelines [37,38 39,40]. Since all platforms provide the underlying DAPI stain and morphology images

524 (in the case of CosMx and MERSCOPE) it is likely that segmentation performance could be

525 improved on a sample-by-sample or tissue-type-by tissue-type basis. Future work should seek to

526 assess cell segmentation tools and their performance across data from each platform to help inform

527 the choice of analytical method where needed.

528   Plex is an important factor in ST experiments which we have not explicitly considered. The

529 kinds of questions that may be answered by a 1,000-plex panel are clearly different than those

530 answered by a 300-plex panel, offering more opportunities to explore intra- and intercellular

531 signaling interactions. Thus, we note that for the right question, the higher false positive rates and

532 lower sensitivities of CosMx relative to Xenium could be tolerated for a broader coverage of the

533 biology. On the other hand, the fully configurable nature of Xenium and MERSCOPE panels could

534 be better suited for branches of biology not well sampled by the 1,000 plex CosMx panel. We

535 recommend subsampling existing atlas data to determine whether the gene set which can be studied

536 will be sufficient to cluster the cell types of interest and identify the necessary biological programs.

537 We note that each of the manufacturers has publicly stated plans to grow their product offerings

538 to increasing panel sizes.

539   There are several limitations of our study. While we attempted to match time post slicing,

540 the unintentional acquisition of MERSCOPE tissues at thinner thicknesses meant that the rerun

541   MERSCOPE data had a longer time on slide than other panels. However, the fact that the increase

542   in counts in the rerun matched the increase in imaging volume, and the fact that Xenium runs

543   showed stable expression levels over time suggests that this contribution was minimal.

544   Additionally, our panel design for MERSCOPE required removal of genes so the panel was

545   compatible with all tissues, lowering the plexity slightly. This could have compromised

546   MERSCOPE's ability to identify cell types relative to Xenium.

547   Most importantly, we only attempted to compare the performance of iST platforms under

548   typical use cases for clinical samples obtained from archival biobanks. Our results don't

549   necessarily extend to non-human samples, frozen samples, and even FFPE samples which have

550   been extensively validated for high RNA integrity. Indeed, there have been reports that

551   MERSCOPE, in previous studies of the mouse brain, shows comparable or even superior results

552   to those reported by 10X Xenium[41]. Given the large change in data quality between the normal

553   and tumor TMA, we cannot exclude the possibility that in the highest quality samples MERSCOPE

554   would provide higher transcript numbers, with the associated downstream benefits relative to

555   Xenium and CosMx. However, the current guidance of $DV_{200} > 0.6$ restricts studies to the upper

556   regime sample quality and limits archival investigations.

557   Despite these limitations, our overall interpretation of these results is that amplification of

558   RNA signal is especially important for recovery of transcript counts by iST in low-quality samples

559   where RNA may be highly degraded and fewer landing sites are available for probes. Platforms

560   (such as Xenium) which rely on small numbers of landing sites and are subsequently heavily

561   amplified are robust to RNA degradation and are thus more broadly compatible with a broad range

562   of samples. On the other hand, when sample quality is high (as in some of our tumor samples) the

563   gap between amplified and unamplified platforms' performance closes and most platforms can

564   yield useful data for subsequent downstream spatial analysis.

565   **Methods**

566   **Sample choice and TMA construction**

567   Two TMAs were constructed using FFPE clinical discards at Brigham and Women's

568   Hospital Pathology Core and were acquired with a waiver of consent for non-sequencing based

569   readouts under IRB 2014 P 001721. The samples included:

570   1. A tumor TMA of 170 cores, 0.6 mm in diameter, including a variety of cancer samples

571   and healthy lymphoid tissue as a positive staining control. The TMA samples were

572   selected from samples previously characterized by ImmunoProfile and were selected

573   to encompass both high and low levels of the biomarkers in the ImmunoProfile panel

574   [*CD8*, *PD-1*, *PD-L1*, *Foxp3*, tumor marker (*Cytokeratin*, *Sox10*, or *PAX8*)].

575   Annotations were performed by KF and SR based on H&E and immunofluorescence

576   staining. Cores included both tumor and healthy control annotation, though for the

577   purpose of this study, all were combined under their tumor label. Tumors were also

578   chosen to be a mixture of *PD-L1* high and *PD-L1* low, a parameter to be analyzed at a

579   future date. This TMA had previously been studied by both H&E, and several highly

580   multiplexed immunostaining approaches, and was known to be of high morphological

581   integrity.

582   2. A normal TMA of 45 cores 1.2 mm in diameter representing a broad range of normal

583   tissues. Samples were sourced from the same patient in either duplicate or triplicate.

584   This TMA was chosen for the breadth of tissue lineages included and the relatively

585   large core size.

586   All samples were fully de-identified before assembly into TMAs. The breakdown of the number

587   of samples per tissue and the number of cores per tissue is included in **Supplementary Table 1-**

588   **2.**

589   **Preparation of sequential sections**

590   Sequential sections were prepared according to manufacturer instructions ("Tissue

591   Preparation Guide Demonstrated Protocol CG000578" for Xenium, "91600112 MERSCOPE User

592   Guide Formalin-Fixed Paraffin-Embedded Tissue Sample Preparation RevB" for Vizgen, and

593   "MAN-10159-01CosMx SMI Manual Slide Preparation Manual" for CosMx) at the Brigham and

594   Women's Hospital Pathology Core. Prior to collecting samples, ~50 μm of each TMA were faced

595   off to reach deeper into the sample where RNA integrity was likely higher. 5 μm sequential

596   sections were then collected, floated in a 37˚C water bath, and adhered to Xenium slides (10X, PN

597   1000460), Vizgen FFPE coverslips (Vizgen, PN 10500102), or standard Superfrost+ slides for

598   CosMx (Leica BOND PLUS slides, Leica Biosystems S21.2113.A). TMAs were sliced as close to

599   the center of the active area as possible for each platform. Samples were baked at 42˚C for 3 hours

600   for Xenium, 55 ˚C for 15 minutes for MERSCOPE, and 60˚C for 16 hours for CosMx. Sections

601   were stored according to manufacturer instructions prior to processing, with 10X Xenium stored

602   in a desicator at room temperature, Vizgen MERSCOPE coverslips stored at -20˚C, and Nanostring

603   CosMx slides stored at 4˚C. Samples for 10X Xenium and Vizgen MERSCOPE were brought to

604   the Spatial Technology Platform at the Broad Institute for processing, while samples for

605   Nanostring CosMx were processed at the Wei lab at Brigham and Women's Hospital.

606

Vizgen MERSCOPE probe selection

Pre-designed probe panels from Vizgen were not available at the time of the experiment. Therefore, we ordered custom gene panels to match the pre-released gene panels from 10X for the human breast and human lung panels. Gene lists were uploaded to the Vizgen panel design portal and were checked against all profiled tissues, removing genes that were overexpressed in any individual tissue based on Vizgen's design guidelines (FPKM > 900), and ensuring that the total panel FPKM did not exceed the allowed limit in any individual sample type. Panels were manufactured at the 300 gene scale as custom panels BP0892 and BP0893. The final gene lists, for all three iST modalities are available in **Supplementary Table 3.**

Vizgen MERSCOPE data acquisition

MERSCOPE samples were imaged according to manufacturer protocol "9160001 MERSCOPE Instrument User Guide RevF". Samples were processed in two batches, the first of four samples, two of each TMA and with each library prepped in parallel; and a follow up sample of each TMA re-run with the breast panel. Samples were first hybridized with anchoring probes overnight before being embedded in a polyacrylamide gel. Samples were incubated for two hours with a digestion solution at 37°C and then overnight at 47°C overnight in a detergent clearing solution and proteinase K to remove native proteins while the anchoring probes kept nucleic acids bound to the gel. After clearing, samples were additionally photobleached using Vizgen's MERSCOPE Photobleacher for three hours at room temperature in the clearing solution. Samples were hybridized with encoding probes and a cell boundary stain (PN 10400118) and then imaged with imaging kits (PN 10400005). Samples were stored at 37°C in clearing solution after hybridization and before final imaging. After an initial examination of the data, a second batch of

630   both TMAs was run a second time with the human breast panel, increasing the set imaging capture

631   thickness from 5 μm to 10 μm to capture more tissue from cores that had lifted during the gel

632   embedding process. Data was processed on premises through the standard Vizgen workflow to

633   generate cell by gene and transcript by location matrices. We segmented the data with a built-in

634   Cellpose method on the most accurate looking cell boundary stain.

635

636   **10X Xenium data acquisition**

637   10X Xenium samples were processed in three batches according to manufacturer protocols

638   "Probe Hybridization, Ligation & Amplification, User Guide CG0000582" and "Decoding &

639   Imaging, User Guide CG000584". Samples were stained utilizing 10X's predesigned Human

640   Breast (10X, PN 1000463), Human Multi-Tissue and Cancer (10X, PN 1000626), and Human

641   Lung panels (10X, PN 1000601), as they became available from the manufacturer. Slides for both

642   TMAs were processed in pairs according to which probe library they were receiving. Slides were

643   stained with a Xenium imaging kit according to manufacturer instructions (10X, PN 1000460).

644   Briefly, padlock probes were incubated overnight before rolling circle amplification and native

645   protein autofluorescence was reduced with a chemical autofluorescence quencher. Slides were

646   processed on a 10 Xenium Analyzer, with ROIs selected to cover the entire TMA region. Data was

647   processed on premises through the standard 10X workflow to generate cell by gene and transcript

648   by location matrices.

649

650   **Nanostring CosMx**

651   Nanostring CosMx samples were prepared with one 1000 plex panel. Samples were

652   hybridized with probes and stained with cell markers. Samples were loaded onto the CosMx SMI

653 at the same time for imaging, during which branched fluorescent probes were hybridized onto the

654 samples to amplify the signal above the background.

655 Nanostring CosMx samples were prepared with Human Universal Cell Characterization

656 1000 Plex Panel (part number 122000157) according to manufacturer protocol "MAN-10159-01

657 CosMx SMI Manual Slide Preparation Manual". Firstly, slides were baked at 60℃ overnight for

658 better tissue adherence. After baking, slides were treated sequentially with deparaffinization, target

659 retrieval (15 min at 100℃), permeabilization (3μg/mL proteinase K, 15 min at 40℃), fiducials

660 application, post-fixation, NHS-acetate application and then hybridized with denatured probes

661 from universal panel and default add-on panel. After *in situ* hybridization (18 hours at 37℃), slides

662 were washed and incubated with DAPI (15 min at RT) and marker stain mix (with *PanCK, CD45,*

663 *CD68* and cell segmentation marker *CD298/B2M*). Slides were washed and loaded onto the

664 CosMx SMI for UV bleaching, imaging, cycling and scanning. Raw images were decoded by

665 default pipeline on Atomx SIP (cloud-based service). Machine: CosMx_0020. Serial Number:

666 INS2301H0020

667 **Data preprocessing**

668 After data acquisition, the resulting outputs were uploaded to a Google bucket associated

669 with a terra.bio Workspace for distribution and follow on analysis.

670 To facilitate standardized data formatting and subsequent analytical processes, we built a

671 data ingestion pipeline with the following objectives: a) to grab cell-level and transcript-level

672 data from diverse platforms and normalize the data structure; b) to tag each cell and transcript

673 with essential metadata including tissue type, tumor status, *PD-L1* status, among others

674 (**Supplementary Fig. 6**); and c) to transform the data into various formats tailored to the

675 requirements of particularized analyses. Specifically, to tag the data, core centers in the TMA

676   were manually identified using DAPI images (Xenium) or cell metadata that contains global

677   coordinates (MERSCOPE and CosMx) using QGIS(version:3.16.10-Hannover). Cells or

678   transcripts within a specified radius were then labeled with core metadata via spatial joining

679   (implemented by GeoPandas, version:0.13.0). In instances where cores are in close proximity or

680   when a uniform radius cannot be applied effectively, we manually generated the core boundary

681   masks.

682

683   **Reproducibility**

684   To evaluate panel to panel reproducibility we summed the expression level of shared

685   genes between indicated panels (breast vs. multi-tissue and breast vs. lung panels from Xenium

686   and breast vs. lung panels from MERCOPE) over an individual core and plotted all cores present

687   in each panel, before calculating a Pearson's correlation. The format of the data used is shown in

688   **Supplementary Table 10.** To evaluate core to core reproducibility, the individual gene counts of

689   core 1 were plotted against those of core 2 and a Pearson's r correlation was calculated.

690 **On target rates and false-discovery measurements**

691 To compare across panels and platforms, we subset all datasets to include only cores

692 assayed in all runs. The fraction of on-target barcodes was calculated as a percentage of all

693 transcripts corresponding to genes relative to the total number of calls (including negative

694 control probes and unused barcodes or blank barcodes). These measurements were performed on

695 individual cores and averaged across all cores of the same tissue type.

696 Because the difference in relative numbers of controls and target barcodes across

697 different platforms, we adopted the false discovery rate (FDR) calculation to evaluate the

698 specificity in a more normalized way (**Fig. 2f-g**). We calculated the FDR of platform p panel m

699 data in tissue t using the following equation and cell level data (**see example in Supplementary**

700 **Table 11**):

701 $$FDR(negative\ control\ probes)_{p,m}^{t} = \ mean\left(\frac{\sum_j neg_{jn}}{\sum_i g_{in} + \sum_j neg_{jn}} \times \frac{I}{J} \times \%\right), \quad n = \{1, \dots, N\}$$

702 Where N is the total number of cores that belong to tissue type t, I is the total number of unique

703 genes, J is the total number of negative control probes, g_{in} is the gene expression of gene i in

704 core n, neg_{jn} is the total calls negative control probe j in core.

705 Since MERSCOPE does not include negative control probes, FDR was recalculated by

706 substituting negative control with blank barcodes (**Fig. 2f**) using the following equation:

707 $$FDR(blank\ barcodes)_{p,m}^{t} = \ mean\left(\frac{\sum_l blank_{ln}}{\sum_i g_{in} + \sum_l blank_{ln}} \times \frac{I}{L} \times \%\right), \quad n = \{1, \dots, N\}$$

708 Where N is the total number of cores that belong to tissue type t, I is the total number of unique

709 genes, L is the total number of unused barcodes or blank barcodes, g_{in} is the gene expression of

710 gene i in core n, $blank_{ln}$ is the total calls of unused barcode or blank barcode l in core n,

711 specifically, we used BLANK for Xenium data, Blank for MERSCOPE data, and SystemControl

712   for CosMx data. We only used the data from matched cores, so N is same for different platform

713   p.

714

715   **Sensitivity comparison**

716   Sensitivity was measured by the percentage of the total number of unique genes detected

717   above noise level, where the noise was estimated as two standard deviation above average

718   expression of the negative control probes.

719

720   **Orthogonal RNA-Seq concordance analysis**

721   RNA TCGA cancer sample gene data summarizes 7,932 samples from 17 different cancer

722   types, and it provides FPKM for each gene documented. We used all samples which were

723   annotated as BRCA (Breast cancer), BLCA (Bladder cancer), COAD and READ (colorectal

724   cancer), HNSC (head and neck squamous cell carcinoma), LUAD and LUSC (non-small cell lung

725   cancer), SKCM (melanoma), and OV (ovarian cancer). For GTEx, we selected the tissue types

726   matching the annotation in our normal tissue TMA. For each panel, the genes probed by iST were

727   averaged across all patients with the matching tissue label from the RNA-seq database.

728   To get pseudo-bulked iST values, the expression level of each gene in each core was normalized

729   to the sum of all genes in that core and scaled by 100,000. We then averaged these scaled pseudo-

730   bulk expression values across cores and plotted them against the averaged FPKMs from reference

731   RNA-seq data sets.

732

733   **Tissue marker enrichment analysis**

734   To determine the assay's ability to specifically identify known lineage markers, we focused

735   on the normal tissue TMA profiled with multi-tissue panel of Xenium, breast panel of MERSCOPE,

736   and 1K panel of CosMx. We selected genes with known canonical expression patterns using based

737   on transcriptomics data from GTEx. If a gene had 20-fold higher expression in a specific tissue

738   than every other tissue combined, this gene was considered to be a tissue marker and was used for

739   assessing specificity for each platform. Counts for each gene were normalized to the total counts

740   within the core, and then the Z-score of this gene across tissue types was plotted in a heatmap **Fig.**

741   **3e**. We calculated average expression of a gene across cores of the same tissue type and normalized

742   to the total averaged expression of all genes. Z-scores were calculated with the mean and standard

743   deviation across all averaged genes.

744

745   **Evaluation of cell segmentation performance**

746   In the absence of ground truth data, we conducted a comparative analysis of cell counts,

747   cell areas, coexpression across various platforms and panels, utilizing the segmentation results

748   supplied by each respective company. To facilitate comparison, cell counts were normalized to a

749   consistent area of 1000 μm^2. Both cell count and cell area were then delineated at two distinct

750   levels of detail: a consolidated assessment encompassing all tissue types (see **Fig. 4b-c**), as well

751   as a segregated evaluation by individual tissue types (refer to **Supplementary Table 6-7**). To

752   evaluate the biological performance of the segmentation, we plotted coexpression plots of gene

753   pairs that are mutually exclusive including *CD3e* vs. *CD19*, *CD4* vs. *CD8*, and *CD3e* vs.

754   *EPCAM*. We pooled all the filtered cells from matched cores and all available panels of each

755   platform, dropped cells which do not express either gene, plotted the expression of one gene

756 against the other, and converted the scatter plot to a 2D histogram showing cell numbers in each

757 co-expression bin (**Fig. 4g**).

758

759 **Cells per area quantification**

760 Segmented cells were aggregated by TMA cores. For Xenium and MERSCOPE data, the

761 estimation of tissue area was performed by calculating the area of a discernible circle, utilizing

762 respective radius of 0.3 µm and 0.6 µm for tumor and normal TMA samples. Conversely, for the

763 CosMx dataset, the tissue area estimation was approached differently due to its square-like data

764 presentation, a result of the FOV selection process. Here, the tissue area was deduced by

765 multiplying the number of FOVs covered by each core with the area of a single FOV.

766 **Clustering**

767 For cell filtering, cells with less than 10 transcript counts in MERFISH and Xenium

768 datasets were removed, and cells with less than 20 transcript counts in CosMx datasets were

769 removed. We followed standard processes to then cluster and annotate cell types across each

770 dataset using Scanpy[42]. Briefly, data was normalized and scaled, dimensionality reduction was

771 performed and cell clusters were identified[43, 44]. To identify the cell type for each cluster, we used

772 a t-test to find the markers for each Leiden cluster and annotated them according to previous

773 literature[29-36]. These are some of the example markers used for cell type annotation: in breast

774 samples, PIGR and KIT for alveolar cells, for B cells, KRT5, DST, and MYLK for basal cells,

775 LUM, MMP2 and CXCL12 for fibroblast, etc. Heatmaps of the top 3 markers for each cluster are

776 drawn for each dataset from all three panels (refer to **Supplementary Figure 5a-c**). For datasets

777 that showed batch effect with patients, Harmony was used to remove this variance[45]. Correlation

778   heatmaps were generated over overlapping genes that exist in both datasets, and the Pearson

779   correlation coefficient was calculated.

780

781   Acknowledgment

792

793   Author contributions

794   Conceptualization: A.Y., S.F., tissue-microarray construction: K.F., K.P., T.B., S.R.,

795   pathological annotation: K.P., S.R., gene selection: A.Y., S.F., Xenium and MERSCOPE data

796   acquisition: J.N., CosMx data acquisition: C.G., M.T.,K.W., analysis: H.W., R.H., B.G.,S.F.,

797   figure generation: H.W., R.H., B.G.,S.F., writing original draft: H.W., R.H., J.N., B.G., S.F., draft

798   reviewing and editing: H.W., R.H., B.G., S.F., supervision: B.G., S.F., funding acquisition: S.F.

799   K.W.

800

Data and Code Availability

Cell level data and gene level data will be made available at the time of final publication. All code used in this manuscript for data processing and analysis will be made available on github prior to final publication.

Declaration of interests

All authors declare that they have no conflicts of interest.

References

1. Chen W.T., Lu A. Croissants K. et al. Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease. Cell. 182 (4), 976–991 (2020).

2. Hunter, M.V., Moncada, R., Weiss, J.M. et al. Spatially resolved transcriptomics reveals the architecture of the tumor-microenvironment interface. Nat Commun 12, 6278 (2021).

3. Lein E, Borm LE, Linnarsson S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. Science. 2017 Oct 6;358(6359):64-69.

4. Lewis SM, Asselin-Labat ML, Nguyen Q, et al. Spatial omics and multiplexed imaging to explore cancer biology. Nat Methods. 2021 Sep;18(9):997-1012.

5. Rao, A., Barkley, D., França, G.S. et al. Exploring tissue architecture using spatial transcriptomics. Nature 596, 211–220 (2021).

6. Haque, A., Engel, J., Teichmann, S.A. et al. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med 9, 75 (2017).

7. Williams, C.G., Lee, H.J., Asatsuma, T. et al. An introduction to spatial transcriptomics for biomedical research. Genome Med 14, 68 (2022).

8. Tian, L., Chen, F. & Macosko, E.Z. The expanding vistas of spatial transcriptomics. Nat Biotechnol 41, 773–782 (2023).

9. Janesick A., Shelansky R., Andrew D.G., et al. High resolution mapping of the breast cancer tumor microenvironment using integrated single cell, spatial and insitu analysis of FFPE tissue. bioRxiv 2022.10.06.510405

10. He S., Bhatt R., Birditt B., et al. High-plex multiomic analysis in FFPE tissue at single-cellular and subcellular resolution by spatial molecular imaging. bioRxiv 2021.11.03.467020

11. Moffitt, J. R. et al. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. Science 362, eaau5324 (2018).

12. Du J., Yang Y.C., An Z.J., et al. Advances in spatial trasnciptomics and related data analysis strategies. J Transl Med. 2023; 21: 330.

13. Phan, H.V., van Gent, M., Drayman, N. et al. High-throughput RNA sequencing of paraformaldehyde-fixed single cells. Nat Commun 12, 5636 (2021).

14. Xu, Z., Zhang, T., Chen, H. et al. High-throughput single nucleus total RNA sequencing of formalin-fixed paraffin-embedded tissues by snRandom-seq. Nat Commun 14, 2734 (2023).

15. Gnanapragasam VJ. Unlocking the molecular archive: the emerging use of formalin-fixed paraffin-embedded tissue for biomarker research in urological cancer. BJU Int. 2010 Jan;105(2):274-8.

16. Matsunaga H., Arikawa K., Yamazaki M., et al. Reproducible and sensitive micro-tissue RNA sequencing from formalin-fixed paraffin-embedded tissues for spatial gene expression analysis. Sci Rep 12, 19511 (2022).

846   17. Kok Hao Chen et al. Spatially resolved, highly multiplexed RNA profiling in single

847   cells.Science 348, aaa 6090 (2015).

848

849   18. Walsh, L.A., Quail, D.F. Decoding the tumor microenvironment with spatial technologies.

850   Nat Immunol 24, 1982–1993 (2023).

851   19. Tomczak K., Czerwinska P., Wiznerowicz M. The cancer genome atlas (TCGA): an

852   immeasuable source of knowledge

853   20. GTEx Consortium. Genetic effects on gene expression across human tissues. Nature 550,

854   204–213 (2017).

855   21. Junttila S., Smolander J., Elo L., et al. Benchmarking methods for detecting differential

856   states between conditions from multi-subject single-cell RNA-seq data. Briefings in

857   Bioinformatics, Volume 23, Issue 5, Sep. 2022.

858   22. Zimmerman S. M., Fropf R., Kulasekara B. R., et al. Spatially resolved whole

859   transcriptome profiling in human and mouse tissue using digital spatial profiling. Genome

860   Res. 2022 Oct;32(10):1892-1905.

861   23. Trevethan. R. Sensitivity, specificity and predictive values: foundations, liabilities, and

862   pitfalls in research and practice. Front Public Health. 2017 Nov 20;5:307.

863   24. He S, Bhatt R, Brown C. et al. High-plex imaging of RNA and proteins at subcellular

864   resolution in fixed tissue by spatial molecular imaging. Nat Biotechnol. 2022

865   Dec;40(12):1794-1806.

866   25. Garrido-Trigo A, Corraliza AM, Veny M et al. Macrophage and neutrophil heterogeneity

867   at single-cell spatial resolution in human inflammatory bowel disease. Nat Commun. 2023

868   Jul 26;14(1):4506. doi: 10.1038/s41467-023-40156-6. PMID: 37495570; PMCID:

869   PMC10372067.

870   26. Nieto P., Elosua-Nayes M., Trincado J. L., et al. A single-cell tumor immune atlas for

871   precision oncology. Genome Res. 2021 Oct; 31(10): 1913–1926.

872   27. Elmentaite R., Conde C. D., Yang L., et al. Single-cell atlases: shared and tissue-specific

873   cell types across human organs.

874   28. Heumos L., Schaar A. C., Lance C., et al. Best practices for single-cell analysis across

875   modalities. Nat Rev Genet 24, 550–572 (2023).

876   29. Kumar, T., Nee, K., Wei, R. et al. A spatially resolved single-cell genomic atlas of the adult

877   human breast. Nature 620, 181–191 (2023).

878   30. Gray G.K., Li C.M., Rosenbluth J.M. et al. A human breast atlas integrating single-cell

879   proteomics and transcriptomics. Developmental Cell, Volume 57, Issue 11, 2022.

880   31. Reed A.D., Pensa S., Steif A. et al. A human breast cell atlas mapping the homeostatic

881   cellular shifts in the adult breast. bioRxiv 2023.04.21.537845

882   32. Travaglini K.J., Nabhan A.N., Penland L. et al. A molecular cell atlas of the human lung

883   from single-cell RNA sequencing. Nature 587, 619–625 (2020).

884   33. Sikkema, L., Ramírez-Suástegui, C., Strobl, D.C. et al. An integrated cell atlas of the lung

885   in health and disease. Nat Med 29, 1563–1577 (2023).

886   34. Wu, S.Z., Al-Eryani, G., Roden, D.L. et al. A single-cell and spatially resolved atlas of

887   human breast cancers. Nat Genet 53, 1334–1347 (2021).

888   35. Dave A., Charytonowicz D., Francoeur N. et al. The breast cancer single-cell atlas: defining

889   cellular heterogeneity within model cell lines and primary tumors to inform disease

890   subtype, stemness, and treatment options. Cell Oncol (Dordr). 2023 Jun;46(3):603-628.

891   36. Yeo S.K., Zhu X., Okamoto T. et al. Single-cell RNA sequencing reveals distinct patterns

892   of cell state heterogeneity in mouse models of breast cancer. eLife 9:e58810.

37. Greenwald, N.F., Miller, G., Moen, E. et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. Nat Biotechnol 40, 555–565 (2022).

38. Stringer, C., Wang, T., Michaelos, M. et al. Cellpose: a generalist algorithm for cellular segmentation. Nat Methods 18, 100–106 (2021).

39. Petukhov, V., Xu, R.J., Soldatov, R.A. et al. Cell segmentation in imaging-based spatial transcriptomics. Nat Biotechnol 40, 345–354 (2022).

40. Chen, H., Li, D. & Bar-Joseph, Z. SCS: cell segmentation for high-resolution spatial transcriptomics. Nat Methods 20, 1237–1243 (2023).

41. Marco Salas, S., Czarnewski, P., Kuemmerle, L. B., Helgadottir, S., Mattsson Langseth, C., Tiesmeyer, S., ... & Nilsson, M. (2023). Optimizing Xenium In Situ data utility by quality assessment and best practice analysis workflows. bioRxiv, 2023-02.

42. Wolf, F., Angerer, P. & Theis, F. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 19, 15 (2018).

43. Dries R., Chen J., Del Rossi N., et al. Advances in spatial transcriptomic data analysis. Genome Res. 2021 Oct;31(10):1706-1718.

44. Traag, V.A., Waltman, L. & van Eck, N.J. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep 9, 5233 (2019).

45. Korsunsky, I., Millard, N., Fan, J. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 16, 1289–1296 (2019).

Supplementary Figures and Supplementary Tables: [link]
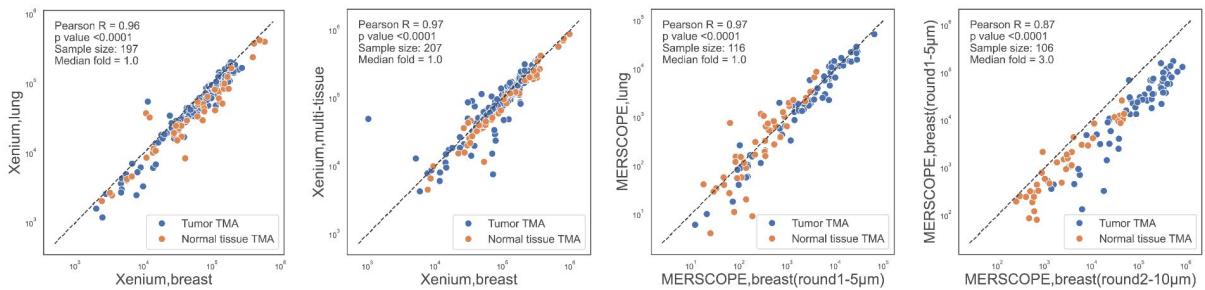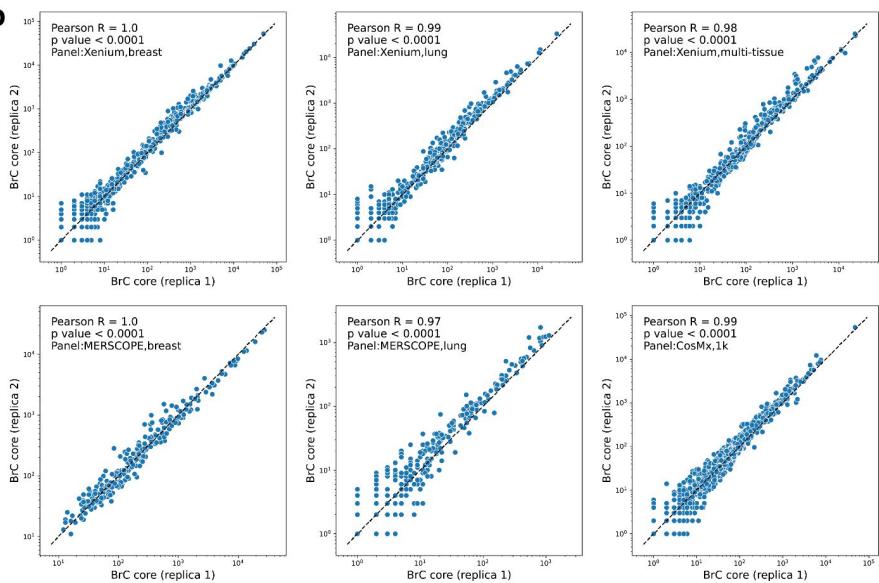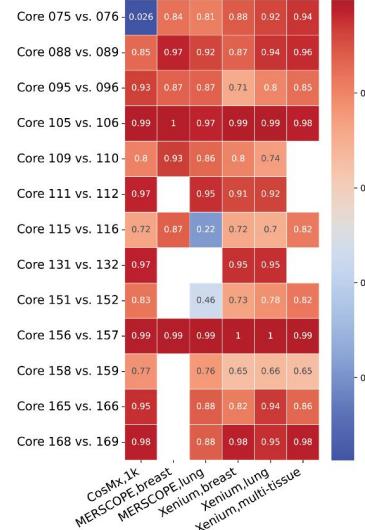
Supplementary Figure 1: Reproducibility across different panels and cores from same patient.
(a) Scatter plots of cumulative gene expression levels (natural log transformed) of shared genes between two panels within each platforms, captured from matched tissue cores. Column 1: Xenium breast vs. Xenium lung; Column 2: Xenium breast vs. Xenium multi-tissue; Column 3: MERSCOPE breast round 1(5 µm) vs. MERSCOPE breast round 2(10 µm). Each data point corresponds to a TMA core. **(b)** Scatter plots of gene expression levels (natural log transformed) of every shared gene between two cores of the same tissue type from the same patient. In this example, cores are from breast cancer tissue. Each data point corresponds to a gene. **(c)** Heatmap of correlation coefficient expressed as Pearson's r values, indicating good core-to-core or sample-to-sample reproducibility. Core pairs are selected from same tissue/tumor type from the same patients.
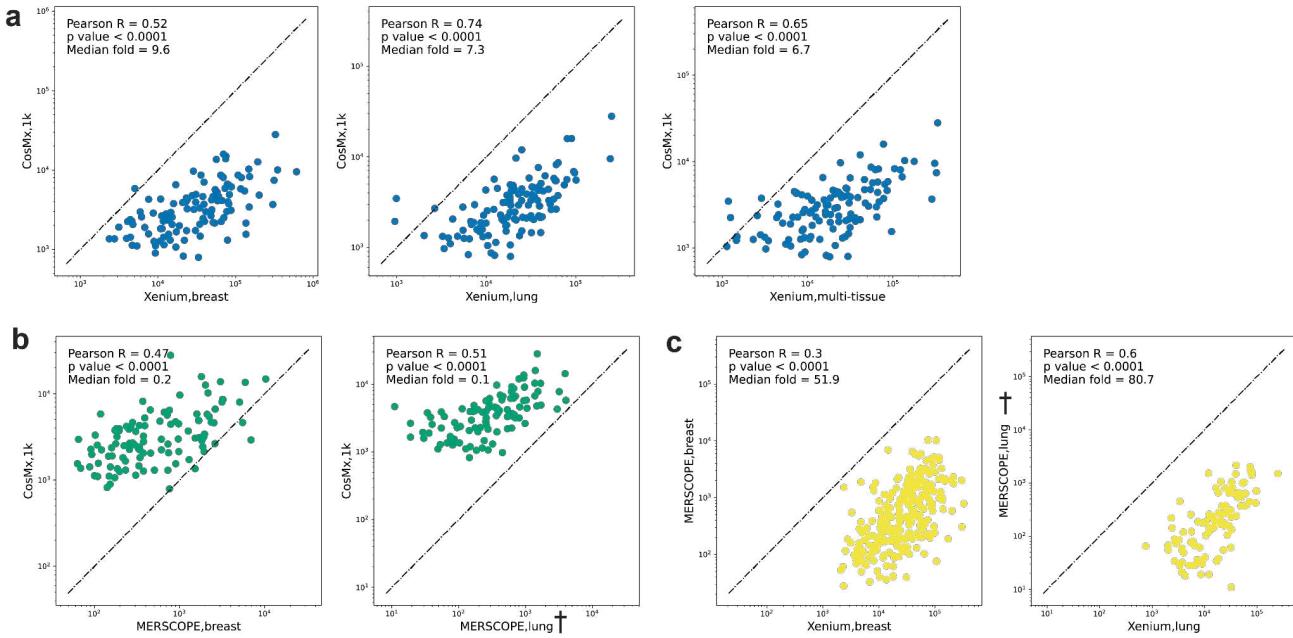
Supplementary Figure 2: Gene by gene plots of iST results by panel and by tissue microarray.
(a) Scatter plots of summed gene expression levels (natural log transformed) of every shared gene between Xenium (breast/lung) and CosMx (1k) data, captured from matched normal tissue TMA cores. Each data point corresponds to a gene. **(b)** Same as (a) but between MERSCOPE (breast/lung) and CosMx(1k). **(c)** Same as (a) but between Xenium(breast/lung) and MERSCOPE(breast/lung). **(d)** Same as (a) but between Xenium(multi-tissue) and CosMx(1k). † denotes the MERSCOPE lung panel acquired with a 5 µm imaging thickness.
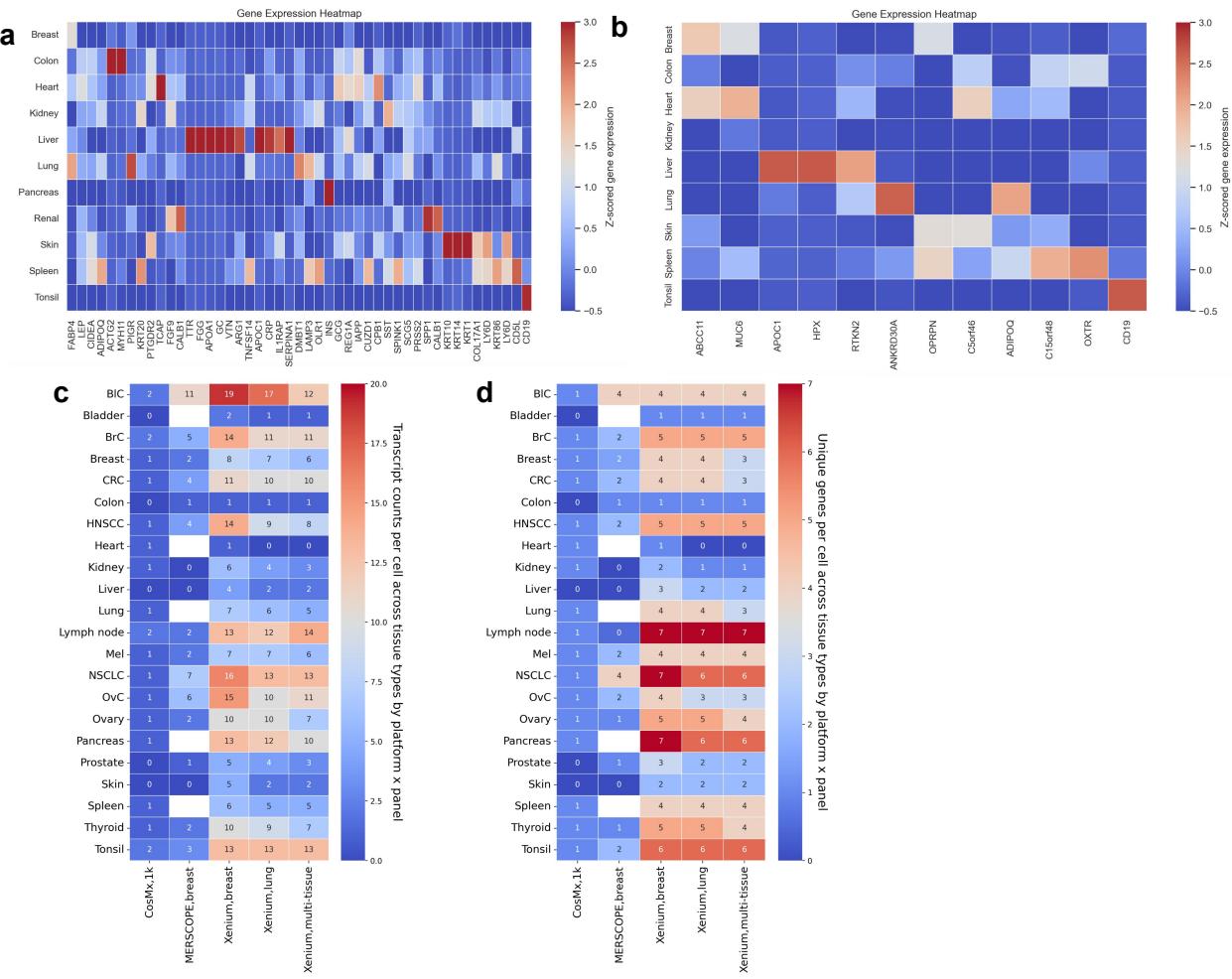
Supplementary Figure 3: Tissue marker analyses and cell level measurements.
(a) Heatmap of Z-scored gene expression showing CosMx's ability to specifically identify known lineage markers. We focused on the normal tissue TMA profiled with multi-tissue panel and selected genes with canonical expression patterns for this analysis. **(b)** Same as (a) but for MERSCOPE (breast panel). **(c)** Heatmap of transcripts per cell after filtration. Only shared genes (40) are considered here for each panel. **(d)** Same as (c) but showing unique transcripts from the same gene set.

Supplementary Fig: 4 Global clustering analyses
(a) Global Clustering results of tumor TMA from Xenium breast panel (top), Xenium lung panel (middle), and Xenium panhuman panel (bottom). **(b)** Global Clustering results of tumor TMA from MERFISH breast panel (top), MERSCOPE lung panel (middle), and CosMx multitissue panel (bottom).
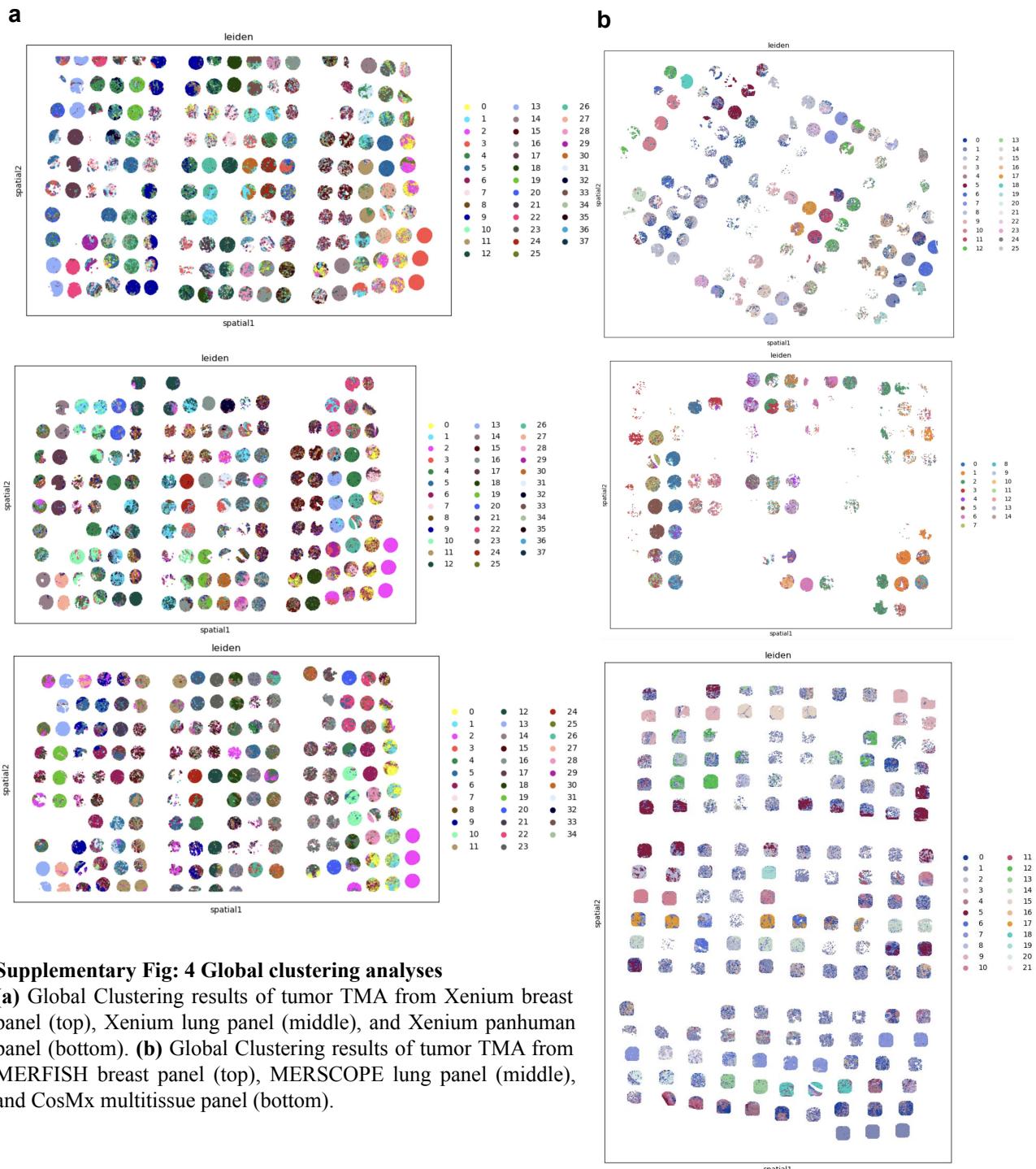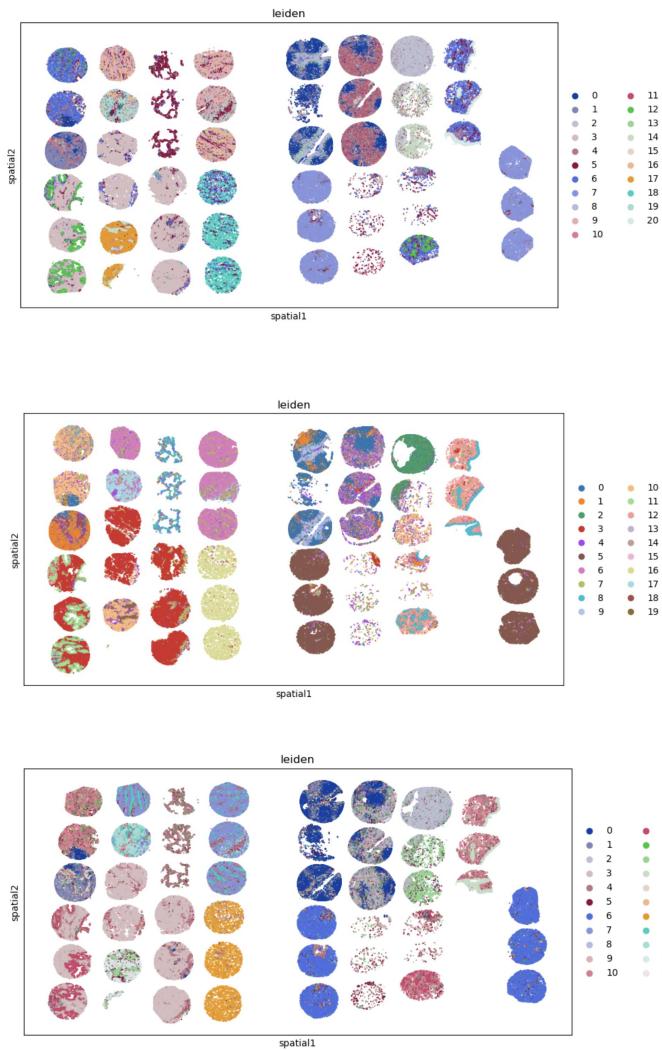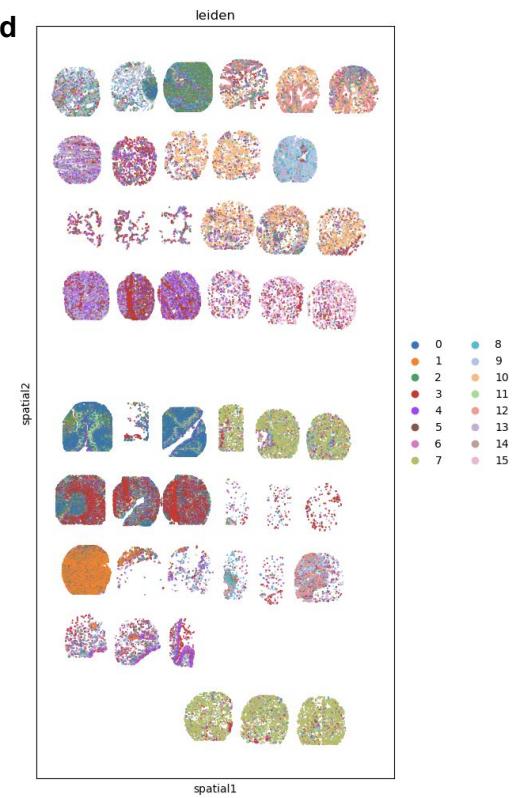
Supplementary Fig: 4 Global clustering analyses
(c) Global Clustering results of normal TMA from Xenium breast panel (top), Xenium lung panel (middle), and Xenium panhuman panel (bottom). **(d)** Global Clustering results of normal TMA from MERSCOPE breast panel (top), MERSCOPE lung panel (middle), and CosMx multitissue panel (bottom).
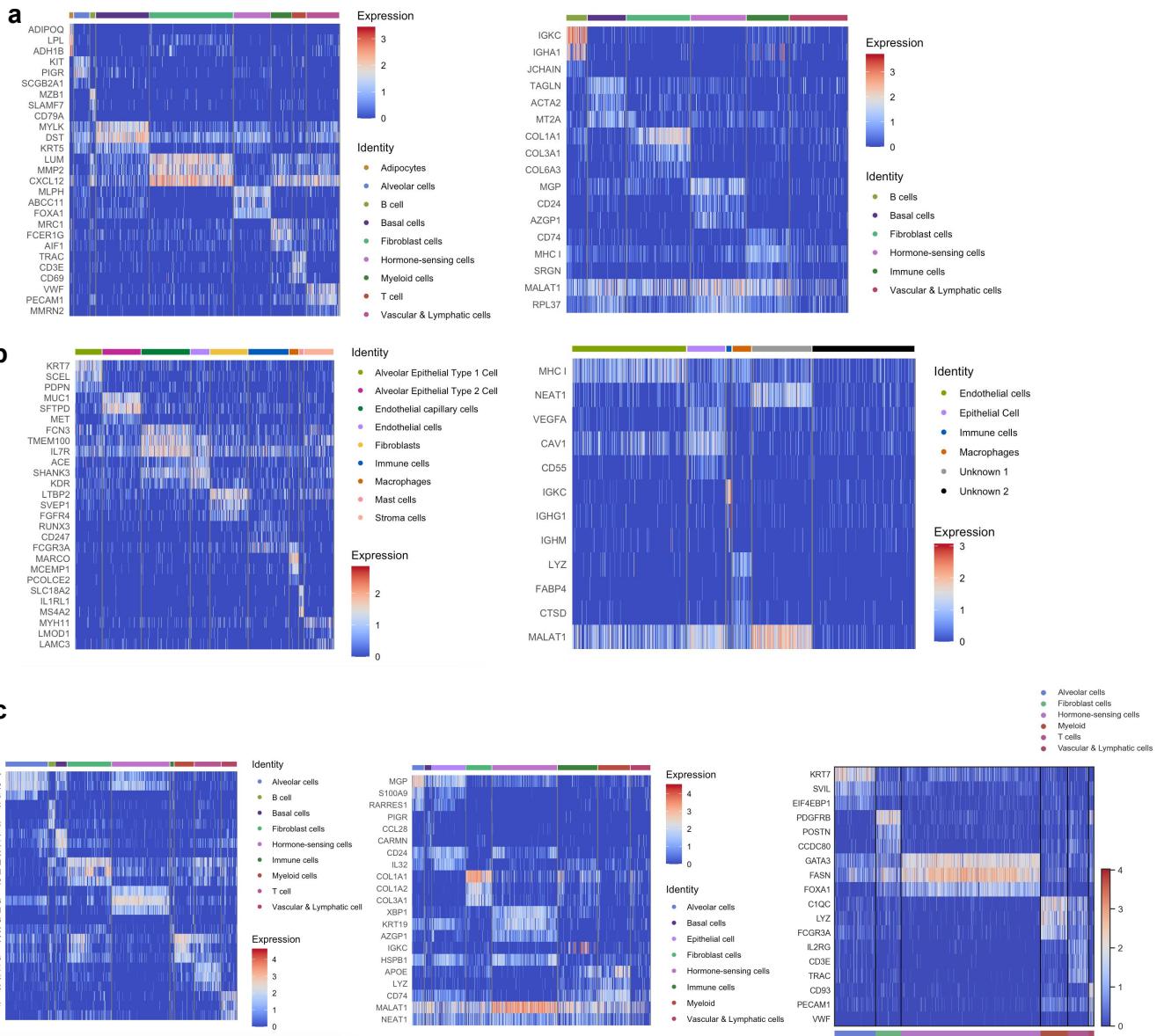
Supplementary Fig: 5. Cell type recovery and UMAPs
(a) Heatmap showing the top gene markers for cell types annotated in breast samples of normal TMA from Xenium breast (left) and Cosmx multitissue (right). **(b)** Heatmap showing the top gene markers for cell types annotated in lung samples of normal TMA from Xenium lung (left) and Cosmx multitissue (right). **(c)** Heatmap showing the top gene markers for cell types annotated in breast cancer samples of tumor TMA from Xenium breast (left), Cosmx multitissue (middle), NERSCOPE breast (right).
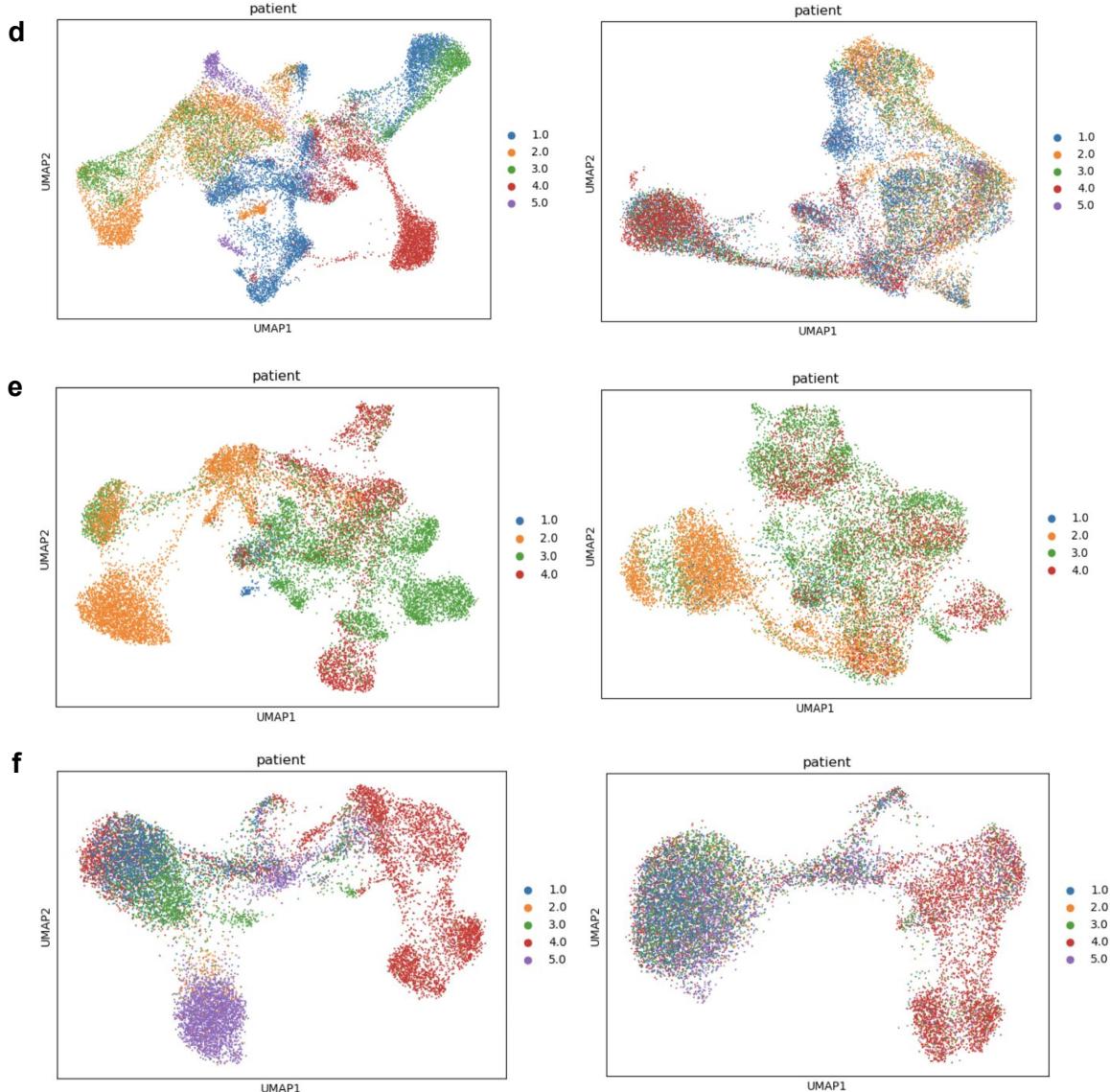
Supplementary Fig: 5. Cell type recovery and UMAPs
(d) UMAP of breast cancer samples of tumor TMA from Xenium breast panel pre (left) and post (right) batch effect removal. **(e)** UMAP of breast cancer samples of tumor TMA from Cosmx multitissue panel pre (left) and post (right) batch effect removal. **(f)** UMAP of breast cancer samples of tumor TMA from MERFISH breast panel pre (left) and post (right) batch effect removal
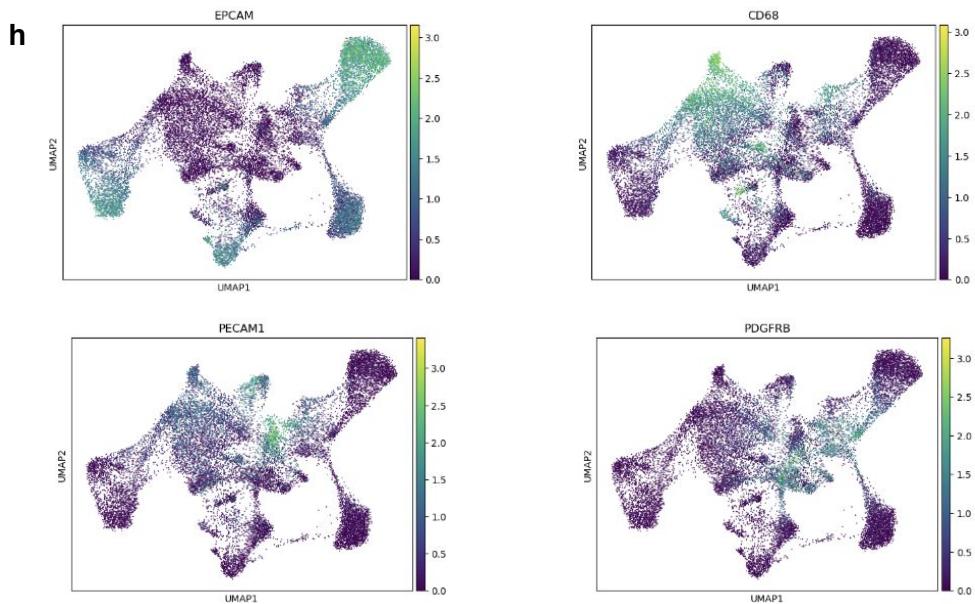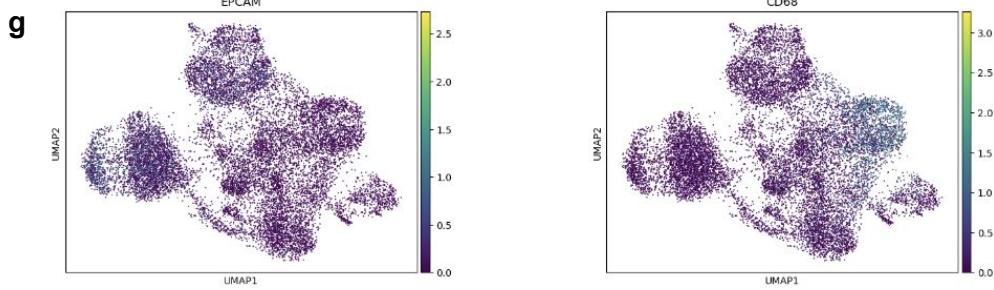
Supplementary Fig: 5. Cell type recovery and UMAPs
(g) UMAP plot of well-known gene markers for BrC, in breast cancer samples of tumor TMA from CosMx multitissue panel. **(h)** UMAP plot of well-known gene markers for BrC, in breast cancer samples of tumor TMA from Xenium multitissue panel.

a | Count table or transcript data with spatial information | Semi-automatically generated core mask joined with metadata | Cell/transcript-level data tagged with metadata

Geometry type: Points

Geometry type: Polygons

Spatial joining
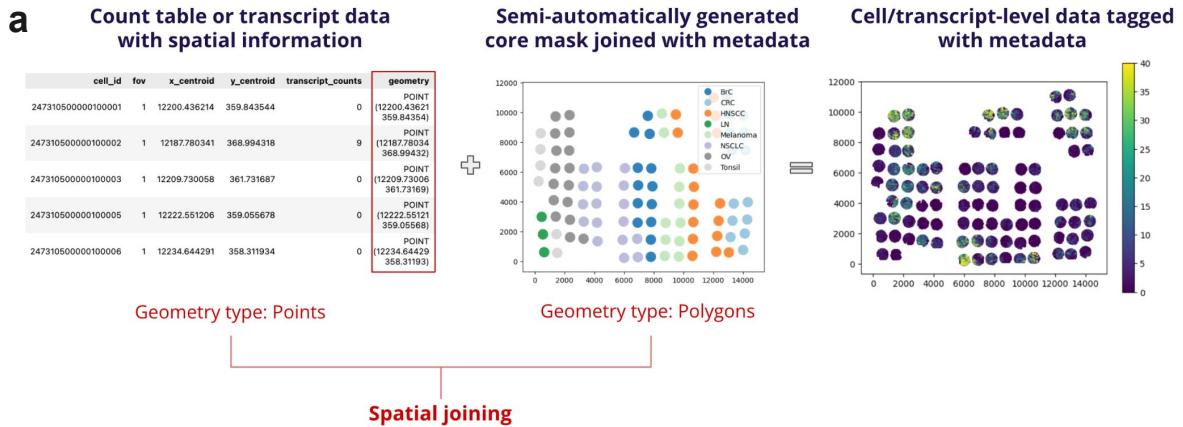
Supplementary Figure 6: Workflow for tagging imaging spatial transcriptomics data

(a) To facilitate standardized data formatting and subsequent analytical processes, we built this data ingestion pipeline with the following objectives: 1) to grab cell-level and transcript-level data from diverse platforms and normalize the data structure; 2) to tag each cell and transcript with essential metadata including tissue type, tumor status, PD-L1 status, among others; and 3) to transform the data into various formats tailored to the requirements of particularized analyses. Specifically, to tag the data, core centers in the TMA were pinpointed using DAPI images (Xenium) or cell metadata that contains global coordinates (MERSCOPE and CosMx) using QGIS(version:3.16.10-Hannover). Cells or transcripts within a specified radius were then labeled with core metadata via spatial joining (implemented by GeoPandas, version:0.13.0). In instances where the cores are in close proximity or when a uniform radius cannot be applied effectively, we manually generated the core boundary masks.