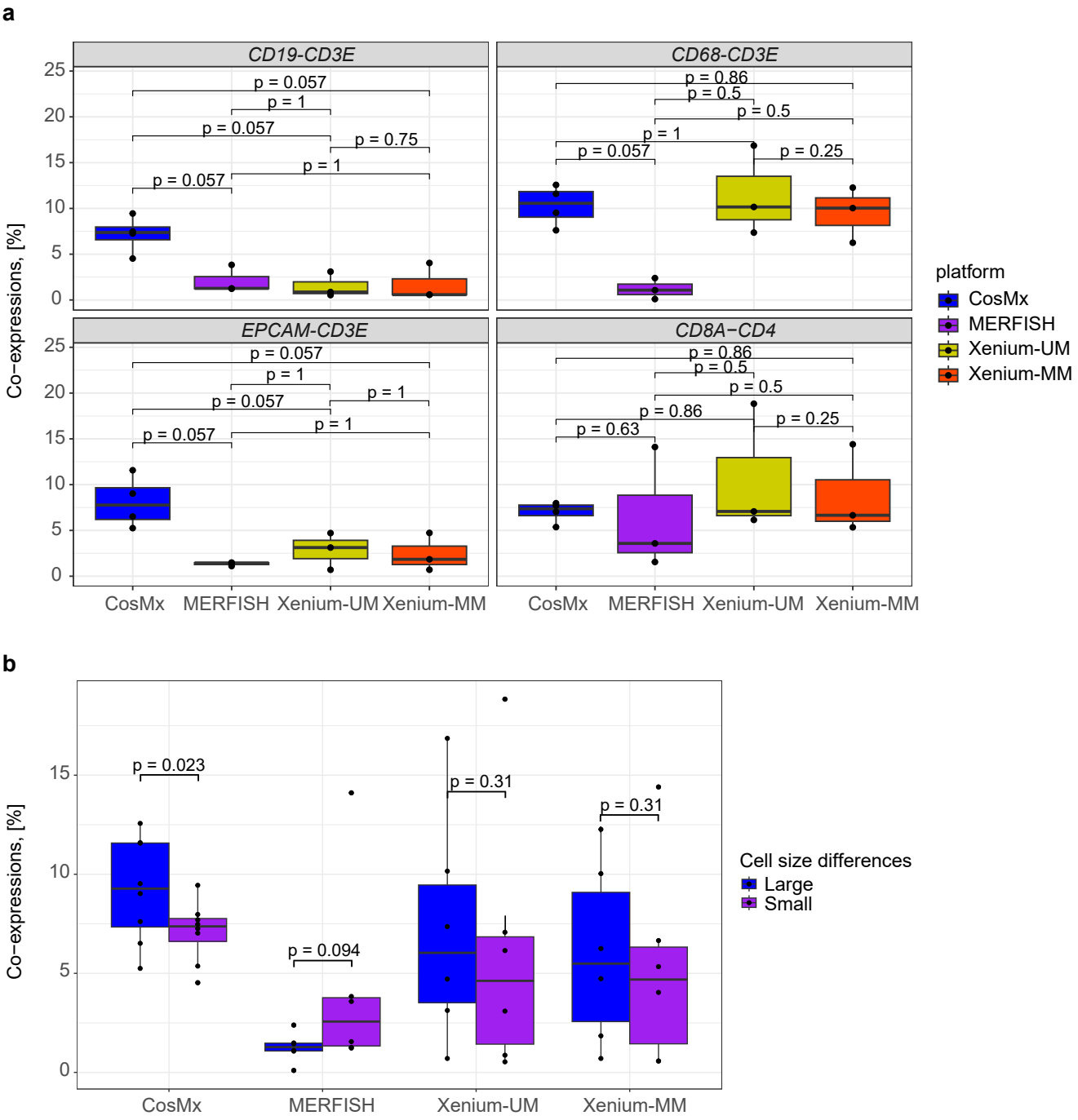
Supplementary Information for
Comparison of imaging based single-cell resolution spatial transcriptomics profiling platforms using formalin-fixed paraffin-embedded tumor samples

Nejla Ozirmak Lermi[1]*, Max Molina Ayala[2]*, Sharia Hernandez[2], Wei Lu[2], Khaja Khan[2], Alejandra Serrano[2], Idania Lubo[2], Leticia Hamana[2], Katarzyna Tomczak[2], Sean Barnes[2], Jinzhuang Dou[1], Qingnan Liang[1], RTI Team, Maria Gabriela Raso[2], Ximing Tang[2], Mei Jiang[2], Beatriz Sanchez-Espiridion[2], Annikka Weissferdt[3], John Heymach[4], Jianjun Zhang[4], Boris Sepesi[5], Tina Cascone[4], Anne Tsao[4], Mehmet Altan[4], Reza Mehran[5], Don Gibbons[4], Ignacio Wistuba[2], Cara Haymaker[2#*], Ken Chen[1#*] & Luisa M. Solis Soto[2#*]

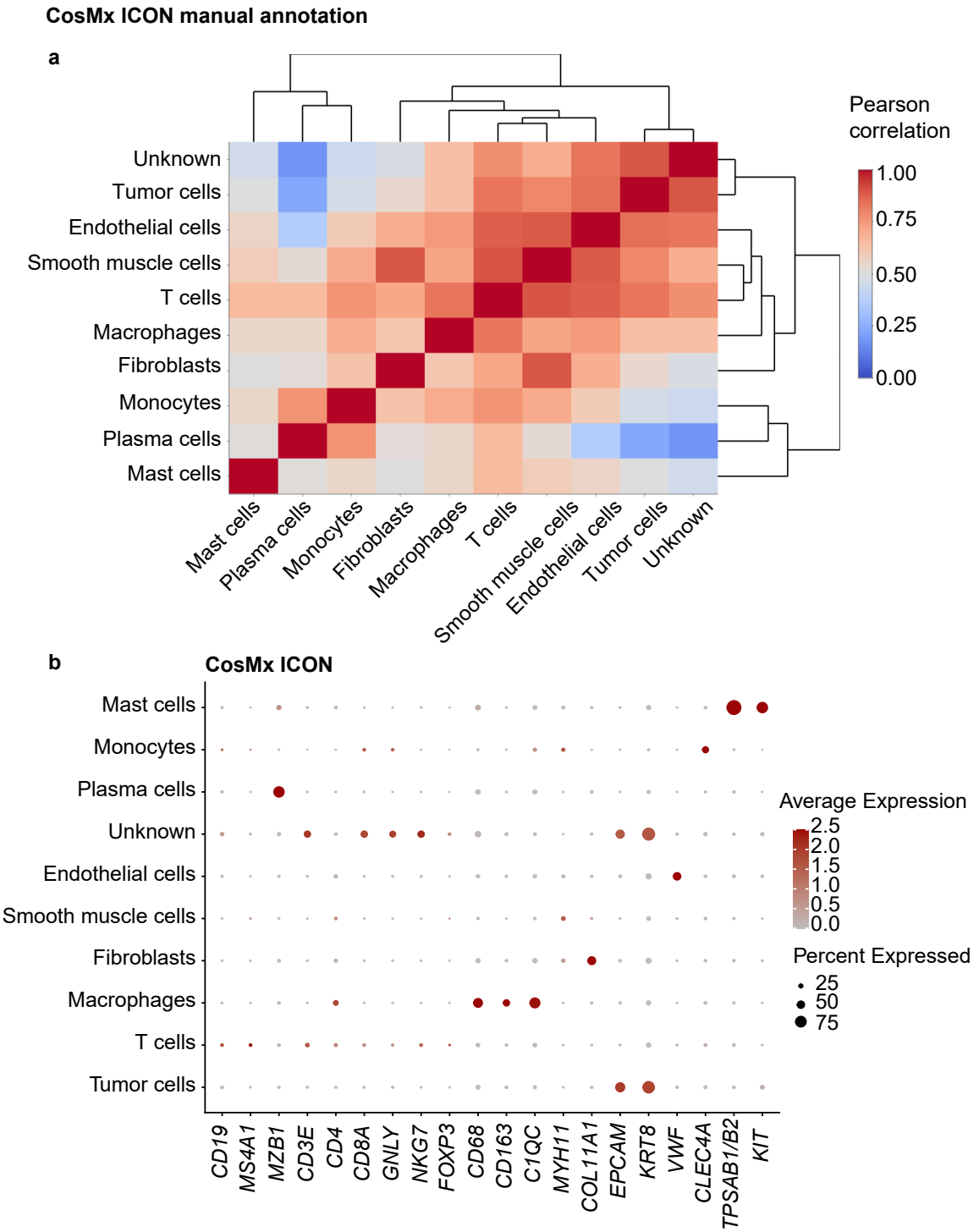
*Authors contributed equally.
#Corresponding authors: C.H. E-mail: chaymaker@mdanderson.org
K.C. E-mail: kchen3@mdanderson.org
L.M.S.S. E-mail: lmsolis@mdanderson.org

The PDF file includes:

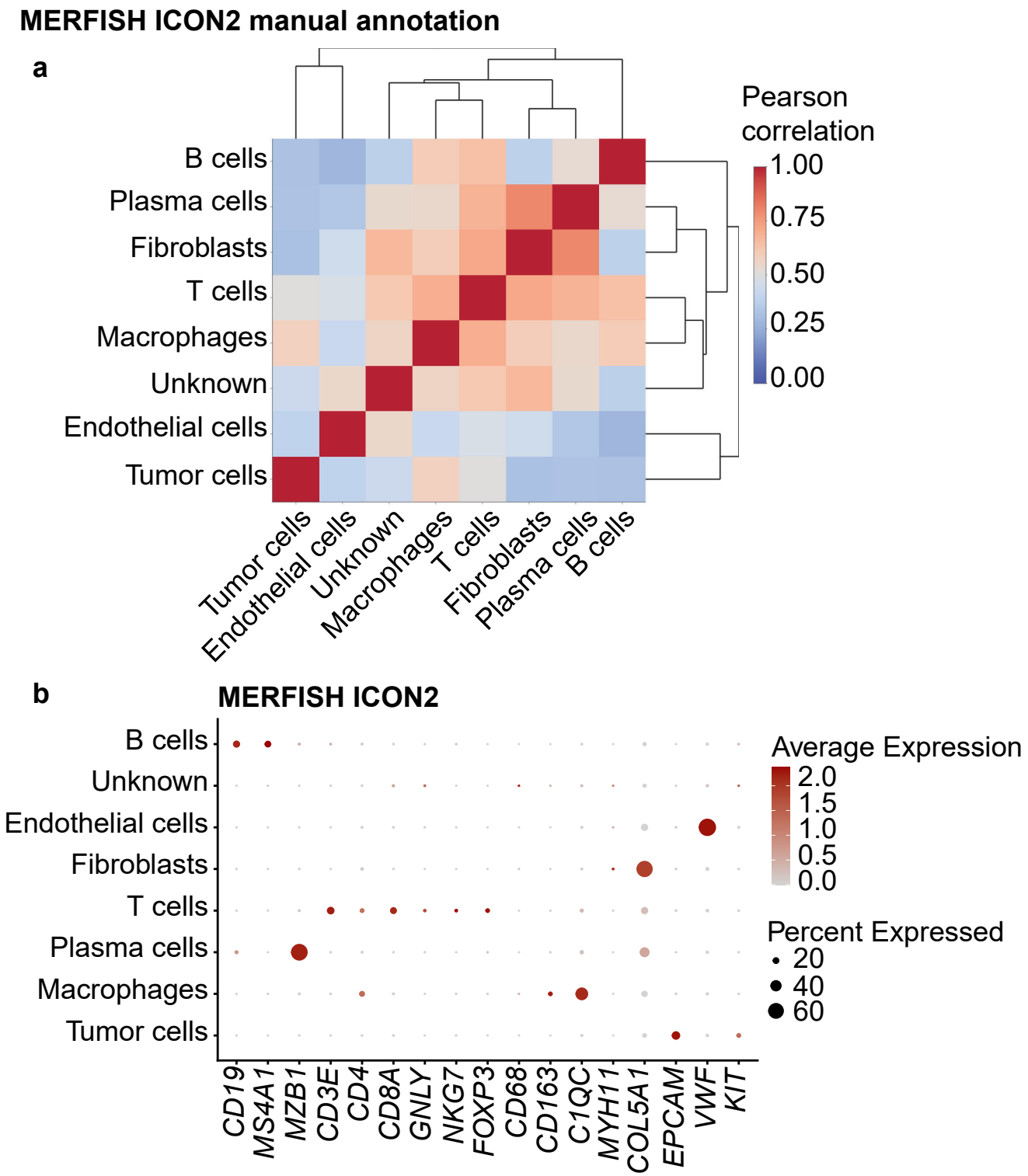
Supplementary Figures



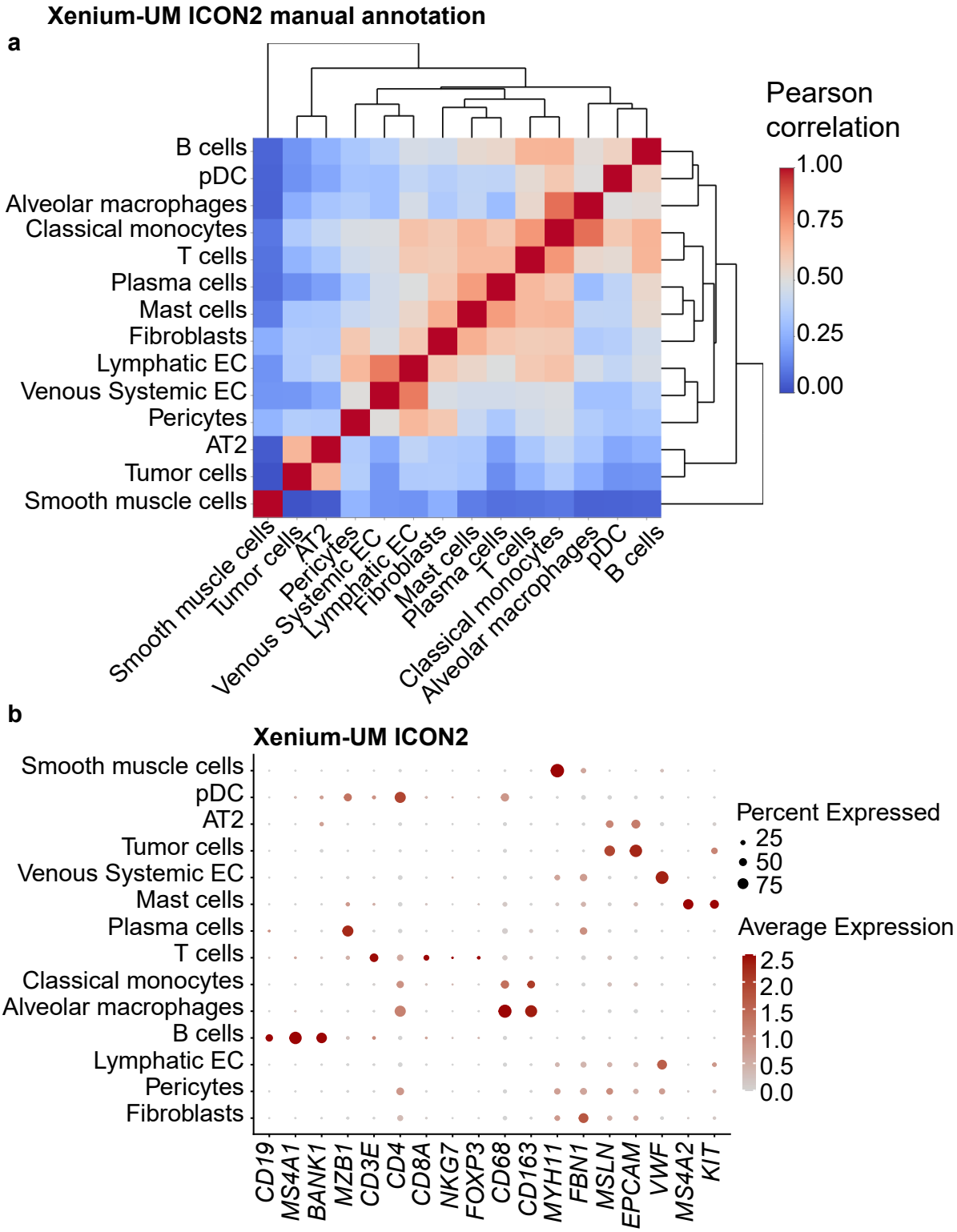
Supplementary Fig. 1: Comparison of co-expressions of disjoint gene pairs. a, Box plots display differences between percentage of cells that co-expressed gene pairs in CosMx (blue) (n = 4 TMAs), MERFISH (purple) (n = 3 TMAs), Xenium-UM (yellow) (n = 3 TMAs) and Xenium-MM (orange) (n = 3 TMAs). The center of the box plot is denoted by the median, a horizontal line dividing the box into two equal halves. The bounds of the box are defined by the lower quartile (25th percentile) and the upper quartile (75th percentile). The whiskers extend from the box and represent the data points that fall within 1.5 times the interquartile range (IQR) from the lower and upper quartiles. Any data point outside this range is considered an outlier and plotted individually. P values of two-sided Wilcoxon signed rank test with paired samples and two-sided Mann Whitney U test with unpaired samples are displayed on images. **b,** Box plots display differences between percentage of cells based on large (blue) and small (purple) cell size differences in CosMx (n = 8 TMAs), MERFISH (n = 6 TMAs), Xenium-UM (n = 6 TMAs) and Xenium-MM (n = 6 TMAs). The center of the box plot is denoted by the median, a horizontal line dividing the box into two equal halves. The bounds of the box are defined by the lower quartile (25th percentile) and the upper quartile (75th percentile). The whiskers extend from the box and represent the data points that fall within 1.5 times the interquartile range (IQR) from the lower and upper quartiles. Any data point outside this range is considered an outlier and plotted individually. Large cell size differences include *EPCAM - CD3E* and *CD68 - CD3E* gene pairs. Small cell size differences include *CD19 - CD3E* and *CD8A - CD4* gene pairs. Source data are provided as a Source Data file.



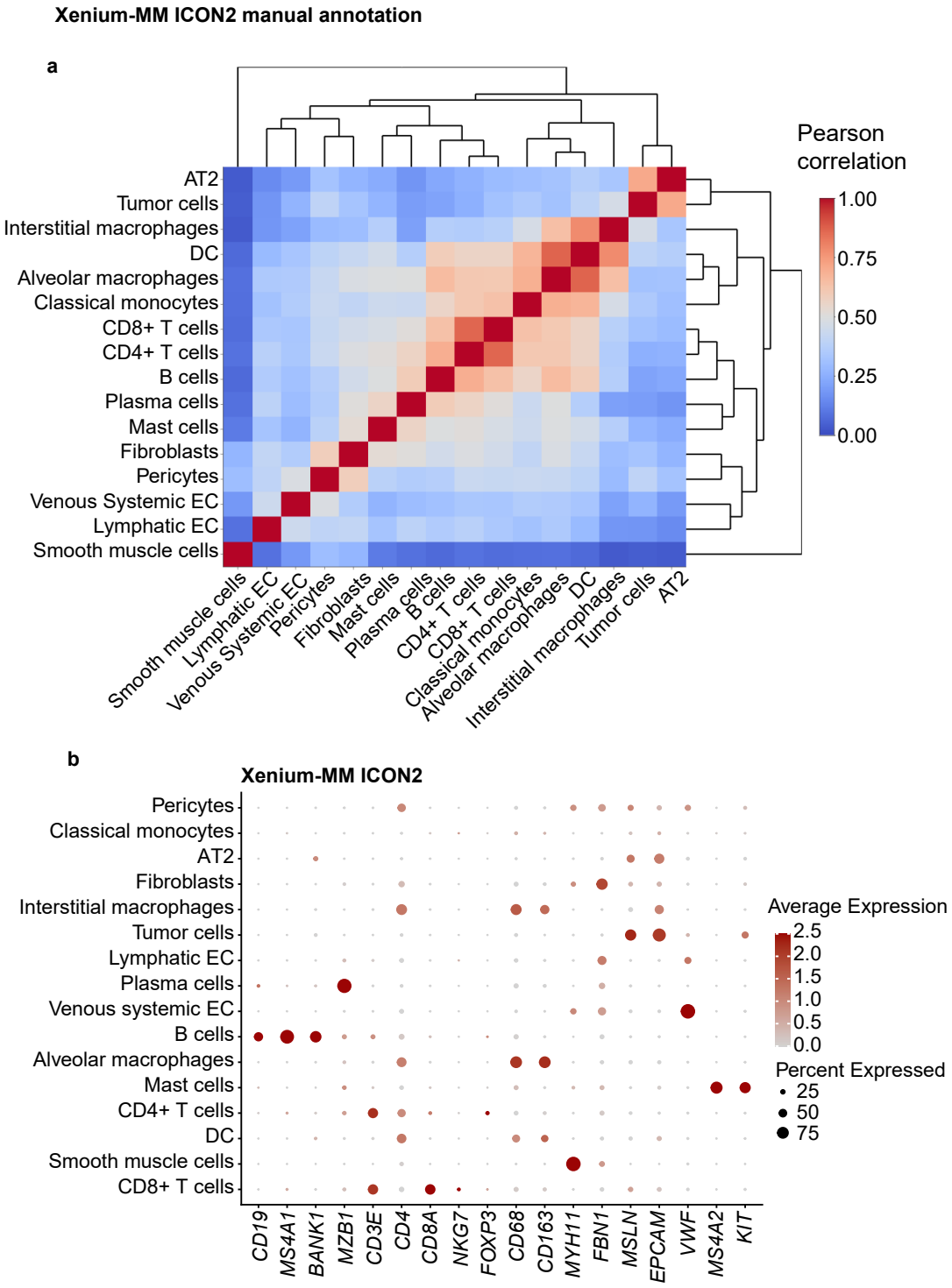
Supplementary Fig. 2: Cell-type annotation for the ICON TMAs on the CosMx platform using a cluster-based manual annotation approach. a, Correlation matrix of the assigned cell types in the ICON TMAs shows high correlation between multiple cell types. Positive Pearson correlations are red and negative Pearson correlations are blue, with intensity representing the strength of the correlation. Hierarchical dendrograms are based on eucledian distance. Clusters were manually annotated with the corresponding cell types based on highly expressed genes in the clusters. **b,** Dot plot displays expression of selected cell marker genes in the assigned cell types of ICON TMAs. High expression of genes are red and low expression of genes are gray, with intensity representing the strength of the expression. High percentage of gene expression in a cell type is big and low percentage of gene expression in a cell type is small size dots, with intensity representing the percentage of the expression.



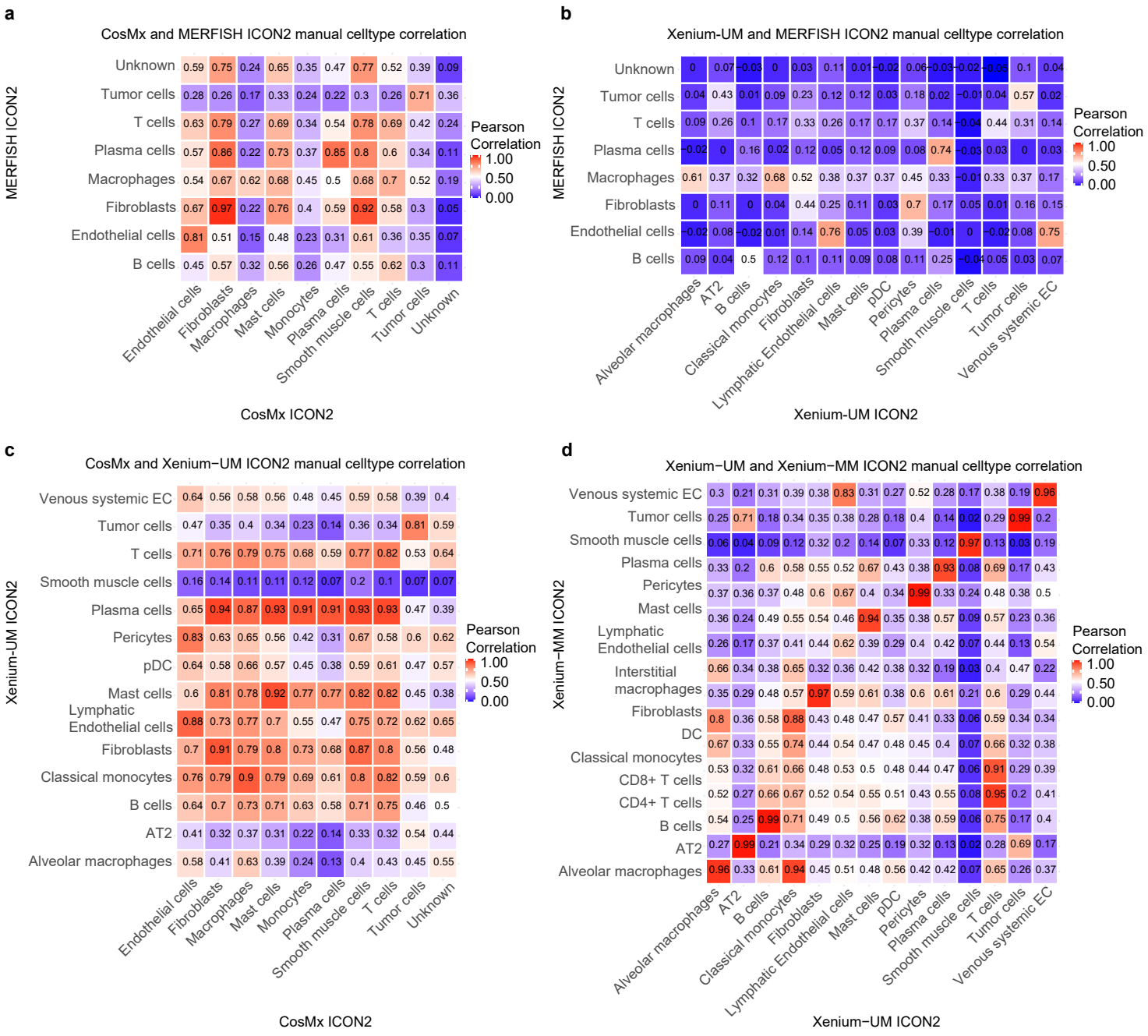
Supplementary Fig. 3: Cell-type annotation for the ICON2 TMA on the MERFISH platform using a cluster-based manual annotation approach. a, Correlation matrix of the assigned cell types in the ICON2 TMA shows high correlation between multiple cell types. Positive Pearson correlations are red and negative Pearson correlations are blue, with intensity representing the strength of the correlation. Hierarchical dendrograms are based on eucledian distance. Clusters were manually annotated with the corresponding cell types based on highly expressed genes in the clusters. **b,** Dot plot displays expression of selected cell marker genes in the assigned cell types of ICON2 TMA. High expression of genes are red and low expression of genes are gray, with intensity representing the strength of the expression. High percentage of gene expression in a cell type is big and low percentage of gene expression in a cell type is small size dots, with intensity representing the percentage of the expression.



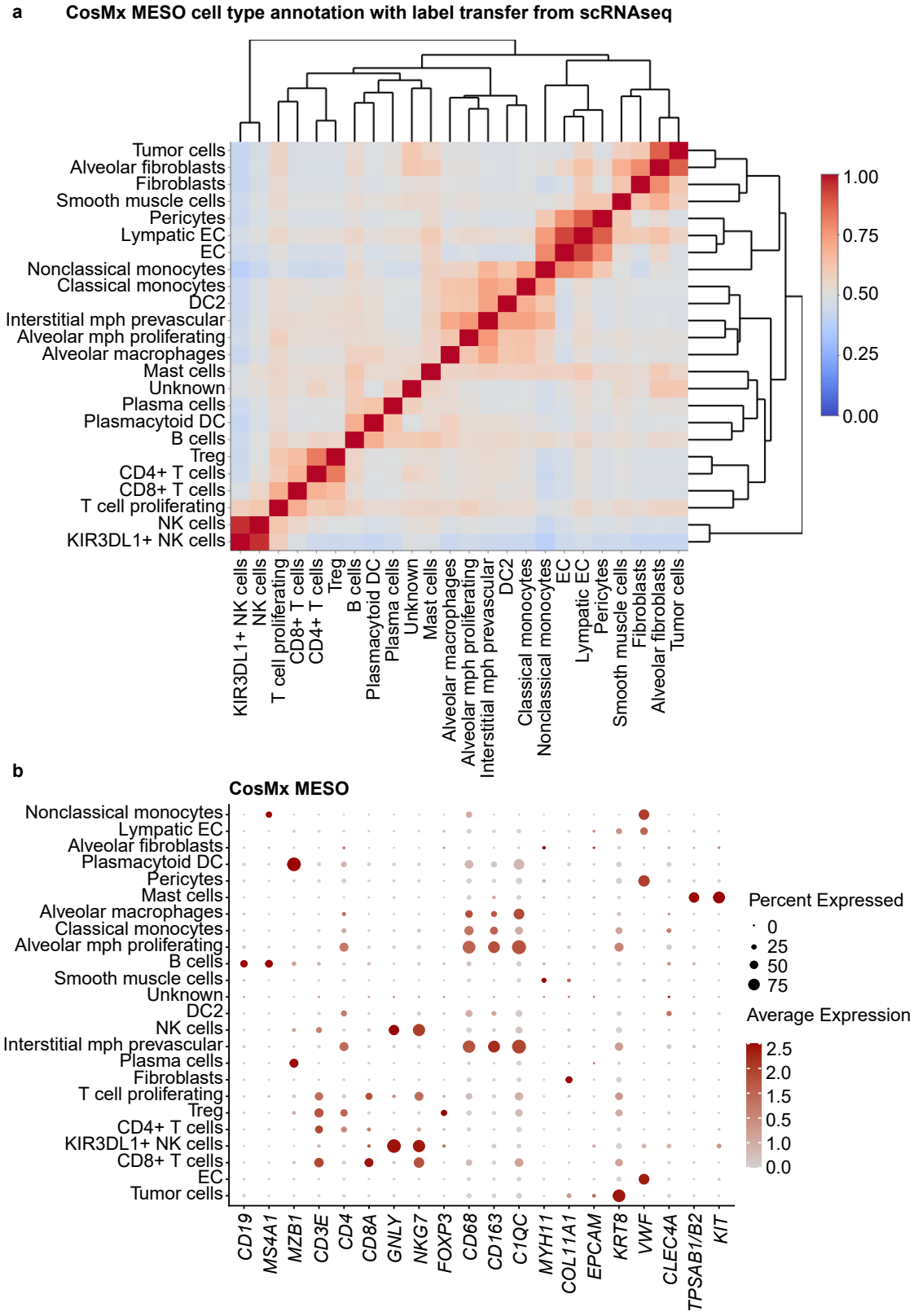
Supplementary Fig. 4: Cell-type annotation for the ICON2 TMA on the Xenium-UM platform using a cluster-based manual annotation approach. a, Correlation matrix of the assigned cell types in the ICON2 TMA shows high correlation between multiple cell types. Positive Pearson correlations are red and negative Pearson correlations are blue, with intensity representing the strength of the correlation. Hierarchical dendrograms are based on eucledian distance. Clusters were manually annotated with the corresponding cell types based on highly expressed genes in the clusters. **b,** Dot plot displays expression of selected cell marker genes in the assigned cell types of ICON2 TMA. High expression of genes are red and low expression of genes are gray, with intensity representing the strength of the expression. High percentage of gene expression in a cell type is big and low percentage of gene expression in a cell type is small size dots, with intensity representing the percentage of the expression.



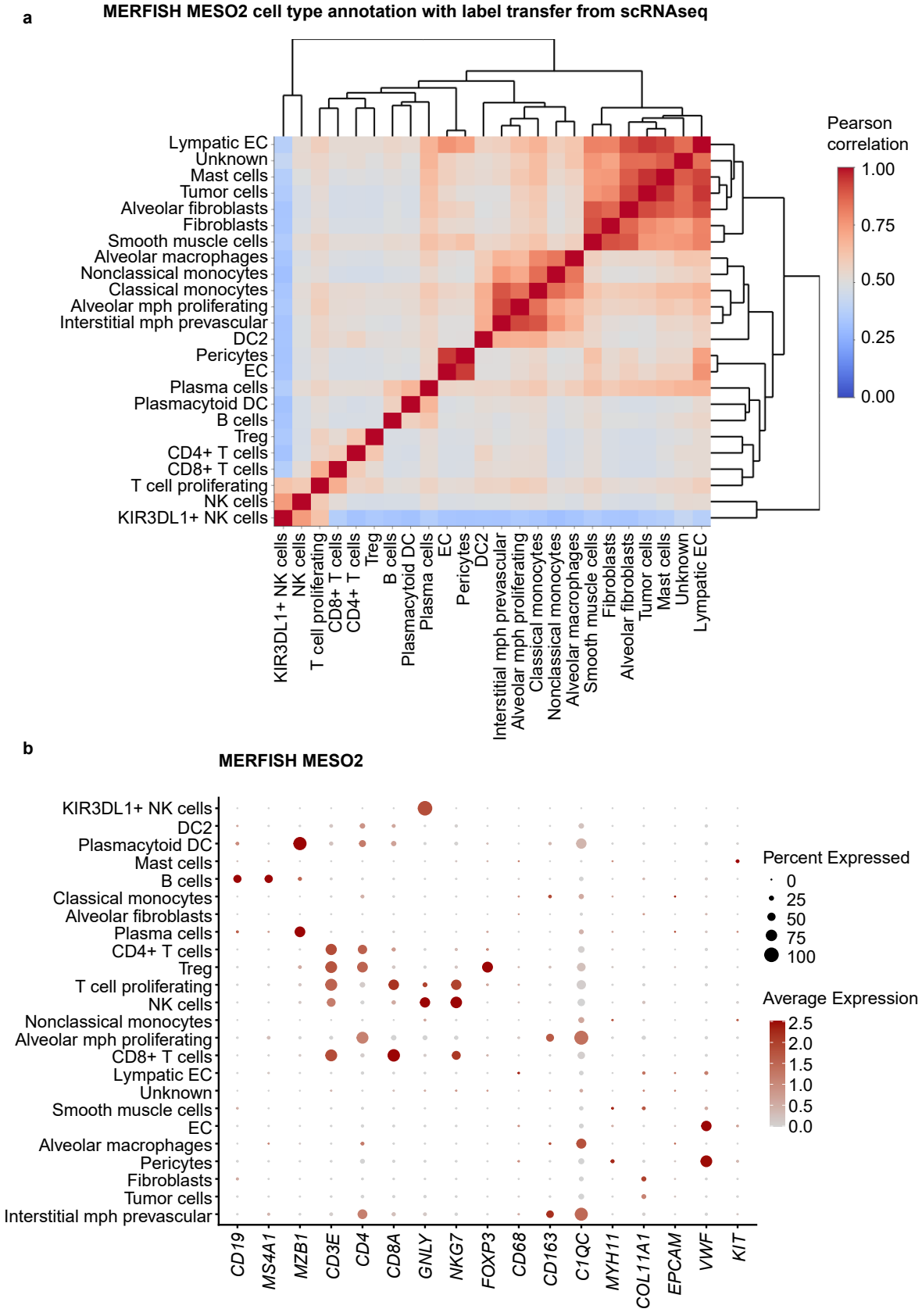
Supplementary Fig. 5: Cell-type annotation for the ICON2 TMA on the Xenium-MM platform using a cluster-based manual annotation approach. a, Correlation matrix of the assigned cell types in the ICON2 TMA shows high correlation between multiple cell types. Positive Pearson correlations are red and negative Pearson correlations are blue, with intensity representing the strength of the correlation. Hierarchical dendrograms are based on eucledian distance. Clusters were manually annotated with the corresponding cell types based on highly expressed genes in the clusters. b, Dot plot displays expression of selected cell marker genes in the assigned cell types of ICON2 TMA. High expression of genes are red and low expression of genes are gray, with intensity representing the strength of the expression. High percentage of gene expression in a cell type is big and low percentage of gene expression in a cell type is small size dots, with intensity representing the percentage of the expression.



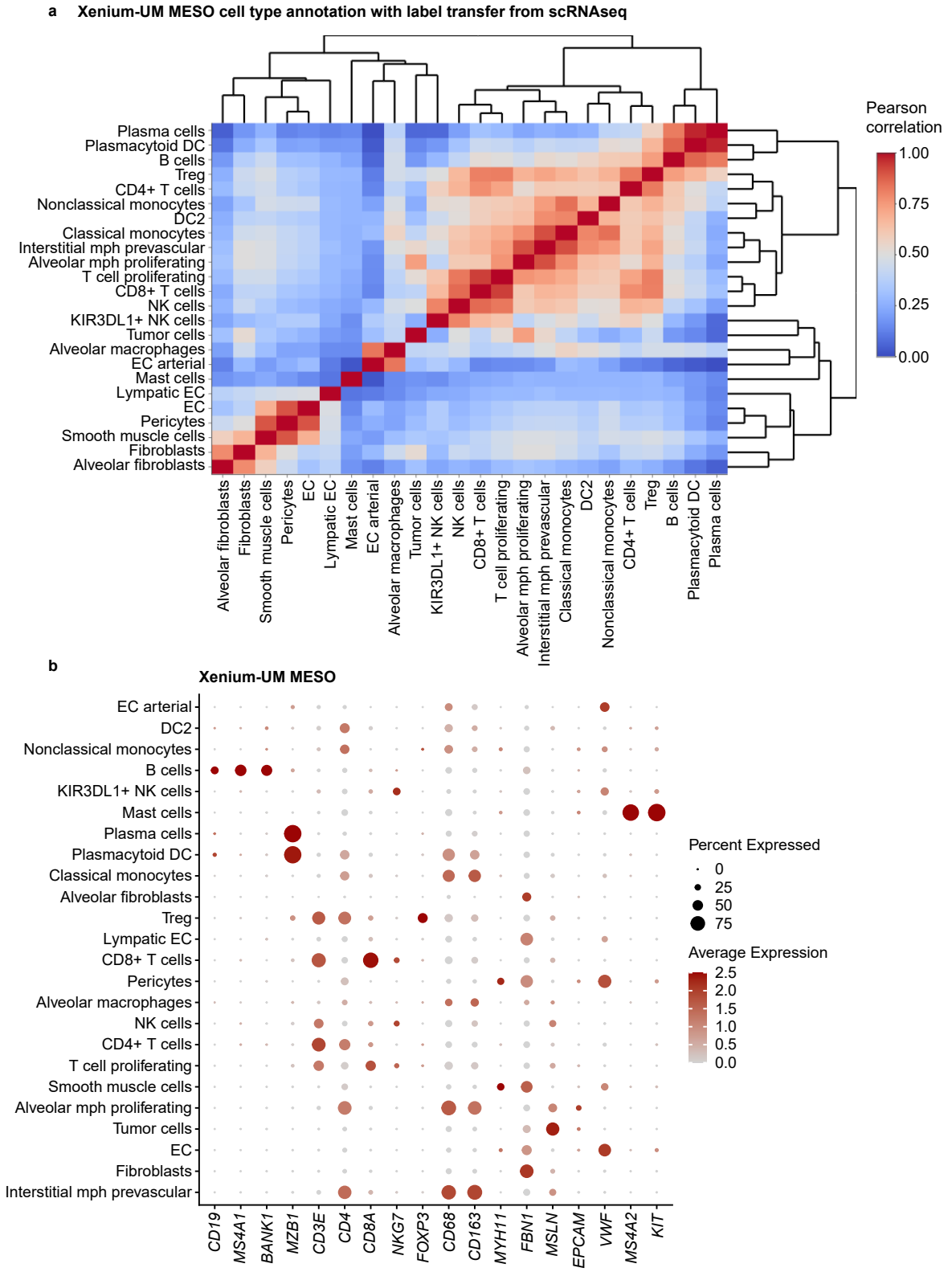
Supplementary Fig. 6: Correlation of annotated cell types between ST platforms in ICON2. Heatmaps show correlation of annotated cell types between **a,** CosMx and MERFISH, **b,** Xenium-UM and MERFISH, **c,** CosMx and Xenium-UM and **d,** Xenium-UM and Xenium-MM of ICON2 based on shared genes. Red, white and blue colors represent high, medium and low Pearson correlations, with intensity representing the strength of the correlation between celltypes, repectively. Correlation coefficients are written on heatmap cells. Source data are provided as a Source Data file.



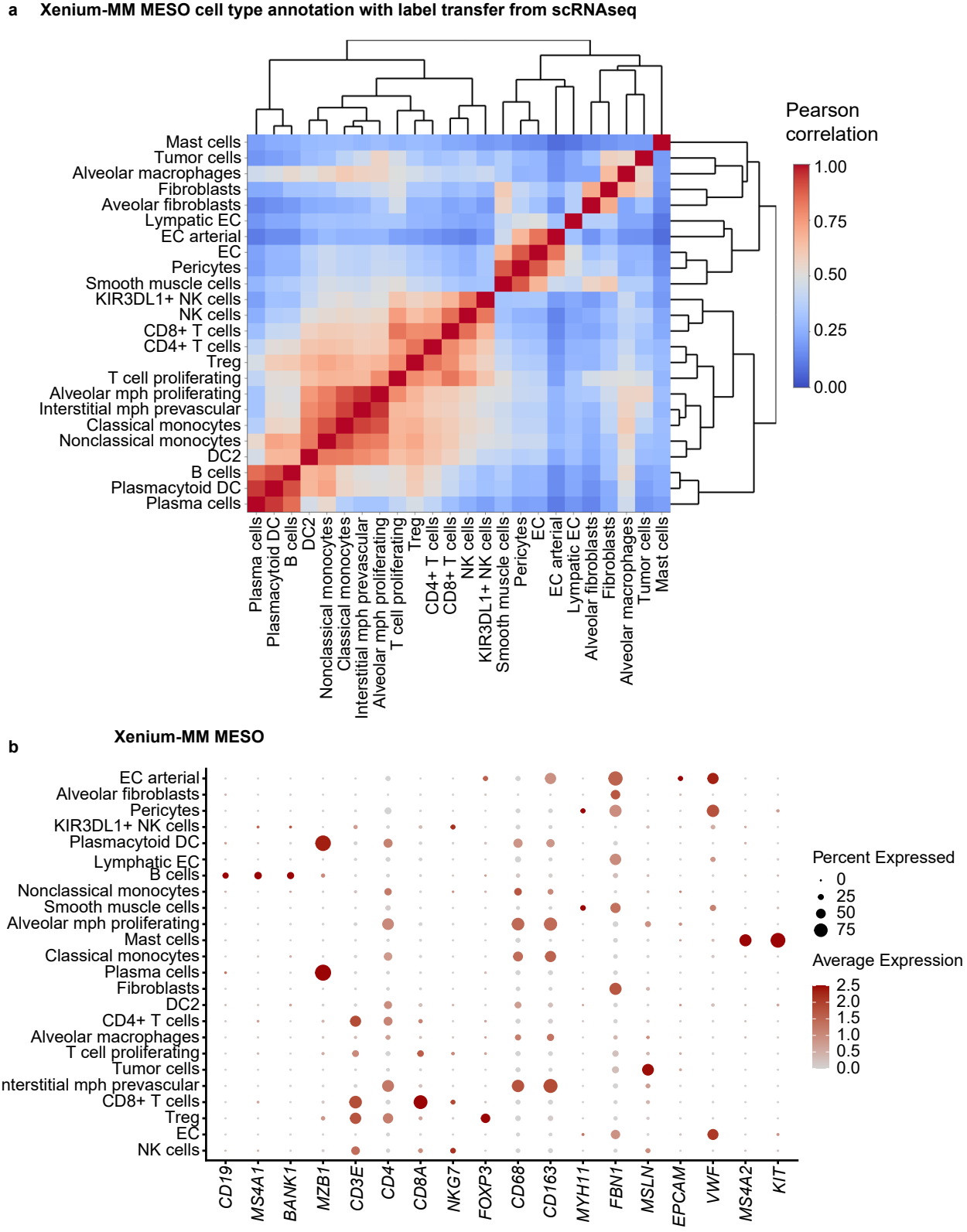
Supplementary Fig. 7: Cell-type annotation for the MESO TMAs on the CosMx platform using a transfer label approach from scRNA-seq. a, Correlation matrix of the assigned cell types in the MESO TMAs shows high correlation between multiple cell types. Positive Pearson correlations are red and negative Pearson correlations are blue, with intensity representing the strength of the correlation. Hierarchical dendrograms are based on eucledian distance. Clusters were manually annotated with the corresponding cell types based on highly expressed genes in the clusters. **b,** Dot plot displays expression of selected cell marker genes in the assigned cell types of MESO TMAs. High expression of genes are red and low expression of genes are gray, with intensity representing the strength of the expression. High percentage of gene expression in a cell type is big and low percentage of gene expression in a cell type is small size dots, with intensity representing the percentage of the expression.



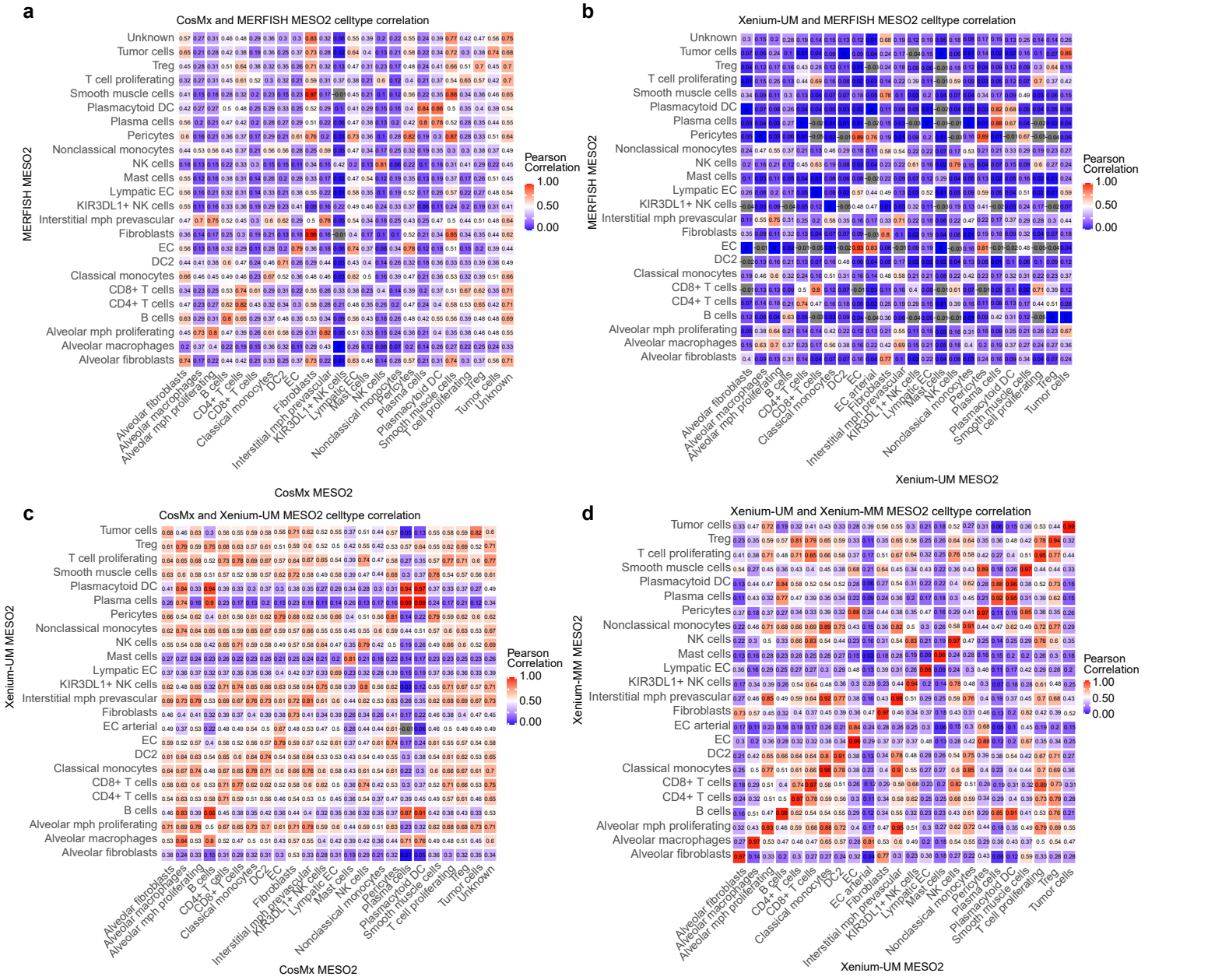
Supplementary Fig. 8: Cell-type annotation for the MESO2 TMA on the MERFISH platform using transferring label approach from scRNA-seq. a, Correlation matrix of the assigned cell types in the MESO2 TMA shows high correlation between multiple cell types. Positive Pearson correlations are red and negative Pearson correlations are blue, with intensity representing the strength of the correlation. Hierarchical dendrograms are based on eucledian distance. Clusters were manually annotated with the corresponding cell types based on highly expressed genes in the clusters. **b,** Dot plot displays expression of selected cell marker genes in the assigned cell types of MESO2 TMA. High expression of genes are red and low expression of genes are gray, with intensity representing the strength of the expression. High percentage of gene expression in a cell type is big and low percentage of gene expression in a cell type is small size dots, with intensity representing the percentage of the expression.



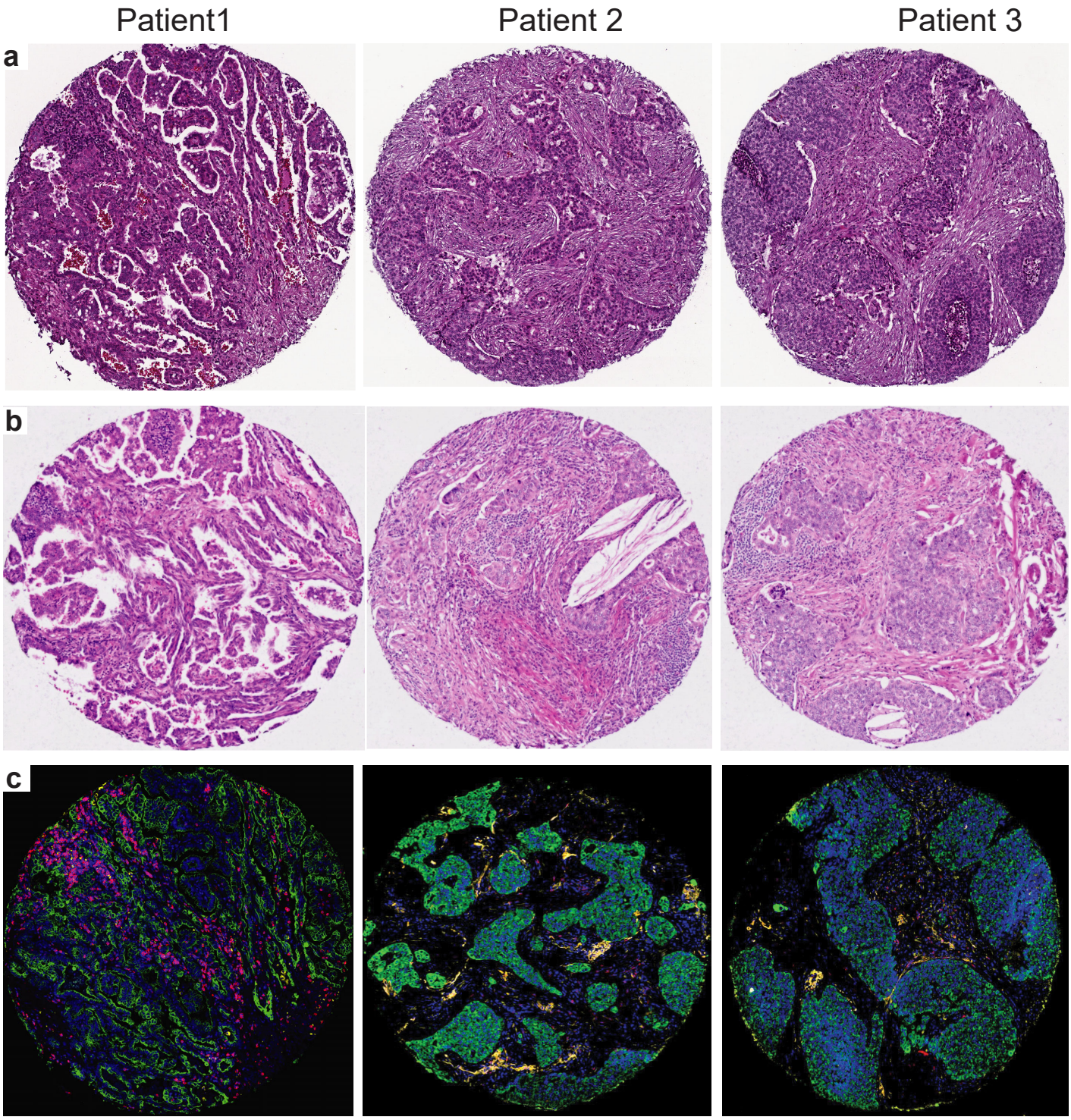
Supplementary Fig. 9: Cell-type annotation for the MESO TMAs on the Xenium-UM platform using a transfer label approach from scRNA-seq. a, Correlation matrix of the assigned cell types in the MESO TMAs shows high correlation between multiple cell types. Positive Pearson correlations are red and negative Pearson correlations are blue, with intensity representing the strength of the correlation. Hierarchical dendrograms are based on eucledian distance. Clusters were manually annotated with the corresponding cell types based on highly expressed genes in the clusters. **b,** Dot plot displays expression of selected cell marker genes in the assigned cell types of MESO TMAs. High expression of genes are red and low expression of genes are gray, with intensity representing the strength of the expression. High percentage of gene expression in a cell type is big and low percentage of gene expression in a cell type is small size dots, with intensity representing the percentage of the expression.



Supplementary Fig. 10: Cell-type annotation for the MESO TMAs on the Xenium-MM platform using a transfer label approach from scRNA-seq. a, Correlation matrix of the assigned cell types in the MESO TMAs shows high correlation between multiple cell types. Positive Pearson correlations are red and negative Pearson correlations are blue, with intensity representing the strength of the correlation. Hierarchical dendrograms are based on eucledian distance. Clusters were manually annotated with the corresponding cell types based on highly expressed genes in the clusters. **b,** Dot plot displays expression of selected cell marker genes in the assigned cell types of MESO TMAs. High expression of genes are red and low expression of genes are gray, with intensity representing the strength of the expression. High percentage of gene expression in a cell type is big and low percentage of gene expression in a cell type is small size dots, with intensity representing the percentage of the expression.



Supplementary Fig. 11: Correlation of annotated cell types between ST platforms in MESO2. Heatmaps show correlation of annotated cell types between **a,** CosMx and MERFISH, **b,** Xenium-UM and MERFISH, **c,** CosMx and Xenium-UM and **d,** Xenium-UM and Xenium-MM of MESO2 based on shared genes. Red, white and blue colors represent high, medium and low Pearson correlations, with intensity representing the strength of the correlation between celltypes, repectively. Correlation coefficients are written on heatmap cells. Source data are provided as a Source Data file.



Supplementary Figure 12: Strategy of pathologists' evaluation on ICON2. Microphotographs showing H&E **a,** reference H&E, **b,** Xenium-MM H&E staining slides, **c,** multiplex Immunofluorescence (mIF) stained with CK (green), CD3 (red) and CD68 (yellow) performed in a serial section. We evaluated the pattern of expression of endothelial cells, fibroblasts, macrophages, plasma cells, T cells and epithelial cells, and compared them with the patterns of the phenotypes obtained for each platform in ICON2. Reference H&E images were scanned at 40x in a Leica AT2 system, using standard bright-field settings. mIF images were captured by the GeoMX DSP system, using 3 channels + DAPI. PanCK (green, Alexa Fluor 532, panCK Clone AE1+AE3), CD3E (red, Alexa Fluor 647, Clone UMAB54), and CD68, (yellow, Alexa Fluor 534, clone KP1), all antibodies in a 0.5 µg/ml dilution. Visualization settings adjusted inside the HALO Digital Image Analysis software (version 3.4, Indica Labs) adjusting only the thresholds for black and white by signal intensity and adjusted for each whole core. Xenium-MM H&E imaging was provided by 10X Genomics with bright-field scanning performed in the Xenium analyzer, on the same slide of the Xenium-MM assay after Xenium assay processing.