Cancer Immune Monitoring and Analysis Center
Precision Immunology Institute
Icahn School of Medicine at Mount Sinai (ISMMS)
Contact PI: Sacha Gnjatic, PhD (Sacha.Gnjatic@mssm.edu)

Performance Lab:
Human Immune Monitoring Center (HIMC)
Miriam Merad, MD, PhD, Director
Sacha Gnjatic, PhD, Professor and Co-Director
Seunghee Kim-Schulze, PhD, Facility Director
1470 Madison Avenue, New York, NY 10028

Assay type	Primary assay outputs	Pre-processing/ Normalization /QC	Initial analyses	Derived data outputs
Spatial RNAseq	~20,000+ gene quantifications and ~5,000 cells per sample.	HIMC Sample QC Sequencing QC Analysis QC	Pathology annotation, Differential gene expression, molecular pathway enrichment.	Gene expression per cluster/cell type per sample.

1. Purpose of assay

High-dimensional single cell monitoring is a key strategy to elucidate complex phenotypic and functional characteristics of heterogeneous immune populations. This is a broadly applicable approach that can provide valuable insights into disease mechanisms and therapeutic responses and potentially identify correlative cellular biomarkers in the context of many different trials. However, existing single cell profiling methods require single cell suspensions purified from the biological source of interest. That leads to tissue disintegration and loss of invaluable information of single cell surroundings, cell-to-cell interactions and large tissue architecture in general. To gain insight into expression programs and functional states of cells inside the preserved tissues at nearly single-cell resolution, several spatial transcriptome profiling technologies have been developed including Visium (10X Genomics).

Visium is a molecular profiling solution which enables whole transcriptome profiling of cells located within Æ 55 mm spots within the morphological context of interest. The technology has been optimized for fresh frozen and FFPE embedded tissues broadening the platform applicability for a diverse of biological context and tissue types. To maximize the resolution, Visium takes advantage of unique molecular identifiers (UMI) as a strategy to barcode individual molecules of mRNA at the point of reverse transcription. To encode the spatial information within sequencing data, the technology uses spatial barcodes uniquely allocated to each spot on the tissue slide. On the experimental side, Visium protocols include multiple QC steps to ensure RNA integrity and high quality of the sequencing library. Finally, FFPE-tailored Visium assay operates with a pre-designed set of probes (targeting up to 20,000 genes) validated to anneal to RNA molecules of interest within paraffin-embedded tissues. Ultimately, Visium enables expression profiling of thousands of genes per tissue within its morphological environment at a 10-20 cell resolution (**Figure 1**).

The MS-CIMAC has invested heavily to establish a comprehensive Visium RNAseq pipeline led by Dr. Seunghee Kim-Schulze. It includes optimized sample processing SOPs and an efficient data processing pipeline. As one of the most active spatial transcriptomic programs in the country.

2. Manufacture's Validation

Visium Spatial Gene Expression for FFPE overcomes the challenges imposed by formalin-fixation and paraffin-embedding for transcriptomic analysis, to enable unbiased whole transcriptome analysis of FFPE tissue sections. Below is a list of tissue types that were tested by 10x and found to perform well with Visium for FFPE. Tissues were acquired from biobanks.

Organism	Tissue	Unknown, Healthy, Diseased	Region, if known
Human	Brain	Unknown	Cerebellum

Organism	Tissue	Unknown, Healthy, Diseased	Region, if known
Human	Brain	Healthy	
Human	Brain	Diseased, Glioblastoma	
Human	Breast	Healthy	
Human	Breast	Diseased, Cancer	
Human	Cervix	Diseased, Cancer	
Human	Intestine	Diseased, Colorectal Cancer	
Human	Spine	Diseased, Glioblastoma	
Human	Heart	Healthy	
Human	Heart	Diseased, Myxoma	
Human	Heart	Diseased, Sarcoma	
Human	Kidney	Healthy	
Human	Large Intestine	Healthy	
Human	Large Intestine	Diseased, Cancer	
Human	Lung	Healthy	
Human	Lung	Diseased, Cancer	
Human	Lymph Node	Healthy	
Human	Lymph Node	Diseased, Inflamed	
Human	Lymph Node	Diseased, Lymphoma	
Human	Ovary	Diseased, Cancer	
Human	Prostate	Healthy	
Human	Prostate	Diseased, Cancer	
Human	Skin	Diseased, Malignant Melanoma	
Human	Skin	Diseased, Squamous Cell Carcinoma	
Human	Skin	Diseased, Merkel Cell Carcinoma	
Human	Spinal Cord	Unknown	
Human	Spleen	Healthy	
Mouse	Brain	Healthy	

Organism	Tissue	Unknown, Healthy, Diseased	Region, if known
Mouse	Brain	Healthy	Cerebellum
Mouse	Brain	Healthy	Olfactory bulb
Mouse	Heart	Healthy	
Mouse	Kidney	Healthy	
Mouse	Liver	Healthy	
Mouse	Spleen	Healthy	
Mouse	Testis	Healthy	
Mouse	Thymus	Healthy	
Mouse	Embryo	Healthy	
Mouse	Eye	Healthy	
Mouse	Muscle	Healthy	Quadricep
Mouse	Ovary	Healthy	
Mouse	Placenta	Healthy	Embryonic (E17)
Mouse	Small Intestine	Healthy	
Mouse	Skin	Healthy	
Mouse	Tongue	Healthy	

Below is a list of tissue types tested by 10x with Visium for FFPE that were found to not perform well. Notes on why these samples were challenging are listed.

Organism	Tissue	Unknown, Healthy, Diseased	Additional Information
Human	Adipose	Unknown	Low yield / Poor sensitivity
Human	Tonsil	Diseased, Reactive Follicular Hyperplasia	Inconsistent permeabilization efficiency
Human	Skin	Normal	High propensity for detachment
Mouse	Bone/Cartilage	Healthy	Requires further optimization

3. Icahn School of Medicine at Mount Sinai Human Immune Monitoring Center Experience and Capacity

MS-CIMAC relies on the experience from ISMMS' Human Immune Monitoring Center (HIMC) which has been performing Visium 10X spatial transcriptomic sequencing since the manufacturer made the method available commercially. Dozens of samples have been analyzed to date, with a strict SOPs and processes for sample handling, sequencing of libraries, and data analysis, as summarized below in Table 1 and Figure 1.

Sample processing	Library prep and sequencing	Data processing and analysis
Strict SOPs and detailed sample processing logs	qPCR validation of library contents and quality	Spaceranger pipeline for read mapping, debarcoding and read counting
Optimized tissue slicing and sample QC (DV200)	Bioanalyzer/Tapestation validation of insert size (Figure TD1)	Filtering of low UMI events, mitochondrial RNA events
Optimized H&E staining and image acquisition	Optimized sequencing protocol for scRNAseq libraries	Integration with pathology annotation, clustering, gene DE analysis, pathway enrichment and single cell deconvolution

Table 1. Key steps of the Visium pipeline established at HIMC, Mount Sinai.

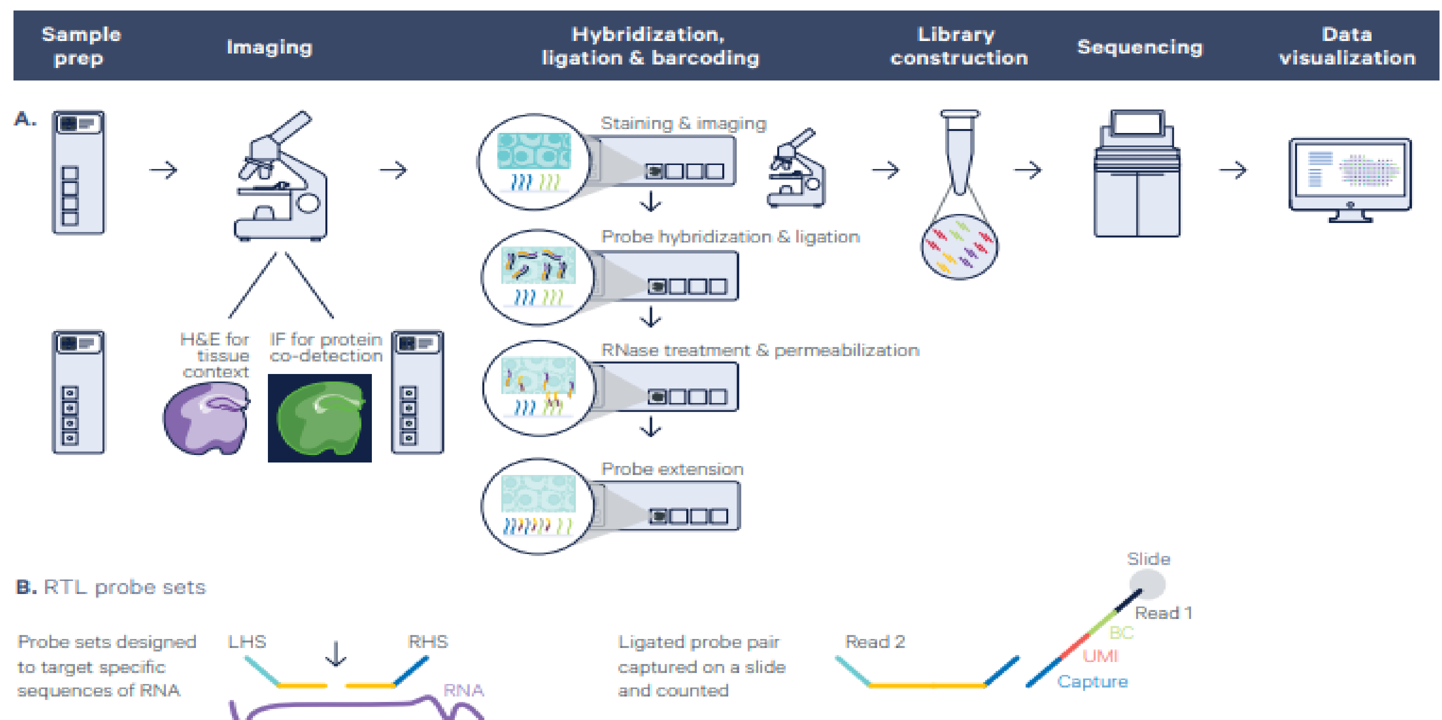
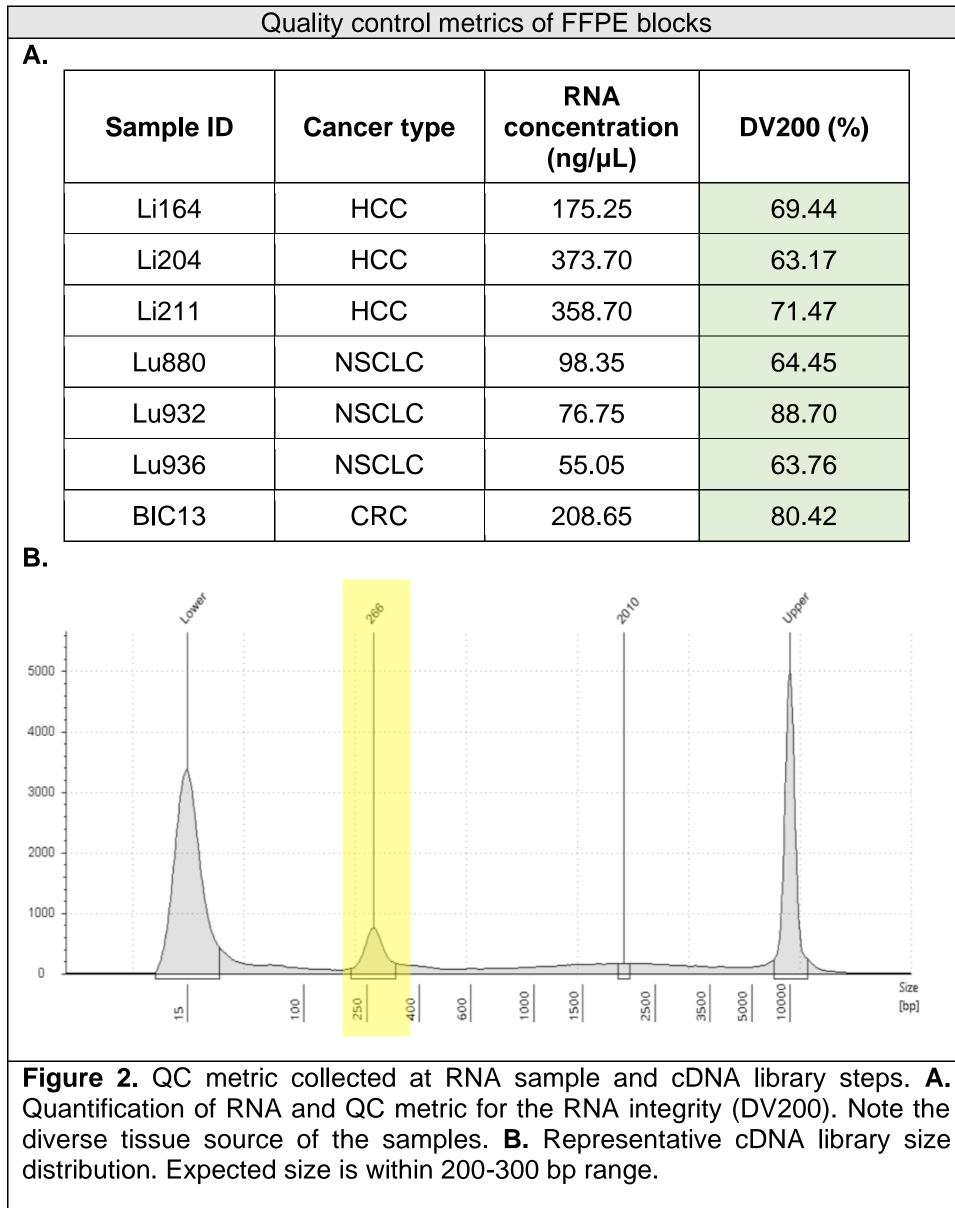


Figure 1: Visium FFPE workflow. (A) Experimental steps including H&E staining and tissue imaging, probe hybridization, library construction and sequencing. (B) Molecular events of probe hybridization and library construction.

3.1. Initial Sample Quality Control (QC) and library QC metrics

Types of biological samples accepted: The 10X Genomics FFPE Visium Gene expression assay is compatible with most Formalin-fixed paraffin-embedded tissue sections.

FFPE blocks requirements: The initial block quality is very important for the success of the 10X Genomics FFPE Visium Gene expression assay. It is essential to assess RNA quality of the tissue block before proceeding with sectioning by calculating the percentage of total RNA fragments >200 nucleotides (DV200) of RNA extracted from tissue sections. FFPE tissue blocks (up to 3 months old) with DV200 greater than 50% can be used for the assay (See Figure 2). Recommended section thickness is 5 µm.



Quality control metrics of FFPE blocks

A.

Sample ID	Cancer type	RNA concentration (ng/µL)	DV200 (%)
Li164	HCC	175.25	69.44
Li204	HCC	373.70	63.17
Li211	HCC	358.70	71.47
Lu880	NSCLC	98.35	64.45
Lu932	NSCLC	76.75	88.70
Lu936	NSCLC	55.05	63.76
BIC13	CRC	208.65	80.42

B.

Figure 2. QC metric collected at RNA sample and cDNA library steps. **A.** Quantification of RNA and QC metric for the RNA integrity (DV200). Note the diverse tissue source of the samples. **B.** Representative cDNA library size distribution. Expected size is within 200-300 bp range.

3.2. Sequencing and read mapping QC metrics

Sequencing results are QC'ed using the SpaceRanger 10X genomics pipeline, following manufacturer recommendations. Briefly, several metrics indicating the quality of the sequencing are collected (**Figure 3**). High quality sample is expected to have: mean reads per spot under the tissue > 25,000; valid barcodes, UMIs and Q30 Bases in Barcode and UMI > 95%; Fraction of reads in spots under the tissue – measure of lateral diffusion of RNA ~ 90%. As shown on representative sample, the established experimental procedures are allowing to obtain sequencing results of excellent quality (**Figure 3**).

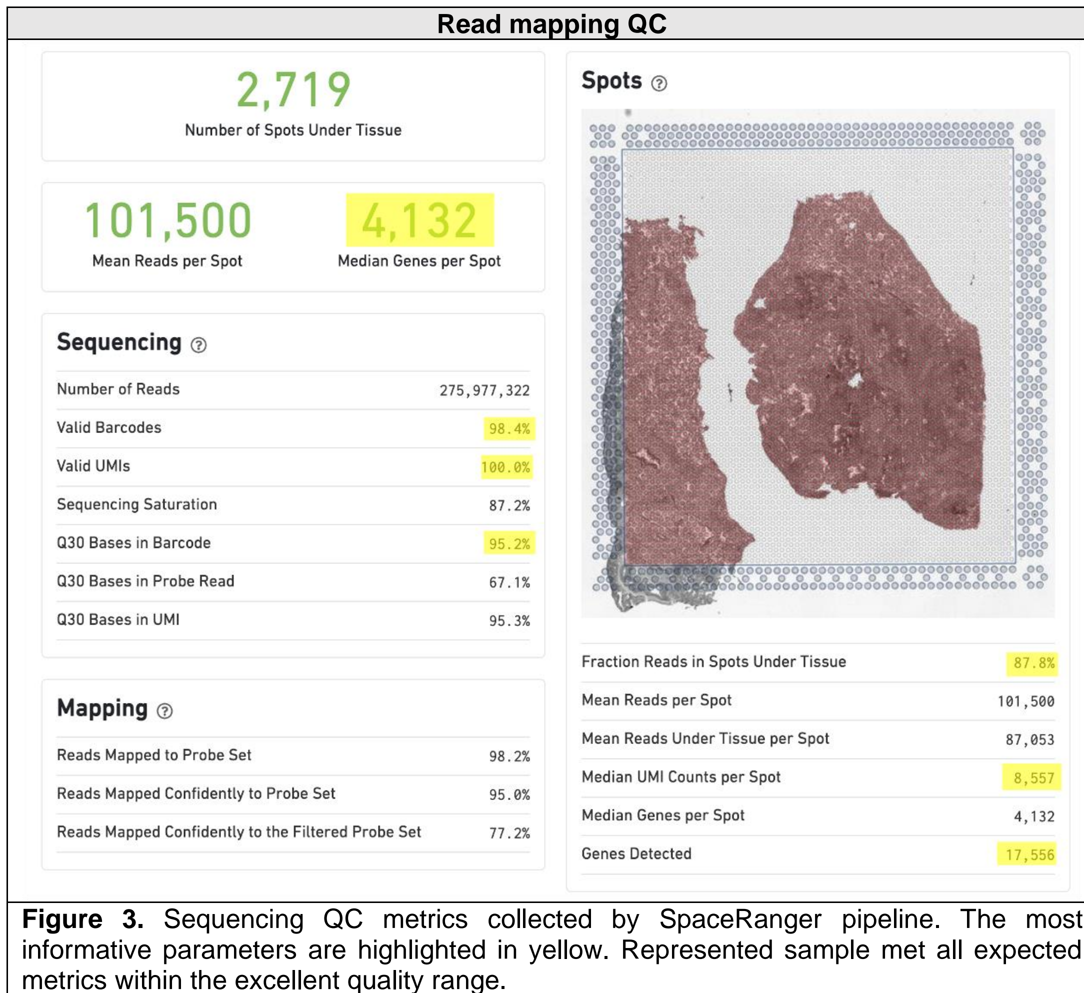


Figure 3. Sequencing QC metrics collected by SpaceRanger pipeline. The most informative parameters are highlighted in yellow. Represented sample met all expected metrics within the excellent quality range.

Sequencing metrics	Lu936_1	Lu936_2	Expected value
Number of Reads	275,977,322	216,740,984	Sequencing output dependent
Median Genes per Spot	4,132	3,814	Dependent on tissue type, RNA quality, and sequencing depth
Median UMI Counts per Spot	8,557	7,532	Dependent on tissue type, RNA quality, and sequencing depth
Genes Detected	17,556	17,604	Dependent on tissue type, RNA quality, and sequencing depth
Reads Mapped to Probe Set	98.2%	98.2%	Ideally > 50%
Valid barcodes	98.4%	98.4	> 75%
Valid UMI	100%	100%	> 75%
Fraction reads in spots under tissue	87.8%	89.5%	Ideally > 50%

3.3. Validation of gene expression reproducibility

To validate Visium transcriptome reproducibility, lung tissue block from one human was used to prepare two consecutive (e.g. “sister”) cuts. Two Visium transcriptome libraries were generated and sequenced on the same NovaSeq S1 sequencer. We performed two correlative analyses: Pearson correlation at the level of normalized gene expression across all sampled tissue and spearman correlation score of transcriptomes between matched Visium spots after affine registration of two H&E images. As shown on **Figure 4A**, overall gene expression from consecutive slides is highly reproducible reaching Pearson correlation > 0.99. The list of top 10 expressed genes is also nearly identical. More granular analysis of correlation between

matched spots after affine registration of the major tissue slices. We detected high level of expression concordance in areas with high cellularity and low concordance – in areas with low cellular density. (Spearman coefficient, **Figure 4B**, cell density annotation: **Figure 5A**). Importantly, upon integration of Visium cell deconvolution pipelines with pathology annotations, we detected that cellular density is a parameter affecting estimated of number of cells and diversity per spot (**Figure 5B**).

Sample reproducibility between consecutive tissue cuts

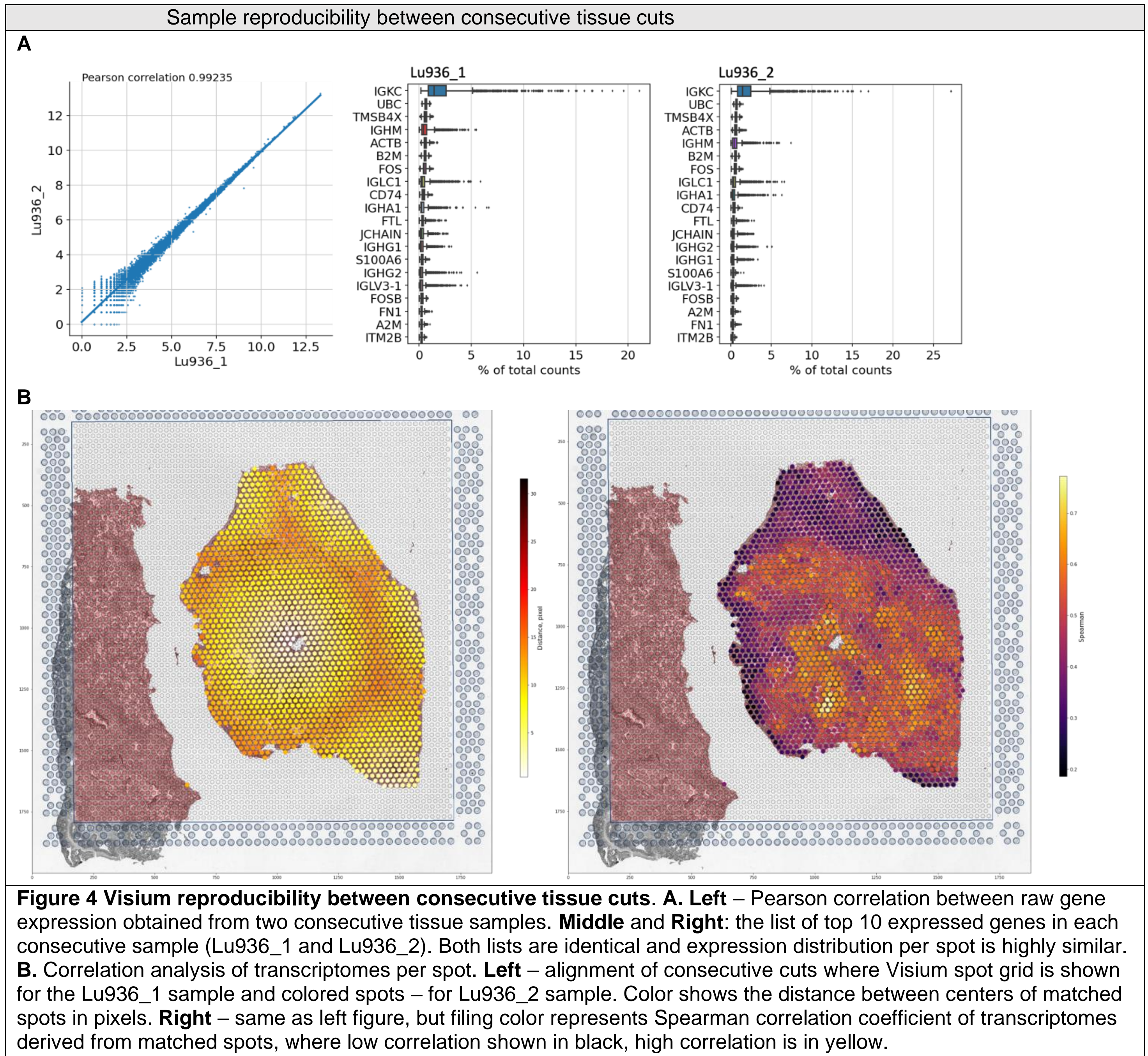


Figure 4 Visium reproducibility between consecutive tissue cuts. **A. Left** – Pearson correlation between raw gene expression obtained from two consecutive tissue samples. **Middle** and **Right**: the list of top 10 expressed genes in each consecutive sample (Lu936_1 and Lu936_2). Both lists are identical and expression distribution per spot is highly similar. **B.** Correlation analysis of transcriptomes per spot. **Left** – alignment of consecutive cuts where Visium spot grid is shown for the Lu936_1 sample and colored spots – for Lu936_2 sample. Color shows the distance between centers of matched spots in pixels. **Right** – same as left figure, but filing color represents Spearman correlation coefficient of transcriptomes derived from matched spots, where low correlation shown in black, high correlation is in yellow.

3.4 Validation of biological correlates and differential gene expression (DE) analysis

FFPE tissue blocks allow for (1) 5 um-thick tissue cuts and (2) high-contrast H&E staining – both being reproducibly established as part of HIMC experimental pipeline. H&E-stained images are annotated for major tissue areas (non-tumor, tumor, stroma, and immune aggregates) and cellular density (low and high) by the professional pathologist manual observations. Manually annotated and colored representations are integrated with Visium coordinate space to allow seamless comparison of Visium transcriptome information with tissue morphology (**Figure 5A**).

Pathology annotation of H&E staining

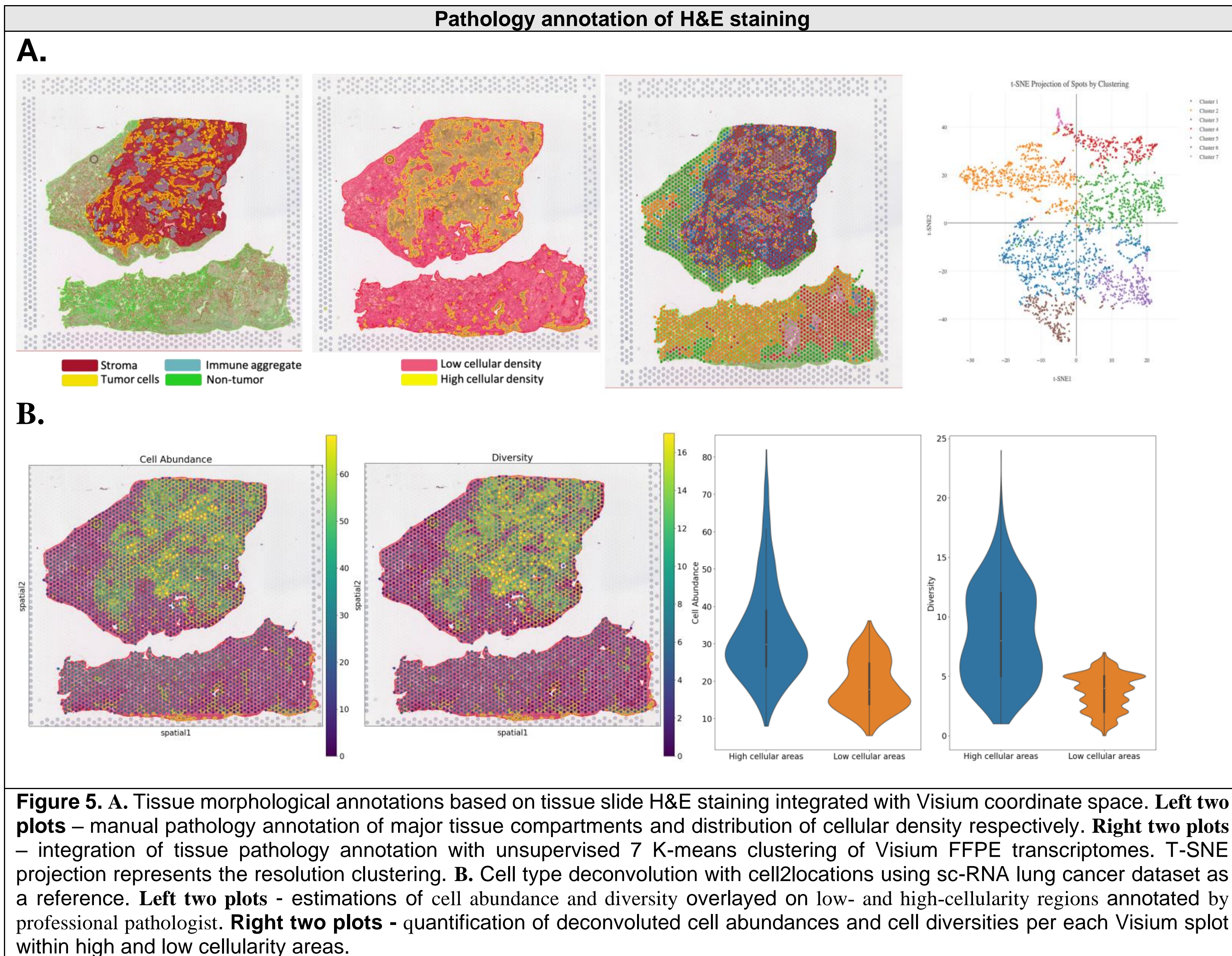


Figure 5. A. Tissue morphological annotations based on tissue slide H&E staining integrated with Visium coordinate space. **Left two plots** – manual pathology annotation of major tissue compartments and distribution of cellular density respectively. **Right two plots** – integration of tissue pathology annotation with unsupervised 7 K-means clustering of Visium FFPE transcriptomes. T-SNE projection represents the resolution clustering. **B.** Cell type deconvolution with cell2locations using sc-RNA lung cancer dataset as a reference. **Left two plots** - estimations of cell abundance and diversity overlayed on low- and high-cellularity regions annotated by professional pathologist. **Right two plots -** quantification of deconvoluted cell abundances and cell diversities per each Visium splot within high and low cellularity areas.

To validate that obtained transcriptome profiles are biologically correlate with tissue areas of origin, we performed differential gene expression analysis of spot clusters matched per each tissue compartment. Unsupervised K-means clustering of Visium spots is able to capture seven clusters which perfectly aligned with pathology supervised annotation (**Figure 5A**). K-means cluster 6 is assigned to immune aggregate area which indeed is enriched in genes signatures of B-cell and other lymphoid cell activity as determined by DE gene analysis of cluster 6 versus all others (**Figure 6A**). Indeed, gene-signature enrichment analysis (GSEA) using obtained ranked gene list against Gene Ontology pathways confirmed the presence of B-cell activity and BCR activation in lymphoid aggregate area (**Figure 6B**). This observation suggests that H&E guided, pathology-driven Visium spot subsetting captures transcriptome variability which is biologically valid and matches with transcriptional programs expected to be present from annotated tissue areas. Similarly, to lymphoid aggregate areas, tumor-specific regions are enriched in gene signatures of tumorigenesis and particularly associated with lung cancer progression (**Figure 7**). Other clusters showed enrichment of their own signatures not shown in this report due to space limitations.

Differential expression and cell type correlations

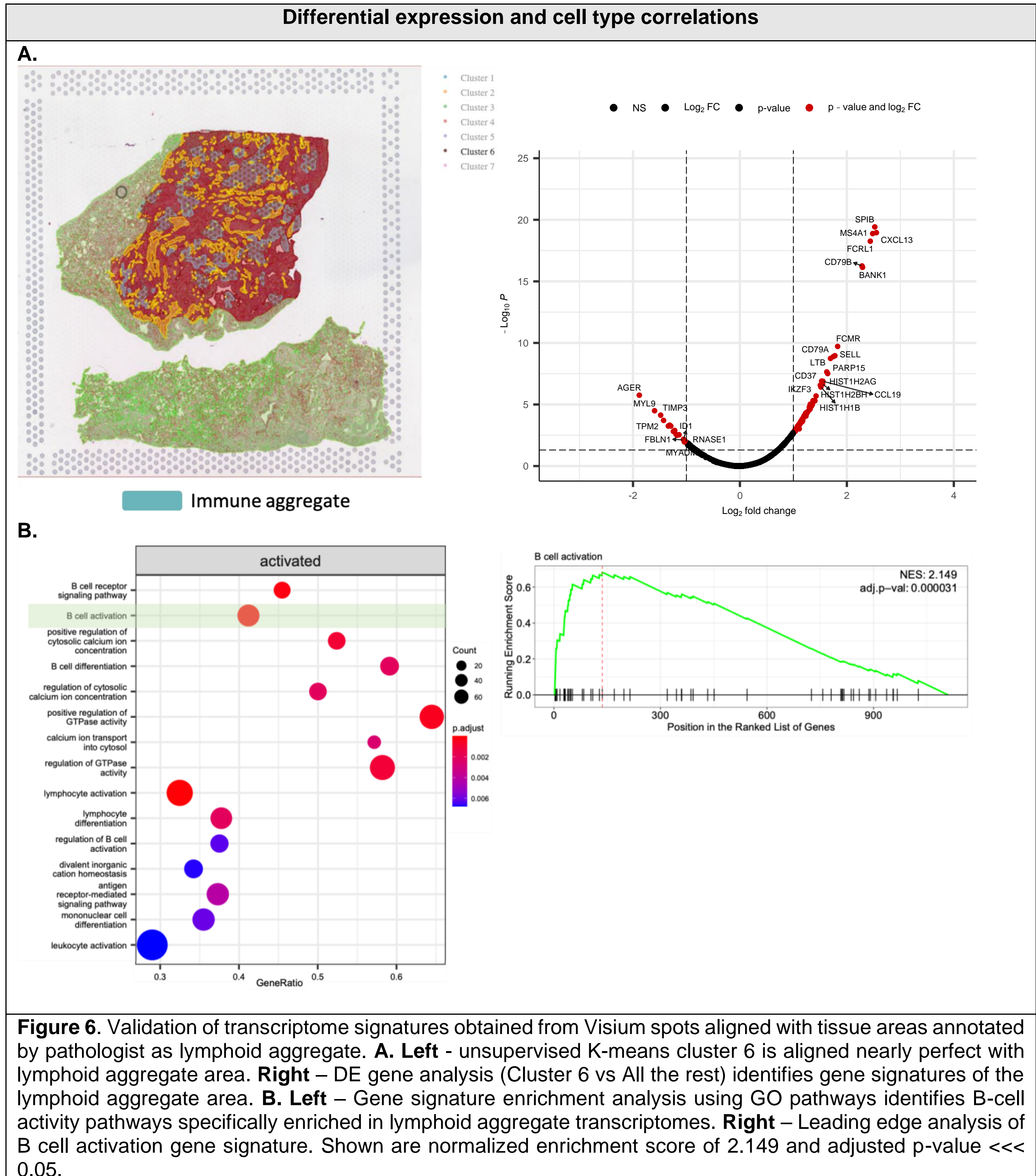


Figure 6. Validation of transcriptome signatures obtained from Visium spots aligned with tissue areas annotated by pathologist as lymphoid aggregate. **A. Left** - unsupervised K-means cluster 6 is aligned nearly perfect with lymphoid aggregate area. **Right** – DE gene analysis (Cluster 6 vs All the rest) identifies gene signatures of the lymphoid aggregate area. **B. Left** – Gene signature enrichment analysis using GO pathways identifies B-cell activity pathways specifically enriched in lymphoid aggregate transcriptomes. **Right** – Leading edge analysis of B cell activation gene signature. Shown are normalized enrichment score of 2.149 and adjusted p-value <<< 0.05.

Major cell types per site

A.

Tumor cells

B.

Figure 7. Validation of transcriptome signatures obtained from Visium spots aligned with tissue areas annotated by pathologist as tumor area. **A. Left** - unsupervised K-means cluster 5 is aligned nearly perfect with tumor area. **Right** – DE gene analysis (Cluster 5 vs All the rest) identifies gene signatures of tumor area. **B. Left** – Gene signature enrichment analysis using C2 oncological pathways from MSigDB identifies tumor-associated pathways specifically enriched in tumor area transcriptomes with apparent lung-cancer tissue source (gene sets of upregulated/amplified genes). **Right** – Leading edge analysis of Shedden Lung Cancer Good Survival A4 gene signature. Shown are normalized enrichment score of 2.0138 and adjusted p-value <<< 0.05.

3.5 Comparison of FFPE vs. OCT performance

The Visium 10X platform offers flexibility in terms of tissue embedding method and source: FFPE or OCT. Although we anticipate using the method primarily for FFPE tissues, due to simpler implementation as part of multi-site clinical trials with a central biorepository and larger availability of such material, we still asked whether assay performance substantially differed between the two embedding sources. For this purpose, we compared normalized UMI expression in a variety of tumor types between samples collected as FFPE vs. OCT. It should be noted that the FFPE method uses targeted probes (18k) vs. greater coverage in OCT (36k), but circumvents reverse transcription. As a result, FFPE was able to typically recover higher UMI/spot even in samples with poorly preserved RNA/poor tissue quality. When comparing spatially resolved information between FFPE and OCT, using a subset of highly spatially variable genes (selected by var/mean plots), we found that the results were concordant across most tissues (**Figure 8**). Consistent with a greater sensitivity to detect rare UMIs, FFPE could resolve reads that were almost undetectable in OCT, which could be useful for identifying small subsets with greater granularity. It cannot be excluded that some of these FFPE-specific genes are due to unresolved cross-reactivity in probes, and investigations are ongoing to address this. Still, overall, FFPE was recommended over OCT, given its advantages for the CIMAC workflow.

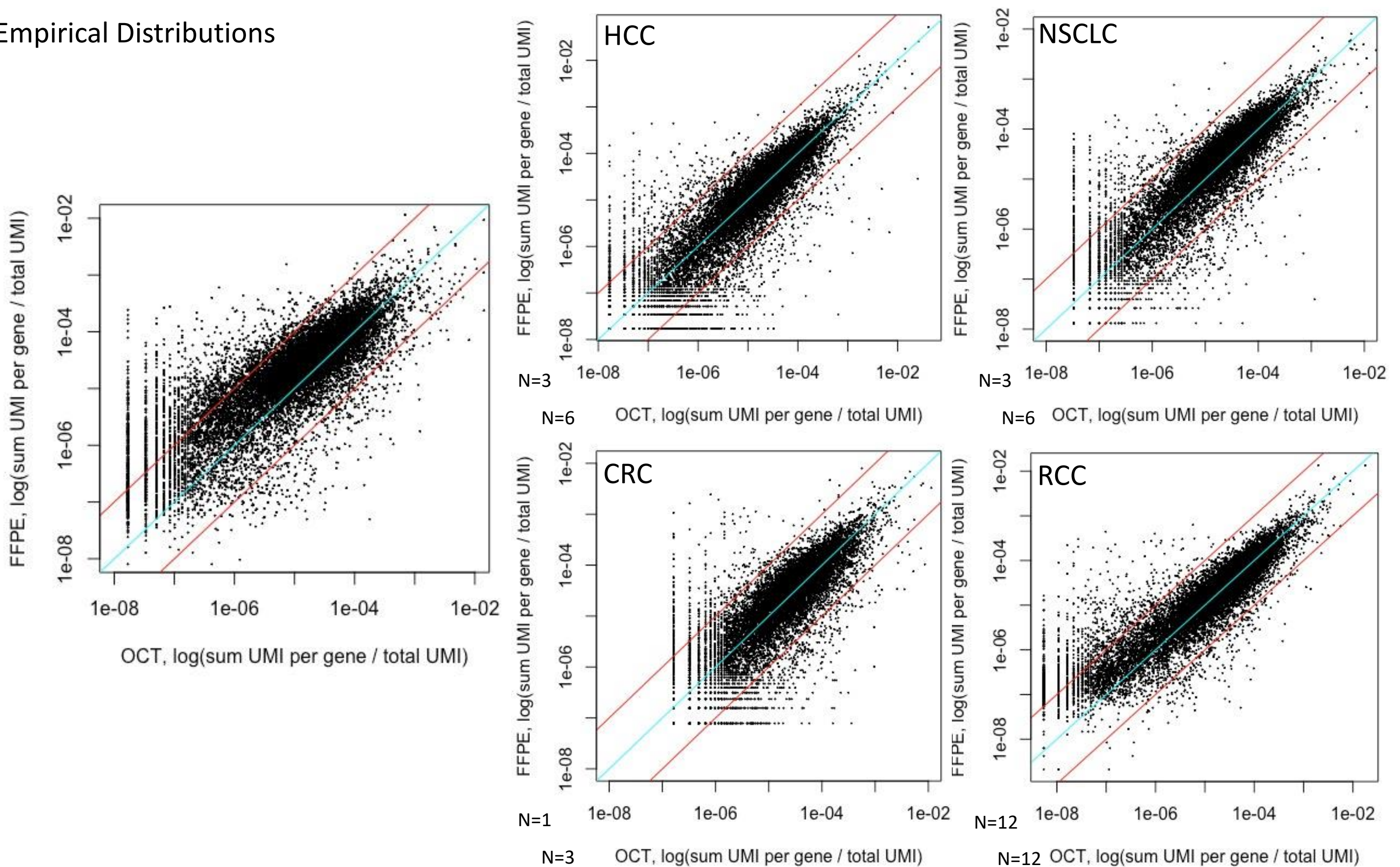


Figure 8. Comparison of FFPE vs. OCT in tissues from hepatocellular carcinoma (HCC), non-small cell lung carcinoma (NSCLC), colorectal carcinoma (CRC), and renal cell carcinoma (RCC). The left panel is an aggregate of all samples (with their n indicated). Plotted are the most spatially variable genes expressed as sum UMI per gene / total UMI, with most genes falling within a comparable range across both methods. Exceptions are noted at the very low range of OCT, with more detection seen in FFPE.

DATA ANALYSIS PIPELINE

Obtained sequencing reads are QC'ed, filtered and aligned to HG38 reference genome according to Spaceranger 10x Genomics pipeline (see references). H&E image is gathered at 20x magnification (Leica X) and manually annotated by professional pathologist using QuPath open source software to characterize major biological structures (stroma, tumor, normal, immune aggregate) and general physical parameters (low and high cellular densities). To map spatial barcodes, annotated images are combined together with sequencing datasets into Spaceranger pipeline. Spatially resolved gene expression is further analysed using Scanpy toolkit developed in python. PCA-UMAP transformation and unsupervised clustering are done with the minimum K-mean clusters to match pathologist review visually. Differential gene expression analysis and gene signature analysis are done in R by ClusterProfiler package using MSigDB, GO-term and KEGG databases as a reference list of biological pathways. To identify cell types and states within each Visium spot, we setup Cell2location deconvolution pipeline. Cell2location (3) takes reference cell type signatures derived from scRNA-seq and spatial transcriptomics data as input. The model then decomposes spatially resolved multi-cell RNA count matrices into the reference cell type signatures, thereby estimating cell abundance of individual cell types across locations. For this dataset, we constructed a reference of approximately 500,000 cells from tumor samples in Non-small cell lung cancer and lung adenocarcinoma. Approximated cell abundance and number of cell types per spot show concordant enrichment in pathology defined high cellular areas.

CONCLUSION STATEMENT

We at HIMC have developed experimental and bioinformatics pipelines for efficient profiling of spatially resolved gene expression using Visium FFPE 10x Genomics methodology. Combined together with available biological databases of annotated biological pathways and gene signatures, it allows for discovery of novel spatially resolved transcriptional biomarkers with the potential predictive (if applied at pre-treatment setting) or mechanistic value (if applied at on-treatment setting).

3. References

1. https://pages.10xgenomics.com/rs/446-PBO-704/images/10x_LIT000128_PS_Spatial_biology_without_limits_Spatial_gene_expression_in_FFPE.pdf

2. https://www.10xgenomics.com/support/spatial-gene-expression-ffpe/documentation/workflows/ffpe-v-1/steps/experimental-design-and-planning/visium-spatial-gene-expression-for-ffpe-tested-tissues

3. Kleshchevnikov, V., Shmatko, A., Dann, E. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. Nat Biotechnol 40, 661–671 (2022).