Visium CytAssist Spatial Gene Expression for FFPE: Robust Data Analysis with Minimal Impact of Genomic DNA

Introduction

The 10x Genomics Spatial Gene Expression technology for FFPE enables spatial transcriptomic insights by analyzing mRNA in tissue sections derived from formalin-fixed, paraffin-embedded (FFPE) tissue samples. While the Visium Spatial Gene Expression assay analyzes FFPE sections placed directly on a Visium slide (Direct placement), the Visium CytAssist Gene Expression for FFPE workflow (CytAssist-enabled) automates the transfer of transcriptomic probes from glass slides to Visium slides. This provides more flexibility in sample preparation and enables spatial profiling of a broader spectrum of samples. Tissue permeabilization in the CytAssist-enabled assay was optimized by increasing the decrosslinking temperature from 70°C in the Direct placement to 95°C (Figure 1). A consequence of this optimization is that probe access to genomic DNA (gDNA) sequences is increased, leading to Unique Molecular Identifier (UMI) counts from probe ligation events on gDNA in addition to the desired probe ligation events on mRNA (Figure 2).

This Technical Note characterizes the presence of gDNA in data derived from the CytAssist-enabled and the Direct placement assays. The results highlight that despite higher gDNA accessibility in the CytAssist-enabled data, the robustness of the biological interpretation is comparable between both assays.

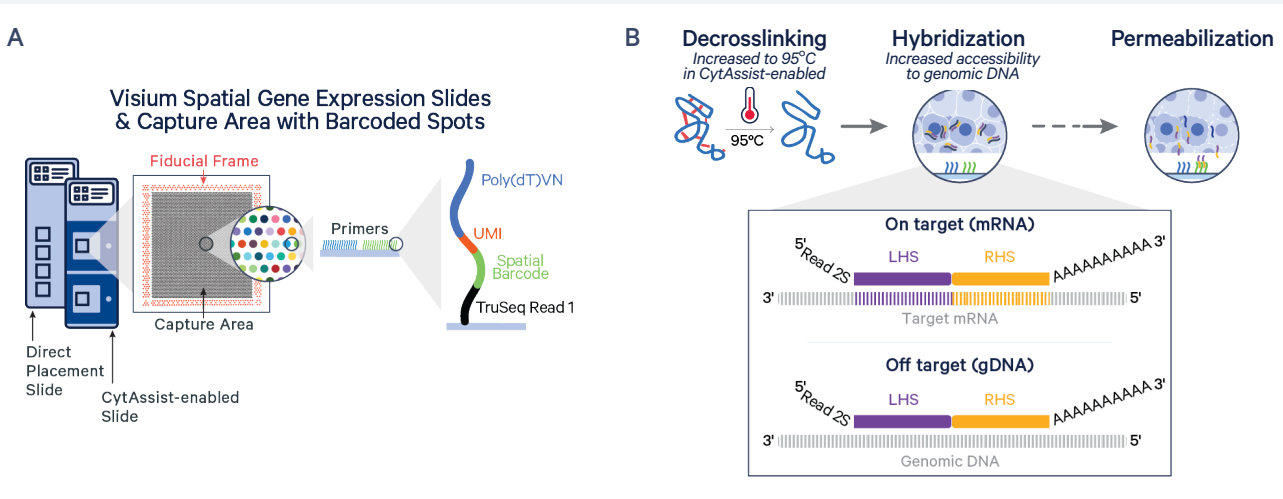


Figure 1. Comparing Direct placement and CytAssist-enabled assays. **A.** The spot size and resolution on the slides are identical between the two assays. **B.** The decrosslinking temperature is increased to 95°C in the CytAssist-enabled assay. Partially single stranded and fragmented gDNA onto which RTL probes can be ligated are released during the hybridization step, followed by probe capture on the Visium slide, leading to a low to moderate gDNA signal.

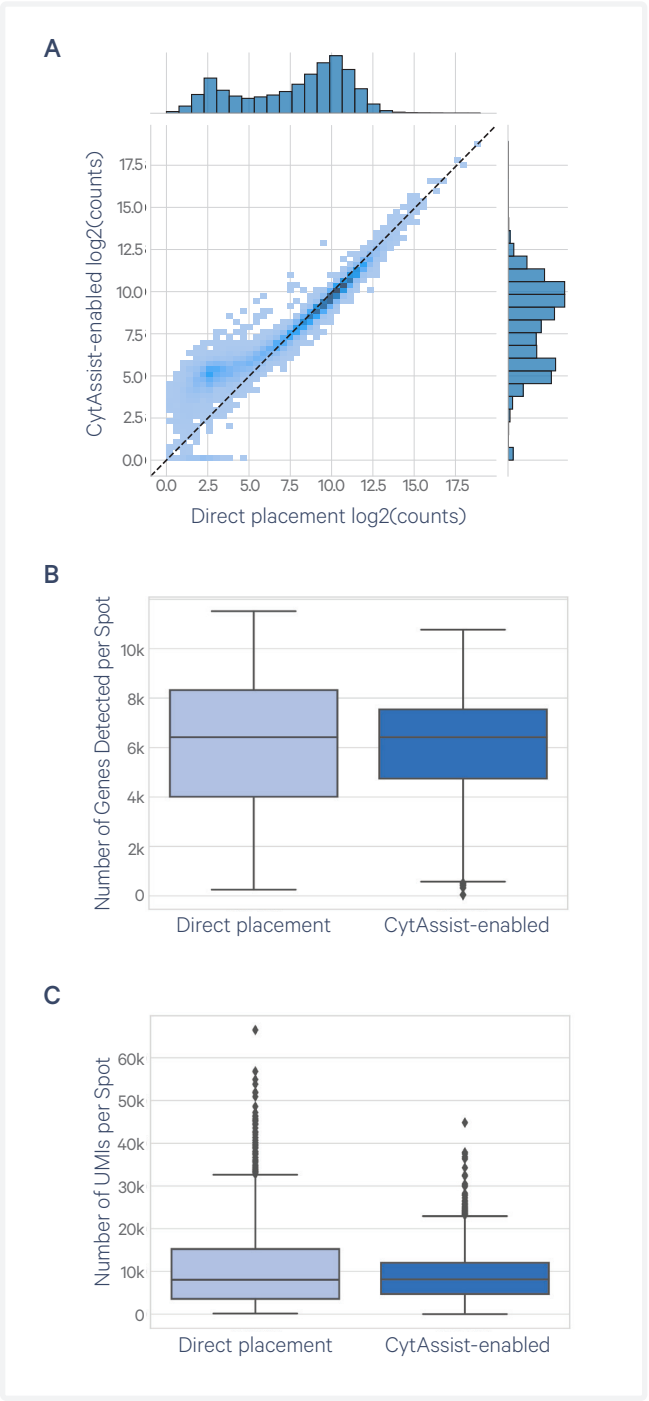


Figure 2. Higher gDNA in CytAssist-enabled assay does not impact key metrics. **A.** Comparison of the Direct placement and CytAssist-enabled human tonsil data (n=2) shows genes with lower counts deviating from the diagonal, indicating an inflated background signal due to gDNA in CytAssist-enabled assay. Each point represents a gene, the x axis is log2(counts) for Direct placement, and the y axis is log2(counts) for CytAssist-enabled assay. If there were no gDNA background, a fit along the diagonal would be expected. Presence of gDNA has little effect on key metrics; **B.** Genes per spot **C.** UMIs per spot, which are comparable between the two assays.

Methods

Experimental Design

An overview of the experimental design is provided in Figure 3. Consecutive sections from human ovarian cancer, lung cancer, tonsil, glioblastoma, and prostate cancer were processed according to the respective assay protocols of both assays (Table 1). The same Visium Human Transcriptome Probes v2 designed for the CytAssist assay were used for both Direct placement and CytAssist-enabled assays for estimating gDNA contribution to gene counts (see Estimating gDNA fraction below). This experimental design enabled quantifying the contribution to UMI counts from ligation events templated by gDNA in both the CytAssist-enabled and the Direct placement assay. Whether the interpretation of the underlying biology was impacted by the presence of increased UMI counts stemming from gDNA in CytAssist-enabled assay, was subsequently examined. Note that the use of Visium Human Transcriptome Probes v2 with Direct placement assay is not a 10x officially supported protocol.

Tissue Processing & Library Preparation

10x Genomics obtained tissues with corresponding disease states from the vendors listed in Table 1.

Tissue Type	Vendor	Disease State
Human Prostate Cancer	Avaden Biosciences	Adenocarcinoma
Human Glioblastoma	BioIVT	Glioblastoma Multiforme
Human Tonsil	Avaden Biosciences	Hyperplasia
Human Ovarian Cancer	Avaden Biosciences	Serous Carcinoma
Human Lung Cancer	Avaden Biosciences	Invasive Mucinous Adenocarcinoma
Human Spleen	Avaden Biosciences	Non-Diseased

Table 1. List of tissues used along with vendor and disease state information.

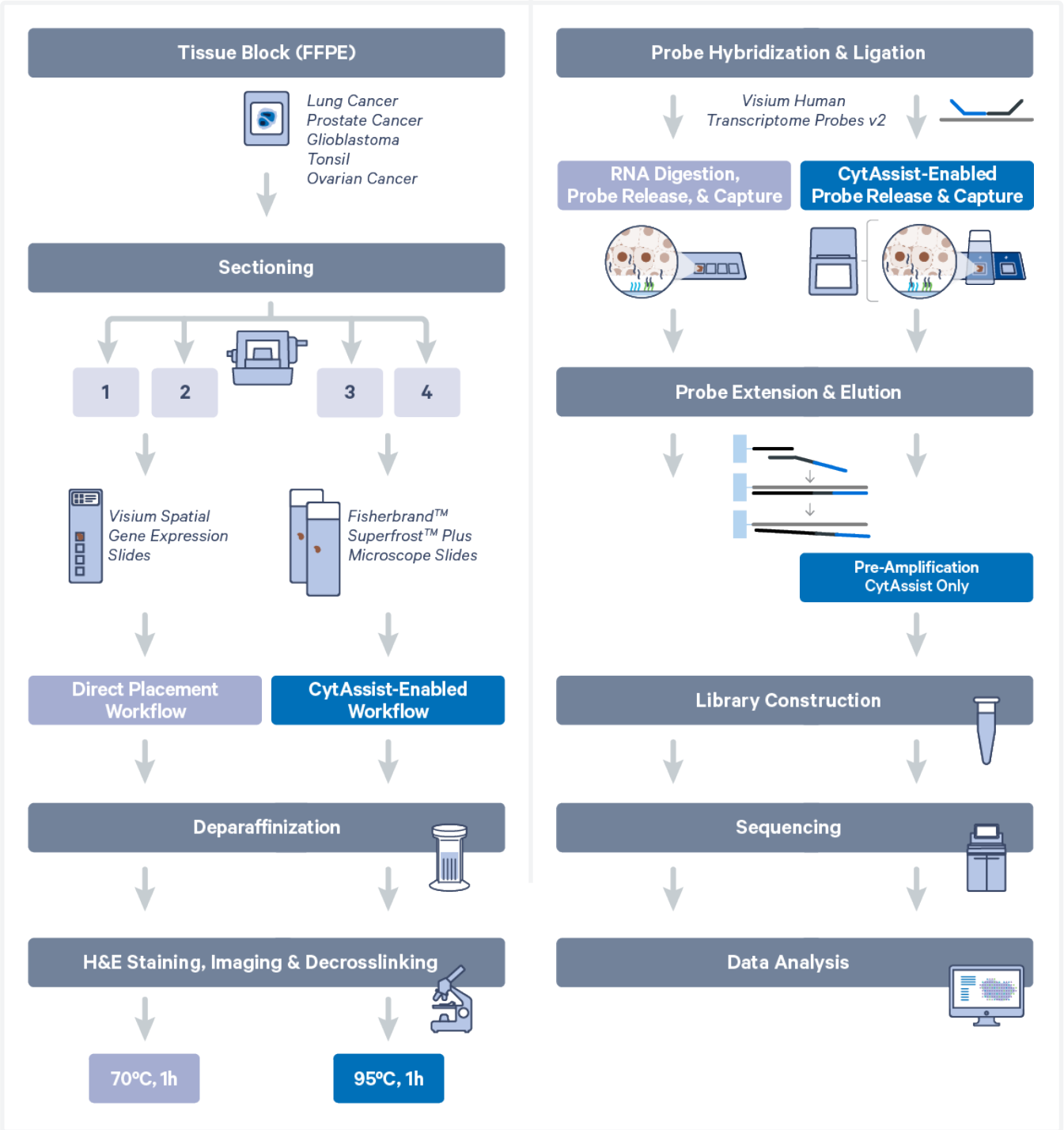


Figure 3. Experimental design overview. Four consecutive sections from each one of the five human tissue FFPE blocks were sectioned and the same number of sections from each block were assessed using the Direct placement and CytAssist-enabled assays. The workflow was executed as described in the respective protocols for each assay including the decrosslinking temperature which was 70°C for the Direct placement assay and 95°C for the CytAssist-enabled assay. Visium Human Transcriptome Probes v2 were used for both the workflows. For the Direct placement assay, the RNA digestion, probe release, and capture occurred directly on the Visium slide. For the CytAssist-enabled assay, a CytAssist instrument facilitated the transfer of probes from two glass slides to a single Visium slide. Note that the spleen tissue sections were processed following the same CytAssist-enabled workflow as shown here.

For the Direct placement assay, 5 µm thick tissue sections were placed on Visium Spatial Gene Expression slides. For the CytAssist-enabled assay, 5 µm tissue sections were placed on Fisherbrand™ Superfrost™ Plus Microscope Slides (Fisher Scientific), and later in the workflow, ligation products were transferred to Visium CytAssist Spatial Gene Expression Slides via the Visium CytAssist instrument. H&E images for both assays were acquired using a V200 Slide Scanner Microscope from Olympus.

Tissue processing, library preparation, and sequencing were performed as described in the respective Demonstrated Protocols and User Guides. For the CytAssist-enabled assay, tissue sections were prepared following Demonstrated Protocol (CG000518), followed by H&E staining (CG000520). Libraries were prepared following the Visium CytAssist Spatial Gene Expression Reagents Kits for FFPE User Guide (CG000495). For the Direct placement assay, tissue sections were prepared as per the Demonstrated Protocol (CG000408), followed by H&E staining (CG000409). Libraries were prepared as per the User Guide (CG000407), with the exception that the Visium Human Transcriptome Probes v2 (PN-1000466) were used for both assays as described above. All libraries were sequenced on an Illumina NovaSeq at a minimum sequencing depth of 45k mean reads per tissue covered spot. A dual indexing sequencing configuration was used: 28 bp read 1 (16 bp Visium spatial barcode, 12 bp UMI), 50 bp read2 (probe insert), 10 bp i7 sample index, and 10 bp i5 sample index.

Data Processing, Clustering, and Differential Gene Expression

Raw FASTQ and imaging data were processed with the Space Ranger 2.0 pipeline. For each tissue type, the replicates from a given assay were aggregated to find common clusters and annotations that help define biological differences. Harmony [1] was used to find a shared embedding, in which spots are grouped by tissue regions rather than dataset-specific conditions. To accommodate analysis across different tissues, to minimize noise, and to avoid selection bias, hyperparameters were kept consistent. Specifically, when using a graph clustered approach, the resolution was optimized to identify six to eight clusters, which accommodates broad biological clusters across all tissues. To assess the biological signal between spots on the Visium slides, differential gene expression (DGE) analysis was conducted between the Harmony clusters followed by comparison of the results from both assays.

Detailed Analysis Workflow

Space Ranger 2.0

1. Preprocess all raw FASTQ from consecutive sections
   - Reduces biological differences in the comparison
2. Downsample both samples to 30k reads per spot
   - Reduces batch effects linked to sequencing depth

Seurat v3.9.9 [2]

3. Aggregate both samples
4. Select top variable genes (n_genes=2000, default value)
5. Run PCA on first 10 (npcs=10, default value) components

Harmony

6. Find common cluster embeddings
7. Find clusters (resolution 0.2)
   - The hyper-parameter value reflects the aim to generate 6-8 clusters per sample

Seurat v3.9.9

8. DGE analysis across all detected genes between clusters
   - Minimum of 30% of spots must express the gene

Estimating gDNA Fraction

The Visium Human Transcriptome Probes v2 contain on average three probe-pairs per transcript (~54k probe-pairs targeting ~18k transcripts). While the majority (~47k probe-pairs) of the probes target non-exon-junction-spanning ("unspliced") regions in the transcriptome, some (~7.5k probe-pairs) have one exon-junction-spanning ("spliced") probe (Figure 4). Spliced probes, defined as having a splice junction within 10 nucleotides of the ligation junction, are unlikely to hybridize and ligate onto gDNA, while unspliced probes can use either gDNA or mRNA as templates. Comparing these two types of probes allows for the estimation of reads originating from gDNA. For genes that have both spliced and unspliced probes, the difference in counts for unspliced probes vs spliced probes is used to estimate the UMIs from gDNA (~5.2k genes). The human and mouse v1 probe sets were designed with one probe-pair per gene, making an accurate estimate of gDNA on a per-gene basis impossible. This necessitated an experimental design using Visium Human Transcriptome Probes v2 with the Direct placement assay.

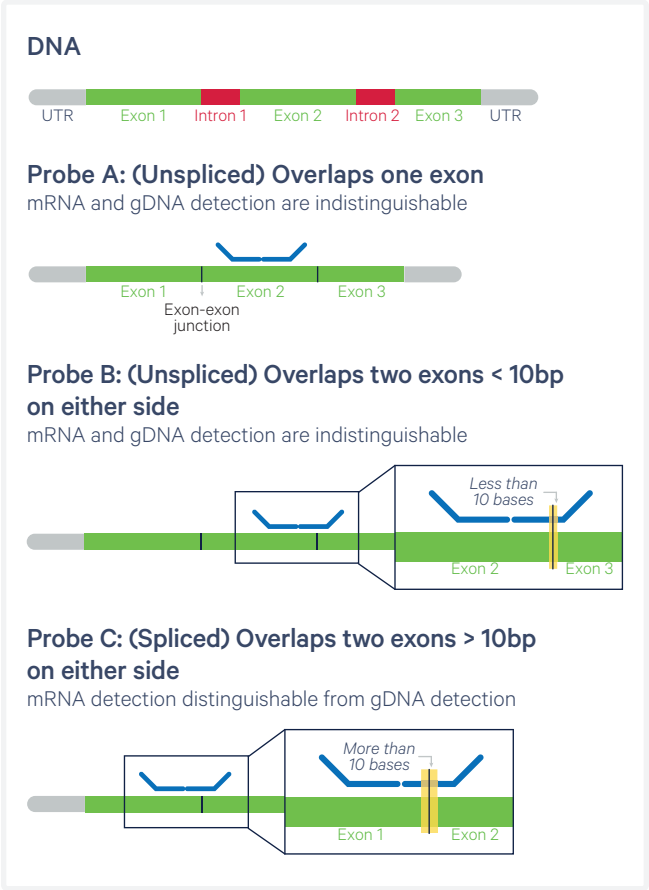


Figure 4. Visium for FFPE human v2 probe design and probe type labeling. To distinguish reads originating from gDNA from those originating from mRNA, measurements from exon-junction-spanning ("spliced") probes, which are unlikely to template onto gDNA, were compared with probes targeting non-exon-junction-spanning ("unspliced") regions that can use either gDNA or mRNA as a template. Probe A is fully overlapping an exon, and mRNA detection cannot be distinguished from gDNA detection. Probe B is partially overlapping two exons, but by fewer than 10 bp on one side; this probe also counts as unspliced. Probe C overlaps two exons by more than 10 bp on either side and is labeled as exon-junction-spanning. Such a ligation event is assumed to come from mRNA only.

The gDNA estimate is obtained by fitting a piecewise linear model on genes with both spliced and unspliced probes. The model predicts the log-transformed gene counts for the unspliced probes as a function of log-transformed gene counts for the spliced probes and log-transformed estimated gDNA counts per gene per 1,000 spots under tissue:

$$\mathbf{x_{unspliced}} = \mathbf{gDNA} + \boldsymbol{B}(\mathbf{x_{spliced}} - \hat{\mathbf{x}}_{\mathbf{gDNA}})\ \mathbf{II}\ (\mathbf{x_{spliced}} > \hat{\mathbf{x}}_{\mathbf{gDNA}})$$

where $\hat{\mathbf{x}}_{\mathbf{gDNA}}$ is the estimated gDNA UMIs per gene, $\mathbf{x_{unspliced}}$ and $\mathbf{x_{spliced}}$ are the average number of UMIs corresponding to unspliced and spliced probes from a single gene, respectively, and $\mathbf{II}\ (\mathbf{x_{spliced}} > \hat{\mathbf{x}}_{\mathbf{gDNA}})$ is an indicator function whose value is one if $\mathbf{x_{spliced}}$ is greater than $\hat{\mathbf{x}}_{\mathbf{gDNA}}$, and zero in all other instances.

For any fixed estimate of $\hat{\mathbf{x}}_{\mathbf{gDNA}}$, the model parameters $\mathbf{gDNA}$ and $\boldsymbol{B}$ are estimated by fitting a linear regression with a two dimensional parameter vector. Only the values of $\mathbf{x_{spliced}}$ seen in the genes under consideration are the relevant $\hat{\mathbf{x}}_{\mathbf{gDNA}}$ values to consider while fitting the model. Therefore, a linear model with every $\mathbf{x_{spliced}}$ seen in the dataset as $\hat{\mathbf{x}}_{\mathbf{gDNA}}$ is fit and the model with the smallest residual sum of squares is used for the estimates.

Results and Discussion

Assessing gDNA Signal

The comparison of the Direct placement and the CytAssist-enabled assays for human tonsil tissue demonstrates the presence of gDNA in CytAssist-enabled data, visualized as the deviation from the diagonal towards the Y axis (Figure 2A). The deviation implies that there are more UMI counts than expected in the CytAssist-enabled data, which particularly impacts the data at the low end of the range of UMI counts. The genes per spot and UMIs per spot from tissue-associated spots are not significantly different between Direct placement and CytAssist-enabled assays (Figure 2B & 2C).

Researchers performing spatial analyses on human tissues can assess the level of gDNA in their Visium CytAssist Spatial Gene Expression for FFPE datasets using a new metric, *Estimated UMIs from Genomic DNA*, in the web_summary.html file output from Space Ranger 2.0 (Figure 5). The web summary also shows a scatterplot comparing spliced to unspliced probe counts, from which the *Estimated per Probe Background UMI Count* metric is derived. The difference in the counts due to background becomes apparent at low probe counts, where the relationship between the sums switches from being linearly correlated (turquoise) to non-correlated (purple). The mean of the discordant, non-linear counts (purple line) gives the estimated UMI background level cutoff (reported as *Estimated per Probe Background UMI Count* in web summary) below which unspliced probe counts have a high probability of stemming from a background probe ligation event.

Summary statistics for the six tissue types are tabulated in Table 2. Estimated UMIs from gDNA as reported in the web_summary.html file output from Space Ranger was 0.9-4.8% for CytAssist-enabled assay*. For the Direct placement assay, estimated gDNA was 0-0.6% (Figure 6A). As a consequence, in the tissues run with CytAssist-enabled assay, there is a slight increase in the detection of lowly expressed genes. This increase in detection is compounded with gDNA signal and cannot be attributed to transcriptomic signal alone (Figure 6C).

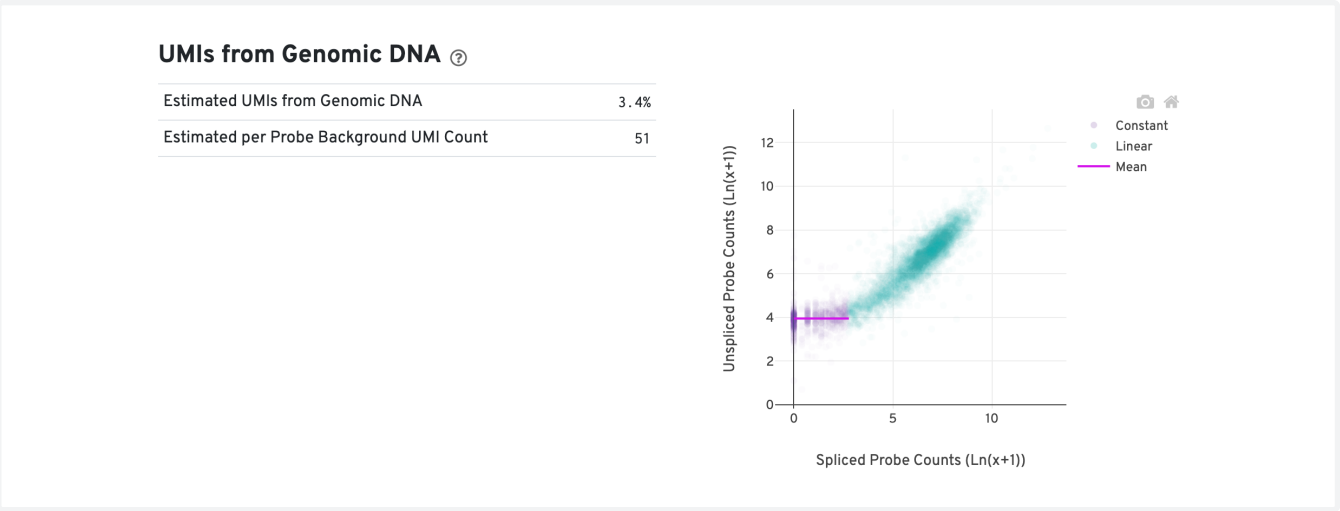


Figure 5. Example web_summary.html metric† and plot showing gDNA signal assessment in the CytAssist-enabled human data. Each point in the plot reflects the relationship between two pseudo-bulk counts for a transcript. The x-axis represents the sum of the spliced probe counts across barcodes per transcript, while the y-axis represents the sum of the unspliced probe counts across barcodes for the same transcript.

*Spleen samples run with CytAssist-enabled assay resulted in ~20% Estimated UMIs from gDNA.

†This metric will not be available with mouse data which uses v1 probes generated using either CytAssist-enabled and Direct placement assay as well as human data that uses v1 probes (Direct placement assay) due to the probeset design (see Estimating gDNA Fraction section).

Tissue Section ID	Assay	Genes Detected	Mean Panel Genes	Median UMI Counts per Spot	Estimated UMIs from gDNA
Ovarian_Cancer_DP_rep1	Direct placement	17999	3020	5691	0.40%
Ovarian_Cancer_DP_rep2	Direct placement	18016	3108	5929	0.40%
Ovarian_Cancer_CE_rep1	CytAssist-enabled	18063	4519	9838	3.20%
Ovarian_Cancer_CE_rep2	CytAssist-enabled	18063	4012	8326	2.90%
Lung_Cancer_DP_rep1	Direct placement	17471	6939	23699	0.00%
Lung_Cancer_DP_rep2	Direct placement	17316	7123	25057	0.00%
Lung_Cancer_CE_rep1	CytAssist-enabled	18068	6450	20332	0.90%
Lung_Cancer_CE_rep2	CytAssist-enabled	18067	6647	21850	0.90%
Tonsil_DP_rep1	Direct placement	18062	6325	16287	0.50%
Tonsil_DP_rep2	Direct placement	18061	6511	17209	0.60%
Tonsil_CE_rep1	CytAssist-enabled	18068	6327	16463	3.40%
Tonsil_CE_rep2	CytAssist-enabled	18067	5640	13076	3.50%
Glioblastoma_DP_rep1	Direct placement	17088	5603	12480	0.10%
Glioblastoma_DP_rep2	Direct placement	17024	5314	11295	0.10%
Glioblastoma_CE_rep1	CytAssist-enabled	18050	4485	8420	4.80%
Glioblastoma_CE_rep2	CytAssist-enabled	18049	5050	10415	4.60%
Prostate_Cancer_DP_rep1	Direct placement	17131	6506	18906	0.00%
Prostate_Cancer_DP_rep2	Direct placement	17476	6358	17869	0.10%
Prostate_Cancer_CE_rep1	CytAssist-enabled	18063	5676	14510	1.40%
Prostate_Cancer_CE_rep2	CytAssist-enabled	18063	6106	16548	1.40%
Spleen_CE_rep1	CytAssist-enabled	18057	2734	3866	23.30%
Spleen_CE_rep2	CytAssist-enabled	18057	2607	3702	20.60%

Table 2. Key metrics to assess the effect of gDNA on biological interpretations of human tissues. Visium Human Transcriptome Probes v2 were used for both the Direct placement and the CytAssist-enabled assays to estimate and directly compare the gDNA fraction in the six tissue types. The data from spleen was not used for any of the plots or analysis shown in this Technical Note and are included to showcase gDNA range in challenging tissue types.

Tissue Section ID: Unique identifier for the tissue section; prefix denotes tissue type, followed by the assay abbreviation; suffix specifies the technical replicate.

Assay: Direct placement = Visium Spatial Gene Expression for FFPE; CytAssist-enabled = Visium CytAssist Spatial Gene Expression for FFPE.

Genes Detected: The number of unique genes from the filtered probe set with at least one UMI count in any tissue covered spot downsampled to 30k raw reads per spot.

Mean Panel Genes: The mean number of genes detected per spot under tissue-associated barcode downsampled to 30k raw reads per spot. Detection is defined as the presence of at least one UMI count.

Median UMI Counts per Spot: The median number of UMI counts per tissue covered spot downsampled to 30k raw reads per spot.

Estimated UMIs from gDNA: The percentage of UMIs estimated to originate from off-target probe binding to gDNA downsampled to 30k raw reads per spot.

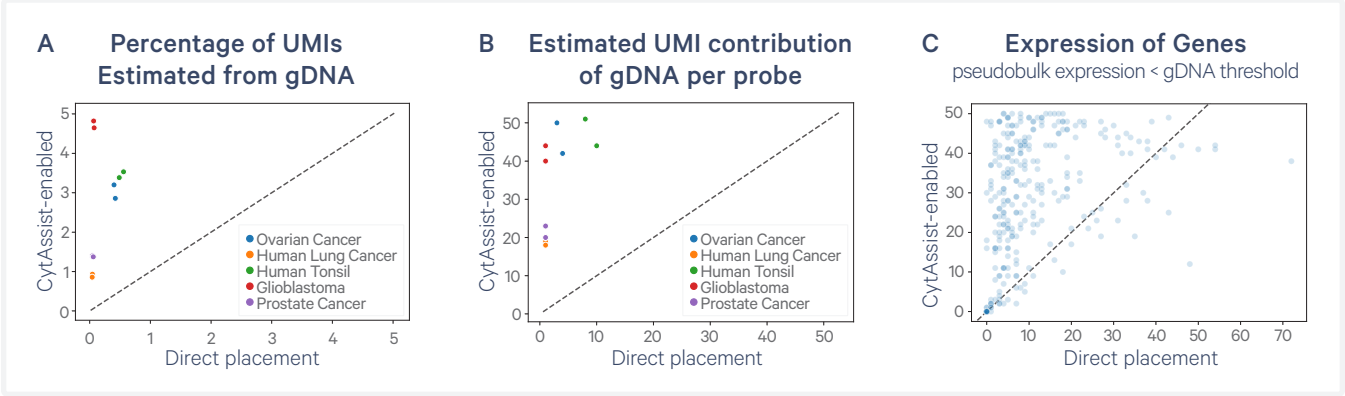


Figure 6. Contribution of estimated UMIs from gDNA are variable across different human tissue types. For each tissue type, paired data from tissue sections run on the Direct placement and CytAssist-enabled assays are compared and corresponding gDNA values are plotted. **A.** Estimated UMIs from Genomic DNA. **B**. Estimated per Probe Background UMI Count. **C**. Genes with low expression with gene counts less than the gDNA threshold in CytAssist-enabled assay in the tonsil paired dataset. These plots show a small increase in background gDNA contribution in the CytAssist-enabled assay relative to Direct placement assay.

Biological Interpretations are Robust to Presence of gDNA in CytAssist-Enabled Assay

The qualitative assessment of the clusters inferred by Harmony showed consistency between the aggregated tissue sections run on Direct placement and CytAssist-enabled assays (Figure 7). The majority of the graph-based clustering generated by Space Ranger 2.0 follow similar anatomical regions between the individual Direct placement and CytAssist-enabled tissue sections and the cluster colocalizations are consistent between the Space Ranger and Harmony generated clusters.

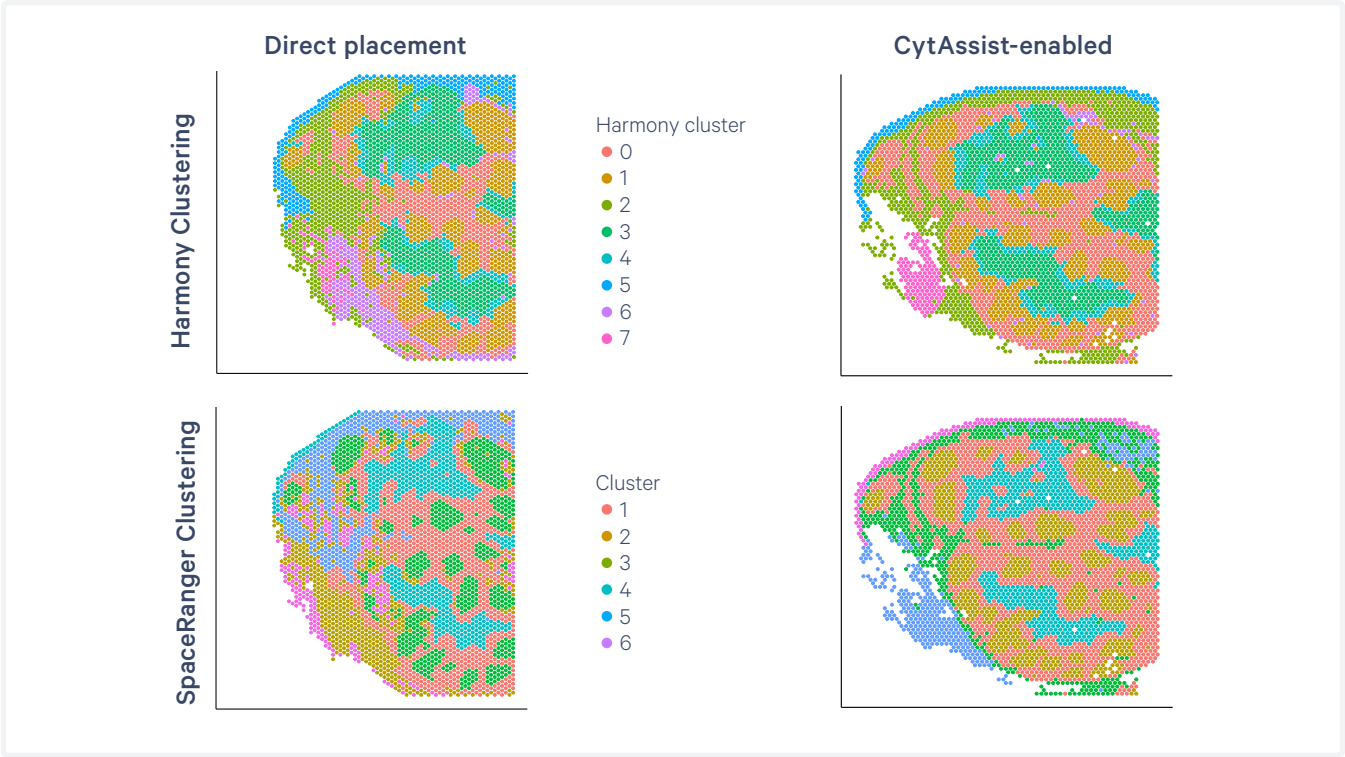


Figure 7. Cluster assignment of tonsil sections generated from Direct placement and CytAssist-enabled assays show similar clustering patterns. Top row is a comparison of Harmony clustering in aggregated tissue sections between Direct placement (top left) and CytAssist-enabled (top right) for the tonsil tissue. Bottom row shows graph-based clustering produced for each individual tissue section separately; Direct placement (bottom left) and CytAssist-enabled (bottom right).

Next, the genes that are differentially expressed in the same clusters on both assays were examined. The DGE analysis is comparable with respect to the genes that are present in the same clusters between the Direct placement and CytAssist-enabled assays. Most differentially expressed upregulated genes detected in the Direct placement are also detected in CytAssist-enabled assay (Figure 8A). Furthermore, the number of uniquely differentially expressed (DE) genes is higher in the CytAssist-enabled assay. Comparisons of log2FC and adjusted p-values for the DGE analyses are shown in Figure 8B. The data indicates that the cluster assignment and the direction of DE gene regulation are consistent between the two assays. If there was no batch difference, a fit along the diagonal would be expected. These results indicate that the two assays are largely producing congruent results despite the presence of UMIs attributed to gDNA counts.

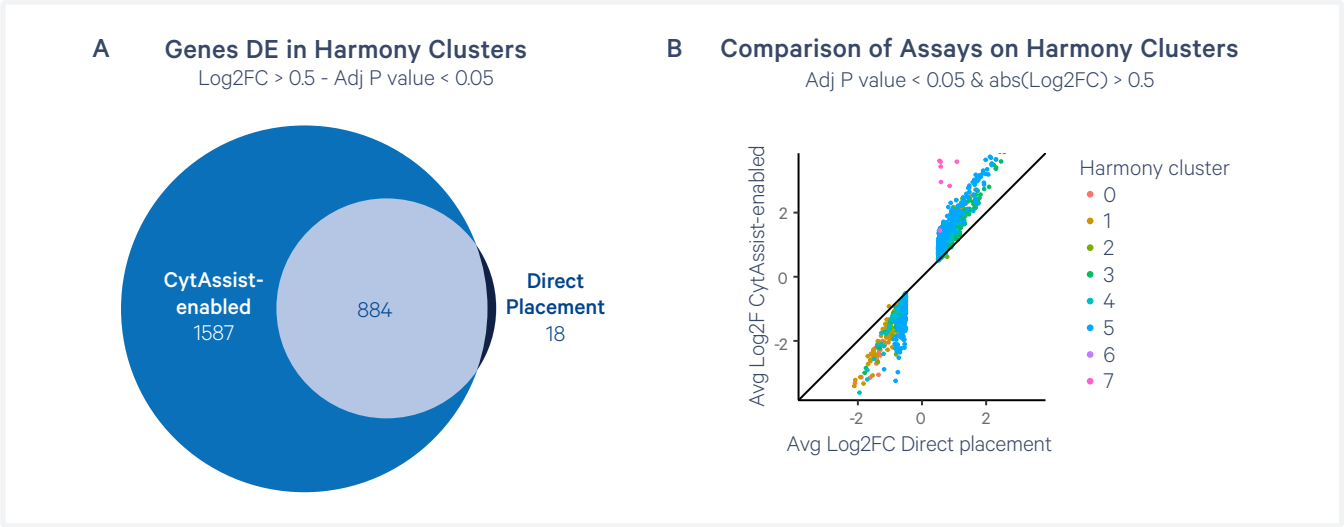
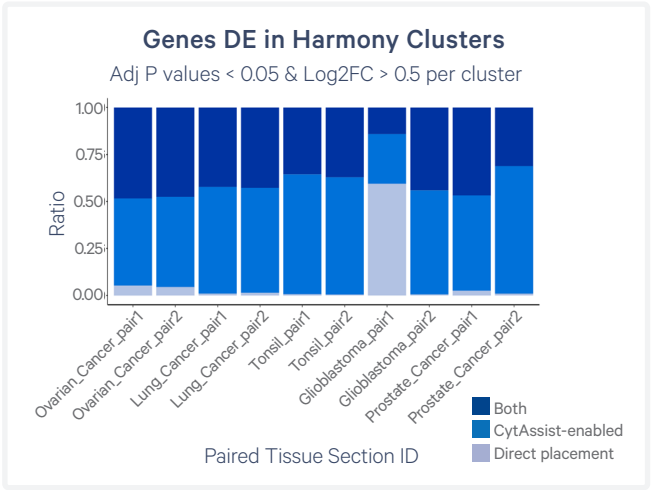


Figure 8. The CytAssist-enabled assay detects most of the differentially expressed (DE) genes detected by the Direct placement assay. **A.** Distribution of the upregulated differentially expressed genes in the human tonsil tissue sections run using the Direct placement and CytAssist-enabled assays, with higher number of unique genes detected in CytAssist-enabled assay. **B.** DGE analysis results measured using log2FC comparisons on clusters inferred from Harmony. Each dot represents one gene in one cluster. The x axis shows the average log2FC for Direct placement, the y axis for CytAssist-enabled assays respectively. The close match along the diagonal indicates congruence among cluster assignments between assays.



The overlap of the differentially expressed genes across all the Harmony clusters between paired tissue sections for each tissue type corresponding to Direct placement and CytAssist-enabled replicates were assessed (Figure 9). Overall a high concordance was observed between both the assays, indicating minimal impact of presence of gDNA to biological interpretation. In the Glioblastoma_pair1, the tissue section run with Direct placement assay contained low UMI count regions, which skewed the clustering results. Hence, the discordance observed in the overlap of the DE genes between the two glioblastoma tissue sections.

Figure 9. Concordant differentially expressed (DE) genes with positive log fold change in each pair of tissue sections (one tissue section each run with Direct placement and CytAssist-enabled assays) across all tissue types based on Harmony clustering.

Conclusion

This Technical Note quantifies the level and impact of reads originating from probe ligation events on gDNA and evaluates if biological interpretations of the data are robust. While an increase in signal from gDNA was observed in data derived using Visium CytAssist Spatial Gene Expression vs. Visium Spatial for FFPE assay (Table 2), the biological interpretations as measured by clustering and DGE are robust to the presence or absence of this signal. Because counts from gDNA are low and are randomly distributed across the spatial extent of the tissue under analysis, gDNA is unlikely to change biological interpretation of most data. However, very lowly expressed genes might be difficult to discern from background noise originating from gDNA. When studying less abundant genes (~1-2 UMIs/spot) or rare cell types in tissues using this assay, additional validation methods may be required to confirm the results (Figure 6C).

In challenging tissue types with lower mRNA content or dense cellular/nuclear content, the estimated gDNA signal could be higher. For example, in spleen tissue, the estimated gDNA is ~20%. It is challenging to extract mRNA from spleen and maintain RNA quality [3]. Furthermore, spleen is cellularly dense tissue which contributes to higher nuclear content per spot, leading to higher than average gDNA estimates. While the biological interpretations are unlikely to change for the vast majority of highly expressed genes, the number of genes with low expression conflated with background signal will be higher in these tissues and hence, the data should be interpreted accordingly.

Researchers can assess gDNA in their datasets using a metric and associated plot in the web_summary.html file output by Space Ranger. Information to determine which probes target exon-junction-spanning (spliced) and non-exon-junction-spanning (unspliced) regions are accessible in the probeset BED or CSV file or with additional metadata information in the probeset TSV file. Thus, use of Visium CytAssist Spatial Gene Expression assay will allow researchers to expand tissue types for interrogation, without compromising biological interpretation, despite the presence of low to moderate gDNA signal.

References

1. 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). https://doi.org/10.1038/s41592-019-0619-0
2. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-3587.e29. doi:10.1016/j.cell.2021.04.048
3. Madissoon, E., Wilbrey-Clark, A., Miragaia, R.J. et al. scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. Genome Biol 21, 1 (2020). https://doi.org/10.1186/s13059-019-1906-x

Document Revision Summary

Document Number	CG000605
Title	Visium CytAssist Spatial Gene Expression for FFPE: Robust data analysis with minimal impact of genomic DNA
Revision	Rev A
Revision Date	August 2022

Contact:
support@10xgenomics.com
10x Genomics
6230 Stoneridge Mall Road
Pleasanton, CA 94588 USA