Interpreting Space Ranger Web Summary Files for Visium Spatial Gene Expression for FFPE Assay

Introduction

The web summary file in the output folder of the Space Ranger analysis software is the initial point of reference for determining sample performance in the Visium Spatial Gene Expression for FFPE assay. This Technical Note presents an overview of web summary file interpretation, including the expected metrics and characteristic plots for libraries generated using the Visium Spatial Gene Expression for FFPE assay.

Interpreting Web Summary File Metrics

Representative summary files for Visium Spatial Gene Expression for FFPE libraries and other Space Ranger output files are available for download on the 10x Genomics Support website. Several metrics in the web summary file can be used to assess the overall success of an experiment, including sequencing, mapping, and spot metrics. A representative web summary file for a Visium Spatial Gene Expression for FFPE library is shown below (Figure 1).

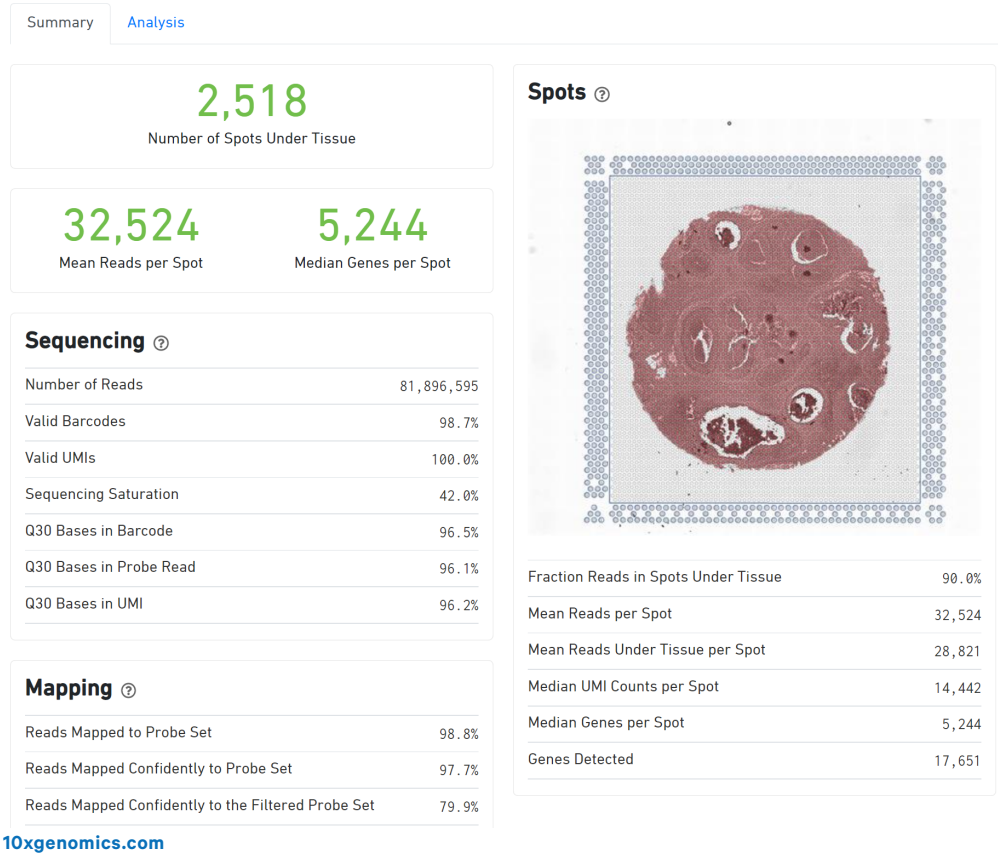


Figure 1. Web summary file from H&E stained breast cancer (ductal carcinoma in situ, invasive carcinoma) sample. The summary tab reports metrics that can be used to assess assay performance. The analysis tab contains secondary analysis results including tissue plots, t-SNE projections, and differential gene expression by cluster. Green text indicates that the key metrics are in the expected range while red/yellow text indicates errors/warnings. Descriptions of the metrics can also be found by clicking the icon ⓘ next to the section header.

Table 1. Metrics in the Space Ranger summary file

Metrics	Definition	Expected Value	Notes
Sequencing Metrics			
Number of reads	Total number of read pairs that were assigned to this library in demultiplexing	Sequencing output dependent	Lower than expected may indicate poor sequencing run (overclustering, underclustering, low % passing filter).
Valid barcodes	Fraction of reads with barcodes that match the whitelist* after barcode correction	>75%	Low valid barcodes may indicate sequencing issues (such as low Read 1 Q30 score).
Valid UMIs	Fraction of reads with valid UMIs; i.e. UMI sequences that do not contain Ns and that are not homopolymers	>75%	Low valid UMIs may indicate issues with sequencing or library quality.
Sequencing saturation	The fraction of reads originating from an already-observed UMI. This is a function of library complexity and sequencing depth	Dependent upon sequencing depth and sample complexity	Dependent on library complexity, sequencing depth, and experiment analysis goals. Lower sequencing saturation indicates a high proportion of the library complexity has not been captured by sequencing.
Q30 bases in barcode, Sample Index, or UMI	Fraction of tissue-associated barcode, Sample Index, or UMI bases with Q-Score ≥30, excluding very low quality/no call (Q≤2) bases from the denominator	Sequencing platform dependent	Low Q30 base percentages could indicate sequencing issue such as sub-optimal loading concentration.
Q30 bases in probe read	Fraction of RNA read bases with Q-score ≥ 30, excluding very low quality/no-call (Q ≤2) bases from the denominator. This is Read 2 for the Visium v1 chemistry	Sequencing platform dependent, ideally >65%	Expected to be lower than Q30 bases in barcode or UMI (Read 1) or Sample Index (i7 or i5 read) and is sequencing platform dependent. Low Q30 base percentages could indicate sequencing issue such as sub-optimal loading concentration.
Spots			
Fraction reads in spots under tissue	The fraction of valid barcode, confidently mapped to transcriptome reads with tissue-associated barcodes	Ideally >50%	Low fraction reads in spots under the tissue indicate that many of the reads were not assigned to tissue covered spots. This could be caused by high levels of ambient RNA resulting from inefficient permeabilization, because the incorrect image or orientation was used, or because of poor tissue detection. The latter case can be addressed by using the manual tissue selection option through Loupe.
Mean reads per spot	The number of reads, both under and outside of tissue, divided by the number of barcodes associated with a spot under tissue	25,000 reads pairs/tissue covered spot minimum recommended	The necessary sequencing depth per tissue covered spot depends on the sample type (high or low RNA) and the desired analysis.

* A whitelist is the list of all known barcode sequences that have been included in the Visium Spatial Gene Expression for FFPE Reagent Kits and are available during library preparation.

Mean reads under tissue per spot	The number of reads under tissue divided by the number of barcodes associated with a spot under tissue	Dependent on tissue type, RNA quality, and sequencing depth	Lower than expected values may be biological (low transcriptional diversity) or may indicate low sequencing depth, library complexity, or quality.
Median UMI counts per spot	The median number of UMI counts per tissue covered spot	Dependent on tissue type, RNA quality, and sequencing depth	
Median genes per spot	The median number of genes detected per tissue covered spot. Detection is defined as the presence of at least one UMI count	Dependent on tissue type, RNA quality, and sequencing depth	
Genes detected	The number of unique genes from the filtered probe set with at least one UMI count in any tissue covered spot	Dependent on tissue type, RNA quality, and sequencing depth	
Mapping Metrics			
Reads mapped to probe set	Fraction of reads that mapped to the probe set	Variable	Lower than expected values could be indicative of low library or sample quality or the use of the wrong probe set.
Reads mapped confidently to probe set	Fraction of reads that mapped uniquely to a probe in the probe set	Ideally >50%	Lower than expected values could be indicative of low aggregate expression, use of the wrong probe set, or inefficient targeting to the probe set.
Reads mapped confidently to the filtered probe set	Fraction of reads from probes that map to a unique gene. These reads are considered for UMI counting by default. This metric will be None when probe filtering is disabled. For more information on probe filtering visit the 10x Genomics Support website	Ideally >50%	

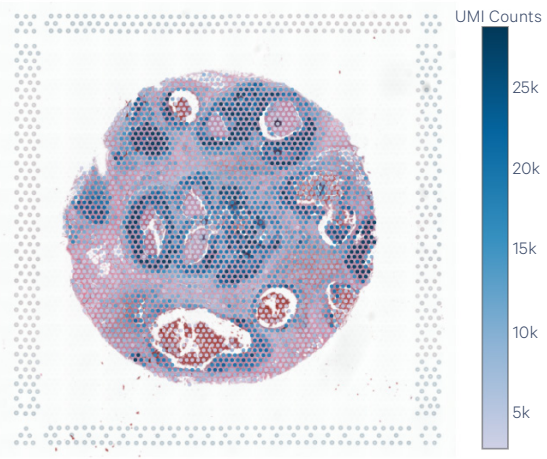
Interpreting the Web Summary File Plots

The summary file also contains tissue plots and t-SNE projection plots (Table 2) in the analysis tab.

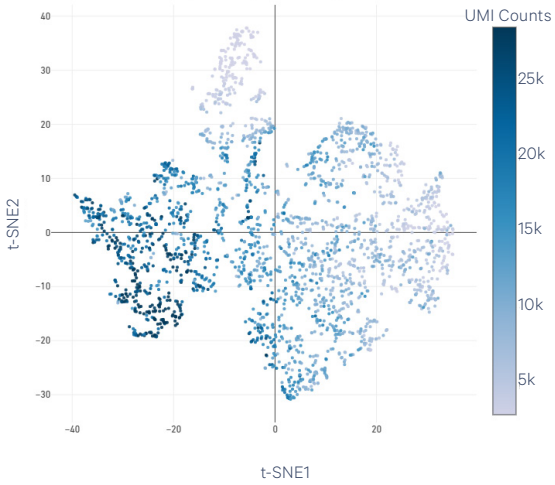
Table 2. Plots in the Space Ranger web summary file

Plots & Interpretation
Tissue Plot with Spots Colored by UMI Count: The UMI plot shows total UMI counts for each spot overlayed on the tissue image. Spots with greater UMI counts likely have higher RNA content than spots with fewer UMI counts. The color scale can be adjusted for different heat map colors.
t-SNE Projection of Spots Colored by UMI Counts: The t-SNE projection shows total UMI counts for spots displayed by a 2-dimensional embedding produced by the t-SNE algorithm. In this space, pairs of spots that are close to each other have more similar gene expression profiles than spots that are distant from each other. The color scale can be adjusted for different heat map colors.
Examples
Typical sample: High UMI counts in regions of high cell density or highly expressing cells and low UMI counts in regions of lower cell density or lowly expressing cells. Clear separation into distinct clusters. UMI counts within each cluster may vary in heterogeneous tissues.

Tissue Plot with Spots Colored by UMI Count

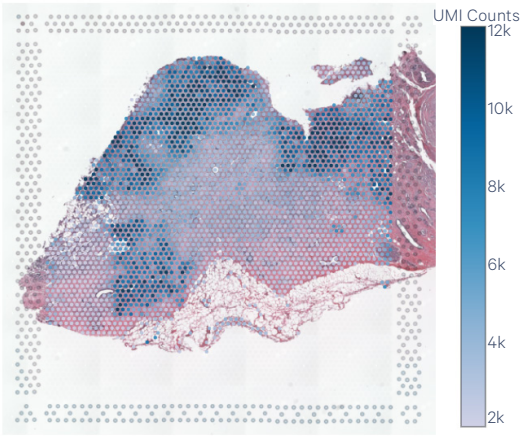


t-SNE Projection of Spots Colored by UMI Counts



Compromised sample: UMIs across tissue do not agree with tissue morphology. t-SNE plot may have a lack of cluster structure, one large cluster, or no separation. May indicate uneven application of reagent, tissue detachment, image alignment issue, or an image swap. See this Q&A article on identifying image or FASTQ swaps.

Tissue Plot with Spots Colored by UMI Count



t-SNE Projection of Spots Colored by UMI Counts

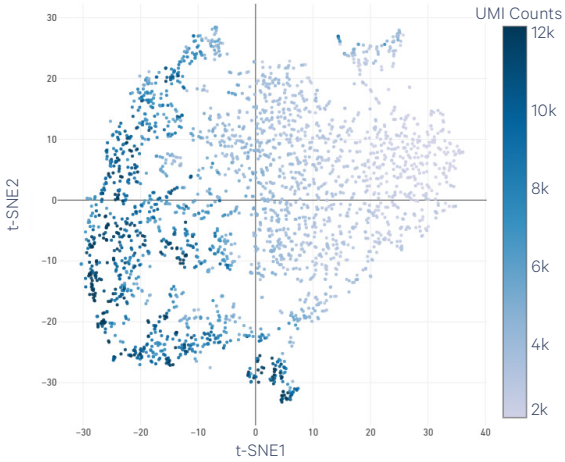
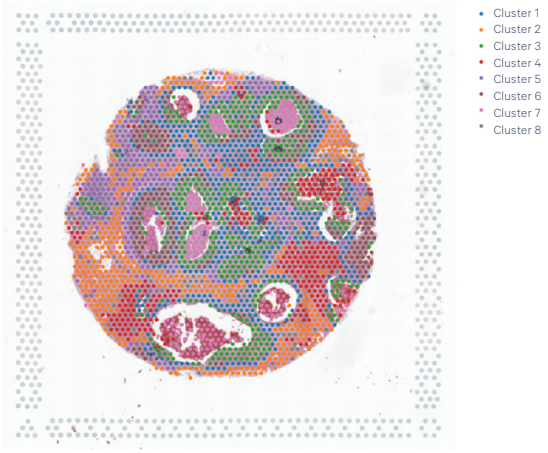


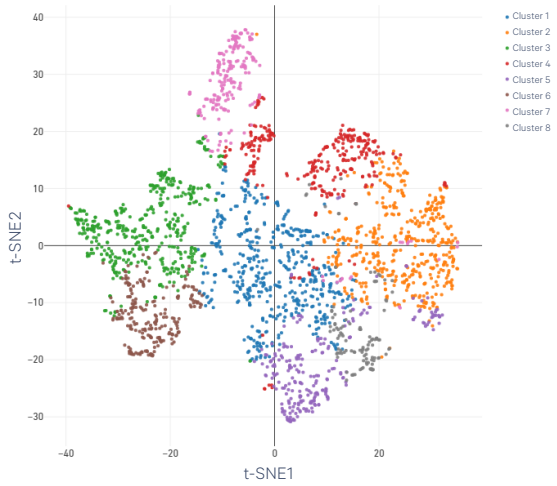
Table 2 contd. Plots in the Space Ranger web summary file

Plots & Interpretation
Tissue Plot with Spots Colored by Clustering: The clustering plot shows assignments of spot-barcodes to clusters by an automated clustering algorithm across the tissue. The clustering groups together spots that have similar expression profiles. Spots are colored according to their cluster assignment and projected onto the tissue image. Only spots under tissue are used in the clustering algorithm.
t-SNE Projection of Spots Colored by Clustering: The t-SNE projection shows clusters for spots displayed by a 2-dimensional embedding produced by the t-SNE algorithm. The axes correspond to the 2-dimensional embedding. In this space, pairs of spots that are close to each other have more similar gene expression profiles than spots that are distant from each other. The display is limited to a random subset of spots.
Examples
Typical sample: Clustering mirrors the tissue morphology. UMI counts within each cluster may vary in heterogeneous tissues. Clustering may not appear discrete with homogeneous tissues.

Tissue Plot with Spots Colored by Clustering

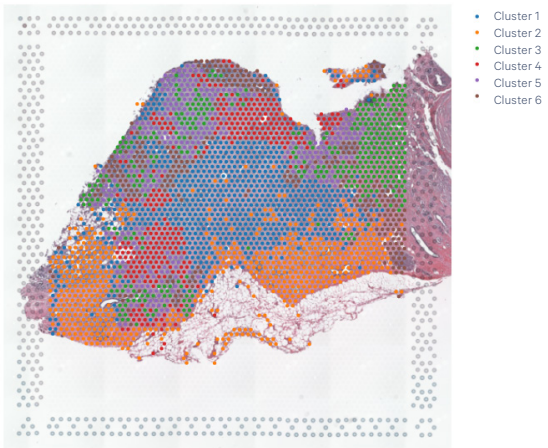


t-SNE Projection of Spots Colored by Clustering



Compromised sample: Clustering across tissue does not agree with tissue morphology. t-SNE plot may have a lack of cluster structure, one large cluster, or no separation. May indicate uneven application of reagent, tissue detachment, image alignment issue, or an image swap.

Tissue Plot with Spots Colored by Clustering



t-SNE Projection of Spots Colored by Clustering

Document Revision Summary

Document Number	CG000499
Title	Interpreting Space Ranger Web Summary Files for Visium Spatial Gene Expression for FFPE Assay
Revision	Rev A
Revision Date	February 2022

References

1. Visium Spatial Gene Expression Reagent Kits for FFPE User Guide (Document CG000407)

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