Q² Solutions®

Single-Cell RNA Profiling of FFPE Tissue

APPLICATION NOTE

Q2LabSolutions.com

Main Takeaways

- Enables the interrogation of FFPE tissue at single cell resolution for highly reproducible gene expression evaluations.
- Demonstrates cross-platform agreement between single cell RNA-seq with bulk RNA-seq from clinically relevant FFPE tissues.
- Characterizes QC expectations and recommendations for enabling successful single-cell RNA-seq processing of FFPE tissues.

Single-cell RNA sequencing (scRNA-seq) uncovers tissue heterogeneity by providing gene expression measurements in individual cells within the tissue. This level of resolution allows for a clearer understanding of specific cellular functions, tissue complexity, and is a powerful tool to discover therapeutic targets and novel biomarkers. A current limitation of scRNA-seq is the requirement to isolate viable cells from fresh or cryopreserved solid tumor biopsies, especially given that formalin-fixed paraffin embedded (FFPE) material are ubiquitously utilized for long-term tissue preservation. Though formalin fixation preserves samples for extended periods of time, it also fragments genetic material, which introduces challenges for many RNA-seq methods due to the reduced size of template material. 10x Genomics recently launched the Chromium Single Cell Gene Expression Flex assay, a probe-based solution for scRNA-seq from fixed cells to complement traditional bulk RNA sequencing methodologies. Herein, this discussion describes and characterizes the utilization of prepared single-cells or isolated RNA material from FFPE blocks of varying tissue types, ages, and quality. Matched samples were evaluated using the Single Cell Gene Expression Flex and a bulk RNA sequencing strategy to provide orthogonal comparisons across the two assays. The total number of genes detected by the Single Cell Flex assay was highly reproducible between replicates and ranged between 14,000 and 18,000 per tissue type, indicating robustness of the assay across tissues. While FFPE curl number and size input variables were found to have no significant impact on single-cell QC metrics, the effect of DV200 was observed within each tissue type, indicating a strong driver of metric variability. Qualitative examination of the correlations between scores reflecting immune cell activity from bulk sequencing and cell type prevalence from scRNA-seq display strong linear association; scRNA-seq and bulk RNA sequencing provide consistent information regarding immune cell presence reflecting high quality cell and RNA capture from the Single Cell Flex assay. These combined findings support the use of the 10x Genomics Single Cell Gene Expression Flex assay to obtain high quality single cell gene expression data from FFPE tissues. The data also demonstrate that meaningful gene expression information at single cell resolution is produced by multiple clinically relevant tissue types and input amounts, but samples with the highest DV200 produce the most robust data.

“The total number of genes detected by the Single Cell Flex assay ranged between 14,000 and 18,000 per tissue type.”

“These findings support the use of the Single Cell Gene Expression Flex assay to obtain high quality single cell gene expression data from FFPE tissues.”

Background and Objectives

- Single cell RNA sequencing (scRNA-seq) uncovers tissue heterogeneity by providing gene expression measurements in individual cells within a tissue.
- A current limitation of scRNA-seq is the requirement to isolate viable cells from fresh or cryopreserved solid tumor biopsies, especially given that formalin-fixed paraffin embedded (FFPE) materials are ubiquitously utilized for long-term tissue preservation.
- Tissue fixation fragments genetic material, which introduces challenges for RNA-seq methods due to the reduced size of template material.
- The objective of this study was to describe and characterize the utilization of single cells prepared from FFPE tissue sections to establish sample requirements, key quality control metrics, and reproducibility of the 10x Genomics Gene Expression Flex workflow.

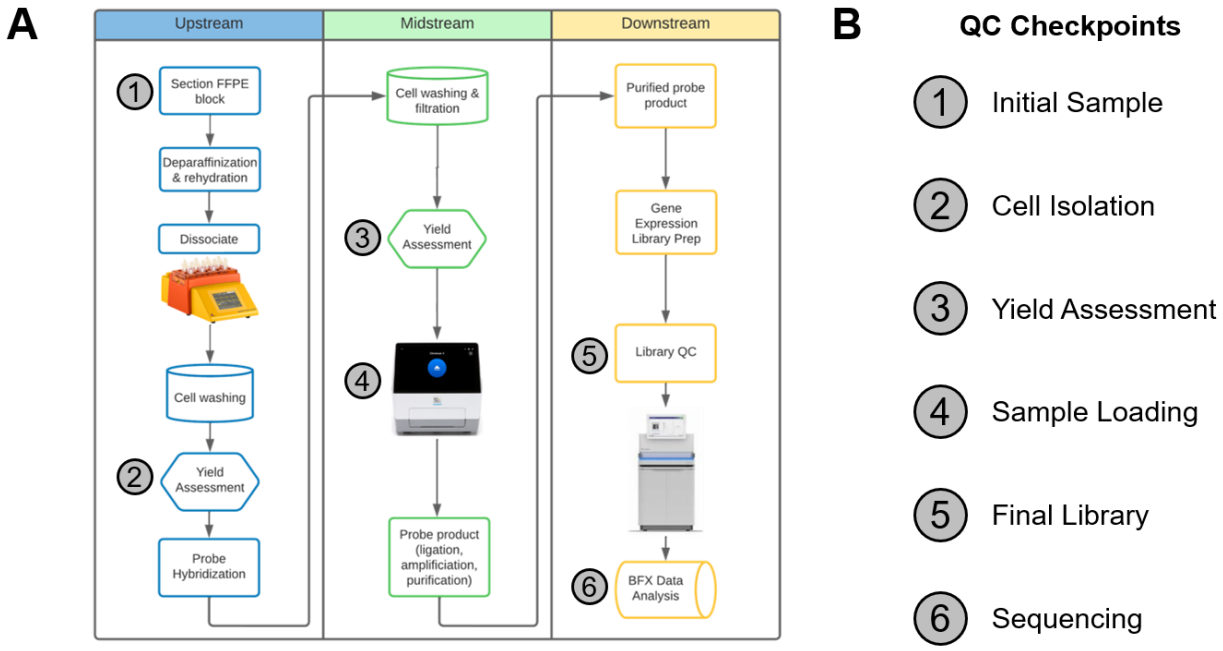
Methods (Figure 1)

- FFPE tissue blocks from varying tissue type, age, and quality (i.e., DV200) were sectioned to either 25 or 50 μm. The ovarian cancer FFPE block is $>$ 25 years old and the testes block is an unknown age. All other FFPE tissues are $<$ 5 years old.
- One or two FFPE tissue curls per sample were sectioned 1-6 weeks prior to processing and were enzymatically digested into single cell suspensions using the 10x Genomics recommended FFPE dissociation workflow with modifications for high throughput scRNA-seq.
- Isolated cells were counted using an automated fluorescent cell counter then probe-hybridized using the 10x Genomics Gene Expression Flex workflow.
- Probe-hybridized cells were washed then processed using the 10x Genomics Chromium X instrument into gel-beads in emulsion (GEMs) by combining barcoded gel beads, master mix, partitioning oil, and a single cell.
- Barcoded and ligated products were recovered, and sequencing libraries generated via sample index PCR. Samples were sequenced using Illumina NovaSeq with a depth of 20,000 read pairs per cell and analyzed by CellRanger 7.0.1.

FFPE blocks of varying ages, ranging from 5-25 years, and quality were evaluated for characterizing assay expectations.

FFPE samples were sectioned 1-6 weeks prior to downstream processing to investigate curl stability.

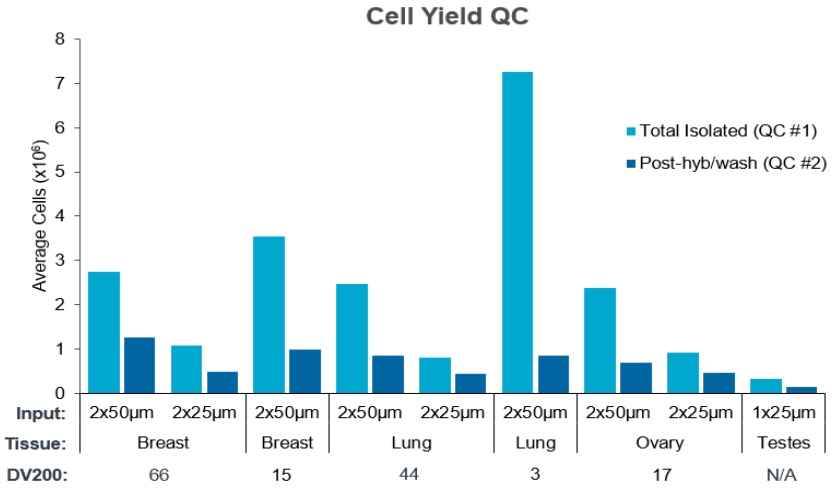
Figure 1. FFPE tissue dissociation and Gene Expression Flex workflow with associated QC checkpoints. The workflow is divided into upstream (blue), midstream (green), and downstream (yellow) components that broadly encompass dissociation and probe hybridization, GEM Generation, and sequencing steps, respectively (**A**). QC checkpoints assess sample quality throughout the workflow for ensuring high quality data outcomes (**B**).



Results

- All samples yielded sufficient cells for probe hybridization and for targeting the capture of 10,000 cells for downstream analysis (**Figure 2**).
- Samples with < 2 sections, or thickness < 50 μm, featured lowest isolated cell yield (**Figure 2**).

Figure 2. Cell yield assessments from dissociated FFPE tissue sections. The maximum number of available isolated cells, or up to 2 x 10^6 cells from each sample, underwent probe hybridization. Post hybridization / washing counts represent the cells available for GEM generation. Data represent the average cell counts (n = 1-3 replicates per tissue).



- The total number of genes detected was highly reproducible between replicates and ranged between 14,000 and 18,000 per tissue type (**Figure 3**).
- DV200 scores strongly correlate with key QC metrics, specifically cell capture efficiency and RNA species diversity (**Figure 3, Figure 4**).

Figure 3. DV200 scores strongly correlate with key sequencing QC metrics. Correlation plots for DV200 score against the estimated number of cells, fraction of reads in cells, and median genes per cell in breast, lung and ovary samples (**A**)**.** Box plots showing estimated number of cells, fraction of reads in cells and median genes per cell in breast, lung, ovary, and testes sample. Also plotted are the amount of tissue input and the DV200 score (**B**). The DV200 scores of testis samples are unknown, but data are consistent with high quality input.

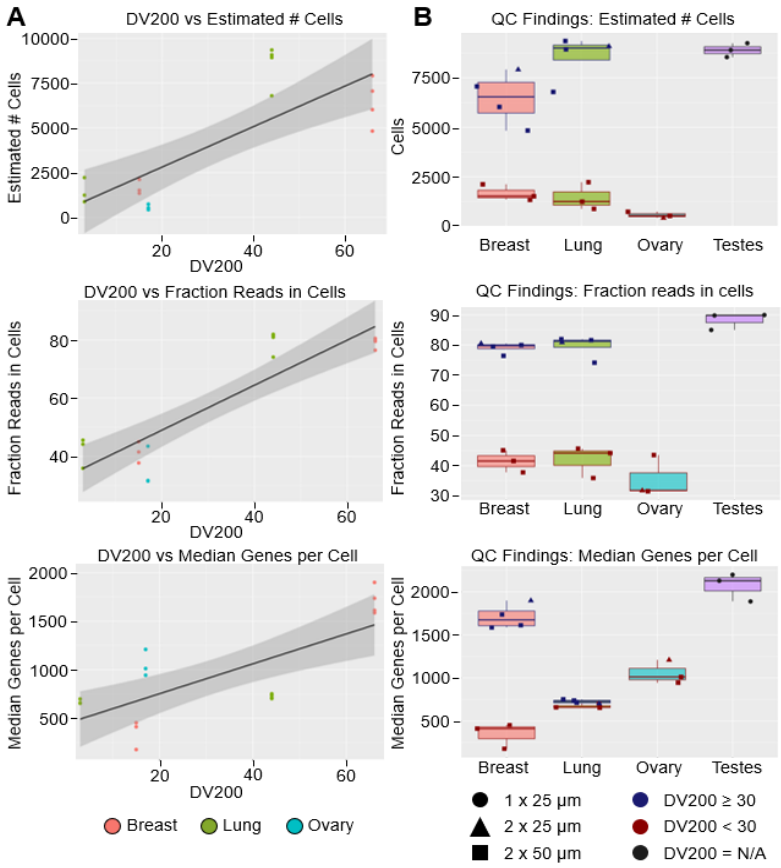
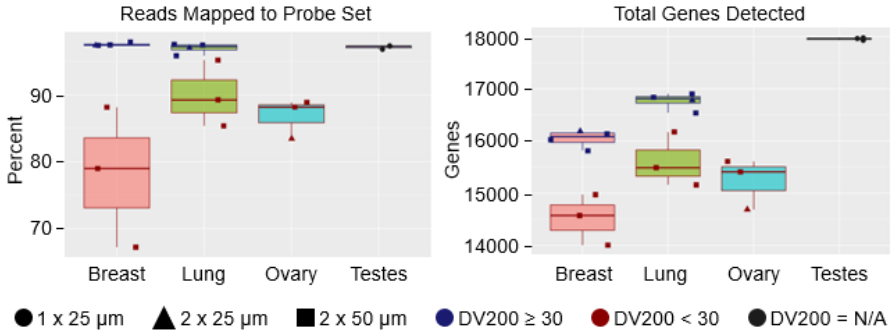
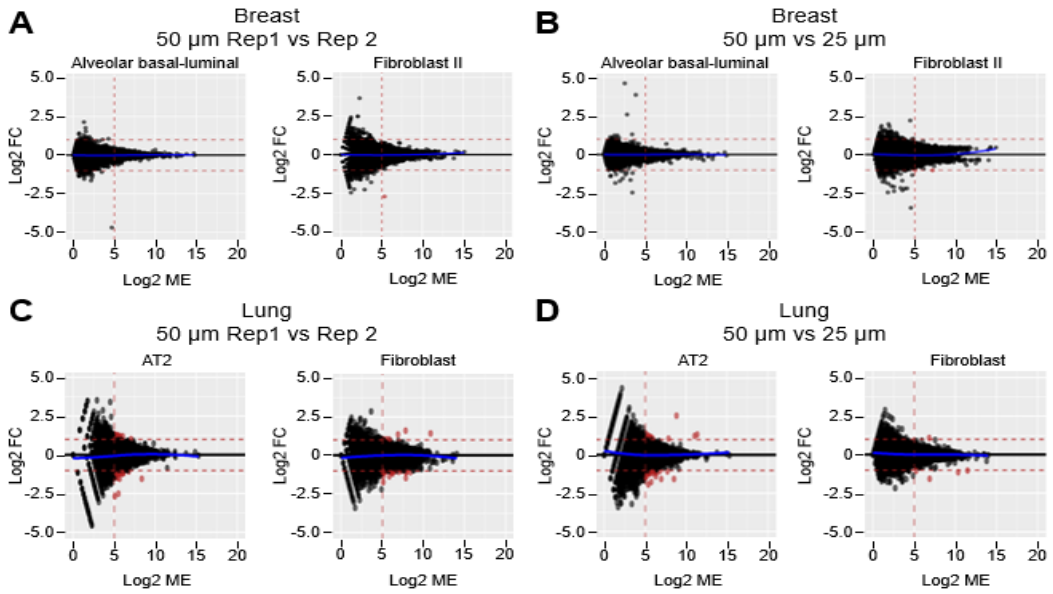


Figure 4. Probe set mapping and gene detection are increased in samples with DV200 scores greater than 30. Box plots showing reads mapped to probe set and total genes detected in breast, lung, ovary, and testis samples. Also plotted are the amount of tissue input and the DV200 score. The DV200 scores of testis samples are unknown, but data are consistent with high quality input.



- Bland-Altman (BA) plots display minimal genes with high expression and high log2 fold-changes across replicates and within multiple tissue types. This implies limited systematic bias in these data related to variability between replicates of the same sample or different input amount. Furthermore, the data display a high degree of reproducibility (**Figure 5**).

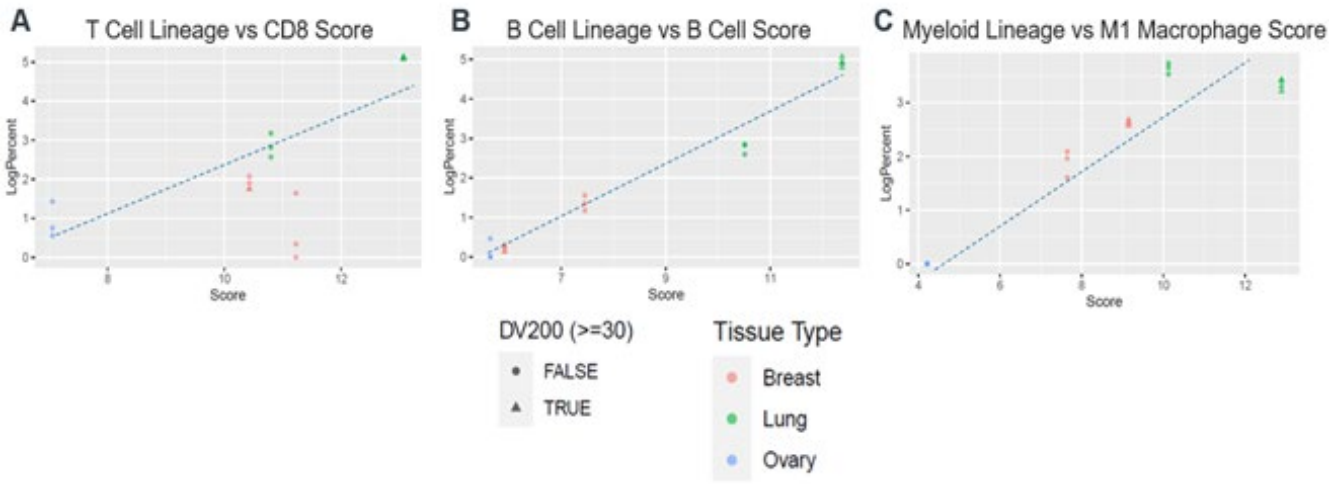
Figure 5. Bland-Altman plots imply a high degree of reproducibility between replicates of the same sample or different curl size. The log2 fold-change (FC) and log2 mean-expression (ME) of gene expression in representative cell types were plotted against one another for breast (**A,B**) or lung (**C,D**) tissue. Shown are comparisons between 50 μm replicates (**A,C**) and differing curl size (**B,D**).



- To assess the biological relevance of information of data resulting from the Flex assay and its agreement with a well-characterized assay, an additional set of matched FFPE samples were processed using bulk RNA-sequencing.
    - Single Cell RNA-Sequencing Data Description:
        - For all single cell experiments, the Azimuth system (v0.4.6) was applied using tissue specific references to annotate cell identity. The cell calls provided by Azimuth were determined to be of high quality in relevant immune cell categories (i.e., T-, B- and myeloid cells) through examination of cell-specific marker gene expressions.
        - For each replicate, the percentage of all cells captured belonging to T-cells (CD4+ and/or CD8+), B-cells (inclusive of plasma cells) and myeloid cells (inclusive of dendritic, macrophage, monocytes, and more) were computed.
    - Bulk RNA-Sequencing Data Description:
        - Each FFPE tissue block was further assessed using bulk RNA-sequencing, one sample per block.
        - Q^2 Solutions Genomics' proprietary Immune Landscape Signature (ILS) scores were computed for each sample. ILS scores reflect immune cell abundance and activity across varied immune categories (e.g., CD8 Score, B Cell Score, M1 Macrophage score). A higher score indicates increased expression of genes relevant to the category which, in turn, indicates increased immune cell presence and/or activity of the implicated cell type.

- Comparison:
  - Percentages of T-, B- and myeloid cells were computed in single cell experiments as described for each replicate. Each replicate was then matched with its paired bulk sequencing sample forming a 6-tuple information vector for each replicate: [T-Cell proportion, B-Cell proportion, Myeloid Proportion, CD8 ILS Score, B Cell ILS Score, M1 Macrophage ILS score]
  - Percentages were transformed to log2-scale and plotted against the associated ILS score as displayed in figure 6; T-cells map to CD8 ILS score, B-Cells to the B Cell ILS score, and myeloid cells to the M1 macrophage score.
- Observations (**Figure 6**):
  - A positive relationship is observed between ILS score and cell type percentage for each of the three immune cell types captured in this experiment.
  - This suggests that cell capture is reasonable representative of the bulk tissue in single cell and that expression information is sufficient for reasonable, accurate cell type identification.
  - While all samples are displayed, we do note that low DV200 is observed to cause high variability and potential deviations from "perfect" agreement across the two platforms.

Figure 6. Cross-platform agreement in immune cell abundance for T-, B- and Myeloid cells. The log2-transformed percentage of immune cells for any given single cell experiment (x-axis) is plotted against the Immune Landscape Signature score from bulk RNA-sequencing (y-axis) to assess consistency in called cell populations across the two platforms. The positive associations observed reflect that single cell and bulk RNA-seq assays are reflecting biologically consistent information regarding these immune populations.



Note: The lines depicted in this figure are illustrative. They are not lines of best fit.

Conclusions

- Fixation stabilizes samples and provides flexibility in study design, decision making, and storage.
- Single cell analysis is now attainable for FFPE samples, including archived tissues and planned studies.
- Fixed single cells can be reliably recovered from FFPE tissue sections.
- The 10x Genomics Gene Expression Flex Kit yields high quality single cell gene expression data from FFPE tissues.
- Data suggests that FFPE curls sectioned and stored 1 to 6 weeks prior to downstream processing meets expectations for key QC metrics.
- Meaningful gene expression data is produced by multiple tissue types and input amounts, but samples with the highest DV200 produce the most robust data.
- Gene expression patterns are highly reproducible across different cell types and replicates of the same sample, and different input amounts.
- scRNA-seq and bulk RNA sequencing provide consistent information regarding immune cell presence reflecting high quality cell and RNA capture from the Single Cell Flex assay.
- Data analysis suggest sample input of 1 x 25 μm with DV200 $\geq$ 30 yields adequate cells for processing, with lower inputs currently being evaluated (**Table 1**).

Table 1. Key QC Metrics for scRNA-seq of FFPE tissue. Q^2 Solutions recommends submitting material as informed in the table to ensure sufficient yield of intact cells for achieving high quality downstream data. Samples will be assessed according to established QC metric expectations throughput processing.

1 - Initial Sample QC	$\geq$ 1 x 25 μm curls with DV200 > 30
2 - Cell Isolation QC	$\geq$ 200,000 cells
3 - Sample Input QC	$\geq$ 60,000 cells
4 - Final Library QC	Library size: 100 – 500 bp Concentration: > 2 nM
5 - Sequencing QC	Target cell capture: up to 5,000 or **10,000** Recommended Mean Reads per Cell: $\geq$ 10,000 for cell population assessment $\geq$ **20,000** for transcript assessment Q30 Bases in Barcode/UMI: > 70%

Authors and References

Q^2 Solutions Authors:

1. Adam Black, Senior Scientist II, Translational Genomics
2. Doug Wilson, Senior Biostatistician, Translational Genomics
3. Joseph Modarelli, Associate Principal Scientist, Translational Genomics

References:

1. Max Karlsson et al., "A single–cell type transcriptomics map of human tissues". Sci.Adv.7, eabh2169(2021). DOI:10.1126/sciadv.abh2169

2. Andrew Butler, Charlotte Darby, Yuhan Hao, Paul Hoffman and Rahul Satija (2022). Azimuth: A Shiny App Demonstrating a Query-Reference Mapping Algorithm for Single-Cell Data. R package version 0.4.6. https://github.com/satijalab/azimuth

3. Stuart, Tim, et al. "Comprehensive Integration of Single Cell Data." Cell, vol. 177, no. 7, 13 June 2019, https://doi.org/10.1101/460147.

4. Sikkema, L., Ramírez-Suástegui, C., Strobl, D.C. et al. An integrated cell atlas of the lung in health and disease. Nat Med 29, 1563–1577 (2023). https://doi.org/10.1038/s41591-023-02327-2

5. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

6. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

7. Jones WD, Michener CM, Biscotti C, Braicu I, Sehouli J, Ganapathi MK, Ganapathi RN. RNA Immune Signatures from Pan-Cancer Analysis Are Prognostic for High-Grade Serous Ovarian Cancer and Other Female Cancers. Cancers (Basel). 2020 Mar 7;12(3):620. doi: 10.3390/cancers12030620. PMID: 32156016; PMCID: PMC7139955.

To learn more about our clinical genomic services, please visit
www.Q2LabSolutions.com/genomics-laboratories

Contact Us

Toll free: +1 855.277.9929
Direct: +1 919.998.7000
International: +44 (0) 1506 814000

Q² Solutions®