Date: August 9, 2024

Cancer Immune Monitoring and Analysis Center
Dana-Farber Cancer Institute
Harvard Medical School
Contact PI: Catherine J. Wu, MD (e-mail: Catherine_wu@dfci.harvard.edu)

Performance lab:
Translational Immunogenomics Lab (TIGL)
Catherine J. Wu, MD (Faculty Advisor)
Kenneth J. Livak, PhD (Technology Lead Scientist)
Shuqiang Li, PhD (Technology Lead Scientist)
Jack Ghannam (Graduate Student)
Jeremy M. Simon, PhD (Senior Research Scientist)

Single Nucleus RNA Sequencing (snRNAseq) Analytical Performance Report, version 2.0

1. Intended use

Single cell RNA sequencing (scRNAseq) has become the definitive method for determining cell types and cell states,[1] but the preference for using fresh tissue makes it challenging to apply scRNAseq to clinical solid tumor samples. The challenges of working with fresh tissue include:[2] (1) Quick, enzymatic dissociation to single cells must be optimized for each tumor type. (2) The dissociation method may lead to differential recovery of cell types or changes in gene expression. (3) There must be close coordination between the tissue acquisition and processing teams, which is often difficult to reconcile with surgical scheduling. (4) Longitudinal samples must be processed on separate days, which may introduce batch effects. The advantage of single nucleus RNA sequencing (snRNAseq) is that it can be performed on frozen samples. This immediately de-couples sample acquisition and sample processing, reducing concerns related to the timing of surgery, enabling analysis of archived frozen samples, and facilitating analysis of longitudinal samples. Also, 10x Genomics has a standardized nuclei isolation kit and protocol that only needs minor modifications to use for different tumor and tissue types. Finally, snRNAseq is often preferred for certain tissues, such as brain[3,4] and retina,[5] that are difficult to dissociate into single cell suspensions. Thus, the intended use of snRNAseq is to obtain single nuclei transcriptome profiles from tissue samples that have been flash frozen in order to identify cell types and cell states in the samples being analyzed.

Table 1. Summary of analytical performance findings for snRNAseq

Accuracy	Accurate assay performance is determined by length distribution of cDNA libraries (**Figure 1**) as compared to reference provided by 10x Genomics. We show that the distribution of cell types observed for a representative glioblastoma sample is consistent with a comprehensive map of cellular states established for glioblastoma[6] (**Figure 10**).
Precision	Correlation coefficients comparing replicates range from 0.80 to 1 (mean 0.97; median 0.99) for gene-by-gene comparison of pseudobulk expression levels in the most prevalent cell types from four glioblastoma samples (**Figures 6-9**).
Analytical sensitivity	We determine this by median detection of transcripts from at least 1000 genes per nucleus. Sensitivity reflects the expression level limit for detecting transcripts. More transcripts detected translates to more genes detected per nucleus.
Analytical specificity including interfering substances	Specificity is determined by the well-established method of RNA sequencing. Cell Ranger uses the alignment tool STAR[7] and filters out any reads that do not uniquely map to a single gene. Interfering substances are not applicable to RNA sequencing.
Reportable range	Detection of 20,000 to >33,000 genes per sample. We consistently achieve a median detection of 1,000 to >3,400 genes per nucleus.

Reference interval (normal range)	The 10x Genomics kit used here targets recovery of results from 2,000 to 20,000 cells or nuclei.
Standardization, harmonization, reproducibility, and ruggedness	Standardization and ruggedness have been achieved by using commercial kits following the manufacturer's protocol. All analyses were performed using the same reference database.
Quality control and improvement procedures	As explained in detail in section 2 **(Materials and methods)**, quality control is incorporated as part of the routine processing of the data. Run metrics after QC filtering are summarized in **Table 2**.
Any other performance data	Not applicable.

2. Materials and methods

Banking frozen tissue samples:

1. Tissue in a cryovial was flash frozen by either submerging in liquid nitrogen/liquid-nitrogen cooled bath (e.g. isopentane) or placing the tube deep in a bucket of dry ice. Wait at least 2-3 minutes for the tissue to freeze all the way through.
2. The tube containing the tissue was transferred to vapor phase liquid nitrogen for long-term storage.
3. Once removed from liquid nitrogen, tissue was maintained at -80°C or on dry ice until just before suspension processing.

We report the results of processing of frozen tissue samples (designated 124_PreVax, 129_PostVax, 133_PreVax, and 134_PostVax) from four glioblastoma (GBM) patients as a representative workflow for this assay. Specimens were thawed and processed to generate single nuclei suspensions using the Chromium Nuclei Isolation with RNase Inhibitor Kit (cat. no. 1000494) from 10x Genomics, following manufacturer's instructions with minor modifications to enhance myelin removal. Per the protocol guidelines of using 3-50mg input, 15-20mg tissue were used to prepare each sample. For each sample, the suspension was split into two replicates before processing in separate lanes using the Chromium Next GEM Single Cell 5' HT Kit v2 (cat. no. 1000356) from 10x Genomics following the manufacturer's protocol. Transcriptome libraries were sequenced on the Illumina NovaSeq using the parameters set by 10x Genomics. The use of commercial kits and processes standardizes the acquisition of snRNAseq data.

Processing of snRNAseq sequencing data was performed in the following manner:

1. The Cell Ranger v7.2.0 pipeline (10x Genomics) was run to generate single-nucleus gene expression matrices.
2. CellBender v0.3.0[8] was used to remove background noise in the gene expression matrices.
3. Quality control and filtering were performed leveraging *ddqc*,[9] a quality control package that applies an adaptive threshold across cell clusters based on the median absolute deviation (MAD) of four QC metrics to set lower bounds for number of UMI counts and number of genes detected and upper bounds for fraction of reads mapping to mitochondrial or ribosomal genes such that nuclei at >2 MADs below their cluster median for counts and genes or above their cluster median for mitochondrial or ribosomal reads are removed from downstream analysis.
4. Resulting filtered gene-barcode matrices were processed using Seurat v5 (https://satijalab.org/seurat/).
    a. Matrices were normalized with the 'NormalizeData' function using the 'LogNormalize' method.
    b. The top 2,000 variable genes were identified using the 'vst' method in the 'FindVariableFeatures' function.
    c. Gene expression matrices were scaled and centered using 'ScaleData'.
    d. Linear dimensionality reduction performed using principal component analysis (PCA).
    e. Cells were clustered using the Louvain algorithm using the top 30 principal components.

f. Non-linear dimensionality reduction performed using uniform manifold approximation and projection (UMAP).

5. scDblFinder v1.16.0 (https://bioconductor.org/packages/release/bioc/html/scDblFinder.html) was applied to identify and remove doublets with an expected doublet rate ~8% based on the loading rate.

Dot plots and correlation plots comparing the replicates were generated using the 'Introduction to scRNA-seq integration' vignette (https://satijalab.org/seurat/articles/integration_introduction.html#identify-conserved-cell-type-markers).

3. Results

Single-cell cDNA Libraries

The initial quality control metric for assessing the likelihood of a successful assay is the size distribution of the full-length cDNA library generated as an intermediate step in the 10x Genomics protocol. **Figure 1** shows these size distributions for the eight samples analyzed in this report.

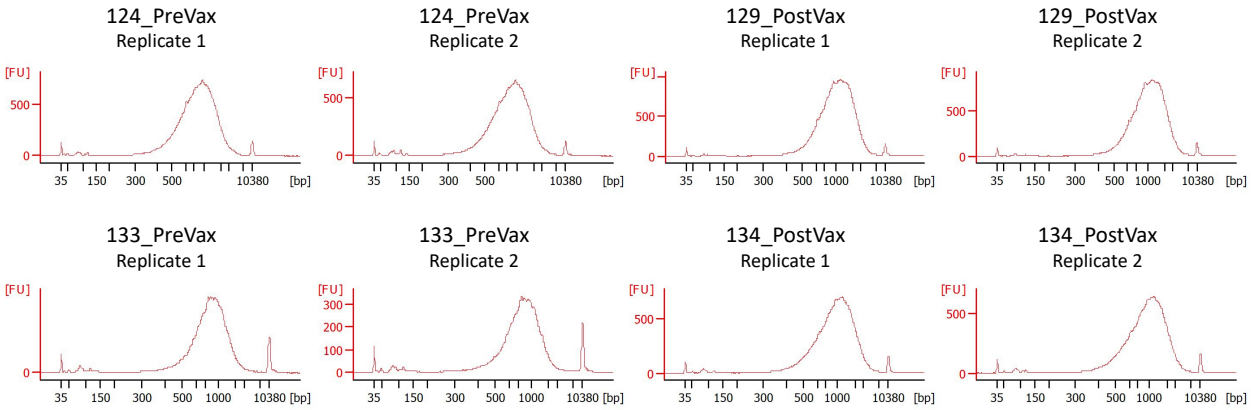

Figure 1. Quality control of snRNAseq cDNA libraries. Agilent Bioanalyzer traces showing the fragment size distribution of the cDNA library for each of the indicated samples.

These results match the representative traces found in the Chromium Next GEM Single Cell 5' HT Kit v2 protocol CG000423 and in a published protocol for preparing snRNAseq libraries from brain tissue.[10]

Nuclei and Gene Recovery

Table 2. Recovery statistics for 10x Genomics Chromium runs

	Replicate	Estimated Number of Nuclei Loaded	Number of Nuclei Recovered	Fraction Reads in Nuclei	Fraction Reads Mapped to Transcriptome	Mean Reads per Nucleus	Median UMI Counts per Nucleus	Median Genes per Nucleus	Total Genes Detected	Sequencing Saturation
124_Prevax	1	31,100	14,058	84.8%	61.8%	38,831	5,752	2,684	31,751	0.603
124_Prevax	2	31,100	29,064	79.6%	58.6%	14,402	1,232	964	31,524	0.512
129_Postvax	1	32,200	15,348	91.7%	65.3%	40,722	7,125	3,460	33,037	0.637
129_Postvax	2	32,200	15,247	91.7%	65.1%	39,676	7,046	3,416	33,048	0.630
133_Prevax	1	30,200	9,185	89.5%	51.5%	56,869	5,433	2,642	30,665	0.729
133_Prevax	2	30,200	15,309	69.6%	47.3%	36,516	1,481	1,110	30,528	0.707
134_Postvax	1	29,400	12,071	93.6%	64.3%	61,473	6,118	2,579	33,359	0.713
134_Postvax	2	29,400	13,338	92.2%	64.7%	46,779	5,366	2,377	33,222	0.673

For replicate 2 of the 124_PreVax and replicate 2 of the 133_Prevax sample, the number of nuclei recovered may be inflated due to the spurious inclusion of droplets containing ambient RNA and cell debris rather than nuclei. This leads to depressed values for number of median genes per nucleus. For replicate 2 of 124_Prevax, this increases the number of nuclei analyzed to greater than the targeted range of 2,000 to 20,000 nuclei and reduces the median number of genes detected to fewer than the minimum expected value of 1,000 genes per nucleus. The results from these spurious "nuclei" are removed during subsequent quality control steps. That is why the results of **Figures 6 and 8** indicate that the expression values for these two replicates are still reliable.

Major Cell Types

Seurat used Louvain clustering to identify major cell types. **Figures 2-5** compares the two replicates of each sample for the expression levels of the following canonical genes:

Oligodendrocytes: *PLP1*, *MAG*, *CLDND1*, *GPM6B*
Pericytes: *LAMA2*, *PRKG1*, *COL4A1*, *VIM*
Endothelial cells: *COL4A1*, *VWF*, *PECAM1*, *VIM*
Neurons: *RBFOX3*, *MAP2*, *SYN1*, *GPM6B*
Lymphocytes: *CD2*, *CD3E*, *CD5*
Myeloid cells: *CD74*, *CD163*, *CD83*
Tumor cells: *GAP43*, *GPM6B*, *VIM*, *MAP2*

In these figures, the nuclei for each of the indicated cell types are assessed for whether they express the indicated gene, with the size of the dot indicating the percent of cell type-specific nuclei expressing the gene. The results indicate that the two replicates for each sample are very similar qualitatively in the distribution of cell types.

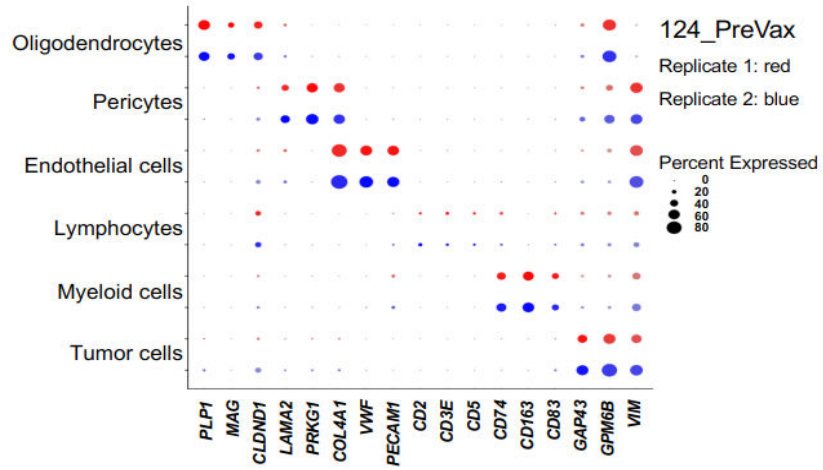

Figure 2. Expression of canonical genes for sample 124_PreVax.

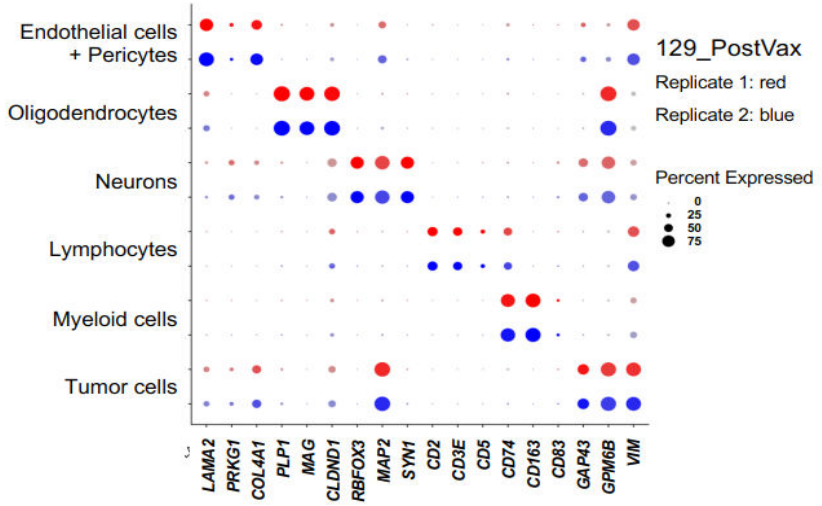

Figure 3. Expression of canonical genes for sample 129_PostVax.

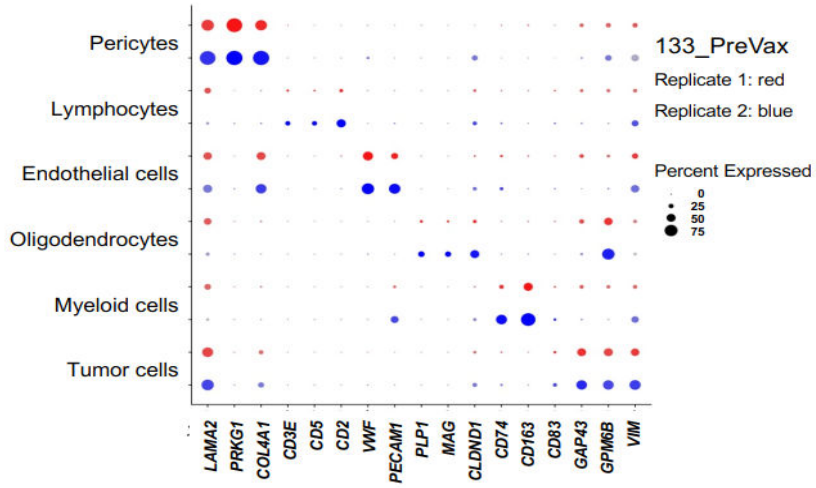

Figure 4. Expression of canonical genes for sample 133_PreVax.

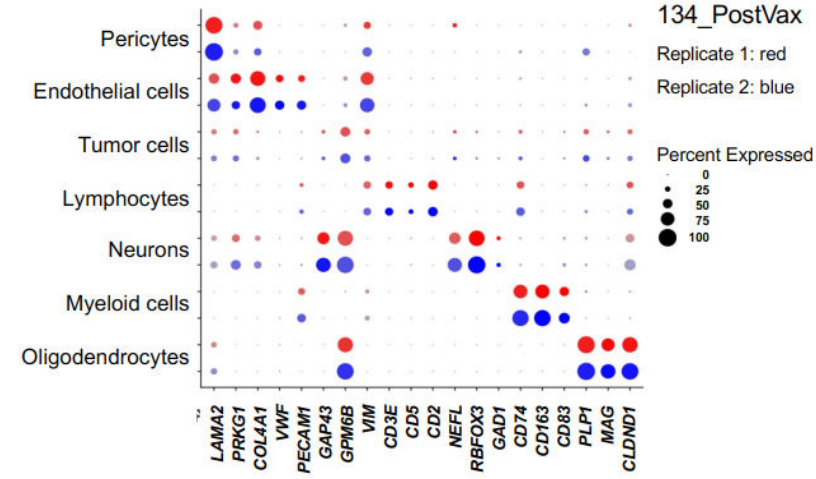

Figure 5. Expression of canonical genes for sample 134_PostVax.

Quantitative Comparison of Replicates

The sparsity of single-cell transcriptome data and the large cell-to-cell fluctuations in transcript levels caused by the bursting nature of eukaryotic transcription[11] mean that reproducibility cannot be meaningfully assessed at the individual cell level. Thus, to perform a quantitative comparison of the two replicates for each of the four samples, a pseudobulk analysis was performed where the expression levels for all the cells of the same cell type were aggregated. In **Figures 6-9**, each point corresponds to a different gene. For each gene, the UMI counts were summed for all the cells from the same sample for each cell type. This sum was normalized by the total UMI counts across all genes and then multiplied by an arbitrary scale factor of 10,000. Each of these values was natural log transformed to generate the expression level graphed in the figures. The *r* values shown are the Pearson correlation coefficient. The results indicate excellent reproducibility for snRNAseq with correlation coefficients ranging from 0.80 to >0.99 (mean 0.97; median 0.99).

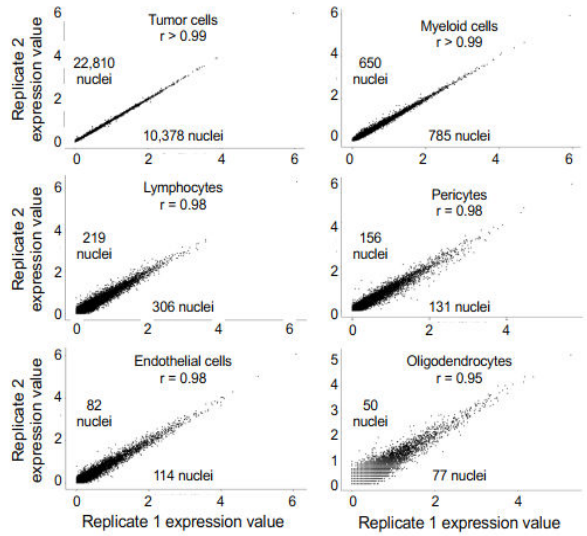

Figure 6. 124_PreVax replicate comparison.

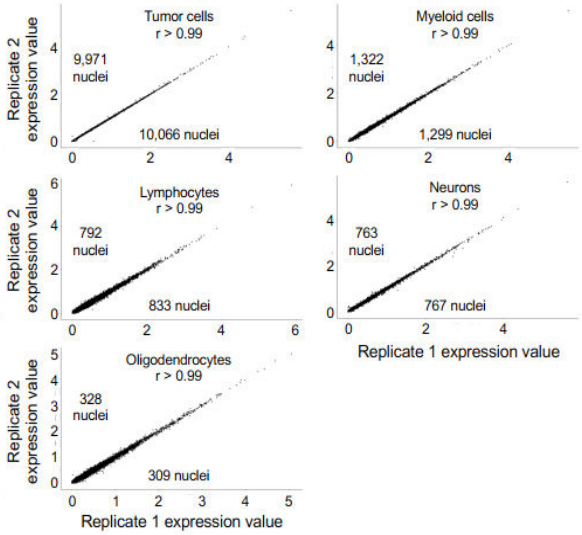

Figure 7. 129_PostVax replicate comparison.

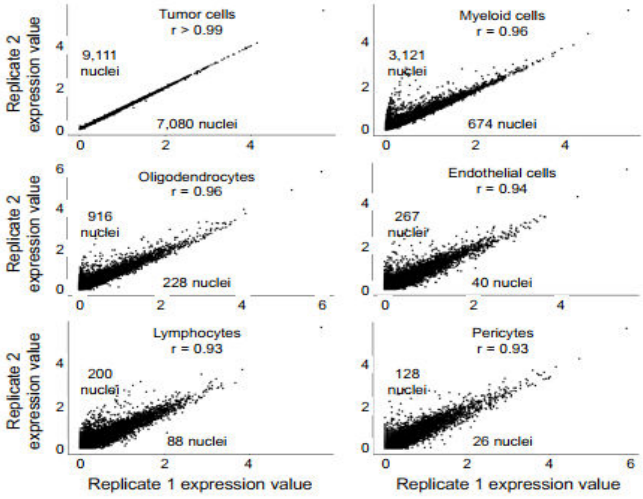

Figure 8. 133_PreVax replicate comparison.

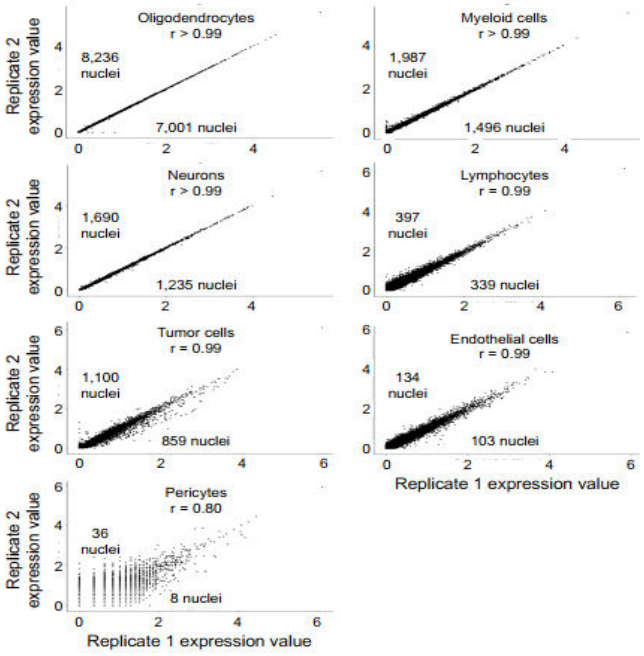

Figure 9. 134_PostVax replicate comparison.

Comparison to Published Glioblastoma Cell Type Results.

Neftel *et al*.[6] leveraged scRNAseq, bulk genetic and expression analysis, functional assays, and single-cell lineage tracing to derive the map of glioblastoma cellular states shown in **Figure 10A**. Malignant cells in glioblastoma that are not actively cycling exist in four cellular states: neural-progenitor-like (NPC-like), oligodendrocyte-progenitor-like (OPC-like), astrocyte-like (AC-like), and mesenchymal-like (MES-like). The plasticity between these states generates heterogeneity such that glioblastoma samples have different relative frequencies of cells in each state. The authors developed a single-cell gene signature score (SC) for each of the four cellular states using scRNAseq results. These scores were combined as explained in the manuscript to generate the values plotted in **Figure 10A** for each single cell. Notably, since their original Neftel *et al*.[6] manuscript, the Tirosh lab has updated their methods of assigning tumor cell state using pre-established signature scores (https://github.com/jlaffy/scalop). The samples in this report were analyzed using this newest approach from the Tirosh group. The analysis involves a more sophisticated method of annotating tumor cells using copy number alterations (CNAs) for enhanced accuracy of distinguishing tumor from normal (identifying cells with a mean absolute CNA signal comparable to high-confidence tumor cells). **Figure 10B** shows that these scores can be applied to snRNAseq data to determine the cellular state distribution for sample 133_PreVax. These results indicate that this sample has a preponderance of MES-like cell states. The two replicates show very similar patterns. **Figure 10C** quantifies the extent of similarity between replicates. **Figure 11** shows the cellular state distributions for 124_PreVax, 129_PostVax, and 134_PostVax. The results confirm that different tumors have different mixes of the four glioblastoma cellular states. Again, the replicates for each sample have very similar patterns. It can be seen in the results for 134_PostVax that the new algorithm for identifying the tumor state significantly decreased the number of tumor nuclei compared to the analyses of **Figures 5** and **9**.

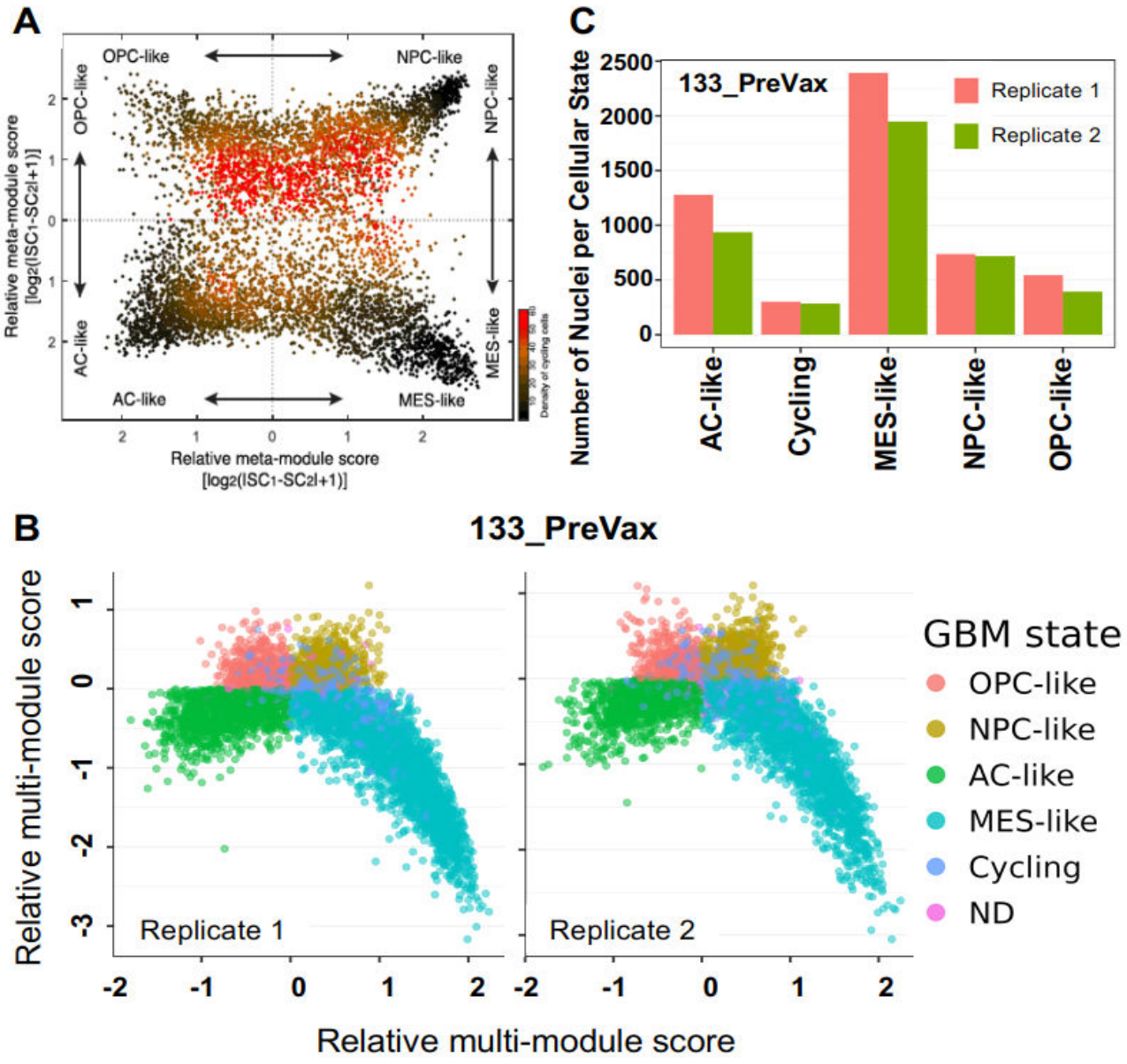

Figure 10. **Cellular state maps for glioblastoma**. **A**. This is Figure 3F in Neftel *et al*.[6] that plots the cellular state scores for single cells using on scRNAseq results. **B**. Cellular state scores based on snRNAseq results were calculated using an updated version of the algorithm used in **A** for sample 133_PreVax. **C**. Number of cells in each cellular state for sample 133_PreVax.

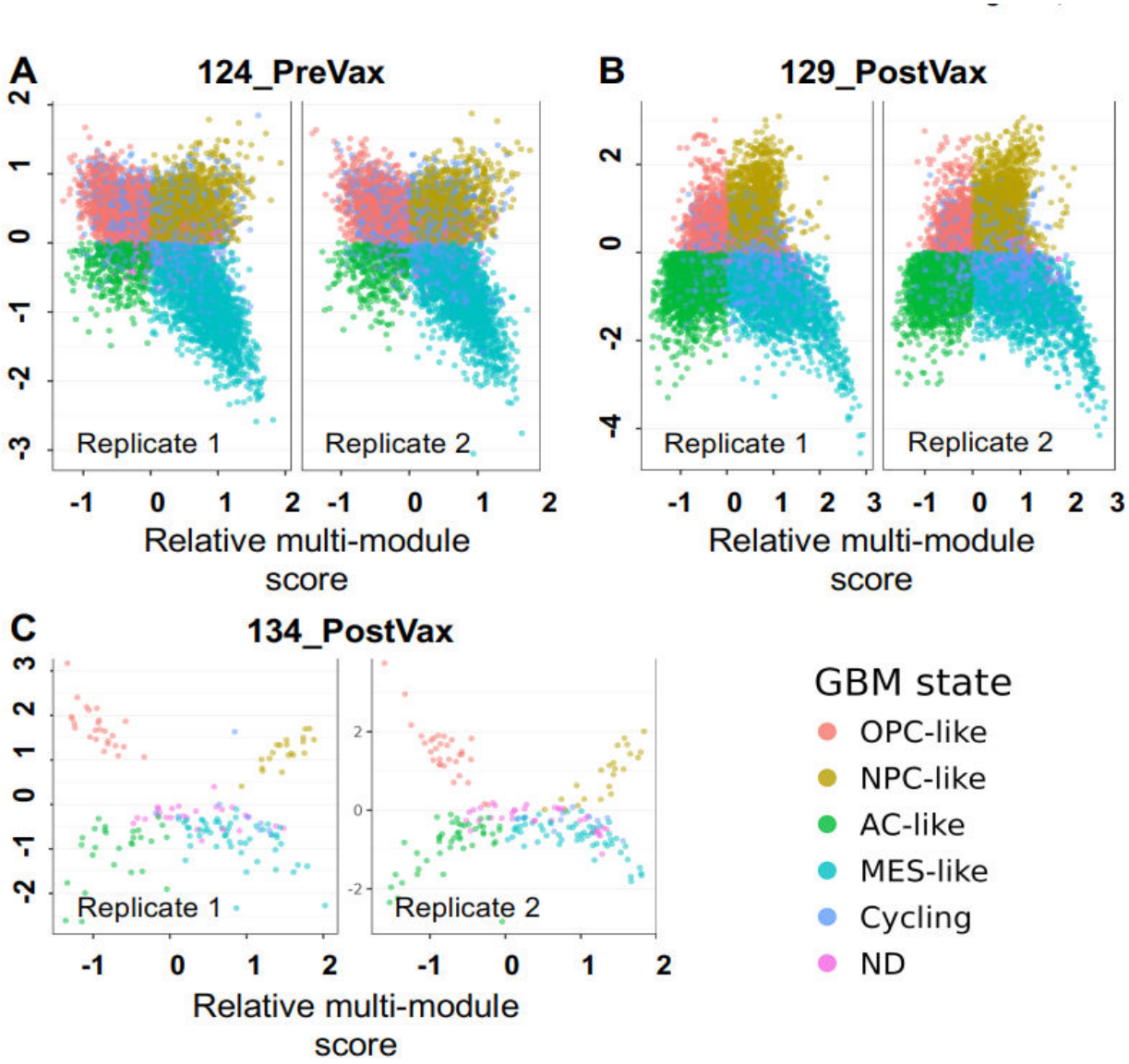

Figure 11. **Cellular state maps for three additional cases of glioblastoma**. **A**. Sample 124_PreVax. **B**. Sample 129_PostVax. **C**. Sample 134_PostVax.

References

1. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, Bodenmiller B, Campbell P, Carninci P, Clatworthy M, Clevers H, Deplancke B, Dunham I, Eberwine J, Eils R, Enard W, Farmer A, Fugger L, Göttgens B, Hacohen N, Haniffa M, Hemberg M, Kim S, Klenerman P, Kriegstein A, Lein E, Linnarsson S, Lundberg E, Lundeberg J, Majumder P, Marioni JC, Merad M, Mhlanga M, Nawijn M, Netea M, Nolan G, Pe'er D, Phillipakis A, Ponting CP, Quake S, Reik W, Rozenblatt-Rosen O, Sanes J, Satija R, Schumacher TN, Shalek A, Shapiro E, Sharma P, Shin JW, Stegle O, Stratton M, Stubbington MJT, Theis FJ, Uhlen M, van Oudenaarden A, Wagner A, Watt F, Weissman J, Wold B, Xavier R, Yosef N, Human Cell Atlas Meeting Participants. The human cell atlas. Elife. 2017 Dec 5;6. PMCID: PMC5762154
2. Slyper M, Porter CBM, Ashenberg O, Waldman J, Drokhlyansky E, Wakiro I, Smillie C, Smith-Rosario G, Wu J, Dionne D, Vigneau S, Jané-Valbuena J, Tickle TL, Napolitano S, Su M-J, Patel AG, Karlstrom A, Gritsch S, Nomura M, Waghray A, Gohil SH, Tsankov AM, Jerby-Arnon L, Cohen O, Klughammer J, Rosen Y, Gould J, Nguyen L, Hofree M, Tramontozzi PJ, Li B, Wu CJ, Izar B, Haq R, Hodi FS, Yoon CH, Hata AN, Baker SJ, Suvà ML, Bueno R, Stover EH, Clay MR, Dyer MA, Collins NB, Matulonis UA, Wagle N, Johnson BE, Rotem A, Rozenblatt-Rosen O, Regev A. A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. Nat Med. 2020 May 11;26(5):792–802. PMCID: PMC7220853
3. Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M, Choudhury SR, Aguet F, Gelfand E, Ardlie K, Weitz DA, Rozenblatt-Rosen O, Zhang F, Regev A. Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat Methods. 2017 Oct;14(10):955–958. PMCID: PMC5623139
4. Bakken TE, Hodge RD, Miller JA, Yao Z, Nguyen TN, Aevermann B, Barkan E, Bertagnolli D, Casper T, Dee N, Garren E, Goldy J, Graybuck LT, Kroll M, Lasken RS, Lathia K, Parry S, Rimorin C, Scheuermann RH, Schork NJ, Shehata SI, Tieu M, Phillips JW, Bernard A, Smith KA, Zeng H, Lein ES, Tasic B. Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. PLoS One. 2018 Dec 26;13(12):e0209648. PMCID: PMC6306246
5. Liang Q, Dharmat R, Owen L, Shakoor A, Li Y, Kim S, Vitale A, Kim I, Morgan D, Liang S, Wu N, Chen K, DeAngelis MM, Chen R. Single-nuclei RNA-seq on human retinal tissue provides improved transcriptome profiling. Nat Commun. 2019 Dec 17;10(1):5743. PMCID: PMC6917696
6. Neftel C, Laffy J, Filbin MG, Hara T, Shore ME, Rahme GJ, Richman AR, Silverbush D, Shaw ML, Hebert CM, Dewitt J, Gritsch S, Perez EM, Gonzalez Castro LN, Lan X, Druck N, Rodman C, Dionne D, Kaplan A, Bertalan MS, Small J, Pelton K, Becker S, Bonal D, Nguyen Q-D, Servis RL, Fung JM, Mylvaganam R, Mayr L, Gojo J, Haberler C, Geyeregger R, Czech T, Slavc I, Nahed BV, Curry WT, Carter BS, Wakimoto H, Brastianos PK, Batchelor TT, Stemmer-Rachamimov A, Martinez-Lage M, Frosch MP, Stamenkovic I, Riggi N, Rheinbay E, Monje M, Rozenblatt-Rosen O, Cahill DP, Patel AP, Hunter T, Verma IM, Ligon KL, Louis DN, Regev A, Bernstein BE, Tirosh I, Suvà ML. An integrative model of cellular states, plasticity, and genetics for glioblastoma. Cell. 2019 Aug 8;178(4):835–849.e21. PMCID: PMC6703186
7. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan 1;29(1):15–21. PMCID: PMC3530905
8. Fleming SJ, Chaffin MD, Arduini A, Akkad A-D, Banks E, Marioni JC, Philippakis AA, Ellinor PT, Babadi M. Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. Nat Methods. 2023 Sep;20(9):1323–1335. PMID: 37550580
9. Subramanian A, Alperovich M, Yang Y, Li B. Biology-inspired data-driven quality control for scientific discovery in single-cell transcriptomics. Genome Biol. 2022 Dec 27;23(1):267. PMCID: PMC9793662
10. Waag R, Bohacek J. Single-Nucleus RNA-Sequencing in Brain Tissue. Curr Protoc. 2023 Nov;3(11):e919. PMID: 37987152
11. Dar RD, Razooky BS, Singh A, Trimeloni TV, McCollum JM, Cox CD, Simpson ML, Weinberger LS. Transcriptional burst frequency and burst size are equally modulated across the human genome. Proc Natl Acad Sci USA. 2012 Oct 23;109(43):17454–17459. PMCID: PMC3491463