APPLICATION NOTE

G4™ Single Cell RNA-Sequencing Performance with the 10x Genomics® Chromium Single Cell 3’ Gene Expression Kit

- Rapid and cost-efficient sequencing of single cell RNA sequencing (scRNA-seq) libraries
- Seamless integration of the G4 Sequencing Platform into the 10x Genomics Chromium® Single Cell 3’ Gene Expression library preparation workflow and bioinformatics pipeline
- Highly accurate scRNA-Seq data using the G4 and 10x Genomics Chromium Single Cell 3’ Gene Expression assay

Introduction

The 10x Genomics Chromium single cell platform has revolutionized basic and translational research in immunology, developmental biology, and cancer by enabling the resolution of distinct cell populations within heterogeneous samples.[1] Single cell resolution of biological samples has accelerated our understanding of the complexity of living organisms and has opened up new possibilities for research in fields such as cancer research, genomics, and evolutionary biology.

Similarly, next-generation sequencing (NGS) is an enabling technology for single cell analysis, providing genomic information about the cell populations. Advancements in NGS over the last decade have contributed to faster, more accurate results while driving down the cost of experiments. The G4 Sequencing Platform combined with the 10x Genomics single cell analysis platform allows researchers to achieve faster and more flexible sequencing while profiling tens of thousands of cells.

In this application note, the G4 Sequencing Platform demonstrates excellent accuracy and technical reproducibility for single cell analysis via the 10x Chromium Next GEM Single Cell 3’ Reagent Kits v3.1 (Dual Index) **(Figure 1)**. Furthermore, Singular Genomics is recognized as a member of the Compatible Partner Program (CPP) for 10x Genomics.

G4 Specifications for scRNA-Seq

The G4 Sequencing Platform is a highly versatile benchtop sequencer that is well suited for demanding scRNA-Seq applications. The G4 Platform leverages a novel, 4-color Rapid sequencing by synthesis (SBS) chemistry to deliver highly accurate reads (single or paired-read format with optional index reads) with a 11-14 hour turnaround for 10x Genomics Chromium Single Cell 3’ Gene Expression experiments (28x91bp reads).

To maximize flexibility, the G4 Platform enables users to load up to four flow cells at a time, with each flow cell comprising four fluidically independent lanes, thereby enabling sample multiplexing without the need for index reads. The G4 Platform outputs FASTQ format files that integrate seamlessly with existing bioinformatics tools, including Cell Ranger. Users may elect to automatically demultiplex samples on-instrument via sample indices provided by the sample sheet or off-instrument using the Singular Genomics rapid demultiplexing tool.

More information about G4 specifications, such as run time, accuracy, and quality metrics, can be found on the Singular Genomics website.

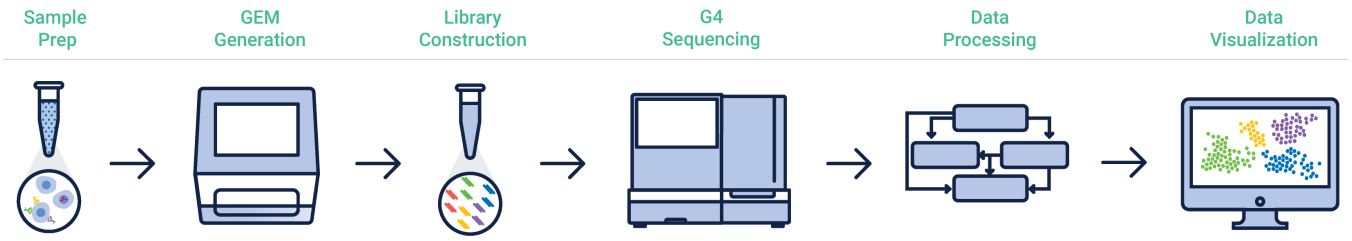


Figure 1 10x Genomics Chromium single cell workflow using the G4 Sequencing Platform.
Icons courtesy of 10x Genomics.

Methods

Library Preparation, Sequencing, and Analysis

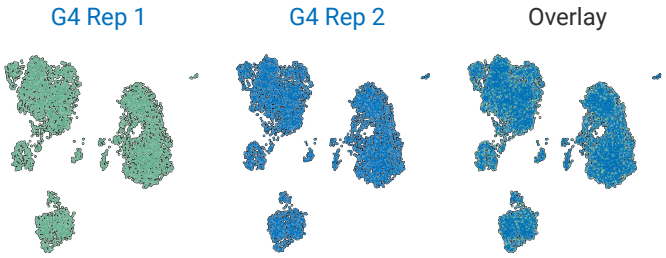
As part of the 10x Genomics CPP validation, 10x Genomics Chromium Single Cell 3' Gene Expression libraries were prepared from 7,000 fresh frozen healthy donor peripheral blood mononuclear cells (PBMCs) using the 10x Chromium Next GEM Single Cell 3' Gene Expression Protocol (User Guide CG000315) as indicated, except that after cDNA amplification and DNA clean-up via SparQ PureMag beads (Cat #95196), the resulting cDNA was split into two fractions. PCR amplification (14 cycles) of each fraction was performed using either Singular Genomics non-indexed PCR primers (2uM final concentration each) for sequencing via the G4 Sequencing Platform, or 10x Genomics Dual Index TT Set A (final concentration per manufacturer's recommendations) for sequencing via the Illumina® NextSeq 2000. Library quality was assessed via Qubit and Tapestation HSD5000 kit. The Singular Genomics library was sequenced in replicate on the G4 using two F2 flow cells, while the Illumina library was sequenced on the NextSeq 2000 using a P2 flow cell (28x91 cycle run format, no index reads).

Filtered, paired FASTQs were processed using Cell Ranger count (v6.0.0) and the resulting h5 file was processed using scanpy[2] (v1.8.2). First, doublets were removed using scanpy's wrapper of Scrublet[3] (v0.2.3) with an expected doublet rate of 7%. Next, cells were removed that fell below the 5th percentile of the number of UMIs per cell and the number of unique genes per cell in the dataset, or that had greater than 10% of their UMIs assigned to mitochondrially encoded genes. Next, genes were removed that were not detected in at least five cells. The top 2,000 highly variable genes (HVGs) were determined using scanpy's implementation of the Seurat v3 HVG detection method.[4] Finally, a latent representation of the datasets was generated using the scvi-tools[5,6] scVI model (v0.14.5; layers = 2, epochs = 400) on the top 2,000 HVGs using raw count data. Nearest-neighbor graphs, UMAPs, and Leiden clusters were generated using scanpy's modules on the scVI latent representation. Automated cell type identification was performed using CellTypist[7] (v1.2.0) and the "Immune_All_High" model with majority voting applied. Differential gene expression analysis was performed using scanpy's rank genes module with the Wilcoxon method on the library size-normalized, log-transformed counts.

Results

Replicate sequencing on two F2 flow cells yielded approximately 371M paired reads (180M and 191M paired reads respectively, replicate 1 and 2). Cell Ranger quality metrics produced for each replicate were highly concordant with respect to the number of identified cells, reads per cell, number of genes identified per cell, and read distribution across genic and intergenic regions (web summary available for download on the Singular Genomics website). Uniform manifold approximation and projection (UMAP) and Leiden clustering revealed a nearly identical grouping of single cell transcriptomes across technical replicates **(Figure 2A)**. Consistent with high technical reproducibility, we observed a strong correlation of pseudo-bulk profiles across the replicates (R^2 = 0.9949) **(Figure 2B)**.

(2A)



(2B)

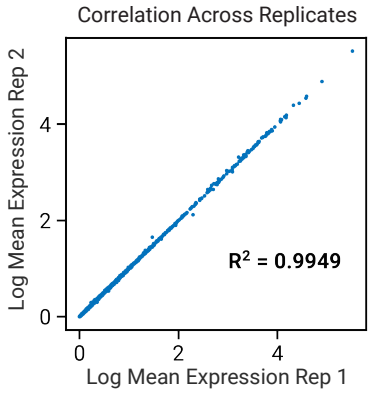
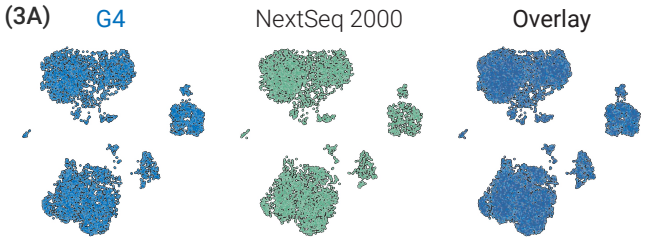


Figure 2 (A) UMAP embedding of single cell gene expression profiles obtained from technical replicate sequencing of a PBMC library. **(B)** Spearman's correlation of average gene expression across technical replicates.



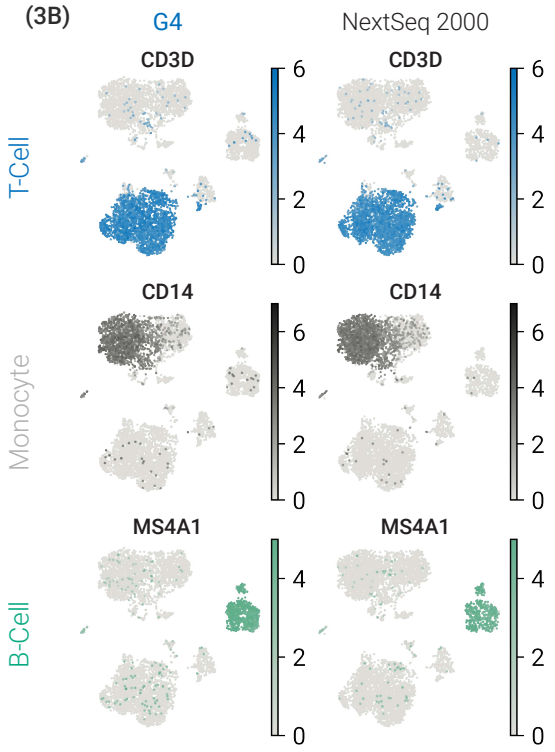
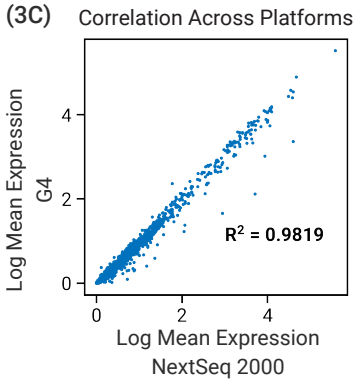
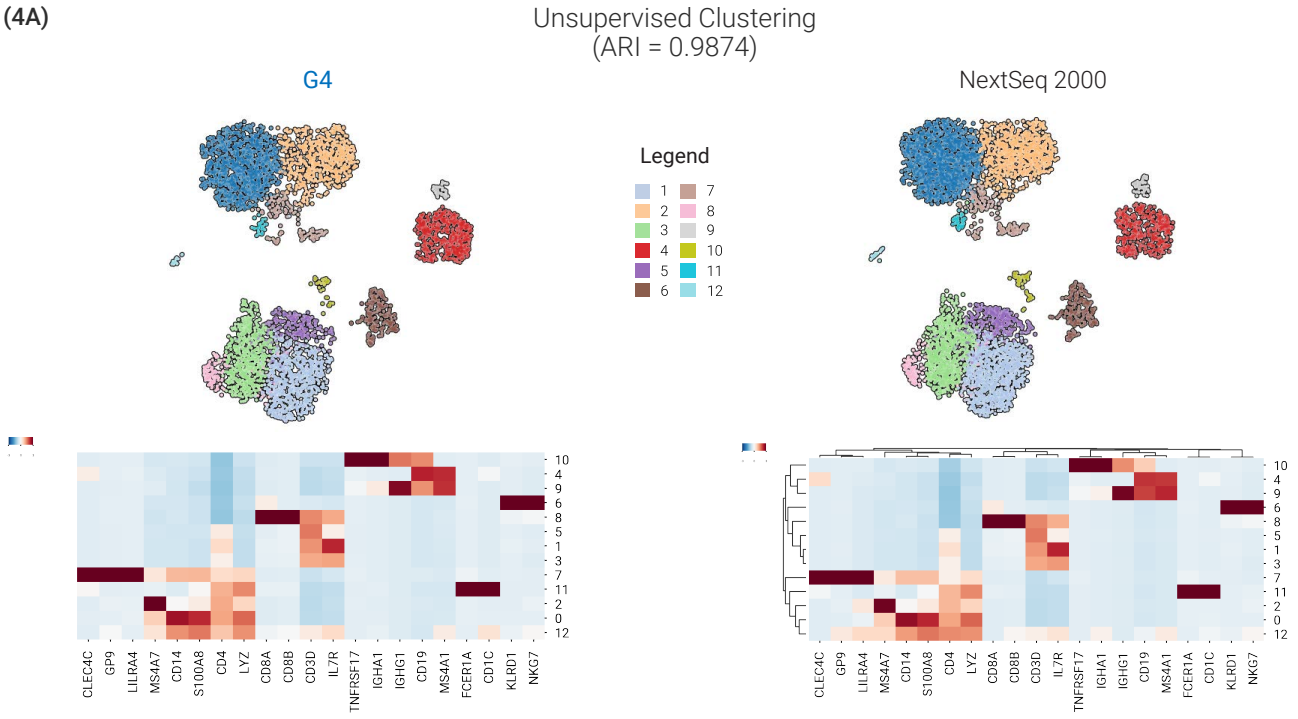


Figure 3 (A) UMAP embedding of single cell gene expression profiles obtained from sequencing of a PBMC library on the G4 and NextSeq 2000 platforms. **(B)** Expression of lineage-defining markers for T-Cells, Monocytes, and B-Cells in G4 and NextSeq 2000 datasets overlaid on the UMAP embedding. **(C)** Spearman's correlation of average gene expression across platforms, calculated as the library-size normalized, log-transformed UMI counts per gene.



To compare data produced on the G4 and NextSeq 2000 sequencers, we performed UMAP dimensional reduction with the combined datasets. We observed nearly identical embeddings across platforms **(Figure 3A)**, with comparable representation of T-Cells, Monocytes, and B-Cells, as inferred by marker gene expression **(Figure 3B)** and strong correlation of pseudo-bulk profiles between the two platforms **(Figure 3C)**. To explore the datasets in greater depth, we used CellTypist to perform unsupervised clustering and automated cell type annotation on the combined dataset. The adjusted rand index (ARI), a measure of similarity between sets of labels, demonstrated nearly identical cell type and unsupervised cluster labels between the G4 and NextSeq 2000 datasets with comparable lineage marker expression profiles **(Figure 4A, 4B)**. The comparability across platforms was further underscored by the strong correlation of pseudo-bulk profiles and differential gene expression analysis results for the major cell types identified **(Figure 4C, 4D)**.



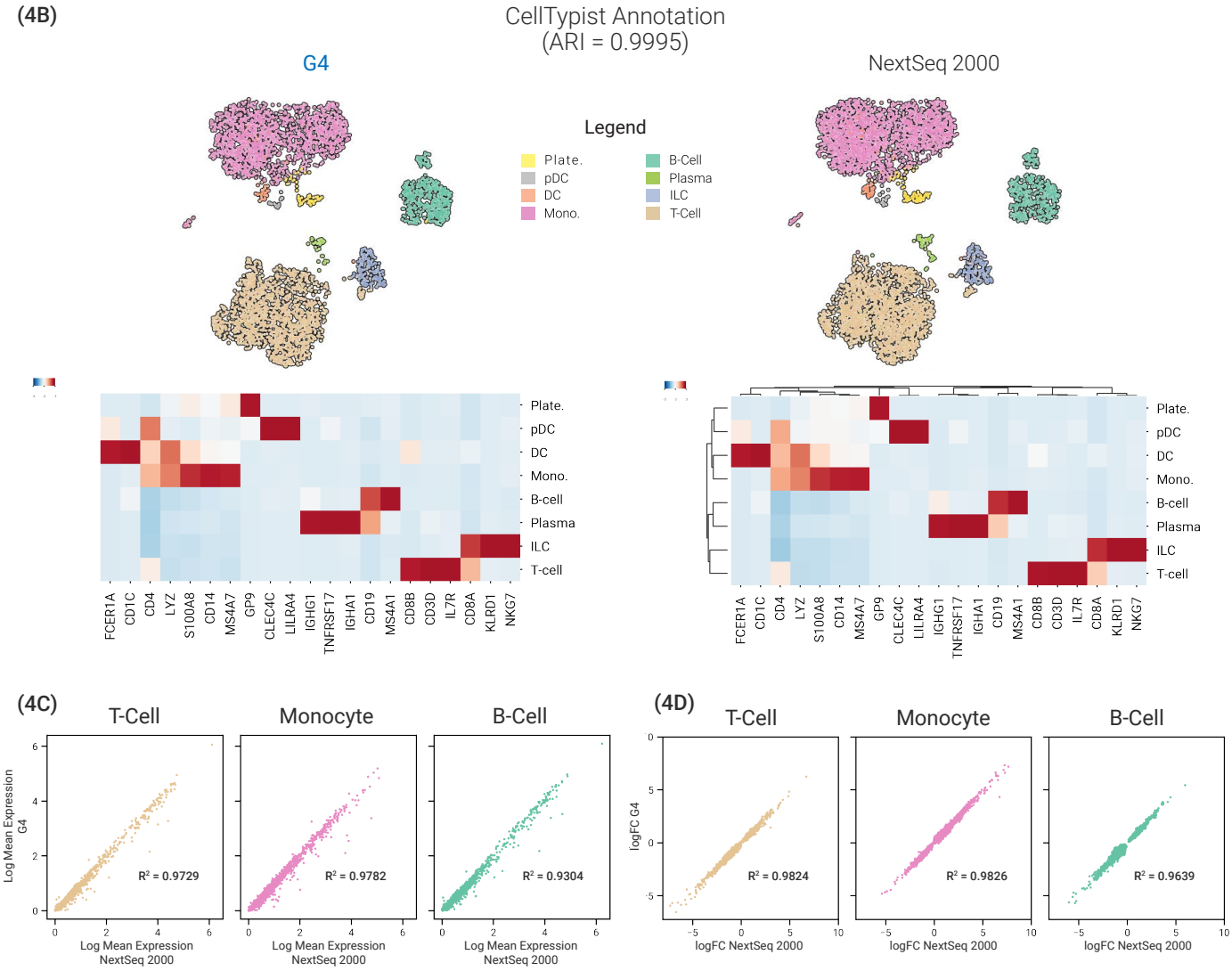


Figure 4 (A) Unsupervised Leiden clustering and **(B)** CellTypist annotations for NextSeq 2000 and G4 datasets overlaid on the UMAP embedding (top) with average expression profiles for a panel of well-known PBMC phenotyping markers for each identified cluster/cell type (bottom). **(C)** Spearman's correlation of average gene expression across platforms for each cell type, calculated as the library-size normalized, log-transformed UMI counts per gene. **(D)** Pearson's correlation across platforms for log fold changes (logFC) obtained from differential gene expression analysis of each cell type versus all other cells in the dataset. Cluster maps were generated by performing hierarchical clustering of the NextSeq 2000 dataset and then applying the same row/column ordering to the G4 dataset. Z-scoring was applied to each column of the cluster map and values were clipped to the range −3 to +3.

Conclusion

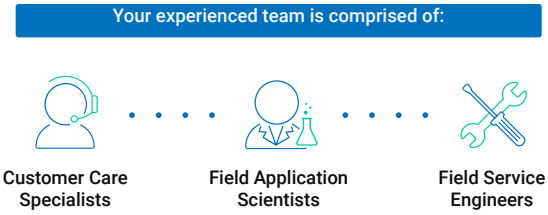
scRNA-Seq data generated using the 10x Chromium Single Cell 3' Gene Expression assay and the G4 Sequencing Platform demonstrates high technical reproducibility and performance comparable to data derived from the Illumina® NextSeq 2000 platform. Notably, G4 and NextSeq 2000 datasets uncovered nearly identical cell types and unsupervised clusters with comparable gene expression profiles consistent with major PBMC cell types. Beyond the results presented here, an equivalently high cross-platform correlation was observed for libraries prepared from 1,000 and 10,000 PBMCs (all datasets available for download at on the Singular Genomics website), indicating that the G4 Platform delivers robust, accurate results over a range of library sizes. In light of these strong results, 10x Genomics has recognized Singular Genomics as a validated member of the 10x Genomics Compatible Partner Program.

Overall, the G4 can be seamlessly integrated into labs running scRNA-Seq workflows. These labs can benefit from unique flow cell flexibility to match sequencing throughput more precisely to the sample set on hand. This, in combination with shorter sequencing cycle times, enables less waste, reduced turnaround times, and controlled costs for labs incorporating the G4 into their 10x Genomics scRNA-Seq operations.

SINGULAR GENOMICS

Get in Touch with the Customer Care Team

The purchase of a G4 comes with the assistance of a world-class experienced team to help you every step of the way. Our customer care team will assist you with order placement and can address any questions you may have. Our field service engineers (FSE) ensure a successful installation and provide instrument support and our field application scientists (FAS) conduct training and validation of your desired application. Our team is committed to support you when you need us.



Begin Your Journey with G4

[Contact our sales team](#) to learn more about the capabilites of the G4 Sequencing Platform

Website: www.singulargenomics.com
Email: care@singulargenomics.com
Call: +1 442-SG-CARES (442-742-2737)
Address: 3010 Science Park Rd, San Diego, CA 92121

REFERENCES

1. Haque, A., Engel, J., Teichmann, S. A. & Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med. 9: 1–12 (2017).
2. Wolf, F.A., Angerer, P., and Theis, F.J. SCANPY: large-scale single-cell gene expression data analysis. Genome Biology 19: 15 (2018).
3. Wolock, S.L., Lopez, R., and Klein, A.M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Systems 8: 281-291 (2019).
4. Stuart, T., Butler, A., et al. Comprehensive Integration of Single-Cell Data. Cell 177(7): 1888-1902 (2019).
5. Lopez, R., Regier, J., Cole, M.B., Jordan, M.I., and Yosef, N. Deep generative modeling for single-cell transcriptomics. Nature Methods 15(12): 10533-1058 (2018).
6. Gayoso, A., Lopez, R., et al. A Python library for probabilistic analysis of single-cell omics data. Nature Biotechnology 40: 163-166 (2022).
7. Conde, C.M., Xu, C., et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. Science 376: 6594 (2022).