SINGULAR GENOMICS

Single Cell RNA Sequencing with the G4™ Sequencing Platform

Kenneth Gouin III, Anne Decker, Sabrina Shore, Ryan Shultzaberger, Martin M Fabani, Timothy Looney, Eli N Glezer
Singular Genomics Systems, Inc., San Diego, California

Abstract

Single cell RNA sequencing (scRNA-Seq) has revolutionized basic and translational research by enabling the resolution of distinct cell populations within heterogeneous samples[1]. The G4™ Sequencing Platform provides plug-and-play performance for scRNA-seq and provides users with added flexibility to tailor run sizes and flow cell configurations to the sample set, rather than accumulating samples to massively pool onto large flow cells (Table 1). Less waste, reduced turnaround times, and controlled costs can be realized by incorporating the G4 into your scRNA-Seq operations.

Table 1. G4 Single Cell RNA-Sequencing Specifications (100 cycles). [a]Single cell RNA-Seq assumptions are based on 150M reads per sample. [b]Assumes 1-5 G4 sequencing runs per week.

Flow Cell Type	F2	F3
Throughput (M Reads)	150-165M per FC 600-660 per run	300-330M per FC 1,200-1,320M per run
Run Time (Hours)	12-15	12-15
Quality	75-90% ≥ Q30	
Accuracy	99.6-99.9%	
Samples / Flow Cell	1	2
Samples / Run[a]	4	8
Samples / Week[b]	1-20	2-40

Methods

Fresh, frozen, healthy donor PBMCs were processed using the Chromium Next GEM Single Cell 3' Protocol (Cat #1000128) and the resulting cDNA was split in two fractions to be sequenced on either the NextSeq 2000 or G4 platforms. Samples were then pre-processed using Cell Ranger (v6.0.0) and downstream analysis was performed using scanpy[2] (v1.8.2) and scvi-tools[3-4].

Results

Replicate sequencing on two G4 F2 flow cells yielded comparable quality control metrics that passed 10x Genomics' specifications for the assay, and sufficient depth was achieved with each individual flow cell (Table 2). Downstream UMAP and Leiden clustering demonstrated consistent results across the replicates with highly concordant (R^2 = 0.99) pseudo-bulk gene expression profiles (Figure 1).

Table 2. (To the Right) G4 Single Cell RNA-Sequencing Metrics.

Single Cell RNA-Seq Metrics	G4 Rep 1	G4 Rep 2
Read Configuration	R1: 28 bp R2: 91 bp	R1: 28 bp R2: 91 bp
Paired-Reads (M)	180M	191M
Number of Cells	9,012	9,025
Number of Reads / Cell	19,987	21,190
Number of Genes / Cell	1,303	1,323
Total Genes Detected	23,578	23,673
Fraction Reads in Cells	91.10%	91.10%
Valid Barcodes	98.20%	98.20%
Reads Mapped to Exon (%)	57.40%	57.00%
Reads Mapped to Transcriptome (%)	54.50%	54.10%

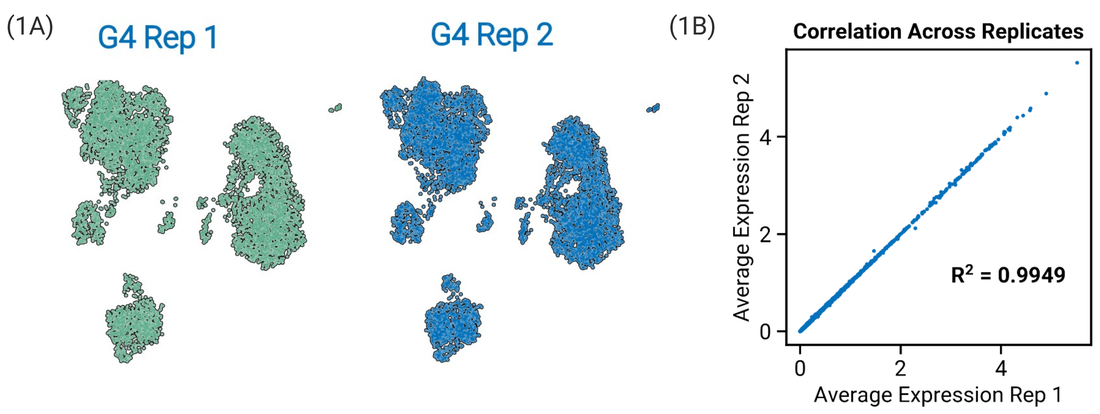

Figure 1. (A) UMAP embedding of single cell gene expression profiles obtained from technical replicate sequencing of a PBMC library. (B) Spearman's correlation of average gene expression across technical replicates.

Results (Continued)

Comparison of data produced on the G4 Sequencing Platform to data produced on the NextSeq 2000 demonstrated nearly identical embeddings across platforms (Figure 2A) and strong correlation of pseudo-bulk profiles between the two platforms (Figure 2B). Additionally, automated cell type identification with CellTypist[5] demonstrated nearly identical cell type labels between the G4 and NextSeq datasets (ARI = 0.99) with comparable lineage marker expression profiles (Figure 3A). The comparability across platforms was further underscored by strong correlation of differential gene expression analysis results for the major cell types identified (Figure 3B).

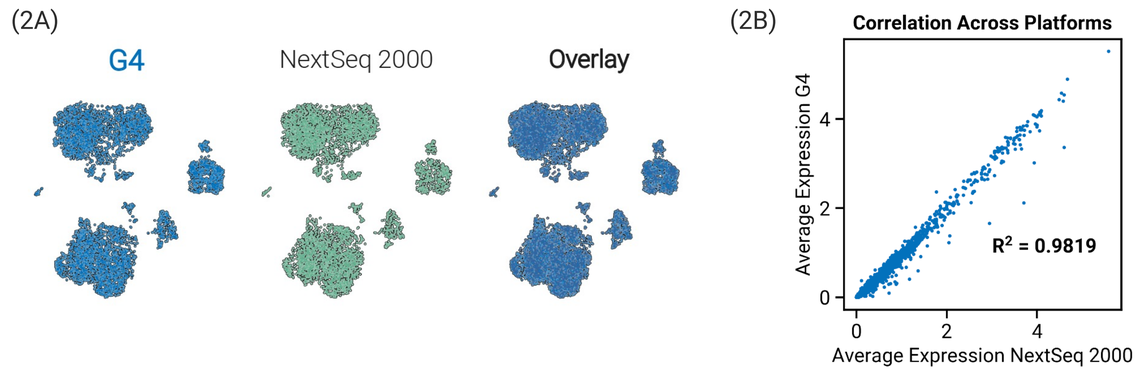

Figure 2. (A) UMAP embedding of single cell gene expression profiles obtained from sequencing of a PBMC library on the G4 and NextSeq platforms. (B) Spearman's correlation of average gene expression across platforms.

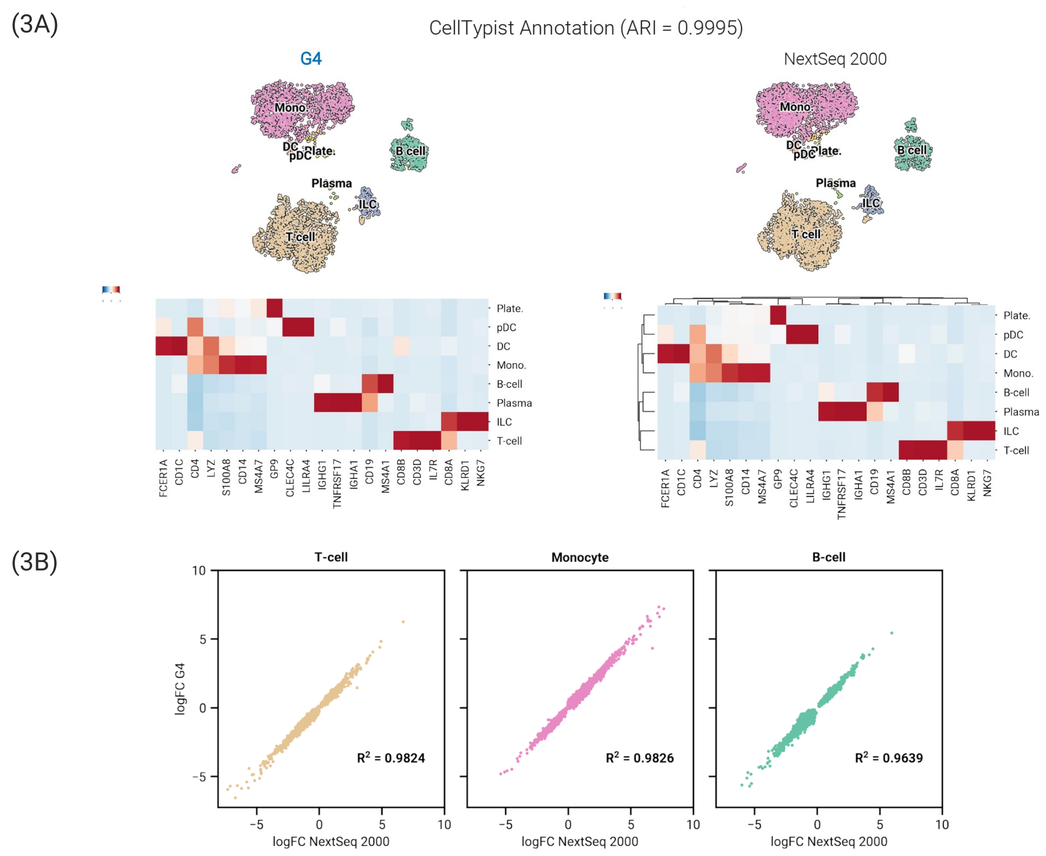

Figure 3. (A) CellTypist annotations for NextSeq and G4 datasets overlaid on the UMAP embedding (top) with average expression profiles for a panel of well-known PBMC phenotyping markers for each identified cluster/cell type (bottom). (B) Pearson's correlation across platforms for log fold changes (logFC) obtained from differential gene expression analysis of each cell type versus all other cells in the dataset.

Conclusion

scRNA-Seq data generated by the G4 demonstrates high technical reproducibility and performance comparable to the Illumina® NextSeq 2000 platform. Notably, G4 and Illumina datasets uncovered nearly identical cell types with comparable gene expression profiles consistent with major PBMC cell types.

The G4 Sequencing Platform is a plug-and-play solution for scRNA-Seq workflows that is compatible with existing laboratory ecosystems. The unique flow cell flexibility and unmatched run times of the G4 unit offer labs the ability to scale operations to match demand and reduce turnaround times on results.

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) 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).
4) Gayoso, A., Lopez, R., et al. A Python library for probabilistic analysis of single-cell omics data. Nature Biotechnology 40: 163-166 (2022).
5) Conde, C.M., Xu, C., et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. Science 376: 6594 (2022).