SINGULAR GENOMICS

APPLICATION NOTE

Single Cell RNA Sequencing on the G4™

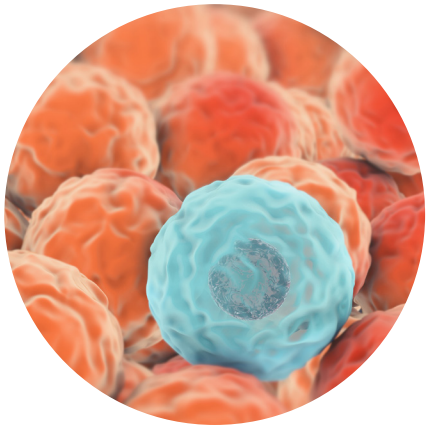
- Rapid SBS enables cost-efficient delivery of 1–8 single cell RNA sequencing samples in 12–15 hours.
- Seamless integration of the G4 into existing library prep workflows and bioinformatics pipelines.
- G4 delivers highly accurate scRNA-Seq data comparable to the leading high-throughput platform.

Introduction

Single cell RNA sequencing (scRNA-Seq) has revolutionized basic and translational research in immunology, developmental biology and cancer by enabling the resolution of distinct cell populations within heterogeneous samples[1]. A growing body of data highlight the potential for scRNA-Seq, and other single cell analysis technologies, to impact clinical decision making, for example by enabling the analysis of circulating tumor cells (CTCs) to determine cancer progression or predict patient response to targeted therapy as part of precision medicine[2].

scRNA-Seq protocols often begin with the enrichment of a cell population of interest (e.g. via Ficoll density gradient centrifugation or fluorescence-activated cell sorting (FACS)). The isolated population is then subjected to single cell encapsulation, RNA isolation and barcoding followed by next generation sequencing to assess the transcriptome of each barcoded cell. Finally, the single cell transcriptomes are compared to reference databases or clustered to determine cell identity and phenotype[3].

The G4 Sequencing Platform is a highly versatile benchtop sequencing platform that is well suited for demanding scRNA-Seq applications. The system leverages a novel, 4 color rapid SBS chemistry to deliver highly accurate reads with versatility. The G4 is compatible with existing upstream scRNA-Seq library preparation kits and outputs demultiplexed FASTQ files that integrate seamlessly with existing bioinformatic pipelines. Here we apply the G4 platform to characterize single cells derived from healthy donor peripheral blood mononuclear cells (PBMCs).



scRNA-Seq Sequencing Parameters

The G4 Sequencing Platform enables users to run 1, 2, 3, or 4 flow cells at a time. Mixing and matching two flow cell types across 4 positions enables users to start the system with 8 different run sizes. Each flow cell has 4 lanes, enabling up to 16 independent lanes per run, providing users flexibility in designing sequencing experiments to maximize turnaround time or throughput. Examples of G4 sequencing output, run time, accuracy, quality, and sample throughput are shown in **Table 1**. Run time, accuracy, and quality metrics reflect G4 specifications as detailed on the Singular Genomics website.

Flow Cell Type	F2	F3
Cycles	100	100
Throughput (M Reads)	150-165M per FC 600-660M per run	300-330M per FC 1,200-1,320M per Run
Run Time (Hours)	12-15	12-15
Quality	75-90% Bases ≥ Q30	
Accuracy	99.6-99.9%	
Samples / Flow Cell	1	2
Samples / Run[a]	4	8
Samples / Week[b]	1-20	2-40

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

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%

Table 2: G4 single cell RNA-Seq metrics.

Methods

Approximately 7,000 fresh frozen healthy donor PBMCs were processed using the Chromium Next GEM Single Cell 3' Protocol (Cat #1000128) followed by DNA clean-up via SparQ PureMag beads (Cat. #95196). PCR amplification was carried out using Singular Genomics non-indexed PCR primers (14 cycles, at 2uM final concentration each). Libraries were assessed via Qubit and Tapestation HSD5000 kit. A single library was sequenced in replicate using two F2 flow cells with a 28X91 cycle run format.

Filtered FASTQs were analyzed using Cellranger count (v6.0.0), then the resultant h5 file was processed using Scanpy (v1.8.2) to remove doublets (via Scublet), cells having a UMI count below the 5th percentile, and cells for which >10% of UMIs corresponded to mitochondrial genes. Next, genes that were not detected in at least 5 cells were removed, and the top 2,000 highly variable genes (HVGs) were determined using Scanpy's implementation of the Seurat v3 HVG detection method. Finally, a latent representation of the datasets was generated using the scvi-tools scVI model (v0.14.5; layers = 2, epochs = 400) on the top 2,000 HVGs (performed on the raw data). Nearest-neighbor graphs, UMAPs, and leiden clusters were generated using Scanpy's tools on the scVI latent representation.

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 (Table 2). Uniform manifold approximations and projection (UMAP) clustering revealed a nearly identical grouping of single cell transcriptomes across technical replicates, where each group represents a distinct putative cell type. Consistent with high technical reproducibility, we observe a strong correlation of average gene expression profiles across the replicates (R^2 = 0.9933).

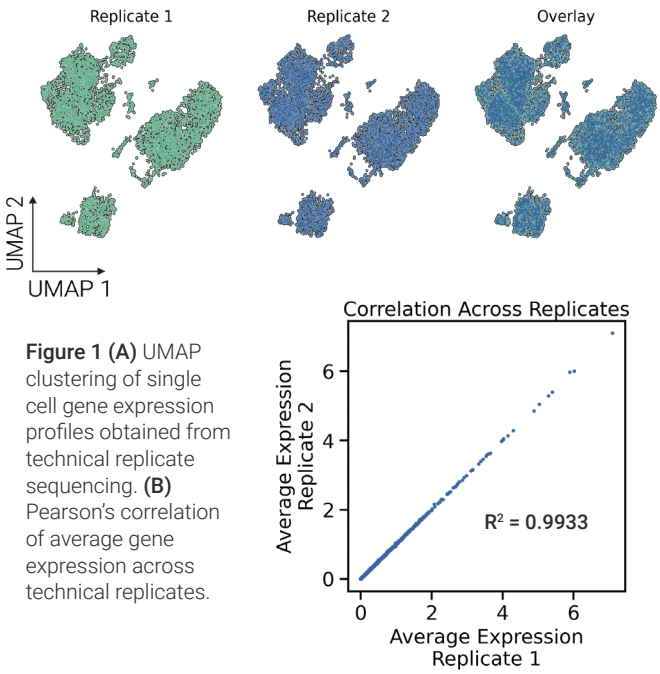


Figure 1 (A) UMAP clustering of single cell gene expression profiles obtained from technical replicate sequencing. **(B)** Pearson's correlation of average gene expression across technical replicates.

In order to compare data produced on the G4 to other common sequencing systems, we performed UMAP clustering of a public PBMC scRNA-Seq dataset derived from the NovaSeq 6000[4], then overlaid the UMAP profile with that obtained from the G4. We observe a highly similar cell grouping pattern across platforms (Figure 2A), with similar representation of T cells, monocytes, and B cells as inferred by marker gene expression. Further, unsupervised clustering of each dataset yielded 11 putative cell type groups having nearly identical expression profiles of key marker genes (Figure 3).

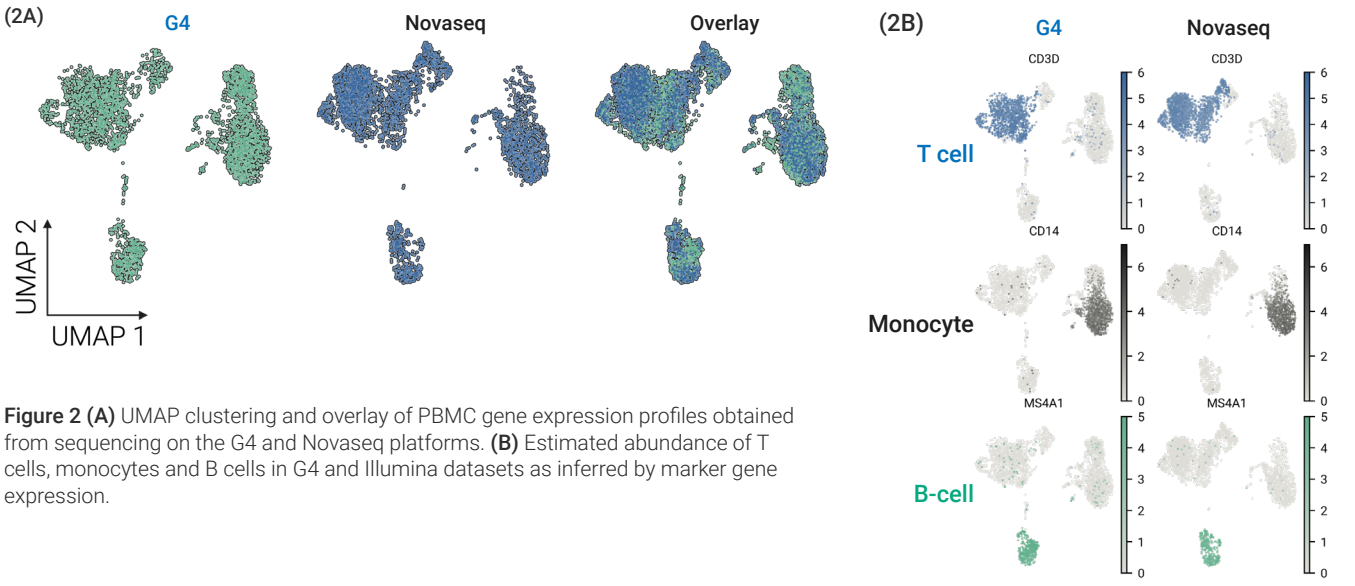


Figure 2 (A) UMAP clustering and overlay of PBMC gene expression profiles obtained from sequencing on the G4 and Novaseq platforms. **(B)** Estimated abundance of T cells, monocytes and B cells in G4 and Illumina datasets as inferred by marker gene expression.

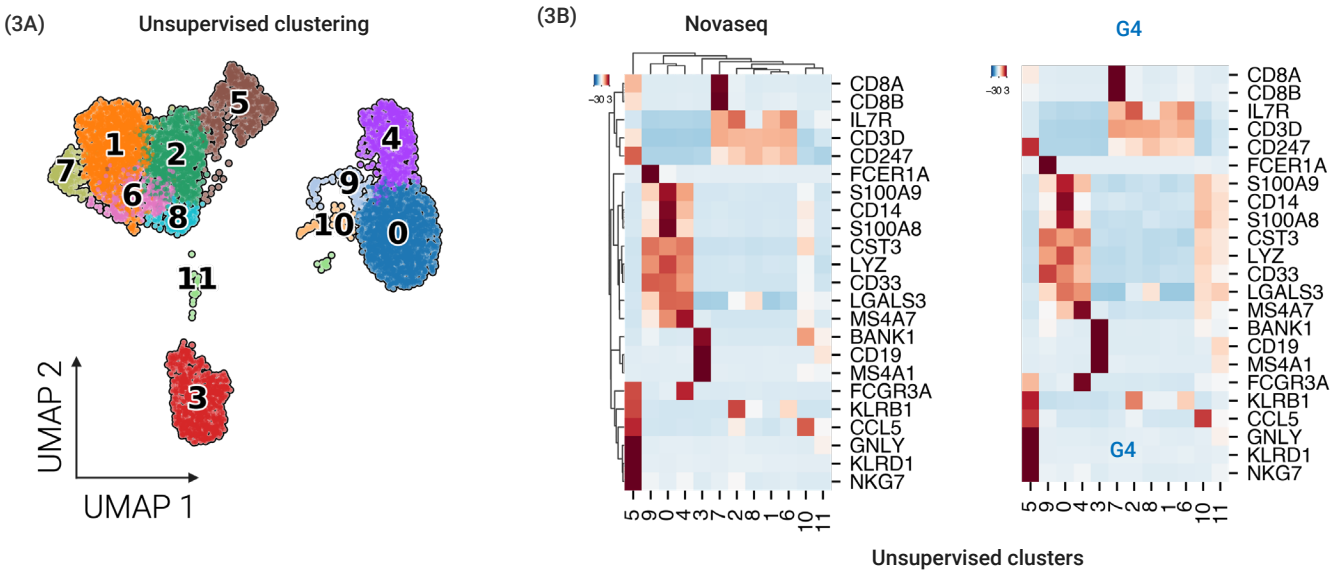


Figure 3 (A) Unsupervised clustering of G4 and NovaSeq datasets into 11 major groups. (B) Cluster gene expression profiles are nearly identical across platforms and are consistent with major PBMC cell types.

Conclusion

scRNA-Seq data generated by the G4 demonstrates high technical reproducibility and performance comparable to the Illumina® NovaSeq 6000 platform. Notably, unsupervised clustering of G4 and Illumina datasets yielded cluster gene expression profiles that were nearly identical across platforms and consistent with major PBMC cell types.

The G4 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 offer labs the ability to scale operations to match demand and reduce turnaround times on results.

scRNA sequencing with the G4 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. Less waste, reduced turnaround times and controlled costs can be realized by incorporating the G4 Sequencer into your scRNA-Seq operations.

*FASTQ files from this study are available by request for additional analysis.

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 how the G4 can transform your sequencing workflows.

Website: www.singulargenomics.com
Email: care@singulargenomics.com
Call: +1 442-SG-CARES (442-742-2737)
Address: 3010 Science Park Rd, San Diego, CA 92122

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. Sun, G. *et al*. Single-cell RNA sequencing in cancer: Applications, advances, and emerging challenges. *Mol. Ther. - Oncolytics* **21**, 183–206 (2021).
3. Conesa, A. *et al*. A survey of best practices for RNA-seq data analysis. *Genome Biology* vol. 17 1–19 (2016).
4. https://www.10xgenomics.com/resources/datasets/5-k-peripheral-blood-mononuclear-cells-pbm-cs-from-a-healthy-donor-v-3-chemistry-3-1-standard-3-0-2

Research Use Only. Not for use in diagnostic procedures.
© Singular Genomics Systems, Inc. All rights reserved. G4 and G4 Sequencing Platform are trademarks of Singular Genomics Systems, Inc.