SINGULAR GENOMICS

The Max Read™ Kit for High-Throughput Single Cell Sequencing on the G4™

Niek van Wietmarschen, Kenneth Gouin III, Bubba Brooks, Liz LaMarca, Ryan Shultzaberger, Timothy Looney, Martin M Fabani, Daan Witters, Eli N Glezer
Singular Genomics Systems, Inc., 3010 Science Park, Rd, San Diego, California

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. However, there remains a need for cost-effective, high throughput sequencing solutions to reduce the cost of scRNA-Seq studies. We previously introduced the Max Read™ Kit, which enables higher output of short reads for the G4™ Sequencing Platform, without a significant impact on read quality. Here, we evaluate the performance of the Max Read Kit for scRNA-seq by sequencing a 10x Genomics 3' RNA-Seq library prepared from human peripheral blood mononuclear cells (PBMC) using the G4, comparing results to those from the Illumina® NextSeq 2000. We demonstrate excellent accuracy, reproducibility and throughput with the Max Read™ sequencing kit.

Methods

Max Read™ Library Preparation and Sequencing

scRNA-Seq library preparation for the Max Read Kit begins with a conventionally prepared scRNA-Seq library. The library is split into three aliquots, each of which is PCR-converted to introduce a unique pair of sequencing primers. The adapted libraries are re-pooled, clustered at high density, then sequenced sequentially over three rounds, whereby each round utilizes one of the three sequencing primer pairs.

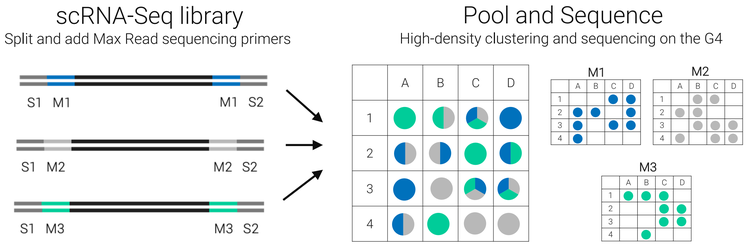


Figure 1 High-density clustering and clonal amplification of Max Read libraries on a patterned flow cell allows multiple template molecules to amplify within the same nanowell of the array (left grid), which are deciphered through sequential rounds of sequencing, thereby yielding multiple high-quality reads per nanowell and greater read throughput.

Single Cell Sequencing on the G4™

The G4 Platform supports three throughput configurations: F2, F3 and M3 (Max Read Kits). This flexibility enables efficient experimental design, rapid turnaround time and low cost per sample.

	F2 Flow Cell	F3 Flow Cell	Max Reads[1]
Reads Delivered / Flow Cell	~200 M[2]	~400 M[2]	~ 800 M
Reads Delivered / Run	~800 M[2]	~1,600 M[2]	~ 3,200 M
# Samples / Flow Cell[3]	1	2	4
# Samples / Run[3]	4	8	16
Price/M Read	~$3.0	~$1.8	$1.0
Price/Sample[3]	$600	$360	$200
Run Time	11-14 hours[4]	11-14 hours[4]	~24 hours[4]
Quality (%Q30)	80-90%	80-90%	80-90%

Numbers are projected and results may vary. Refer to kit specifications for more detail.

[1] Max Read kits specifications are projected, and kits are currently only compatible with 10x Genomics Chromium™ 3 and 5' Gene Expression assays and Visium™ Spatial Gene Expression. Kits allow for 1 sample per lane
[2] Throughputs listed are approximations and not guaranteed above kit specifications. Results may vary based on experimental design and sample type.
[3] Assumes 28x91 read configuration, 10,000 cells/sample and 20,000 reads/cell
[4] Run time includes clustering, sequencing and instrument wash for non-indexed reads

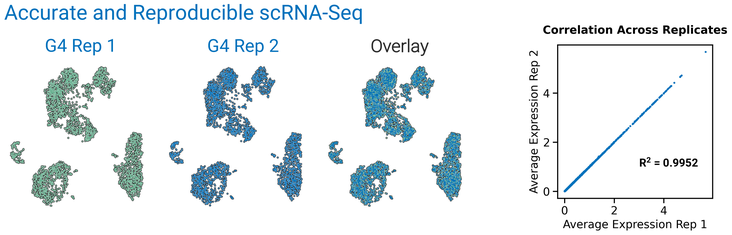
Methods

Library Preparation, Sequencing and Analysis

A 10x Genomics 3' RNA-Seq library was prepared from ~7000 total cells comprising human healthy donor PBMCs spiked with 10% Jurkat and 10% Ramos cells following standard 10x Genomics protocol, then split into two aliquots. One aliquot was converted to Max Read format for sequencing in replicate on the G4 Platform with the F3 flow cell, while the other aliquot was sequenced on the NextSeq 2000. 200M read pairs were used for downstream analysis following processing via Cell Ranger, scanpy and scVI tools.

Results

Accurate and Reproducible scRNA-Seq



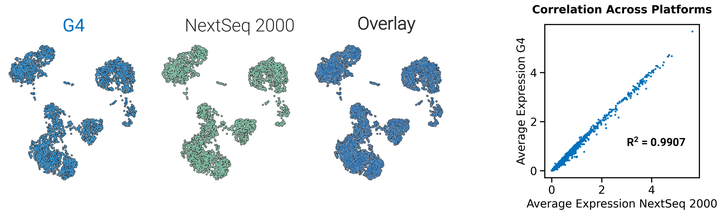


Figure 2 UMAP clustering and overlay of Max Read technical replicates (top) and Max Read vs NextSeq 2000 (bottom). Scatterplots indicate Pearson's correlation of average gene expression across conditions, calculated as the library size normalized, log transformed UMI counts per gene. The high technical and cross-platform correlation indicates that Max Read format data is comparable to conventional sequencing data.

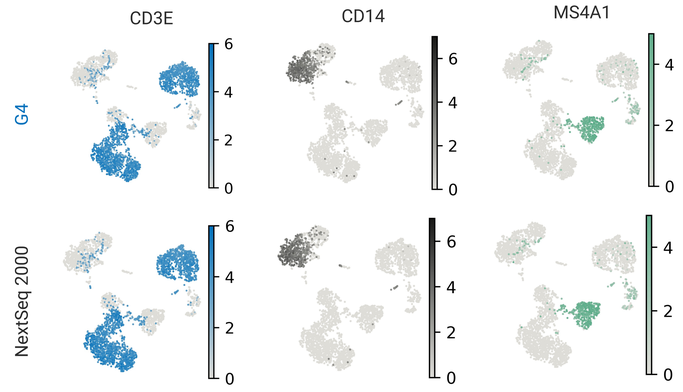


Figure 3 UMAP clustering and overlay of G4 Max Read and NextSeq2000, annotated by expression of T-Cell, Monocyte, and B-Cell markers (CD3E, CD14, and MS4A1, respectively). Expression patterns are nearly identical across platforms, yielding nearly identical estimated abundances of key immune cell types.

Results

Continued

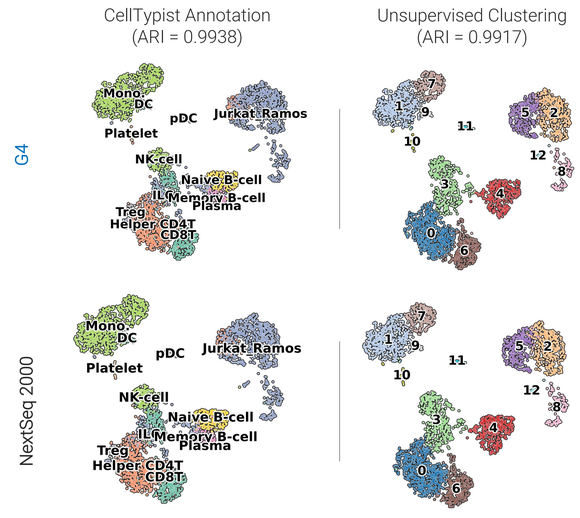


Figure 4 Celltypist and scVI leiden clustering labels for G4 Max Read and NextSeq 2000 datasets. The cell type labels are nearly identical across platforms, as indicated by the high Adjusted Rand Index (ARI). The estimated frequency of Jurkat and Ramos cells was 20% and 4% respectively, within each platform dataset.

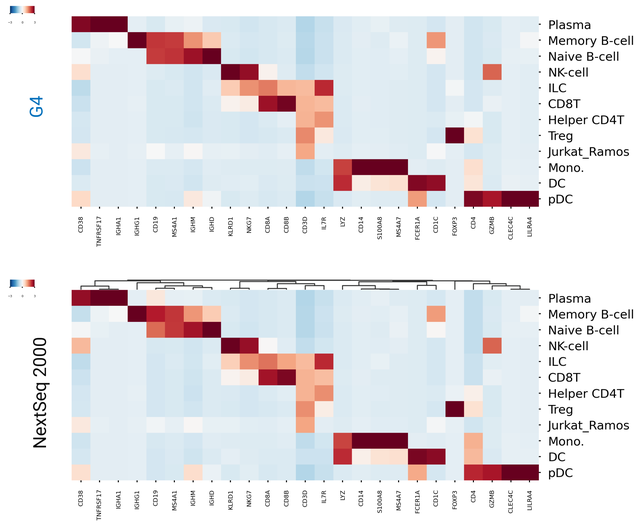


Figure 5 Heatmap indicating gene expression of key cell type-specific markers. Z-scoring was applied to each column of the cluster map and values were clipped to the range −3 to +3. Gene expression profiles are nearly identical across platforms.

Conclusion

In this study, we demonstrate excellent accuracy, reproducibility and throughput of the Max Read format when applied to scRNA-seq. Using Max Read Kits for single cell analysis (28x91bp paired read format) customers may expect ~800M reads per F3 flow cell, sufficient for one typical ~10,000 cell scRNA-seq library per each of the four flow cell lanes.

In summary, the Max Read format significantly increases the sequencing throughput for short read applications such as scRNA-Seq to 3.2 billion reads per run on the G4, enabling cost-effective sequencing without compromising data quality.