SINGULAR GENOMICS

Performance Assessment of the Novel G4™ Sequencing Platform for Cancer Research

Kenneth Gouin III, Niek van Wietmarschen, Bubba Brooks, Liz LaMarca, Ryan Shultzaberger, Mark Kalaj, Sabrina Shore, Daan Witters, Timothy Looney, Martin M Fabani
Singular Genomics Systems, Inc, 3010 Science Park Road, San Diego, California, 92121

Introduction

Next-generation sequencing (NGS) has achieved widespread adoption as a tool for cancer research. Despite this success, traditional NGS systems are limited by long analysis times, labor intensive protocols, and the need for extensive sample batching to achieve cost-effective use. Designed to address the current limitations, the G4™ Sequencing Platform delivers highly-accurate, rapid results with a single-day turnaround time. Here, we assess the performance of the G4 for whole exome sequencing, methylome profiling, and single cell RNA sequencing, comparing results to Illumina® platforms.

Methods

G4™ Sequencing Platform

G4 is an innovative benchtop sequencer designed to deliver rapid sequencing results with throughput flexibility to reduce batching-related delays. The G4 Sequencing Platform supports single or paired end reads of up to 150 bp, including the ability to include dual index reads for sample multiplexing. Users may analyze up to four flow cells of two types (F2: up to 250M reads, F3: up to 450M reads) in a single run. To facilitate multiplexing, each flow cell comprises four fluidically independent lanes. Further, the G4 Max Read™ format delivers up to 3.2B reads per run for single cell applications, for a cost of ~$1 per million reads. We benchmarked platform performance by performing whole exome sequencing (IDT exome) and methylome profiling (NEB EM-Seq) of NA12878 using the F2 flow cell. We further performed scRNA-seq analysis of human PBMCs (10x Genomics 3' Gene Expression Kit) using the G4 Max Read format.

Figure 1 G4 Sequencing Platform and Performance Specifications

Fast, Flexible Sequencing on the G4

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

	Reagent Configuration[1]	Run Time[2]	Reads / Flow Cell[3]	Reads / Run[3]	Quality[4]
F2 Flow Cell	100 cycles	~11 hours	Up to 250 M	Up to 1,000 M	80-90% Bases ≥ Q30
	200 cycles	~15 hours			
	300 cycles	~19 hours			
F3 Flow Cell	50 cycles	8 - 11 hours	Up to 450 M	Up to 1,800 M	
	100 cycles	11 - 14 hours			
	200 cycles	15 - 19 hours			
	300 cycles	19 - 24 hours			
Max Read[5]	28x91 Single Cell	~ 24 hours	800 M	3,200 M	
	28x50 Spatial FFPE				

[1] Reagents include 50 additional cycles above what is represented to account for adapters and indices.
[2] Run time includes clustering, sequencing and instrument wash for non-indexed reads.
[3] Paired reads passing filter for F2 and F3 are dependent on application and read length.
[4] Performance metrics may be impacted by application, sample quality, library preparation, loading concentration, and other sequencing considerations. Metrics as generated on reference bacterial and human genomes.
[5] 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.

Results

Exome Sequencing and Analysis of NA12878

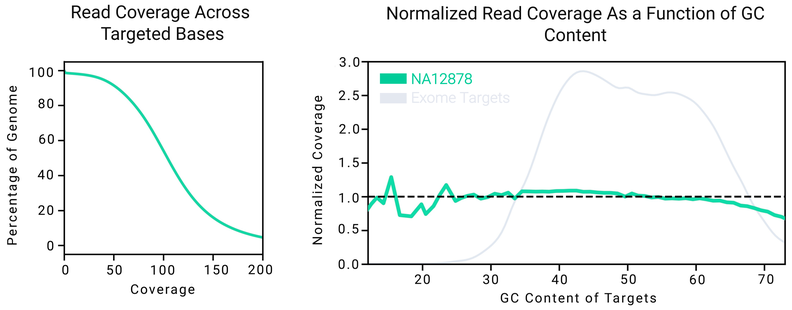

Figure 2 Read coverage across targeted exome regions and coverage as a function of GC content following 2x150bp sequencing of an exome library prepared from enzymatically sheared NA12878 gDNA to achieve ~110x coverage.

Table 1 Variant detection performance over high confidence regions of NA12878 was assessed via hap.py following 100x sequencing via the G4 F2 flow cell. Variants were detected using a custom DeepVariant v1.4 whole exome model, available for download on the Singular Genomics website.

Metric	Value 100x Coverage
SNP Precision	99.67%
SNP Recall	98.60%
SNP F1-Score	99.13%
Indel (<50bp) Precision	98.87%
Indel (<50bp) Recall	93.09%
Indel F1-Score	95.89%

EM-Seq Analysis of NA12878

An EM-Seq library was prepared from NA12878 and sequenced on the G4 to 191M 2x100bp reads depth, then analyzed via bwa-meth.

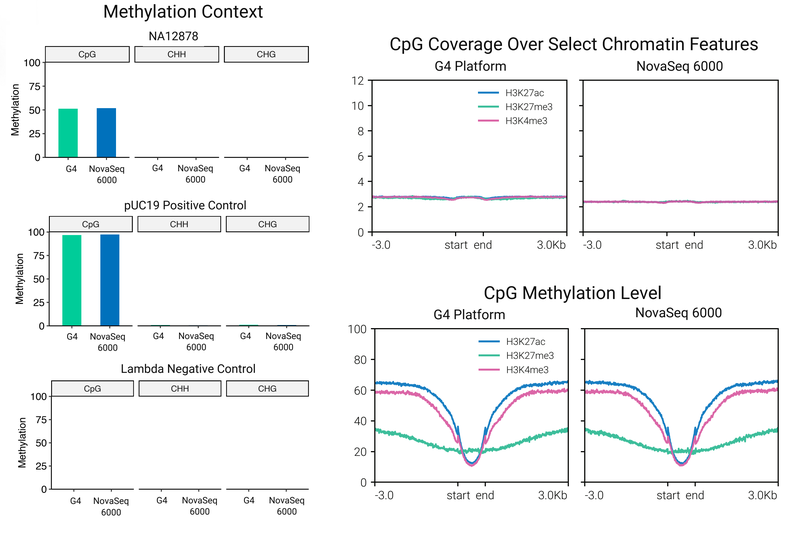

Figure 3. Methylation context, CpG coverage and methylation level over select chromatin features. For comparison, an equivalent number of NA12878 EM-Seq reads were analyzed from the NovaSeq 6000 (SRA database record SRR10532144).

Results (Continued)

Accurate and Reproducible scRNA-Seq

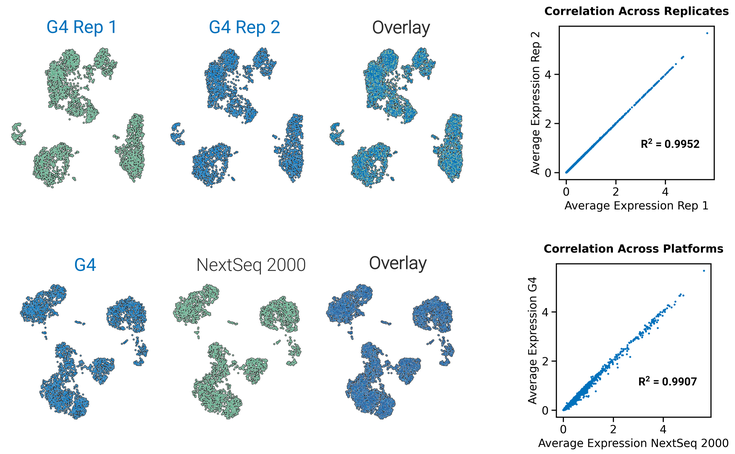

Figure 4 UMAP clustering and overlay of Max Read technical replicates (top) and Max Read vs NextSeq 2000 (bottom) for a human PBMC library spiked with 10% Jurkat and Ramos cells. Scatterplots indicate Pearson's correlation of average gene expression across conditions, calculated as the library size normalized, log transformed UMI counts per gene.

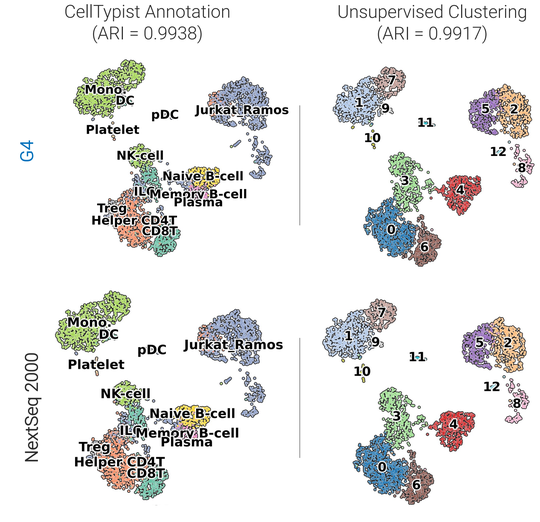

Figure 5 Celltypist and scVI leiden clustering labels for G4 Max Read and NextSeq 2000 datasets. The cell type labels are nearly identical across platforms, as indicated by the high Adjusted Rand Index (ARI). The estimated frequency of Jurkat and Ramos cells was 20% and 4% respectively, within each platform dataset.

Conclusion

The G4 Sequencing Platform delivers highly concordant results to the NextSeq 2000 and NovaSeq 6000 for human variant detection, methyl-seq, and scRNA-Seq. With high accuracy, combined with rapid turnaround times, scalable sequencing capacity, and the flexibility to run 4 flow cells and 16 lanes at a time, the G4 Sequencing Platform is well suited to have broad utility for cancer research applications.