SINGULAR GENOMICS

Performance Assessment of the G4™ Sequencing Platform, A Novel Platform for Rapid and Flexible Next-Generation Sequencing

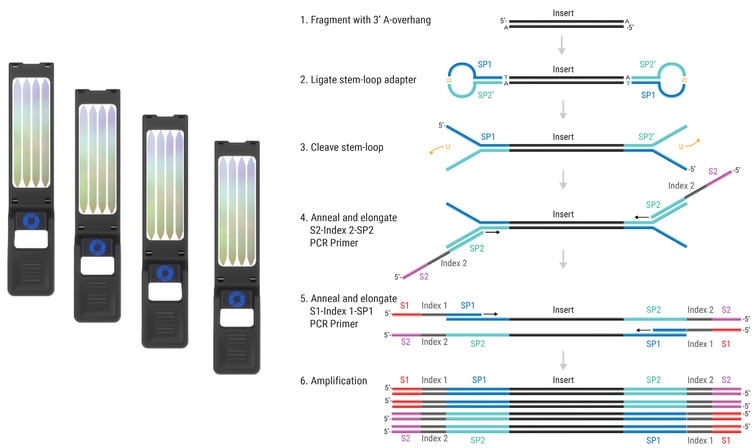
Kenneth Gouin III[1], Ann Tong[1], Sabrina Shore[1], Ryan Schultzaberger[1], Martin Fabani[1], Timothy Looney[1], Eli Glezer[1]
[1]Singular Genomics Systems, Inc., San Diego, California

Background

Next-generation sequencing (NGS) has become a foundational tool for both biological research and in-vitro diagnostics, particularly in oncology, immunology, and detection of genetic disorders. Despite its success, there is a need for new DNA sequencing platforms that combine high accuracy, speed, and flexible throughput to provide timely results and cost-effective operations for research and clinical applications. Here, we evaluate the performance of the novel Singular Genomics G4™ Sequencing Platform for rapid sequencing-by-synthesis (SBS). We demonstrate its utility for whole genome sequencing (WGS), bulk and single cell RNA-seq, and high fidelity demultiplexing via unique dual indices (UDIs).

Methods – Performance Characterization

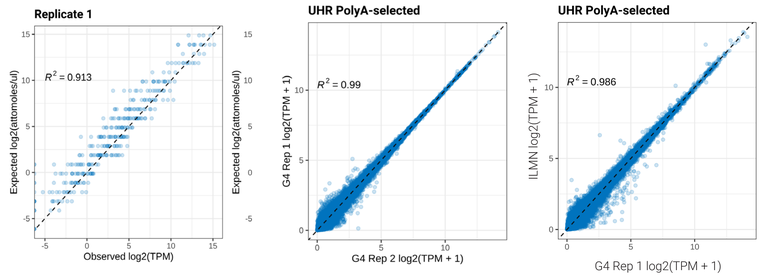
Above Left: Libraries evaluated in this performance study. Each library was sequenced using the G4™ and the NextSeq 2000, where applicable.
Above Right: The G4 Sequencing Platform and key value propositions. The G4 delivers up to four ~30x human genomes in 19 hours.



Above Left: To reduce batching-related delays, up to four flow cells of two types (F2: 150M, F3: 300M; 2x150bp reads) may be analyzed in parallel. To facilitate multiplexing, each flow cell comprises four fluidically-independent lanes.
Above Right: Workflow for library preparation on the G4. Loop adapters are used to attach sequencing primer sites. A second PCR step adds index sequences and G4-specific flow cell adapter sequences.

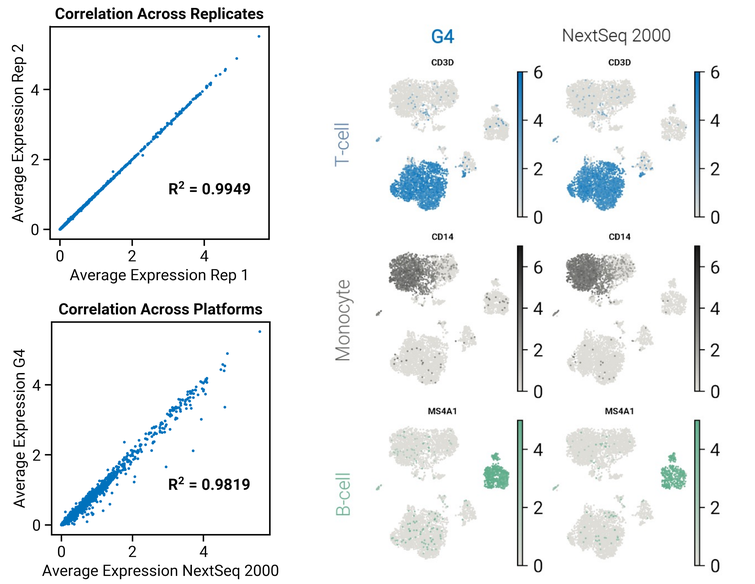
Results

Bulk RNA-Seq



PolyA-selected RNA-Seq libraries were prepared from UHR RNA (Thermo Fisher Scientific) with an ERCC spike-in, sequenced via 2x100bp reads on G4 and Illumina NextSeq 2000 platforms, then downsampled to 25M read pairs for analysis via STAR[1].
Above left: Observed vs expected ERCC counts for a G4 sequencing library.
Middle: Correlation in transcript counts across G4 technical replicates.
Above right: Correlation in transcript counts across platforms.

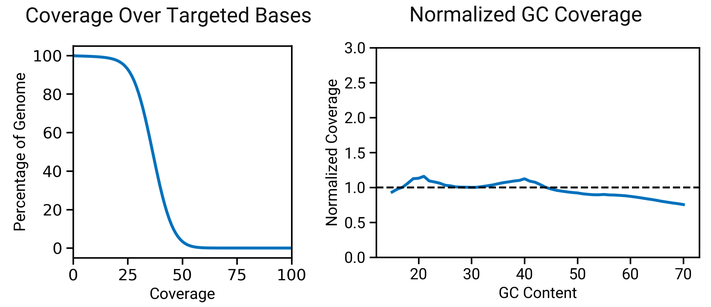
Single Cell RNA-Seq



Top left: Scatterplot illustrating pseudo-bulk RNA-Seq analysis of scRNA-Seq libraries sequenced in replicate on the G4.
Lower left: Pseudo-bulk RNA-seq correlation of libraries sequenced on the G4 and NextSeq 2000.
Top right: Immune cell annotations overlayed onto UMAP visualization of G4 and NextSeq data following processing by Cell Ranger, scanpy[2] and scvi-tools[3]. The gene expression profiles and cell annotations are nearly identical across platforms.

Results (Continued)

Whole Genome Sequencing of HG002

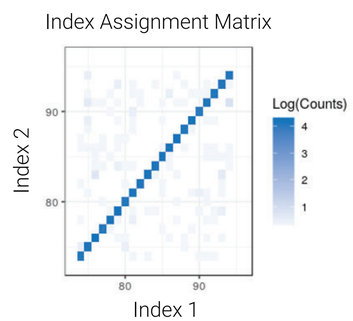


Metrics	31x Coverage Default WGS model	31x Coverage Singular WGS model
SNP Precision	99.86%	99.86%
SNP Recall	99.12%	99.10%
SNP F1-Score	99.49%	99.48%
Indel (<50bp) Precision	98.37%	98.56%
Indel (<50bp) Recall	96.27%	96.81%
Indel F1-Score	97.31%	97.68%

Above: Performance of default and custom DeepVariant v1.4 WGS models applied to HG002 high confidence regions, as determined by hap.py. Both default and custom models show robust performance, reflecting the compatibility of G4 data with tools developed for market leading reversible terminator nucleotide sequencing systems.

Sample Demultiplexing

To assess the fidelity of demultiplexing, 96 libraries were prepared using the Singular Genomics UDI barcoding kit, then sequenced in multiplex.
Right: matrix of index assignment. Barcode misassignment presents as colored cells outside of the identity line. The single index misassignment rate was <0.5% for all indices, implying a dual index misassignment rate of $<2.5x10^{-5}$.



Conclusion

The G4 Sequencing Platform delivers high accuracy and technical reproducibility over a range of applications, with a faster turnaround than traditional reversible terminated nucleotide sequencing systems. Notably, the error profile of the G4 platform closely matches that of Illumina platforms, yielding highly correlated results for RNA expression profiling and WGS. We expect the rapid turnaround and flexible throughput will be especially relevant for future translational and clinical research.

References: 1. Dobin, 2016; 2. Wolf, 2018; 3. Lopez, 2018