SINGULAR GENOMICS

APPLICATION NOTE

RNA Sequencing on the G4®

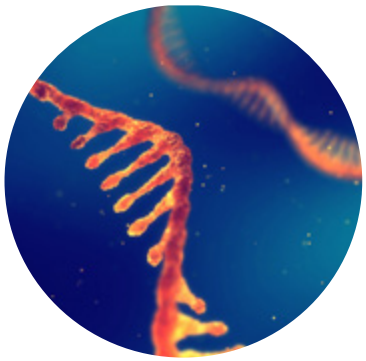
- The unique flow cell design and scalable capacity of the G4 empowers labs to run 4–32 samples per run across 1–4 flow cells and 4–16 lanes.
- The G4 Sequencing Platform delivers highly-accurate RNA sequencing results that demonstrate a strong correlation with the NextSeq 2000 in less than 24 hours.

Introduction

RNA sequencing (RNA-Seq) utilizes next-generation sequencing (NGS) technology to assess differential gene expression, detect novel transcripts and characterize new splice variants or cell types.[1] These methodologies provide greater insight into the genetic mechanisms that differentiate normal and diseased cells. RNA-Seq is used in clinical oncology research and diagnostics as a tool to characterize tumor phenotypes and to develop more effective therapies. Additionally, RNA-Seq is useful in other clinical applications, such as transplant medicine and infectious diseases.[3]

RNA-Seq workflows begin with the isolation of RNA from fresh frozen or formalin fixed and paraffin embedded (FFPE) derived tissue, or blood samples. After isolation, NGS library preparation begins with reverse transcription of RNA into cDNA with the addition of adapter sequences. Strand-specificity in cDNA libraries enables the preservation of antisense transcripts.[4] After sequencing, reads are processed for transcriptome profiling, which involves comparing reads to existing annotations for transcript discovery or mapping and quantifying reads to a reference genome or transcriptome.[5] The analysis of differential expression and alternative splicing enables characterization of cell-types, cellular activity mechanisms, and other phenotypic information.

The G4 Sequencing Platform is a highly versatile benchtop sequencer suitable for RNA sequencing applications. The G4 Platform leverages a novel, 4-color Rapid sequencing-by-synthesis (SBS) chemistry to deliver highly accurate reads (single or paired-read format with optional index reads) in less than 24 hours. The G4 is compatible with existing upstream RNA-Seq library preparation kits and outputs demultiplexed FASTQ files used with existing bioinformatic pipelines.



G4 RNA Sequencing Parameters

The G4 Sequencing Platform supports two different flow cell densities, the F2 flow cell and the F3 flow cell. The G4 features a 4-flow cell design, enabling users to run 1, 2, 3, or 4 flow cells at a time. Each flow cell has 4 independent lanes, enabling up to 16 independent lanes per run, providing users flexibility in designing sequencing experiments. Sequencing output, run time, quality, and throughput by flow cell and run are shown in **Table 1**.

Flow Cell Type	F2	F3
Read Length	Up to 2x150	
Run Time (Hours)[a]	8–24	
Reads / Flow Cell	200M	400M
Reads / Run	800M	1,600M
Quality	80–90% bases ≥ Q30	
Samples / FC[b]	4	8
Samples / Run[b]	16	32

Table 1 RNA sequencing parameters.
[a]Run time is dependent on read length.
[b]RNA-Seq assumptions are based on 2x100 bp, 50M read sequencing per sample.
Throughputs listed are approximations and not guaranteed above kit specifications. Results may vary based on experimental design and sample type. Contact Support for more details.

Methods

All human RNA samples are from Thermo Fisher Scientific Universal Human Reference RNA (Cat #QS0639, lot #301095-000). Poly(A)-selection (NEBNext PolyA mRNA Magnetic Isolation Module; Cat #E7490) steps were carried out per kit guidelines, using 1 µg of input RNA from each sample containing ERCC spike-in Mix 1 (Cat #4456740; 2 µL of 1:100 dilution).

Library Preparation and Sequencing

Poly-A selected libraries were made using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (Cat #E7760), with minor modifications. PCR amplification (8 cycles) was carried out using Singular Genomics Indexed PCR primers (Cat #700,134) at 2 µM final concentration each. Libraries were purified using SparQ PureMag beads (Cat #95196) with a 0.9X ratio, quantified with a Qubit 4 Fluorometer and an Agilent TapeStation. Libraries (331.63 bp ± 17.06 bp) were multiplexed and sequenced on the G4 Sequencing Platform with a 2x100 cycle format on F2 flow cells.

Analysis

Reads were demultiplexed using fgbio's DemuxFastqs (v1.4.0-c00bb7f-SNAPSHOT) with min-mismatch-delta and max-mismatches set to 1. Nf-Core/RNASeq v3.6 (10.5281/zenodo.1400710) was used to perform all downstream bioinformatic processing with remove_ribo_rna turned on for rRNA removal. All other parameters were left as their defaults. All default containers were used except STAR's was changed to quay.io/biocontainers/star:2.7.10a--h9ee0642_0 and Salmon to quay.io/biocontainers/salmon:1.7.0--h10bb6b4_1. Summary statistics and plotting were conducted in R version 4.1.2, using tidyverse_1.3.1, ggpubr_0.4.0, and ggplot2_3.3.5.

Results

Replicate Library Sequencing Results on G4

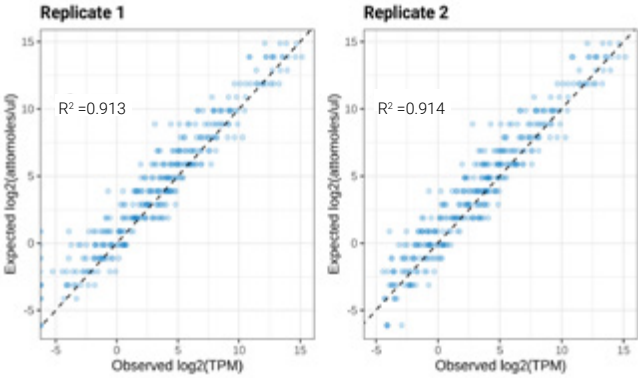
RNA sequencing run metrics from a study comparing replicate Universal Human Reference (UHR) RNA samples is summarized in **Table 2**. Sample replicates were run using a 2x100 bp configuration. Replicates yielded 37M and 55M paired-reads respectively with Q30 base quality scores for all replicates exceeding 86%.

Run Metrics	UHR Rep 1	UHR Rep 2
Read Configuration (bp)	2x100	2x100
Paired-Reads (M)	37	55
Bases ≥ Q30 R1	87.50%	88.25%
Bases ≥ Q30 R2	86.20%	89.22%

Table 2 Replicate run metrics on the G4 Sequencing Platform.

Spike-recovery studies were conducted using ERCC spike-in Mix 1 into UHR replicates. Results are shown in Figures 2A and 2B. **Figure 1A** shows that a high correlation between expected and observed ERCC counts was observed in both replicates (R1, R^2=0.913, R2, R^2=0.914). **Figure 1B** shows very high correlation (R^2=0.990) across transcript counts in replicate libraries generated by PolyA-selection and sequenced on the G4.

(1A) High Expected vs. Observed ERCC Count Correlation



(1B) High Correlation Acress Transcript Counts

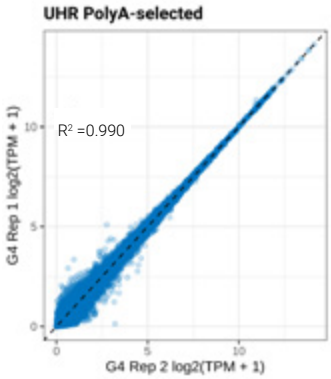


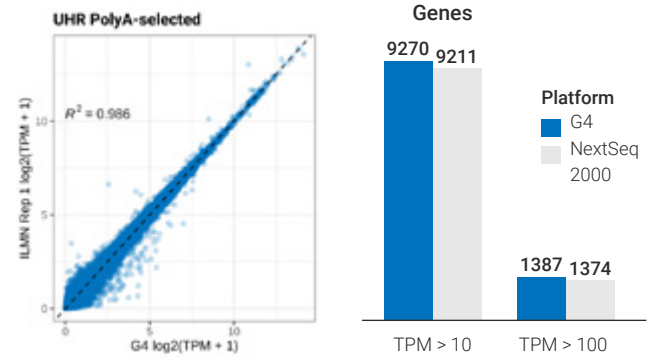
Figure 1 High correlation across replicates. **(1A)** High expected vs. observed ERCC count correlation. **(1B)** High correlation across transcript counts.

Comparison of G4 and NextSeq 2000

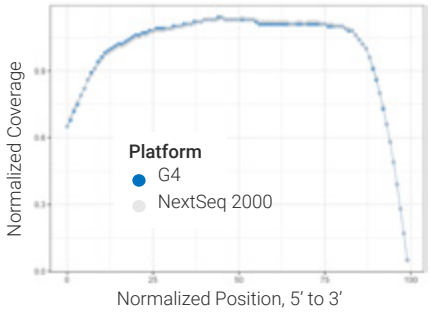
UHR Poly-A selected replicate libraries were run on the G4 and Illumina NextSeq 2000 platforms. **Figure 2A** shows a high correlation of transcript counts (R^2=0.986) between the platforms. The number of detected genes between the platforms was also measured and compared **(Figure 2B)** in transcripts per million (TPM) of >10 and >100. Very similar totals were detected by each platform and at each expression level.

Read coverage uniformity across a gene was compared between platforms **(Figure 2C)**. The normalized coverage plots by gene position show nearly identical coverage uniformity between platforms. Finally, read distribution across exonic, intergenic, and intronic regions were compared **(Figure 2D)**. The percentage of reads obtained from the various genomic regions was highly comparable in transcripts per million (TPM) of >10 and >100. Very similar totals were detected by each platform and at each expression level between platforms.

(2A) Correlation of Transcript Count **(2B) Number of Detected Genes**



(2C) Gene Body Coverage Uniformity



(2D) Read Distribution Across Genic and Intergenic Regions

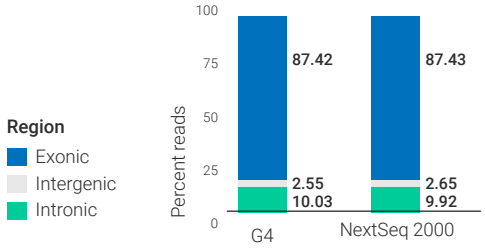


Figure 2 High correlation between G4 and NextSeq 2000. **(3A)** Correlation of transcript counts. **(3B)** Number of detected genes. **(3C)** Gene body coverage uniformity. **(3D)** Read distribution across genic and intergenic regions.

Conclusion

RNA sequencing data generated on the G4 Sequencing Platform demonstrates performance comparable to the Illumina NextSeq 2000 platform. Notably, the gene expression data generated on the G4 and NextSeq 2000 are highly correlated.

The G4 Sequencing Platform is a plug-and-play solution for RNA-Seq workflows that is compatible with existing laboratory operations. The unique flow cell flexibility and run times under 24 hours offer labs the ability to scale operations to match demand and reduce turnaround times on results.

RNA sequencing with the G4 provides users with added flexibility to tailor run sizes and flow cell configurations to the sample set, rather than batching samples to pool onto large flow cells. The G4 Sequencing Platform streamlines lab operations, reduces turnaround times, and delivers highly accurate RNA sequencing results.

*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 support of a world-class experienced team, consisting of industry veterans, 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 system installation and provide instrument service and support. Our field application scientists (FAS) conduct user training and provide support for your applications. Our team is committed to support you when you need us.

Your experienced team is comprised of:

Customer Care Specialists | **Field Application Scientists** | **Field Service Engineers**

Begin Your Journey with G4

[Contact our sales team](#) to learn more about the capabilities of the G4 Sequencing Platform.

Website: www.singulargenomics.com
Email: care@singulargenomics.com
Call: +1 442-SG-CARES (442-742-2737)
Address: 3010 Science Park Rd, San Diego, CA 92121

REFERENCES

1. Geraci *et al.* Editorial: RNA-seq analysis: Methods, applications and challenges. Frontiers in Genetics. 2020;11. doi:10.3389/fgene.2020.00220
2. Buzdin *et al.* RNA sequencing for Research and Diagnostics in Clinical Oncology. Seminars in Cancer Biology. 2020;60:311–23. doi:10.1016/j.semcancer.2019.07.010
3. Byron *et al.* Translating RNA sequencing into Clinical Diagnostics: Opportunities and challenges. Nature Reviews Genetics. 2016;17(5):257–71. doi:10.1038/nrg.2016.10
4. Levin *et al.* Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nature Methods. 2010;7(9):709–15. doi:10.1038/nmeth.1491
5. Conesa *et al.* A survey of best practices for RNA-seq data analysis. Genome Biology. 2016;17(1). doi:10.1186/s13059-016-0881-8