Application Note

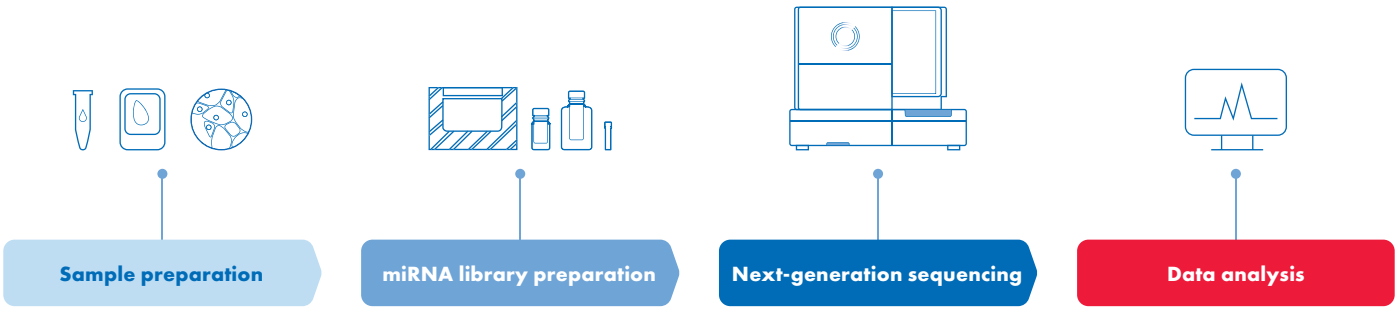
Sequencing of QIAseq® miRNA-derived libraries on the Singular Genomics® G4™ Platform

Introduction

MicroRNAs (miRNAs) are short (approximately 22–24 bp long) non-coding RNAs that regulate RNA expression. Involved in nearly every aspect of biology, miRNAs regulate mRNA expression by binding to the 3′ untranslated region (UTR) of their mRNA targets, which subsequently either prevents translation or promotes degradation of the target mRNA. More recently, miRNAs have also been demonstrated to bind long non-coding RNAs, circular RNAs and even other miRNAs (1). This array of potential interactions has implications for both normal and disease biology, driving an intense interest in miRNAs for most areas of research, particularly in translational regulation. In addition to their broad biological functionality, miRNAs are robustly and stably present in most tissues and biofluids. This ubiquitous functionality and expression profile has brought miRNAs into focus as promising, robust biomarker candidates. Blood-based biomarkers are particularly interesting due to their minimally invasive accessibility (2).

In the most recent version of miRBase (V22.1), there are 2656 annotated mature human miRNAs, and across all species, there are nearly 50,000 annotated mature miRNAs (hereafter referred to as miRNAs). In addition, miRNAs exhibit both 5′ and 3′ "raggedness," presenting an extra layer of complexity for proper quantification and identification of miRNAs. This vast number of potential sequences places one technology – next-generation sequencing (NGS) – at the forefront of miRNA biomarker discovery.

As a massively parallel sequencing technology, NGS enables rapid quantification and identification of known and novel miRNAs from hundreds of samples simultaneously. Critical for miRNA NGS are robust sample preparation, miRNA-specific library preparation, sequencing and bioinformatics. miRNA library preparation must enable sensitive and reliable detection of differentially expressed miRNAs (2, 3), and the sequencing must be ▷



accurate, fast and, most importantly, flexible. miRNA NGS routinely demands anywhere from 5–8 million reads per sample for standard experiments and up to 10–20 million reads per sample for ultra-deep discovery experiments. Adding an extra layer of complexity, the number of samples ready for sequencing at a given time often varies, as samples typically converge from different laboratories.

The QIAseq miRNA Library Kit is optimized to detect miRNAs down to ultralow input levels. This kit integrates unique molecular indices (UMIs) into the reverse-transcription process, enabling unbiased and accurate miRNome-wide quantification of mature miRNAs using NGS. The standard QIAseq miRNA procedure does not require gel purification, excision and elution, saving sample and time. In addition, this kit has a fully optimized library process to prevent adapter dimerization – even from very low inputs of total RNA – and eliminate contaminating RNAs that steal reads. Together, these features of the QIAseq miRNA Library Kit are designed to maximize the yield of miRNA available to sequence (2,3).

The Singular Genomics G4 Sequencing Platform is an innovative benchtop sequencer combining novel 4-color rapid sequencing by synthesis (SBS) chemistry with advanced engineering to provide single-day turnaround times for a broad range of applications. The G4's ability to deliver fast results and run 1–4 flow cells in parallel, each with 4 independently addressable lanes, enables laboratories with highly efficient operations. More information about G4 specifications, such as run time, accuracy and quality metrics, can be found on the Singular Genomics website.

In this Application Note, we evaluated the performance of QIAseq miRNA-derived libraries sequenced on the G4 Sequencing Platform.

Materials and Methods

miRNA library preparation

The QIAseq miRNA Library Kit (QIAGEN) was used to prepare miRNA NGS libraries from 100 ng aliquots of XpressRef Human Universal Total RNA (QIAGEN). In total, 11 technical replicates were performed. The cDNA was split for subsequent Singular Genomics indexing and Illumina® indexing. For the Singular Genomics workflow, SG UDI Primers Set A [1-24] (Singular Genomics) were used to introduce unique dual indexes (UDIs) and Singular Genomics flow cell binding sequences. Similarly, for the Illumina workflow, the QIAseq miRNA 12 Index Kit IL UDI (QIAGEN) was used to add UDIs and Illumina flow cell binding sequences to the QIAseq libraries via PCR.

NGS

Sequencing was carried out using both the Singular Genomics G4 Sequencing Platform and the Illumina MiSeq®, following the corresponding vendor's recommendation for library dilution and flow cell seeding concentration. The sequencing format was set to a single read of 72 bp, plus dual index reads of 12 bp for the G4 Platform, and 8 bp for the MiSeq.

Data analysis

For both the G4 and MiSeq FASTQ files, the GeneGlobe® Data Analysis Portal (QIAGEN) was used for primary analysis. Briefly, during primary analysis, trimmed reads are sequentially aligned to a variety of RNA categories (miRBase mature, miRBase hairpin, piRNA, tRNA, rRNA, mRNA and otherRNA) using bowtie. Once a read is aligned, it cannot be aligned to a second category. For reads not aligned in the first round, a second mapping to miRBase mature occurs, where up to two mismatches are tolerated. miRBase V21 was used for the miRNA reference. Once all reads assigned to a particular RNA are counted, the associated UMIs are clustered to count unique molecules of each RNA. The total miRNA unique molecular index (UMI) count was used for data normalization.

Results and Discussion

Mapping metrics: Singular Genomics G4 Platform vs. Illumina MiSeq

For both Singular Genomics G4 and Illumina MiSeq reads, the GeneGlobe Data Analysis Portal provides an overview of trimming and mapping statistics, and the results are presented as a percentage of total reads (Figure 1).

The results of the eleven libraries derived from the XpressRef Universal RNA are of high quality and indistinguishable from each other. While miRNA mapping results are approximately 70% of total reads in each sample for both sequencing systems, the average miRNA mapping for the G4 is slightly greater than 70% across the samples. In contrast, the average miRNA mapping for the MiSeq is slightly greater than 67% across the samples. Notably, the total background signals (read sets including no_adapter_reads, too_short_reads, UMI_defective_reads, notCharacterized_notMappable) were nearly equal at 18%. Collectively, these high-quality mapping metrics for the Singular Genomics G4, which align with the Illumina MiSeq results, suggest that the G4 Platform chemistry is optimal for miRNA sequencing. The slightly higher miRNA mapping percentage for G4 will be interesting to evaluate in more challenging samples for miRNA NGS, such as biofluids.

Individual miRNA mapping: Singular Genomics G4 Sequencing Platform vs. Illumina MiSeq

The Singular Genomics G4 and Illumina MiSeq NGS systems detected an average of 426 miRNAs in common. Once normalized and averaged across the eleven replicates, the results were log2 transformed and represented on a scatterplot (Figure 2).

When compared, there is a nearly perfect R^2 value (0.9959) between the two NGS systems suggesting that the sequencing results are nearly identical. The influence of read depth was removed from consideration here by selecting miRNAs that were commonly expressed, but this could be evaluated in future studies.

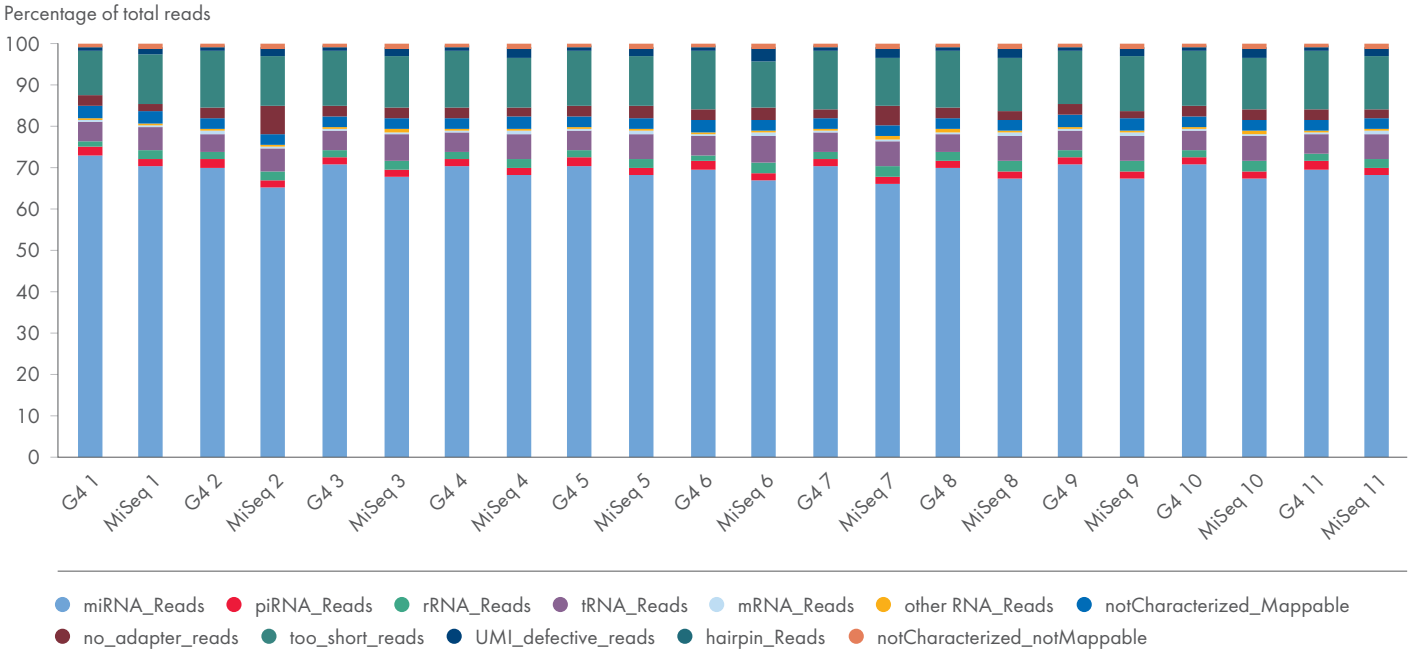


Figure 1. GeneGlobe Data Analysis Portal mapping metrics for QIAseq miRNA libraries prepared from XpressRef Universal RNA and sequenced on the Singular Genomics G4 and Illumina MiSeq NGS systems. Singular Genomics G4 is designated as G4, Illumina MiSeq is designated as MiSeq, and each library replicate is designated as 1 through 11.

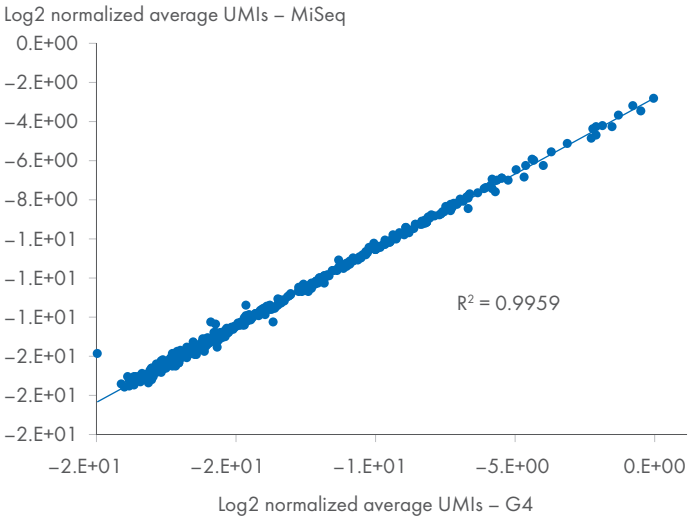


Figure 2. Scatterplot results of 426 miRNA detected by both the Singular Genomics G4 Sequencing Platform and Illumina MiSeq. Results represent the average normalized UMI count across the eleven replicates sequenced for each sample. The results have been log2 transformed to visualize the dynamic range of miRNA expression better.

Run-to-run mapping metrics: Singular Genomics G4 Sequencing Platform

To assess run-to-run sequencing reproducibility, two independent G4 sequencing runs were performed on the same set of eleven QIAseq miRNA XpressRef libraries. As with the G4 Platform comparison to the MiSeq system, the results were analyzed on the GeneGlobe Data Analysis Portal and presented as a percentage of total reads (Figure 3).

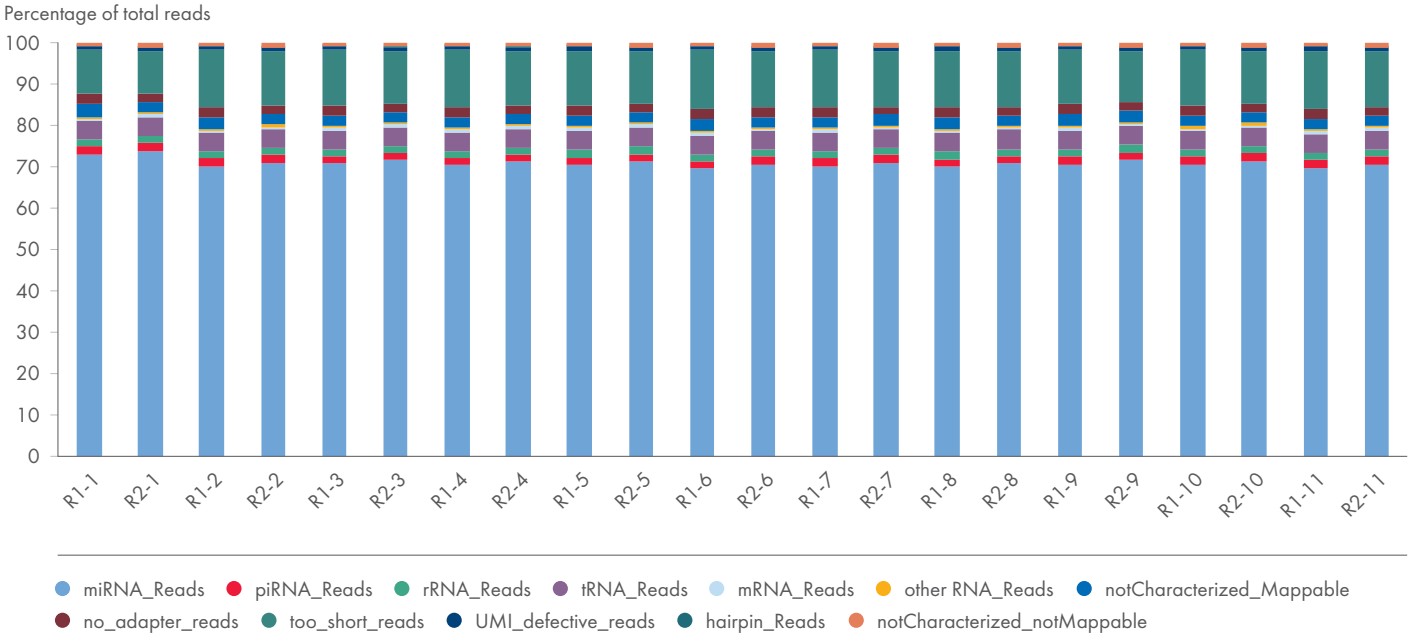


Figure 3. GeneGlobe Data Analysis Portal mapping metrics for QIAseq miRNA libraries prepared from XpressRef Universal RNA and sequenced on two Singular Genomics G4 Platform runs. Run one is designated as R1, run two is designated as R2, and each library replicate is designated as 1 through 11.

Run-to-run individual miRNA mapping: Singular Genomics G4 Sequencing Platform

Between the two independent Singular Genomics G4 runs, both had an average of 747 miRNAs in common. Compared to the G4 vs. MiSeq assessment, the increased capture was most likely due to the higher sequencing read depth for the G4 runs. Once normalized and averaged across the eleven replicates, the results were log2 transformed and represented on a scatterplot (Figure 4).

When compared, there is a nearly perfect R^2 value (0.9932) between the two G4 runs, suggesting that the sequencing results are nearly identical.

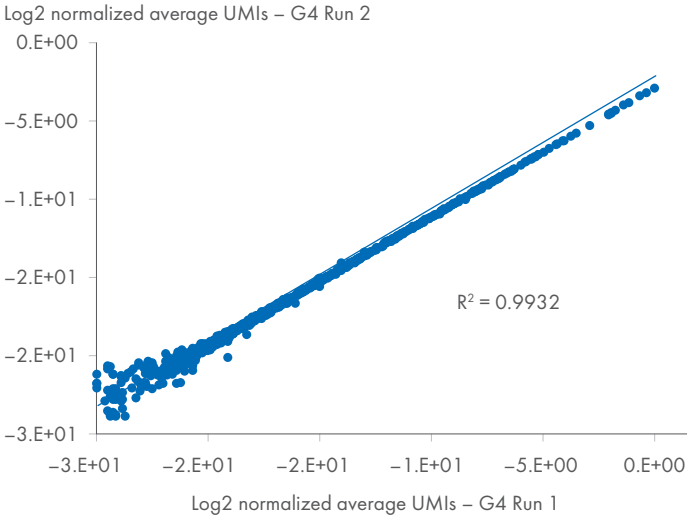


Figure 4. Scatterplot results of 747 miRNAs in common between the two Singular Genomics G4 Sequencing Platform runs. Results represent the average normalized UMI count across the eleven replicates. The results have been log2 transformed to visualize the dynamic range of miRNA expression better.

Since the miRNA mapping percentage was 1% higher in run 2 vs. run 1, we assessed whether this resulted in a greater number of miRNAs being captured. For this assessment, we used 10 UMI counts as an expression cutoff (Figure 5) to remove potential PCR founder effects. In contrast, we did not apply this cutoff for the above described G4 vs. MiSeq and G4 run 1 vs. run 2 comparisons.

The number of detected miRNAs in run 2 increased by 7.5%. This increase is potentially promising, as the detection of known and novel miRNAs is paramount for miRNA biomarker detection. However, that the number of reads per sample were 31% higher in run 2. We cannot rule out effects of read depth on detection; however, the number of reads for both run 1 and 2 were within the acceptable range for deep miRNA sequencing (on average, 16 million reads per sample for run 1 and 21 million reads per sample for run 2). Future studies will be aimed at determining whether sequencing parameter changes improve sensitivity in difficult samples such as biofluids. ▷

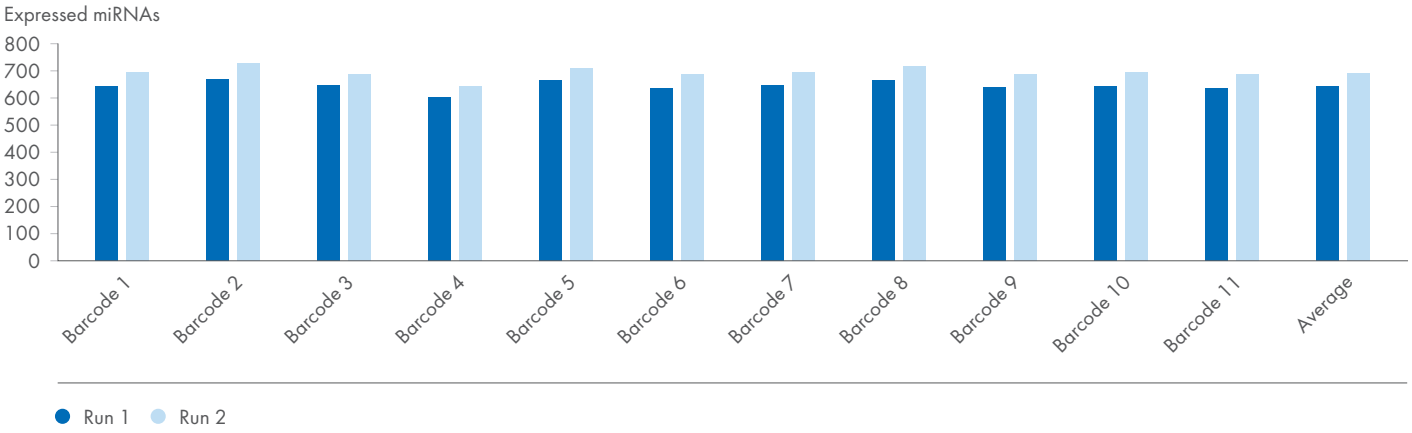


Figure 5. Number of miRNAs exhibiting at least 10 UMI counts in Singular Genomics G4 Sequencing Platform runs 1 and 2. The cutoff for reliable expression detection is generally considered to be 10 UMI counts.

Conclusion

Over the past thirty years, miRNAs have been characterized as a key player in nearly all normal- and disease-related biological pathways. Combining this broad functionality with their robust and stable expression in minimally invasive biofluids, miRNAs are at the forefront of potential biomarker candidates. For miRNA biomarker discovery, NGS is ideally suited as a first step. The differential expression profiles of known and unknown miRNAs can be rapidly characterized in unparalleled high-throughput capacity.

This study evaluated QIAseq miRNA-derived libraries on the Singular Genomics G4 Sequencing Platform. The results demonstrate that the QIAseq miRNA Library Kit achieves optimal performance with the G4 Sequencing Platform. The unique flow cell flexibility and unmatched run times of the G4 allow labs to scale operations to match demand and reduce turnaround times on results. Following sequencing, data was analyzed using GeneGlobe integrated RNA-seq Analysis Portal, access to which is included with the QIAseq miRNA Library Kit. Overall, the QIAseq miRNA Library Kit, in combination with the G4 Platform, is sensitive, robust and reproducible for miRNA biomarker discovery.

References

1. Hill M, Tran N. miRNA interplay: mechanisms and consequences in cancer. *Dis Model Mech.* 2021;14(4).

2. Coenen-Stass AML, et al. Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. *RNA Biol.* 2018;15(8):1133–1145.

3. Heinicke F, Zhong X, Zucknick M, et al. Systematic assessment of commercially available low-input miRNA library preparation kits. *RNA Biol.* 2020;17(1):75–86.

Ordering Information

Product	Supplier	Cat. no.
QIAseq miRNA Library Kit (12)	QIAGEN	331502
QIAseq miRNA Library Kit (96)	QIAGEN	331505
QIAseq miRNA Library Auto Kit (384)	QIAGEN	331509
SG UDI – Set of 96	Singular Genomics	700110
SG UDI – Set of 24 (Set A)	Singular Genomics	700111
SG UDI – Set of 24 (Set B)	Singular Genomics	700112

The QIAseq miRNA Library Kit is intended for molecular biology applications and the Singular Genomics G4 sequencer is intended for research use only.

For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit instructions for use or user operator manual. QIAGEN kit instructions for use and user manuals are available at **www.qiagen.com** or can be requested from QIAGEN Technical Services or your local distributor.

Looking to optimize your miRNA-seq? Visit: **www.qiagen.com/QIAseq-mirna**

Learn more about the capabilities of the G4 Sequencing Platform here: **www.singulargenomics.com/g4**