Streamlined Whole Transcriptome sequencing using Singular Genomics G4 platform and Quantabio sparQ RNA-Seq HMR Kit

Keywords: Whole Transcriptome sequencing, Singular Genomics G4 platform, sparQ RNA-Seq HMR Kit

Abstract

Next-Generation Sequencing with massive parallel short read sequencing technology has proven to be one of the most useful technologies developed to date. The G4™ sequencing platform from Singular Genomics is a powerful and versatile benchtop sequencer that can be used for a wide range of genomic applications. In this application note we demonstrate a fully optimized total RNA-seq workflow using the sparQ RNA-Seq HMR Kit (Quantabio) and the Singular Genomics G4 platform.

Introduction

RNA-seq studies carried out using high-throughput sequencing of cDNA have provided tremendous insight into cellular transcript studies on a large and comprehensive scale. Methods for the focused directional sequencing of RNA provide unparalleled sequence data that has advanced the molecular analysis of disease states and biological traits by allowing identification of both known and novel RNA structural features and isoforms, and accurate quantification of transcripts from both orientations. However, technical challenges such as laborious and lengthy workflows, affordability, compromised accuracy, read coverage biases, and limited transcript diversity have impeded implementation of the technology in many labs. There are additional challenges when working with degraded RNA samples and RNA derived from FFPE, which are commonly used in precision medicine research.

While short-read sequencing technologies have become widely adopted and play a dominant role in NGS analysis, there remains a need for improvements in sequencing speed, flexibility and accuracy to fulfill the emerging demands of genomics and translational research. The Singular Genomics G4™ sequencing platform leverages a novel 4-color sequencing-by-synthesis chemistry to deliver high accuracy data with a rapid turnaround. To maximize flexibility, the G4 can process up to four flow cells at a time (F2: 150M reads; F3:300M reads), each comprising four fluidically- independent lanes.

Here we present a streamlined library prep workflow for total RNA-seq on the G4 platform. These fully optimized protocols utilize Quantabio's simple and robust sparQ RNA-seq HMR Kit with integrated rRNA and globin mRNA depletion. Protocols for these kits were modified to accommodate the looped adapter and barcoded S1-S2 primers used by Singular Genomics.

The Quantabio RNA-Seq protocol presented here will be helpful for genomics and translational research labs in adopting the Singular Genomics platform to achieve high quality sequence data with improved accuracy, flexibility and faster time to result.

Methods

Samples

High quality Universal Human Reference RNA (UHRR) from Qiagen (RIN ~ 9.5) and human adult normal liver tissue FFPE RNA from BioChain (RIN ~ 3.5) were used for this study.

Library Preparation

RNA-seq libraries were prepared using the sparQ RNA-Seq HMR Kit (Quantabio). RNA inputs of 1000 ng and 5 ng UHR RNA or 100 ng and 10 ng FFPE RNA samples were used. UHR RNA samples of 1000 ng and 5 ng were fragmented for 8 min and 4 min respectively. Both FFPE samples were fragmented for 1 min. For the total RNA-Seq library preparation, sparQ RNA-Seq HMR Kit product manual was followed with the described modifications: for the adapter ligation step, we substituted a cleavable stem looped adapter from Singular Genomics, replacing the Y-shaped adapters the kit was originally designed to use. Libraries were then purified using 0.8X bead cleanup and eluted in 15.5 µl elution buffer (10 mM Tris-HCl, pH 8.0). The looped adapter was cleaved followed by HiFi PCR amplification of the libraries was performed using barcoded S1 and S2 primers from Singular Genomics.

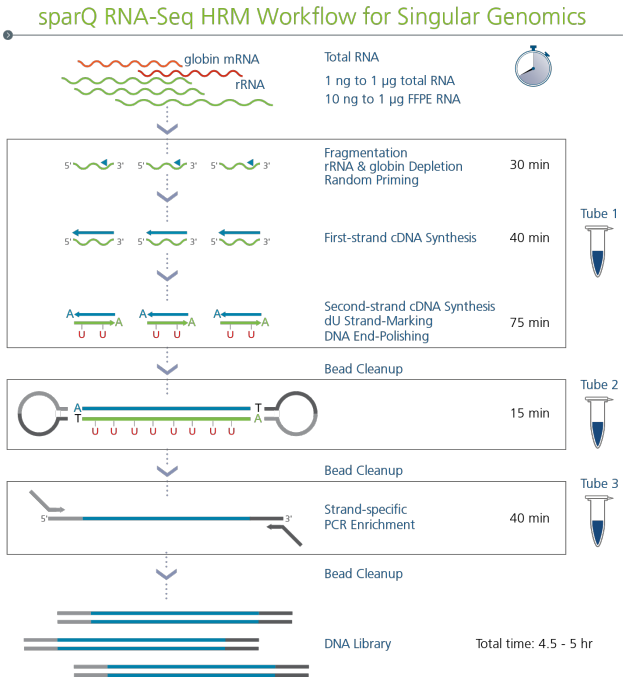

Figure 1: sparQ RNA-Seq HMR workflow for Singular Genomics G4 platform: sparQ RNA-Seq HMR workflow was modified for compatibility with the G4 platform. Looped adapters were used during the ligation step, then cleaved, followed by PCR amplification of the libraries using indexed primers containing S1 and S2 sequences.

Library quantification and validation

To verify the size of the cDNA libraries, 1 µl of the purified libraries was run on a High Sensitivity D1000 ScreenTape on the 4200 TapeStation® System (Agilent).

Libraries were quantified using Qubit® 1X dsDNA High Sensitivity (HS) Assay Kit (Thermo Fisher Scientific) using manufacturer's recommendations: 2 µl purified library was added to 198 µl of Qubit® working solution and mixed by vortexing. The tubes were incubated at room temperature for 2 min before loading the tubes into a Qubit 3.0 Fluorometer.

Sequencing

To verify the size of the PCR products, 1 ul of the DNA was run on a D5000 ScreenTape on the 4200 TapeStation System (Agilent).

Data analysis

FASTQ files generated from the G4 platform were used as an input for data analysis using the CLC Genomics Workbench 20.0.4 software (Qiagen). First, a QC report was generated to assess the quality of the sequencing reads including quality distribution, coverage, nucleotide contribution, GC content, duplication rate etc. The reads were then mapped onto the Human (GRCh38/hg38) reference transcriptome for further analysis.

Results

High quality cDNA libraries

RNA-Seq libraries were prepared using 1000 ng and 5 ng UHR RNA (RIN ~9.5) or 100 ng and 10 ng FFPE RNA (RIN ~3.5) in order to evaluate both high and low input amounts, as well as high and low quality inputs.

All the libraries prepared with the modified sparQ RNA-Seq HMR protocol showed a single narrow distribution pattern in TapeStation with a peak value around 300 bp (data not shown).

Strand specificity

The sparQ RNA-Seq HMR Kit generates directional or strand-specific libraries. In a strand-specific library preparation the reads generated after sequencing should be mapped in reverse (3' to 5') orientation. For all four samples, the percentage of reads mapped in the reverse orientation was 98% or higher (Figure 2A). This high strandedness enhances the mapping of sense and anti-sense reads leading to better transcript quantification and detection of low-level or rare transcripts.

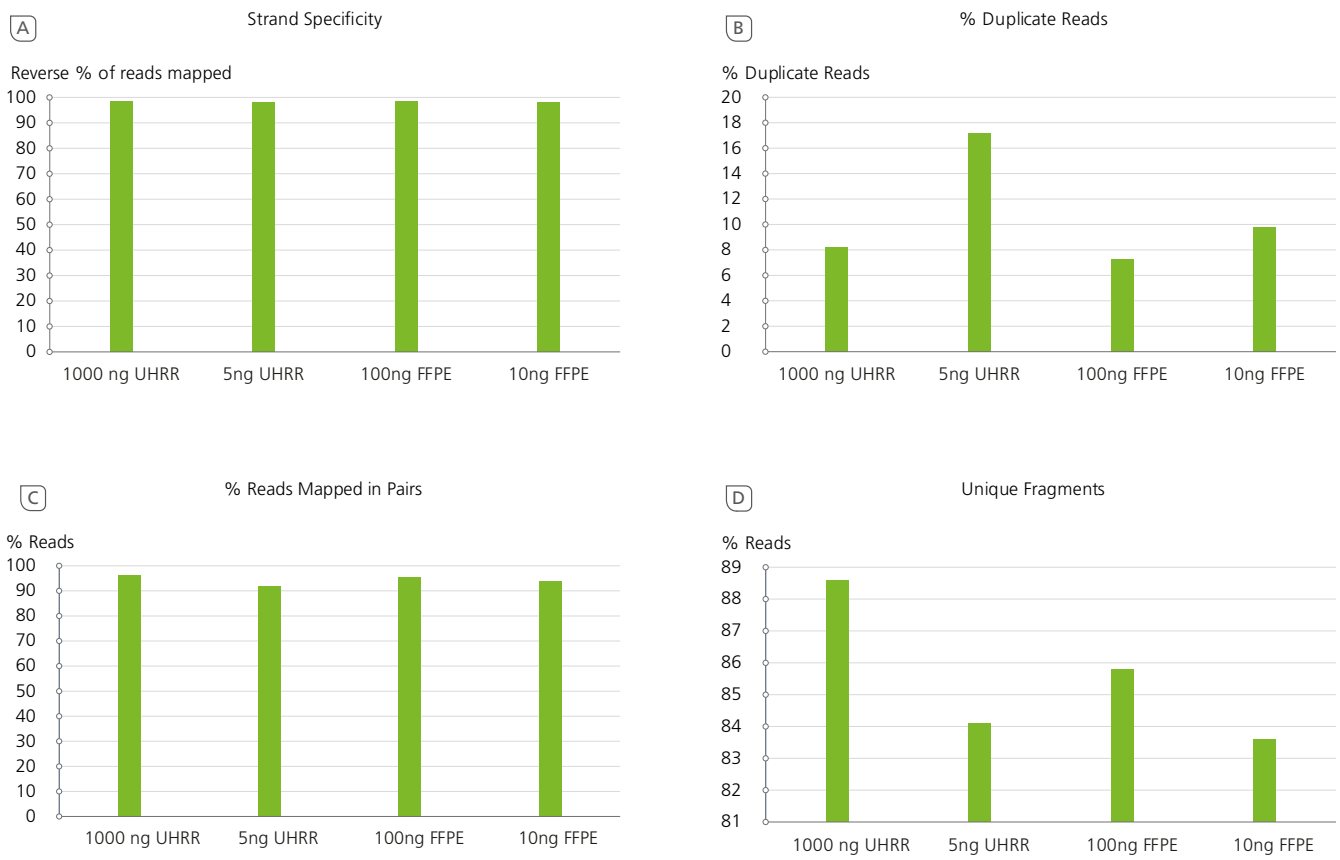

Figure 2: **Sequencing quality metrics:** CLC Genomics Workbench was used to determine (**A**) strand specificity, (**B**) percent duplicate reads, (**C**) percent reads mapped in pairs and (**D**) unique fragments for each library.

Duplication rate

In order to achieve enough library yield for sequencing, RNA-Seq library preparation requires amplification of the cDNA libraries using high fidelity polymerases. A higher number of PCR cycles generates duplicate reads that are not usable for downstream data analysis. Considering this, the 1000 ng UHR RNA libraries were only amplified with 8 cycles whereas all the other libraries were amplified with14 cycles. The duplication rate of all the samples (except 5 ng UHR RNA) was below 10% (Figure 2B) independent of the input amount and quality of the samples.

Mapped reads to the reference transcriptome

To investigate the quality of the reads, we also checked the percentage of reads mapped in pairs. Because strand-specific settings were used for mapping and further data analysis, the percentage of unmapped reads also includes reads that were ignored due to mapping to the wrong strand. The average percentage of reads mapped in pairs was 95% (Figure 2C). Higher input RNA samples showed minor improvement over libraries prepared with lower input samples.

Library diversity

The diversity of each library was assessed by measuring the quantity of unique sequenced fragments. Libraries prepared with sparQ RNA-Seq Kit showed high rates (≥84%) of unique fragments (Figure 2D) indicating high library diversity regardless of RNA input quantity and sample type which enables identification of full length and novel transcripts.

RNA biotype

Finally, we looked at the RNA biotypes within the mapped reads. Out of all mapped reads, approximately 80% of the reads mapped to protein coding or exonic regions. Around 3% of the reads mapped to different non-coding RNA like snRNA and lncRNA which are of great interest to studies of gene regulation pathways. In order to determine the percentage of rRNA reads, all reads from Singular G4

platform were mapped onto human 28S, 18S, 5.8S and mitochondrial 12 and 16S rRNA sequences. The percentage of reads mapped onto the ribosomal reference gene was between 3 to 5% independent of the input amount and quality of the RNA samples.

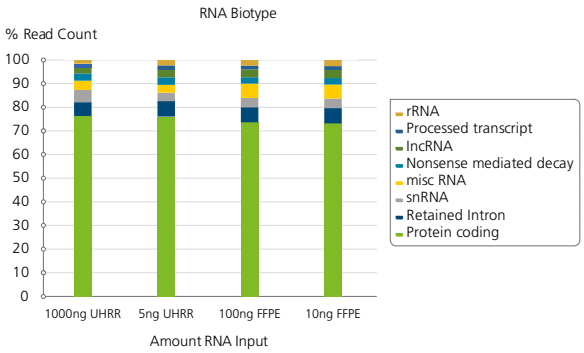

Figure 3: **RNA biotype:** Libraries were prepared from UHR and FFPE RNA samples at high and low input amounts, then sequenced using the Singular Genomics G4 platform. Reads were mapped to the human transcriptome reference sequence. Libraries prepared with this optimized protocol showed a high proportion of protein coding reads and a low proportion of rRNA reads. Diverse non-coding RNA biotypes were also detected.

Conclusions

Transcriptome-wide RNA sequencing is not only useful to understand the gene expression profile of various sample types and possible genetic variations present in transcripts, but also to study the different non-coding RNA sequences involved in gene regulation. Here we demonstrated a streamlined workflow for the sparQ RNA-Seq HMR kit that generates stranded RNA-seq libraries for the Singular Genomics™ G4 platform. In particular, the sparQ RNA-Seq HMR kit is combined with Singular Genomics universal looped adapters and indexed primers containing compatible sequences required to bind with Singular Genomics flow cells. Analysis of the whole transcriptomic data using CLC Genomics Workbench shows high mapping and low duplication rates resulting in more valuable data, eventually lowering sequencing costs. Additionally, this optimized workflow efficiently removes ribosomal RNA, maintaining almost 80% exonic reads which are extremely valuable for a wide range of applications. High quality sequencing results were not only obtained from high quality and higher input samples; similar results were also achieved from degraded and low input samples.

Overall, the compatibility of the sparQ RNA-Seq Kit with the powerful G4 sequencing platform will be helpful for future transcriptomic research and diagnostics.

Trademarks:
TapeStation® is a registered trademark of Agilent Technologies, Inc. QUBIT® is a registered trademark of Thermo Fisher Scientific. Singular Genomics™ and G4™ are trademarks of Singular Genomics Systems, Inc.