

Supplementary Protocol

QIAseq[®] miRNA Primary Quantification

The QIAseq miRNA sequencing files are uploaded to the GeneGlobe[®] Data Analysis Center, and the reads are processed as follows:

1. Calibrate miRBase entries.

For miRNA entries with identical or near-identical sequences in the miRBase mature database, a manual calibration was created and a new, combined miRNA entry was made for each particular miRNA set. For example, the sequence of hsa-miR-151b is entirely contained within the sequence of hsa-miR-151a-5p. As a result, the miRNA entry is hsa-miR-151b/151a-5p.

2. Trim adapter and low-quality bases.

Reads are first processed by trimming off the 3' adapter and low-quality bases using cutadapt (cutadapt.readthedocs.io/en/stable/guide.html). Reads with no adapter sequence are tallied (no_adapter_reads).

3. Identify insert sequences and unique molecular index (UMI) sequences.

Following trimming, the insert sequences and UMI sequences are identified. Reads with less than 16 bp insert sequences (too_short_reads) or less than 10 bp UMI sequences (UMI_defective_reads) are discarded.

4. Align insert sequences.

To annotate the insert sequences, a unique sequence set is made for all readsets/samples in a submitted job. Following this, a sequential alignment strategy is followed to map to different databases (perfect match to miRBase mature, miRBase hairpin, noncoding RNA, mRNA and otherRNA, and ultimately a second mapping to miRBase mature, where up to two mismatches are tolerated) using bowtie (bowtie-bio.sourceforge.net/index.shtml). At each step, only unmapped sequences pass to the next step. Read counts for each RNA category (miRBase mature, miRBase hairpin, piRNA, tRNA, rRNA, mRNA and otherRNA) are calculated from the mapping results (miRNA_Reads, hairpin_Reads, piRNA_Reads, etc.). miRBase V21 is used for miRNA, and piRNABank is used for piRNA.

For human, mouse or rat, a species-specific miRBase mature database is used, and all remaining unmapped sequences are aligned to the respective genome (human: Genome Reference Consortium GRCh38; mouse: Genome Reference Consortium GRCm38; rat: RGSC Rnor_6.0) to identify possible novel miRNA molecules. For species other than human, mouse or rat (referred to as "Other"), a database that includes all miRBase mature entries is used.

5. Count reads and unique molecules assigned to miRNAs, piRNAs, tRNAs and otherRNAs.

For each sample in a submitted job, all reads assigned to a particular miRNA or piRNA ID are counted, and the associated UMIs are clustered to count unique molecules. Read counts and clustered UMI counts are presented in the output Excel® file "miR_piRNA" sheet. For sequences aligned with tRNAs or otherRNAs, these results are displayed in the "tRNA" or "otherRNA" sheet, respectively. For sequences aligned to genome at the last alignment step (this is performed for human, mouse and rat using the most recent genome version), the same information (read counts and clustered UMIs) are output to "notCharacterized_mappable" sheet. Remaining reads are also tallied (notCharacterized_notMappable).

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

Trademarks: QIAGEN®, Sample to Insight®, QIAseq®, GeneGlobe® (QIAGEN Group); Excel® (Microsoft Corporation). 11/2018 HB-2608-001 © 2018 QIAGEN, all rights reserved.