

## **QIAseq miRNA Quantification Steps:**

To quantify miRNA molecules, the uploaded sequence reads are processed using the following steps:

### **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 (<http://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 sequences**

Following trimming, the insert sequences and Unique Molecular Index (UMI) sequences are identified. Reads with less than 16 bp insert sequences (too\_short\_reads) or with 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, non-coding RNA, mRNA, otherRNAs, and ultimately a second mapping to miRBase mature where up to two mismatches are tolerated) using bowtie (<http://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 and rat, a species-specific miRBase mature database is used, and all remaining unmapped sequences are aligned to the respective genome to identify possible

novel miRNA molecules. For species other than human, mouse, and rat (referred to as "Other"), a database including all miRBase mature entries is used.

## **5. Count reads and unique molecules for miRNAs and piRNAs**

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 "otherRNAs", these results are displayed in the "otherRNA" sheet. 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).