

miRNA Analysis Pipeline

Introduction

The GDC miRNA quantification analysis makes use of a modified version of the profiling pipeline that the British Columbia Genome Sciences Centre developed. The pipeline generates TCGA-formatted miRNAseq data. The first step is read alignment. The tool then compares the individual reads to sequence feature annotations in miRBase v21 and UCSC. Of note, however, the tool only annotates those reads that have an exact match with known miRNAs in miRBase and should therefore not be considered for novel miRNA identification or mismatched alignments.

For more information see [BCGSC's GitHub](#) or the [original publication](#).

Data Processing Steps

Alignment Workflow

The miRNA pipeline begins with the [Alignment Workflow](#), which in the case of miRNA uses BWA-aln. This outputs one BAM file for each read group in the input.

| I/O | Entity | Format |
|--------|--|--------------|
| Input | Submitted Unaligned Reads or Submitted Aligned Reads | FASTQ or BAM |
| Output | Aligned Reads | BAM |

miRNA Expression Workflow

Following alignment, BAM files are processed through the [miRNA Expression Workflow](#).

The outputs of the miRNA profiling pipeline report raw read counts and counts normalized to reads per million mapped reads (RPM) in two separate files `mirnas.quantification.txt` and `isoforms.quantification.txt`. The former contains summed expression for all reads aligned to known miRNAs in the miRBase reference. If there are multiple alignments to different miRNAs or different regions of the same miRNA, the read is flagged as cross-mapped and every miRNA annotation is preserved. The latter contains observed isoforms.

| I/O | Entity | Format |
|--------|----------------------------------|--------|
| Input | Aligned Reads | BAM |
| Output | miRNA Expression | TXT |

File Access and Availability

| Type | Description | Format |
|-----------------------------------|--|--------|
| Aligned Reads | miRNA-Seq reads that have been aligned to the GRCh38 build. Reads that were not aligned are included to facilitate the availability of raw read sets. | BAM |
| miRNA Expression Quantification | A table that associates miRNA IDs with read count and a normalized count in reads-per-million-miRNA-mapped. | TXT |
| Isoform Expression Quantification | A table with the same information as the miRNA Expression Quantification files with the addition of isoform information such as the coordinates of the isoform and the type of region it constitutes within the full miRNA transcript. | TXT |

[Site Home](#) | [Policies](#) | [Accessibility](#) | [FOIA](#) | [HHS Vulnerability Disclosure](#)

[U.S. Department of Health and Human Services](#) | [National Institutes of Health](#) | [National Cancer Institute](#) | [USA.gov](#)

NIH... Turning Discovery Into Health ®

Version 1.0