**Materials &** **Methods**

The Zymo Research Aladdin small RNA-seq pipeline is built with [Nextflow](https://www.nextflow.io/)1 and was originally adapted from the [nf-core/smallrnaseq pipeline](https://nf-co.re/smrnaseq)2. In the first pipeline step, [FastQC](https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) (v0.11.9) generated quality control metrics from raw Illumina sequencing reads. Reads were then adapter trimmed and quality filtered with [Trim Galore!](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) (v0.6.6).

Trimmed reads were subsequently aligned using [Bowtie](https://bowtie-bio.sourceforge.net/index.shtml)3 (v1.3.0) against miRNA hairpin reference sequences from [miRBase](https://www.mirbase.org/)4. Resulting miRNA alignments were annotated and quantified using [mirtop](https://github.com/miRTop/mirtop) (v0.4.23). Further miRNA-specific quality control and computational analysis, such as contamination checks, was done with [miRTrace](https://github.com/friedlanderlab/mirtrace)5 (v1.0.1). [isomiRs](https://www.bioconductor.org/packages/release/bioc/html/isomiRs.html)6 (v1.18.1) then used mirtop output to complete differential expression analysis of miRNAs. isomiRs uses the [DESeq2](https://bioconductor.org/packages/release/bioc/html/DESeq2.html)7 R package under the hood for differential expression analysis.

In addition to miRNA, this pipeline used a separate track to quantify the following small RNA types from samples: tRNA, rRNA, lncRNA, scaRNA, snoRNA, snRNA, and miscellaneous RNA. tRNA quantified by this pipeline included both mitochondrial tRNA and mature tRNA. Trimmed reads were aligned with Bowtie (v1.3.0) to mature tRNA reference sequences from [GtRNAdb](http://gtrnadb.ucsc.edu/)8, miRNA hairpin sequences from miRBase, rRNA sequences from [Ensembl](https://useast.ensembl.org/index.html)9 and UCSC repeatmasker, and mitochondrial tRNA, lncRNA, scaRNA, snoRNA, snRNA, and miscellaneous RNA sequences from Ensembl. Bowtie smallRNAseq alignments were then quantified with [RSEM](http://deweylab.github.io/RSEM/README.html)10 (v1.2.28) for an overall perspective of RNA type composition in the sample. After removal of miRNA and rRNA counts, RSEM quantifications were formatted with [tximport](https://bioconductor.org/packages/release/bioc/html/tximport.html)11 (v1.18.0) and input into [DESeq2](https://bioconductor.org/packages/release/bioc/html/DESeq2.html) (v1.30.1) to complete differential expression analysis.

Please consult our smallRNA-Seq report documentation for more details and example data explanations: <https://github.com/Zymo-Research/pipeline-resources/blob/smrnaseq/report_docs/smRNAseq_documentation.md>

**smallRNAseq Reference File Generation**

High confidence mature tRNA sequences are downloaded from GtRNAdb. An example GtRNAdb access link for GRCh38 can be viewed [here](http://gtrnadb.ucsc.edu/genomes/eukaryota/Hsapi38/Hsapi38-seq.html). miRNA hairpin sequences for all species in miRBase are [provided](https://mirbase.org/ftp.shtml) in a single FASTA file. [seqkit](https://bioinf.shenwei.me/seqkit/)12 was used to extract species-specific sequences from the miRBase hairpin file, as well as convert GtRNAdb and miRBase reference sequences from RNA to DNA format. For some assemblies, Ensembl provides a noncoding RNA file with mitochondrial tRNA, lncRNA, snoRNA, scaRNA, snRNA, miscellaneous RNA, and rRNA sequences. An example file for GRCh38 can be viewed [here](https://ftp.ensembl.org/pub/release-105/fasta/homo_sapiens/ncrna/). Sequences with desired gene\_biotypes were all extracted with seqkit from the noncoding RNA file. Transcripts that did not belong to the primary assembly GTF were removed. We supplemented rRNA reference sequences with Repeatmasker sequences from [UCSC’s Table Browser](https://genome.ucsc.edu/cgi-bin/hgTables) with identifiers *5S*, *LSU-rRNA\_Hsa*, and *SSU-rRNA\_Hsa*.

**References**

1. Di Tommaso, P.; Chatzou, M.; Floden, E. W.; Barja, P. P.; Palumbo, E.;  
Notredame, C., Nextflow enables reproducible computational workflows. *Nat Biotechnol*  
**2017**, *35* (4), 316-319.

2. Wang, C.; Pantano, L.; Ewels, P.; Peltzer, A.; Patel, H.; Hammaren, R.; Jemt, A., nf-core/smrnaseq: nf-core/smRNAseq v1.0.0 (Version 1.0.0). Zenodo: 2019.

3. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **2009**, *10* (3).

4. Kozomara, A.; Birgaoanu, M.; Griffiths-Jones, S., miRBase: from microRNA sequences to function. *Nucleic Acids Res* **2019**, *47* (1), 155-162.

5. Kang, W.; Eldfjell, Y.; Fromm, B.; Estivill, X.; Biryukova, I.; Friedländer M. R., miRTrace reveals the organismal origins of microRNA sequencing data. *Genome Biol* **2018**, *19* (1), 213.

6. Pantano, L.; Escaramis, G.,isomiRs: Analyze isomiRs and miRNAs from small RNA-seq. (Version 1.18.1). R package: 2022.

7. Love, M. I.; Huber, W.; Anders, S., Moderated estimation of fold change and  
dispersion for RNA-seq data with DESeq2. *Genome Biol* **2014**, *15* (12), 550.

8. Thornlow, B. P.; Armstrong, J.; Holmes, A.; Howard, J.; Corbett-Detig, R.; Lowe, T., Predicting transfer RNA gene activity from sequence and genome context. *Genome Res* **2019**, *30* (1), 85-94.

9. Cunningham, F.; Allen, J.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.; Armean, I. et al., Ensembl 2022. *Nucleic Acids Res* **2022**, *50* (1), 988-995.

10. Li, B.; Dewey, C., RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **2011**, *12*, 323.

11. Soneson, C.; Love, M. I.; Robinson, M., Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **2015**, *4*, 1521.

12. **Shen, W.**; Le, S; Li, Y; Hu, F., SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLOS ONE **2016**, *11* (10).

**How to Cite Us**

If you use the Aladdin smallRNAseq pipeline for your publications, please cite it as below:

Aladdin smallRNAseq Pipeline (https://github.com/Zymo-Research/aladdin-smrnaseq), Aladdin Bioinformatics Platform, 2022.