**Methods**

The Zymo Research RNAseq pipeline was originally adapted from nf-core/rnaseq pipeline v1.4.2 (<https://github.com/nf-core/rnaseq>).1 The pipelines was built using Nextflow (<https://www.nextflow.io/>).2 Briefly, quality control of raw reads was carried out using FastQC v0.11.9 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Adapter and low-quality sequences were trimmed from raw reads using Trim Galore! v0.6.6 (<https://www.bioinformatics.babraham.ac.uk/projects/trim_galore>). Trimmed reads were aligned to the reference genome using STAR v2.6.1d (<https://github.com/alexdobin/STAR>).3 BAM file filtering and indexing was carried out using SAMtools v1.9 (https://github.com/samtools/samtools).4 RNAseq library quality control was implemented using RSeQC v4.0.0 (<http://rseqc.sourceforge.net/>)5 and QualiMap v2.2.2-dev (<http://qualimap.conesalab.org/>)6. Duplicate reads were marked using Picard tools v2.23.9 (<http://broadinstitute.github.io/picard/>)7. Library complexity was estimated using Preseq v2.0.3 (<https://github.com/smithlabcode/preseq>).8 Duplication rate quality control was performed using dupRadar v1.18.0 (<https://bioconductor.org/packages/dupRadar/>).9 Reads overlapping

with exons were assigned to genes using featureCounts v2.0.1 (<http://bioinf.wehi.edu.au/featureCounts/>).10 Classification of rRNA genes/exons and their reads were based on annotations and RepeatMasker rRNA tracks from UCSC genome browser when applicable. Differential gene expression analysis was completed using DESeq2 v1.28.0 (<https://bioconductor.org/packages/DESeq2/>).11 Functional enrichment analysis was achieved using g:Profiler python API v1.0.0 (https://biit.cs.ut.ee/gprofiler/gost).12 Quality control and analysis results plots were visualized using MultiQC v1.9 (<https://github.com/ewels/MultiQC>).13

**References**

1. Patel, H.; Ewels, P.; Peltzer, A.; Hammarén, R.; Botvinnik, O.; Sturm, G.; Davenport, C., nf-core/rnaseq: nf-core/rnaseq v3.0 - Silver Shark (Version 3.0). Zenodo: 2020.

2. 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.

3. Dobin, A.; Davis, C. A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T. R., STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **2013,** *29* (1), 15-21.

4. Danecek, P.; Bonfield, J. K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M. O.; Whitwham, A.; Keane, T.; McCarthy, S. A.; Davies, R. M.; Li, H., Twelve years of SAMtools and BCFtools. *Gigascience* **2021,** *10* (2).

5. Wang, L.; Wang, S.; Li, W., RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **2012,** *28* (16), 2184-5.

6. García-Alcalde, F.; Okonechnikov, K.; Carbonell, J.; Cruz, L. M.; Götz, S.; Tarazona, S.; Dopazo, J.; Meyer, T. F.; Conesa, A., Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* **2012,** *28* (20), 2678-9.

7. “Picard Toolkit.” 2019. Broad Institute, GitHub Repository. http://broadinstitute.github.io/picard/; Broad Institute.

8. Daley, T.; Smith, A. D., Predicting the molecular complexity of sequencing libraries. *Nat Methods* **2013,** *10* (4), 325-7.

9. Sayols, S.; Scherzinger, D.; Klein, H., dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data. *BMC Bioinformatics* **2016,** *17* (1), 428.

10. Liao, Y.; Smyth, G. K.; Shi, W., featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014,** *30* (7), 923-30.

11. 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.

12. Raudvere, U.; Kolberg, L.; Kuzmin, I.; Arak, T.; Adler, P.; Peterson, H.; Vilo, J., g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* **2019,** *47* (W1), W191-W198.

13. Ewels, P.; Magnusson, M.; Lundin, S.; Käller, M., MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **2016,** *32* (19), 3047-8.

**How to Cite Us**

If you use the Zymo Research RNAseq pipeline for your publications, please cite it as below.

Zymo Research RNAseq Pipeline (<https://github.com/Zymo-Research/aladdin-rnaseq>), Aladdin Open Bioinformatics, 2021.