# RNA-seq Analysis

## Sequencing

RNA was isolated using the mirVana RNA isolation kit and following the Total RNA isolation protocol (Thermo-Fisher Scientific, Waltham MA). RNA library preparation and RNA-sequencing (single-end, 50 bp read length) were performed by the University of Michigan DNA Sequencing Core using the Illumina Hi-Seq 2500 platform. All sequences were deposited in the EMBL-EBI ArrayExpress database using Annotare 2.0 and are cataloged under the accession number E-MTAB-XXXX.

## Alignment

Pseudoalignment of raw Illumina sequence reads was computed using kallisto v0.44.0 (*1*). All sequences were aligned to *Homo sapiens* Genome Reference Consortium human build 38 release 79 (GRCh38.rel79) and the astrovirus VA1 genome (GenBank: 4731478). Aligned reads against each genome were tabulated separately.

## Quantification and Differential Expression analysis

Differential expression of pseudoaligned sequences was calculated using the R package DEseq2 (*2*). The multiple testing-adjusted p-value was calculated using the DESeq2 implementation of the Wald test (*2*) to compare viral-infected HIOs to mock-infected HIOs by time point.

## Gene Set Enrichment Analysis

Gene pathway over-representation tests and Gene Set Enrichment Analysis (*3*) were implemented using the R packages clusterProfiler (*4*) and ReactomePA (*5*). Conserved gene pathways were retrieved from the Gene Ontology Consortium (GO) database (*6*) and REACTOME database (*7*).

## Statistical analysis

All RNA-seq analysis was conducted in R (*8*) using GNU Emacs v25.1 (*9*) on the 64-bit Debian Linux stable version 9 operating system. Plots were constructed using the R package ggplot2 (*10*). Data analysis scripts and further documentation for the RNA-seq analysis are available <https://github.com/hilldr/astrovirus>

# References

1. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic rna-seq quantification. *Nat Biotechnol*. **34**, 525–7 (2016).

2. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biology*. **15**, 550 (2014).

3. A. Subramanian *et al.*, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. **102**, 15545–50 (2005).

4. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, ClusterProfiler: An r package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology*. **16**, 284–287 (2012).

5. G. Yu, Q.-Y. He, ReactomePA: An r/bioconductor package for reactome pathway analysis and visualization. *Molecular BioSystems*. **12**, 477–479 (2016).

6. Gene Ontology Consortium, Gene ontology consortium: Going forward. *Nucleic Acids Res*. **43**, D1049–56 (2015).

7. A. Fabregat *et al.*, The reactome pathway knowledgebase. *Nucleic Acids Res*. **46**, D649–D655 (2018).

8. R Core Team, *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria, 2017; <https://www.R-project.org/>).

9. R. M. Stallman, EMACS the extensible, customizable self-documenting display editor. *ACM SIGOA Newsletter*. **2**, 147–156 (1981).

10. H. Wickham, *Ggplot2: Elegant graphics for data analysis* (Springer-Verlag New York, 2009; [http://ggplot2.org](http://ggplot2.org/)).