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, and L. Pachter. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol, 34(5):525–7, May 2016. PMID: 27043002.

[2] M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15:550, 2014.

[3] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, 102(43):15545–50, Oct 2005. PMID: 16199517.

[4] G. Yu, L.-G. Wang, Y. Han, and Q.-Y. He. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS: A Journal of Integrative Biology, 16(5):284–287, 2012. PMID: 22455463.

[5] G. Yu and Q.-Y. He. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Molecular BioSystems, 12(12):477–479, 2016. PMID: 26661513.

[6] Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res, 43(Database issue):D1049–56, Jan 2015. PMID: 25428369.

[7] A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May, M. Milacic, C. D. Roca, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, G. Viteri, J. Weiser, G. Wu, L. Stein, H. Hermjakob, and P. D'Eustachio. The Reactome Pathway Knowledgebase. Nucleic Acids Res, 46(D1):D649–D655, Jan 2018. PMID: 29145629.

[8] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2017.

[9] R. M. Stallman. EMACS the Extensible, Customizable Self-documenting Display Editor. ACM SIGOA Newsletter, 2(1-2):147–156, April 1981.

[10] H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.