Essential Genes

*Saccharomyces cerevisiae* genes were determined to be essential or non-essential for spaceflight using data derived from Nislow et al. (2015). The experiment consisted of pooling strains from both a heterogenous and homogenous knockout library, exposing the pooled cultures microgravity conditions at the International Space Station (ISS) in an Opticell Processing Module (OPM) and then amplifying and sequencing the barcoded regions. Barcodes were then mapped to a to specific knockout strain and quantified at specific time points. A linear fit was then computed for counts across time points (7-21 generations) and an F-test was performed against a null model. Significant fitness defects were defined as a robust Z score for each time comparison as follows:

Where *MAD* is the median absolute deviation and *R* is the ratio of abundance for the *ith* strain across time points. It is defined as:

Where *g* is the generation of the sampled strain. A P value for each strain can then be computed using the Z score. Strains that were shown to have significant fitness defects at a specific time point are defined as having a log­2R ≥ 1 and a p-value ≤ 0.001 and or having the count drop below background threshold. The genes associated with each knockout strain exhibiting significant fitness defects at any later time point are listed as essential for spaceflight within the database. This includes both heterogenous and homologous knockout samples.

Microarrays

Microarray experiments were pulled from the NCBI Gene Expression Omnibus (GEO) database using the R package GeoQuery (Davis and Meltzer, 2007). Samples were removed if they significantly differed in density distribution and or involved treatments not related to microgravity such as gene knockouts experiments. A log2 transformation and or cyclic loess normalization was applied if needed. The arrays were then fitted to a linear mixed model defined and provided by the R package Limma and a moderated *t* statistic was generated for each gene in addition to a log2 Fold – Change value after the appropriate contrasts were performed (Smyth et al., 2020).

RNAseq

All gene expression arrays derived from high throughput sequencing were processed from raw sequence data provided by the NCBI Sequence Read Archive (SRA) (Leinonen et al., 2011). The processing was performed on the cloud computing environment Galaxy (Afgan et al., 2018). Reads were aligned to their respective reference genome using HISAT after preprocessing and trimming the reads (Kim et al., 2015). The tool featureCounts was then used to quantify exons using the respective organism’s annotated genome (Liao et al., 2014). Count files were then filtered of low-expressed exons and normalized using the R package edgeR and then voom transformed for differential expression analysis using Limma (Chen et al., 2020; Smyth et al., 2020). All metadata was fetched from NCBI GEO.

Citations

Afgan, E., D. Baker, B. Batut, M. van den Beek, D. Bouvier, M. Čech, J. Chilton, D. Clements, N. Coraor, B.A. Grüning, A. Guerler, J. Hillman-Jackson, S. Hiltemann, V. Jalili, H. Rasche, N. Soranzo, J. Goecks, J. Taylor, A. Nekrutenko, and D. Blankenberg. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46:W537–W544. doi:10.1093/nar/gky379.

Chen, Y., A.T. Lun, D.J. McCarthy, M.E. Ritchie, B. Phipson, Y. Hu, X. Zhou, M.D. Robinson, and G.K. Smyth. 2020. edgeR: Empirical Analysis of Digital Gene Expression Data in R. Bioconductor version: Release (3.11).

Davis, S., and P.S. Meltzer. 2007. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*. 23:1846–1847. doi:10.1093/bioinformatics/btm254.

Kim, D., B. Langmead, and S.L. Salzberg. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*. 12:357–360. doi:10.1038/nmeth.3317.

Leinonen, R., H. Sugawara, and M. Shumway. 2011. The Sequence Read Archive. *Nucleic Acids Res.* 39:D19–D21. doi:10.1093/nar/gkq1019.

Liao, Y., G.K. Smyth, and W. Shi. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 30:923–930. doi:10.1093/bioinformatics/btt656.

Nislow, C., A.Y. Lee, P.L. Allen, G. Giaever, A. Smith, M. Gebbia, L.S. Stodieck, J.S. Hammond, H.H. Birdsall, and T.G. Hammond. 2015. Genes Required for Survival in Microgravity Revealed by Genome-Wide Yeast Deletion Collections Cultured during Spaceflight. *BioMed Research International*. 2015:e976458. doi:<https://doi.org/10.1155/2015/976458>.

Smyth, G., Y. Hu, M. Ritchie, J. Silver, J. Wettenhall, D. McCarthy, D. Wu, W. Shi, B. Phipson, A. Lun, N. Thorne, A. Oshlack, C. de Graaf, Y. Chen, M. Langaas, E. Ferkingstad, M. Davy, F. Pepin, and D. Choi. 2020. limma: Linear Models for Microarray Data. Bioconductor version: Release (3.11).