Differential Expression Analysis Description
Processing Statistics
Terms in the following summary statistics explained
DeSeq2 Results
Differential Expression Results: NHD13 vs WT
Software
References

Sample RNAseq Report: NHD13-vs-WT

GRC Bioinformatics Team

8 August 2022

Differential Expression Analysis Description

DESeq2 (-lfc 1 -lfdr 0.1) was used to perform data normalization and differential expression analysis with an adjusted p-value threshold of 0.20 on each set of raw expression measures. The 'RUVinv' method was applied, which mediates log2 fold changes for less expressing genes. The analysis was performed using dreamtoolsp version 1.0.0 (https://github.com/bmc-grc/dreamtoolsp/tree/master). Transcripts with 1-4 CpM in the DESeq2 normalized data and results.

Processing Statistics

For additional information about each sample, please see "multi_qc.html" within the results directory. Full description of all Fastqc figures, fastqc website.

Terms in the following summary statistics explained

LFC = Log2FoldChange

outliers = "The results function automatically flags genes which contain a Cook's distance above a cutoff for samples which have 3 or more replicates. The p-values and adjusted p-values for these genes are set to NA. At least 3 replicates are required for flagging, as it is difficult to judge which sample might be an outlier with only 2 replicates. With many degrees of freedom – i.e., many more samples than the number of coefficients being fitted – a Cook's distance cutoff of 3 is more conservative than the default cutoff of 0.99 that is used by default. These cutoffs can be changed by modifying the parameters to the cooksCutoff argument of results(). Cook's distance plots can be generated using plotCooks() after running DESeq() or nbinomWaldTest(). Cook's distances are stored in assays(dds)[['cooks']]. Cook's distances are also used by default to perform outlier replace counts with large Cook's distance with the trimmed mean over all samples, scaled up by the size factor or normalization factor for that sample. This approach is recommended, it will not lead to false positives, as it raises the Cook's distance only for those samples for which the value 20 Differential analysis of count data - the DESeq2 package predicted by the null hypothesis." DESeq2 Documentation (November 30, 2016)

low counts = "The results function of the DESeq2 package performs independent filtering by default using the mean of normalized counts as a filter statistic. A threshold on the filter statistic is found which optimizes the number of adjusted p values lower than a significance level alpha (we use the standard variable name for significance level, though it is unrelated to the dispersion parameter α). The theory behind independent filtering is discussed in more detail in Section 4.7. The adjusted p values for the genes which do not pass the filter threshold are set to NA." DESeq2 Documentation (November 30, 2016)

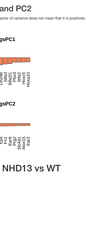
DeSeq2 Results

Sample-to-sample Distances

A heatmap showing the hierarchically clustered Euclidean distances between samples from the regularized log transformation of the normalized count data. This plot is useful for visualizing the variability within and between condition groups.

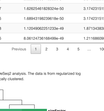
PCA plot

The samples shown in the 2D plane spanned by their first two principal components which are the two components explaining most of the variance. This plot is useful for visualizing the overall effect of experimental covariates and batch effects.



Top/Bottom loading genes for PC1 and PC2

The top and bottom contributing factors of variance. A gene designated as a factor of variance does not mean that it is positively expressed across all samples, rather that it varies widely across samples.



Differential Expression Results: NHD13 vs WT

Summary

```
ran set ZINFA with nonzero total read count
adjusted p-value < 0.20
LFC > 0 (up) : 1842, 6.0%
LFC < 0 (down) : 1776, 6.0%
outliers [1] : 3,268 , 16.42%
low counts [2] : 18845, 63%
(mean count < 2)
```

Raw differential expression results are already included within the Results directory

Only displaying the top 1000 most significantly differentially expressed genes.

CSV

Search:

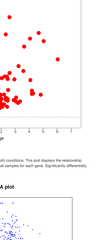
	BaseMean ⇕	logFoldChange ⇕	stat ⇕	pvalue ⇕	padj ⇕
Hova8	635.805	2.554	21.334	5.50825133501588e-101	8.27940237915492e-97
Pfn3	456.874	3.088	20.921	3.45271737985747e-97	2.59499325312723e-93
Pfn1	402.025	3.298	18.425	1.26725408472861e-80	6.34492171442284e-77
H19a	294.801	2.598	16.291	1.34834296554111e-59	5.06673583591739e-56
Vcam1	994.47	-1.868	-15.183	4.55860054052578e-52	1.37049184942984e-48
Ccr2	1462.182	-2.175	-15.078	2.23566601733024e-51	5.57602556954594e-48
F5	806.854	-2.132	-14.947	1.62635481683024e-50	3.17423131592492e-47
Slfn2	299.801	4.619	14.945	1.68841348239618e-50	3.17423131592492e-47
Hova7	253.443	2.520	14.818	1.12649682031233e-49	1.87124082853204e-46
Tanc1	236.21	-2.336	-14.665	8.36124738184599e-49	1.21188693905457e-45

Showing 1 to 10 of 1,000 entries

Previous 1 2 3 4 5 ... 100 Next

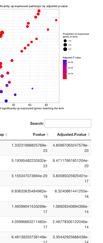
Heatmap

The expression data for all significant differentially expressed genes from the DeSeq2 analysis. The data is from regularized log transformation of the normalized count data. Samples and genes are hierarchically clustered.



Volcano plot

Genes with multiple test corrected p-values <0.05 are colored either green or red according to the direction of the fold change. A line at zero on the x-axis indicates no change in reference. This plot is useful for visualizing the magnitude of the fold changes seen between the two groups being compared.



MA plot

The comparison between the log fold change (M) and log average (A) across both conditions. This plot displays the relationship between differential expression magnitude and the average expression across all samples for each gene. Significantly differentially expressed genes (p-adj <0.05) are colored red.

Enrichr Results

Significant up-regulated and down-regulated genes based on p-adj <0.05 and abs(logFoldChange) > 0 are submitted to Enrichr to identify significantly enriched pathways and transcription factor that could contribute to the observed phenotype. For questions regarding enrichr please see the Enrichr website. Gene set enrichment will not be run on salmon results, only on human or mouse experiments, and will only be run if a given test results in > 50 significantly differentially up/down expressed genes. Even if there are over 50 genes heading into enrichment, it may not result in enriched pathways due to the combinations of genes involved.

Significantly up-regulated pathways

The table below is also included within enrichr_NHD13_vs_WT_up.txt

Show 10 entries

Search:

	database ⇕	Term ⇕	Overlap ⇕	Pvalue ⇕	Adjusted.Pvalue ⇕
1	GO_Biological_Process_2018	neutrophil processing (GO:0001776)	71/227	1.22021998653799e-25	4.69847065247579e-20
2	GO_Biological_Process_2018	ribosome biogenesis (GO:0042254)	70/226	5.16595452329202e-25	9.47417881872356e-20
3	GO_Biological_Process_2018	translation (GO:0006412)	67/232	3.15534737389e-23	3.65850255875401e-18
4	GO_Biological_Process_2018	gene expression (GO:0010467)	92/411	6.50923535459682e-23	6.32408614412026e-18
5	GO_Biological_Process_2018	rRNA processing (GO:0006364)	58/202	1.46589041532289e-19	1.39835934084288e-15
6	GO_Biological_Process_2018	rRNA metabolic process (GO:0016072)	57/200	6.05899665211140e-19	2.49736351120498e-15
7	GO_Biological_Process_2018	cellular macromolecule biosynthetic process (GO:0034645)	82/367	6.48126053737184e-19	3.30442565966108e-15
8	GO_Biological_Process_2018	rRNA aminoacylation for protein translation (GO:0006418)	26/46	7.05802156617509e-19	3.30442565966108e-15
9	GO_Biological_Process_2018	tRNA aminoacylation (GO:0006438)	24/40	1.75880954417739e-18	6.92544755257472e-15
10	GO_Biological_Process_2018	mRNA processing (GO:0006397)	68/283	6.37388436879936e-18	2.32410606363598e-14

Showing 1 to 10 of 284 entries

Previous 1 2 3 4 5 ... 39 Next

Significantly down-regulated pathways

The table below is also included within enrichr_NHD13_vs_WT_down.txt

Show 10 entries

Search:

	database ⇕	Term ⇕	Overlap ⇕	Pvalue ⇕	Adjusted.Pvalue ⇕
1	GO_Biological_Process_2018	neutrophil mediated immunity (GO:0002446)	120/487	2.29244968978447e-25	7.29767758237584e-22
2	GO_Biological_Process_2018	neutrophil activation involved in immune response (GO:0002283)	118/483	3.70738083188449e-25	7.29767758237584e-22
3	GO_Biological_Process_2018	neutrophil degranulation (GO:0043312)	118/479	5.99414936906083e-25	7.88632319080731e-22
4	GO_Biological_Process_2018	cytokine-mediated signaling pathway (GO:0019221)	118/503	8.19196264739074e-25	8.06261107272932e-12
5	GO_Biological_Process_2018	cellular response to cytokine stimulus (GO:0071345)	79/406	8.07256501388181e-18	5.93006350699123305554
6	GO_Biological_Process_2018	actin filament organization (GO:0007015)	31/130	4.60527233817160e-8	0.00005921569671537544
7	GO_Biological_Process_2018	enzyme-linked receptor protein signaling pathway (GO:0007167)	50/120	1.60806569187633e-7	0.00016160602934388
8	GO_Biological_Process_2018	pattern recognition receptor signaling pathway (GO:0002221)	17/48	4.06133119956557e-7	0.00016160602934388
9	GO_Biological_Process_2018	response to cytokine (GO:0034097)	32/138	4.15611328243567e-7	0.00016160602934388
10	GO_Biological_Process_2018	positive regulation of lymphocyte activation (GO:0050871)	21/71	5.85268152922597e-7	0.00023430701766296

Showing 1 to 10 of 48 entries

Previous 1 2 3 4 5 ... 47 Next

Software

```
## R version 4.1.1 (2021-08-10)
## Platform: x86_64-redhat-linux-gnu (64-bit)
## Running under: Red Hat Enterprise Linux Server 7.9 (Maipo)
```

References

```
## Matrix products: default
## BLAS/LAPACK: /gpfs/fs2/scratch/grc_group/.conda/envs/grc_rnaSeq_3.1/lib/libopenblasp-r0.3.18.so
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats4    stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
##  [1] DOSE_3.20.0                 DT_0.20
##  [3] plyr_1.8.6                  gridExtra_2.3
##  [5] ggplot2_3.3.5               enrichR_3.0
##  [7] biomaRt_2.50.0              readr_2.1.0
##  [9] tximport_1.22.0             knitr_1.35
## [11] BiocStyle_2.22.0            pcaExplorer_2.20.0
## [13] genefilter_1.76.0           RColorBrewer_1.1-2
## [15] pheatmap_1.0.12             foreach_1.5.1
## [17] BiocParallel_1.28.0         DESeq2_1.34.0
## [19] SummarizedExperiment_1.24.0 Biobase_2.54.0
## [21] MatrixGenerics_1.6.0        matrixStats_0.61.0
## [23] GenomicRanges_1.46.0        GenomeInfoDb_1.30.0
## [25] IRanges_2.28.0              S4Vectors_0.32.0
## [27] BiocGenerics_0.40.0
##
## loaded via a namespace (and not attached):
##   [1] GOstats_2.60.0         fastmatch_1.1-3        BiocFileCache_2.2.0
##   [4] NMF_0.21.0             igraph_1.2.8           lazyeval_0.2.2
##   [7] GSEABase_1.56.0        shinydashboard_0.7.2   splines_4.1.1
##  [10] crosstalk_1.2.0        gridBase_0.4-7         digest_0.6.28
##  [13] GOSemSim_2.20.0        htmltools_0.5.2        viridis_0.6.2
##  [16] GO.db_3.14.0           fansi_0.4.2            magrittr_2.0.1
##  [19] memoise_2.0.0          cluster_2.1.2          doParallel_1.0.16
##  [22] tzdb_0.2.0             limma_3.50.0           Biostrings_2.62.0
##  [25] annotate_1.72.0        prettyunits_1.1.1      colorspace_2.0-2
##  [28] blob_1.2.2             rappdirs_0.3.3         ggrepel_0.9.1
##  [31] xfun_0.28              dplyr_1.0.7            crayon_1.4.2
##  [34] RCurl_1.98-1.5         jsonlite_1.7.2         graph_1.72.0
##  [37] survival_3.2-13        iterators_1.0.13       glue_1.5.0
##  [40] registry_0.5-1         gtable_0.3.0           zlibbioc_1.40.0
##  [43] XVector_0.34.0         webshot_0.5.2          DelayedArray_0.20.0
##  [46] Rgraphviz_2.38.0       SparseM_1.81           scales_1.1.1
##  [49] DBI_1.1.1              rngtools_1.5.2         Rcpp_1.0.7
##  [52] viridisLite_0.4.0      xtable_1.8-4           progress_1.2.2
##  [55] bit_4.0.4              AnnotationForge_1.36.0 htmlwidgets_1.5.4
##  [58] httr_1.4.2             fgsea_1.20.0           threejs_0.3.3
##  [61] shinyAce_0.4.1         ellipsis_0.3.2         farver_2.1.0
##  [64] pkgconfig_2.0.3        XML_3.99-0.8           sass_0.4.0
##  [67] dbplyr_2.1.1           locfit_1.5-9.4         utf8_1.2.2
##  [70] labeling_0.4.2         tidyselect_1.1.1       rlang_0.4.12
##  [73] reshape2_1.4.4         later_1.2.0            AnnotationDbi_1.56.1
##  [76] munsell_0.5.0          tools_4.1.1            cachem_1.0.6
##  [79] generics_0.1.1         RSQLite_2.2.8          shinyBS_0.61
##  [82] evaluate_0.14          stringr_1.4.0          fastmap_1.1.0
##  [85] heatmaply_1.3.0        yaml_2.2.1             bit64_4.0.5
##  [88] purrr_0.3.4            KEGGREST_1.34.0        dendextend_1.15.2
##  [91] RBGL_1.70.0            mime_0.12              DO.db_2.9
##  [94] xml2_1.3.2             compiler_4.1.1         plotly_4.10.0
##  [97] filelock_1.0.2         curl_4.3.2             png_0.1-7
## [100] tibble_3.1.6           geneplotter_1.72.0     bslib_0.3.1
## [103] stringi_1.7.5          highr_0.9              lattice_0.20-45
## [106] Matrix_1.3-4           vctrs_0.3.8            pillar_1.6.4
## [109] lifecycle_1.0.1        BiocManager_1.30.16    jquerylib_0.1.4
## [112] data.table_1.14.2      bitops_1.0-7           seriation_1.3.1
## [115] qvalue_2.26.0          httpuv_1.6.3           R6_2.5.1
## [118] TSP_1.1-11             promises_1.2.0.1       topGO_2.46.0
## [121] codetools_0.2-18       assertthat_0.2.1       rjson_0.2.20
## [124] Category_2.60.0        pkgmaker_0.32.2        withr_2.4.2
## [127] GenomeInfoDbData_1.2.7 parallel_4.1.1         hms_1.1.1
## [130] grid_4.1.1             tidyr_1.1.4            rmarkdown_2.11
## [133] shiny_1.7.1            base64enc_0.1-3
```

References

1. Love MI, Huber W and Anders S (2014). “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.” Genome Biology, 15, pp. 550. doi: 10.1186/s13059-014-0550-8.
2. R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
3. Orchestrating high-throughput genomic analysis with Bioconductor. W. Huber, V.J. Carey, R. Gentleman, …, M. Morgan Nature Methods, 2015:12, 115.
4. Marini F (2017). pcaExplorer: Interactive Visualization of RNA-seq Data Using a Principal Components Approach. R package version 2.4.0, https://github.com/federicomarini/pcaExplorer.
5. Dundar F, Skrabanek L, Zumbo P (2017). Introduction to differential gene expression analysis using RNA-seq. Weill Cornell Medical College. URL http://chagall.med.cornell.edu/RNASEQcourse/Intro2RNAseq.pdf
6. Love MI, Anders S, Kim V, Huber W (2017). rnaseqGene: RNA-seq workflow: gene-level exploratory analysis and differential expression. R package version 3.4.2. URL http://www.bioconductor.org/help/workflows/rnaseqGene/