Next-generation sequencing - analysis with DESeq2

Comparison of gene expression in DSCs compared to melanocytes - functional analysis

Marc Bender

2024-06-13

Contents

| Resources and introduction | 2 |
|---|----|
| Pathway analysis (Wikipathways) | 3 |
| Master regulators | 11 |
| Analysis of donor effects (high dispersion genes) | 14 |
| Appenix | 15 |
| Session Info | 15 |

Resources and introduction

This document includes the functional analysis for the comparison of RNASeq data from dermal stem cells (DSCs) and Melanocytes. The exploratory data analysis and differential expression analysis are included in separate files. Statistical methods are described in a separate .docx file.

 $\label{lem:packages} For additional info on analysis workflows check: DESEQ workflow: https://www.bioconductor.org/packages/release/bioc/vignettes/DESeq2/inst/doc/DESeq2.html https://bioconductor.org/packages/release/workflows/vignettes/rnaseqGene/inst/doc/rnaseqGene.html https://www.bigbioinformatics.org/r-and-rnaseq-analysis https://github.com/bigbioinformatics/r-programming-and-rnaseq-workshop ClusterProfiler: https://pubmed.ncbi.nlm.nih.gov/34557778/$

Pathway analysis (Wikipathways)

Pathway analysis was conducted with the R package pathfindR combining a classical overrepresentation analysis with the information gained from protein interaction networks to take into account semantic relationships between differentially expressed genes.

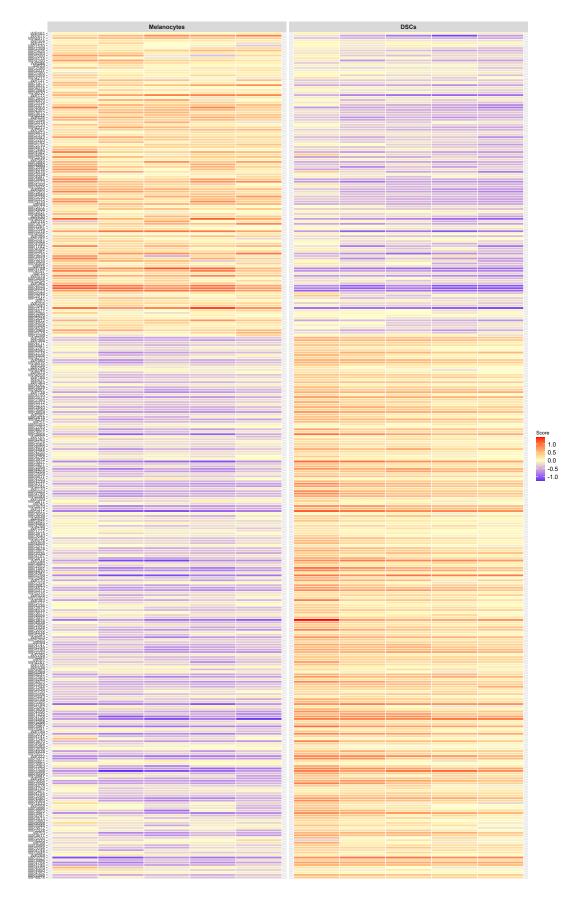


Figure 1: Pathway scores between melanocytes and DSCs. All pathways. $\begin{tabular}{c} 4 \end{tabular}$

In total 342 pathways were significantly altered. This high amount of regulated pathways impedes interpretaion especially due to a substantial overlap between certain pathways and biological processes. Therefore we chose to analyse only representative pathways. PathfindR offers an option to cluster pathways based on shared genes and assign a representative pathway to each cluster based on the lowest adjusted pvalue. This leads to reduction of superfluous information. Clusters were formed separately for all upregulated and all downregulated pathways, respectively.

The top 30 activated and top 30 repressed pathways (based on a rank which was calculated as the square root of the fold-enrichment and the inverse logarithm of the pvalue) are shown in a wordcloud to facilitate biological inferences.

Nitric oxide metabolism in cystic fibrosis

Male infertility HDAC6 interactions in the central nervous system

Chemokine signaling pathway Network map of SARS CoV 2 signaling pathway

Microglia pathogen phagocytosis pathway Photodynamic therapy induced NFE2L2 NRF2 survival signaling

TNF alpha signaling pathway TROP2 regulatory signaling
Familial hyperlipidemia type 3

NRF2 pathway
Measles virus infection

NRF2 pathway
Measles virus infection

NRF3 pathway
Measles virus infection

NRF4 pathway
NRF5 pathway
Measles virus infection

NRF6 pathway
NRF6 pathway
NRF7 pathway

Measles virus intection
of secretase in neurodegenerative diseases
Transcriptional cascade regulating adipogenesis
Cytoplasmic ribosomal proteins Parkin ubiquitin proteasomal system pathway Acute myeloid leukemia

UDP derived sugars synthesis in fibroblasts

SARS CoV 2 and COVID 19 pathway

miRNA regulation of DNA damage response

Female steroid hormones in cardiomyocyte energy metabolism

Mammary gland development pathway Embryonic development Stage 1 of 4

Ectoderm differentiation Development of ureteric collection system

Endoplasmic reticulum stress response in coronavirus infection NRP1 triggered signaling pathways in pancreatic cancer MAPK signaling pathway TGF beta signaling in thyroid cells for epithelial mesenchymal transition

Apoptosis Mammary gland development pathway Puberty Stage 2 of 4
Hippo Merlin signaling dysregulation Endo

Endochondral ossification

Pathogenic Escherichia coli infection

Adipogenesis Ebola virus infection in host

EGF EGFR signaling pathway Focal adhesion Wnt signaling pathway

miRNA targets in ECM and membrane receptors
Spinal cord injury

TGF beta signaling pathway Primary focal segmental glomerulosclerosis FSGS

Spinal cord injury
TGF beta receptor signaling Glioblastoma signaling pathways
Influence of laminopathies on Wnt signaling
Metabolic pathways of fibroblasts Neovascularization processes

Small cell lung cancer Neural crest differentiation

Hair follicle development cytodifferentiation part 3 of 3

Brain derived neurotrophic factor BDNF signaling pathway

Figure 2: Wordcloud of representative pathways. Top30 activated and repressed pathways are shown.

As representative pathways might be less known than other pathways in each cluster we also printed the clusters with the highest number of pathways in each respective cluster. For upregulated pathways only clusters with more than 3 pathways were kept, for downregulated pathways only clusters with more than 5 pathways were kept.



Figure 3: Activated pathways clustered by amount of shared genes.

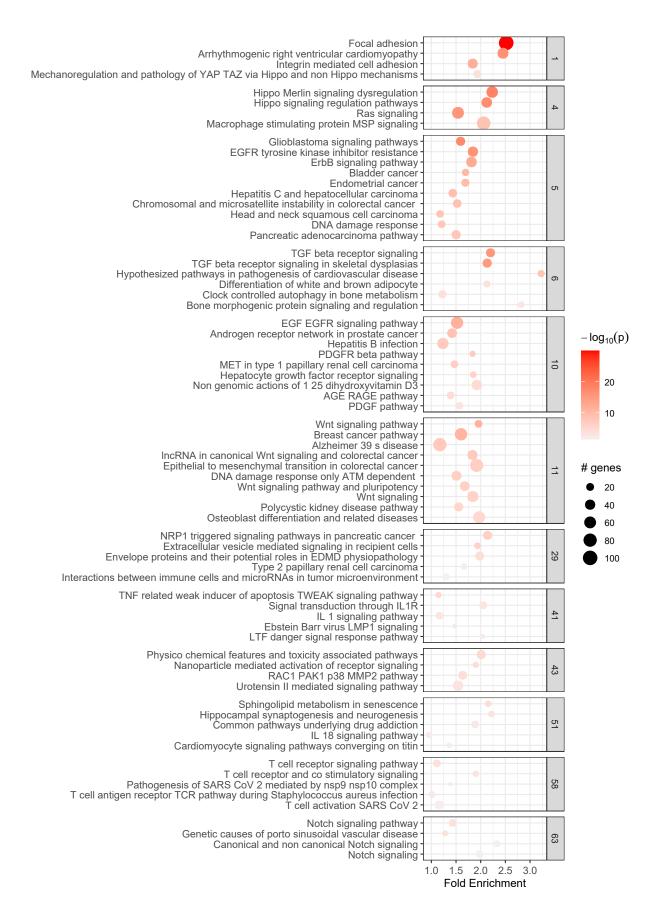


Figure 4: Repressed pathways clustered by amount of shared genes.

Master regulators

Master regulators are transcription factors or regulatory proteins at the top of a gene regulation hierarchy which target downstream effector proteins and orchestrate gene sets. The master regulator analysis was performed with the corto package. In a first approach 294 suggested master regulators (extracted from the corto package) and 36 skin marker genes (for melanocytes, DSCs and fibroblasts) were used as input for the algorithm. In a second approach only the 294 suggested master regulators without skin specific markers were used to identify novel regulatory genes.

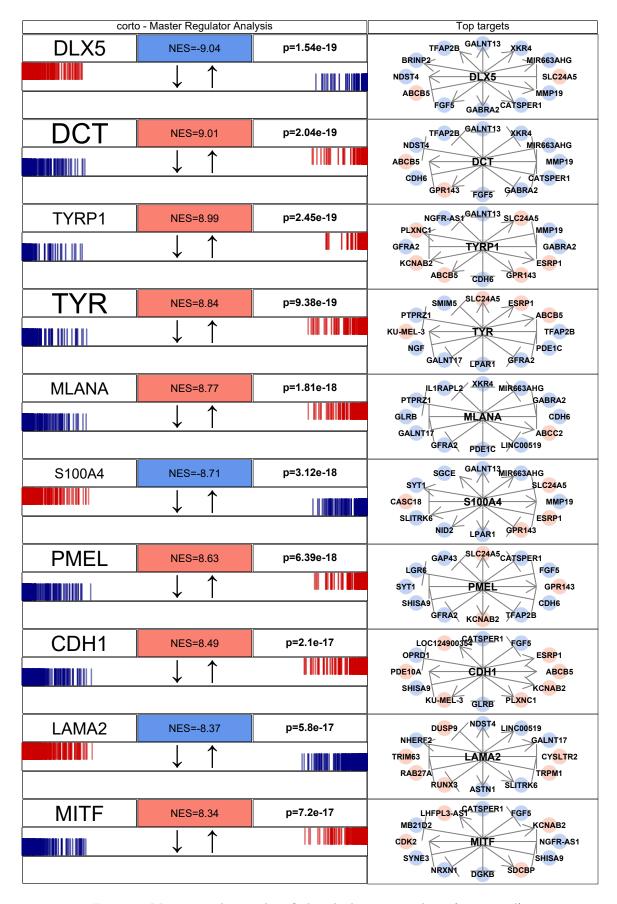


Figure 5: Master regulators identified with the corto package (supervised). $12\,$

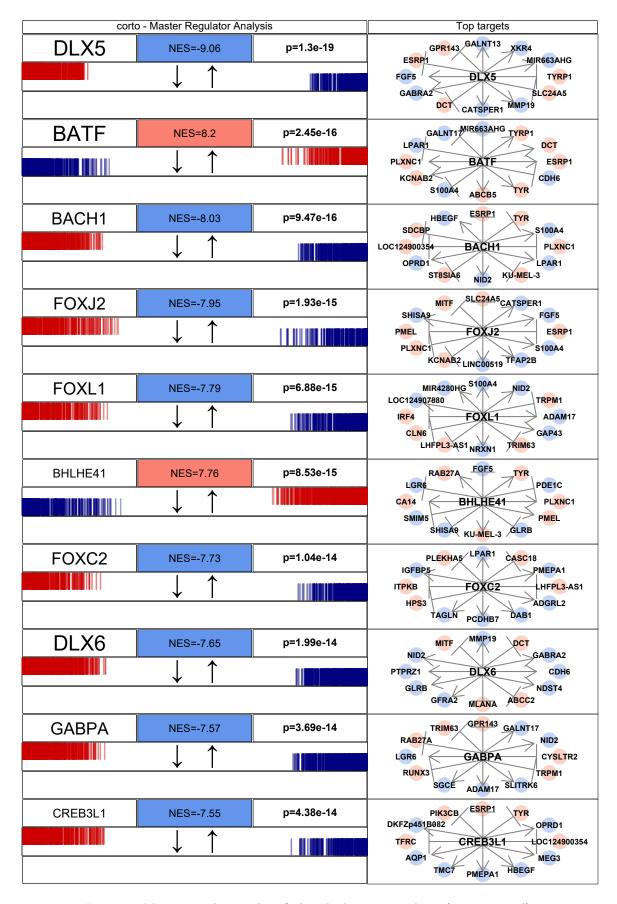


Figure 6: Master regulators identified with the corto package (unsupervised). 13

Analysis of donor effects (high dispersion genes)

Analysis has been conducted by using high dispersion genes as input for the goseq pipeline. Related terms were reduced with the rrvgo function reduceSimMatrix() with a similarity threshold of 0.7. The top 30 reduced terms are depoited in the wordcloud plot.

> regulation of cell activation myeloid dendritic cell antigen processing and presentation positive regulation of developmental process anatomical structure development circulatory system development positive regulation of intracellular signal transduction tissue development peptide antigen assembly with MHC class II protein complex

> > tube developmentresponse to stimulus

regulation of cell population proliferation multicellular organismal process cell division attachment of spindle microtubules to kinetochore

positive regulation of ERK1 and ERK2 cascade regulation of nuclear division response to chemical

regulation of mitotic sister chromatid separation positive regulation of smooth muscle cell migration

symbiont entry into host cell regulation of reproductive process cell adhesion regulation of cellular process negative regulation of chromosome organization

renal system vasculature morphogenesis positive regulation of chemotaxis

positive regulation of phospholipase C activity negative regulation of protein localization to cell periphery positive regulation of protein modification process vascular endothelial growth factor signaling pathway

Figure 7: Wordcloud of reduced GOBP-terms. Top30 GOBP terms for high dispersion genes are shown.

Appenix

Session Info

```
## R version 4.4.0 (2024-04-24 ucrt)
## Platform: x86_64-w64-mingw32/x64
## Running under: Windows 10 x64 (build 19041)
## Matrix products: default
##
##
## locale:
## [1] LC_COLLATE=German_Germany.utf8 LC_CTYPE=German_Germany.utf8
## [3] LC_MONETARY=German_Germany.utf8 LC_NUMERIC=C
## [5] LC_TIME=German_Germany.utf8
##
## time zone: Europe/Berlin
## tzcode source: internal
## attached base packages:
## [1] stats4
                 stats
                           graphics grDevices utils
                                                         datasets methods
## [8] base
##
## other attached packages:
## [1] ggwordcloud_0.6.1
                                    pathfindR_2.4.1
## [3] pathfindR.data_2.1.0
                                    corto_1.2.4
## [5] clusterProfiler_4.12.0
                                    lubridate_1.9.3
## [7] forcats_1.0.0
                                    stringr_1.5.1
## [9] dplyr_1.1.4
                                    purrr_1.0.2
## [11] readr_2.1.5
                                    tidyr_1.3.1
## [13] tibble_3.2.1
                                    ggplot2_3.5.1
## [15] tidyverse_2.0.0
                                    DESeq2_1.44.0
## [17] SummarizedExperiment_1.34.0 Biobase_2.64.0
## [19] MatrixGenerics_1.16.0
                                    matrixStats 1.3.0
## [21] GenomicRanges_1.56.0
                                    GenomeInfoDb_1.40.0
## [23] IRanges 2.38.0
                                    S4Vectors_0.42.0
## [25] BiocGenerics_0.50.0
## loaded via a namespace (and not attached):
##
     [1] fs_1.6.4
                                  bitops_1.0-7
                                                            enrichplot_1.24.0
##
     [4] HDO.db_0.99.1
                                  httr_1.4.7
                                                            RColorBrewer_1.1-3
     [7] doParallel_1.0.17
                                  tools_4.4.0
                                                            utf8_1.2.4
## [10] R6_2.5.1
                                  lazyeval_0.2.2
                                                            mgcv_1.9-1
## [13] withr_3.0.0
                                  prettyunits_1.2.0
                                                            gridExtra_2.3
## [16] cli_3.6.2
                                  textshaping_0.3.7
                                                            scatterpie_0.2.2
## [19] officer_0.6.6
                                  labeling_0.4.3
                                                            slam_0.1-50
## [22] tm_0.7-13
                                  goseq_1.56.0
                                                            pbapply_1.7-2
## [25] askpass_1.2.0
                                                            Rsamtools_2.20.0
                                  commonmark_1.9.1
## [28] systemfonts_1.0.6
                                  yulab.utils_0.1.4
                                                            gson 0.1.0
## [31] txdbmaker_1.0.0
                                  gfonts_0.2.0
                                                            DOSE_3.30.0
## [34] plotrix_3.8-4
                                  rstudioapi_0.16.0
                                                            RSQLite_2.3.6
                                  treemap_2.4-4
## [37] httpcode_0.3.0
                                                            generics_0.1.3
## [40] gridGraphics_0.5-1
                                                            gtools_3.9.5
                                  BiocIO_1.14.0
## [43] zip_2.3.1
                                  GO.db_3.19.1
                                                            Matrix_1.7-0
```

```
[46] fansi 1.0.6
                                  abind 1.4-5
                                                            lifecycle 1.0.4
##
  [49] yaml_2.3.8
                                  gplots_3.1.3.1
                                                            qvalue_2.36.0
## [52] SparseArray 1.4.1
                                  BiocFileCache 2.12.0
                                                            grid 4.4.0
## [55] blob_1.2.4
                                  promises_1.3.0
                                                            crayon_1.5.2
## [58] lattice 0.22-6
                                  cowplot_1.1.3
                                                            GenomicFeatures_1.56.0
## [61] KEGGREST 1.44.0
                                  magick 2.8.3
                                                            pillar 1.9.0
## [64] knitr 1.46
                                  fgsea 1.30.0
                                                            rjson 0.2.21
                                  fastmatch 1.1-4
## [67] codetools 0.2-20
                                                            glue_1.7.0
## [70] ggfun_0.1.4
                                  fontLiberation_0.1.0
                                                            data.table_1.15.4
## [73] vctrs_0.6.5
                                  png_0.1-8
                                                            treeio_1.28.0
## [76] gtable_0.3.5
                                  cachem_1.0.8
                                                            xfun_0.43
## [79] S4Arrays_1.4.0
                                  mime_0.12
                                                            tidygraph_1.3.1
## [82] pheatmap_1.0.12
                                  iterators_1.0.14
                                                            tinytex_0.51
                                  ggtree_3.12.0
                                                            bit64_4.0.5
## [85] nlme_3.1-164
## [88] fontquiver_0.2.1
                                  progress_1.2.3
                                                            filelock_1.0.3
## [91] KernSmooth_2.23-22
                                  colorspace_2.1-0
                                                            rrvgo_1.16.0
## [94] DBI_1.2.2
                                                            bit_4.0.5
                                  tidyselect_1.2.1
## [97] compiler 4.4.0
                                  curl 5.2.1
                                                            httr2 1.0.1
## [100] BiasedUrn_2.0.11
                                  flextable_0.9.6
                                                            NLP_0.2-1
## [103] xml2 1.3.6
                                  fontBitstreamVera 0.1.1
                                                            DelayedArray_0.30.0
## [106] shadowtext_0.1.3
                                  rtracklayer_1.64.0
                                                            scales_1.3.0
## [109] caTools 1.18.2
                                  rappdirs_0.3.3
                                                            digest 0.6.35
## [112] rmarkdown_2.26
                                  XVector_0.44.0
                                                            htmltools_0.5.8.1
## [115] pkgconfig 2.0.3
                                  umap 0.2.10.0
                                                            highr 0.10
## [118] dbplyr 2.5.0
                                  fastmap_1.1.1
                                                            rlang_1.1.3
## [121] UCSC.utils_1.0.0
                                  shiny_1.8.1.1
                                                            farver 2.1.1
## [124] jsonlite_1.8.8
                                  BiocParallel_1.38.0
                                                            GOSemSim_2.30.0
## [127] RCurl_1.98-1.14
                                  magrittr_2.0.3
                                                            GenomeInfoDbData_1.2.12
## [130] ggplotify_0.1.2
                                  wordcloud_2.6
                                                            patchwork_1.2.0
## [133] munsell_0.5.1
                                  Rcpp_1.0.12
                                                            reticulate_1.37.0
## [136] ape_5.8
                                  viridis_0.6.5
                                                            gdtools_0.3.7
## [139] stringi_1.8.4
                                  ggraph_2.2.1
                                                            zlibbioc_1.50.0
## [142] MASS_7.3-60.2
                                  plyr_1.8.9
                                                            parallel_4.4.0
## [145] ggrepel_0.9.5
                                  Biostrings_2.72.0
                                                            graphlayouts_1.1.1
## [148] splines 4.4.0
                                  gridtext_0.1.5
                                                            hms 1.1.3
## [151] geneLenDataBase_1.40.1
                                  ekbSeq_0.0.3
                                                            locfit_1.5-9.9
## [154] igraph 2.0.3
                                  uuid 1.2-0
                                                            markdown 1.12
## [157] reshape2_1.4.4
                                  biomaRt_2.60.0
                                                            crul_1.4.2
## [160] XML_3.99-0.16.1
                                  evaluate_0.23
                                                            tzdb_0.4.0
## [163] foreach_1.5.2
                                  tweenr_2.0.3
                                                            httpuv_1.6.15
## [166] openssl 2.1.2
                                                            gridBase 0.4-7
                                  polyclip 1.10-6
## [169] ggforce 0.4.2
                                  xtable_1.8-4
                                                            restfulr 0.0.15
## [172] RSpectra_0.16-1
                                  tidytree_0.4.6
                                                            later 1.3.2
## [175] viridisLite_0.4.2
                                  ragg_1.3.1
                                                            aplot_0.2.2
## [178] memoise_2.0.1
                                  AnnotationDbi_1.66.0
                                                            GenomicAlignments_1.40.0
## [181] timechange_0.3.0
```