

Next-generation sequencing - analysis with DESeq2

Comparison of gene expression in DSCs after irradiation - functional analysis

Marc Bender

2024-07-04

Contents

Resources and introduction	2
Pathway analysis (Wikipathways)	3
Appenix	11
Session Info	11

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.

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. In total 16888 genes were expressed above a threshold of 10 counts in at least 5 samples. Out of these 2148 did not have a documented gene symbol, leaving 14740 genes as input for the pathfindR workflow. Out of `nrow(sigRes)` differentially expressed genes, 485 genes were kept in the workflow. xxx (have to be manually curated from the pathfindR output) genes did not have any documented protein interaction in the PIN (protein interaction network), leaving xxx (have to be manually curated from the pathfindR output) genes as final input for the pathfindR enrichment.

(For the publication: pathfindR messages with number of genes into results.)

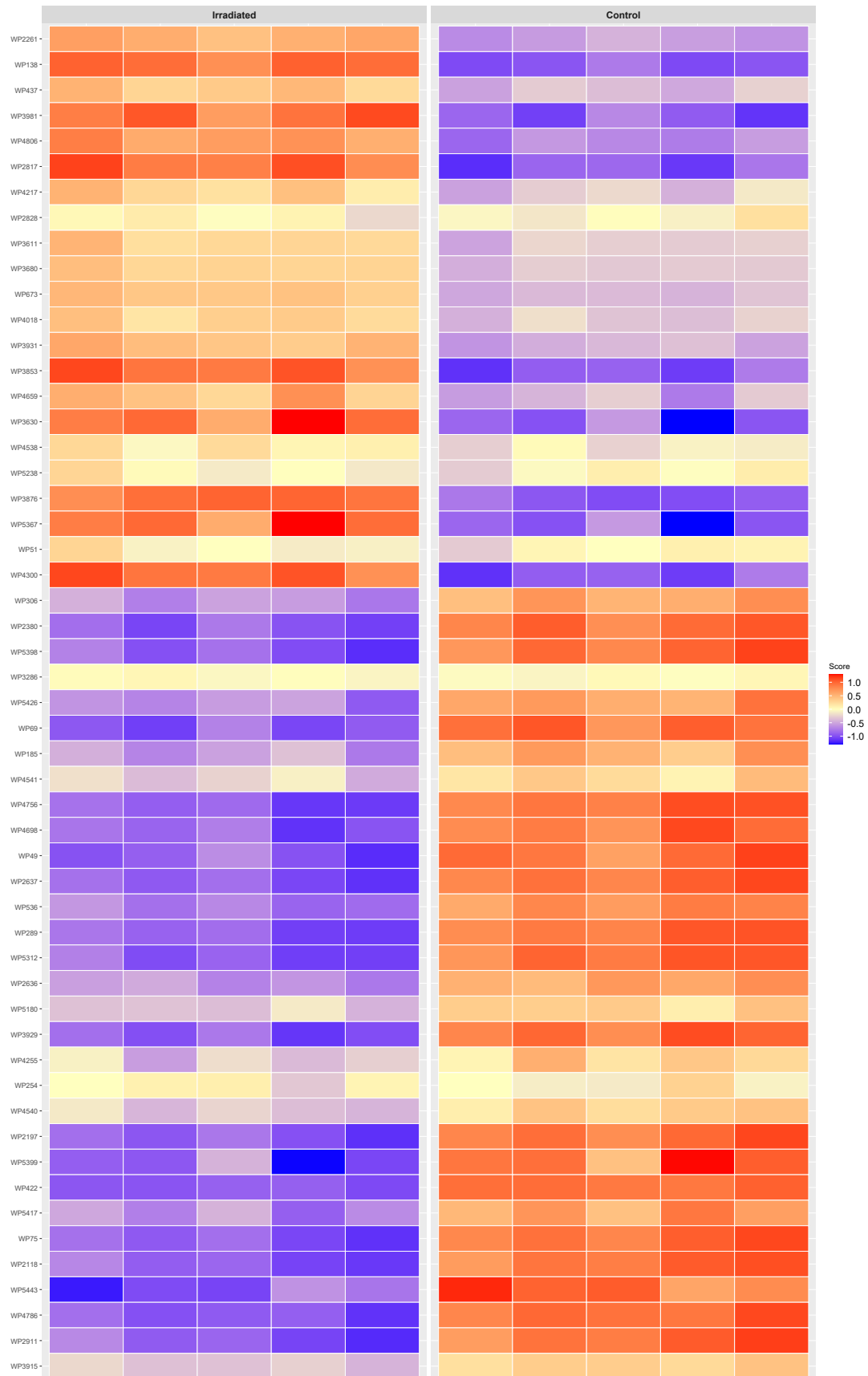


Figure 1: Pathway scores between irradiated and unirradiated DSCs. All pathways.

In total 53 pathways were significantly altered. This high amount of regulated pathways impedes interpretation 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.

Activated

Cholestasis
Glioblastoma signaling pathways

Repressed

Renin angiotensin aldosterone system RAAS
Metabolic pathways of fibroblasts
Interleukin 1 IL 1 structural pathway
Focal adhesion
NRXN1 deletion syndrome
Copper homeostasis
Brain derived neurotrophic factor BDNF signaling pathway
Hippo Merlin signaling dysregulation
Cell lineage map for neuronal differentiation
Endothelin pathways
Non small cell lung cancer

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.

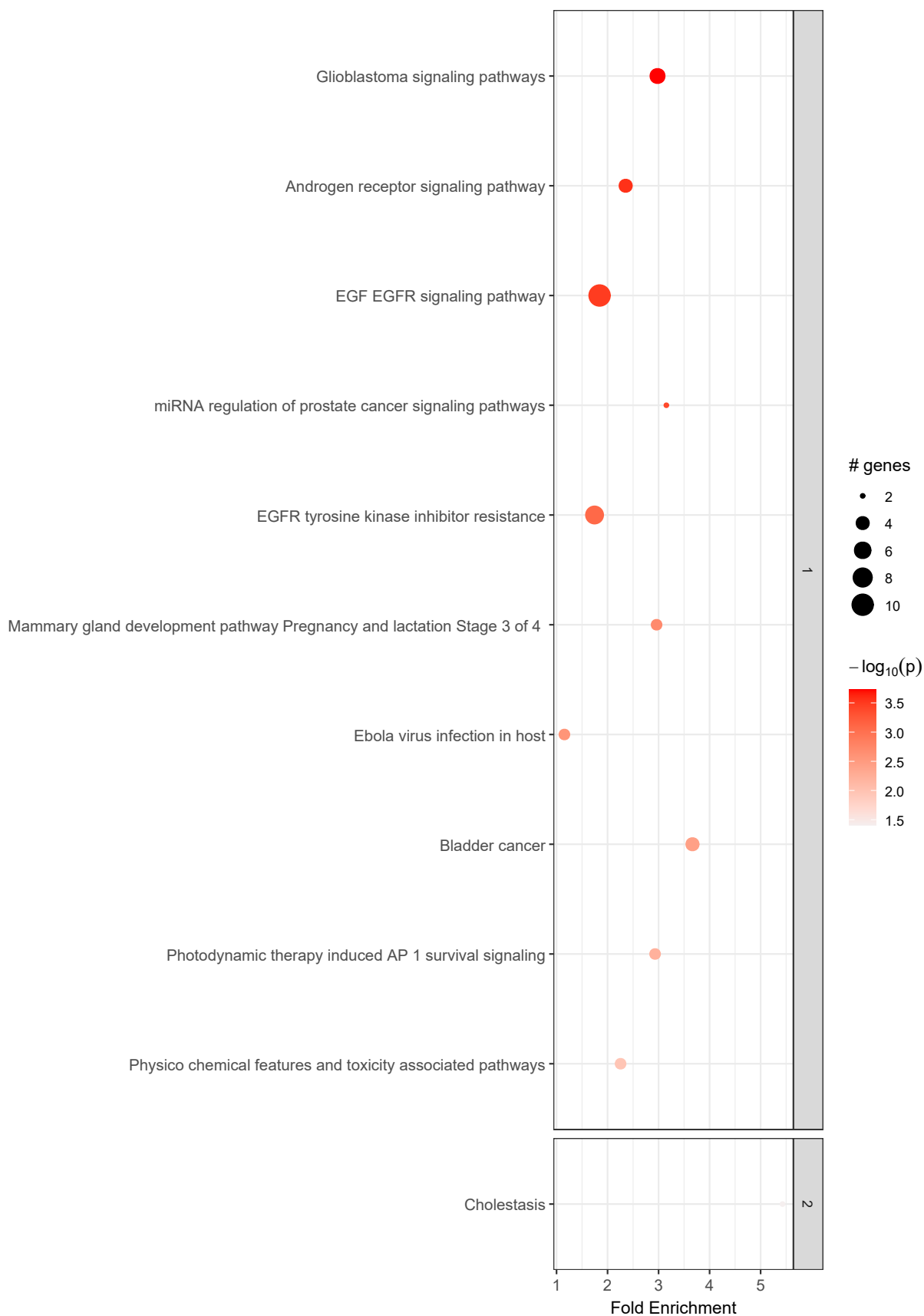


Figure 3: Activated pathways clustered by amount of shared genes.

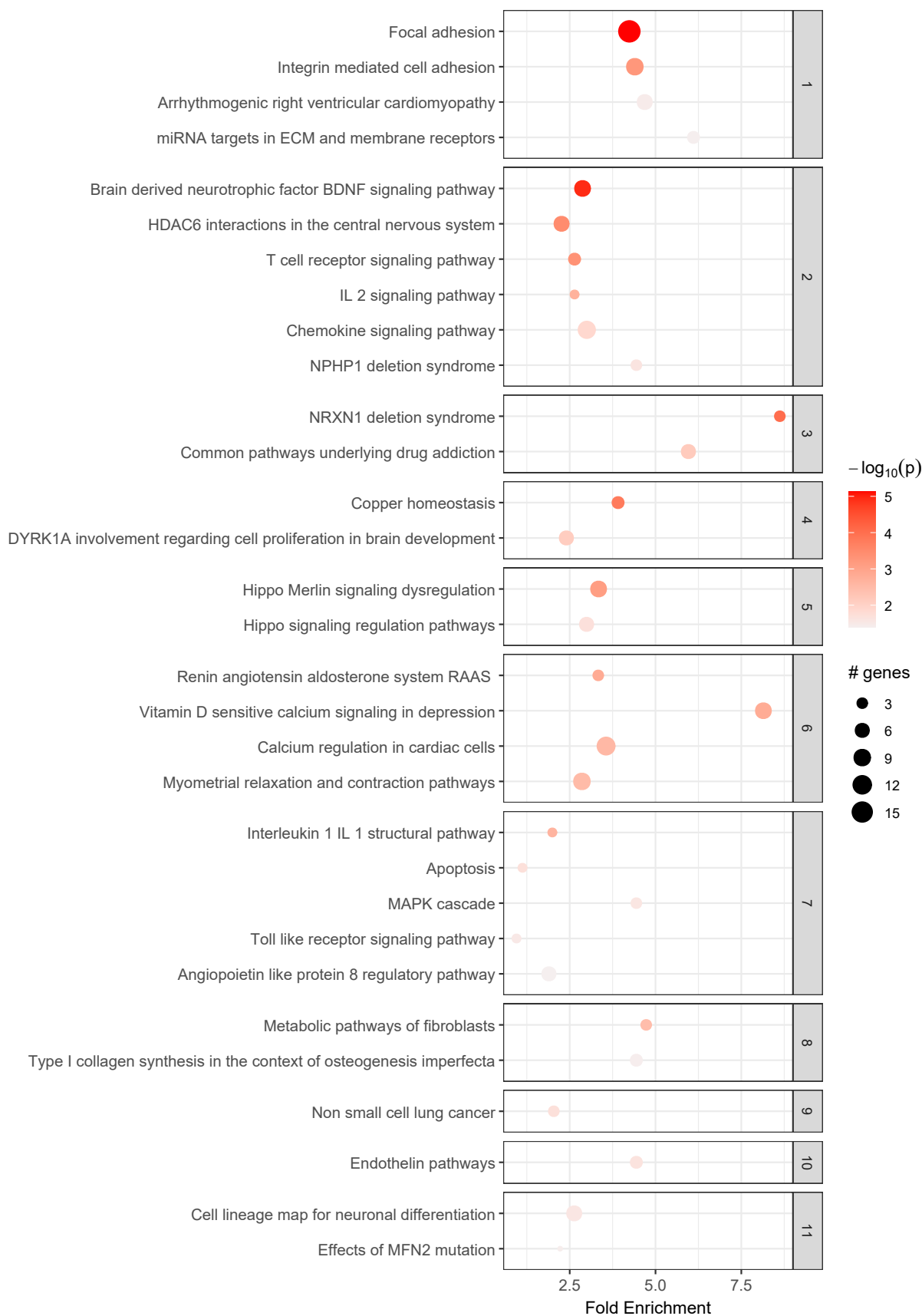


Figure 4: Repressed pathways clustered by amount of shared genes.

Appendix

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       rrvgo_1.16.0
## [5] ekbSeq_0.0.4               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] GetoptLong_1.0.5           withr_3.0.0               prettyunits_1.2.0
## [16] gridExtra_2.3              cli_3.6.2                 textshaping_0.3.7
## [19] scatterpie_0.2.2           officer_0.6.6             labeling_0.4.3
## [22] slam_0.1-50                tm_0.7-13                 goseq_1.56.0
## [25] askpass_1.2.0              commonmark_1.9.1          Rsamtools_2.20.0
## [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] rstudioapi_0.16.0          RSQLite_2.3.6             httpcode_0.3.0
## [37] treemap_2.4-4              shape_1.4.6.1             generics_0.1.3
## [40] gridGraphics_0.5-1         BiocIO_1.14.0             zip_2.3.1
## [43] GO.db_3.19.1               Matrix_1.7-0              fansi_1.0.6
```

## [46]	abind_1.4-5	lifecycle_1.0.4	yaml_2.3.8
## [49]	qvalue_2.36.0	SparseArray_1.4.1	BiocFileCache_2.12.0
## [52]	grid_4.4.0	blob_1.2.4	promises_1.3.0
## [55]	crayon_1.5.2	lattice_0.22-6	cowplot_1.1.3
## [58]	GenomicFeatures_1.56.0	KEGGREST_1.44.0	ComplexHeatmap_2.20.0
## [61]	pillar_1.9.0	knitr_1.46	fgsea_1.30.0
## [64]	rjson_0.2.21	codetools_0.2-20	fastmatch_1.1-4
## [67]	glue_1.7.0	ggfun_0.1.4	fontLiberation_0.1.0
## [70]	data.table_1.15.4	vctrs_0.6.5	png_0.1-8
## [73]	treeio_1.28.0	gtable_0.3.5	cachem_1.0.8
## [76]	xfun_0.43	S4Arrays_1.4.0	mime_0.12
## [79]	tidygraph_1.3.1	pheatmap_1.0.12	iterators_1.0.14
## [82]	tinytex_0.51	nlme_3.1-164	ggtree_3.12.0
## [85]	bit64_4.0.5	fontquiver_0.2.1	progress_1.2.3
## [88]	filelock_1.0.3	colorspace_2.1-0	DBI_1.2.2
## [91]	tidyselect_1.2.1	bit_4.0.5	compiler_4.4.0
## [94]	curl_5.2.1	httr2_1.0.1	BiasedUrn_2.0.11
## [97]	flextable_0.9.6	xml2_1.3.6	NLP_0.2-1
## [100]	fontBitstreamVera_0.1.1	DelayedArray_0.30.0	shadowtext_0.1.3
## [103]	rtracklayer_1.64.0	scales_1.3.0	rappdirs_0.3.3
## [106]	digest_0.6.35	rmarkdown_2.26	XVector_0.44.0
## [109]	htmltools_0.5.8.1	pkgconfig_2.0.3	umap_0.2.10.0
## [112]	highr_0.10	dbplyr_2.5.0	fastmap_1.1.1
## [115]	GlobalOptions_0.1.2	rlang_1.1.3	UCSC.utils_1.0.0
## [118]	shiny_1.8.1.1	farver_2.1.1	jsonlite_1.8.8
## [121]	BiocParallel_1.38.0	GOSemSim_2.30.0	RCurl_1.98-1.14
## [124]	magrittr_2.0.3	GenomeInfoDbData_1.2.12	ggplotify_0.1.2
## [127]	wordcloud_2.6	patchwork_1.2.0	munsell_0.5.1
## [130]	Rcpp_1.0.12	ape_5.8	viridis_0.6.5
## [133]	gdtools_0.3.7	reticulate_1.37.0	stringi_1.8.4
## [136]	ggraph_2.2.1	zlibbioc_1.50.0	MASS_7.3-60.2
## [139]	plyr_1.8.9	parallel_4.4.0	ggrepel_0.9.5
## [142]	Biostrings_2.72.0	graphlayouts_1.1.1	splines_4.4.0
## [145]	gridtext_0.1.5	circlize_0.4.16	hms_1.1.3
## [148]	geneLenDataBase_1.40.1	locfit_1.5-9.9	igraph_2.0.3
## [151]	uuid_1.2-0	markdown_1.12	reshape2_1.4.4
## [154]	biomaRt_2.60.0	crul_1.4.2	XML_3.99-0.16.1
## [157]	evaluate_0.23	tzdb_0.4.0	foreach_1.5.2
## [160]	tweenr_2.0.3	httpuv_1.6.15	openssl_2.1.2
## [163]	polyclip_1.10-6	clue_0.3-65	gridBase_0.4-7
## [166]	ggforce_0.4.2	xtable_1.8-4	restfulr_0.0.15
## [169]	RSpectra_0.16-1	tidytrees_0.4.6	later_1.3.2
## [172]	viridisLite_0.4.2	ragg_1.3.1	clusterProfiler_4.12.0
## [175]	aplot_0.2.2	memoise_2.0.1	AnnotationDbi_1.66.0
## [178]	GenomicAlignments_1.40.0	cluster_2.1.6	timechange_0.3.0