The Innovation, Volume *2*

Supplemental Information

clusterProfiler 4.0: A universal enrichment tool for interpreting omics data

Tianzhi Wu, Erqiang Hu, Shuangbin Xu, Meijun Chen, Pingfan Guo, Zehan Dai, Tingze Feng, Lang Zhou, Wenli Tang, Li Zhan, Xiaocong Fu, Shanshan Liu, Xiaochen Bo, and Guangchuang Yu

Supplemental File of

clusterProfiler 4.0: A universal enrichment tool for interpreting omics data

Tianzhi Wu[§], Erqiang Hu[§], Shuangbin Xu, Meijun Chen, Pingfan Guo, Zehan Dai, Tingze Feng, Lang Zhou, Wenli Tang, Li Zhan, Xiaocong Fu, Shanshan Liu, Xiaochen Bo[*] and Guangchuang Yu[*]

[*]correspondence: Guangchuang Yu <gcyu1@smu.edu.cn> and Xiaochen Bo <boxc@bmi.ac.cn>

1 Installation

To install `clusterProfiler` package, please enter the following command in R:

```
if (!requireNamespace("BiocManager", quietly = TRUE))
    install.packages("BiocManager")
BiocManager::install("clusterProfiler")
```

To reproduce examples in this document, you need to install several extra packages:

```
install.packages(c("forcats", "ggplot2", "ggnewscale", "ggupset"))
BiocManager::install(c("org.Hs.eg.db", "enrichplot",
            "ChIPseeker", "TxDb.Hsapiens.UCSC.hg19.knownGene"))
```

2 Docker image

To help users to build the computing environment, we also provided a docker image[1]. Users can pull and run it according to the following commands. They don't need to install the dependency packages.

1. Install Docker (https://www.docker.com/). For example:

```
# Terminal of Ubuntu
sudo apt-get install docker.io
```

2. Pull the Docker image from Docker Hub:

```
# Terminal of Ubuntu
sudo docker pull xushuangbin/clusterprofilerdocker:latest
```

3. Run the image:

```
# Terminal of Ubuntu
sudo docker run -e PASSWORD=yourpassword -p 8787:8787 xushuangbin/clusterprofilerdocker
```

4. Log in to RStudio at `http://localhost:8787` using username `rstudio` and password `yourpassword`. For Windows users, you also need to provide your IP address, you can find it using `docker-machine ip default`. Inside the RStudio, you can run the examples provided in this document.

Besides, the clusterProfiler package can be installed in virtual environment using `conda`, see also https://anaconda.org/bioconda/bioconductor-clusterprofiler.

3 Bioinformatics tools that depends on clusterProfiler

The `clusterProfiler` library is one of the fundamental packages and it had been incorporated in more than thirty R packages (in CRAN or Bioconductor) to perform functional enrichment analysis for different topics, especially for cancer research (Table S1).

```
db <- utils::available.packages(repo=BiocManager::repositories())
pkgs <- tools::package_dependencies('clusterProfiler', db=db,
              which = c("Depends", "Imports"), reverse=TRUE)[[1]]
sort(pkgs)
```

```
##  [1] "AutoPipe"        "bioCancer"       "CEMiTool"        "CeTF"
##  [5] "conclus"         "DAPAR"           "debrowser"       "eegc"
##  [9] "enrichTF"        "esATAC"          "ExpHunterSuite"  "famat"
```

[1]https://hub.docker.com/r/xushuangbin/clusterprofilerdocker

```
## [13] "fcoex"           "GDCRNATools"     "immcp"           "IRISFGM"
## [17] "maEndToEnd"      "MAGeCKFlute"     "methylGSA"       "miRspongeR"
## [21] "MoonlightR"      "multiSight"      "netboxr"         "PFP"
## [25] "recountWorkflow" "RNASeqR"         "RVA"             "signatureSearch"
## [29] "TCGAbiolinksGUI" "TCGAWorkflow"    "TimiRGeN"
```

Table S1: R packages that rely on clusterProfiler to perform functional analysis.

Package	Description
AutoPipe	Automated Transcriptome Classifier Pipeline: Comprehensive Transcriptome Analysis
bioCancer	Interactive Multi-Omics Cancers Data Visualization and Analysis
CEMiTool	Co-expression Modules identification Tool
CeTF	Coexpression for Transcription Factors using Regulatory Impact Factors and Partial Correlation and Information Theory analysis
conclus	ScRNA-seq Workflow CONCLUS - From CONsensus CLUSters To A Meaningful CONCLUSion
DAPAR	Tools for the Differential Analysis of Proteins Abundance with R
debrowser	Interactive Differential Expresion Analysis Browser
eegc	Engineering Evaluation by Gene Categorization (eegc)
enrichTF	Transcription Factors Enrichment Analysis
esATAC	An Easy-to-use Systematic pipeline for ATACseq data analysis
ExpHunterSuite	Package For The Comprehensive Analysis Of Transcriptomic Data
famat	Functional analysis of metabolic and transcriptomic data
fcoex	FCBF-based Co-Expression Networks for Single Cells
GDCRNATools	an R/Bioconductor package for integrative analysis of lncRNA, mRNA, and miRNA data in GDC
immcp	Candidate Prescriptions Discovery Based on Pathway Fingerprint
IRISFGM	Comprehensive Analysis of Gene Interactivity Networks Based on Single-Cell RNA-Seq
maEndToEnd	An end to end workflow for differential gene expression using Affymetrix microarrays
MAGeCKFlute	Integrative Analysis Pipeline for Pooled CRISPR Functional Genetic Screens
methylGSA	Gene Set Analysis Using the Outcome of Differential Methylation
miRspongeR	Identification and analysis of miRNA sponge interaction networks and modules
MoonlightR	Identify oncogenes and tumor suppressor genes from omics data
multiSight	Multi-omics Classification, Functional Enrichment and Network Inference analysis
netboxr	netboxr
PFP	Pathway Fingerprint Framework in R
recountWorkflow	recount workflow: accessing over 70,000 human RNA-seq samples with Bioconductor
RNASeqR	an R package for automated two-group RNA-Seq analysis workflow
RVA	RNAseq Visualization Automation
signatureSearch	Environment for Gene Expression Searching Combined with Functional Enrichment Analysis
TCGAbiolinksGUI	TCGAbiolinksGUI: A Graphical User Interface to analyze cancer molecular and clinical data
TCGAWorkflow	TCGA Workflow Analyze cancer genomics and epigenomics data using Bioconductor packages
TimiRGeN	Time sensitive microRNA-mRNA integration, analysis and network generation tool

Moreover, clusterProfiler has been incorporated into different workflows and analysis websites (including shiny apps).

Workflows that incorporates `clusterProfiler`:

- TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages[2]
- Microbe-Flow: a comprehensive workflow for bacterial genomics, pathogenomics and genomic epidemiology[3]
- ViralLink: An integrated workflow to investigate the effect of SARS-CoV-2 on intracellular signalling and regulatory pathways[4]
- Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute[5]
- MUSIC: Model-based Understanding of SIngle-cell CRISPR screening[6]
- An end to end workflow for differential gene expression using Affymetrix microarrays[7]
- recount workflow: Accessing over 70,000 human RNA-seq samples with Bioconductor[8]
- RNAseq workflow[9]
- RNAseq Analysis[10]

[2]https://f1000research.com/articles/5-1542
[3]https://neatseq-flow.readthedocs.io/projects/neatseq-flow-modules/en/latest/Workflow_docs/Microbe-Flow.html
[4]https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008685
[5]https://www.nature.com/articles/s41596-018-0113-7
[6]https://github.com/bm2-lab/MUSIC
[7]https://f1000research.com/articles/5-1384/v2
[8]https://f1000research.com/articles/6-1558
[9]https://github.com/twbattaglia/RNAseq-workflow
[10]https://learn.gencore.bio.nyu.edu/rna-seq-analysis/gene-set-enrichment-analysis/

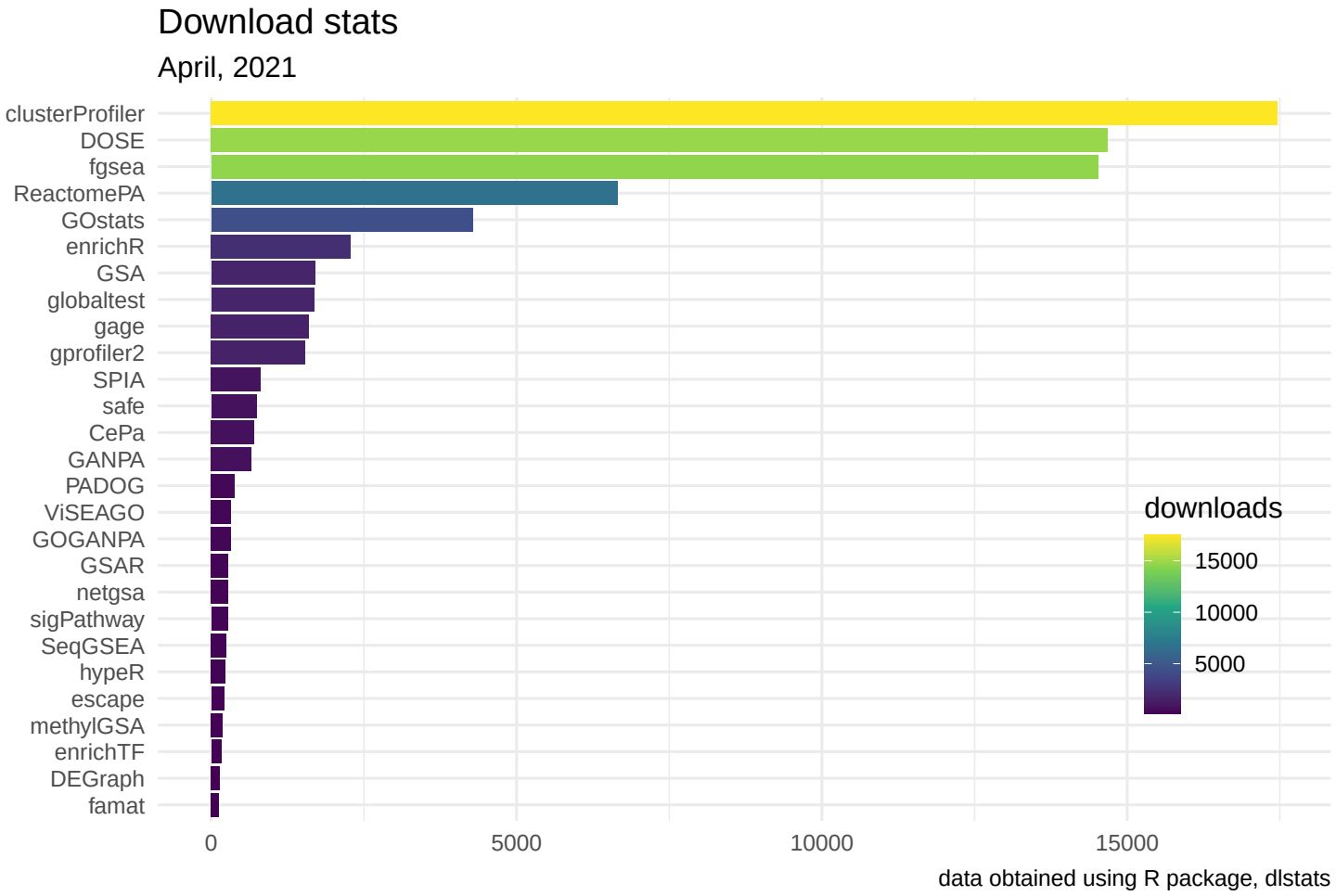
- Automated transcriptomics data analysis workflow using pathway and network analysis approaches[11]

Analysis websites (or shiny apps) that incorporates `clusterProfiler`:

- NASQAR: a web-based platform for high-throughput sequencing data analysis and visualization[12]
- Shiny-Seq: advanced guided transcriptome analysis[13]
- ProteoRE: A biologist-oriented Galaxy platform for proteomics data exploration[14]
- Netpredictor: R and Shiny package to perform drug-target network analysis and prediction of missing links[15]
- ABioTrans: A Biostatistical Tool for Transcriptomics Analysis[16]
- SigBio-Shiny: A standalone interactive application for detecting biological significance on a set of genes[17]

4 Comparing clusterProfiler with other tools

Here, we compare `clusterProfiler` with other R packages that also can perform functional enrichment analysis (Table S2). The packages in Table S2 were ordered by monthly download stats (April 2021).



Focus on the R ecosystem, `clusterProfiler` is the most popular package for functional enrichment analysis. Compare to other tools, `clusterProfiler` has many good features. It internally supports GO and KEGG for thousands of species, allows users to specify background gene set, provides general interface for external annotation data, works with GMT files, and supports comparing functional profiles among different conditions.

Several R packages ouput tabular result (e.g., data frame). Data frame is simple and easy to process and visualize using tidy tools (e.g., dplyr) and ggplot2. However, many useful information including input data, parameter setting and gene set, are missing. These information maybe useful for further interpretation and visualization. Instead, most of the R packages

[11] https://fairdomhub.org/studies/837
[12] https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-020-03577-4
[13] https://bmcresnotes.biomedcentral.com/articles/10.1186/s13104-019-4471-1
[14] https://github.com/vloux/ProteoRE
[15] https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2254-7
[16] https://www.frontiersin.org/articles/10.3389/fgene.2019.00499/full
[17] https://github.com/sk-sahu/sig-bio-shiny

encapsulate enrichment result into more complicated R object (S3, S4 or R6) to include enrichment result with associated data. This will prevent users to explore the result using tidy tools and ggplot2. The `clusterProfiler` and its sub-packages (including `DOSE` and `ReactomePA`) provide tidy interface to process enrichment result and directly supports of visualizing enrichment result in ggplot2. To our knowledge, this feature cannot be found in other R packages that also output enrichment result as complicated R object.

Table S2: Comparing clusterProfiler with other tools

		Input and annotation						Method			Interpretation				
Software	Repo	Annotation	Supported organisms	ID conversion	Updated KEGG	External annotation data	Support GMT file	Algorithm	Selection of background set	Profile comparison	Output	Tidy interface	Support ggplot2	Visualization methods	Remove redundant terms
clusterProfiler	2	GO, KEGG, WikiPathways	plenty	Y	Y	Y	Y	ORA, GSEA	Y	Y	enrichResult, gseaResult, compareClusterResult (S4)	Y	Y	11	Y
DOSE	2	DisGeNE, DO, NCG	1	N	NA	N	N	ORA, GSEA	Y	N	enrichResult, gseaResult (S4)	Y	Y	11	Y
fgsea	2	NA	NA	Y	NA	Y	Y	ORA, GSEA	Y	N	data.table	Y	Y	2	N
ReactomePA	2	Reactome	7	N	NA	N	N	ORA, GSEA	Y	N	enrichResult, gseaResult (S4)	Y	Y	11	N
GOstats	2	NA	NA	N	NA	Y	N	ORA	Y	N	GOHyperGResult (S4)	N	N	1	N
enrichR	1	GO, KEGG, WikiPathways, BioCarta, Reactome, GEO, GeneSigDB, HPO, KEA, MSigDB, COVID-19 Related Gene Sets	5	Y	N	N	N	ORA	N	N	list	N	N	3	N
GSA	1	NA	NA	N	NA	Y	Y	Gene set analysis	N	N	GSA (S3)	N	N	1	N
globaltest	2	GO, KEGG, MSigDB, Anni	21	N	N	N	N	regression analysis	N	N	gt (S4)	N	N	2	N
gage	2	GO, KEGG	plenty	Y	Y	Y	N	GSEA	N	N	list	N	N	0	Y
gprofiler2	1	GO, KEGG, Reactome,WikiPathways, miRTarBase, TRANSFAC, Human Protein Atlas, protein complexes from CORUM, HPO	plenty	Y	N	Y	Y	ORA	N	Y	list	N	N	1	N
SPIA	2	KEGG	plenty	N	Y	N	N	Signaling Pathway Impact Analysis	Y	N	data.frame	Y	Y	1	N
safe	2	GO, KEGG, PFAM, Reactome	20	N	N	Y	N	ORA, Wilcoxon rank sum, Pearson's chi-squared type statistic, t-statistic	N	N	SAFE (S4)	N	N	2	N
CePa	1	NCI_Nature, KEGG, BioCarta, Reactome	1	N	N	N	N	CePa	Y	N	cepa (S3)	N	N	3	N
GANPA	1	NA	NA	N	NA	Y	N	GANPA	N	N	.csv files	N	N	0	N
PADOG	2	KEGG	1	N	Y	Y	N	PADOG	N	Y	data.frame	Y	Y	0	N
ViSEAGO	2	GO	21	N	N	N	N	ORA, GSEA	N	Y	fgsea, enrich_GO_terms (S4)	N	N	2	N
GOGANPA	1	NA	NA	N	NA	Y	N	GO-Functional-Network-based Gene-Set-Analysis	N	N	.csv file	N	N	0	N
GSAR	2	NA	NA	N	NA	Y	N	two-sample Nnparametric multivariate test	N	N	list	N	N	0	N
netgsa	1	NA	NA	N	NA	Y	N	netgsa	N	N	list	N	N	3	N
sigPathway	2	NA	NA	N	NA	Y	N	GSEA, sigPathway	N	N	list	N	N	0	N
SeqGSEA	2	NA	NA	Y	NA	Y	Y	GSEA	N	N	SeqGeneSet (S4)	N	N	0	N
hypeR	2	MSigDB, KEGG, Reactome, MetaboAnalyst	11	N	N	Y	N	ORA, GSEA	Y	N	hyp (R6)	N	N	3	N
escape	2	MSigDB	11	N	N	Y	N	GSEA	N	N	data.frame	N	N	6	N
methylGSA	2	GO, KEGG, Reactome	1	Y	N	N	N	ORA, GSEA	N	N	data.frame	Y	Y	0	N
enrichTF	2	Transcription factor information	2	N	NA	N	N	t-tests, ORA	N	N	list	N	N	0	N
DEGraph	2	KEGG	plenty	N	Y	Y	N	t-tests	N	N	list	N	N	1	N
famat	2	GO, KEGG, Wikipathways, Reactome	1	N	Y	N	N	ORA	N	N	list	N	N	0	N

[1] Repo: 1 for CRAN and 2 for Bioconductor

[2] Supported organisms: 'NA' for not applicable as there is no species annotation data internally supported by the package; 'plenty' for hundreds or thousands species supported (mostly for KEGG and/or GO)

[3] Tidy interface: whether the output object can be processed directly using tidy tools such as dplyr

[4] Support ggplot2: whether the output object can be visualized directly using ggplot2 command

[5] Y for supported, N for not supported and NA for not applicable

5 Data sets

Three data sets were used in this document, including:

- `geneList` provided by the DOSE package
- `DE_GSE8057` provided by the clusterProfiler package
- `GSM1295076_CBX6_BF_ChipSeq_mergedReps_peaks.bed.gz` provided by the ChIPseeker package

The `geneList` was derived from the R package breastCancerMAINZ that contains 200 breast cancer samples, including 29 samples in grade I, 136 samples in grade II and 35 samples in grade III. The ratio of geometric mean of grade III samples versue geometric mean of grade I samples for each gene was computed. The `geneList` data set contains logarithm of these ratios (base 2).

The `DE_GSE8057` data set was derived from the `GSE8057` data set which can be downloaded in GEO and the expreimental design was documented in https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE8057. All the treated samples were compared with control samples by different conditions using the limma package. The `DE_GSE8057` data set contains differential expressed genes (DEGs) for each condition and the DEGs were selected in case of expression values with fold change > 1 or adjusted p value < 0.05.

The GSM1295076_CBX6_BF_ChipSeq_mergedReps_peaks.bed.gz file can be accessed via `ChIPseeker::getSampleFiles()[[4]]` or downloaded using the command `ChIPseeker::downloadGSMbedFiles("GSM1295076")`. The experimental design was documented in https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1295076.

In addition to GO and KEGG, two additional gene sets were used in the manuscript, including:

- ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X
- WikiPathways

The `ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X` was downloaded from https://maayanlab.cloud/Enrichr/geneSetLibrary?mode=text&libraryName=ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X. This gene set was used to identify transcriptional factors associated with genomic regions obtained from a ChIPseq experiment.

The WikiPathways, `wikipathways-20210310-gmt-Homo_sapiens.gmt` was downloaded from https://wikipathways-data.wmcloud.org/current/gmt/. This gene set was used to identify biological pathways using community curated knowledge.

6 Examples of using clusterProfiler

This session provides source codes to reproduce the figures presented in the manuscript.

6.1 GO enrichment analysis

```
library(clusterProfiler)
library(enrichplot)

## geneList for GSEA examples
data(geneList, package="DOSE")

## fold change > 2 as DE genes, for ORA examples
de <- names(geneList)[abs(geneList) > 2]


ego <- enrichGO(de, OrgDb = "org.Hs.eg.db", ont="BP", readable=TRUE)

## use simplify to remove redundant terms
ego2 <- simplify(ego, cutoff=0.7, by="p.adjust", select_fun=min)


## visualization
ego <- pairwise_termsim(ego)
ego2 <- pairwise_termsim(ego2)

p1 <- emapplot(ego, cex_label_category=.8, cex_line=.5) + coord_cartesian()
p2 <- emapplot(ego2, cex_label_category=.8, cex_line=.5) + coord_cartesian()

p1 <- p1 + scale_fill_continuous(low = "#e06663", high = "#327eba", name = "p.adjust",
                                 guide = guide_colorbar(reverse = TRUE, order=1), trans='log10')
p2 <- p2 + scale_fill_continuous(low = "#e06663", high = "#327eba", name = "p.adjust",
```

```
                    guide = guide_colorbar(reverse = TRUE, order=1), trans='log10')

cowplot::plot_grid(p1, p2, labels=c("A", "B"), rel_widths=c(1, 1.2))
```

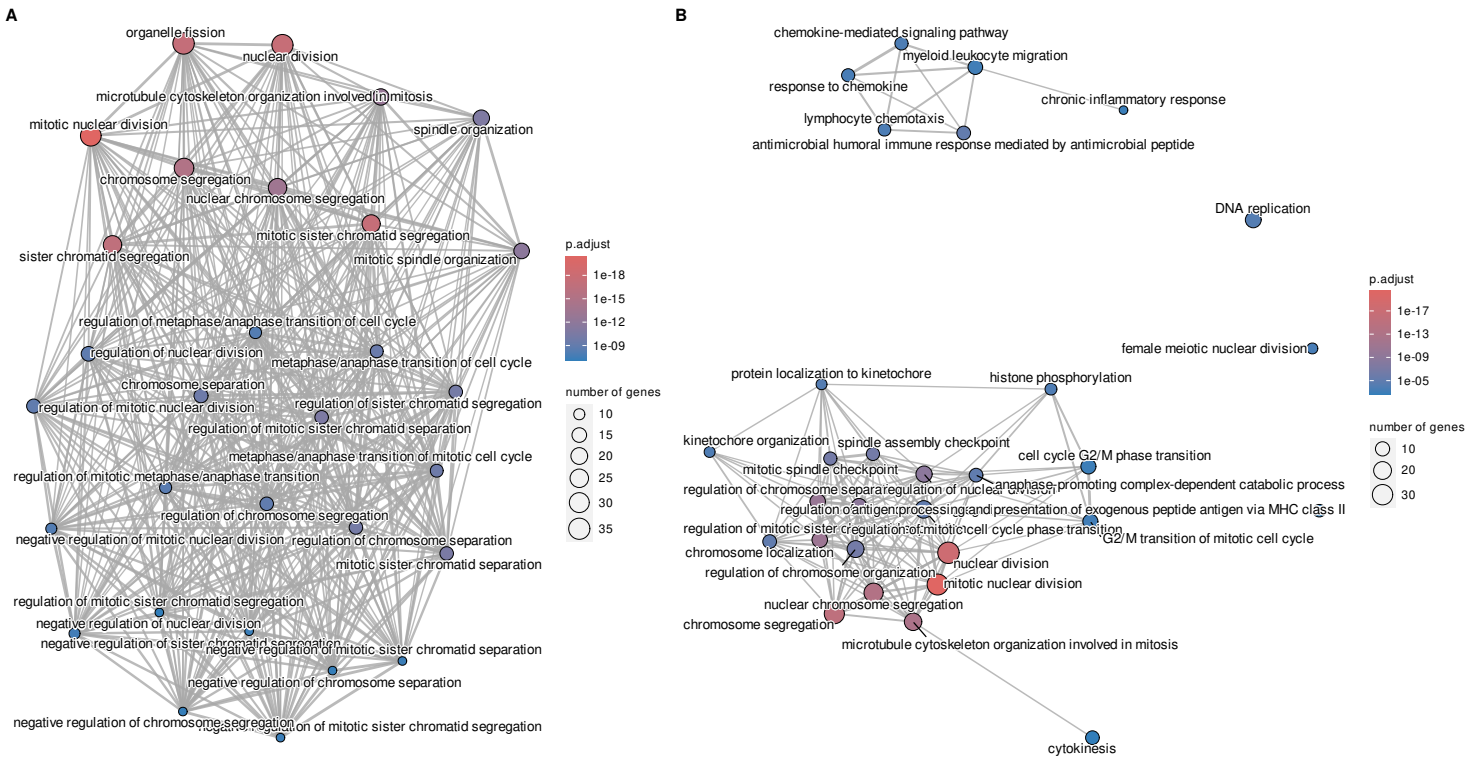
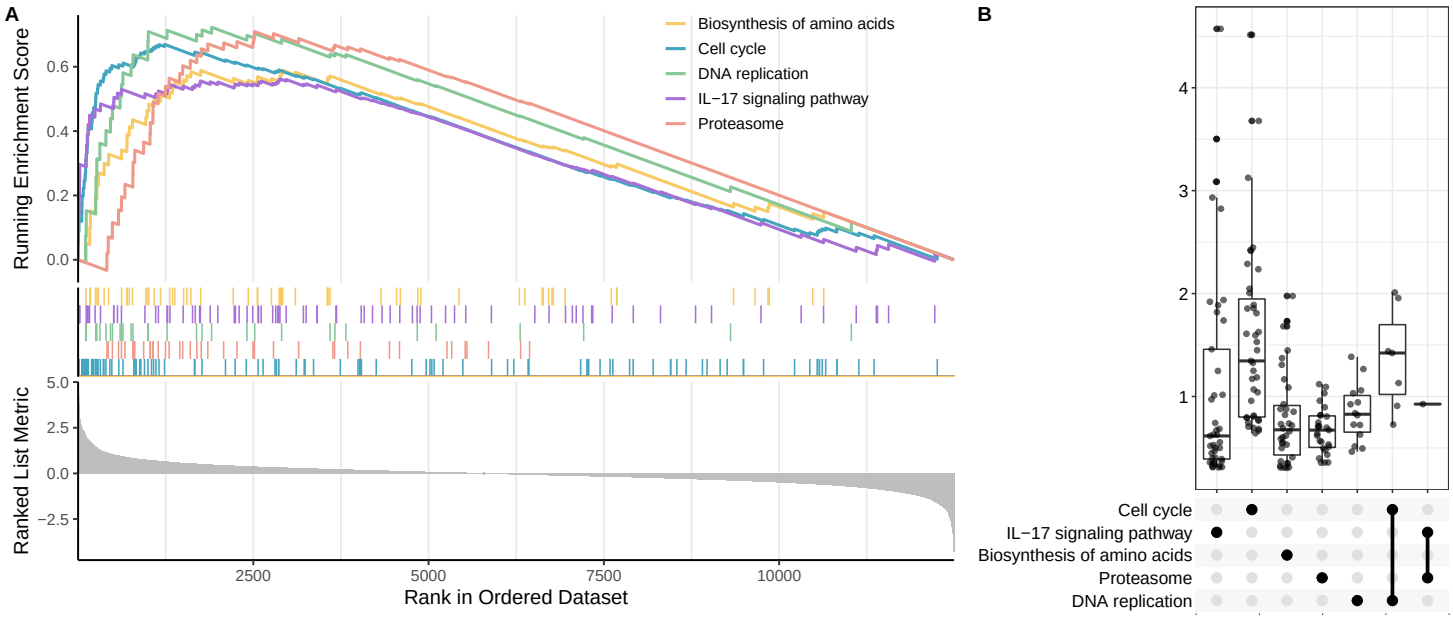


Fig. 1: Gene ontology enrichment analysis.

6.2 KEGG enrichment analysis

```
data(geneList, package="DOSE")
kk <- gseKEGG(geneList, organism = "hsa", eps=0)

## sorted by absolute values of NES
kk2 <- arrange(kk, desc(abs(NES)))

## visualization
color <- c("#f7ca64", "#43a5bf", "#86c697", "#a670d6", "#ef998a")
kp1 <- gseaplot2(kk2, 1:5, color = color, pvalue_table=F, base_size=14)
kp2 <- upsetplot(kk2, n=5)
cowplot::plot_grid(kp1, kp2, rel_widths=c(1, .5), labels=c("A", "B"))
```



6.3 Functional interpretation of genomic regions of interest

```
library(ChIPseeker)
## the file can be downloaded using `downloadGSMbedFiles("GSM1295076")`
file <- "GSM1295076_CBX6_BF_ChipSeq_mergedReps_peaks.bed.gz"
gr <- readPeakFile(file)

library(TxDb.Hsapiens.UCSC.hg19.knownGene)
TxDb <- TxDb.Hsapiens.UCSC.hg19.knownGene
genes <- seq2gene(gr, tssRegion=c(-1000, 1000), flankDistance = 3000, TxDb)

library(clusterProfiler)
## downloaded from 'https://maayanlab.cloud/Enrichr/geneSetLibrary?mode=
## text&libraryName=ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X'
encode <- read.gmt("ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X.txt")
g <- bitr(genes, 'ENTREZID', 'SYMBOL', 'org.Hs.eg.db')
```

```
## Warning in bitr(genes, "ENTREZID", "SYMBOL", "org.Hs.eg.db"): 5.32% of input
## gene IDs are fail to map...
```

```
x <- enricher(g$SYMBOL, TERM2GENE=encode)
cnetplot(x, cex_label_gene=0.6,
       color_category = "#97c497", color_gene = "#c4c4c4") +
       guides(size = guide_legend(override.aes=list(shape=1)))
```

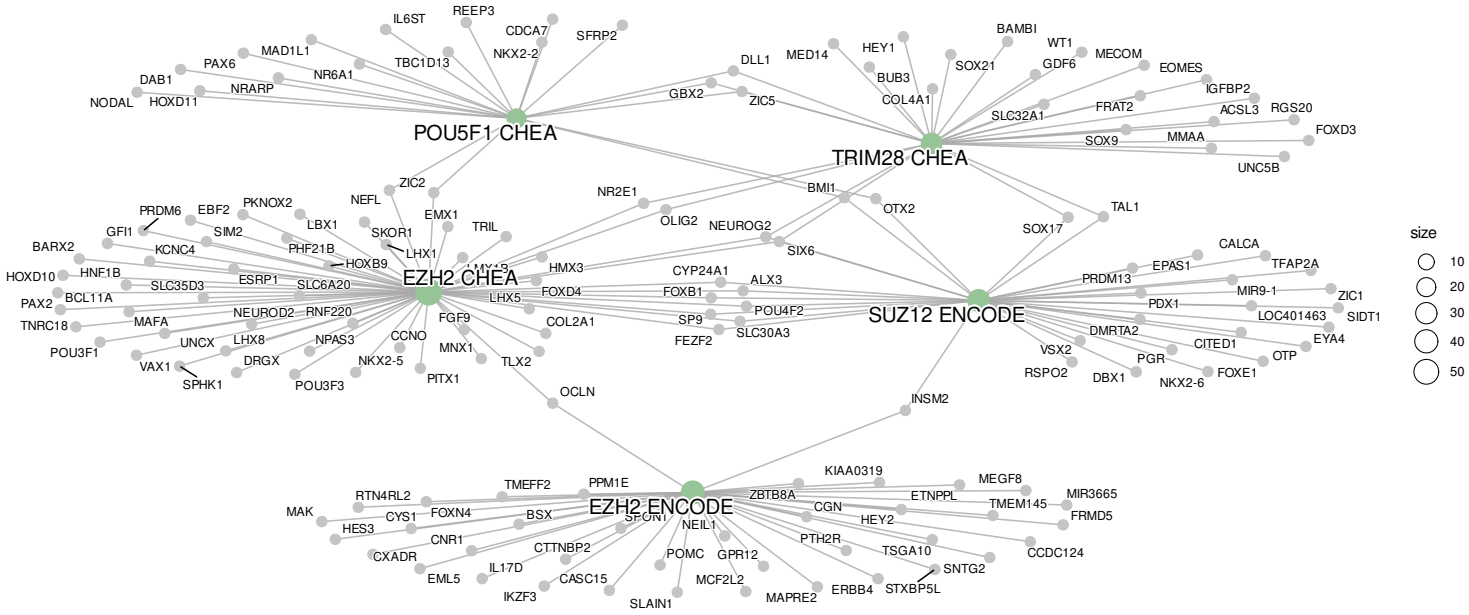


Fig. 2: Functional enrichment analysis of genomic regions of interest.

6.4 Comparison for different conditions

```
## downloaded from https://wikipathways-data.wmcloud.org/current/gmt/
gmt <- 'wikipathways-20210310-gmt-Homo_sapiens.gmt'
wp <- read.gmt.wp(gmt)

data(DE_GSE8057)

xx <- compareCluster(Gene~time+treatment, data=DE_GSE8057, fun = enricher,
              TERM2GENE=wp[,c("wpid", "gene")], TERM2NAME=wp[,c("wpid", "name")])

pp <- dotplot(xx, x="time") + facet_grid(~treatment) +
    aes(x=fct_relevel(time, c('0h', '2h', '6h', '24h'))) + xlab(NULL) +
    scale_color_gradientn(colours=c("#b3eebe", "#46bac2", "#371ea3"),
                    guide=guide_colorbar(reverse=TRUE, order=1)) +
    guides(size = guide_legend(override.aes=list(shape=1))) +
    theme(panel.grid.major.y = element_line(linetype='dotted', color='#808080'),
          panel.grid.major.x = element_blank())
```

```
print(pp)
```

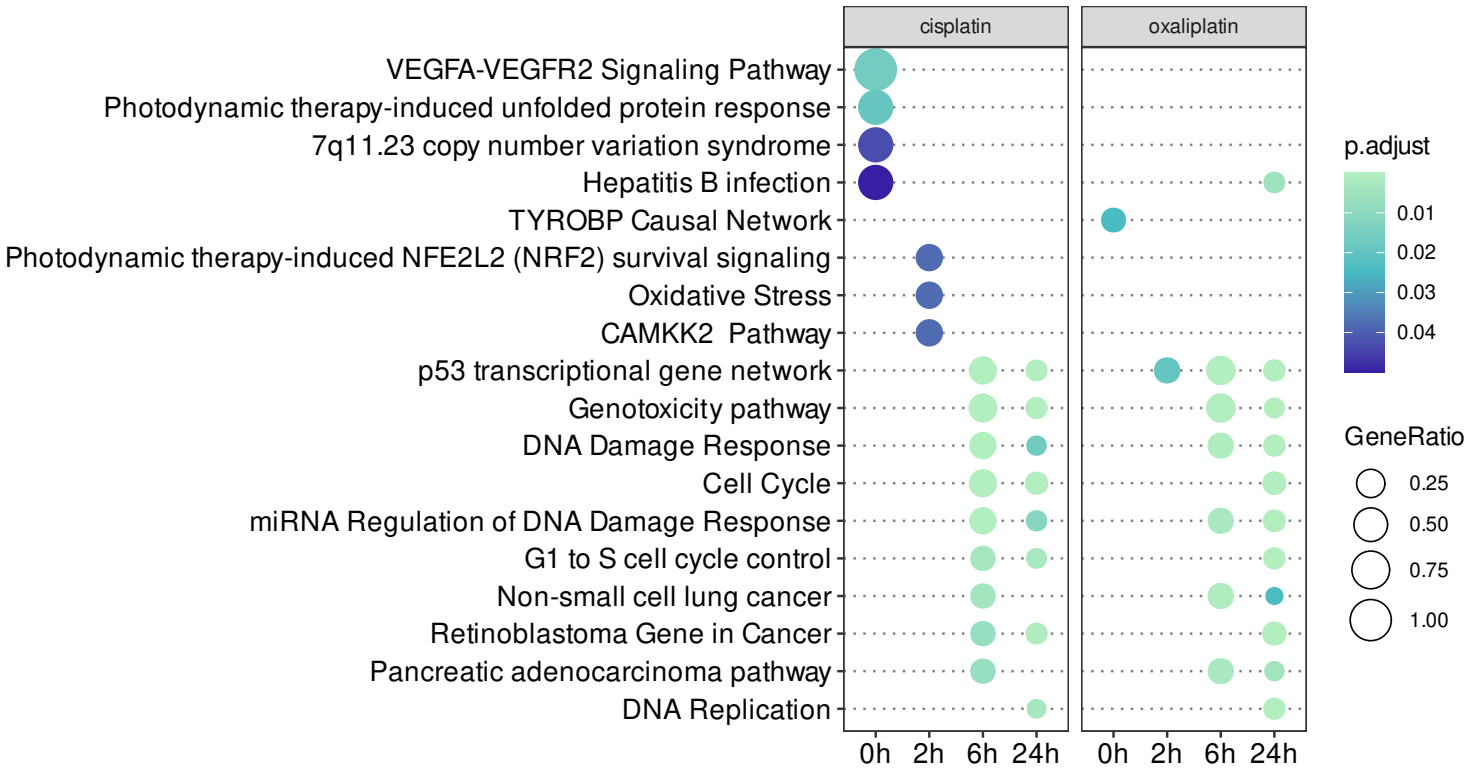
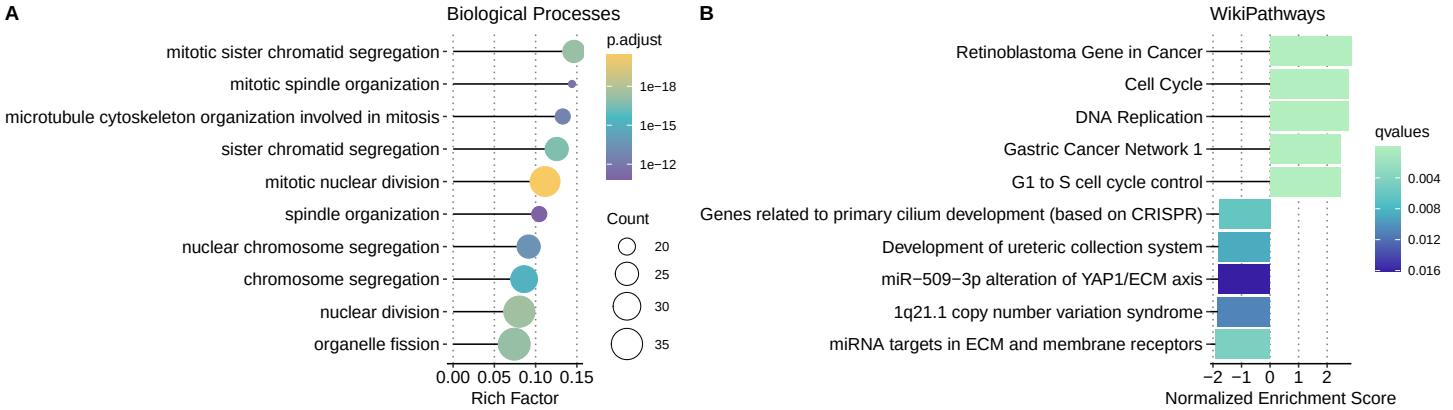


Fig. 3: Comparing functional profiles among different levels of conditions.

6.5 Visualization using ggplot2

```
library(forcats)
library(ggplot2)

ewp <- GSEA(geneList, TERM2GENE=wp[,c("wpid", "gene")], TERM2NAME=wp[,c("wpid", "name")])

ewp2 <- arrange(ewp, desc(abs(NES))) %>%
        group_by(sign(NES)) %>%
        slice(1:5)
ego3 <- mutate(ego, richFactor = Count / as.numeric(sub("/\\d+", "", BgRatio)))

mytheme <- theme(panel.border=element_blank(),
        panel.grid.major=element_line(linetype='dotted', colour='#808080'),
        panel.grid.major.y=element_blank(),
        panel.grid.minor=element_blank(),
        axis.line.x = element_line())

g1 <- ggplot(ego3, showCategory = 10,
  aes(richFactor, fct_reorder(Description, richFactor))) +
  geom_segment(aes(xend=0, yend = Description)) +
  geom_point(aes(color=p.adjust, size = Count)) +
  scale_color_gradientn(colours=c("#f7ca64", "#46bac2", "#7e62a3"),
                        trans = "log10",
                        guide=guide_colorbar(reverse=TRUE, order=1)) +
  scale_size_continuous(range=c(2, 10)) +
  scale_x_continuous(expand=c(0,0)) +
  theme_dose(12) +
  mytheme +
  xlim(NA, 0.15) +
  guides(size = guide_legend(override.aes=list(shape=1))) +
  xlab("Rich Factor") +
  ylab(NULL) +
  ggtitle("Biological Processes")

g2 <- ggplot(ewp2, showCategory=10,
        aes(NES, fct_reorder(Description, NES), fill=qvalues)) +
    geom_col() +
    geom_segment(mapping=aes(x=-2.1,
                             xend=ifelse(sign(NES)>0, 0, NES),
                             yend=Description)) +
    scale_x_continuous(expand=c(0,0)) +
    scale_fill_gradientn(colours=c('#b3eebe', "#46bac2", '#371ea3'),
                         guide=guide_colorbar(reverse=TRUE)) +
    theme_dose(12) +
    mytheme +
    xlab("Normalized Enrichment Score") +
    ylab(NULL) +
    ggtitle("WikiPathways")

cowplot::plot_grid(g1, g2, labels=c("A", "B"))
```



NOTE:

1. source codes and datasets to produce this file can be obtained online[18].
2. setting font and adding rounded rectangular as background for each of the legends, as presented in the manuscript, can be done by the `ggfun`[19] package.

[18] https://github.com/YuLab-SMU/supplemental-clusterProfiler-v4
[19] https://cran.r-project.org/package=ggfun

7 Session information

Here is the output of `sessionInfo()` of the system on which the Supplemental file was compiled:

```
## - Session info ---------------------------------------------------------------
##  setting  value
##  version  R version 4.1.0 (2021-05-18)
##  os       Arch Linux
##  system   x86_64, linux-gnu
##  ui       X11
##  language (EN)
##  collate  en_US.UTF-8
##  ctype    en_US.UTF-8
##  tz       Asia/Chongqing
##  date     2021-07-05
##
## - Packages -------------------------------------------------------------------
##  package                       * version    date       lib source
##  AnnotationDbi                 * 1.54.0     2021-05-19 [1] Bioconductor
##  ape                             5.5        2021-04-25 [1] CRAN (R 4.1.0)
##  aplot                           0.0.6      2020-09-03 [1] CRAN (R 4.1.0)
##  assertthat                      0.2.1      2019-03-21 [1] CRAN (R 4.1.0)
##  Biobase                       * 2.52.0     2021-05-19 [1] Bioconductor
##  BiocFileCache                   2.0.0      2021-05-19 [1] Bioconductor
##  BiocGenerics                  * 0.38.0     2021-05-19 [1] Bioconductor
##  BiocIO                          1.2.0      2021-05-19 [1] Bioconductor
##  BiocManager                     1.30.15    2021-05-11 [1] CRAN (R 4.1.0)
##  BiocParallel                    1.26.0     2021-05-19 [1] Bioconductor
##  biomaRt                         2.48.0     2021-05-19 [1] Bioconductor
##  Biostrings                      2.60.0     2021-05-19 [1] Bioconductor
##  bit                             4.0.4      2020-08-04 [1] CRAN (R 4.1.0)
##  bit64                           4.0.5      2020-08-30 [1] CRAN (R 4.1.0)
##  bitops                          1.0-7      2021-04-24 [1] CRAN (R 4.1.0)
##  blob                            1.2.1      2020-01-20 [1] CRAN (R 4.1.0)
##  bookdown                        0.22       2021-04-22 [1] CRAN (R 4.1.0)
##  boot                            1.3-28     2021-05-03 [1] CRAN (R 4.1.0)
##  cachem                          1.0.5      2021-05-15 [1] CRAN (R 4.1.0)
##  caTools                         1.18.2     2021-03-28 [1] CRAN (R 4.1.0)
##  ChIPseeker                    * 1.28.3     2021-05-21 [1] Bioconductor
##  cli                             2.5.0      2021-04-26 [1] CRAN (R 4.1.0)
##  clusterProfiler               * 4.1.1      2021-07-05 [1] Bioconductor
##  colorspace                      2.0-1      2021-05-04 [1] CRAN (R 4.1.0)
##  conflicted                    * 1.0.4      2019-06-21 [1] CRAN (R 4.1.0)
##  cowplot                         1.1.1      2020-12-30 [1] CRAN (R 4.1.0)
##  crayon                          1.4.1      2021-02-08 [1] CRAN (R 4.1.0)
##  curl                            4.3.1      2021-04-30 [1] CRAN (R 4.1.0)
##  data.table                      1.14.0     2021-02-21 [1] CRAN (R 4.1.0)
##  DBI                             1.1.1      2021-01-15 [1] CRAN (R 4.1.0)
##  dbplyr                          2.1.1      2021-04-06 [1] CRAN (R 4.1.0)
##  DelayedArray                    0.18.0     2021-05-19 [1] Bioconductor
##  digest                          0.6.27     2020-10-24 [1] CRAN (R 4.1.0)
##  DO.db                           2.9        2021-05-22 [1] Bioconductor
##  DOSE                          * 3.19.1     2021-06-13 [1] Bioconductor
##  downloader                      0.4        2015-07-09 [1] CRAN (R 4.1.0)
##  dplyr                           1.0.6      2021-05-05 [1] CRAN (R 4.1.0)
##  ellipsis                        0.3.2      2021-04-29 [1] CRAN (R 4.1.0)
##  enrichplot                    * 1.13.1     2021-06-30 [1] Bioconductor
##  evaluate                        0.14       2019-05-28 [1] CRAN (R 4.1.0)
##  fansi                           0.5.0      2021-05-25 [1] CRAN (R 4.1.0)
##  farver                          2.1.0      2021-02-28 [1] CRAN (R 4.1.0)
##  fastmap                         1.1.0      2021-01-25 [1] CRAN (R 4.1.0)
##  fastmatch                       1.1-0      2017-01-28 [1] CRAN (R 4.1.0)
##  fgsea                           1.18.0     2021-05-19 [1] Bioconductor
##  filelock                        1.0.2      2018-10-05 [1] CRAN (R 4.1.0)
##  forcats                       * 0.5.1      2021-01-27 [1] CRAN (R 4.1.0)
##  generics                        0.1.0      2020-10-31 [1] CRAN (R 4.1.0)
##  GenomeInfoDb                  * 1.28.0     2021-05-19 [1] Bioconductor
##  GenomeInfoDbData                1.2.6      2021-05-21 [1] Bioconductor
##  GenomicAlignments               1.28.0     2021-05-19 [1] Bioconductor
##  GenomicFeatures               * 1.44.0     2021-05-19 [1] Bioconductor
##  GenomicRanges                 * 1.44.0     2021-05-19 [1] Bioconductor
##  ggforce                         0.3.3      2021-03-05 [1] CRAN (R 4.1.0)
##  ggnewscale                      0.4.5      2021-01-11 [1] CRAN (R 4.1.0)
##  ggplot2                       * 3.3.4      2021-06-16 [1] CRAN (R 4.1.0)
##  ggraph                          2.0.5      2021-02-23 [1] CRAN (R 4.1.0)
##  ggrepel                         0.9.1      2021-01-15 [1] CRAN (R 4.1.0)
```

```
##  ggtree                        3.1.1.992  2021-06-13 [1] Bioconductor
##  glue                          1.4.2      2020-08-27 [1] CRAN (R 4.1.0)
##  GO.db                         3.13.0     2021-05-22 [1] Bioconductor
##  GOSemSim                      2.18.0     2021-05-19 [1] Bioconductor
##  gplots                        3.1.1      2020-11-28 [1] CRAN (R 4.1.0)
##  graphlayouts                  0.7.1      2020-10-26 [1] CRAN (R 4.1.0)
##  gridExtra                     2.3        2017-09-09 [1] CRAN (R 4.1.0)
##  gtable                        0.3.0      2019-03-25 [1] CRAN (R 4.1.0)
##  gtools                        3.8.2      2020-03-31 [1] CRAN (R 4.1.0)
##  hms                           1.1.0      2021-05-17 [1] CRAN (R 4.1.0)
##  htmltools                     0.5.1.1    2021-01-22 [1] CRAN (R 4.1.0)
##  httr                          1.4.2      2020-07-20 [1] CRAN (R 4.1.0)
##  igraph                        1.2.6      2020-10-06 [1] CRAN (R 4.1.0)
##  IRanges                     * 2.26.0     2021-05-19 [1] Bioconductor
##  jsonlite                      1.7.2      2020-12-09 [1] CRAN (R 4.1.0)
##  kableExtra                  * 1.3.4      2021-02-20 [1] CRAN (R 4.1.0)
##  KEGGREST                      1.32.0     2021-05-19 [1] Bioconductor
##  KernSmooth                    2.23-20    2021-05-03 [1] CRAN (R 4.1.0)
##  knitr                         1.33       2021-04-24 [1] CRAN (R 4.1.0)
##  labeling                      0.4.2      2020-10-20 [1] CRAN (R 4.1.0)
##  lattice                       0.20-44    2021-05-02 [1] CRAN (R 4.1.0)
##  lazyeval                      0.2.2      2019-03-15 [1] CRAN (R 4.1.0)
##  lifecycle                     1.0.0      2021-02-15 [1] CRAN (R 4.1.0)
##  magrittr                    * 2.0.1      2020-11-17 [1] CRAN (R 4.1.0)
##  MASS                          7.3-54     2021-05-03 [1] CRAN (R 4.1.0)
##  Matrix                        1.3-3      2021-05-04 [1] CRAN (R 4.1.0)
##  MatrixGenerics                1.4.0      2021-05-19 [1] Bioconductor
##  matrixStats                   0.58.0     2021-01-29 [1] CRAN (R 4.1.0)
##  memoise                       2.0.0      2021-01-26 [1] CRAN (R 4.1.0)
##  munsell                       0.5.0      2018-06-12 [1] CRAN (R 4.1.0)
##  nlme                          3.1-152    2021-02-04 [1] CRAN (R 4.1.0)
##  org.Hs.eg.db                * 3.13.0     2021-05-22 [1] Bioconductor
##  patchwork                     1.1.1      2020-12-17 [1] CRAN (R 4.1.0)
##  pillar                        1.6.1      2021-05-16 [1] CRAN (R 4.1.0)
##  pkgconfig                     2.0.3      2019-09-22 [1] CRAN (R 4.1.0)
##  plotrix                       3.8-1      2021-01-21 [1] CRAN (R 4.1.0)
##  plyr                          1.8.6      2020-03-03 [1] CRAN (R 4.1.0)
##  png                           0.1-7      2013-12-03 [1] CRAN (R 4.1.0)
##  polyclip                      1.10-0     2019-03-14 [1] CRAN (R 4.1.0)
##  prettyunits                   1.1.1      2020-01-24 [1] CRAN (R 4.1.0)
##  progress                      1.2.2      2019-05-16 [1] CRAN (R 4.1.0)
##  purrr                         0.3.4      2020-04-17 [1] CRAN (R 4.1.0)
##  qvalue                        2.24.0     2021-05-19 [1] Bioconductor
##  R6                            2.5.0      2020-10-28 [1] CRAN (R 4.1.0)
##  rappdirs                      0.3.3      2021-01-31 [1] CRAN (R 4.1.0)
##  RColorBrewer                  1.1-2      2014-12-07 [1] CRAN (R 4.1.0)
##  Rcpp                          1.0.6      2021-01-15 [1] CRAN (R 4.1.0)
##  RCurl                         1.98-1.3   2021-03-16 [1] CRAN (R 4.1.0)
##  reshape2                      1.4.4      2020-04-09 [1] CRAN (R 4.1.0)
##  restfulr                      0.0.13     2017-08-06 [1] CRAN (R 4.1.0)
##  rjson                         0.2.20     2018-06-08 [1] CRAN (R 4.1.0)
##  rlang                         0.4.11     2021-04-30 [1] CRAN (R 4.1.0)
##  rmarkdown                   * 2.8        2021-05-07 [1] CRAN (R 4.1.0)
##  Rsamtools                     2.8.0      2021-05-19 [1] Bioconductor
##  RSQLite                       2.2.7      2021-04-22 [1] CRAN (R 4.1.0)
##  rstudioapi                    0.13       2020-11-12 [1] CRAN (R 4.1.0)
##  rtracklayer                   1.52.0     2021-05-19 [1] Bioconductor
##  rvcheck                     * 0.1.8      2020-03-01 [1] CRAN (R 4.1.0)
##  rvest                         1.0.0      2021-03-09 [1] CRAN (R 4.1.0)
##  S4Vectors                   * 0.30.0     2021-05-19 [1] Bioconductor
##  scales                        1.1.1      2020-05-11 [1] CRAN (R 4.1.0)
##  scatterpie                    0.1.6      2021-04-23 [1] CRAN (R 4.1.0)
##  sessioninfo                   1.1.1      2018-11-05 [1] CRAN (R 4.1.0)
##  shadowtext                    0.0.8      2021-04-23 [1] CRAN (R 4.1.0)
##  stringi                       1.6.2      2021-05-17 [1] CRAN (R 4.1.0)
##  stringr                       1.4.0      2019-02-10 [1] CRAN (R 4.1.0)
##  SummarizedExperiment          1.22.0     2021-05-19 [1] Bioconductor
##  svglite                       2.0.0      2021-02-20 [1] CRAN (R 4.1.0)
##  systemfonts                   1.0.2      2021-05-11 [1] CRAN (R 4.1.0)
##  tibble                        3.1.2      2021-05-16 [1] CRAN (R 4.1.0)
##  tidygraph                     1.2.0      2020-05-12 [1] CRAN (R 4.1.0)
##  tidyr                         1.1.3      2021-03-03 [1] CRAN (R 4.1.0)
##  tidyselect                    1.1.1      2021-04-30 [1] CRAN (R 4.1.0)
##  tidytree                      0.3.4      2021-05-22 [1] CRAN (R 4.1.0)
##  treeio                        1.17.1.992 2021-06-13 [1] Bioconductor
##  tweenr                        1.0.2      2021-03-23 [1] CRAN (R 4.1.0)
```

```
##  TxDb.Hsapiens.UCSC.hg19.knownGene * 3.2.2      2021-05-22 [1] Bioconductor
##  utf8                                1.2.1      2021-03-12 [1] CRAN (R 4.1.0)
##  vctrs                               0.3.8      2021-04-29 [1] CRAN (R 4.1.0)
##  viridis                             0.6.1      2021-05-11 [1] CRAN (R 4.1.0)
##  viridisLite                         0.4.0      2021-04-13 [1] CRAN (R 4.1.0)
##  webshot                             0.5.2      2019-11-22 [1] CRAN (R 4.1.0)
##  wget                              * 0.0.1      2020-04-27 [1] local
##  withr                               2.4.2      2021-04-18 [1] CRAN (R 4.1.0)
##  xfun                                0.23       2021-05-15 [1] CRAN (R 4.1.0)
##  XML                                 3.99-0.6   2021-03-16 [1] CRAN (R 4.1.0)
##  xml2                                1.3.2      2020-04-23 [1] CRAN (R 4.1.0)
##  XVector                             0.32.0     2021-05-19 [1] Bioconductor
##  yaml                                2.2.1      2020-02-01 [1] CRAN (R 4.1.0)
##  zlibbioc                            1.38.0     2021-05-19 [1] Bioconductor
##
## [1] /home/ygc/R/library
## [2] /usr/lib/R/library
```