

P259: Gene set enrichment analysis (GSEA)

Dendritic cells (pDC)

Kim Dill-McFarland, kadm@uw.edu

version May 18, 2021

Contents

Background	1
Setup	1
Load data	2
RNA-seq	2
Broad gene sets	2
Gene set enrichment analysis (GSEA)	2
Run GSEA	2
Significant GSEA	3
Hypergeometric enrichment of DEGs	3
Run enrichment	6
Significant enrichment	7
R session	7

Background

The purpose of this workflow is to perform GSEA for the impacts of human rhinovirus (RV) infection, eosinophil (EOS) supernatant, and Anti-IL5 therapy.

Setup

Load packages

```
# Data manipulation and figures
library(tidyverse)
# Multipanel figures
library(cowplot)

#GSEA
library(fgsea)
library(gage)

#Print pretty tables to Rmd
library(knitr)
library(kableExtra)
```

Set seed

```
set.seed(589)
```

Load data

RNA-seq

Contrast model results.

```
pval_2 <- read_csv("results/gene_level/P259.2_gene_pval.csv") %>%  
  filter(model=="contrasts")  
pval_1 <- read_csv("results/gene_level/P259.1_gene_pval.csv") %>%  
  filter(model=="contrasts")
```

Extract and format fold change (FC) for each contrast.

```
gene.ls <- list()  
  
for (contrast in unique(pval_2$group)){  
  #Subset to contrast of interest  
  pval.temp <- pval_2 %>% filter(group == contrast)  
  
  genes.temp <- pval.temp$logFC  
  names(genes.temp) <- pval.temp$hgnc_symbol  
  
  list.name <- paste(gsub(" - ", ".", contrast), 2, sep=".")  
  gene.ls[[list.name]] <- genes.temp  
}  
  
for (contrast in unique(pval_1$group)){  
  #Subset to contrast of interest  
  pval.temp <- pval_1 %>% filter(group == contrast)  
  
  genes.temp <- pval.temp$logFC  
  names(genes.temp) <- pval.temp$hgnc_symbol  
  
  list.name <- paste(gsub(" - ", ".", contrast), 1, sep=".")  
  gene.ls[[list.name]] <- genes.temp  
}
```

Broad gene sets

From <https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp>. Downloaded in data_clean/Broad_gmt/

Gene set enrichment analysis (GSEA)

The following function performs GSEA using both fast gene set enrichment analysis (fgsea) and generally applicable gene-set enrichment (gage).

```
source("https://raw.githubusercontent.com/kdillmcfarland/R_bioinformatic_scripts/master/GSEA_fxn.R")
```

Run GSEA

```
GSEA(gene_list = gene.ls,
     gmt_file="data_clean/Broad_gmt/h.all.v7.4.symbols.gmt")
```

Hallmark (H)

```
GSEA(gene_list = gene.ls,
     gmt_file="data_clean/Broad_gmt/c2.cp.v7.4.symbols.gmt")
```

Curated canonical pathway gene sets (C2:CP)

```
GSEA(gene_list = gene.ls,
     gmt_file="data_clean/Broad_gmt/c5.go.bp.v7.4.symbols.gmt")
```

GO biological process gene sets (C5:GO:BP)

Significant GSEA

Gene sets of interest are those significant for both virus AND EOS supernatant or Anti-IL5 therapy. Results are only considered significant if both fgsea and gage methods meet the FDR threshold in the same fold change direction.

Hypergeometric enrichment of DEGs

Test for Broad gene set enrichment in differentially expressed genes (DEG) as defined by change with virus (*e.g.* significant for virus in untreated and/or treated donors) AND different between untreated and treated donors in media OR RV. DEGs are defined at FDR < 0.1

```
#Define change with virus
DEG1.v <- read_csv("results/gene_level/P259.1_gene_pval.csv") %>%
  filter(group %in% c("none_HRV - none_none", "EOS.supp_HRV - EOS.supp_none") &
         adj.P.Val <= fdr.cut)

##
## -- Column specification -----
## cols(
##   geneName = col_character(),
##   logFC = col_double(),
##   AveExpr = col_double(),
##   t = col_double(),
##   P.Value = col_double(),
##   adj.P.Val = col_double(),
##   B = col_double(),
##   group = col_character(),
##   FC.group = col_character(),
##   model = col_character(),
##   hgnc_symbol = col_character(),
##   `Previous symbols` = col_character(),
##   `Alias symbols` = col_character(),
##   gene_biotype = col_character()
## )
```

set	pathway	group	fgsea.FDR	fgsea
HALLMARK	EPITHELIAL_MESENCHYMAL_TRANSITION	AntiIL5_HRVAntiIL5_none	0.0263992	-1.68
		EOSsupp_HRVEOSsupp_none	0.0257202	-1.57
		EOSsupp_HRVnone_HRV	0.1633394	-1.39
		none_HRVnone_none	0.1247661	-1.29
		none_HRVnone_none	0.3113446	-1.21
		AntiIL5_HRVnone_HRV	0.4429134	-1.16
		EOSsupp_nonenone_none	0.8136095	-1.06
		AntiIL5_nonenone_none	0.0155510	1.47
	INFLAMMATORY_RESPONSE	EOSsupp_HRVnone_HRV	0.3452781	-1.18
		EOSsupp_nonenone_none	0.7834331	-1.10
		AntiIL5_HRVAntiIL5_none	1.0000000	-0.87
		AntiIL5_HRVnone_HRV	0.7255245	0.98
		AntiIL5_nonenone_none	0.0275724	1.39
		none_HRVnone_none	0.0241313	1.70
		none_HRVnone_none	0.0242131	1.73
		EOSsupp_HRVEOSsupp_none	0.0257202	1.76
	INTERFERON_ALPHA_RESPONSE	EOSsupp_nonenone_none	0.0394945	-1.82
		EOSsupp_HRVnone_HRV	0.0917431	-1.80
		AntiIL5_HRVnone_HRV	1.0000000	0.68
		AntiIL5_nonenone_none	0.0233266	1.55
		none_HRVnone_none	0.0242131	2.94
		none_HRVnone_none	0.0241313	2.98
		EOSsupp_HRVEOSsupp_none	0.0257202	2.99
		AntiIL5_HRVAntiIL5_none	0.1048218	3.35
	INTERFERON_GAMMA_RESPONSE	EOSsupp_nonenone_none	0.0394945	-1.62
		EOSsupp_HRVnone_HRV	0.1385809	-1.41
		AntiIL5_HRVnone_HRV	0.4429134	1.17
		AntiIL5_nonenone_none	0.0155510	1.45
		none_HRVnone_none	0.0242131	2.80
		AntiIL5_HRVAntiIL5_none	0.1838235	2.81
		none_HRVnone_none	0.0241313	2.84
		EOSsupp_HRVEOSsupp_none	0.0257202	3.02
NABA	CORE_MATRISOME	AntiIL5_HRVAntiIL5_none	0.0764426	-1.83
		EOSsupp_nonenone_none	0.7379868	-1.45
		AntiIL5_HRVnone_HRV	0.8939063	-1.26
		none_HRVnone_none	0.7781424	-1.12
		EOSsupp_HRVnone_HRV	0.9452543	1.19
		EOSsupp_HRVEOSsupp_none	0.6651976	1.21
		none_HRVnone_none	0.6772921	1.21
		AntiIL5_nonenone_none	0.0625027	1.66
	ECM_AFFILIATED	AntiIL5_HRVAntiIL5_none	0.0764426	-1.80
		EOSsupp_nonenone_none	0.9367200	-1.40
		none_HRVnone_none	0.4178844	-1.37
		EOSsupp_HRVEOSsupp_none	0.5691744	-1.29
		EOSsupp_HRVnone_HRV	1.0000000	-1.17
		AntiIL5_HRVnone_HRV	0.9820625	-1.12
		none_HRVnone_none	0.8358530	-1.09
		AntiIL5_nonenone_none	0.0625027	1.66
	ECM_REGULATORS	AntiIL5_HRVAntiIL5_none	0.0764426	-1.86
		none_HRVnone_none	0.1267318	-1.75
		EOSsupp_HRVEOSsupp_none	0.2253595	-1.42
		EOSsupp_nonenone_none	0.8758220	-1.39
		none_HRVnone_none	0.4695960	-1.28
		AntiIL5_HRVnone_HRV	0.9276141	-1.24
		EOSsupp_HRVnone_HRV	1.0000000	1.07
		AntiIL5_nonenone_none	0.0625027	1.55
REACTOME	GPCR_LIGAND_BINDING	AntiIL5_HRVAntiIL5_none	0.0764426	-1.58
		EOSsupp_HRVnone_HRV	0.4309555	-1.35
		EOSsupp_nonenone_none	0.9527990	-1.31

```

#Define change with treament
DEG1.t <- read_csv("results/gene_level/P259.1_gene_pval.csv") %>%
  filter(group %in% c("EOS.supp_none - none_none", "EOS.supp_HRV - none_HRV") &
    adj.P.Val <= fdr.cut)

##
## -- Column specification -----
## cols(
##   geneName = col_character(),
##   logFC = col_double(),
##   AveExpr = col_double(),
##   t = col_double(),
##   P.Value = col_double(),
##   adj.P.Val = col_double(),
##   B = col_double(),
##   group = col_character(),
##   FC.group = col_character(),
##   model = col_character(),
##   hgnc_symbol = col_character(),
##   `Previous symbols` = col_character(),
##   `Alias symbols` = col_character(),
##   gene_biotype = col_character()
## )

#Define change with virus
DEG2.v <- read_csv("results/gene_level/P259.2_gene_pval.csv") %>%
  filter(group %in% c("none_HRV - none_none", "AntiIL5_HRV - AntiIL5_none") &
    adj.P.Val <= fdr.cut)

##
## -- Column specification -----
## cols(
##   geneName = col_character(),
##   logFC = col_double(),
##   AveExpr = col_double(),
##   t = col_double(),
##   P.Value = col_double(),
##   adj.P.Val = col_double(),
##   B = col_double(),
##   group = col_character(),
##   FC.group = col_character(),
##   model = col_character(),
##   hgnc_symbol = col_character(),
##   `Previous symbols` = col_character(),
##   `Alias symbols` = col_character(),
##   gene_biotype = col_character()
## )

#Define change with treament
DEG2.t <- read_csv("results/gene_level/P259.2_gene_pval.csv") %>%
  filter(group %in% c("AntiIL5_none - none_none", "AntiIL5_HRV - none_HRV") &
    adj.P.Val <= fdr.cut)

##
## -- Column specification -----

```

```
## cols(
##   geneName = col_character(),
##   logFC = col_double(),
##   AveExpr = col_double(),
##   t = col_double(),
##   P.Value = col_double(),
##   adj.P.Val = col_double(),
##   B = col_double(),
##   group = col_character(),
##   FC.group = col_character(),
##   model = col_character(),
##   hgnc_symbol = col_character(),
##   `Previous symbols` = col_character(),
##   `Alias symbols` = col_character(),
##   gene_biotype = col_character()
## )

#Format to ENTREZ

#Format in list
DEG.ls <- list()
DEG.ls[["EOSsup"]] <- intersect(DEG1.v$hgnc_symbol, DEG1.t$hgnc_symbol)
DEG.ls[["AntiIL5"]] <- intersect(DEG2.v$hgnc_symbol, DEG2.t$hgnc_symbol)

#Script for running term enrichment
source("https://raw.githubusercontent.com/kdillmcfarland/R_bioinformatic_scripts/master/hypergeo_enrichment.R")
```

Run enrichment

```
enrich.fxn(gene.list = DEG.ls, ID.type="SYMBOL",
           category = "H",
           genome = "org.Hs.eg.db",
           basename = "DEG",
           outdir = "results/enrichment/")
```

Hallmark (H)

```
enrich.fxn(gene.list = DEG.ls, ID.type="SYMBOL",
           category = "C2", subcategory="CP",
           genome = "org.Hs.eg.db",
           basename = "DEG",
           outdir = "results/enrichment/")
```

Curated canonical pathway gene sets (C2:CP)

```
enrich.fxn(gene.list = DEG.ls, ID.type="SYMBOL",
           category = "C5", subcategory = "GO:BP",
           genome = "org.Hs.eg.db",
           basename = "DEG",
           outdir = "results/enrichment/")
```

GO biological process gene sets (C5:GO:BP)

Significant enrichment

Gene sets enriched in DEGs at FDR < 0.05

R session

```
sessionInfo()
```

```
## R version 4.0.2 (2020-06-22)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS 10.16
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] kableExtra_1.3.4.9000 knitr_1.33      gage_2.38.3
## [4] fgsea_1.14.0          cowplot_1.1.1   forcats_0.5.1
## [7] stringr_1.4.0         dplyr_1.0.6     purrr_0.3.4
## [10] readr_1.4.0           tidyr_1.1.3     tibble_3.1.1
## [13] ggplot2_3.3.3         tidyverse_1.3.1
##
## loaded via a namespace (and not attached):
## [1] fs_1.5.0              lubridate_1.7.10    bit64_4.0.5
## [4] webshot_0.5.2         httr_1.4.2          tools_4.0.2
## [7] backports_1.2.1       utf8_1.2.1          R6_2.5.0
## [10] DBI_1.1.1             BiocGenerics_0.34.0 colorspace_2.0-1
## [13] withr_2.4.2           tidysselect_1.1.1   gridExtra_2.3
## [16] bit_4.0.4             compiler_4.0.2       graph_1.66.0
## [19] cli_2.5.0             rvest_1.0.0         Biobase_2.48.0
## [22] xml2_1.3.2            scales_1.1.1        systemfonts_1.0.2
## [25] digest_0.6.27         rmarkdown_2.8       svglite_2.0.0
## [28] XVector_0.28.0        pkgconfig_2.0.3     htmltools_0.5.1.1
## [31] dbplyr_2.1.1          fastmap_1.1.0       rlang_0.4.11
## [34] readxl_1.3.1          rstudioapi_0.13     RSQlite_2.2.7
## [37] generics_0.1.0        jsonlite_1.7.2      BiocParallel_1.24.1
## [40] magrittr_2.0.1        GO.db_3.11.4        Matrix_1.3-3
## [43] Rcpp_1.0.6            munsell_0.5.0       S4Vectors_0.26.1
## [46] fansi_0.4.2           lifecycle_1.0.0     stringi_1.6.1
## [49] yaml_2.2.1            zlibbioc_1.34.0     grid_4.0.2
## [52] blob_1.2.1            parallel_4.0.2      crayon_1.4.1
## [55] lattice_0.20-44       Biostrings_2.56.0   haven_2.4.1
## [58] hms_1.0.0             KEGGREST_1.28.0     pillar_1.6.0
## [61] stats4_4.0.2          fastmatch_1.1-0     reprex_2.0.0
## [64] glue_1.4.2            evaluate_0.14       data.table_1.14.0
## [67] modelr_0.1.8          vctrs_0.3.8         png_0.1-7
## [70] cellranger_1.1.0      gtable_0.3.0        assertthat_0.2.1
```

group	set	pathway
AntiIL5	GOBP	DEFENSE_RESPONSE_TO_VIRUS
		RESPONSE_TO_VIRUS
		REGULATION_OF_CYTOSOLIC_CALCIUM_ION_CONCENTRATION
		RESPONSE_TO_TYPE_I_INTERFERON
		NEGATIVE_REGULATION_OF_VIRAL_GENOME_REPLICATION
		POSITIVE_REGULATION_OF_CATION_TRANSMEMBRANE_TRANSPORT
		REGULATION_OF_CALCIUM_ION_TRANSMEMBRANE_TRANSPORT
		DIVALENT_INORGANIC_CATION_HOMEOSTASIS
		POSITIVE_REGULATION_OF_CALCIUM_ION_TRANSMEMBRANE_TRANSPORT
		ANTIMICROBIAL_HUMORAL_IMMUNE_RESPONSE_MEDIATED_BY_ANTIMICROBIAL
		REGULATION_OF_RELEASE_OF_SEQUESTERED_CALCIUM_ION_INTO_CYTOSOL
		REGULATION_OF_VIRAL_GENOME_REPLICATION
		POSITIVE_REGULATION_OF_TRANSMEMBRANE_TRANSPORT
		NEGATIVE_REGULATION_OF_VIRAL_PROCESS
		REGULATION_OF_CALCIUM_ION_TRANSPORT_INTO_CYTOSOL
		RESPONSE_TO_CHEMOKINE
		RESPONSE_TO_INTERFERON_ALPHA
		NEUTROPHIL_CHEMOTAXIS
		REGULATION_OF_METAL_ION_TRANSPORT
		REGULATION_OF_CARDIAC_MUSCLE_CELL_ACTION_POTENTIAL
		T_CELL_CHEMOTAXIS
		POSITIVE_REGULATION_OF_CALCIUM_ION_TRANSPORT
		SEQUESTERING_OF_CALCIUM_ION
		GRANULOCYTE_CHEMOTAXIS
		NEUTROPHIL_MIGRATION
		VIRAL_GENOME_REPLICATION
		VENTRICULAR_CARDIAC_MUSCLE_CELL_ACTION_POTENTIAL
	HALLMARK	INTERFERON_ALPHA_RESPONSE
		INTERFERON_GAMMA_RESPONSE
	KEGG	CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION
	REACTOME	INTERFERON_ALPHA_BETA_SIGNALING
		PEPTIDE_LIGAND_BINDING_RECEPTORS
		CHEMOKINE_RECEPTORS_BIND_CHEMOKINES
		GPCR_LIGAND_BINDING
		CLASS_A_1_RHODOPSIN_LIKE_RECEPTORS
EOSsup	PID	INTERFERON_SIGNALING
		AR_PATHWAY


```
## [73] cachem_1.0.4      xfun_0.22          broom_0.7.6
## [76] viridisLite_0.4.0 AnnotationDbi_1.50.3 memoise_2.0.0
## [79] IRanges_2.22.2    ellipsis_0.3.2
```
