P259: Gene set enrichment analysis (GSEA)
Dendritic cells (pDC)

Kim Dill-McFarland, kadm@uw.edu

version September 09, 2020

Contents

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.v7.1.symbols.gmt")
```

Hallmark (H)

```
GSEA(gene_list = gene.ls,
     gmt_file="data_clean/Broad_gmt/c2.all.v7.1.symbols.gmt")
```

Curated gene sets (C2)

```
GSEA(gene_list = gene.ls,
     gmt_file="data_clean/Broad_gmt/c5.all.v7.1.symbols.gmt")
```
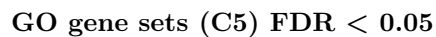
GO gene sets (C5)

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.

Hallmark FDR < 0.1

Curated gene sets (C2) FDR < 0.05

GO gene sets (C5) FDR < 0.05
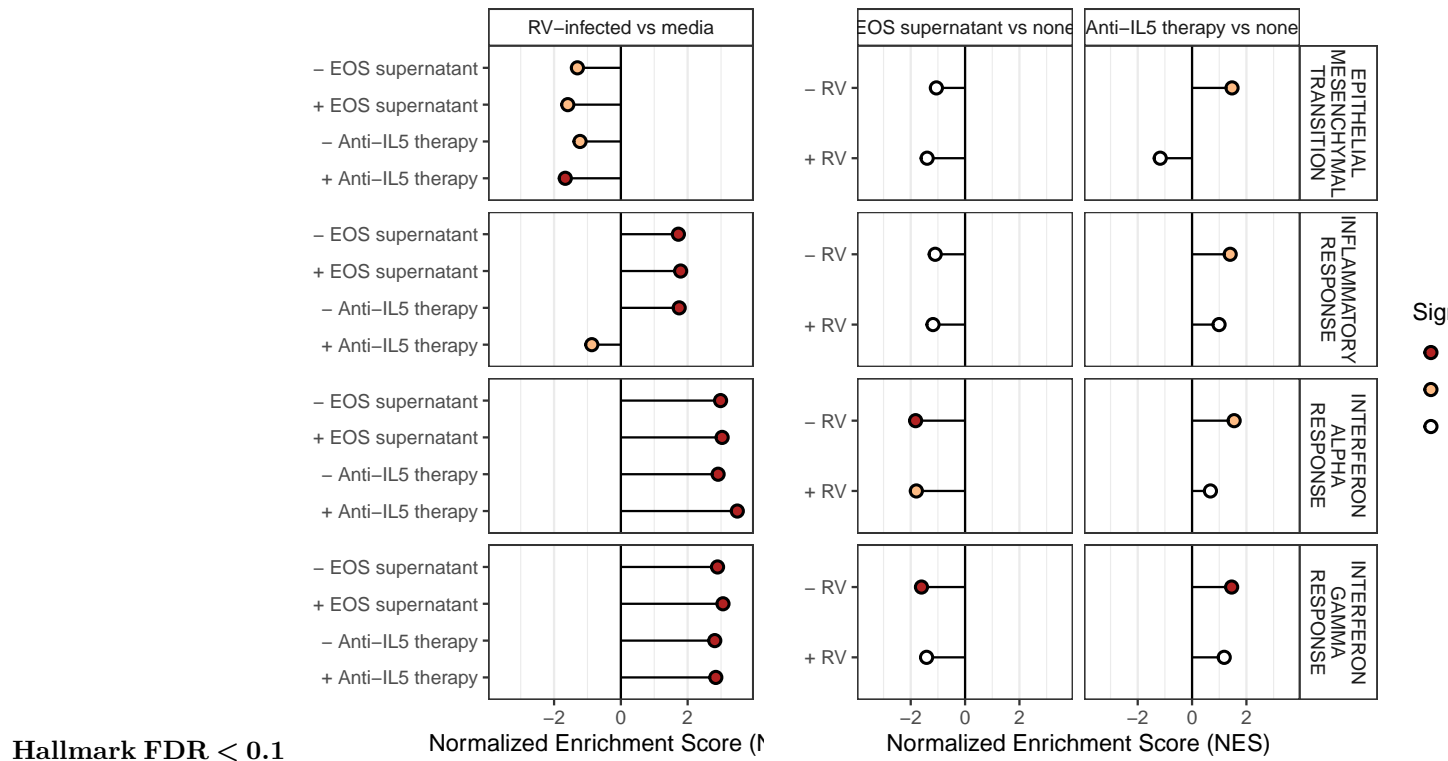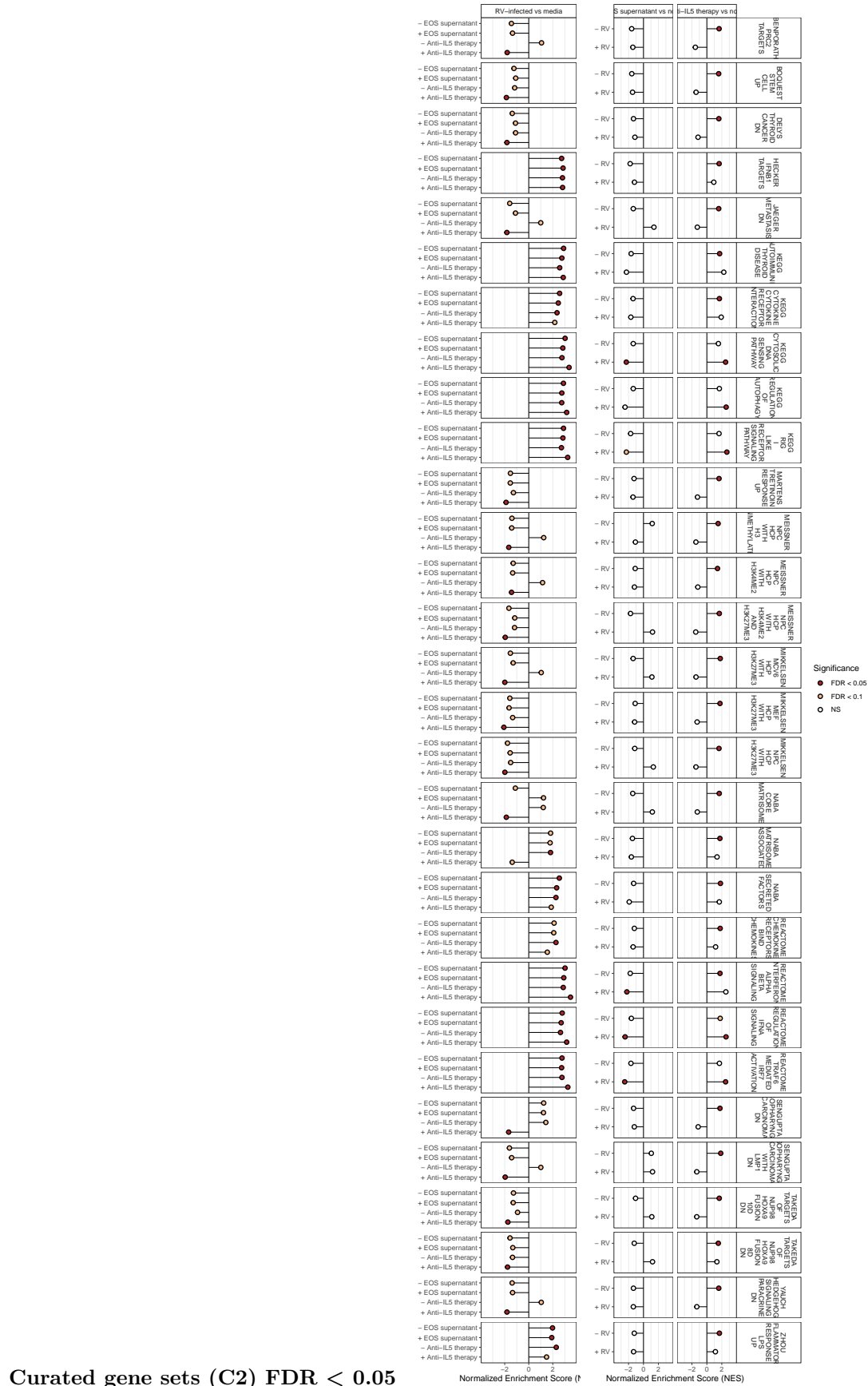
R session

```
sessionInfo()
```

```
## R version 4.0.0 (2020-04-24)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Catalina 10.15.5
##
## 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_1.1.0 knitr_1.29       gage_2.38.3      fgsea_1.14.0
##  [5] cowplot_1.0.0    forcats_0.5.0   stringr_1.4.0    dplyr_1.0.0
##  [9] purrr_0.3.4      readr_1.3.1     tidyr_1.1.0      tibble_3.0.3
## [13] ggplot2_3.3.2    tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
##  [1] Biobase_2.48.0      httr_1.4.2          viridisLite_0.3.0
##  [4] bit64_0.9-7.1       jsonlite_1.7.0      modelr_0.1.8
##  [7] assertthat_0.2.1    stats4_4.0.0        blob_1.2.1
## [10] cellranger_1.1.0    yaml_2.2.1          pillar_1.4.6
## [13] RSQLite_2.2.0       backports_1.1.8     lattice_0.20-41
## [16] glue_1.4.1          digest_0.6.25       XVector_0.28.0
## [19] rvest_0.3.6         colorspace_1.4-1    htmltools_0.5.0
## [22] Matrix_1.2-18       pkgconfig_2.0.3     broom_0.7.0
## [25] haven_2.3.1         zlibbioc_1.34.0     GO.db_3.11.4
## [28] webshot_0.5.2       scales_1.1.1        BiocParallel_1.22.0
## [31] KEGGREST_1.28.0     farver_2.0.3        generics_0.0.2
## [34] IRanges_2.22.2      ellipsis_0.3.1      withr_2.2.0
## [37] BiocGenerics_0.34.0 cli_2.0.2           magrittr_1.5
## [40] crayon_1.3.4        readxl_1.3.1        memoise_1.1.0
## [43] evaluate_0.14       fs_1.4.2            fansi_0.4.1
## [46] xml2_1.3.2          graph_1.66.0        tools_4.0.0
## [49] data.table_1.13.0   hms_0.5.3           lifecycle_0.2.0
## [52] S4Vectors_0.26.1    munsell_0.5.0       reprex_0.3.0
## [55] AnnotationDbi_1.50.3 Biostrings_2.56.0   compiler_4.0.0
## [58] rlang_0.4.7         grid_4.0.0          rstudioapi_0.11
## [61] labeling_0.3        rmarkdown_2.3       gtable_0.3.0
## [64] DBI_1.1.0           R6_2.4.1            gridExtra_2.3
## [67] lubridate_1.7.9     bit_1.1-15.2        fastmatch_1.1-0
## [70] stringi_1.4.6       parallel_4.0.0      Rcpp_1.0.5
## [73] vctrs_0.3.2         png_0.1-7           dbplyr_1.4.4
## [76] tidyselect_1.1.0    xfun_0.16
```