Step 3.1: Comparison of gene expression changes to NASA twin study

Carlos Gallardo & Christian Oertlin

17 April, 2023

```
# Import libraries and helper functions
source("code/helper_functions.R")
library(tidyverse)
library(magrittr)
library(patchwork)
library(RColorBrewer)
library(openxlsx)
library(ComplexHeatmap)

# Colors
colPals <- vector(mode = "list")
colPals$time <- setNames(c("#FBAA3E", "#2C83BE", "#3EB6BD", "#A3D5B3", "#CD71A8"),
                         nm = c("day0", "day7", "day14", "day21", "day28"))
colPals$time_light <- setNames(c("#FDD6A1", "#A2CDE9", "#AFE2E5", "#DDF0E3", "#E8BDD6"),
                               nm = c("day0", "day7", "day14", "day21", "day28"))
colPals$time_dark <- setNames(c("#D87E04", "#174564", "#1F5C60", "#49A065", "#AA3C7E"),
                              nm = c("day0", "day7", "day14", "day21", "day28"))
colPals$inferno <- c("#000004", "#420A68", "#932667", "#DD513A", "#FCA50A", "#FCFFA4")
colPals$blood_cells <- setNames(c("#E54D34","#77A2D5","#B58B80"),
                                nm = c("granulocytes", "lymphocytes", "monocytes"))
colPals$cell_types <- setNames(c("#83D1F6","#FBAA3E","#FCCA7C","#B58B80","#E54D34",
                                 "#B3177E","#9A509F","#77A2D5","#CAC1DD","#36B449","#C1C1C1"),
                               nm = c("B cell", "Macrophage M1", "Macrophage M2",
                                      "Monocyte", "Neutrophil", "NK cell",
                                      "T cell CD4+ (non-regulatory)", "T cell CD8+",
                                      "T cell regulatory (Tregs)", "Myeloid dendritic cell",
                                      "uncharacterized cell"))
colPals$RdBu <- brewer.pal(11, name = "RdBu")
colPals$biotype <- setNames(c("#395982","#49BED9","#18A38A","#36B449","#826F99",
                              "#9852A5","#FBAA3E","#FCCA7C","#FCFFA4","#C1C1C1"),
                            nm = c("protein_coding", "lncRNA", "miRNA", "snoRNA",
                                   "IG_C_gene", "IG_V_gene", "TR_C_gene",
                                   "TR_J_gene", "TR_V_gene", "other"))
colPals$rtqpcr_rnaseq <- setNames(c("#B80D48","#2B6A6C"),
                                  nm = c("rtqpcr", "rnaseq"))
```

Load data

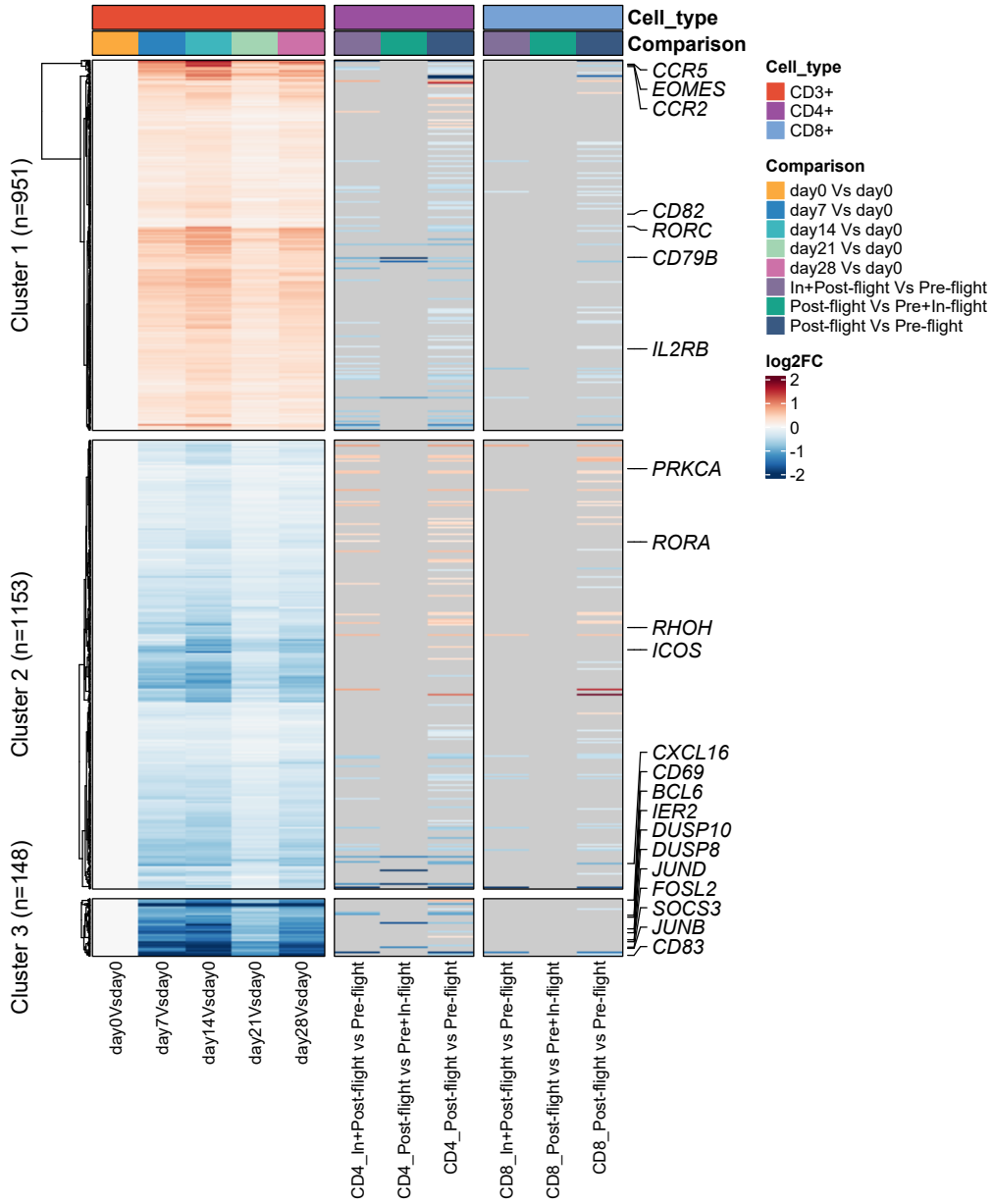
```
# RNA-seq
RNAseq <- readRDS(file='data/rnaseq/rnaseq_volunteers_9&10_excl.rds')
DESeq2_DEGs <- readRDS(file='data/rnaseq/DESeq2_DEGs_unfilt_volunteers_9&10_excl.rds')
DESeq2_DEGs_filt <- readRDS(file='data/rnaseq/DESeq2_DEGs_filt_volunteers_9&10_excl.rds')
DEGs_clusters <- readRDS(file='data/rnaseq/DEGs_kmeans_clusters.rds')

# DEGs from comparisons in NASA twin study
NASA_twin <- read.table(file = 'data/resources/nasa_twin_study_DEGs.csv',
                        skip = 1,
                        stringsAsFactors = FALSE,
                        sep = "\t",
                        header = TRUE,
                        fill = FALSE,
                        quote = "")
```

Heatmap comparing DI changes to NASA study

```
# T cell DEG clusters in dry immersion
# filter genes without gene symbol (ENSID) and also duplicated entries for a gene symbol
DEG_clusters_FC_filt <- DEGs_clusters$DEGs_clusters_FC %>%
  filter(!grepl('^ENS', GeneSymbol)) %>%
  mutate(GeneSymbol = gsub('(.+)_\\d+','\\1',GeneSymbol)) %>%
  filter(!duplicated(GeneSymbol))


# relevant comparisons in NASA study for CD4+ and CD8+ T cells
# we average log2FC values of the different RNA-seq experiment types
NASA_twin_selected <- NASA_twin %>%
  filter(CellType %in% c('CD4','CD8')) %>%
  filter(Coefficient %in% c('Post-flight vs Pre-flight',
                            'In+Post-flight vs Pre-flight',
                            'Post-flight vs Pre+In-flight')) %>%
  dplyr::rename(GeneSymbol = Gene, log2FC = log2.Fold.Change) %>%
  group_by(CellType, Coefficient, GeneSymbol) %>%
  summarise(log2FC = mean(log2FC)) %>%
  unite(col = 'condition', CellType, Coefficient, sep = '_', remove = T) %>%
  pivot_wider(names_from = condition, values_from = log2FC, names_prefix = '')

# combine t cell dry immersion data with NASA study
DEGs_DI_NASA_combined <- DEG_clusters_FC_filt %>%
  left_join(NASA_twin_selected, by = 'GeneSymbol') %>%
  select(-Cluster) %>%
  relocate(Biotype, .after = last_col())
```

```
# relevant genes to annotate
mark.genes <- c("IL7R","ETS1","GATA3","TCF7","TCF1","BCL11B",
                "SPI1","HES1","BCL11A","TCF12","BCL6","BCL2",
                "IER2",
                "CD27", # activation marker
                "CD3G",
                "CD69", # Early activation marker
                "CCR10","CCR2","CCR5",
                "CD160", # inhibits t cell activation
                "CD79B",
                "CD82","CD83",
                "RORA","RORC",
                "FOXP3",
                "CTLA4",
                "PDCD1",
                "CXCR4",
                "CXCL16",
                "ICOS",
                "IL2RA","IL2RB",
                "IL10RA",
                "EOMES",
                "SOCS1","SOCS3",
                "RHOH",
                "DUSP1","DUSP2","DUSP4","DUSP8","DUSP10",
                "FOS","FOSL2",
                "JUN","JUNB","JUND",
                "STAT5",
                "PRKCA",
                "ATF2"
)

# genes changed in at least 1 of the NASA comparisons
detected_NASA_comparisons <- DEGs_DI_NASA_combined %>%
  column_to_rownames(var = 'GeneSymbol') %>%
  select(`CD4_In+Post-flight vs Pre-flight`:`CD8_Post-flight vs Pre-flight`) %>%
  mutate_all(~replace(., is.na(.), 0)) %>%
  mutate_all(abs) %>%
  mutate(sum_change = rowSums(.)) %>%
  filter(sum_change > 0) %>%
  rownames_to_column(var = 'GeneSymbol') %>%
  pull(GeneSymbol)

mark.genes <- intersect(mark.genes, detected_NASA_comparisons)


m <- DEGs_DI_NASA_combined %>%
  column_to_rownames(var = 'GeneSymbol') %>%
  select(day0Vsday0:`CD8_Post-flight vs Pre-flight`)

set.seed(25)
clust_mat <- m
```

```r
clust_mat[is.na(clust_mat)] <- 0
clust = kmeans(clust_mat, centers = 3)

cell_type <- c(rep('CD3+',5), rep('CD4+',3), rep('CD8+',3))

comparison <- gsub('Vs', ' Vs ', colnames(m))
comparison <- gsub(' vs ', ' Vs ', comparison)
comparison <- gsub('CD\\d+_', '', comparison)

lbls <- c("1" = paste0("Cluster 1 (n=",sum(clust$cluster=='1'),")"),
          "2" = paste0("Cluster 2 (n=",sum(clust$cluster=='2'),")"),
          "3" = paste0("Cluster 3 (n=",sum(clust$cluster=='3'),")"))

ha_top <- HeatmapAnnotation(
  Cell_type = factor(cell_type, levels = unique(cell_type)),
  Comparison = factor(comparison, levels = unique(comparison)),
  col = list(
    Cell_type = setNames(c('#E54D34','#9A509F','#77A2D5'),
                         nm = unique(cell_type)),
    Comparison = setNames(c(colPals$time,'#826F99','#18A38A','#395982'),
                         nm = unique(comparison))
  ),
  annotation_name_gp = gpar(fontface = 'bold'),
  border = T
)

ha_right <- rowAnnotation(
  mark = anno_mark(at=which(rownames(m) %in% mark.genes),
                   labels = rownames(m)[which(rownames(m) %in% mark.genes)],
                   padding = unit(1,"mm"),
                   labels_gp = gpar(fontface = 'italic'))
)

p <- Heatmap(m, name = "log2FC",
             row_split = clust$cluster, cluster_row_slices = F, cluster_rows = T,
             column_title = NULL, column_split=factor(cell_type, levels = unique(cell_type)), cluster_columns = F,
             col = circlize::colorRamp2(breaks=seq(-2, 2, length.out=21),
                                        colors=colorRampPalette(rev(colPals$RdBu))(21)),
             top_annotation = ha_top,
             right_annotation = ha_right,
             width = unit(110, "mm"),
             na_col = "grey80",
             show_row_names = F,
             row_title = lbls,
             show_row_dend = T, row_dend_width=unit(10, "mm"), row_gap = unit(2, "mm"),
             show_column_names = T, column_names_gp = gpar(fontsize = 10), column_gap = unit(2, "mm"),
             border = T)

draw(p, merge_legend = T, align_heatmap_legend = "heatmap_top")
```



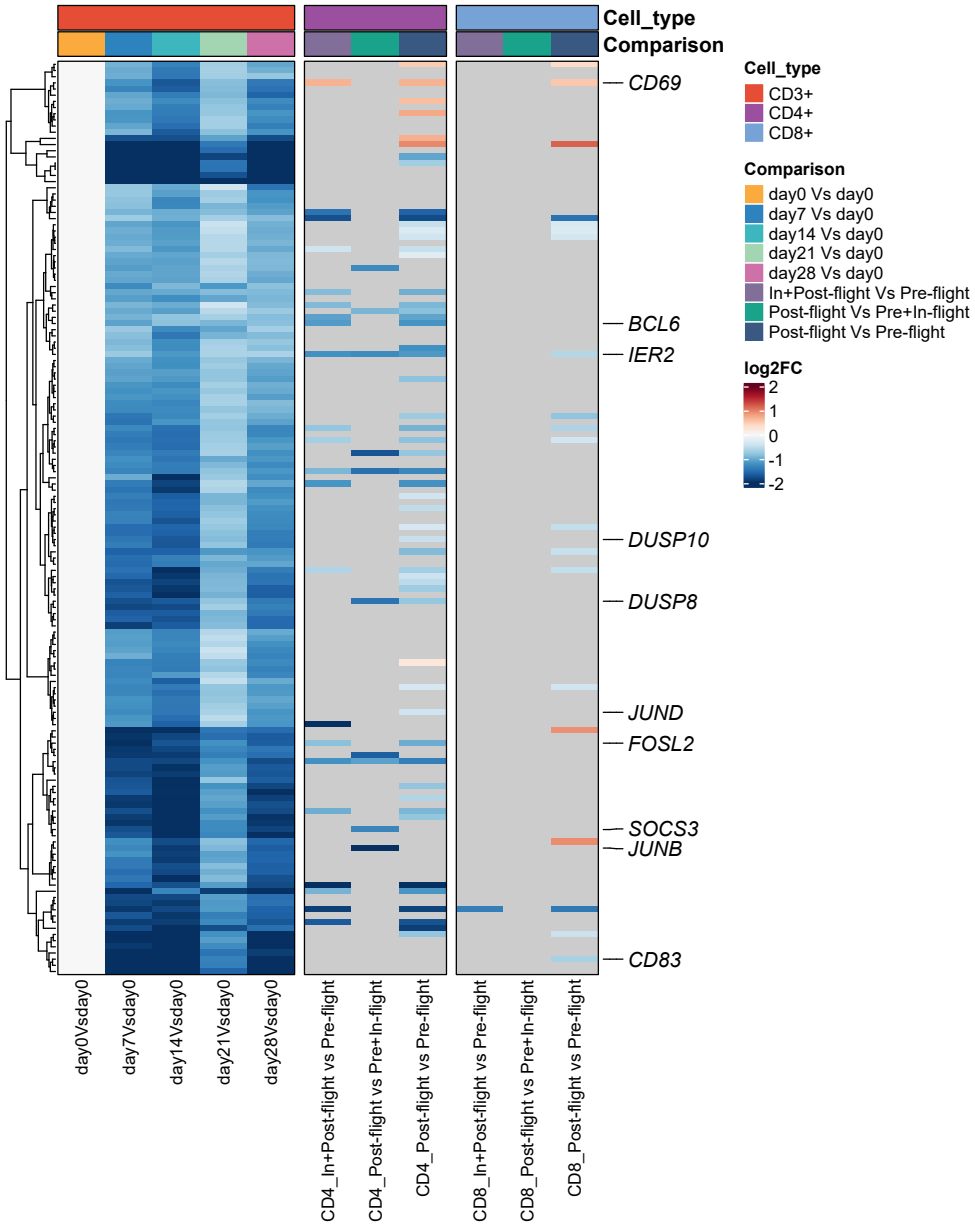

```
pdf("plots/figS9_DEG_NASA_twin_study_comparison.pdf", width = 10, height = 10)
draw(p, merge_legend = T, align_heatmap_legend = "heatmap_top")
dev.off()
```

```
## cairo_pdf
##         2
```

Heatmap of heavily downregulated genes (cluster 3)

```
# add cluster information
DEGs_DI_NASA_combined <- DEGs_DI_NASA_combined %>%
  add_column(Cluster = recode(.$GeneSymbol, !!!clust$cluster), .after = 'GeneSymbol') %>%
  arrange(Cluster)

# subset cluster 3
DEGs_DI_NASA_combined_cl3 <- DEGs_DI_NASA_combined %>%
  filter(Cluster == '3')

# generate heatmap
```

```r
m <- DEGs_DI_NASA_combined_cl3 %>%
  column_to_rownames(var = 'GeneSymbol') %>%
  select(day0Vsday0:`CD8_Post-flight vs Pre-flight`)

cell_type <- c(rep('CD3+',5), rep('CD4+',3), rep('CD8+',3))

comparison <- gsub('Vs', ' Vs ', colnames(m))
comparison <- gsub(' vs ', ' Vs ', comparison)
comparison <- gsub('CD\\d+_', '', comparison)

ha_top <- HeatmapAnnotation(
  Cell_type = factor(cell_type, levels = unique(cell_type)),
  Comparison = factor(comparison, levels = unique(comparison)),
  col = list(
    Cell_type = setNames(c('#E54D34','#9A509F','#77A2D5'),
                         nm = unique(cell_type)),
    Comparison = setNames(c(colPals$time,'#826F99','#18A38A','#395982'),
                         nm = unique(comparison))
  ),
  annotation_name_gp = gpar(fontface = 'bold'),
  border = T
)

ha_right <- rowAnnotation(
  mark = anno_mark(at=which(rownames(m) %in% mark.genes),
                   labels = rownames(m)[which(rownames(m) %in% mark.genes)],
                   padding = unit(1,"mm"),
                   labels_gp = gpar(fontface = 'italic'))
)

p <- Heatmap(m, name = "log2FC",
             cluster_row_slices = F, cluster_rows = T,
             column_title = NULL, column_split=factor(cell_type, levels = unique(cell_type)), cluster_columns = F,
             col = circlize::colorRamp2(breaks=seq(-2, 2, length.out=21),
                                        colors=colorRampPalette(rev(colPals$RdBu))(21)),
             top_annotation = ha_top,
             right_annotation = ha_right,
             width = unit(110, "mm"),
             na_col = "grey80",
             show_row_names = F,
             show_row_dend = T, row_dend_width=unit(10, "mm"), row_gap = unit(2, "mm"),
             show_column_names = T, column_names_gp = gpar(fontsize = 10), column_gap = unit(2, "mm"),
             border = T)

draw(p, merge_legend = T, align_heatmap_legend = "heatmap_top")
```



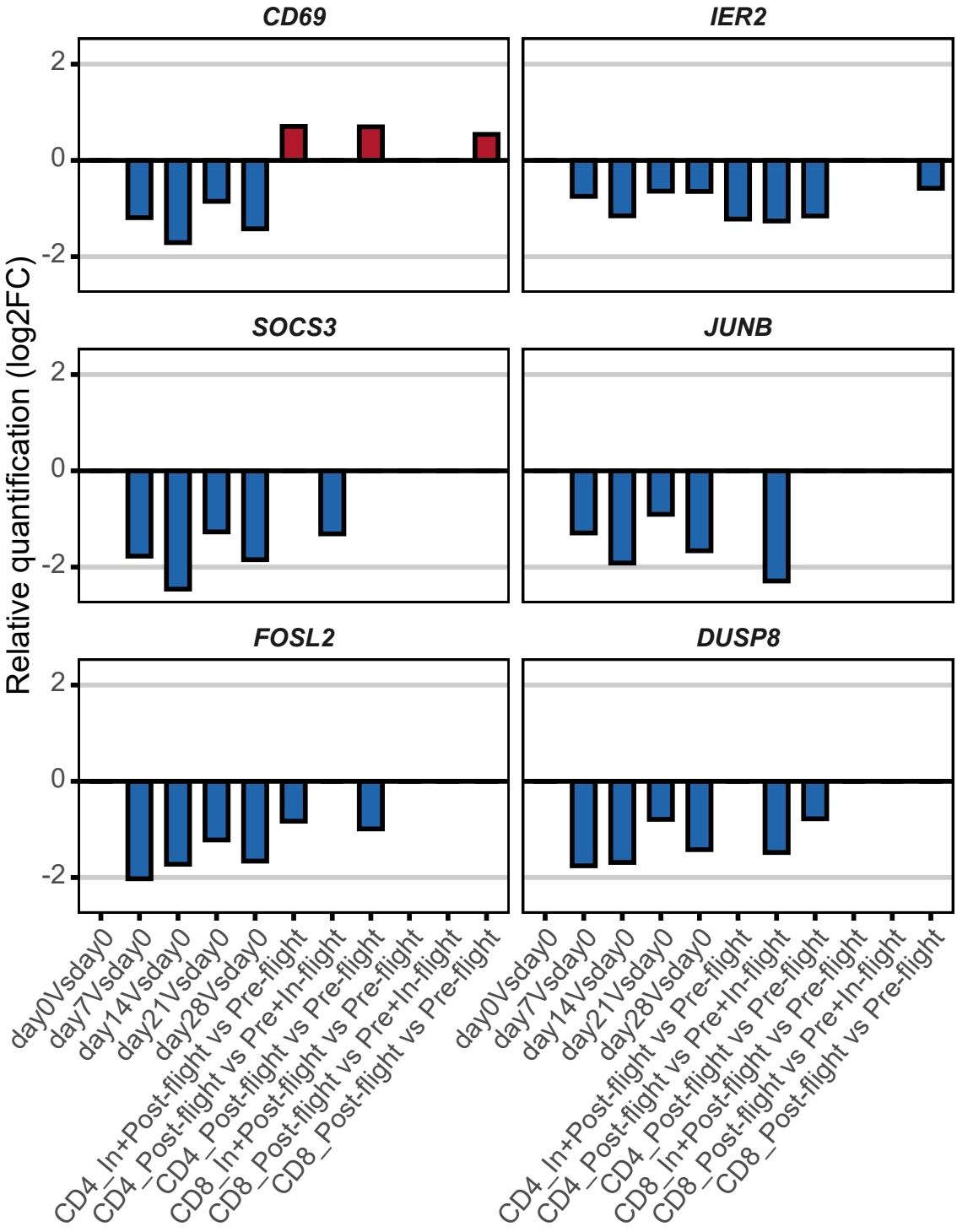

```
pdf("plots/fig3A_DEG_NASA_twin_study_comparison_downregulated.pdf", width = 10, height = 10)
draw(p, merge_legend = T, align_heatmap_legend = "heatmap_top")
dev.off()
```

```
## cairo_pdf
##         2
```

Expression change of selected genes

```
selected_genes <- c('CD69','IER2','SOCS3','JUNB','FOSL2','DUSP8')

df <- DEGs_DI_NASA_combined_cl3 %>%
  filter(GeneSymbol %in% selected_genes) %>%
  pivot_longer(day0Vsday0:`CD8_Post-flight vs Pre-flight`,
               names_to = 'comparison',
               values_to = 'log2FC') %>%
  mutate(GeneSymbol = factor(GeneSymbol, levels = selected_genes),
         comparison = factor(comparison, levels = unique(comparison)),
         log2FC = ifelse(is.na(log2FC), 0, log2FC),
         type = 'unchanged') %>%
```

```
  mutate(type = ifelse(log2FC > 0, 'up', 'down')) %>%
  mutate(type = factor(type, levels = c('up', 'down', 'unchanged')))


ggplot(df, aes(x=comparison, y=log2FC, fill=type)) +
  geom_bar(stat='identity', size=1, width=0.6, color='black') +
  geom_hline(yintercept = 0, color='black', size=1) +
  scale_y_continuous(breaks = c(-2,0,2),
                     limits = c(-2.5,2.5),
                     expand = expansion(mult = c(.05, .01))) +
  scale_fill_manual(values = colPals$RdBu[c(2,10)]) +
  facet_wrap(~GeneSymbol, ncol=2) +
  xlab('') +
  ylab('Relative quantification (log2FC)') +
  theme_bw() +
  theme(
    text = element_text(family = 'Arial', size = 14),
    axis.text.x.bottom = element_text(angle=50, size = 12, hjust = 1, vjust = 1),
    axis.text.y.left = element_text(size = 12, hjust = 1, vjust = 0.3),
    panel.border = element_rect(color = "black", fill = NA, size = 1),
    axis.ticks = element_line(color = "black", size = 1),
    axis.ticks.length = unit(1.1, 'mm'),
    panel.grid.major.y = element_line(color = "grey80", linetype = "solid", size = 1),
    panel.grid.minor.y = element_blank(),
    panel.grid.major.x = element_blank(),
    panel.grid.minor.x = element_blank(),
    legend.position = 'none',
    strip.background = element_blank(),
    strip.text = element_text(face = "bold.italic")
  )
```




```
ggsave(filename = "plots/fig3B_DEG_NASA_twin_study_comparison_selected_log2FC.pdf", width = 6, height = 8, units = "in", dpi = 300, device = cairo
```

Exports

```
openxlsx::write.xlsx(
  list(DEGs_DI_NASA_combined = DEGs_DI_NASA_combined),
  file = "tables/dataS7_DEGs_DI_NASA_combined.xlsx",
  rowNames=F,
  overwrite=T
)
```

SessionInfo

```
sessionInfo()
```

```
## R version 4.2.1 (2022-06-23 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19044)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.utf8
## [2] LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## attached base packages:
## [1] grid      stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
##  [1] ComplexHeatmap_2.14.0 openxlsx_4.2.5.1      RColorBrewer_1.1-3
##  [4] patchwork_1.1.2       magrittr_2.0.3        forcats_1.0.0
##  [7] stringr_1.5.0         dplyr_1.1.1           purrr_1.0.1
## [10] readr_2.1.4           tidyr_1.3.0           tibble_3.2.1
## [13] ggplot2_3.4.2         tidyverse_1.3.2
##
## loaded via a namespace (and not attached):
##  [1] httr_1.4.5          jsonlite_1.8.4      foreach_1.5.2
##  [4] modelr_0.1.11       highr_0.10          stats4_4.2.1
##  [7] googlesheets4_1.1.0 cellranger_1.1.0    yaml_2.3.7
## [10] pillar_1.9.0        backports_1.4.1     glue_1.6.2
## [13] digest_0.6.31       rvest_1.0.3         colorspace_2.1-0
## [16] htmltools_0.5.5     pkgconfig_2.0.3     GetoptLong_1.0.5
## [19] broom_1.0.4         haven_2.5.2         scales_1.2.1
## [22] tzdb_0.3.0          timechange_0.2.0    googledrive_2.1.0
## [25] farver_2.1.1        generics_0.1.3      IRanges_2.32.0
## [28] withr_2.5.0         BiocGenerics_0.44.0 cli_3.6.1
## [31] crayon_1.5.2        readxl_1.4.2        evaluate_0.20
## [34] fs_1.6.1            fansi_1.0.4         doParallel_1.0.17
## [37] xml2_1.3.3          tools_4.2.1         hms_1.1.3
## [40] GlobalOptions_0.1.2 gargle_1.3.0        lifecycle_1.0.3
## [43] matrixStats_0.63.0  S4Vectors_0.36.2    munsell_0.5.0
## [46] reprex_2.0.2        cluster_2.1.3       zip_2.2.2
## [49] compiler_4.2.1      rlang_1.1.0         iterators_1.0.14
## [52] rstudioapi_0.14     circlize_0.4.15     rjson_0.2.21
## [55] rmarkdown_2.21      gtable_0.3.3        codetools_0.2-18
## [58] DBI_1.1.3           R6_2.5.1            lubridate_1.9.2
## [61] knitr_1.42          fastmap_1.1.1       utf8_1.2.3
## [64] clue_0.3-64         shape_1.4.6         stringi_1.7.12
## [67] parallel_4.2.1      Rcpp_1.0.10         vctrs_0.6.1
## [70] png_0.1-8           dbplyr_2.3.2        tidyselect_1.2.0
## [73] xfun_0.38
```