Javi-plots-2

2022-04-29

Libraries

```
library("tidyverse")
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Loading Data

```
large_dataset <- read_tsv(file = "../data/02_large_w_meta_clean.tsv")
```

```
## Rows: 218 Columns: 25054
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr     (4): id, disease, sex, acc_num
## dbl (25050): age, 5S_rRNA, 7SK, 7SK:ENSG00000260682, A1BG, A1BG-AS1, A1CF, A...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
treatment_dataset <- read_tsv(file = "../data/02_treatment_w_meta_clean.tsv")
```

```
## Rows: 38 Columns: 25233
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr     (3): id, sex, acc_num
## dbl (25229): age, disease_duration, 5S_rRNA, 7SK, 7SK:ENSG00000260682, A1BG,...
## lgl     (1): treatment
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
logfold_treatment <- read_tsv(file = "../data/03_treatment_log2fc.tsv")
```

```
## Rows: 19 Columns: 25228
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr     (1): id
```

```
## dbl (25227): 5S_rRNA, 7SK, 7SK:ENSG00000260682, A1BG, A1BG-AS1, A1CF, A2M, A...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Fold change normal with normal mean

Using the mean of normal patients expression for each gene (Is this really an appropiate way to compare data?)

Calculating the mean

selecting only normal samples, transposing the dataset, calculating the mean for each gene

```
normal_dataset <- large_dataset %>%
  filter(grepl("normal", id)) %>%
  select(-disease, -acc_num, -sex, -age) %>%
  pivot_longer(cols = -c(id),
               names_to = "Genes" ,
               values_to = "Reads") %>%
  pivot_wider(names_from = id, values_from = Reads) %>%
  mutate(mean_reads = rowMeans(across(where(is.numeric))))
```

log fold change (using Mads code)

FC = Normal/Normal_mean

```
logfold_normal <- normal_dataset %>%
  rowwise() %>%
  mutate("Normal_2fc_1" = log2(normal_tissue_1+1)-log2(mean_reads+1),
         "Normal_2fc_2" = log2(normal_tissue_2+1)-log2(mean_reads+1),
         "Normal_2fc_3" = log2(normal_tissue_3+1)-log2(mean_reads+1),
         "Normal_2fc_4" = log2(normal_tissue_4+1)-log2(mean_reads+1),
         "Normal_2fc_5" = log2(normal_tissue_5+1)-log2(mean_reads+1),
         "Normal_2fc_6" = log2(normal_tissue_6+1)-log2(mean_reads+1),
         "Normal_2fc_7" = log2(normal_tissue_7+1)-log2(mean_reads+1),
         "Normal_2fc_8" = log2(normal_tissue_8+1)-log2(mean_reads+1),
         "Normal_2fc_9" = log2(normal_tissue_9+1)-log2(mean_reads+1),
         "Normal_2fc_10" = log2(normal_tissue_10+1)-log2(mean_reads+1),
         "Normal_2fc_11" = log2(normal_tissue_11+1)-log2(mean_reads+1),
         "Normal_2fc_12" = log2(normal_tissue_12+1)-log2(mean_reads+1),
         "Normal_2fc_13" = log2(normal_tissue_13+1)-log2(mean_reads+1),
         "Normal_2fc_14" = log2(normal_tissue_14+1)-log2(mean_reads+1),
         "Normal_2fc_15" = log2(normal_tissue_15+1)-log2(mean_reads+1),
         "Normal_2fc_16" = log2(normal_tissue_16+1)-log2(mean_reads+1),
         "Normal_2fc_17" = log2(normal_tissue_17+1)-log2(mean_reads+1),
         "Normal_2fc_18" = log2(normal_tissue_18+1)-log2(mean_reads+1),
         "Normal_2fc_19" = log2(normal_tissue_19+1)-log2(mean_reads+1),
         "Normal_2fc_20" = log2(normal_tissue_20+1)-log2(mean_reads+1),
         "Normal_2fc_21" = log2(normal_tissue_21+1)-log2(mean_reads+1),
         "Normal_2fc_22" = log2(normal_tissue_22+1)-log2(mean_reads+1),
         "Normal_2fc_23" = log2(normal_tissue_23+1)-log2(mean_reads+1),
         "Normal_2fc_24" = log2(normal_tissue_24+1)-log2(mean_reads+1),
         "Normal_2fc_25" = log2(normal_tissue_25+1)-log2(mean_reads+1),
```

```
         "Normal_2fc_26" = log2(normal_tissue_26+1)-log2(mean_reads+1),
         "Normal_2fc_27" = log2(normal_tissue_27+1)-log2(mean_reads+1),
         "Normal_2fc_28" = log2(normal_tissue_28+1)-log2(mean_reads+1)) %>%
  select(-starts_with("normal_t"), -mean_reads) %>%
  pivot_longer(cols = starts_with("Normal_2fc"), names_to = "id") %>%
  pivot_wider(names_from = "Genes", values_from = "value")
```

Fold change normal with baseline RA (RA_pre)

Join RA baseline and normal data

```
normal_mean <- normal_dataset %>%
  select(Genes, mean_reads) %>%
  pivot_longer(cols = -c(Genes),
               names_to = "id" ,
               values_to = "Reads") %>%
  relocate(id, .before = Genes)

aux_dataset <- treatment_dataset %>%
  select(-treatment, -sex, -age, -acc_num, -disease_duration) %>%
  filter(grepl("RA_pre", id)) %>%
  pivot_longer(cols = -c(id),
               names_to = "Genes" ,
               values_to = "Reads") %>%
  bind_rows(normal_mean) %>%
  pivot_wider(names_from = id, values_from = Reads)
```

log fold change (using Mads code)

FC = Baseline_RA/Normal

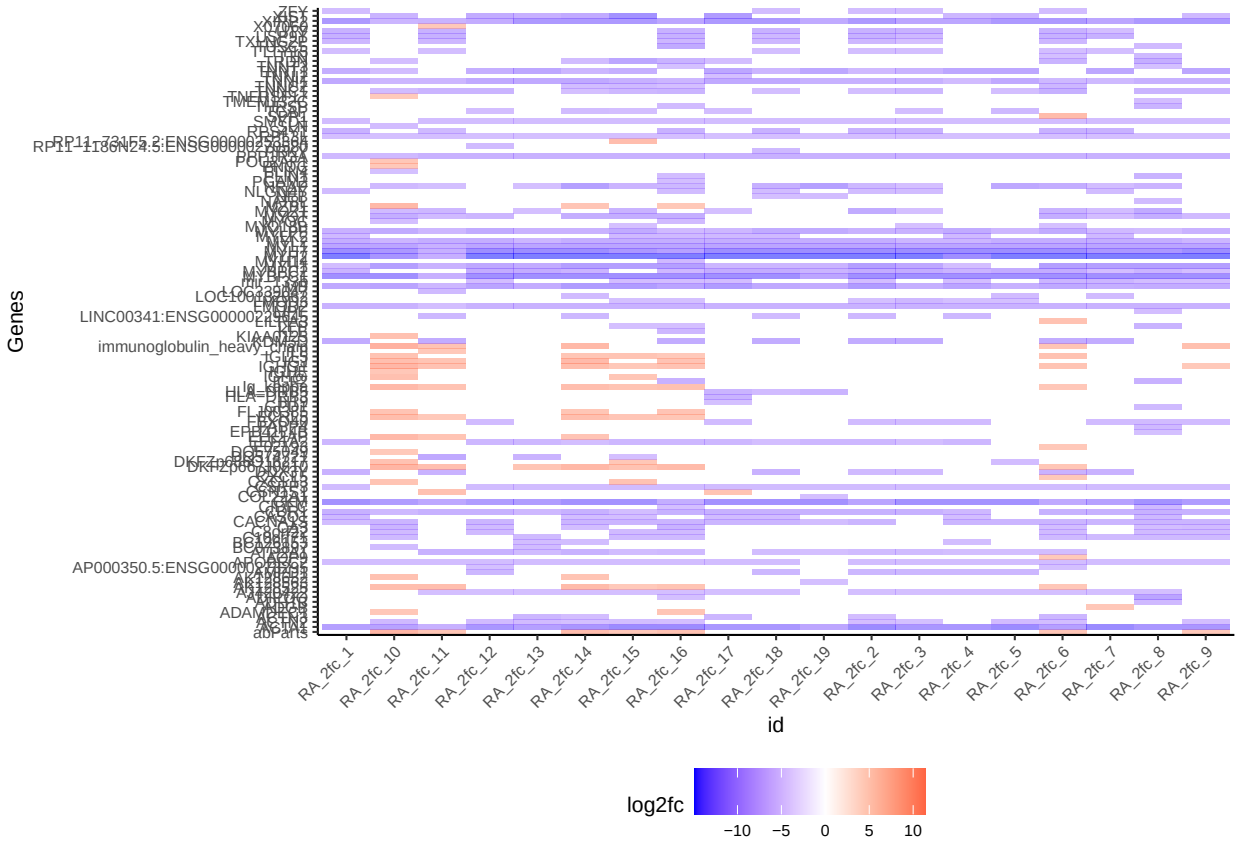
```
logfold_RA <- aux_dataset %>%
  rowwise() %>%
  mutate("RA_2fc_1" = log2(RA_pre_1+1)-log2(mean_reads+1),
         "RA_2fc_2" = log2(RA_pre_2+1)-log2(mean_reads+1),
         "RA_2fc_3" = log2(RA_pre_3+1)-log2(mean_reads+1),
         "RA_2fc_4" = log2(RA_pre_4+1)-log2(mean_reads+1),
         "RA_2fc_5" = log2(RA_pre_5+1)-log2(mean_reads+1),
         "RA_2fc_6" = log2(RA_pre_6+1)-log2(mean_reads+1),
         "RA_2fc_7" = log2(RA_pre_7+1)-log2(mean_reads+1),
         "RA_2fc_8" = log2(RA_pre_8+1)-log2(mean_reads+1),
         "RA_2fc_9" = log2(RA_pre_9+1)-log2(mean_reads+1),
         "RA_2fc_10" = log2(RA_pre_10+1)-log2(mean_reads+1),
         "RA_2fc_11" = log2(RA_pre_11+1)-log2(mean_reads+1),
         "RA_2fc_12" = log2(RA_pre_12+1)-log2(mean_reads+1),
         "RA_2fc_13" = log2(RA_pre_13+1)-log2(mean_reads+1),
         "RA_2fc_14" = log2(RA_pre_14+1)-log2(mean_reads+1),
         "RA_2fc_15" = log2(RA_pre_15+1)-log2(mean_reads+1),
         "RA_2fc_16" = log2(RA_pre_16+1)-log2(mean_reads+1),
         "RA_2fc_17" = log2(RA_pre_17+1)-log2(mean_reads+1),
         "RA_2fc_18" = log2(RA_pre_18+1)-log2(mean_reads+1),
         "RA_2fc_19" = log2(RA_pre_19+1)-log2(mean_reads+1)) %>%
  select(-starts_with("RA_p"), -mean_reads) %>%
```

```
  pivot_longer(cols = starts_with("RA_2fc"), names_to = "id") %>%
  pivot_wider(names_from = "Genes", values_from = "value")
```

Heatplot differential expression between Baseline RA and normal values

```
logfold_RA_long <- logfold_RA %>%
  pivot_longer(cols = -c(id),
               names_to = "Genes" ,
               values_to = "log2fc")

logfold_RA_long %>% filter(log2fc >= 8 | log2fc <= -8) %>%
  ggplot(mapping = aes(x = id,
                       y = Genes,
                       fill = log2fc )) +
  geom_tile(alpha = 0.5) +
  scale_fill_gradient2(low = "blue",
                       mid = "white",
                       high = "red",
                       midpoint = 0) +
  theme_classic(base_size = 8) +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle = 45,
                                   hjust = 1))
```



recreating Figure 2.B treatment paper

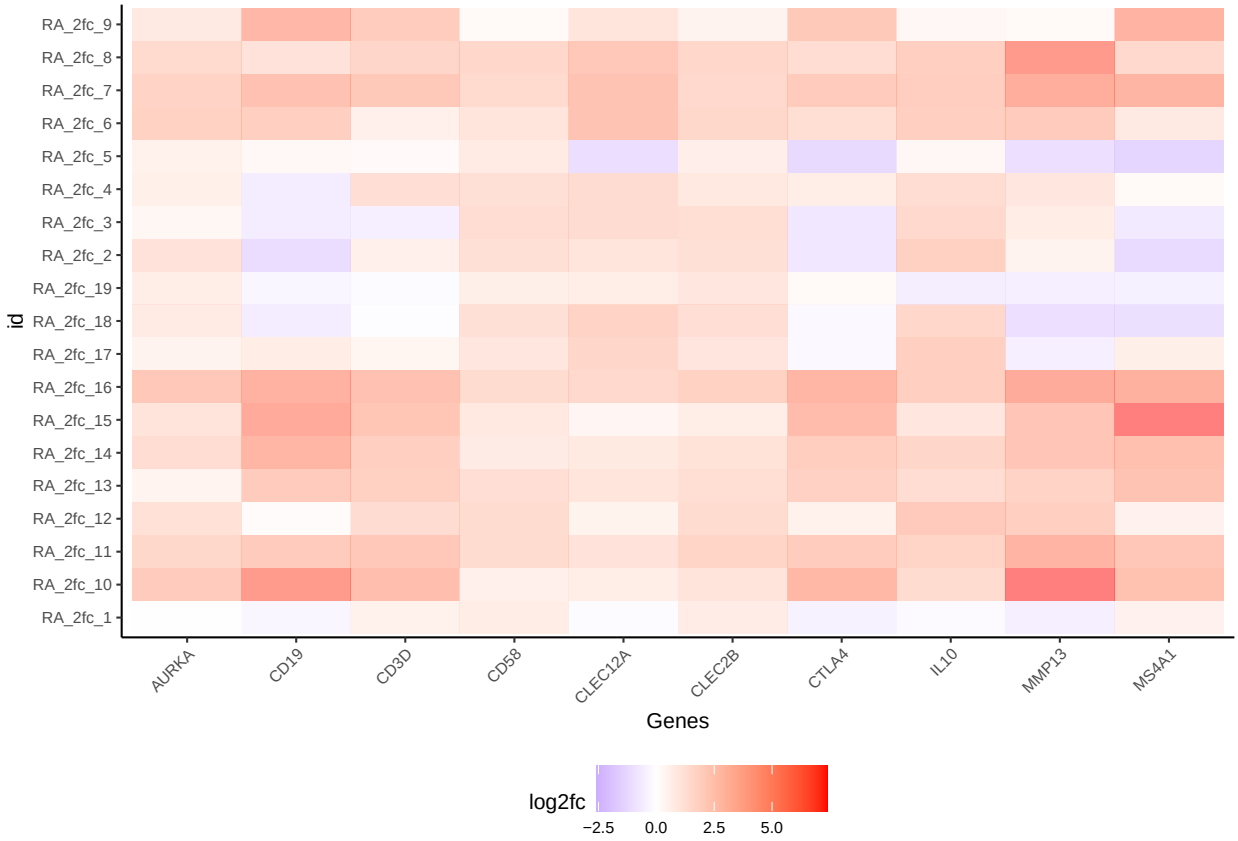
```
logfold_treatment <- logfold_treatment%>%
  mutate(Type = "RA post-tDMARD") %>%
  relocate(Type, .after = id)
```

recreating Figure 2.B and 2.C treatment paper

Just checking if the FC(baseline/normal) are upregulated as expected

```
aux_dataset <- logfold_RA %>%
  select(id, "CD3D", "CTLA4", "MS4A1", "CD19", "IL10", "MMP13", "CLEC12A", "CLEC2B",
         "AURKA", "CD58") %>%
  pivot_longer(cols = -c(id),
               names_to = "Genes" ,
               values_to = "log2fc")

aux_dataset %>%
  ggplot(mapping = aes(x = Genes,
                       y = id,
                       fill = log2fc )) +
  geom_tile(alpha = 0.5) +
  scale_fill_gradient2(low = "blue",
                        mid = "white",
                        high = "red",
                        midpoint = 0) +
  theme_classic(base_size = 8) +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle = 45,
                                   hjust = 1))
```


RA_2fc_9
RA_2fc_8
RA_2fc_7
RA_2fc_6
RA_2fc_5
RA_2fc_4
RA_2fc_3
RA_2fc_2
RA_2fc_19
RA_2fc_18
RA_2fc_17
RA_2fc_16
RA_2fc_15
RA_2fc_14
RA_2fc_13
RA_2fc_12
RA_2fc_11
RA_2fc_10
RA_2fc_1
id
AURKA
CD19
CD3D
CD58
CLEC12A
CLEC2B
CTLA4
IL10
MMP13
MS4A1
Genes
log2fc
−2.5
0.0
2.5
5.0