sldsc_volcano_fig3f

Kate Isaac

2022-09-23

```r
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.10
## v tidyr   1.2.1     v stringr 1.4.1
## v readr   2.1.2     v forcats 0.5.2
## -- Conflicts ---------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(ggplot2)
library(ggthemes)
library(ggrepel)
```

```r
full_enrichment_df = read.delim("~/mccoyLab/VISION/sldsc/withplotting_raw_cCRE_all_UKBB_traits.enrichmen

logged_cutoff <- 2.347514159253684
```

```r
full_enrichment_df$label <- full_enrichment_df$description
full_enrichment_df <- full_enrichment_df %>%
  mutate(across("label", str_replace, "White blood cell \\(leukocyte\\)", "WBC")) %>%
  mutate(across("label", str_replace, "Red blood cell \\(erythrocyte\\)", "RBC")) %>%
  mutate(across("label", str_replace, "haemoglobin", "Hb")) %>%
  mutate(across("label", str_replace, "Haemoglobin", "Hb")) %>%
  mutate(across("label", str_replace, "Neutrophill", "NEU")) %>%
  mutate(across("label", str_replace, "Platelet", "PLT")) %>%
  mutate(across("label", str_replace, "platelet \\(thrombocyte\\)", "PLT")) %>%
  mutate(across("label", str_replace, "Monocyte", "MON")) %>%
  mutate(across("label", str_replace, 'reticulocyte', "RTIC")) %>%
  mutate(across("label", str_replace, 'Reticulocyte', "RTIC")) %>%
  mutate(across("label", str_replace, "Lymphocyte", "LYM")) %>%
  mutate(across("label", str_replace, "Basophil", "BASO")) %>%
  mutate(across("label", str_replace, "Eosinophill", "EOS")) %>%
  mutate(across("label", str_replace, "Mean corpuscular volume", "MCV")) %>%
  mutate(across("label", str_replace, "Mean", "avg")) %>%
  mutate(across("label", str_replace, "percentage", "%")) %>%
  mutate(across("label", str_replace, "distribution", "dist")) %>%
  mutate(across("label", str_replace, "volume", "vol")) %>%
  mutate(across("label", str_replace, "concentration", "conc")) %>%
  mutate(across("label", str_replace, "count", "ct")) %>%
  mutate(across("label", str_replace, "Immature", "Im")) %>%
  mutate(across("label", str_replace, "Average", "avg")) %>%
```

```r
  mutate(across("label", str_replace, "Glycated haemoglobin (mmol/mol)", "Glycated Hb"))
```

add _f and _m for male and female to labels

```r
full_enrichment_df$Sex <- unlist(lapply(1:nrow(full_enrichment_df), function(x) strsplit(row.names(full_
full_enrichment_df <- full_enrichment_df %>%
  mutate(across("Sex", str_replace, "female", "f")) %>%
  mutate(across("Sex", str_replace, "male", "m")) %>%
  unite("full_label", c("label", "Sex"), remove = FALSE)
```

Keep certain labels nonsig: Longest period of depression, Vitamin D, Standing height, Age at death

Keep certain labels sig: Top 14, BASOl %

> colnames(full_enrichment_df) [1] "Enrichment" "Enrichment_p" "Enrichment_p_adj" "description"
> [5] "enrichments_list" "enrichment_ps_list" "colors" "logged_enrichments"
> [9] "logged_enrichments_ps" "full_label" "label" "Sex"

```r
notsig <- which(full_enrichment_df$logged_enrichments_ps < logged_cutoff)
#keep1 <- which(full_enrichment_df$label == "Longest period of depression")
keep2 <- which(full_enrichment_df$label == "Vitamin D")
keep3 <- which(full_enrichment_df$label == "Standing height")
keep4 <- which(full_enrichment_df$label == "Age at death")

sigenrichpanel <- which(full_enrichment_df$logged_enrichments_ps >= logged_cutoff & full_enrichment_df$

sig <- which(full_enrichment_df$logged_enrichments_ps >= logged_cutoff)
keepsig <- 1:17
keepsigpt2 <- which(full_enrichment_df$label == "BASOl %")
keepsigpt3 <- which(full_enrichment_df$label == "EOS %")
keepsigpt4 <- which(full_enrichment_df$label == "PLT dist width")
keepsigpt5 <- which(full_enrichment_df$label == "Glycated Hb")

full_enrichment_df$label[setdiff(notsig,c(keep2,keep3,keep4))] <- ""
full_enrichment_df$full_label[setdiff(notsig,c(keep2,keep3,keep4))] <- ""
full_enrichment_df$label[sigenrichpanel] <- ""
full_enrichment_df$full_label[sigenrichpanel] <- ""
full_enrichment_df$label[setdiff(sig, c(keepsig, keepsigpt2, keepsigpt3, keepsigpt4, keepsigpt5))] <- "
full_enrichment_df$full_label[setdiff(sig, c(keepsig, keepsigpt2, keepsigpt3, keepsigpt4, keepsigpt5))]
```

```r
g <- ggplot(full_enrichment_df, aes(x=logged_enrichments, y=logged_enrichments_ps, color=colors, label=
  geom_point(aes(shape = Sex), alpha=0.5, size=1) +
  geom_hline(yintercept = logged_cutoff, color = "gray", linetype = "dashed") +
  geom_vline(xintercept = 0, color = "gray", linetype = "dashed") +
  geom_text_repel(max.overlaps = 25, nudge_y = 0.75) +
  theme_bw() + theme(panel.background = element_blank(), panel.grid = element_blank()) +
  scale_color_manual(values = c("black" = "black", "red" = "red"), labels = c("not sig", "sig"), guide=
  geom_text(data=data.frame(), aes(label = '587 traits total', x = -5, y = 14),
            hjust = 0, vjust = 1, color = "black") +
  xlab('Log2(enrichment)') +
  ylab('Log10(p-value)')
g
```

```
## Warning: ggrepel: 2 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```
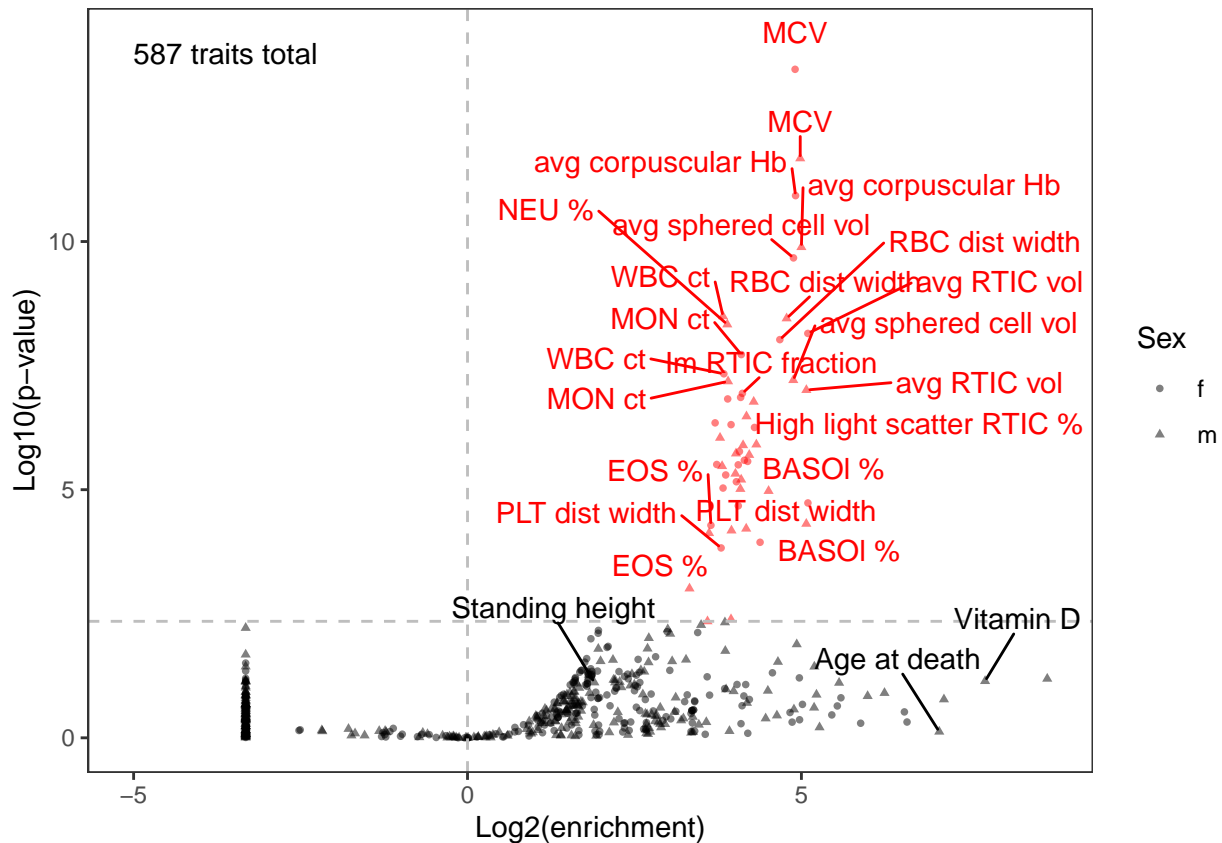
```
ggsave("~/mccoyLab/VISION/sldsc/enrichment_volcano_fmshape.png", plot=g)
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: ggrepel: 2 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

```
ggsave("~/mccoyLab/VISION/sldsc/enrichment_volcano_fmshape.pdf", plot=g)
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: ggrepel: 2 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```