ACE2 Regulatory Network Analysis

Tayaza Fadason [t.fadason@auckland.ac.nz]

April, 2020

Contents

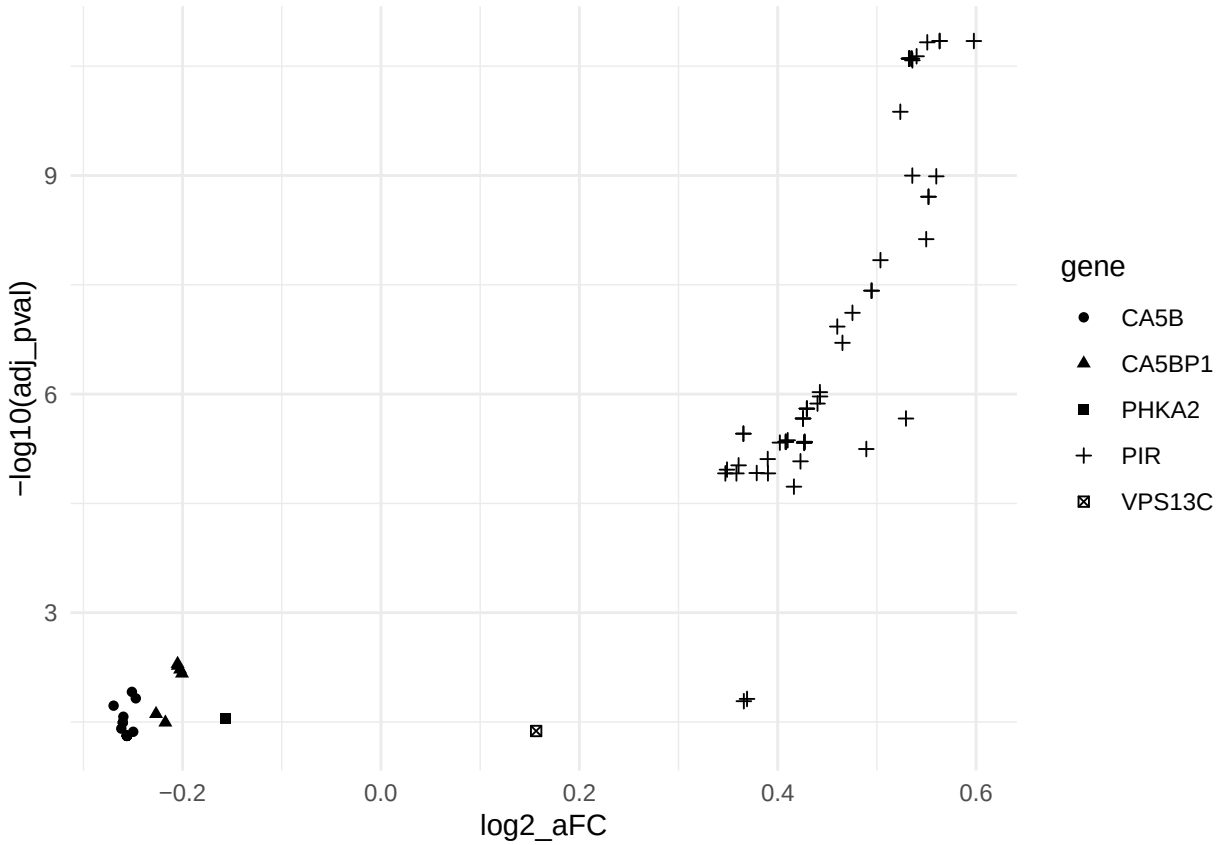
```
setwd('/Volumes/GoogleDrive/My\ Drive/covid19/github/scripts')
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.4
## v tidyr   1.0.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

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

```
lung <- read.csv('../data/significant_eqtls.txt', sep='\t', stringsAsFactors =F)

ggplot(lung,
       aes(x=log2_aFC, y=-log10(adj_pval), shape=gene))+
  geom_point()+
  theme_minimal()
```

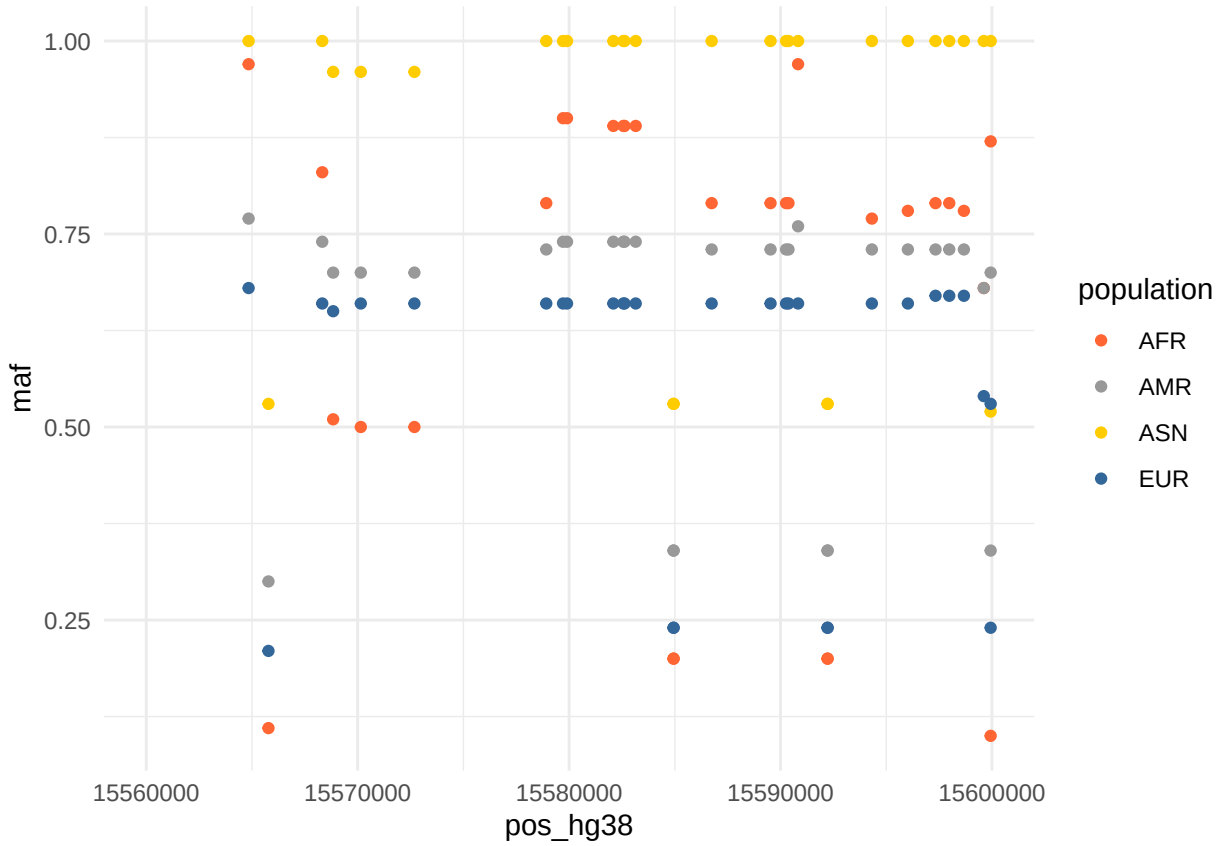



```
### Add the following lines at the beginning of the file produced by this chunk
#track type=interact name="Lung interactions" description="Lung chromatin interactions" useScore=on max
#browser position chrX:15,000,000-16,000,000

lung %>%
  dplyr::select(snp, gene, tissue, snp_chr, snp_locus, gene_chr, gene_start, gene_end, total_hic_score)
  mutate(
  chrom = paste0('chr', gene_chr),
  chromStart = case_when(
    ((as.integer(snp_chr) == as.integer(gene_chr)) & (as.integer(snp_locus)-1 < as.integer(gene_start))
    TRUE ~ as.integer(gene_start)
  ),
  chromEnd = case_when(
    ((as.integer(snp_chr) == as.integer(gene_chr)) & (as.integer(snp_locus) > as.integer(gene_start))) 
    TRUE ~ as.integer(gene_end)
  ),
  name = paste0(as.character(snp), '_', as.character(gene), '_', tissue),
  score = as.integer(0),
  value = as.numeric(total_hic_score),
  exp = tissue,
  color = case_when(
    tissue == 'Lung' ~ '#E41A1C',
    grepl(tolower('Heart'), tolower(tissue)) ~ '#377EB8',
    grepl(tolower('Artery'), tolower(tissue)) ~ '#4DAF4A',
    # grepl(tolower('Heart'), to.lower(tissue)) ~ '#4daf4a',
    TRUE ~ '#333333'
  ),
```

```
  sourceChrom = paste0('chr', snp_chr),
  sourceStart = as.integer(snp_locus-1),
  sourceEnd = as.integer(snp_locus),
  sourceName = snp,
  sourceStrand = as.character('.'),
  targetChrom = paste0('chr', gene_chr),
  targetStart = as.integer(gene_start),
  targetEnd = as.integer(gene_end),
  targetName = gene,
  targetStrand = as.character('.')
) %>%
dplyr::select(-snp, -gene, -tissue, -snp_chr, -snp_locus, -gene_chr, -gene_start, -gene_end, -total_hic
  arrange(exp) %>%
  dplyr::filter(exp == 'Lung') %>%
  rename(`#chrom`=chrom) %>%
  write_tsv('../data/interact.txt')
```

```
## Warning in eval_tidy(pair$lhs, env = default_env): NAs introduced by coercion

## Warning in eval_tidy(pair$lhs, env = default_env): NAs introduced by coercion

## Warning in eval_tidy(pair$lhs, env = default_env): NAs introduced by coercion

## Warning in eval_tidy(pair$lhs, env = default_env): NAs introduced by coercion
```

```
# Go to haploreg site https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php, load SNP list, sel
#Original file downloaded 04-04-20

haploreg <- read.csv('../data/haploreg.txt', sep='\t', stringsAsFactors =F)

mafs <- haploreg %>%
  filter(rsID %in% lung$snp) %>%
  select(rsID, pos_hg38, AFR,   AMR,    ASN,    EUR) %>%
   pivot_longer(AFR:EUR, names_to="population", values_to="maf")


ggplot(mafs %>% filter(pos_hg38>=15560000 & pos_hg38<=15600000), # Plot only eQTLs within the ACE2 gene
       aes(x=pos_hg38, y=maf, colour=population))+
  geom_point()+
  scale_x_continuous(limits=c(15560000,15600000))+
  scale_color_manual(values=c("#ff6633", "#999999", "#ffcc00", "#336699"))+
  theme_minimal()
```




```
# Extract TF binding sites overlapping eQTLs

motifs <- haploreg %>%
  select(rsID, Chromatin_Marks, Motifs, GENCODE_name, dbSNP_functional_annotation, query_snp_rsid, pos_
  dplyr::filter(as.character(rsID) == as.character(query_snp_rsid)) %>%
  select(rsID, Motifs, GENCODE_name, pos_hg38, dbSNP_functional_annotation) %>%
  separate_rows(Motifs, sep=";") %>%
  separate_rows(Motifs, sep="_") %>%
  dplyr::filter(Motifs != "." & is.na(as.integer(Motifs)) & !(grepl('disc', Motifs)) & !(grepl('known',
```

```
## Warning: NAs introduced by coercion
```

```
 merge(lung %>% select(snp,gene) %>% distinct(), motifs, by.x="snp", by.y="rsID") %>%
  # filter(gene=='PIR' & GENCODE_name=='ACE2') %>%
  distinct() %>%
  arrange(Motifs) %>%
  select(snp, gene, Motifs, pos_hg38, dbSNP_functional_annotation, GENCODE_name) %>%
  rename(motif=Motifs, pos=pos_hg38, motif_gene=GENCODE_name) %>%
  write_tsv('../data/motifs.txt')
```