

main

October 1, 2021

1 Exploring the eigen values and correlation with phenotypes

```
[1]: suppressPackageStartupMessages({  
      library(ComplexHeatmap)  
      library(tidyverse)  
      library(ggpubr)  
    })
```

1.1 Prepare data

1.1.1 Load phenotypes data

```
[2]: pheno = data.table::fread("../.../input/phenotypes/merged/_m/  
      ↪merged_phenotypes.csv")  
pheno %>% head(2)
```

	V1	BrNum	RNum	Region	RIN	Age	Sex	Race	Dx	mito
	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<chr>	<chr>	<chr>	<dbl>
A data.table: 2 × 12	R12864	Br1303	R12864	Caudate	9.6	42.98	F	AA	Schizo	0.032
	R12865	Br1320	R12865	Caudate	9.5	53.12	M	AA	Schizo	0.019

```
[3]: ancestry = data.table::fread("../.../input/ancestry_structure/structure.  
      ↪out_ancestry_proportion_raceDemo_compare")  
ancestry %>% head(2)
```

	id	Afr	Eur	group
	<chr>	<dbl>	<dbl>	<chr>
A data.table: 2 × 4	Br2374	0.007	0.993	CAUC
	Br1857	0.001	0.999	CAUC

1.1.2 Load eigen values

```
[4]: eigen = data.table::fread("../.../_m/eigengenes.csv")  
modules = eigen %>% select(-V1) %>% colnames  
eigen %>% dim  
eigen[1:2, 1:5]
```

1. 233 2. 39

	V1	MEblack	MEblue	MEbrown	MEcyan
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
A data.table: 2 × 5	R11135	-0.002453113	0.12140793	-0.08433039	0.14720121
	R11137	-0.056503053	-0.04728863	0.02602025	0.07250302

1.1.3 Merge data

```
[5]: dt = eigen %>% left_join(pheno, by="V1") %>%
      inner_join(ancestry, by=c("BrNum"="id"))
dt %>% dim
dt[1:2, 1:5]
```

1. 233 2. 53

	V1	MEblack	MEblue	MEbrown	MEcyan
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
A data.table: 2 × 5	R11135	-0.002453113	0.12140793	-0.08433039	0.14720121
	R11137	-0.056503053	-0.04728863	0.02602025	0.07250302

1.2 Examine correlation of modules with ancestry

1.2.1 Linear model

```
[6]: pvals = c()
for(mod in modules){
  model = paste0("Eur ~ ", mod)
  res = anova(lm(model, data=dt))
  pvals = c(pvals, res[mod, "Pr(>F)"])
}
fdr <- p.adjust(pvals, method="fdr")
df1 = data.frame("Modules"=modules, "Pvalue"=pvals, "FDR"=fdr)
df1 %>% filter(FDR < 0.05)
```

	Modules	Pvalue	FDR
	<chr>	<dbl>	<dbl>
A data.frame: 1 × 3	MEbrown	0.0009956639	0.03783523

```
[7]: df1 %>% mutate(Tissue="Hippocampus") %>%
      data.table::fwrite("eigen_correlation_ancestry.tsv", sep='\t')
```

1.2.2 Pearson correlation

```
[8]: pvals = c(); est = c()
for(mod in modules){
  res = cor.test(dt[["Eur"]], dt[[mod]], method="pearson")
  pvals = c(pvals, res$p.value)
  est = c(est, res$estimate[[1]])
}
fdr <- p.adjust(pvals, method="fdr")
df2 = data.frame("Modules"=modules, "Rho"=est, "Pvalue"=pvals, "FDR"=fdr)
```

```
df2 %>% filter(Pvalue < 0.05)
```

	Modules <chr>	Rho <dbl>	Pvalue <dbl>	FDR <dbl>
A data.frame: 4 × 4	MEbrown	0.2142915	0.0009956639	0.03783523
	MEgrey	-0.1887614	0.0038296291	0.07276295
	MEroyalblue	-0.1439254	0.0280519150	0.26649319
	MEsaddlebrown	0.1473345	0.0245025391	0.26649319

1.3 Clustering modules

1.3.1 Eigengene clustering

```
[9]: set.seed(13)
mat = eigen %>% column_to_rownames("V1") %>% as.matrix
pdf("eigengene_heatmap.pdf", width=10, height=8)
Heatmap(mat, name="Eigengene", column_dend_height=unit(3, "cm"),
        show_row_names=FALSE, show_row_dend=FALSE, column_km=4)
dev.off()
```

png: 2

1.3.2 clustering within significant module violet

```
[10]: load("../_m/01.RData", verbose=TRUE)
biomart_file = "../input/biomart/biomart.csv"
biomart = data.table::fread(biomart_file)
```

Loading objects:

```
datExpr
sample_table
datTraits
```

```
[11]: get_module_mat <- function(mod){
  MEmodule = data.table::fread("../_m/modules.csv", header=TRUE) %>%
    filter(module == mod)
  mat0 = datExpr %>% as.data.frame %>%
    select(any_of(MEmodule$V1)) %>% as.matrix
  annot = data.frame("Geneid"=colnames(mat0),
                    "ensembl_gene_id"=gsub("\\.*", "", colnames(mat0))) %>%
    left_join(biomart, by="ensembl_gene_id") %>%
    distinct(ensembl_gene_id, .keep_all=TRUE) %>%
    mutate(external_gene_name=coalesce(external_gene_name,ensembl_gene_id))
  #annot$external_gene_name %>% is.na %>% sum
  colnames(mat0) <- annot$external_gene_name
  return(mat0)
}

plot_module_heatmap <- function(mod, w, h){
```

```
pdf(paste0(mod, "_module_heatmap.pdf"), width=w, height=h)
set.seed(13)
row_ha = rowAnnotation(EA = dt$Eur, AA = dt$Afr)
ht = Heatmap(get_module_mat(mod),
              name="Residualized\nExpression",
              column_dend_height=unit(1, "cm"), row_km=2,
              show_row_names=FALSE, show_row_dend=TRUE,
              row_dend_width=unit(2, "cm"),
              right_annotation = row_ha)
draw(ht)
dev.off()
}
```

```
[12]: df1 %>% filter(FDR < 0.05)
```

	Modules	Pvalue	FDR
A data.frame: 1 × 3	<chr>	<dbl>	<dbl>
	MEbrown	0.0009956639	0.03783523

```
[13]: plot_module_heatmap("brown", 24, 12)
```

png: 2

1.4 Reproducibility Information

```
[14]: Sys.time()
proc.time()
options(width = 120)
sessioninfo::session_info()
```

```
[1] "2021-10-01 13:42:53 EDT"
```

```
  user  system elapsed
15.670   1.090   17.111
```

```
Session info
```

```
setting  value
```

```
version  R version 4.0.3 (2020-10-10)
```

```
os       Arch Linux
```

```
system   x86_64, linux-gnu
```

```
ui       X11
```

```
language (EN)
```

```
collate  en_US.UTF-8
```

```
ctype    en_US.UTF-8
```

```
tz       America/New_York
```

```
date     2021-10-01
```

```
Packages
```

```
package      * version  date      lib source
```

```
abind         1.4-5    2016-07-21 [1] CRAN (R 4.0.2)
```

assertthat	0.2.1	2019-03-21	[1]	CRAN	(R 4.0.2)
backports	1.2.1	2020-12-09	[1]	CRAN	(R 4.0.2)
base64enc	0.1-3	2015-07-28	[1]	CRAN	(R 4.0.2)
BiocGenerics	0.36.1	2021-04-16	[1]	Bioconductor	
broom	0.7.9	2021-07-27	[1]	CRAN	(R 4.0.3)
Cairo	1.5-12.2	2020-07-07	[1]	CRAN	(R 4.0.2)
car	3.0-11	2021-06-27	[1]	CRAN	(R 4.0.3)
carData	3.0-4	2020-05-22	[1]	CRAN	(R 4.0.2)
cellranger	1.1.0	2016-07-27	[1]	CRAN	(R 4.0.2)
circlize	0.4.13	2021-06-09	[1]	CRAN	(R 4.0.3)
cli	3.0.1	2021-07-17	[1]	CRAN	(R 4.0.3)
clue	0.3-59	2021-04-16	[1]	CRAN	(R 4.0.3)
cluster	2.1.0	2019-06-19	[2]	CRAN	(R 4.0.3)
colorspace	2.0-2	2021-06-24	[1]	CRAN	(R 4.0.3)
ComplexHeatmap	* 2.6.2	2020-11-12	[1]	Bioconductor	
crayon	1.4.1	2021-02-08	[1]	CRAN	(R 4.0.3)
curl	4.3.2	2021-06-23	[1]	CRAN	(R 4.0.3)
data.table	1.14.2	2021-09-27	[1]	CRAN	(R 4.0.3)
DBI	1.1.1	2021-01-15	[1]	CRAN	(R 4.0.2)
dbplyr	2.1.1	2021-04-06	[1]	CRAN	(R 4.0.3)
digest	0.6.28	2021-09-23	[1]	CRAN	(R 4.0.3)
dplyr	* 1.0.7	2021-06-18	[1]	CRAN	(R 4.0.3)
ellipsis	0.3.2	2021-04-29	[1]	CRAN	(R 4.0.3)
evaluate	0.14	2019-05-28	[1]	CRAN	(R 4.0.2)
fansi	0.5.0	2021-05-25	[1]	CRAN	(R 4.0.3)
fastmap	1.1.0	2021-01-25	[1]	CRAN	(R 4.0.2)
forcats	* 0.5.1	2021-01-27	[1]	CRAN	(R 4.0.2)
foreign	0.8-80	2020-05-24	[2]	CRAN	(R 4.0.3)
fs	1.5.0	2020-07-31	[1]	CRAN	(R 4.0.2)
generics	0.1.0	2020-10-31	[1]	CRAN	(R 4.0.2)
GetoptLong	1.0.5	2020-12-15	[1]	CRAN	(R 4.0.2)
ggplot2	* 3.3.5	2021-06-25	[1]	CRAN	(R 4.0.3)
ggpubr	* 0.4.0	2020-06-27	[1]	CRAN	(R 4.0.2)
ggsignif	0.6.3	2021-09-09	[1]	CRAN	(R 4.0.3)
GlobalOptions	0.1.2	2020-06-10	[1]	CRAN	(R 4.0.2)
glue	1.4.2	2020-08-27	[1]	CRAN	(R 4.0.2)
gtable	0.3.0	2019-03-25	[1]	CRAN	(R 4.0.2)
haven	2.4.3	2021-08-04	[1]	CRAN	(R 4.0.3)
hms	1.1.1	2021-09-26	[1]	CRAN	(R 4.0.3)
htmltools	0.5.2	2021-08-25	[1]	CRAN	(R 4.0.3)
httr	1.4.2	2020-07-20	[1]	CRAN	(R 4.0.2)
IRanges	2.24.1	2020-12-12	[1]	Bioconductor	
IRdisplay	1.0	2021-01-20	[1]	CRAN	(R 4.0.2)
IRkernel	1.2	2021-05-11	[1]	CRAN	(R 4.0.3)
jsonlite	1.7.2	2020-12-09	[1]	CRAN	(R 4.0.2)
lifecycle	1.0.1	2021-09-24	[1]	CRAN	(R 4.0.3)
lubridate	1.7.10	2021-02-26	[1]	CRAN	(R 4.0.3)
magrittr	2.0.1	2020-11-17	[1]	CRAN	(R 4.0.2)

matrixStats	0.61.0	2021-09-17	[1]	CRAN	(R 4.0.3)
modelr	0.1.8	2020-05-19	[1]	CRAN	(R 4.0.2)
munsell	0.5.0	2018-06-12	[1]	CRAN	(R 4.0.2)
openxlsx	4.2.4	2021-06-16	[1]	CRAN	(R 4.0.3)
pbdZMQ	0.3-5	2021-02-10	[1]	CRAN	(R 4.0.3)
pillar	1.6.3	2021-09-26	[1]	CRAN	(R 4.0.3)
pkgconfig	2.0.3	2019-09-22	[1]	CRAN	(R 4.0.2)
png	0.1-7	2013-12-03	[1]	CRAN	(R 4.0.2)
purrr	* 0.3.4	2020-04-17	[1]	CRAN	(R 4.0.2)
R6	2.5.1	2021-08-19	[1]	CRAN	(R 4.0.3)
RColorBrewer	1.1-2	2014-12-07	[1]	CRAN	(R 4.0.2)
Rcpp	1.0.7	2021-07-07	[1]	CRAN	(R 4.0.3)
readr	* 2.0.2	2021-09-27	[1]	CRAN	(R 4.0.3)
readxl	1.3.1	2019-03-13	[1]	CRAN	(R 4.0.2)
repr	1.1.3	2021-01-21	[1]	CRAN	(R 4.0.2)
reprex	2.0.1	2021-08-05	[1]	CRAN	(R 4.0.3)
rio	0.5.27	2021-06-21	[1]	CRAN	(R 4.0.3)
rjson	0.2.20	2018-06-08	[1]	CRAN	(R 4.0.2)
rlang	0.4.11	2021-04-30	[1]	CRAN	(R 4.0.3)
rstatix	0.7.0	2021-02-13	[1]	CRAN	(R 4.0.3)
rstudioapi	0.13	2020-11-12	[1]	CRAN	(R 4.0.2)
rvest	1.0.1	2021-07-26	[1]	CRAN	(R 4.0.3)
S4Vectors	0.28.1	2020-12-09	[1]	Bioconductor	
scales	1.1.1	2020-05-11	[1]	CRAN	(R 4.0.2)
sessioninfo	1.1.1	2018-11-05	[1]	CRAN	(R 4.0.2)
shape	1.4.6	2021-05-19	[1]	CRAN	(R 4.0.3)
stringi	1.7.4	2021-08-25	[1]	CRAN	(R 4.0.3)
stringr	* 1.4.0	2019-02-10	[1]	CRAN	(R 4.0.2)
tibble	* 3.1.4	2021-08-25	[1]	CRAN	(R 4.0.3)
tidyr	* 1.1.4	2021-09-27	[1]	CRAN	(R 4.0.3)
tidyselect	1.1.1	2021-04-30	[1]	CRAN	(R 4.0.3)
tidyverse	* 1.3.1	2021-04-15	[1]	CRAN	(R 4.0.3)
tzdb	0.1.2	2021-07-20	[1]	CRAN	(R 4.0.3)
utf8	1.2.2	2021-07-24	[1]	CRAN	(R 4.0.3)
uuid	0.1-4	2020-02-26	[1]	CRAN	(R 4.0.2)
vctrs	0.3.8	2021-04-29	[1]	CRAN	(R 4.0.3)
withr	2.4.2	2021-04-18	[1]	CRAN	(R 4.0.3)
xml2	1.3.2	2020-04-23	[1]	CRAN	(R 4.0.2)
zip	2.2.0	2021-05-31	[1]	CRAN	(R 4.0.3)

[1] /home/jbenja13/R/x86_64-pc-linux-gnu-library/4.0

[2] /usr/lib/R/library