

main

July 9, 2021

1 Boxplot of top five up- and downregulated genes

```
[1]: library(ggpubr)
library(ggsignif)
library(tidyverse)
```

Loading required package: ggplot2

```
Attaching packages: tidyverse
1.3.1
```

```
tibble 3.1.2    dplyr  1.0.7
tidyr  1.1.3    stringr 1.4.0
readr  1.4.0    forcats 0.5.1
purrr  0.3.4
```

Conflicts

```
tidyverse_conflicts()
dplyr::filter() masks stats::filter()
dplyr::lag()    masks stats::lag()
```

1.1 Functions

```
[2]: add_symnum <- function(res){
  symnum.args <- list(cutpoints = c(0, 0.0001, 0.001, 0.01, 0.05, 1),
    symbols = c("****", "***", "**", "*", "ns"))
  symnum.args$x <- res$adj.P.Val
  pvalue.signif <- do.call(stats::symnum, symnum.args) %>%
    as.character()
  pvalue.format <- format.pval(res$adj.P.Val, digits = 2)
  res <- res %>%
    dplyr::ungroup() %>%
    mutate(FDR = pvalue.format, p.signif = pvalue.signif)
  return(res)
}

save_ggplots <- function(fn, p, w, h){
```

```

for(ext in c('.pdf', '.png', '.svg')){
  ggsave(paste0(fn, ext), plot=p, width=w, height=h)
}
}

```

1.2 All genes

1.2.1 Differential expressed genes

```

[3]: genes = read.delim('../_m/genes/diffExpr_maleVfemale_full.txt')
genes = genes %>% filter(adj.P.Val < 0.05) %>%
  select(gencodeID, ensemblID, Symbol, logFC, t, adj.P.Val)
genes = add_symnum(genes)

up_genes = genes %>% filter(t > 0) %>% mutate('Direction'='Male Bias') %>%
  head(5)
down_genes = genes %>% filter(t < 0) %>% mutate('Direction'='Female Bias') %>%
  head(5)

sig_genes = bind_rows(up_genes, down_genes) %>% mutate_at(vars(Direction), as.
  factor)

for(xx in seq_along(sig_genes$Symbol)){
  sig_genes$New_ID[xx] <- ifelse(sig_genes$Symbol[xx] == '',
    as.character(sig_genes$ensemblID[xx]),
    as.character(sig_genes$Symbol[xx]))
}

sig_genes

```

A data.frame: 10 × 10

	gencodeID <chr>	ensemblID <chr>	Symbol <chr>	logFC <dbl>	t <dbl>	adj.P.Val <dbl>
	ENSG00000229236.1	ENSG00000229236	TTY10	6.9199039	87.48145	5.18666
	ENSG00000154620.5	ENSG00000154620	TMSB4Y	7.0178451	84.48351	4.94205
	ENSG00000226555.1	ENSG00000226555	AGKP1	7.0831120	83.11117	9.80757
	ENSG00000176728.7	ENSG00000176728	TTY14	8.0904908	80.49290	4.89560
	ENSG00000260197.1	ENSG00000260197		6.3029086	79.24522	8.72681
	ENSG00000229807.10	ENSG00000229807	XIST	-9.5895959	-68.61006	1.03474
	ENSG00000005889.15	ENSG00000005889	ZFX	-0.6193958	-29.51527	6.73309
	ENSG00000130021.13	ENSG00000130021	PUDP	-0.5567180	-20.79622	4.14648
	ENSG00000225470.6	ENSG00000225470	JPX	-0.5877969	-19.78297	6.53551
	ENSG00000126012.11	ENSG00000126012	KDM5C	-0.4398548	-18.17526	3.13408

1.2.2 Load residualized expression

```
[4]: res_file = '../_m/genes/residualized_expression.tsv'
tmp_df = data.table::fread(res_file, data.table=F)

resdf0 = tmp_df %>% filter(V1 %in% sig_genes$gencodeID) %>%
  ↪column_to_rownames("V1")
resdf0 <- data.frame(t(resdf0))
resdf0[1:2, 1:5]
```

Warning message in data.table::fread(res_file, data.table = F):
 "Detected 375 column names but the data has 376 columns (i.e. invalid file).
 Added 1 extra default column name for the first column which is guessed to be
 row names or an index. Use setnames() afterwards if this guess is not correct,
 or fix the file write command that created the file to create a valid file."

A data.frame: 2 × 5		ENSG00000130021.13	ENSG00000005889.15	ENSG00000126012.11	ENSG00000126012.11
		<dbl>	<dbl>	<dbl>	<dbl>
	R11135	-0.5047014	-0.2673090	-0.36192377	-0.372271
	R11137	0.1769958	0.3478564	-0.08106981	-0.569806

1.2.3 Load pheno data

```
[5]: pheno_file = '/ceph/projects/v3_phase3_paper/inputs/phenotypes/_m/
  ↪hippo_phenotypes.csv'
pheno <- data.table::fread(pheno_file) %>%
  mutate_if(is.character, as.factor) %>%
  column_to_rownames("V1")
levels(pheno$Sex) <- c("Female", "Male")
head(pheno, 2)
```

A data.frame: 2 × 8		BrNum	RNum	Region	RIN	Age	Sex	Race	Dx
		<fct>	<fct>	<fct>	<dbl>	<dbl>	<fct>	<fct>	<fct>
	R11135	Br2063	R11135	HIPPO	5.9	18.77	Male	CAUC	Control
	R11137	Br2582	R11137	HIPPO	9.2	41.44	Male	CAUC	Control

1.2.4 Merge dataframe

```
[6]: resdf <- inner_join(rownames_to_column(pheno),
  rownames_to_column(resdf0),
  by="rowname")
dim(resdf)
resdf[1:2, 1:10]
```

1. 375 2. 19

A data.frame: 2 × 10		rowname	BrNum	RNum	Region	RIN	Age	Sex	Race	Dx
		<chr>	<fct>	<fct>	<fct>	<dbl>	<dbl>	<fct>	<fct>	<fct>
	1	R11135	Br2063	R11135	HIPPO	5.9	18.77	Male	CAUC	Control
	2	R11137	Br2582	R11137	HIPPO	9.2	41.44	Male	CAUC	Control

1.2.5 Melt data frame

```
[7]: df = resdf %>% select(c('rowname', 'Sex', starts_with('ENSG'))) %>%
  pivot_longer(-c(rowname, Sex), names_to = "gencodeID", values_to = "Res")
  #>%
  inner_join(sig_genes, by='gencodeID') %>% mutate_at(vars("New_ID",
  #>"Symbol", "gencodeID"), as.factor)

df$New_ID = with(df, reorder(New_ID, adj.P.Val, median))
head(df, 2)
```

A tibble: 2 × 13

	rowname	Sex	gencodeID	Res	ensemblID	Symbol	logFC
	<chr>	<fct>	<fct>	<dbl>	<chr>	<fct>	<dbl>
	R11135	Male	ENSG00000130021.13	-0.5047014	ENSG00000130021	PUDP	-0.5567
	R11135	Male	ENSG00000005889.15	-0.2673090	ENSG00000005889	ZFX	-0.6193

1.2.6 Initial ggplot with ggpubr

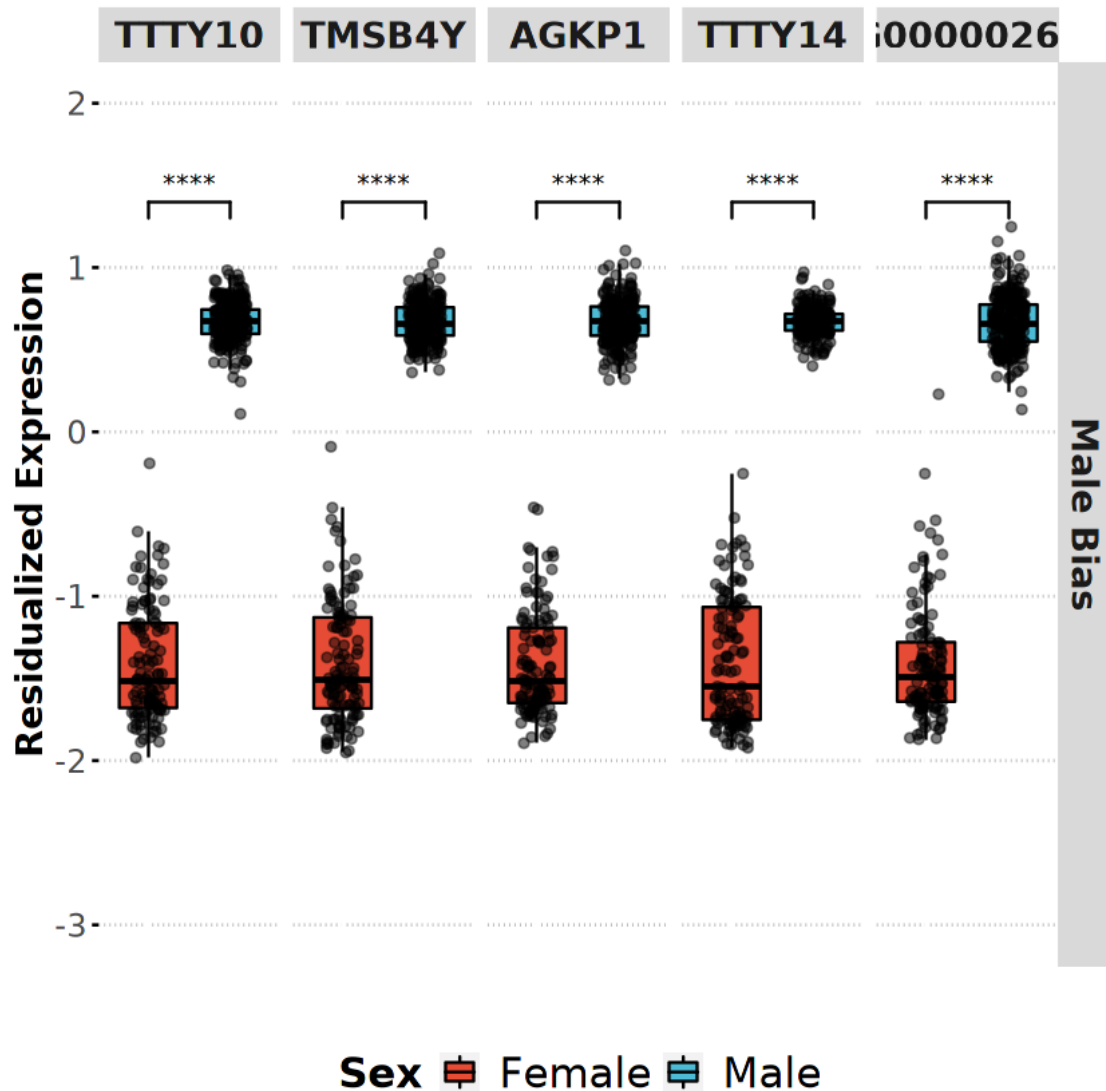
```
[8]: tmp = sig_genes %>% mutate(group1='Female', group2='Male', y_pos=1.4) %>%
  #>filter(Direction == 'Male Bias') %>%
  mutate_if(is.character, as.factor)

bxp_up <- df %>% filter(Direction=="Male Bias") %>%
  ggboxplot(x="Sex", y="Res", facet.by=c("Direction", "New_ID"),
  #>fill="Sex", xlab='',
  panel.labs.font=list(face='bold', size = 16), palette="npg",
  #>outlier.shape=NA,
  ylab='Residualized Expression', add='jitter', ylim=c(-3, 2),
  add.params=list(alpha=0.5), legend="bottom",
  #>ggtheme=theme_pubclean()) +
  geom_signif(data=tmp,
  aes(xmin=group1, xmax=group2, annotations=p.signif,
  #>y_position = y_pos),
  manual=TRUE) +
  font("xy.title", size=16, face="bold") + font("xy.text", size=14) +
  font("legend.title", size=18, face="bold") +
  font("legend.text", size=18) + rremove("x.text") + rremove("x.ticks")

bxp_up
```

Warning message:

"Ignoring unknown aesthetics: xmin, xmax, annotations, y_position"



```
[9]: tmp = sig_genes %>% mutate(group1='Female', group2='Male', y_pos=3.4) %>%
  ↳ filter(Direction == 'Female Bias') %>%
    mutate_if(is.character, as.factor)

bxp_down <- df %>% filter(Direction=='Female Bias') %>%
  ggboxplot(x="Sex", y="Res", fill="Sex", palette="npg", xlab='', outlier.
  ↳ shape=NA,
            facet.by=c("Direction", "New_ID"), panel.labs.
  ↳ font=list(face='bold', size = 16),
            ylab='Residualized Expression', add='jitter', ylim=c(-3, 4),
            add.params=list(alpha=0.5), legend="bottom",
  ↳ ggtheme=theme_pubclean()) +
```

```

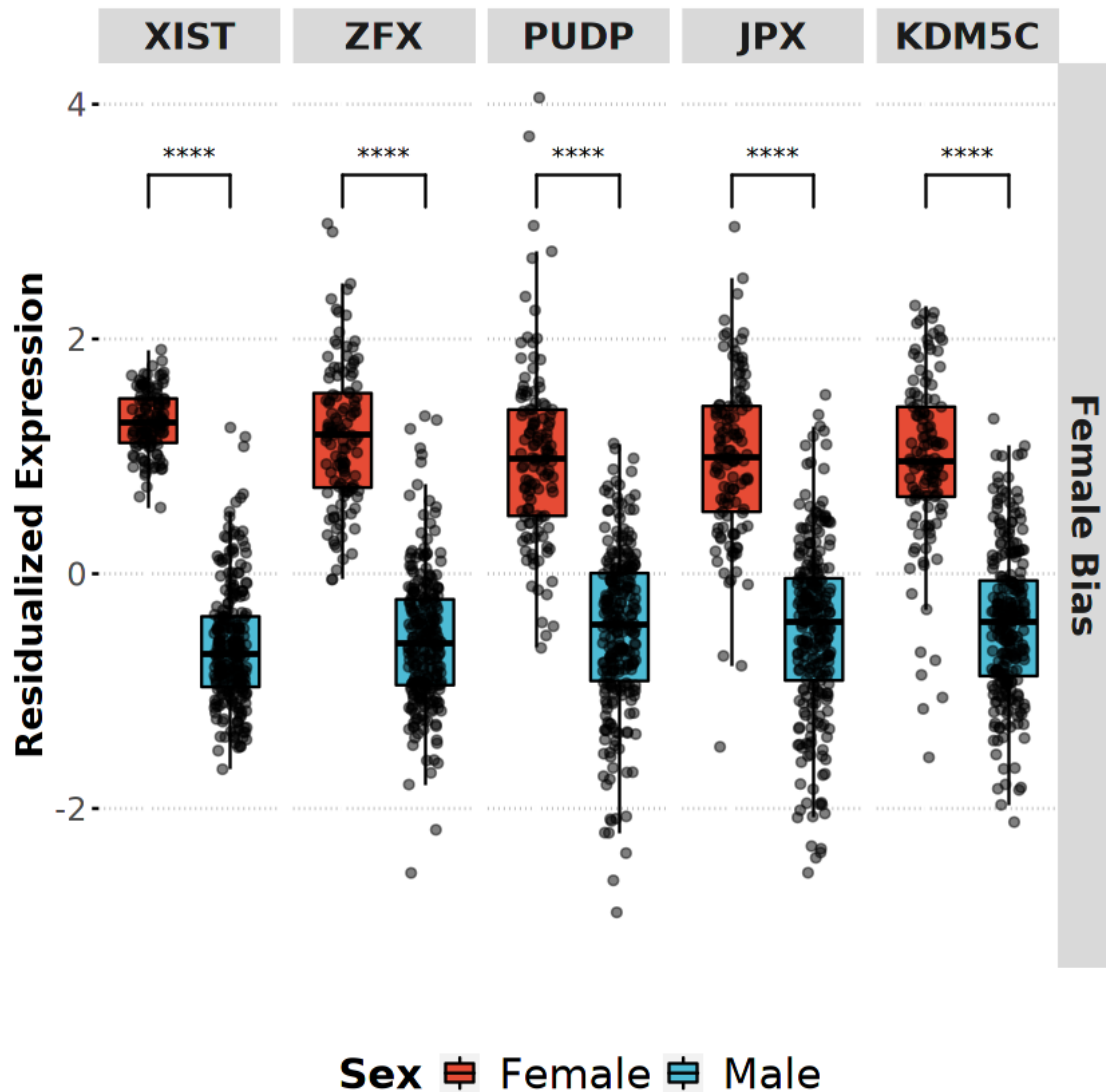
geom_signif(data=tmp,
            aes(xmin=group1, xmax=group2, annotations=p.signif,
→ y_position = y_pos),
            manual=TRUE) +
font("xy.title", size=16, face="bold") + font("xy.text", size=14) +
font("legend.title", size=18, face="bold") +
font("legend.text", size=18) + rremove("x.text") + rremove("x.ticks")

```

bxp_down

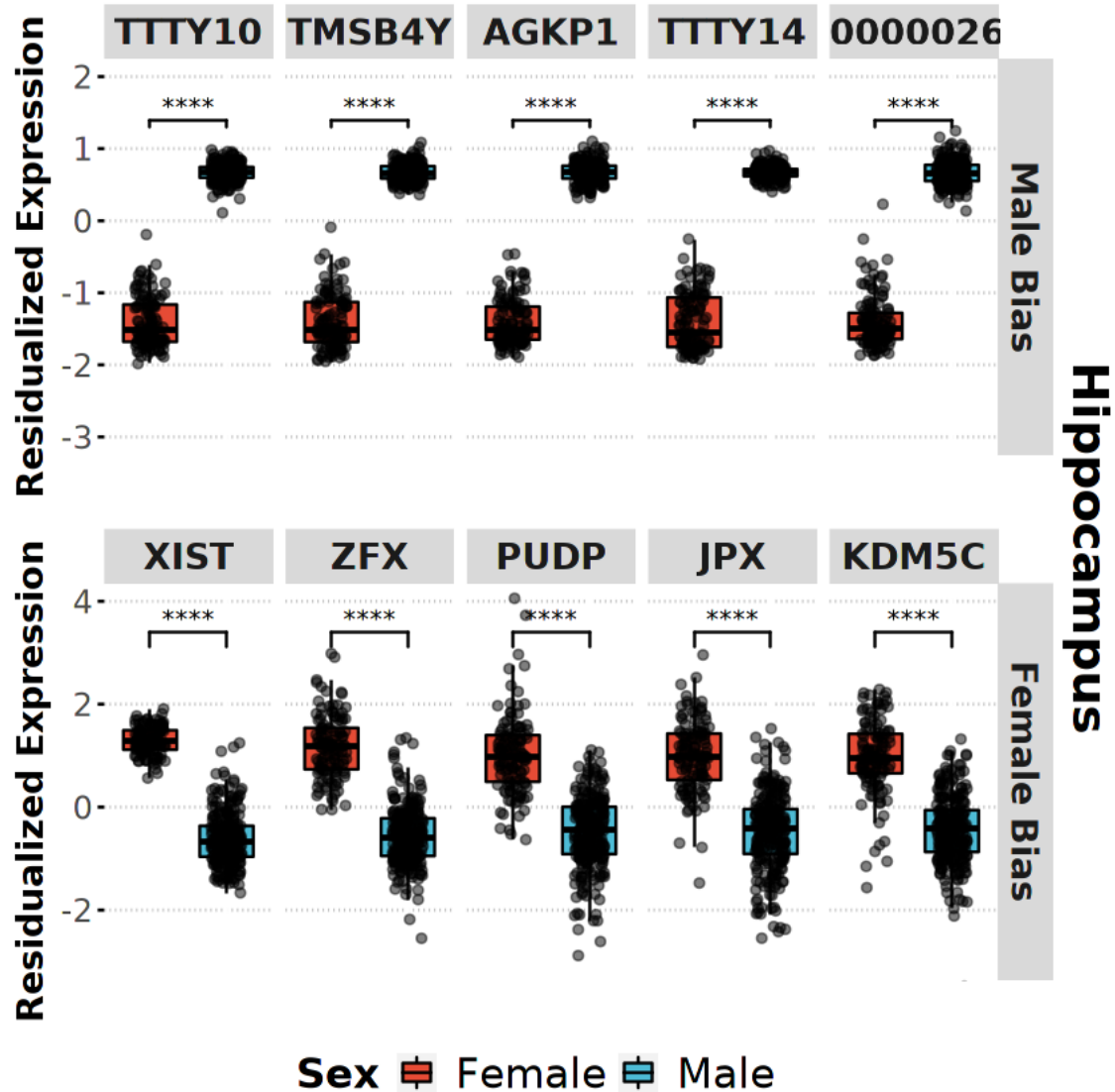
Warning message:

"Ignoring unknown aesthetics: xmin, xmax, annotations, y_position"



1.2.7 Annotate figure

```
[10]: figure <- ggarrange(bxp_up, bxp_down, ncol = 1,
                           nrow=2, common.legend=TRUE,
                           legend='bottom')
figure <- annotate_figure(figure, right=text_grob("Hippocampus", size=22,
↪rot=-90, face='bold'))
figure
```



```
[11]: save_ggplots('hippocampus_topSig_boxplot', figure, 12, 8)
```

1.3 Autosomal only

1.3.1 Differential expressed genes

```
[12]: genes = read.csv('.././metrics_summary/_m/autosomal_DEG.csv')
genes = genes %>% rename("gencodeID" = "Feature") %>%
  filter(adj.P.Val < 0.05) %>%
  select(gencodeID, ensemblID, Symbol, logFC, adj.P.Val)
genes = add_symnum(genes)

up_genes = genes %>% filter(logFC > 0) %>% mutate('Direction'='Male Bias') %>%
  head(5)
down_genes = genes %>% filter(logFC < 0) %>% mutate('Direction'='Female Bias')
  head(5)

sig_genes = bind_rows(up_genes, down_genes) %>% mutate_at(vars(Direction), as.
  factor)

for(xx in seq_along(sig_genes$Symbol)){
  sig_genes$New_ID[xx] <- ifelse(as.character(sig_genes$Symbol[xx]) == '',
    as.character(sig_genes$ensemblID[xx]),
    as.character(sig_genes$Symbol[xx]))
}
sig_genes
```

	gencodeID <chr>	ensemblID <chr>	Symbol <chr>	logFC <dbl>	adj.P.Val <dbl>
A data.frame: 10 × 9	ENSG00000205611.4	ENSG00000205611	LINC01597	1.1778543	1.668358e-14
	ENSG00000283443.1	ENSG00000283443		1.2307236	4.320235e-14
	ENSG00000282826.1	ENSG00000282826	FRG1CP	0.5550113	8.233830e-14
	ENSG00000149531.15	ENSG00000149531	FRG1BP	0.6492567	2.029046e-13
	ENSG00000258484.3	ENSG00000258484	SPESP1	0.7599005	2.382040e-10
	ENSG00000172460.14	ENSG00000172460	PRSS30P	-0.7600790	1.254613e-08
	ENSG00000180574.3	ENSG00000180574		-0.2642184	1.104438e-07
	ENSG00000267265.5	ENSG00000267265		-0.9704551	2.359616e-07
	ENSG00000182912.6	ENSG00000182912	TSPEAR-AS2	-0.7836724	5.779847e-07
	ENSG00000237268.2	ENSG00000237268		-0.9052018	1.531856e-05

1.3.2 Load residualized expression

```
[13]: resdf0 = tmp_df %>% filter(V1 %in% sig_genes$gencodeID) %>%
  column_to_rownames("V1")
resdf0 <- data.frame(t(resdf0))
resdf0[1:2, 1:5]
```

		ENSG00000237268.2 <dbl>	ENSG00000180574.3 <dbl>	ENSG00000258484.3 <dbl>	ENSG00000182912.6 <dbl>
A data.frame: 2 × 5	R11135	-0.1027035	0.4159911	-0.23704639	0.5595973
	R11137	-1.0453669	-0.3354074	0.05510561	0.4118647

1.3.3 Merge dataframe

```
[14]: resdf <- inner_join(rownames_to_column(pheno),
                        rownames_to_column(resdf0),
                        by="rowname")
dim(resdf)
resdf[1:2, 1:10]
```

1. 375 2. 19

	rowname	BrNum	RNum	Region	RIN	Age	Sex	Race	Dx
	<chr>	<fct>	<fct>	<fct>	<dbl>	<dbl>	<fct>	<fct>	<fct>
A data.frame: 2 × 10	1	R11135	Br2063	R11135	HIPPO	5.9	18.77	Male	CAUC
	2	R11137	Br2582	R11137	HIPPO	9.2	41.44	Male	CAUC

1.3.4 Melt data frame

```
[15]: df = resdf %>% select(c('rowname', 'Sex', starts_with('ENSG')) %>%
  pivot_longer(-c(rowname, Sex), names_to = "gencodeID", values_to = "Res")
  %>%
  mutate_if(is.character, as.factor) %>% inner_join(sig_genes,
  by='gencodeID') %>%
  mutate_at(vars("New_ID", "Symbol", "gencodeID"), as.factor)

df$New_ID = with(df, reorder(New_ID, adj.P.Val, median))
head(df, 2)
```

	rowname	Sex	gencodeID	Res	ensemblID	Symbol	logFC
	<fct>	<fct>	<fct>	<dbl>	<chr>	<fct>	<dbl>
A tibble: 2 × 12	R11135	Male	ENSG000000237268.2	-0.1027035	ENSG000000237268		-0.90520
	R11135	Male	ENSG000000180574.3	0.4159911	ENSG000000180574		-0.26421

1.3.5 Initial ggplot with ggpubr

```
[16]: tmp = sig_genes %>% mutate(group1='Female', group2='Male', y_pos=3.5) %>%
  filter(Direction == 'Male Bias') %>%
  mutate_if(is.character, as.factor)

byp_up <- df %>% filter(Direction == "Male Bias") %>%
  ggboxplot(x="Sex", y="Res", fill="Sex", palette="npg", xlab='', outlier.
  shape=NA,
            facet.by=c("Direction", "New_ID"), panel.labs.
  font=list(face='bold', size = 16),
            ylab='Residualized Expression', add='jitter', ylim=c(-6.5,
  5),
            add.params=list(alpha=0.5), legend="bottom",
  ggtheme=theme_pubclean()) +
  geom_signif(data=tmp,
```

```

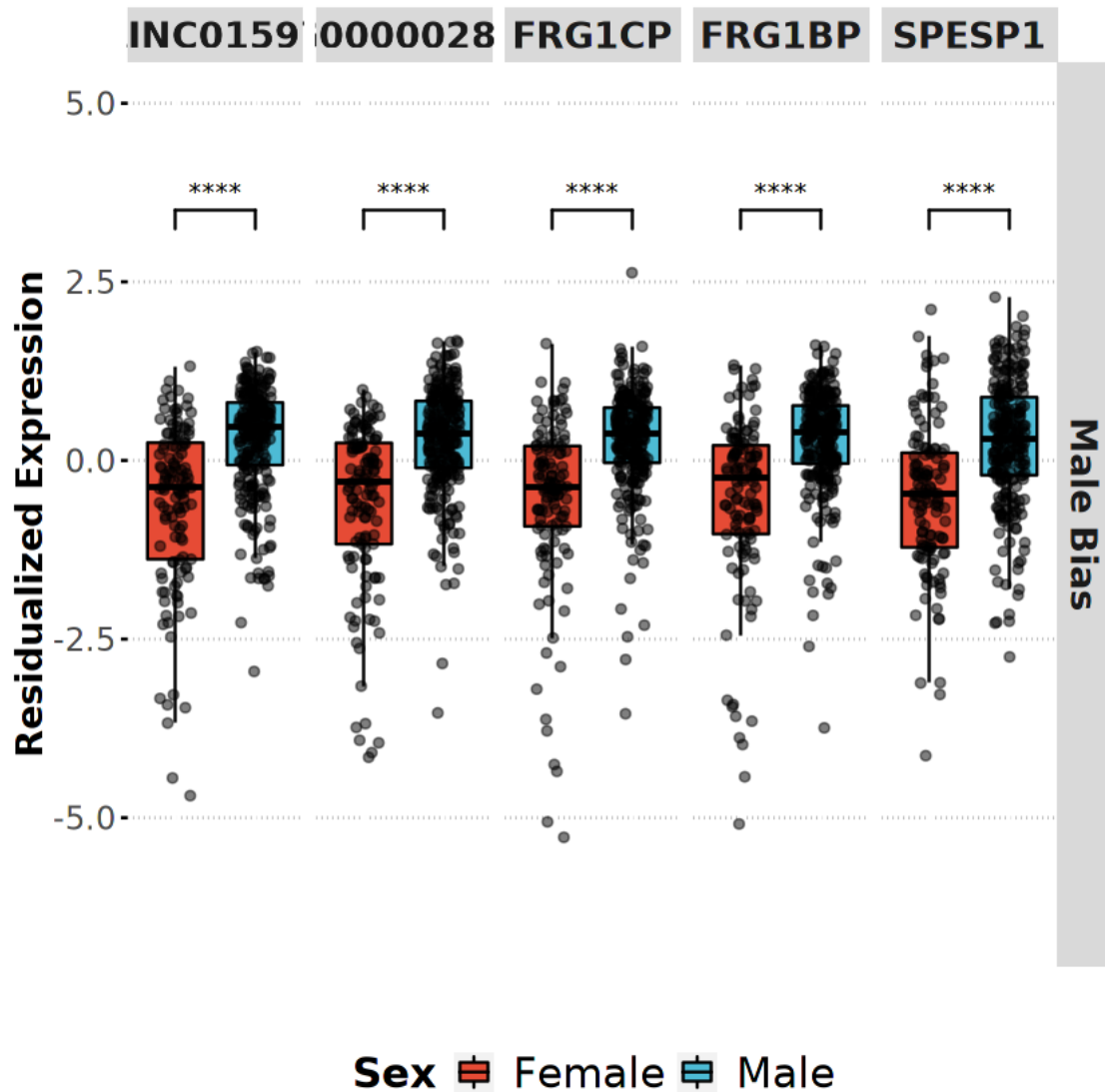
aes(xmin=group1, xmax=group2, annotations=p.signif,
→y_position = y_pos),
manual=TRUE) +
font("xy.title", size=16, face="bold") + font("xy.text", size=14) +
font("legend.title", size=18, face="bold") +
font("legend.text", size=18) + rremove("x.text") + rremove("x.ticks")

```

bxp_up

Warning message:

"Ignoring unknown aesthetics: xmin, xmax, annotations, y_position"



```

[17]: tmp = sig_genes %>% mutate(group1='Female', group2='Male', y_pos=4.5) %>%
  ↪filter(Direction == 'Female Bias') %>%
    mutate_if(is.character, as.factor)

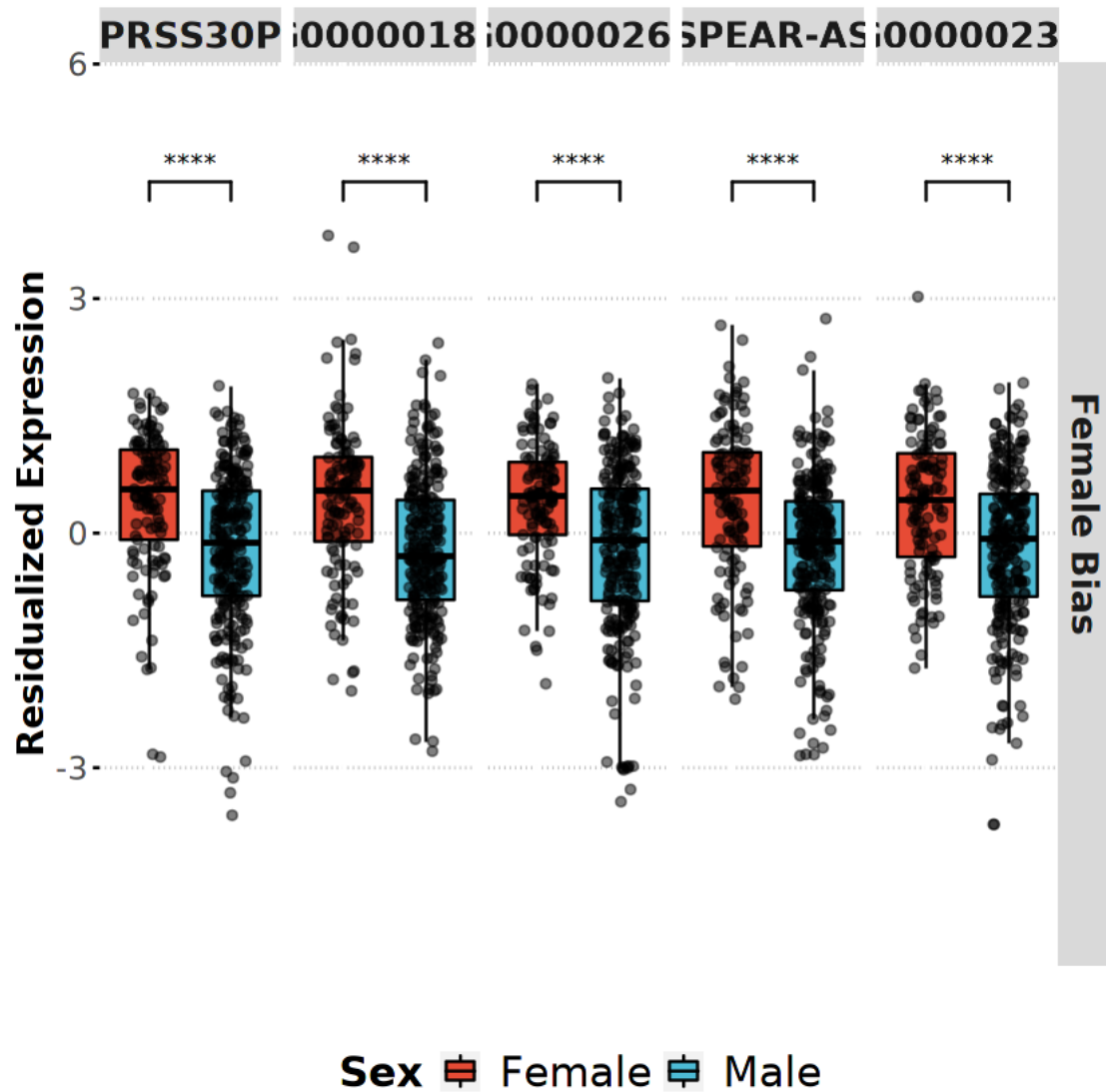
bxp_down <- df %>% filter(Direction=="Female Bias") %>%
  ggboxplot(x="Sex", y="Res", fill="Sex", palette="npg", xlab='', outlier.
  ↪shape=NA,
    facet.by=c("Direction", "New_ID"), panel.labs.
  ↪font=list(face='bold', size = 16),
    ylab='Residualized Expression', add='jitter', ylim=c(-5, 5.5),
    add.params=list(alpha=0.5), legend="bottom",
  ↪ggtheme=theme_pubclean()) +
    geom_signif(data=tmp,
      aes(xmin=group1, xmax=group2, annotations=p.signif,
  ↪y_position = y_pos),
      manual=TRUE) +
    font("xy.title", size=16, face="bold") + font("xy.text", size=14) +
    font("legend.title", size=18, face="bold") +
    font("legend.text", size=18) + rremove("x.text") + rremove("x.ticks")

bxp_down

```

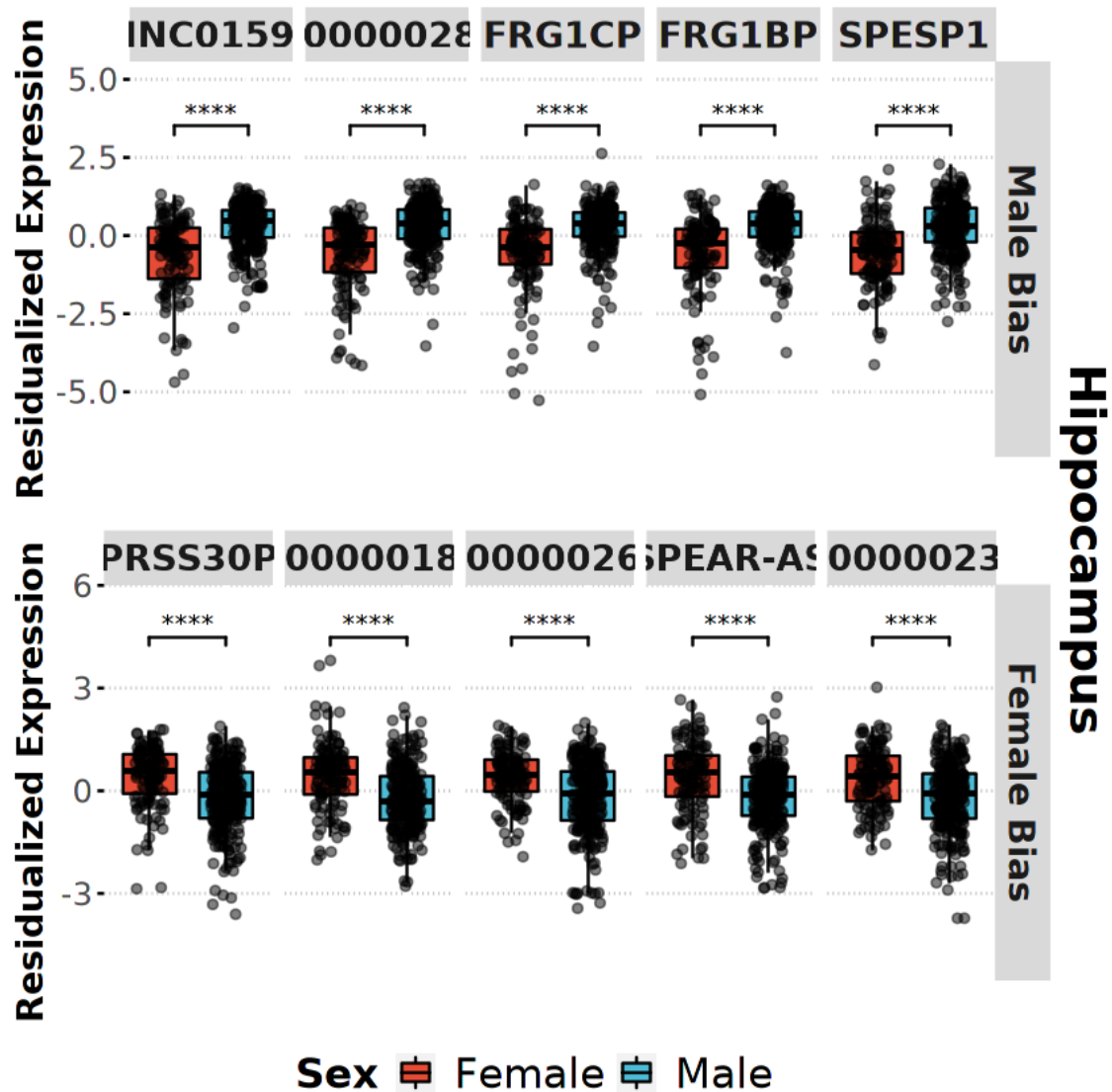
Warning message:

"Ignoring unknown aesthetics: xmin, xmax, annotations, y_position"



1.3.6 Annotate figure

```
[18]: figure2 <- ggarrange(bxp_up, bxp_down, ncol=1, nrow=2, common.legend=TRUE,
  ↪ legend='bottom')
figure2 <- annotate_figure(figure2,
  ↪ right=text_grob("Hippocampus", size=22, rot=-90,
  ↪ face='bold'))
figure2
```



```
[19]: save_ggplots('hippocampus_topSig_autosomal_boxplot', figure2, 12, 8)
```

1.4 Session Info

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

```
[1] "2021-07-09 22:56:30 EDT"
```

```
user system elapsed
19.794 1.892 21.969
```

```

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     2021-07-09

```

```

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)
broom         0.7.8     2021-06-24 [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)
cli           3.0.0     2021-06-30 [1] CRAN (R 4.0.3)
colorspace    2.0-2     2021-06-24 [1] CRAN (R 4.0.3)
cowplot       1.1.1     2020-12-30 [1] CRAN (R 4.0.2)
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    1.14.0    2021-02-21 [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        0.6.27    2020-10-24 [1] CRAN (R 4.0.2)
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)
fans          0.5.0     2021-05-25 [1] CRAN (R 4.0.3)
farver        2.1.0     2021-02-28 [1] CRAN (R 4.0.3)
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)
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)
ggsci         2.9       2018-05-14 [1] CRAN (R 4.0.2)
ggsignif      * 0.6.2     2021-06-14 [1] CRAN (R 4.0.3)
glue          1.4.2     2020-08-27 [1] CRAN (R 4.0.2)
gridExtra     2.3       2017-09-09 [1] CRAN (R 4.0.2)
gtable        0.3.0     2019-03-25 [1] CRAN (R 4.0.2)

```

haven	2.4.1	2021-04-23	[1]	CRAN	(R 4.0.3)
hms	1.1.0	2021-05-17	[1]	CRAN	(R 4.0.3)
htmltools	0.5.1.1	2021-01-22	[1]	CRAN	(R 4.0.2)
httr	1.4.2	2020-07-20	[1]	CRAN	(R 4.0.2)
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)
labeling	0.4.2	2020-10-20	[1]	CRAN	(R 4.0.2)
lifecycle	1.0.0	2021-02-15	[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)
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	1.6.1	2021-05-16	[1]	CRAN	(R 4.0.3)
pkgconfig	2.0.3	2019-09-22	[1]	CRAN	(R 4.0.2)
purrr	* 0.3.4	2020-04-17	[1]	CRAN	(R 4.0.2)
R6	2.5.0	2020-10-28	[1]	CRAN	(R 4.0.2)
Rcpp	1.0.7	2021-07-07	[1]	CRAN	(R 4.0.3)
readr	* 1.4.0	2020-10-05	[1]	CRAN	(R 4.0.2)
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	2.0.0	2021-04-02	[1]	CRAN	(R 4.0.3)
rio	0.5.27	2021-06-21	[1]	CRAN	(R 4.0.3)
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	1.0.0	2021-03-09	[1]	CRAN	(R 4.0.3)
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)
stringi	1.6.2	2021-05-17	[1]	CRAN	(R 4.0.3)
stringr	* 1.4.0	2019-02-10	[1]	CRAN	(R 4.0.2)
svglite	2.0.0	2021-02-20	[1]	CRAN	(R 4.0.3)
systemfonts	1.0.2	2021-05-11	[1]	CRAN	(R 4.0.3)
tibble	* 3.1.2	2021-05-16	[1]	CRAN	(R 4.0.3)
tidyr	* 1.1.3	2021-03-03	[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)
utf8	1.2.1	2021-03-12	[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