

# check\_gender\_distribution\_xlinked

ls760

11/02/2021

## Gender distribution of X-linked diseases

I selected the genes that are cause of X-linked diseases (plus NBEAL2 as control):

Gene	disease	MOI
F8	FVIII deficiency	XLR
F9	FIX deficiency	XLR
FLNA	thrombocytopenia with giant platelets	
NBEAL2	GPS	Not x-linked. used as control
PIGA	Paroxysmal nocturnal hemoglobinuria	Somatic
WAS	Wiskott-Aldrich syndrome	XLR

```
require(reshape2)
require(tidyverse)
load('variopath_MDT.RData')
# create variant IDs
df_clean$ID = paste(df_clean$CHROM, df_clean$POS, df_clean$REF, df_clean$ALT, sep = '_')
# reshape the table to have long format
df_gender = unique(df_gender)
df_gender = melt(df_clean[ (df_clean$CHROM == 'X' | df_clean$GENE == 'NBEAL2'),
                        c('ID', 'GENE', 'male', 'female') ])

# Check where are the majority of x-linked variants in male
ggplot(df_gender, aes(x = ID,
                      y = value,
                      fill = variable)) +
  geom_bar(stat = 'identity', position = 'dodge2') +
  facet_wrap('GENE', scales = 'free') +
  theme(axis.text.x = element_text(angle = 90))
```



## Stats gene-level

Look at how many people are hemizigous/male in x-linked genes

```
gene_list = df_clean[( df_clean$CHROM == 'X' | df_clean$GENE == 'NBEAL2'),
                      c('GENE', 'male', 'female') ] %>%
  group_by(GENE) %>%
  summarise('n_male' = sum(male), 'n_female' = sum(female))
knitr::kable(gene_list, format = "pipe")
```

GENE	n_male	n_female
F8	232	1086
F9	97	754
FLNA	21	78
NBEAL2	13	17
PIGA	1	1
WAS	7	23

## Fisher exact test

```
gene_list$fisher.pvalue = "1"
for (gene in 1:dim(gene_list)[1]) {
  total_obs_gender = (gene_list[gene,2] + gene_list[gene,3]) / 2
  # assuming 50:50 distribution between the genders
  fisher.matrix = matrix(c(as.integer(gene_list[gene,2]),
                           as.numeric(gene_list[gene,3]),
                           as.numeric(total_obs_gender),
                           as.numeric(total_obs_gender) ),
                        nrow = 2,
                        ncol = 2,
                        byrow = T,
                        dimnames =
                          list(c("observed", "expected"),
                               c("male", "female"))))
  gene_list$fisher.pvalue[gene] = fisher.test(fisher.matrix)$p.value
}
knitr::kable(gene_list, format = 'pipe')
```

GENE	n_male	n_female	fisher.pvalue
F8	232	1086	5.68744325145525e-71
F9	97	754	2.71503040412085e-70
FLNA	21	78	2.94539195214109e-05
NBEAL2	13	17	0.79610715858958
PIGA	1	1	1
WAS	7	23	0.0596015801738957

## Example of a fisher matrix

	GENE	
	WAS	
	male	female
observed	7	23
expected	15	15

## — End of the script —

```
## R version 3.6.1 (2019-07-05)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS 10.16
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_GB.UTF-8/en_GB.UTF-8/en_GB.UTF-8/C/en_GB.UTF-8/en_GB.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] forcats_0.5.0  stringr_1.4.0  dplyr_1.0.3    purrr_0.3.4
## [5] readr_1.4.0    tidyr_1.1.2    tibble_3.0.5   ggplot2_3.3.3
## [9] tidyverse_1.3.0 reshape2_1.4.4
##
## loaded via a namespace (and not attached):
## [1] tidyselect_1.1.0 xfun_0.20      haven_2.3.1    colorspace_2.0-0
## [5] vctrs_0.3.6      generics_0.1.0 htmltools_0.5.1 yaml_2.2.1
## [9] rlang_0.4.10     pillar_1.4.7   glue_1.4.2     withr_2.4.0
## [13] DBI_1.1.0        dbplyr_2.0.0   modelr_0.1.8    readxl_1.3.1
## [17] lifecycle_0.2.0  plyr_1.8.6     munsell_0.5.0   gtable_0.3.0
## [21] cellranger_1.1.0 rvest_0.3.6     evaluate_0.14    labeling_0.4.2
## [25] knitr_1.30       fansi_0.4.2     highr_0.8       broom_0.7.3
## [29] Rcpp_1.0.6       scales_1.1.1    backports_1.2.1 jsonlite_1.7.2
## [33] farver_2.0.3     fs_1.5.0        hms_1.0.0       digest_0.6.27
## [37] stringi_1.5.3    grid_3.6.1      cli_2.2.0       tools_3.6.1
## [41] magrittr_2.0.1    crayon_1.3.4    pkgconfig_2.0.3 ellipsis_0.3.1
## [45] xml2_1.3.2        reprex_0.3.0    lubridate_1.7.9.2 assertthat_0.2.1
## [49] rmarkdown_2.6     httr_1.4.2      rstudioapi_0.13 R6_2.5.0
## [53] compiler_3.6.1
```