Class 19 Lab

Divyanshu Kawankar

12/6/2021

###Population Scale Analysis

Q13: Read this file into R and determine the sample size for each genotype and their corresponding median expression levels for each of these genotypes. Hint: The read.table(), summary() and boxplot() functions will likely be useful here. There is an example R script online to be used ONLY if you are struggling in vein. Note that you can find the medium value from saving the output of the boxplot() function to an R object and examining this object. There is also the medium() and summary() function that you can use to check your understanding. - Refer to data below

```
library(ggplot2)

gene <- read.table("rs8067378_ENSG00000172057.6.txt")

table(gene$geno)
```

```
##
## A/A A/G G/G
## 108 233 121
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
summ <- gene %>%
    group_by(geno) %>%
    summarize(mean = mean(exp), median = median(exp), sd = sd(exp))

#median values for A/A, A/G, And G/G respectively
summ$median
```

```
## [1] 31.24847 25.06486 20.07363
```

Q14: Generate a boxplot with a box per genotype, what could you infer from the relative expression value between A/A and G/G displayed in this plot? Does the SNP effect the expression of ORMDL3? Hint: An example boxplot is provided overleaf – yours does not need to be as polished as this one - From the relative expression values between A/A and G/G we could infer that since they are all expressed in different values, their difference may be a product of being affected by asthma associated SNPs. Thus, we could infer that SNP could effect the expression of ORMDL3.

```
library(ggplot2)

library(dplyr)

ggplot(gene) + aes(x=geno,y=exp, fill=geno) +
    geom_boxplot(notch=TRUE)
```