Lab_11

RUNQI ZHANG

Table of contents

Section 1. Proportion og G/G in a population

Download a CSV file from Emsemble

```
mxl <- read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")
table(mxl$Genotype..forward.strand.) / nrow(mxl)
```

```
    A|A     A|G     G|A     G|G
0.343750 0.328125 0.187500 0.140625
```

** Interpreting Base Qualities in R

```
library(seqinr)
library(gtools)
phred <- asc( s2c("DDD") )-33
```

Section 4. Population Scale Analysis

whether there is any association of the 4 asthma-associated SNPs (rs8067378) on ORMDL3

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.

A: the file contains 462 samples. The median expression levels are 20.1, 31.2, 25.1 for genotype G/G, A/A, and A/G, respectively.

```
expr <- read.table("rs8067378.txt")
nrow(expr)
```

```
[1] 462
```

```
table(expr$geno)
```

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

```
round(median(expr$exp[expr$geno=="G/G"]), 1)
```

```
[1] 20.1
```

```
round(median(expr$exp[expr$geno=="A/A"]), 1)
```

```
[1] 31.2
```

```
round(median(expr$exp[expr$geno=="A/G"]), 1)
```

```
[1] 25.1
```

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? A: the relative expression level in A/A subjects are greater than G/G subjects. The SNP does have effect on the expression of ORMDL3, namely that the presence of A leads to higher expression of ORMDL3 and the presecne of G leads to lower expression.

```
library(ggplot2)
ggplot(expr)+
  aes(geno, exp, fill=geno)+
  geom_boxplot(notch=TRUE)
```