lab12 population scale analysis

Longmei Zhang A17012012

```
tb <- read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")
```

```
table(tb$Genotype..forward.strand.) * 100 / nrow(tb)
```

```
    A|A     A|G     G|A     G|G
34.3750 32.8125 18.7500 14.0625
```

Population Scale Analysis One sample is obviously not enough to know what is happening in a population. You are interested in assessing genetic differences on a population scale. So, you processed about ~230 samples and did the normalization on a genome level. Now, you want to find whether there is any association of the 4 asthma-associated SNPs (rs8067378...) on ORMDL3 expression.

read the data

```
expr <- read.table("rs8067378_ENSG00000172057.6.txt")
nrow(expr)
```

```
[1] 462
```

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

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

```
ag <- expr$geno == "A/G"
ag_expr = expr[ag,]
summary(ag_expr$exp)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  7.075  20.626  25.065  25.397  30.552  48.034
```

```
gg <- expr$geno == "G/G"
gg_expr = expr[gg,]
summary(gg_expr$exp)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  6.675  16.903  20.074  20.594  24.457  33.956
```

```
aa <- expr$geno == "A/A"
aa_expr = expr[aa,]
summary(aa_expr$exp)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  11.40   27.02   31.25   31.82   35.92   51.52
```

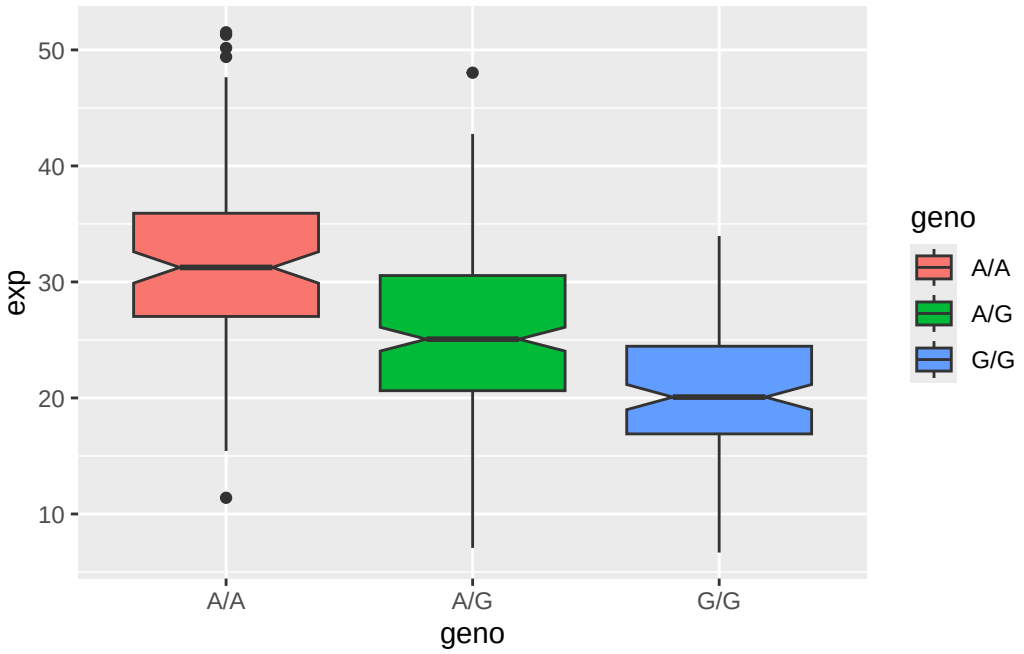
> Q13 Read this file into R and determine the sample size for each genotype and their corresponding median expression levels for each of these genotypes.

There are 462 samples in total, 108 samples with genotype A/A, 233 samples with genotype A/G, and 121 samples with genotype G/G.The median expression level for A/G is 25.065, for G/G is 20.074, and for A/A is 31.25

Make box plot

```
library(ggplot2)
```

```
ggplot(expr, aes(geno, exp, fill = geno)) +
  geom_boxplot(notch = T)
```



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?

Individuals with genotype A/A tend to have higher expression of ORMDL3. It seems like SNP effect the expression of this gene.