

# Class 12 HW

AUTHOR

Charlie Rezanka (A15837296)

## Proportion of G/G in data

Downloaded a CSV file from ensemble ([https://useast.ensembl.org/Homo\\_sapiens/Variation/Sample?db=core;r=17:39905111-40015112;v=rs8067378;vdb=variation;vf=105535077#373531\\_tablePanel](https://useast.ensembl.org/Homo_sapiens/Variation/Sample?db=core;r=17:39905111-40015112;v=rs8067378;vdb=variation;vf=105535077#373531_tablePanel))

```
mx1 <- read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv", row.names =  
head(mx1)
```

	Genotype..forward.strand.	Population.s.	Father	Mother
NA19648 (F)	A A	ALL, AMR, MXL	-	-
NA19649 (M)	G G	ALL, AMR, MXL	-	-
NA19651 (F)	A A	ALL, AMR, MXL	-	-
NA19652 (M)	G G	ALL, AMR, MXL	-	-
NA19654 (F)	G G	ALL, AMR, MXL	-	-
NA19655 (M)	A G	ALL, AMR, MXL	-	-

```
table(mx1$Genotype..forward.strand.)
```

A A	A G	G A	G G
22	21	12	9

```
(table(mx1$Genotype..forward.strand.) / nrow(mx1))*100
```

A A	A G	G A	G G
34.3750	32.8125	18.7500	14.0625

Lets look at a different population, such as Great Britin

```
gbr <- read.csv("373522-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv", row.names =  
head(gbr)
```

	Genotype..forward.strand.	Population.s.	Father	Mother
HG00096 (M)	A A	ALL, EUR, GBR	-	-
HG00097 (F)	G A	ALL, EUR, GBR	-	-

HG00099 (F)	G G ALL, EUR, GBR	-	-
HG00100 (F)	A A ALL, EUR, GBR	-	-
HG00101 (M)	A A ALL, EUR, GBR	-	-
HG00102 (F)	A A ALL, EUR, GBR	-	-

Find the portion of G/G

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

A A	A G	G A	G G
25.27473	18.68132	26.37363	29.67033

The haplotype most associated with childhood asthma is more common in the GB population than MXL

## Population Scale Analysis Homework

```
exp <- read.table("rs8067378_ENSG00000172057.6.txt", row.names = 1)
head(exp)
```

	sample	geno	exp
1	HG00367	A/G	28.96038
2	NA20768	A/G	20.24449
3	HG00361	A/A	31.32628
4	HG00135	A/A	34.11169
5	NA18870	G/G	18.25141
6	NA11993	A/A	32.89721

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.

```
nrow(exp)
```

```
[1] 462
```

```
table(exp$geno)
```

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

```
medians <- aggregate(exp ~ geno, data = exp, FUN = median)
head(medians)
```

	geno	exp
1	A/A	31.24847

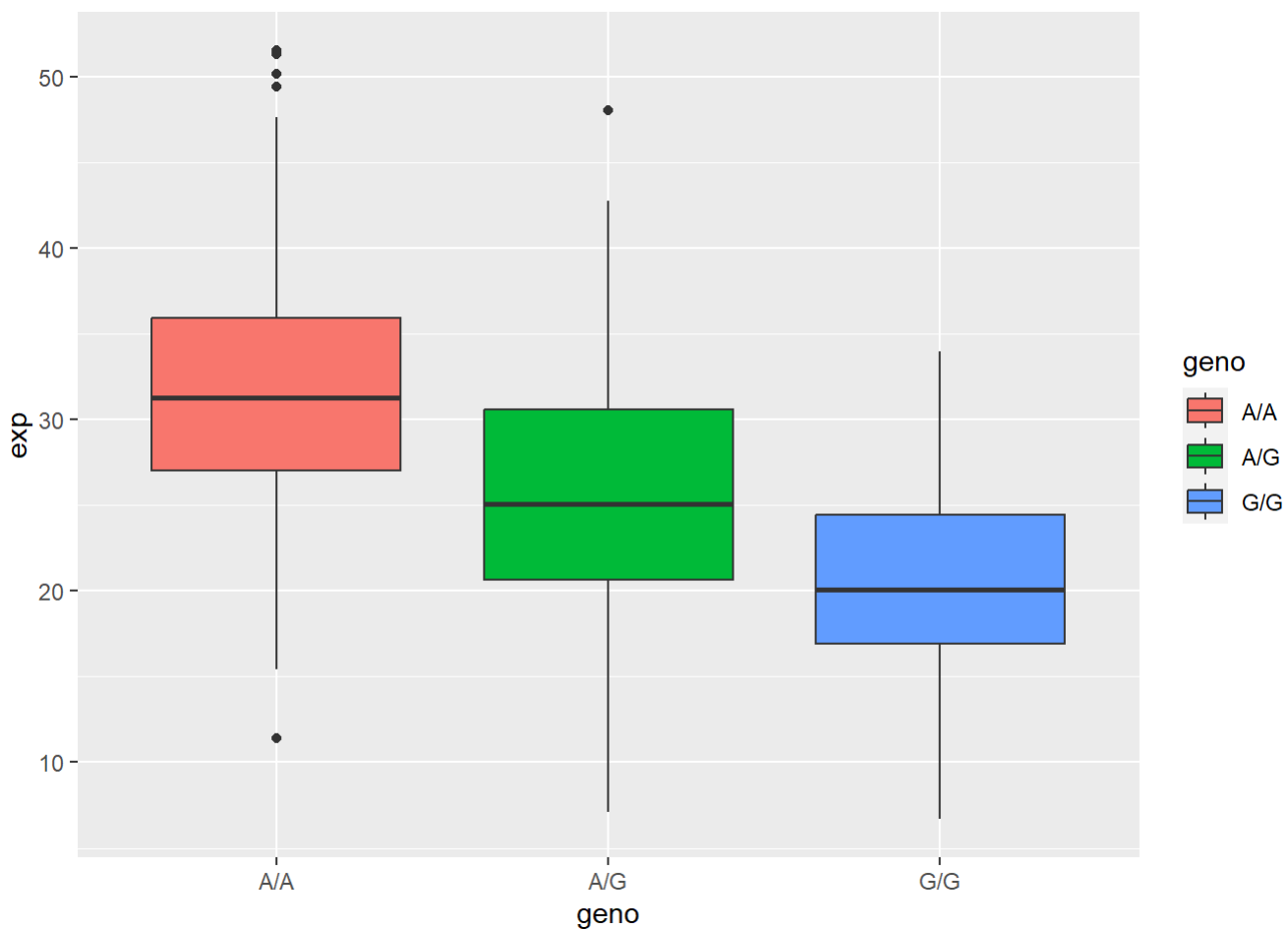
2 A/G 25.06486

3 G/G 20.07363

Out of a total of 462 samples, the A/A genotype has a sample size of 108, A/G of 233, and G/G of 121. Median expression for A/A is 31.2, for A/G is 25.1, and G/G is 20.1.

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?

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



The presence of the G allele in either heterozygous or homozygous genotypes will result in a drop in ORMDL3 expression, with the homozygous G/G genotype having the lowest expression.