

Class 12 Lab

Hugh (Trevor) Redford (PID:A17067426)

2025-02-16

Propotion on G/G in a population

```
# Read in Mexican Ancestry in LA csv file
mxl <- read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")
head(mxl)
```

	Sample..Male.Female.Unknown.	Genotype..forward.strand.	Population.s.	Father
1		NA19648 (F)	A A ALL, AMR, MXL	-
2		NA19649 (M)	G G ALL, AMR, MXL	-
3		NA19651 (F)	A A ALL, AMR, MXL	-
4		NA19652 (M)	G G ALL, AMR, MXL	-
5		NA19654 (F)	G G ALL, AMR, MXL	-
6		NA19655 (M)	A G ALL, AMR, MXL	-
	Mother			
1	-			
2	-			
3	-			
4	-			
5	-			
6	-			

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

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

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 median value from saving the output of the `boxplot()` function to an R object and examining this object. There is also the `median()` and `summary()` function that you can use to check your understanding.

The boxplot below displays the median expression levels corresponding to each genotype.

```
#Read in file
gene_expression <- read.table("rs8067378_ENSG00000172057.6.txt")
head(gene_expression)
```

	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

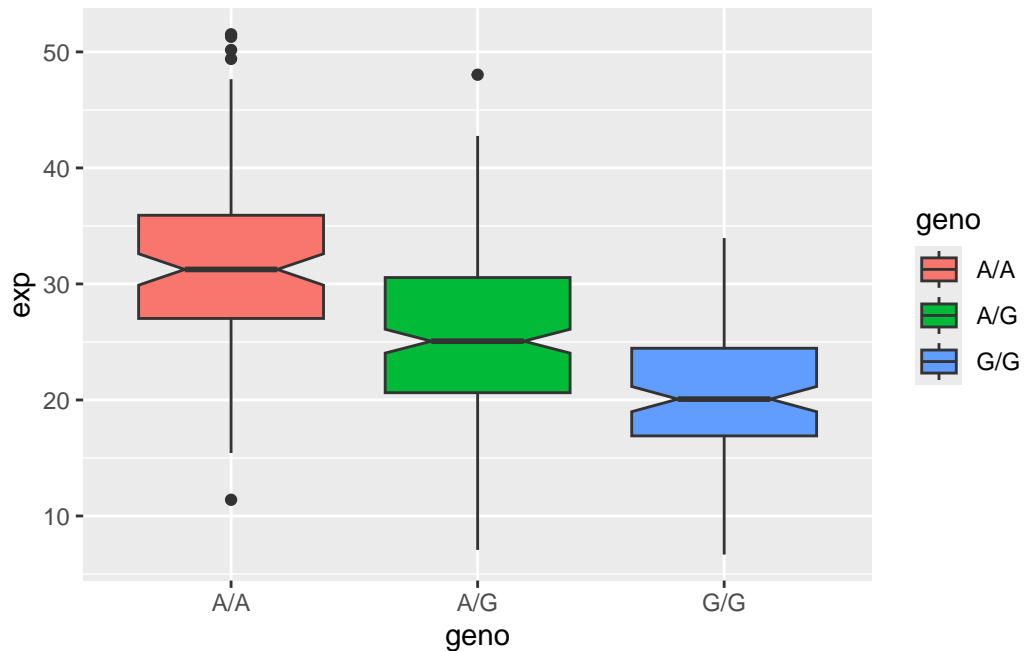
```
nrow(gene_expression)
```

```
[1] 462
```

```
table(gene_expression$geno)
```

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

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



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.

The boxplot shows ORMDL3 expression across genotypes (A/A, A/G, G/G). Individuals with A/A have higher expression than G/G, while A/G has the highest median. This suggests the SNP influences ORMDL3 expression, with the G allele linked to lower levels, highlighting its potential regulatory role.