

# Lab11

#Section 1. Proportion of G/G in a population

Downloaded csv file from Ensembl [https://uswest.ensembl.org/Homo\\_sapiens/Variation/Sample?db=core;r=17:39894595-39895595;v=rs8067378;vdb=variation;vf=105535077](https://uswest.ensembl.org/Homo_sapiens/Variation/Sample?db=core;r=17:39894595-39895595;v=rs8067378;vdb=variation;vf=105535077)

Here we read this 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	-			

```
View(mxl)
```

```
table(mxl$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

Let's look at a different population (GBR)

```
gbr <- read.csv("373522-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")
head(gbr)
```

	Sample..Male.Female.Unknown.	Genotype..forward.strand.	Population.s.	Father
1		HG00096 (M)	A A ALL, EUR, GBR	-
2		HG00097 (F)	G A ALL, EUR, GBR	-
3		HG00099 (F)	G G ALL, EUR, GBR	-
4		HG00100 (F)	A A ALL, EUR, GBR	-
5		HG00101 (M)	A A ALL, EUR, GBR	-
6		HG00102 (F)	A A ALL, EUR, GBR	-
	Mother			
1	-			
2	-			
3	-			
4	-			
5	-			
6	-			

Find proportion of G|G

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

A A	A G	G A	G G
23	17	24	27

```
round(table(gbr$Genotype..forward.strand.) / nrow(gbr) *100 , 2)
```

```
  A|A   A|G   G|A   G|G
25.27 18.68 26.37 29.67
```

This variant that is associated with childhood asthma is more frequent in the GBR population than the MXL population.

## Section 4: 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.

How many samples do we have?

```
#download file from our classroom page
#rename to expr
expr <- read.table("rs8067378_ENSG00000172057.6.txt")
#View the first 6 rows of the file named expr
head(expr)
```

```
  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
```

```
#summary() tells us more information about the data imported
summary(expr)
```

```
      sample              geno              exp
Length:462      Length:462      Min.   : 6.675
Class :character  Class :character  1st Qu.:20.004
Mode  :character  Mode  :character  Median :25.116
```

Mean :25.640  
3rd Qu.:30.779  
Max. :51.518

```
#Subsets all the rows that are A/A genotype  
expr$geno == "A/A"
```

```
[1] FALSE FALSE TRUE TRUE FALSE TRUE FALSE TRUE FALSE FALSE FALSE FALSE  
[13] TRUE FALSE TRUE TRUE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE  
[25] FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE  
[37] TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE  
[49] FALSE FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE FALSE FALSE FALSE  
[61] FALSE TRUE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE  
[73] FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE  
[85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE  
[97] TRUE FALSE TRUE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE  
[109] FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
[121] TRUE TRUE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE  
[133] TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
[145] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE  
[157] FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE TRUE FALSE  
[169] TRUE FALSE FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE TRUE TRUE  
[181] FALSE TRUE FALSE FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE  
[193] FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE TRUE FALSE FALSE  
[205] TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE  
[217] TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE  
[229] TRUE TRUE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE TRUE  
[241] FALSE TRUE TRUE FALSE TRUE TRUE FALSE TRUE FALSE FALSE FALSE FALSE  
[253] FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE TRUE TRUE  
[265] TRUE TRUE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE TRUE FALSE  
[277] FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
[289] FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE  
[301] FALSE FALSE FALSE TRUE TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE  
[313] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE  
[325] FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE  
[337] TRUE FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE TRUE FALSE TRUE  
[349] FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE TRUE TRUE  
[361] FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE TRUE TRUE FALSE  
[373] FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE TRUE  
[385] FALSE TRUE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE  
[397] FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE TRUE FALSE FALSE TRUE
```

```
[409] TRUE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE
[421] FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE
[433] TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE
[445] FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE FALSE FALSE TRUE
[457] FALSE FALSE TRUE FALSE TRUE FALSE
```

```
##Subsets in column exp by all the rows that are A/A genotype
expr$exp[expr$geno == "A/A"]
```

```
[1] 31.32628 34.11169 32.89721 47.64556 30.71355 25.44983 34.24915 35.15014
[9] 24.54684 49.39612 28.20755 28.56199 51.51787 38.10956 31.13741 39.12999
[17] 32.44173 34.03260 36.27151 26.10355 19.48106 38.77623 40.82922 46.50527
[25] 33.07320 25.47283 29.50655 27.48438 35.69719 32.42236 28.91526 19.14544
[33] 22.85572 31.17067 27.87464 27.91580 16.86780 23.66127 28.03403 36.51922
[41] 27.43637 35.63983 38.85161 22.44576 37.06379 35.67637 50.16704 37.94544
[49] 29.15536 28.85309 38.57101 30.89365 39.31537 31.43743 35.26739 22.37043
[57] 28.46943 30.15636 35.88457 32.59723 36.02549 23.86454 28.38114 25.14243
[65] 51.30170 28.14811 29.18390 27.09760 33.89656 33.48253 15.43178 29.45277
[73] 26.80283 43.51943 45.80808 26.55972 24.87330 29.99549 26.68589 39.37193
[81] 38.39523 41.23635 27.13936 36.55643 28.34688 15.42908 47.50579 30.15754
[89] 34.88439 29.29955 32.87767 43.43665 26.56050 21.87746 33.42582 25.38406
[97] 18.20442 21.62336 35.80093 32.26844 34.40756 23.99631 11.39643 35.24632
[105] 23.38376 34.94395 31.92036 44.27738
```

```
#Obtains the median of A/A genotype
median(expr$exp[expr$geno == "A/A"])
```

```
[1] 31.24847
```

```
expr$geno == "A/G"
```

```
[1] TRUE TRUE FALSE FALSE FALSE FALSE TRUE FALSE FALSE TRUE TRUE TRUE
[13] FALSE TRUE FALSE FALSE FALSE FALSE TRUE FALSE TRUE TRUE FALSE FALSE
[25] TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE TRUE
[37] FALSE TRUE TRUE TRUE FALSE TRUE TRUE TRUE TRUE TRUE FALSE FALSE FALSE
[49] FALSE FALSE TRUE TRUE FALSE TRUE FALSE FALSE FALSE TRUE TRUE TRUE
[61] FALSE FALSE TRUE TRUE TRUE TRUE FALSE TRUE TRUE TRUE TRUE TRUE FALSE
[73] FALSE TRUE TRUE FALSE FALSE TRUE FALSE TRUE TRUE TRUE FALSE TRUE TRUE
```

```

[85] FALSE TRUE TRUE TRUE FALSE TRUE TRUE FALSE FALSE TRUE TRUE FALSE
[97] FALSE TRUE FALSE TRUE TRUE FALSE TRUE FALSE FALSE FALSE TRUE TRUE
[109] FALSE FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE FALSE TRUE
[121] FALSE FALSE TRUE TRUE TRUE TRUE FALSE FALSE TRUE TRUE TRUE FALSE
[133] FALSE TRUE FALSE TRUE TRUE TRUE TRUE FALSE TRUE TRUE FALSE TRUE
[145] TRUE TRUE TRUE TRUE TRUE FALSE TRUE TRUE FALSE FALSE FALSE FALSE
[157] TRUE TRUE FALSE TRUE TRUE FALSE FALSE TRUE TRUE FALSE FALSE TRUE
[169] FALSE FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE FALSE FALSE
[181] TRUE FALSE TRUE TRUE TRUE TRUE FALSE FALSE TRUE FALSE TRUE TRUE
[193] FALSE FALSE FALSE TRUE TRUE FALSE FALSE TRUE FALSE FALSE TRUE TRUE
[205] FALSE TRUE FALSE TRUE TRUE TRUE FALSE TRUE TRUE FALSE FALSE TRUE
[217] FALSE FALSE TRUE TRUE TRUE TRUE TRUE FALSE FALSE FALSE FALSE FALSE
[229] FALSE FALSE TRUE FALSE FALSE TRUE FALSE TRUE TRUE TRUE FALSE FALSE
[241] FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE TRUE FALSE TRUE TRUE
[253] FALSE FALSE TRUE TRUE FALSE TRUE FALSE TRUE FALSE TRUE FALSE FALSE
[265] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE TRUE
[277] TRUE FALSE TRUE FALSE TRUE TRUE FALSE TRUE FALSE TRUE FALSE FALSE
[289] TRUE FALSE TRUE FALSE FALSE TRUE TRUE TRUE TRUE TRUE FALSE FALSE TRUE
[301] TRUE TRUE TRUE FALSE FALSE FALSE FALSE FALSE TRUE TRUE TRUE TRUE
[313] TRUE FALSE TRUE FALSE TRUE TRUE FALSE TRUE TRUE FALSE TRUE TRUE
[325] TRUE TRUE TRUE TRUE TRUE FALSE TRUE FALSE TRUE TRUE TRUE TRUE TRUE
[337] FALSE TRUE TRUE FALSE FALSE TRUE FALSE FALSE TRUE FALSE TRUE FALSE
[349] TRUE TRUE FALSE TRUE TRUE FALSE FALSE FALSE FALSE TRUE FALSE FALSE
[361] FALSE FALSE TRUE FALSE TRUE TRUE TRUE FALSE FALSE FALSE FALSE TRUE
[373] FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[385] FALSE FALSE TRUE TRUE FALSE FALSE FALSE TRUE FALSE TRUE TRUE TRUE
[397] TRUE TRUE TRUE FALSE FALSE TRUE TRUE TRUE TRUE FALSE TRUE TRUE FALSE
[409] FALSE TRUE TRUE FALSE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE FALSE
[421] FALSE TRUE TRUE FALSE TRUE FALSE TRUE FALSE TRUE TRUE TRUE TRUE TRUE
[433] FALSE TRUE FALSE FALSE TRUE TRUE TRUE TRUE TRUE TRUE FALSE FALSE TRUE
[445] TRUE FALSE TRUE FALSE TRUE TRUE FALSE TRUE TRUE FALSE TRUE FALSE
[457] FALSE FALSE FALSE TRUE FALSE TRUE

```

```
expr$exp[expr$geno == "A/G"]
```

```

[1] 28.96038 20.24449 31.48736 33.85374 16.29854 19.73450 13.96175 18.40351
[9] 34.21985 23.32404 23.18606 18.15997 28.03580 30.65270 20.51327 29.72309
[17] 22.50789 31.68959 28.76435 35.85543 37.62403 20.54885 31.04941 18.94583
[25] 32.29483 27.81775 19.89903 48.03410 27.90313 36.47949 20.03116 29.65063
[33] 34.92257 16.71385 16.68151 25.71008 11.60808 19.30953 20.49040 22.37224
[41] 15.20045 20.07459 35.42982 19.10884 24.81087 33.22193 33.40835 22.38996

```

```
[49] 31.31626  9.36055 23.17937 30.63079 19.10420 25.70962 38.34531 19.03976
[57] 28.81770 42.75662 25.61662 26.48467 28.02486 24.23377 21.03717 33.51752
[65] 20.73493 33.55650 36.62034 40.06318 33.92744 21.02084 29.95687 25.80393
[73] 21.49924 34.79575 34.57705 30.03549 16.59638 27.22300 24.35621 22.86793
[81] 23.19511 33.74015 24.08401 26.39419 32.33359 11.36287 21.16515 17.21277
[89] 22.73049 25.57669 11.97590 25.35846 33.32411 22.09122 25.19977 30.83577
[97] 23.63709 19.66773 24.64870 17.03159 22.27101 15.92557 32.54150 22.40203
[105] 27.41638 31.99645 30.05415 20.01602 34.47373 28.75978 29.91249 15.71646
[113] 21.76610  9.62656 28.53965 29.49417 29.55520 19.44178 21.43751 30.40382
[121] 28.50982 21.48847 31.10134 35.99067 26.42877 26.28329 20.72639 21.09140
[129] 23.79292 18.79569 29.54042 18.78700 21.41071 21.29782 33.91853 27.20808
[137] 17.14895 28.72738 37.73840 34.14567 15.36874 29.50350 25.70400 20.07219
[145] 24.53928 17.84487 23.35766 15.71243 19.89034 25.37234 16.12745 24.18529
[153] 21.97051 13.08172 32.00764 30.59653 29.74443 27.88354 11.12451  7.07505
[161] 23.13726 21.09331  8.20002 20.62572 18.24345 11.13478 24.43943 22.24289
[169] 24.67325 41.03924 33.31795 27.98844 31.57994 36.73337 18.45322 23.60431
[177] 24.33489 35.74662 22.75684 34.42403 25.34866 25.06486 36.78028 30.67131
[185] 37.82860 25.44086 29.49548 23.27255 28.69506 26.52036 27.49975 24.66196
[193] 25.81562 22.49429 26.88264 28.98200 34.31875 26.40231 21.67621 27.56805
[201] 16.16277 18.28963 30.55183 20.70871 17.66476 22.79543 24.03419 18.02351
[209] 21.63102 27.54976 14.79717 25.34921 21.34916 21.20080 28.36006 22.43773
[217] 25.56306 19.50634 28.93651 31.79897 30.80067 33.95134 39.43243 23.91465
[225] 20.97560 20.02507 33.44170 17.32813 32.61856 36.77906 16.84776 31.53069
[233] 23.18250
```

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

```
[1] 25.06486
```

```
expr$geno == "G/G"
```

```
[1] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
[13] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE
[25] FALSE FALSE FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE
[37] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE
[49]  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE
[61]  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
[73]  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
[85]  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE
[97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE
```

```

[109] TRUE TRUE TRUE FALSE FALSE TRUE TRUE FALSE TRUE TRUE TRUE FALSE
[121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE TRUE
[133] FALSE FALSE TRUE FALSE FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE
[145] FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE TRUE
[157] FALSE FALSE TRUE FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE
[169] FALSE TRUE TRUE TRUE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE
[181] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE
[193] TRUE TRUE TRUE FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE FALSE
[205] FALSE FALSE TRUE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE
[217] FALSE TRUE FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE
[229] FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[241] TRUE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE
[253] TRUE TRUE FALSE FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE FALSE
[265] FALSE FALSE TRUE TRUE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE
[277] FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE TRUE FALSE TRUE TRUE
[289] FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[301] FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE
[313] FALSE TRUE FALSE TRUE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE
[325] FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE
[337] FALSE FALSE FALSE TRUE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE
[349] FALSE FALSE TRUE FALSE FALSE FALSE TRUE TRUE TRUE FALSE FALSE FALSE
[361] TRUE TRUE FALSE TRUE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE
[373] TRUE FALSE TRUE TRUE FALSE TRUE TRUE TRUE TRUE FALSE TRUE FALSE
[385] TRUE FALSE FALSE FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE FALSE
[397] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[409] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[421] TRUE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE
[433] FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[445] FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE
[457] TRUE TRUE FALSE FALSE FALSE FALSE

```

```
expr$exp[expr$geno == "G/G"]
```

```

[1] 18.25141 17.67473 18.55622 23.10383 30.94554 21.14387 18.39547 12.02809
[9] 17.44761 29.82254 23.01983 13.42470 22.65437 11.07445 28.35841 28.79371
[17] 27.08956 16.11138 26.61928 30.18323 19.40790 19.52301 26.56808 17.34076
[25] 10.74263 16.66051 29.01720 20.69333 21.15677 18.58691 19.04962 22.81974
[33] 32.01142 21.12823 18.61268 19.37093 31.42162 16.67764 19.08659 21.55001
[41] 8.29591 12.58869 17.34109 28.23642 19.99979 25.55413 24.45672 23.53572
[49] 22.48273 14.66862 33.95602 18.26466 16.06661 17.32504 19.14766 12.57599
[57] 22.28749 17.29261 24.18141 16.07627 14.80495 23.46573 28.97074 27.78837

```



```
[65] 23.92355  9.55902 12.35836 22.53910 21.98118 16.40569 25.21931 24.32857
[73] 19.42882 26.56993 13.34557 16.60507 24.85165 21.56943 23.95528 16.18962
[81] 22.53720 26.04123  6.67482 20.07363 19.76527 18.50772 20.14146 18.07151
[89]  6.94390 22.14277 14.23742 19.85388 27.73467 19.02064 14.49816 26.78940
[97] 20.84709 10.77316 12.82128 16.90256 29.60045 14.81945 17.46326 23.26922
[105] 21.39806 18.06320 15.91528 24.80823 26.04514 18.28089 23.24907 17.91118
[113] 21.09502 24.74366 27.40521 24.85772 23.08482 16.56929 16.69044 25.08880
[121] 32.78519
```

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

```
[1] 20.07363
```

##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 108 samples with the A/A genotype. There are 233 samples with the A/G genotype. There are 121 samples with the G/G genotype.

A/A: 31.24847 A/G: 25.06486 G/G: 20.07363

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

```
[1] 31.24847
```

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

```
[1] 25.06486
```

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

```
[1] 20.07363
```

```
#Function used to determine sample size per genotype.
table(expr$geno)
```

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

```
#What to know how many sample are in the dataframe? This function will tell us the number  
nrow(expr)
```

```
[1] 462
```

```
#Specifically tells us how many samples are found in each type:  
table(expr$geno)
```

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

```
#Install ggplot2 so we could make a boxplot
```

```
library(ggplot2)
```

##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?

Genotype A/A had higher levels of expression than genotype G/G. According to the boxplot the SNP does effect the expression of ORMDL3.

##Following code is used to make ggplot

```
#input the dataframe first  
#Aes(include x, y data, fill(is to indicate to fill in the box with the color)=color based  
#geom_boxplot: function indicates to make a boxplot  
ggplot(expr) + aes(x=geno, exp, fill=geno) + geom_boxplot(notch=TRUE)
```

