

lab12

Karina Cardenas, A16742606

2025-05-11

Table of contents

| | |
|---|---|
| Section 1: Gene proportions in population | 1 |
| Section 4: plotting | 2 |

Section 1: Gene proportions in population

Finding the Proportion of Homozygous G|G genes in AMR,MXL descent

```
GG_proportion <- read.csv("sample_genotypes.csv")
table(GG_proportion$Genotype..forward.strand.)
```

```
A|A  A|G  G|A  G|G
22   21   12   9
```

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

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

Now lets look at a different population, I picked GBR.

```
GBR_pop <- read.csv("GBR_pop.csv")
head(GBR_pop)
```

| | 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 the proportion of G|G

```
round(table(GBR_pop$Genotype..forward.strand.)/nrow(GBR_pop) * 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: plotting

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.

The file below contains the Genotype expression results of one individual male with childhood asthma. Read this file into R and determine the sample size for each genotype and their corresponding median expression levels for each of these genotypes.

```
expression <- read.table("Expression_genotype.txt")
head(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 |

Q. how many samples do we have ?

```
nrow(expression)
```

```
[1] 462
```

Q. sample size of genotype expression?

```
table(expression$geno)
```

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

Q. Median Expression of genotypes?

```
exp_sum <- tapply(expression$exp, expression$geno, median, na.rm = TRUE)
exp_sum
```

```
A/A      A/G      G/G
31.24847 25.06486 20.07363
```

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(expression) +
  aes(x = geno, y = exp, fill = geno) +
  geom_boxplot(notch = TRUE) +
  geom_jitter(width = 0.2, size = 1, alpha = 0.2) +
  labs(x = "Genotype", y = "Expression")
```

