# Name- Nitin S Mali Ruid: 176007106

# Questions

1. **Using all the control and affected individuals, calculate for each SNP locus the number of individuals having each of the three possible genotypes (“AA”, “Aa”, and “aa”). Write a custom R script that determines the likelihood that the genotype at the locus is significantly different in Parkinson’s Disease individuals versus control individuals at each locus, using chi-squared testing. You will need to decide what files will provide the**

information you need to compute the test. List the top ten SNP loci associated with Parkinson’s Disease, ordered by chi-squared test p-value.

**Solution:**

**sorted\_df\_comb\_by\_p\_values[1:10,]**

chromosome locus AA\_cc Aa\_cc aa\_cc AA\_pd Aa\_pd aa\_pd chi\_sq P\_Values

6925 1 rs1556982 270 0 0 0 0 0 NA 1.134774e-60

26186 1 rs6540798 270 0 0 0 0 0 NA 1.134774e-60

40871 2 rs3771218 270 0 0 0 0 0 NA 1.134774e-60

45290 2 rs13395005 270 0 0 0 0 0 NA 1.134774e-60

76420 3 rs1074985 270 0 0 0 0 0 NA 1.134774e-60

127149 5 <NA> 270 0 0 0 0 0 NA 1.134774e-60

164863 6 rs7749176 270 0 0 0 0 0 NA 1.134774e-60

178331 7 rs2371543 270 0 0 0 0 0 NA 1.134774e-60

335623 16 rs8055783 270 0 0 0 0 0 NA 1.134774e-60

343346 17 rs7225160 270 0 0 0 0 0 NA 1.134774e-60

1. Why is chi-square an appropriate statistic to use for this analysis?

**Solution:**

chi-square is an appropriate statistic test which is used to determine whether the genotype distribution seen in Parkinson’s Disease patients is different than control individuals (i.e. a case-control study). A chi-squared test for these data with 2 degrees of freedom yields a p-value of 6.301 x 10-6, indicating that it is highly unlikely that the Parkinson’s Disease distribution matches the control distribution this would indicate that the genotype distributions are significantly associated with the presence of Parkinson’s Disease.

1. Draw out a representative chi-square table for the SNP locus (rs3741411) on chromosome 11, manually calculate the chi-squared statistic, and use R to get the p-value. Show all work.

**Solution:**

SNP Locus= rs3741411 ,

The chi sq sorted\_df\_comb\_by\_p\_values

**[sorted\_df\_comb\_by\_p\_values$locus == "rs3741411", ]**

**y = matrix(c(184,13,161,9),byrow =TRUE,nrow=2)**

**y**

chisq.test(y) $p.value

[,1] [,2]

[1,] 184 13

[2,] 161 9

The Manual calculation of chi square is :-

Chi-square = (observed – Expected)^2/Expected

= (161 – 184)^2 / 184

= 2.875

=(9 – 13)^2 / 13

= 1.230

Chi-square = 4.105

1. How many SNP loci have a p-value of < 0.05? What does a p-value of 0.05 mean ?

**Solution:**

**nrow(sorted\_df\_comb\_by\_p\_values[sorted\_df\_comb\_by\_p\_values$P\_Values<0.05,])**

[1] 21741

1. You computed a chi-squared test to associate genotypes (AA, Aa, aa) to disease state vs. control. We can also compute the chi-squared test associating individual alleles (A or a) to disease. Compute the allele-specific chi-squared test for chromosome 11 and draw out a representative table for the locus above (rs3741411). Why would we want to conduct this test versus one on the genotype level? Also, list the top ten SNP loci associated with Parkinson’s Disease (for chromosome 11), ordered by chi-squared test p-value.

**Solution:**

**> data <- read.table("parkinsons.assoc", sep = "" , header = T , na.strings ="",**

**stringsAsFactors= F)**

**> data1=data[ order(data[,9]), ]**

**> data1[1:10,]**

CHR SNP BP A1 F\_A F\_U A2 CHISQ P OR

133821 5 rs6596287 135507247 C 0.3537 0.2454 A 15.13 0.0001004 1.683

124182 5 rs1374065 61099359 G 0.4703 0.3537 C 15.11 0.0001012 1.622

133832 5 rs6596293 135544606 G 0.3569 0.2481 A 15.1 0.0001018 1.681

280362 12 rs2470378 69512747 C 0.1685 0.2657 T 15.02 0.0001064 0.5602

23137 1 rs842796 189634072 A 0.4627 0.3469 G 15.01 0.000107 1.621

76961 3 rs4396907 83989494 C 0.3327 0.2269 T 15 0.0001077 1.699

343820 17 rs2525569 26700097 T 0.4331 0.3192 A 14.93 0.0001116 1.629

302920 13 rs562992 109706778 C 0.1123 0.2015 A 14.88 0.0001145 0.5013

343814 17 rs7503922 26666061 C 0.4368 0.3229 A 14.88 0.0001146 1.626

206639 8 rs3802266 124054889 G 0.4 0.5166 A 14.81 0.0001186 0.6238