hla\_analysis

David Martino

2 July 2015

stage 1: exploration of data

setwd("/Users/david/Documents/projects/healthnuts/hla/imputationData/imp2.v2/HN\_hla")  
  
rawdata=read.delim('HN2.hla.csv',header=T,sep=",")  
imputes=read.delim('HN2.hla.txt',header=T,sep="")  
phenotypes=read.delim('PEANUT\_ALLERGY.4PLINK.phen',sep="",header=T)  
  
head(rawdata)

## IndividualID Chromosome Gene Allele Posterior  
## 1 2678 1 A 201 1.0000000  
## 2 2678 2 A 201 1.0000000  
## 3 3583 1 A 101 1.0000000  
## 4 3583 2 A 201 1.0000000  
## 5 127 1 A 6801 1.0000000  
## 6 127 2 A 3303 0.9933333

tail(rawdata)

## IndividualID Chromosome Gene Allele Posterior  
## 7931 2876 1 DRB1 101 1.0000000  
## 7932 2876 2 DRB1 101 1.0000000  
## 7933 781 1 DRB1 1101 1.0000000  
## 7934 781 2 DRB1 1102 0.9666667  
## 7935 2653 1 DRB1 1001 1.0000000  
## 7936 2653 2 DRB1 405 0.9800000

table(rawdata$Gene)

##   
## A B C DPA1 DPB1 DQA1 DQB1 DRB1   
## 992 992 992 992 992 992 992 992

table(rawdata$Chromosome)

##   
## 1 2   
## 3968 3968

summary(rawdata$Posterior)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.1067 0.9800 1.0000 0.9512 1.0000 1.0000

merge imputations with phenotypes

data=merge(imputes,phenotypes,by.x="IID",by.y="IID")

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.