

Linkage Disequilibrium: Lab 1 Solutions

Exercise 1

1. Pull out the site combination from sites 75 & 76 from the hap1 object.
 - Demonstrate the expected and observed haplotype distributions for these two sites?

```
load("haplotypes_chr1.RData")
hapsite75 <- hap1[, 75]
hapsite76 <- hap1[, 76]
chisq.test(hapsite75, hapsite76)$expected

##           hapsite76
## hapsite75      A      B
##           A 55.068 1.9322
##           B  1.932 0.0678

chisq.test(hapsite75, hapsite76)$observed

##           hapsite76
## hapsite75      A      B
##           A  57      0
##           B   0      2
```

- What is the value of χ_1^2 and r^2 ?

```
X2 <- chisq.test(hapsite75, hapsite76, correct = FALSE)$statistic
X2

## X-squared
##           59

N <- sum(chisq.test(hapsite75, hapsite76)$observed)
N

## [1] 59

r2 <- X2/N
r2

## X-squared
##           1
```

- What can you say about the LD of these two sites?

These two sites have VERY HIGH LD. In fact from this sample they seem perfectly correlated.

2. Find the r^2 values of on site combination from columns 75 & 77, and 75 & 80 of the hap1 object. What do you notice about the LD relationship between these two combinations (recall that they are in order by position)?

```
hapsite77 <- hap1[, 77]
hapsite80 <- hap1[, 80]
r2sites75and77 <- chisq.test(hapsite75, hapsite77, correct = FALSE)$statistic/sum(chisq.test(hap
  hapsite77)$observed)
```

```

r2sites75and77

## X-squared
##          0

r2sites75and80 <- chisq.test(hapsite75, hapsite80, correct = FALSE)$statistic/sum(chisq.test(hap
  hapsite80)$observed)
r2sites75and80

## X-squared
##      0.0481

```

The r^2 value of site 77 is smaller than that for site 80, even though site 80 is further apart.

```

ld.Rsq <- function(hapmat) {
  numvec <- as.matrix(c(1:ncol(hapmat)))
  ld <- function(hapcol, hapmat) {
    x1 <- hapmat[, hapcol]
    rsq <- function(z) {
      N <- sum(chisq.test(x1, z)$observed)
      chisq.test(x1, z, correct = FALSE)$statistic/N
    }
    apply(hapmat, 2, rsq)
  }
  map <- apply(numvec, 1, ld, hapmat)
  colnames(map) <- rownames(map)
  bot <- upper.tri(map, diag = TRUE)
  map[bot] <- NA
  signif(map, 4)
}

```

Exercise 2

1. Run `ld.Rsq()` on sites from columns 75-80 from the `hap1` object and compare the map with your values from Exercise 1.

```

hapsites75to80 <- hap1[, 75:80]
LDhapsites75to80 <- ld.Rsq(hapsites75to80)
LDhapsites75to80

##          rs161826  rs161822  rs2453021  rs9657980  rs2493215  rs226473
## rs161826          NA          NA          NA          NA          NA          NA
## rs161822  1.000000          NA          NA          NA          NA          NA
## rs2453021  0.000000  1.248e-05          NA          NA          NA          NA
## rs9657980  0.002646  2.552e-03  0.0231000          NA          NA          NA
## rs2493215  0.002836  3.126e-03  0.5709000  0.0352500          NA          NA
## rs226473   0.048100  4.514e-02  0.0002722  0.0003344  0.002622          NA

```

```

require(reshape)

## Loading required package: reshape
## Loading required package: plyr

```

```
##
## Attaching package: 'reshape'
##
## The following objects are masked from 'package:plyr':
##
##   rename, round_any

require(ggplot2)

## Loading required package: ggplot2

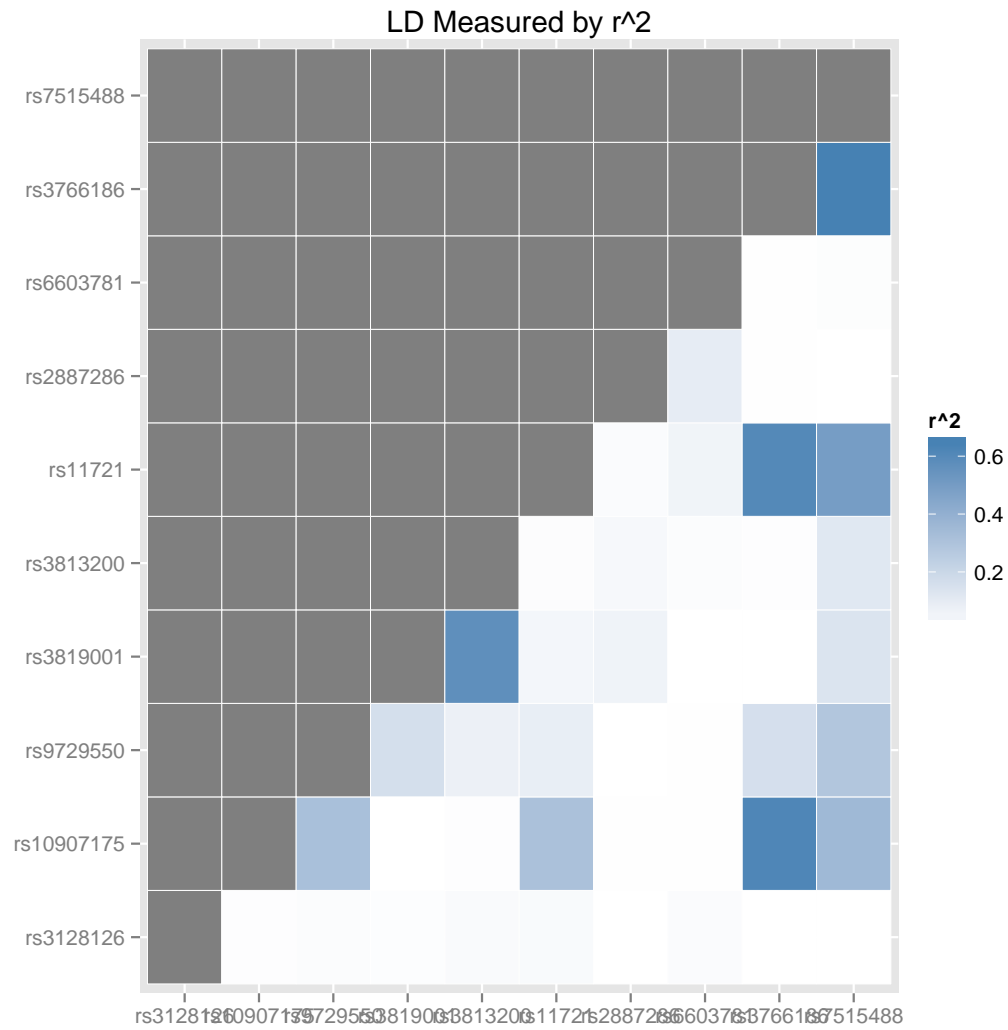
heatmapHaps <- function(hapmap) {
  meltedr2 <- melt(hapmap)

  meltedr2$X1 <- factor(meltedr2$X1, as.character(meltedr2$X1))
  meltedr2$X2 <- factor(meltedr2$X2, as.character(meltedr2$X2))

  p <- ggplot(meltedr2, aes(X1, X2)) + geom_tile(aes(fill = value), colour = "white") +
    scale_fill_gradient(low = "white", high = "steelblue", name = "r^2") +
    xlab(NULL) + ylab(NULL) + ggtitle("LD Measured by r^2")
  p
}
```

2. Built a new LD map of the first ten SNPs, and then input it into heatmapHaps(). What site appears to have the highest average LD with the other sites in the LD map?

```
hapsites1to10 <- hap1[, 1:10]
LDhapsites1to10 <- ld.Rsq(hapsites1to10)
heatmapHaps(LDhapsites1to10)
```



```
save.image("LD_Lab1_Workspace.RData")
```