### Background

In this exercise, we will examine a region of chromosome 8 (8q24) that has been implicated in colon cancer. Per the Stanford Personal Genomics lectures, we started with rs6983267 from the ceu population, expanding the window somewhat to see what we see.

library(knitr)

str <-("../Data/rs6983267.ceu.vcf.gz")  
dat <-read.vcf(str)

## File apparently not yet accessed:  
## Scanning file ../Data/rs6983267.ceu.vcf.gz   
##   
 1.762995 Mb  
## Done.  
##   
Reading 100 / 10000 loci  
Reading 200 / 10000 loci  
Reading 300 / 10000 loci  
Reading 400 / 10000 loci  
Reading 500 / 10000 loci  
Reading 600 / 10000 loci  
Reading 700 / 10000 loci  
Reading 800 / 10000 loci  
Reading 900 / 10000 loci  
Reading 1000 / 10000 loci  
Reading 1100 / 10000 loci  
Reading 1200 / 10000 loci  
Reading 1300 / 10000 loci  
Reading 1400 / 10000 loci  
Reading 1500 / 10000 loci  
Reading 1600 / 10000 loci  
Reading 1700 / 10000 loci  
Reading 1800 / 10000 loci  
Reading 1900 / 10000 loci  
Reading 2000 / 10000 loci  
Reading 2100 / 10000 loci  
Reading 2200 / 10000 loci  
Reading 2300 / 10000 loci  
Reading 2400 / 10000 loci  
Reading 2500 / 10000 loci  
Reading 2600 / 10000 loci  
Reading 2700 / 10000 loci  
Reading 2800 / 10000 loci  
Reading 2900 / 10000 loci  
Reading 3000 / 10000 loci  
Reading 3100 / 10000 loci  
Reading 3179 / 3179 loci.  
## Done.

dat.info <-VCFloci(str)

## Scanning file ../Data/rs6983267.ceu.vcf.gz   
##   
 1.762995 Mb  
## Done.

Pretty good set of loci. Now, how to visualize?

Let's start by subsetting, taking a random sample of 100 SNPs and 20 inviduals:

snp.sub <-sample(1:ncol(dat),100)  
samp.sub <-sample(1:nrow(dat),20)  
dat.sub <-dat[samp.sub,snp.sub]

sm <-summary(dat.sub)

Here's a possibility. Start with the whole database and get the ones with more than one allele

als <-getAlleles(dat)  
 poly2 <-sapply(als,length)

OK. Now get a subset of polymorphisms

loc.poly <-which(poly2==2)

Superb. Now subset dat by that

dat.sub <-dat[,loc.poly]

Now we can visualize a chunk

kable(data.frame(dat.sub[1:20,1:10]))

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | rs112983285 | rs283727 | rs72728054 | rs283728 | rs72728056 | rs111478046 | rs541641381 | rs16902085 | rs4871015 | rs34882299 |
| NA06984 | G|G | G|A | T|T | T|A | C|C | TG|TG | G|G | A|G | G|A | A|G |
| NA06985 | G|G | G|A | T|T | T|A | C|C | TG|TG | G|G | A|A | G|A | A|G |
| NA06986 | G|G | G|A | T|T | T|A | C|C | TG|TG | G|G | A|G | G|A | A|G |
| NA06989 | G|G | A|G | T|T | A|T | T|C | TG|TG | G|G | A|A | A|G | G|A |
| NA06994 | G|G | G|A | T|T | T|A | C|C | TG|TG | G|G | A|A | G|A | A|G |
| NA07000 | G|G | G|G | T|T | T|T | C|C | TG|TG | G|G | A|A | G|G | A|A |
| NA07037 | G|G | G|A | T|T | T|A | C|C | TG|TG | G|G | A|A | G|A | A|G |
| NA07048 | G|G | A|A | T|T | A|A | C|C | TG|TG | G|G | A|A | A|A | G|A |
| NA07051 | G|G | G|G | T|T | T|T | C|C | TG|T | G|G | A|A | G|G | A|A |
| NA07056 | G|G | G|A | T|T | T|A | C|C | TG|TG | G|G | A|A | A|A | A|G |
| NA07347 | G|G | G|A | T|T | A|A | C|C | TG|TG | G|G | A|G | G|A | G|G |
| NA07357 | G|G | A|A | T|T | A|A | C|C | TG|TG | G|G | A|A | A|A | G|G |
| NA10847 | G|G | A|G | T|T | A|T | C|C | TG|TG | G|G | A|A | A|G | G|A |
| NA10851 | G|G | A|A | T|T | A|A | C|C | TG|TG | G|G | G|A | A|A | G|G |
| NA11829 | G|G | G|G | T|T | T|T | C|C | TG|TG | G|G | A|A | G|G | A|A |
| NA11830 | G|G | A|G | T|T | A|T | C|C | TG|TG | G|G | A|A | A|G | G|A |
| NA11831 | G|G | G|G | T|T | T|T | C|C | TG|TG | G|G | A|A | G|G | A|A |
| NA11832 | G|G | A|A | T|T | A|A | C|C | TG|TG | G|G | G|A | A|A | G|G |
| NA11840 | G|G | G|A | T|T | T|A | C|C | TG|TG | G|G | A|A | G|A | A|G |
| NA11843 | G|G | G|A | T|T | T|A | C|C | TG|TG | G|G | A|A | G|A | A|G |