

Targeted Sequencing

Samuel G. Younkin
syounkin@jhsph.edu

November 21, 2012

```
> vcfSamples <- scan(gzfile("/thumper/ctsa/beaty/targeted_seq/snvs.fixed_header.vcf.gz"),
  what = character(0), skip = 1422, nlines = 1, sep = "\t")
> vcfSampNames <- gsub(".*-", "", vcfSamples[-(1:9)])
> pedigreeInfoTrios <- read.xlsx("/home/bst/student/syounkin/targeted-sequencing/data/deepseq_alltrios_
  sheetIndex = 1, colIndex = 1:16, stringsAsFactors = FALSE)

> ped.df <- with(pedigreeInfoTrios, {
  Pop.fac <- as.factor(Population)
  k <- which(levels(Pop.fac) == "Iowa")
  levels(Pop.fac)[k] <- "IOWA"
  data.frame(Population = Pop.fac, Sample.Name = as.character(Sample.Name),
    PI = as.factor(PI))
})

> index <- which(pedigreeInfoTrios$Sample.Name %in% vcfSampNames)
> ped.vcf.df <- ped.df[index, ]
> with(ped.vcf.df, table(Population, useNA = "ifany"))
```

Population				
BEIJING	DENMARK	GUATEMALA	HUNGARY	IOWA
171	11	24	131	410
MADRID	PHILIPPINES	PITTSBURGH	SHANG	SOFC
62	1656	80	69	15
TEXAS	TURKEY	<NA>		
218	21	1129		

```
> with(ped.vcf.df, table(PI, useNA = "ifany"))
```

PI					
Beaty	Hecht	Lidral	Marazita	Murray	Wehby
1129	218	81	584	1861	124

```
> nrow(ped.vcf.df)

[1] 3997

> library("vcf2R")

> file = "/home/bst/student/syounkin/targeted-sequencing/data/snvs.fixed_header.hwe"
> hwe.df <- hwe2R(file = file)
```

```
> with(hwe.df, hist(p.value, breaks = 100))
```

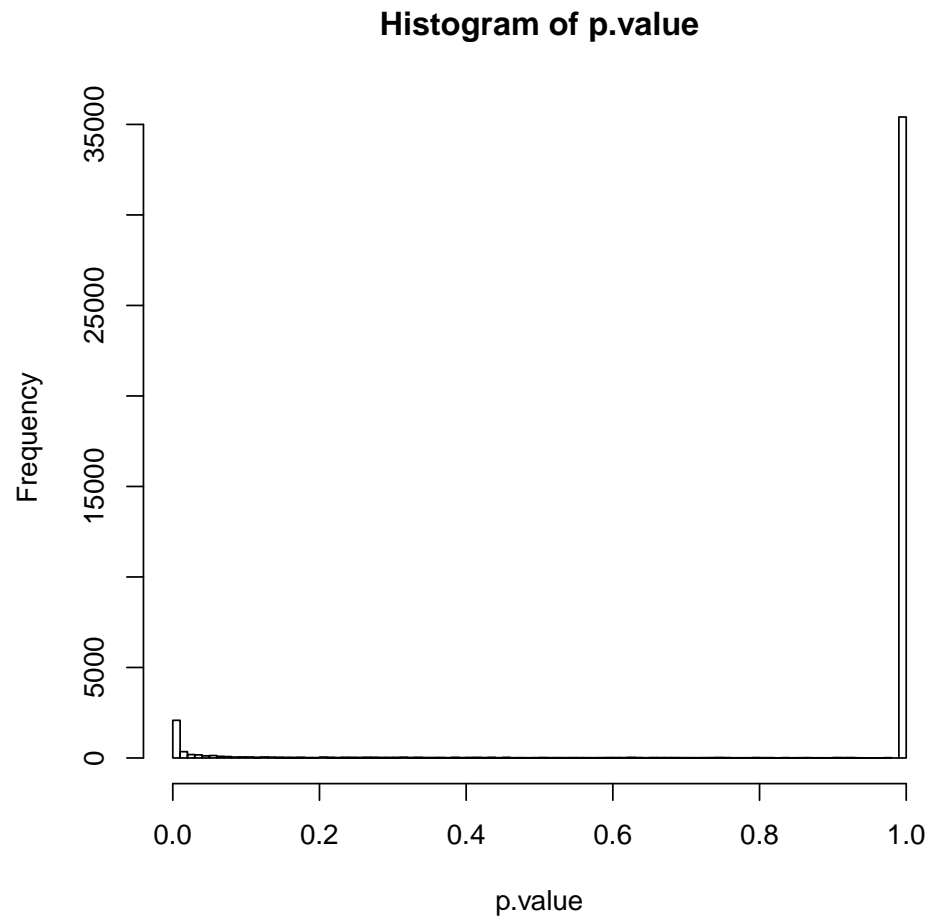
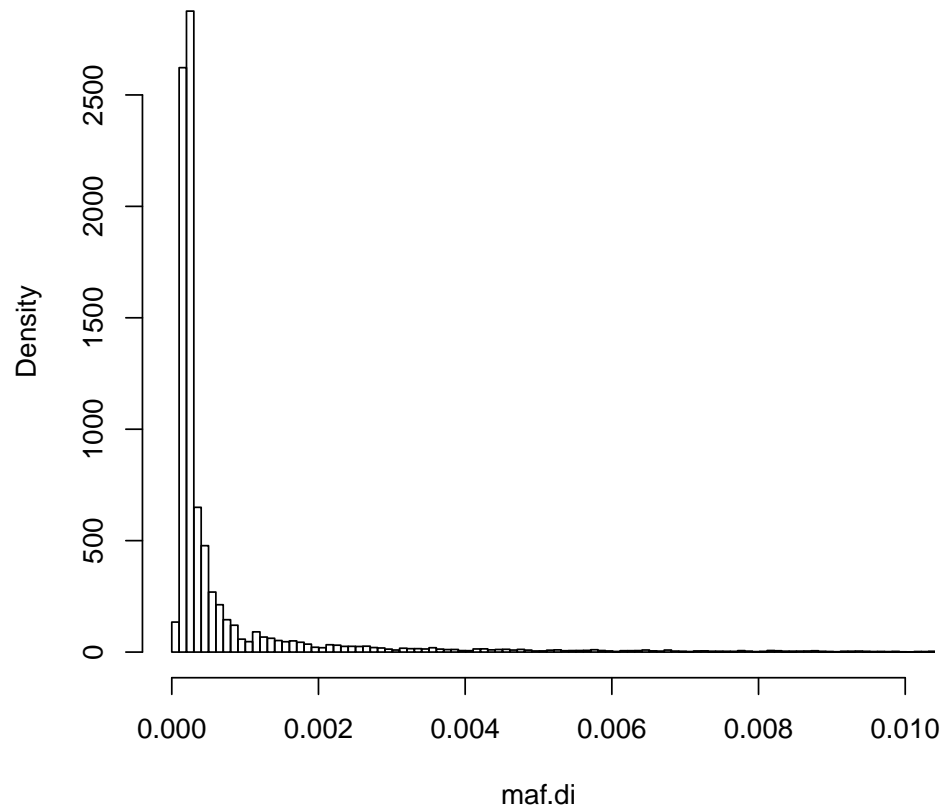


Figure 1: Histogram of Hardy-Weinberg p -values

Histogram of maf.di



```
> file = "/home/bst/student/syounkin/targeted-sequencing/data/snvs.fixed_header.frq"
> maf.di <- frq2R(file = file)

> summary(maf.di)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
0.0000 0.0001  0.0002  0.0195 0.0009  0.4997   1737
```

Histogram of maf.di

