Penetrance Probability Distributions

Brett

January 04, 2018

Contents

Introduction:

This file is used to visualize probability distributions derived from the SCN5A dataset. several representative variants' probability distributions are plotted. At the end, probability distributions are calculated for representative "classified variants" using the classification scheme put forward by the ACMG. The final calculations are the probabilities variants classifiable as (likely) pathogenic, VUS, or (likley) benign have a true penetrance (individuals presenting with either BrS1 or LQT3) greater than 20%. There is also a sensitivity analysis done at the end.

Read in data

```
con = dbConnect(SQLite(),
dbname="/Users/B/Dropbox/SCN5A/BrettsSandbox/paper/data/VariantSCN5A-new.db")
alltables = dbListTables(con)
my.data <- dbReadTable(con, 'VariantSCN5A')
my.data[my.data=='NA'] <- NA
d<-my.data
dbDisconnect(con)

d$resnum<-as.integer(d$resnum)
d$gnomAD[is.na(d$gnomAD)] <- 0
```

```
d$gnomAD<-as.numeric(d$gnomAD)
d$ipeak<-100*as.numeric(d$ipeak)
```

```
## Warning: NAs introduced by coercion
```

```
d$vhalfact<-as.numeric(d$vhalfact)
```

```
## Warning: NAs introduced by coercion
```

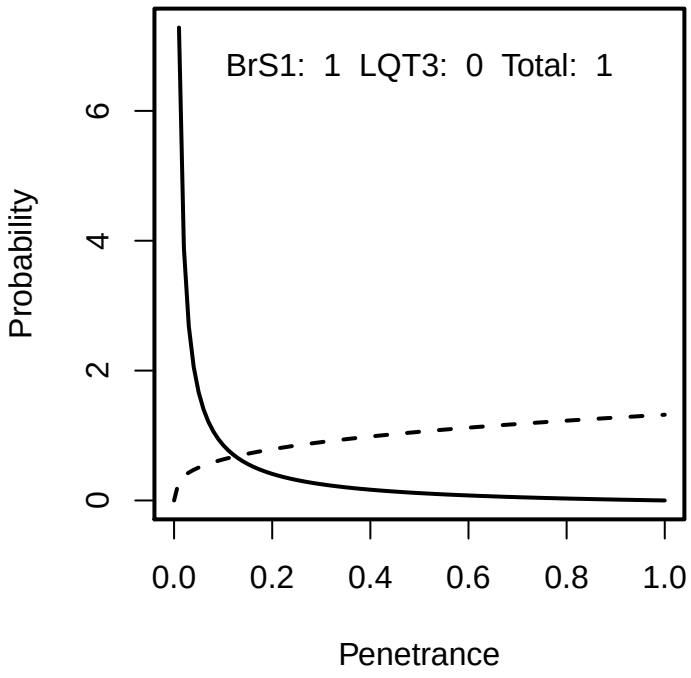
```
d$vhalfinact<-as.numeric(d$vhalfinact)
d$recovfrominact<-log10(100*as.numeric(d$recovfrominact))
d$ilate[as.numeric(d$ilate)==0]<-NA
d$ilate_norm<-log10(d$ipeak*as.numeric(d$ilate)+0.00001)
d$ilate<-log10(100*as.numeric(d$ilate)+0.00001)
d$total_carriers<-d$lqt3+d$brs1+d$unaff+d$gnomAD
d$weight = 1-1/(0.1+d$total_carriers) #weights
d$weightsMilder = 1-1/(1+d$total_carriers) #weights
d$noweights = rep(1,length(d$total_carriers))


servers<-read.csv("/Users/B/Dropbox/SCN5A/BrettsSandbox/paper/data/annotated_variants-trim.txt", sep =
provean <-read.csv("/Users/B/Dropbox/SCN5A/BrettsSandbox/paper/data/provean.txt", sep = "\t")
pph2 <-read.csv("/Users/B/Dropbox/SCN5A/BrettsSandbox/paper/data/pph2-short.txt", sep = "\t")
sift <-read.csv("/Users/B/Dropbox/SCN5A/BrettsSandbox/paper/data/SIFT.txt", sep = "\t")
d <- merge(d, servers, all = TRUE)
d <- merge(d, provean, all = TRUE)
d <- merge(d, sift, all = TRUE)
d <- merge(d, pph2, all = TRUE)
d<-d[!is.na(d$var), ]

d$eaRate<-as.numeric(d$eaRate)


# Adding in penetrance variables
abrs0=0.32
alqt0=0.11
beta0=1
d$LQT_penetranceBayesian<-(d$lqt3+alqt0)/(d$total_carriers+beta0+alqt0)
d$BrS_penetranceBayesian<-(d$brs1+abrs0)/(d$total_carriers+1+alqt0)
d$all_penetranceBayesian<-(d$brs1+abrs0+d$lqt3+alqt0)/(d$total_carriers+beta0+alqt0+abrs0)
e<-d
```

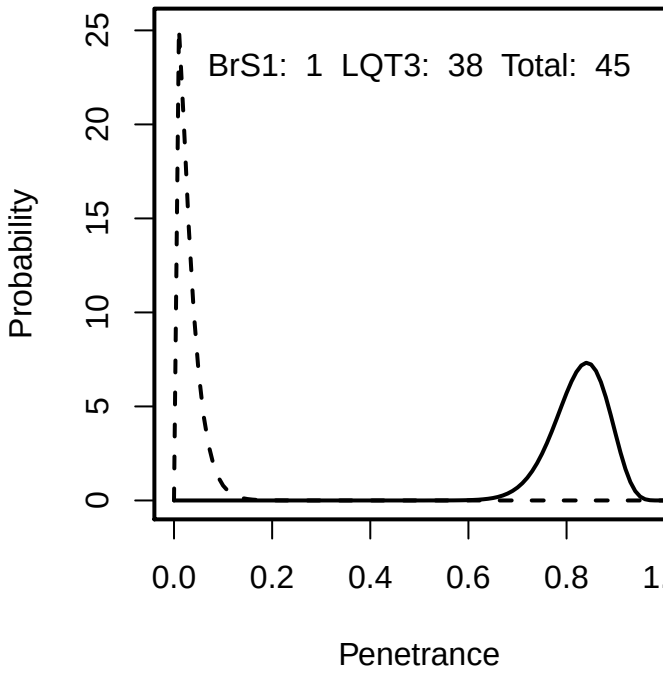
plot probability distributions for BrS1/LQT3 included in main text

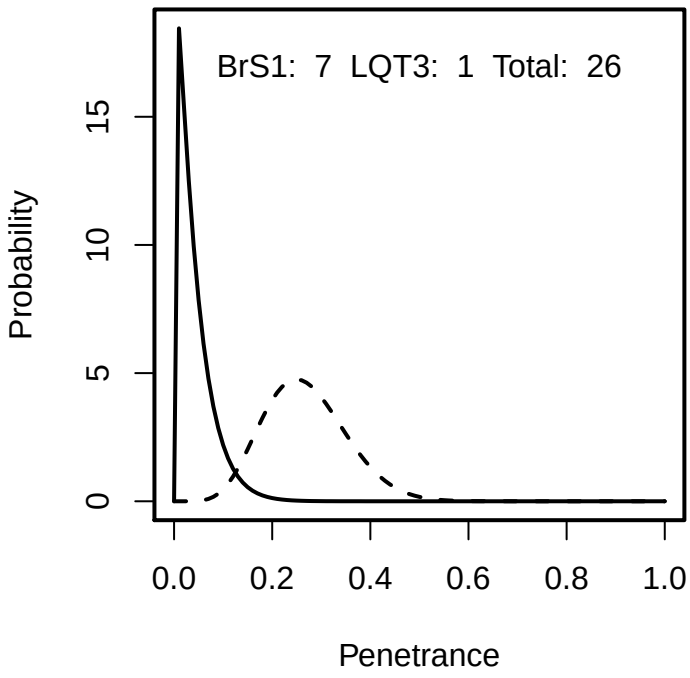
G1748D, D1790G, R965C, L1501V, I1660V, R1644C, T1304M, and E1784K

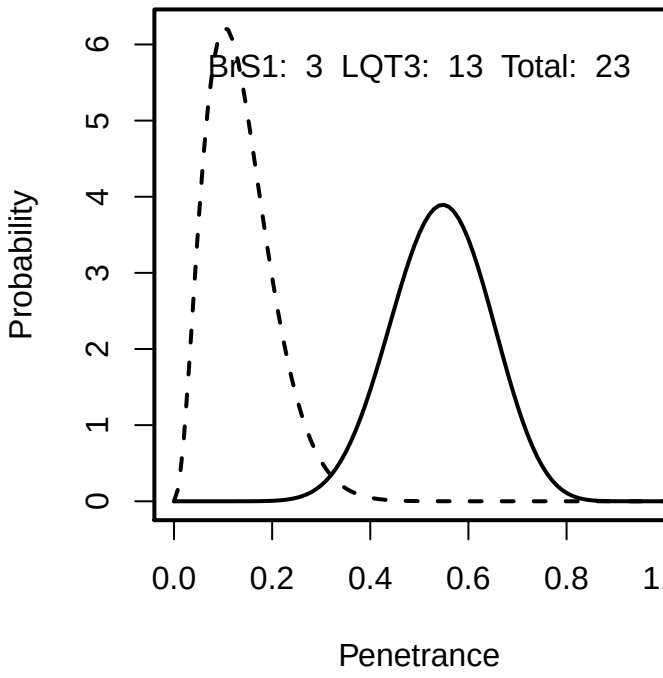
```
uqs<-c("G1748D","D1790G","R965C","L1501V","I1660V","R1644C","T1304M","E1784K")
vois<-c(0)
for(v in 1:length(uqs)){
  vois[v]<-match(uqs[v],d$var)
}
x <- seq(0,.99,0.01)
par(lwd = 2)#, mfrow=c(1,3))
for(i in vois){
  mb <- max(dbeta(x[2:(length(x)-1)],abrs0+d$brs1[i], beta0+(d$total_carriers[i]-d$brs1[i])))
```

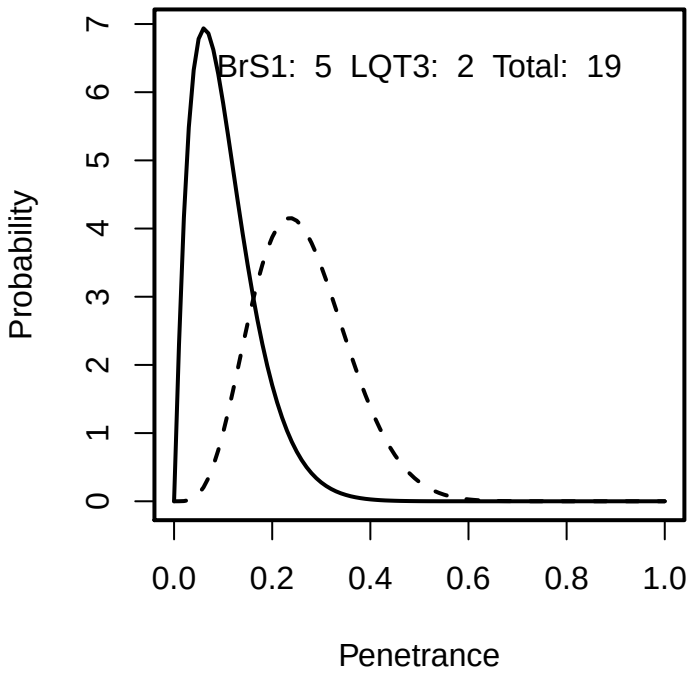
```
  ml <- max(dbeta(x[2:(length(x)-1)],alqt0+d$lqt3[i], beta0+(d$total_carriers[i]-d$lqt3[i])))
  mm <- max(ml,mb)
  curve(dbeta(x, alqt0+d$lqt3[i], beta0+(d$total_carriers[i]-d$lqt3[i])),
        ylab="Probability",
        xlab = "Penetrance", add = FALSE, lty = 1, ylim = c(0,mm))
  curve(dbeta(x, abrs0+d$brs1[i], beta0+(d$total_carriers[i]-d$brs1[i])),
        add = TRUE, lty=2)
  text(0.5,mm*0.92,paste("BrS1: ", d$brs1[i],
                         " LQT3: ", d$lqt3[i],
                         " Total: ", d$total_carriers[i]))
}
```

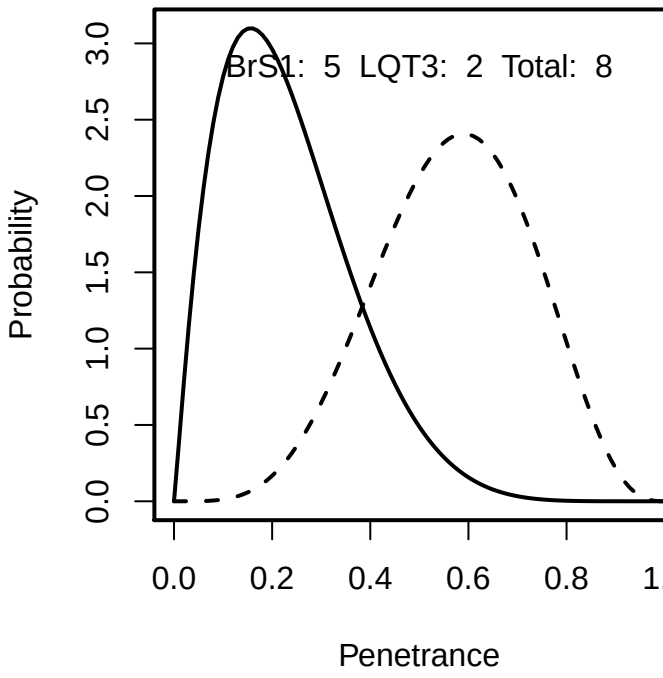


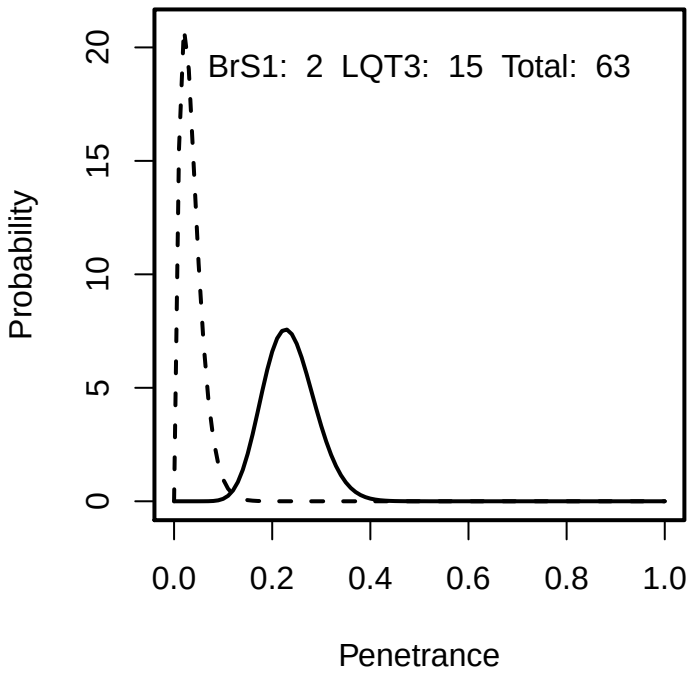


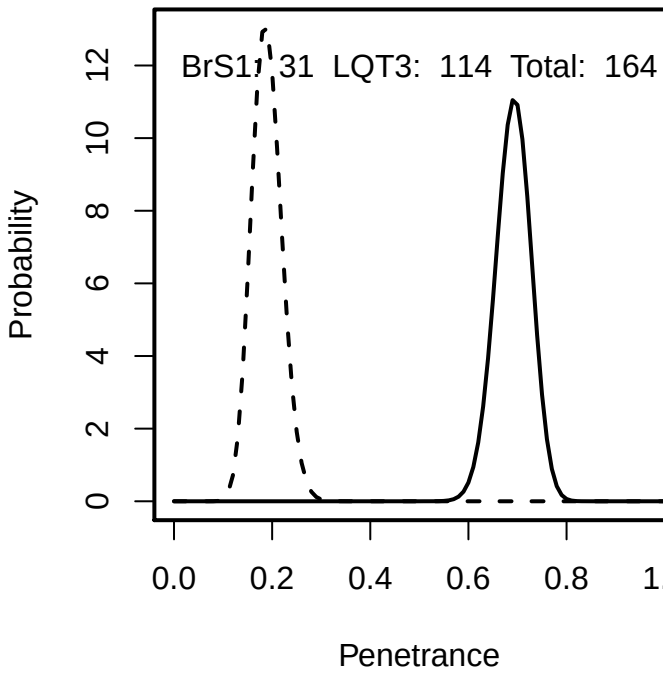

BrS1: 7 LQT3: 1 Total: 26
Probability
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
Penetrance


BrS1: 3 LQT3: 13 Total: 23
Probability
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
Penetrance


BrS1: 5 LQT3: 2 Total: 19
Probability
Penetrance
0 1 2 3 4 5 6 7
0.0 0.2 0.4 0.6 0.8 1.0


BrS1: 5 LQT3: 2 Total: 8
Probability
Penetrance
0.0 0.5 1.0 1.5 2.0 2.5 3.0
0.0 0.2 0.4 0.6 0.8





Plot integrated probability distributions for variants classifiable as (likely) pathogenic, VUS, or (likley) benign

These classifications were made based on the ACMG guidelines from the following website: http://www.medschool.umaryland.edu/Genetic_variant_Interpretation_Tool1.html/

variables used to classfy: PS3/BS3 in vitro functional studies COMPROMIZED/NORMAL PS4 enriched in affected population BS1/PM2 TOO HIGH/LOW in gnomad PP2 Missense var in gene with low rate of benign missense vars PP3/BP4 Multiple lines of computational evivence FOR/AGAINST pathogenicity

Set disease rate and enrichment threasholds

```
dr = 15 # disease rate threashold in gnomad. I'm just going to select one, 8 for BrS1 and 7 for LQT3
et = 0.2 # enrichment threashold. The fraction of carriers who need to present
         # with a disease in order to use PS4. This is subjective
```

Likely pathogenic (III)

given PS4, PP2, PM2

```
d <- e[e$Prediction=="*DAMAGING" &
         (e$prediction=="possiblydamaging" | e$prediction=="probablydamaging"),] # PP3
d <- d[d$gnomAD<=dr | is.na(d$gnomAD),] # PM2
d <- d[(d$lqt3+d$brs1)/(d$total_carriers)>et,] # PS4
d <- d[!is.na(d$var),]
```

```
del <- matrix(0, nrow=length(x), ncol=2)
new.mat.d <- matrix(nrow=length(d$var), ncol=4)
for(i in 1:length(d$var)){
  new.mat.d[i,] <- c(as.character(d$var[i]),
                     d$lqt3[i],d$brs1[i],d$total_carriers[i])

  for(j in 1:length(x)){
    del[j,] <- c(x[j], del[j,2]+integrate(dbeta, shape1 = abrs0+alqt0+as.numeric(new.mat.d[i,2])+as.num
                                          shape2 = (beta0+as.numeric(new.mat.d[i,4])-
                                          as.numeric(new.mat.d[i,2])-as.numeric(new.mat.d[i,3])), x[j],
                 )
  }
}
colnames(new.mat.d) <- c("variant", "lqtpos", "brspos", "vartot")
head(new.mat.d)
```

```
##      variant  lqtpos brspos vartot
## [1,] "A124D"  "0"    "1"    "1"
## [2,] "A178G"  "0"    "1"    "1"
## [3,] "A204V"  "0"    "1"    "1"
## [4,] "A1288G" "0"    "1"    "1"
## [5,] "A1326S" "1"    "0"    "1"
## [6,] "A1330D" "1"    "0"    "1"
```

Likely benign (I)

given PP2, BP4, BS1

```
b <- e[e$Prediction!="*DAMAGING" &
         (e$prediction!="possiblydamaging" & e$prediction!="probablydamaging"),] # BP4
b <- b[b$gnomAD>=dr | is.na(b$gnomAD),] # BS1
b <- b[!is.na(b$var),]

ben <- matrix(0, nrow=length(x), ncol=2)
new.mat.b <- matrix(nrow=length(b$var), ncol=4)
for(i in 1:length(b$var)){
  new.mat.b[i,] <- c(as.character(b$var[i]),b$lqt3[i], b$brs1[i], b$total_carriers[i])
  for(j in 1:length(x)){
    ben[j,] <- c(x[j], ben[j,2]+integrate(dbeta, shape1 = abrs0+alqt0+as.numeric(new.mat.b[i,2])+as.num
                                          shape2 = (beta0+as.numeric(new.mat.b[i,4])-
                                          as.numeric(new.mat.b[i,2])-as.numeric(new.mat.b[i,3])), x[j],
  }
}
colnames(new.mat.b) <- c("variant", "lqtpos", "brspos", "vartot")
head(new.mat.b)
```

```
##      variant lqtpos brspos vartot
## [1,] "A123V" "0"    "0"    "19"
## [2,] "A286S" "0"    "0"    "84"
## [3,] "A572D" "5"    "1"    "1512"
## [4,] "A572S" "1"    "0"    "72"
## [5,] "A572V" "1"    "0"    "72"
## [6,] "A672T" "1"    "0"    "202"
```

VUS

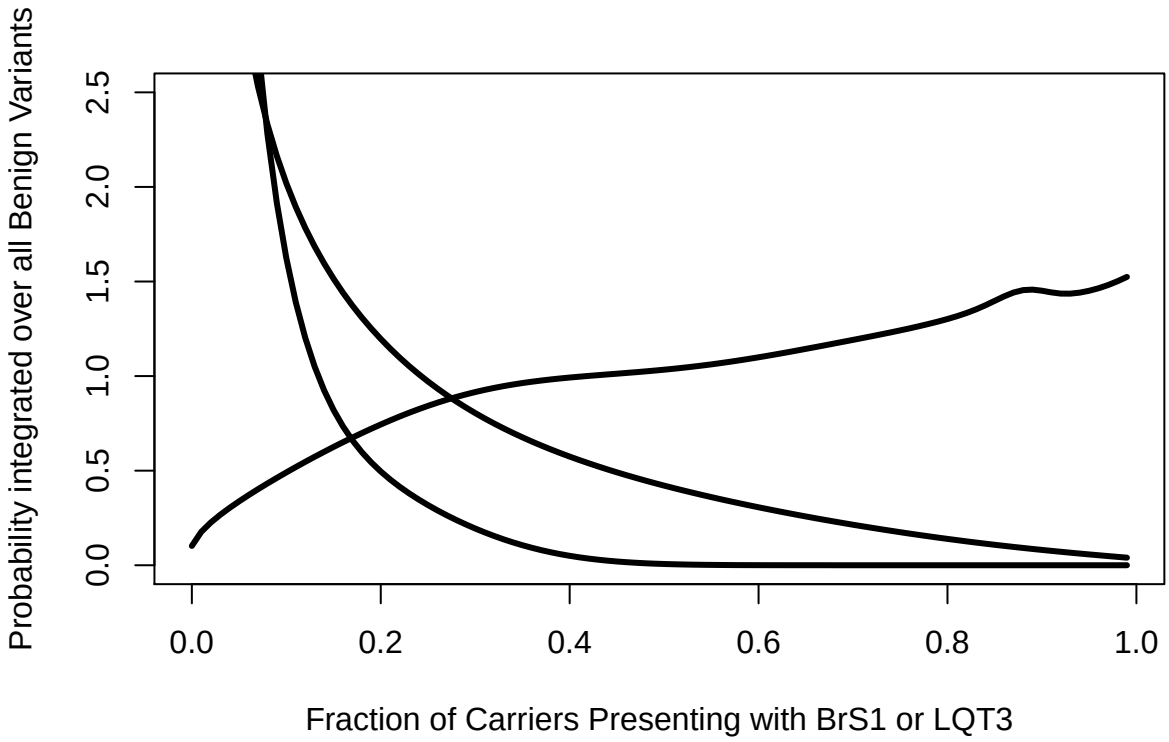
given PM2, BP4, PP2

```
v <- e[e$Prediction!="*DAMAGING" &
         (e$prediction!="possiblydamaging" & e$prediction!="probablydamaging"),] # BP4
v <- v[v$gnomAD<=dr | is.na(v$gnomAD),] # PM2
v <- v[!is.na(v$var),]

# vUSs
vus <- matrix(0, nrow=length(x), ncol=2)
new.mat.v <- matrix(nrow=length(v$var), ncol=4)
for(i in 1:length(v$var)){
  new.mat.v[i,] <- c(as.character(v$var[i]),v$lqt3[i], v$brs1[i], v$total_carriers[i])
  for(j in 1:length(x)){
    vus[j,] <- c(x[j], vus[j,2]+integrate(dbeta, shape1 = abrs0+alqt0+as.numeric(new.mat.v[i,2])+as.num
                                          shape2 = (beta0+as.numeric(new.mat.v[i,4])-
                                          as.numeric(new.mat.v[i,2])-as.numeric(new.mat.v[i,3])), x[j],
                 )
  }
}
colnames(new.mat.v) <- c("variant", "lqtpos", "brspos", "vartot")
head(new.mat.v)
```

```
##      variant lqtpos brspos vartot
## [1,] "A286V" "0"    "0"    "2"
## [2,] "A586T" "0"    "1"    "5"
## [3,] "A606T" "0"    "0"    "1"
## [4,] "A647D" "0"    "1"    "1"
## [5,] "A647S" "0"    "0"    "5"
## [6,] "A647V" "0"    "0"    "3"
```

Plot probability distribution

```
scaled<-length(x)/(length(d$var))
scaleb<-length(x)/(length(b$var))
scalev<-length(x)/(length(v$var))

#plot all variants classified as benign
plot(ben[,1],ben[,2]*scaleb,ylim = c(0,2.5),type = "l", ylab = "Probability integrated over all Benign V

#plot all variants classified as VUS
lines(vus[,1],vus[,2]*scalev,ylim = c(0,2.5),type = "l", ylab = "Probability integrated over all VUS Va

#plot all variants classified as pathogenic
lines(del[,1],del[,2]*scaled,ylim = c(0,2.5),type = "l", ylab = "Probability integrated over all Pathog
```



Calculate the probability the "average/typical" variant has a posterior mean penetrance fraction of carriers presenting with either Brs1 or LQT3 $> 20\%$

```
# calculate integrated probability of penetrance > 20% (1:5).
delh<-0
for (j in 1:(length(x)-length(x)/5)){
  delh <- delh+del[j+length(x)/5,2]*.01*scaled
}
vush<-0
for (j in 1:(length(x)-length(x)/5)){
  vush <- vush+vus[j+length(x)/5,2]*.01*scalev
}
benh<-0
for (j in 1:(length(x)-length(x)/5)){
  benh <- benh+ben[j+length(x)/5,2]*.01*scaleb
}

print("probability a variant classified as (likely) pathogenic has a penetrance > 20%")

## [1] "probability a variant classified as (likely) pathogenic has a penetrance > 20%"

delh

## [1] 0.9082202

print("probability a variant classified as VUS has a penetrance > 20%")
```

```
## [1] "probability a variant classified as VUS has a penetrance > 20%"
```

```
vush
```

```
## [1] 0.3175086
```

```
print("probability a variant classified as (likely) benign has a penetrance > 20%")
```

```
## [1] "probability a variant classified as (likely) benign has a penetrance > 20%"
```

```
benh
```

```
## [1] 0.04879566
```

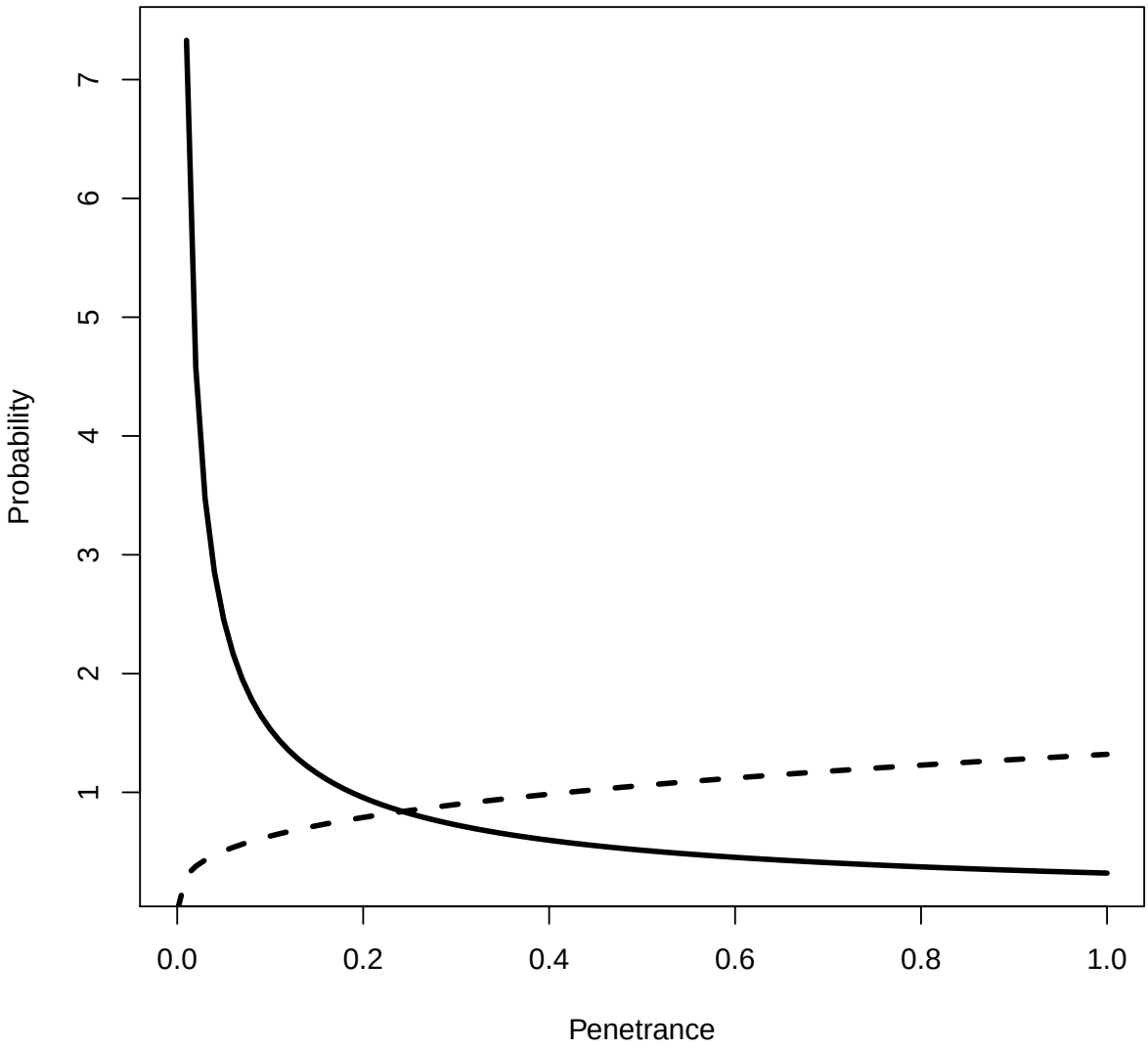
```
#run the following in console:\n
#pdf("~/../Dropbox/Andrew-Brett/scn5a_annotation/paper/images/class_<number>.pdf")\n
#<all the desired plots written below>\n
#dev.off()\n
```

Sensitivity Analysis

Solid lines are the prior, dashed lines are posterior if one affected carrier is observed

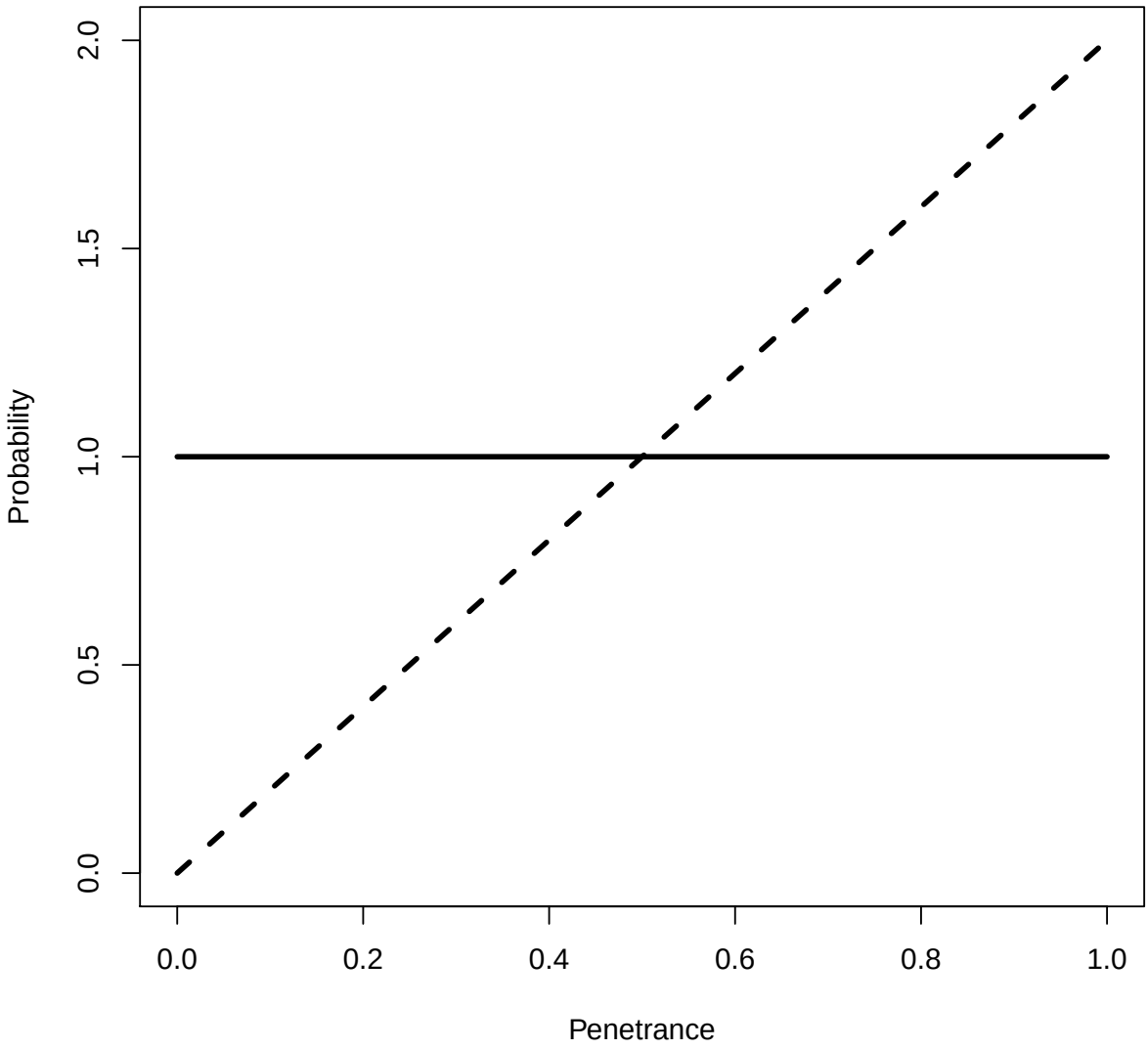
Empirical Bayes (used in manuscript)

```
# Penetrance calculation used (Manuscript Version)
abrs0=0.32
alqt0=0.11
beta0=1
curve(dbeta(x, abrs0, beta0), ylab="Probability", xlab="Penetrance", lwd = 3)
curve(dbeta(x, abrs0+1, beta0), ylab="Probability", xlab="Penetrance", lwd = 3, lty = 2, add = TRUE)
```



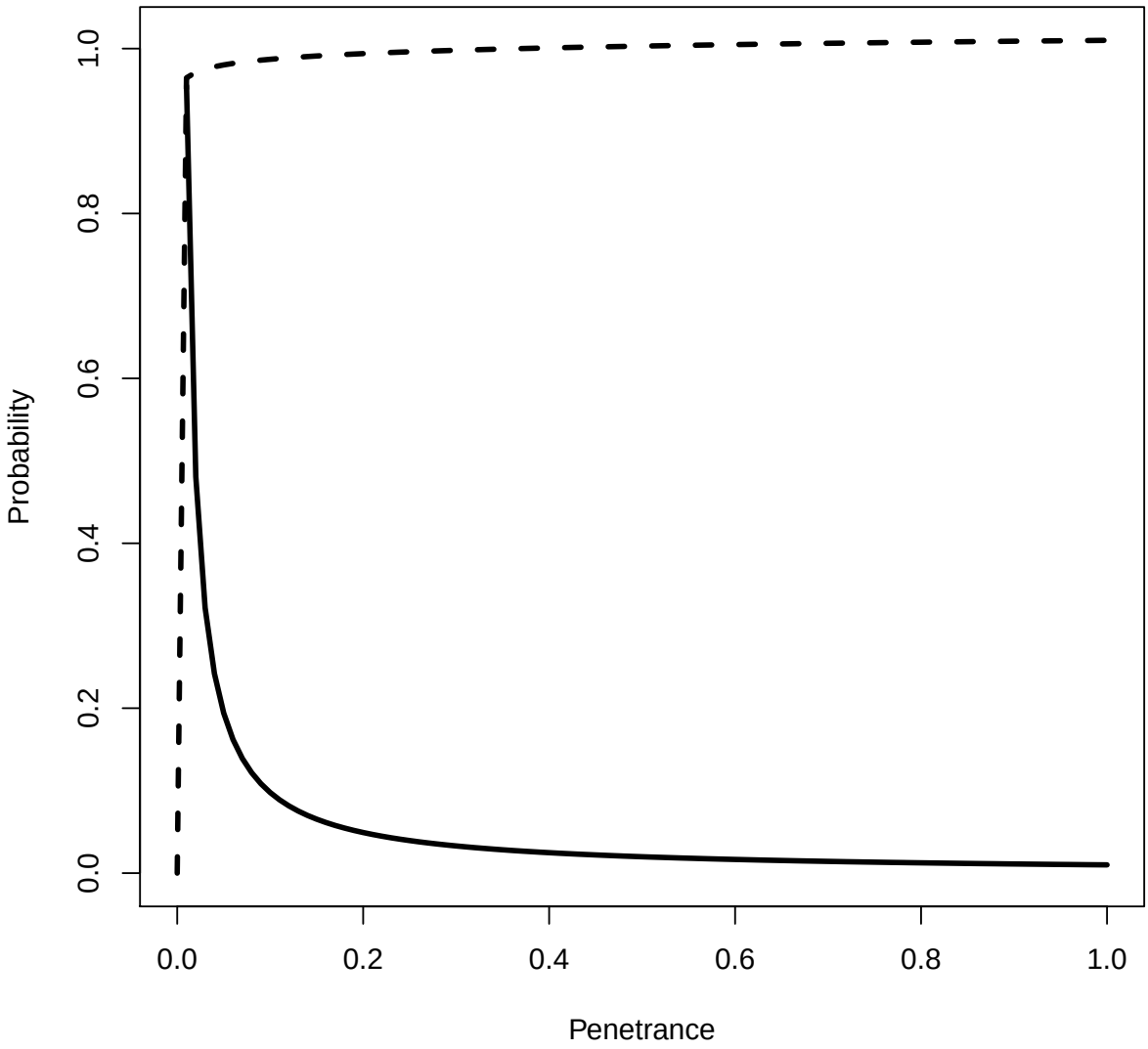
Uninformative Prior

```
# Changing penetrance calculation to uninformative prior
abrs0=1
alqt0=1
beta0=1
curve(dbeta(x, abrs0+1, beta0), ylab="Probability", xlab="Penetrance", lwd = 3, lty = 2)
curve(dbeta(x, abrs0, beta0), ylab="Probability", xlab="Penetrance", lwd = 3, add = TRUE)
```



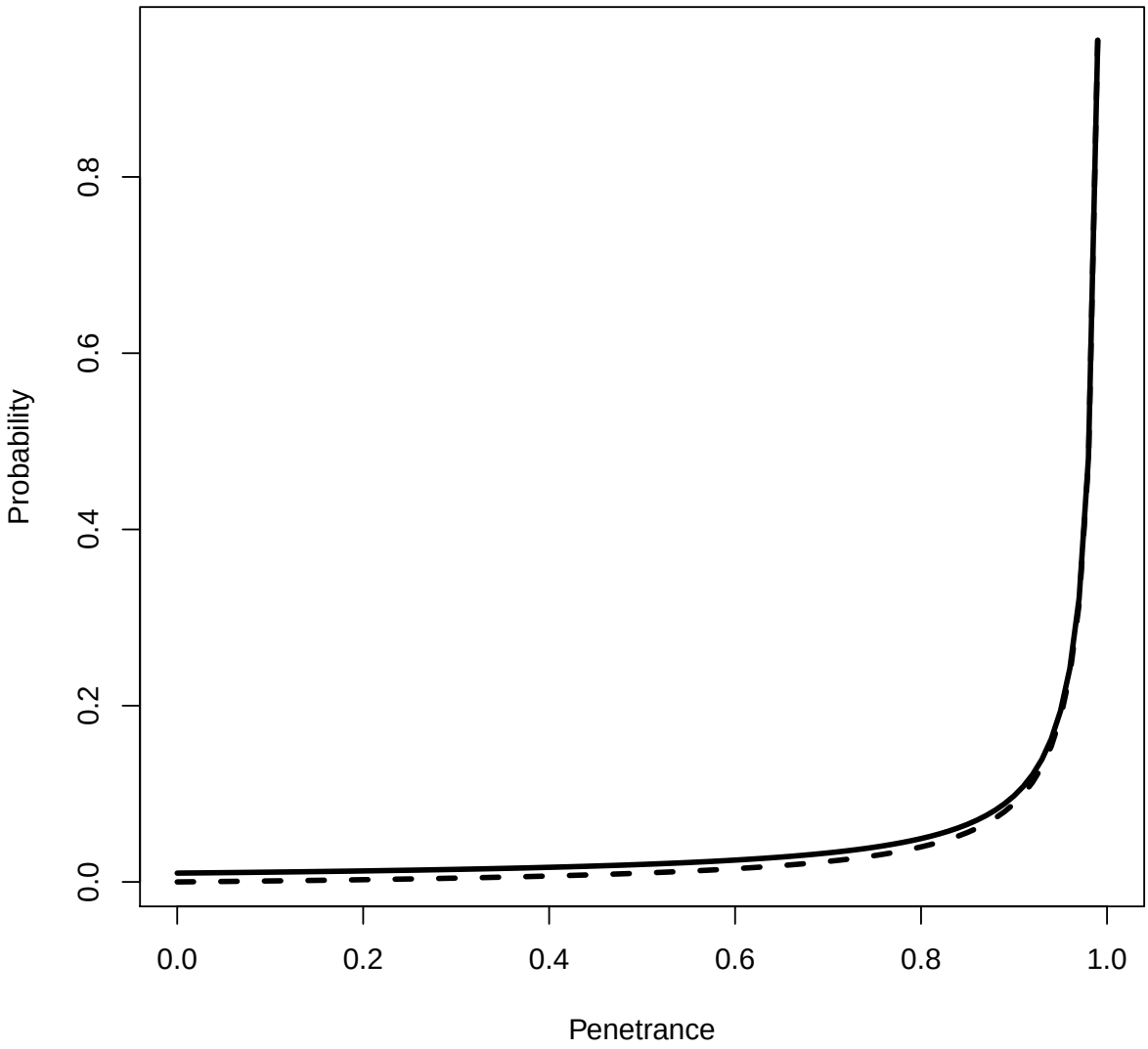
Optimistic Prior

```
# Changing penetrance calculation to optimistic (no affected carriers)
abrs0=0.01
alqt0=0.01
beta0=1
curve(dbeta(x, abrs0+1, beta0), ylab="Probability", xlab="Penetrance", lwd = 3, lty = 2)
curve(dbeta(x, abrs0, beta0), ylab="Probability", xlab="Penetrance", lwd = 3, add = TRUE)
```



Pessimistic Prior

```
# Changing penetrance calculation to pessimistic (one affected carrier)
abrs0=1
alqt0=1
beta0=0.01
curve(dbeta(x, abrs0, beta0), ylab="Probability", xlab="Penetrance", lwd = 3)
curve(dbeta(x, abrs0+1, beta0), ylab="Probability", xlab="Penetrance", lwd = 3, lty = 2, add = TRUE)
```


Probability
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
1.0
Penetrance