Predict LQTS Diagnosis Probability Using Structure, Function, and *In Silico* Features

Laura Bertolami, Shoshana Kelner, and Brett Kroncke

Summer 2021

Table of Contents

# Introduction

# Part 1: Calculate probability of LQT2 diagnosis and LQT2 Probability Density using Various Subsets of the Literature and Cohort Data

## Literature and Cohort Combined (for final predictions)

This only needs to be run if you are not loading an RData that is already clean.

# d <- d[d$mut\_type == "missense" & d$isoform=="A",]  
# d[is.na(d$total\_carriers),"total\_carriers"] <- 0  
# d[is.na(d$lqt2),"lqt2"] <- 0  
# # set initial weighting and penetrance  
# d$weight = 1-1/(0.01+d$total\_carriers)  
# d$penetrance\_lqt2 <- d$lqt2/d$total\_carriers  
# d[d$total\_carriers < 1,"weight"] <- 0.000 # This is changed to "< 2" here to evaluate ROC-AUC of n=1 variants from the literature

### LQT2 empirical diagnosis probability prior

Use observed LQT2 diagnosis probability to calculate “LQTS probability density” as described in previous publication. Plot diagnosis probability density versus residue

This only needs to be run if you are not loading an RData that is already clean.

# Mean squared error  
  
# mse <- function(sm) {  
# mean((sm$residuals)^2\*(sm$weights))  
# }  
# # Derive alpha and beta from weighted mean and MSE (estimated variance)  
# estBetaParams <- function(mu, var) {  
# alpha <- ((1 - mu) / var - 1 / mu) \* mu ^ 2  
# beta <- alpha \* (1 / mu - 1)  
# return(params = list(alpha = alpha, beta = beta))  
# }  
# # Weighted mean to determine LQT2 penetrance empirical prior  
# newdata = data.frame(wt=1)  
# model <- lm(penetrance\_lqt2 ~ 1, data=d, weights = d$weight)  
#   
# summary(model)  
# p<-predict(model, newdata)  
# dev<- mse(model) #p\*(1-p)  
# # Estimated shape parameters for LQT2 empirical prior  
# alpha0 = estBetaParams(p,dev)$alpha  
# beta0 = estBetaParams(p,dev)$beta  
# print(paste("alpha0 = ", alpha0, " beta0 = ", beta0))  
# # Bayesian LQT2 penetrance estimates from empirical priors  
# # and observed affected/unaffected counts:  
# d$lqt2\_penetranceBayesian\_initial <- (alpha0 + d[,"lqt2"])/((alpha0 + beta0 + d[,"total\_carriers"]))  
# d$lqt2\_penetranceBayesian<-d$lqt2\_penetranceBayesian\_initial  
#   
# combined.data<- d  
# #All data is clean in Updated.Data, because the cleaned data was saved as such  
# # this chunk is redundant for us but necessary if new data need to run through chunk 13

### Calculate LQTS probability densities and annotate function and structural location

With the updated empirical priors applied to carrier counts, calculate “LQTS probability density” as described in previous publication. !!! NOTE: since these data are truly the “best estimates” we include all variants in the calculation such that unique scores are by residue not by variant.

# Part 3: Variance explained

## Pearson R\^2 and Spearman Rho Against EM Posterior from Cohort  
  
  
#1   
  
calcPval=function(xName,yName,weightName,nPerms,new.mat2){  
 # Pulls out variables  
 x=new.mat2[,xName]   
 y=new.mat2[,yName]   
 w=new.mat2[,weightName]  
 x2=x[!is.na(x)]  
 y2=y[!is.na(x)]  
 w2=w[!is.na(x)]  
 # Calculate the real correlation  
 realCorr=weightedCorr(x2,y2,method='spearman',weights=w2)  
 # Do permutations, calculate fake correlations  
 permutedCorrList=c()  
 for(permNum in 1:nPerms){  
 permutedX=sample(x2,length(x2),replace=FALSE)  
 wCorrSim=weightedCorr(permutedX,y2,method='spearman',weights=w2)  
 permutedCorrList=c(permutedCorrList,wCorrSim)  
 }  
 permutedCorrList2=abs(permutedCorrList)  
 realCorr2=abs(realCorr)  
   
 # Calculate pvalue  
 summ=sum(realCorr2<permutedCorrList2)  
 pValue=summ/nPerms  
 return(list(realCorr,pValue,length(x2)))  
}  
  
  
#2  
  
  
calcAllPvals=function(yList,xList,nPerms,weightName,new.mat2){  
 i=0  
 resultTable=data.frame()  
 for(yName in yList){  
 for(xName in xList){  
 i=i+1  
 result=calcPval(xName,yName,weightName, nPerms, new.mat2)  
 resultTable[i,'x']=xName  
 resultTable[i,'y']=yName  
 resultTable[i,'nPerms']=nPerms  
 resultTable[i,'weightedCorr']=result[[1]]  
 resultTable[i,'pValue']=result[[2]]  
 resultTable[i,'n']=result[[3]]  
 #print(resultTable[i,'pValue'])  
 }  
 }  
 print(resultTable)  
 return(resultTable)  
}

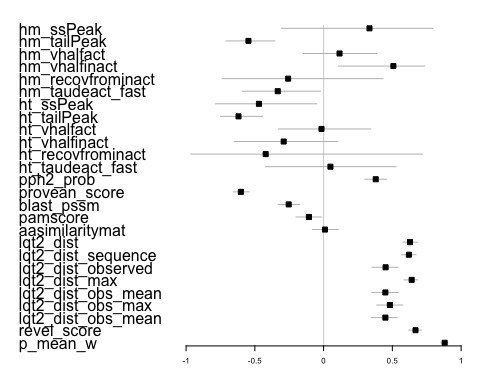
#Forest Plot Data Manipulation (add lqt2\_dist from all sets)

# Data Tables and Forest plot

load('all.lqt2.data.trunc.Rdata')  
herg.combined.data <- herg.combined.data[,!names(herg.combined.data) %in% "cardiacboost"]  
herg.combined.data <- unique(herg.combined.data)  
results.1.data<-herg.combined.data[!is.na(herg.combined.data$provean\_score) & !is.na(herg.combined.data$revel\_score) & !is.na(herg.combined.data$pamscore),]  
  
penetrance\_lqt2 <- na.omit(herg.combined.data$penetrance\_lqt2)  
  
yList=c("penetrance\_lqt2")  
  
xList=c('hm\_ssPeak','hm\_tailPeak','hm\_vhalfact','hm\_vhalfinact','hm\_recovfrominact', 'hm\_taudeact\_fast',   
 'ht\_ssPeak','ht\_tailPeak','ht\_vhalfact','ht\_vhalfinact','ht\_recovfrominact', 'ht\_taudeact\_fast',  
 'pph2\_prob', 'provean\_score', "blast\_pssm",  
 'pamscore', 'aasimilaritymat', "lqt2\_dist",'lqt2\_dist\_sequence', 'lqt2\_dist\_observed', 'lqt2\_dist\_max', 'lqt2\_dist\_obs\_mean','lqt2\_dist\_obs\_max', 'lqt2\_dist\_obs\_mean', "revel\_score", 'p\_mean\_w')  
  
weight <- c(herg.combined.data$weight)  
weight <- weight[weight!=0]  
  
resultTable<-calcAllPvals(yList, xList, 1000, 'weight', results.1.data[results.1.data$total\_carriers>0,])

## x y nPerms weightedCorr pValue n  
## 1 hm\_ssPeak penetrance\_lqt2 1000 0.333618489 0.191 22  
## 2 hm\_tailPeak penetrance\_lqt2 1000 -0.504475727 0.000 109  
## 3 hm\_vhalfact penetrance\_lqt2 1000 0.094556650 0.524 77  
## 4 hm\_vhalfinact penetrance\_lqt2 1000 0.626300479 0.004 31  
## 5 hm\_recovfrominact penetrance\_lqt2 1000 -0.355863754 0.316 12  
## 6 hm\_taudeact\_fast penetrance\_lqt2 1000 -0.330523484 0.040 48  
## 7 ht\_ssPeak penetrance\_lqt2 1000 -0.439847206 0.036 33  
## 8 ht\_tailPeak penetrance\_lqt2 1000 -0.616747843 0.000 91  
## 9 ht\_vhalfact penetrance\_lqt2 1000 -0.056238987 0.712 46  
## 10 ht\_vhalfinact penetrance\_lqt2 1000 -0.261679452 0.203 29  
## 11 ht\_recovfrominact penetrance\_lqt2 1000 -0.421438569 0.361 7  
## 12 ht\_taudeact\_fast penetrance\_lqt2 1000 0.049726258 0.860 21  
## 13 pph2\_prob penetrance\_lqt2 1000 0.377412096 0.000 850  
## 14 provean\_score penetrance\_lqt2 1000 -0.600999848 0.000 850  
## 15 blast\_pssm penetrance\_lqt2 1000 -0.252583636 0.000 850  
## 16 pamscore penetrance\_lqt2 1000 -0.111299475 0.012 850  
## 17 aasimilaritymat penetrance\_lqt2 1000 0.007978436 0.873 850  
## 18 lqt2\_dist penetrance\_lqt2 1000 0.633518856 0.000 850  
## 19 lqt2\_dist\_sequence penetrance\_lqt2 1000 0.624553517 0.000 850  
## 20 lqt2\_dist\_observed penetrance\_lqt2 1000 0.455666339 0.000 494  
## 21 lqt2\_dist\_max penetrance\_lqt2 1000 0.647820173 0.000 850  
## 22 lqt2\_dist\_obs\_mean penetrance\_lqt2 1000 0.455365491 0.000 494  
## 23 lqt2\_dist\_obs\_max penetrance\_lqt2 1000 0.491061828 0.000 494  
## 24 lqt2\_dist\_obs\_mean penetrance\_lqt2 1000 0.455365491 0.000 494  
## 25 revel\_score penetrance\_lqt2 1000 0.669183351 0.000 850  
## 26 p\_mean\_w penetrance\_lqt2 1000 0.891596614 0.000 850

rm(results.1.data)  
  
#4  
  
i=0  
FP.data<-data.frame()  
for (x in xList){  
 i=i+2  
 FP.data[i,"Feature"]<-x  
 t<-herg.combined.data[!is.na(herg.combined.data[,x]) & herg.combined.data$total\_carriers>0,]  
   
 #<-t[!is.na(t[,"var"]),]  
 #FP.data[i,"Feature"]<-paste(x,"\_cohort")  
 #t<-herg.combined.data[!is.na(herg.combined.data[,x]) & herg.combined.data$total\_carriers>0,]  
   
 t<-t[!is.na(t[,"var"]),]  
 foo <- boot(t, function(data,indices)  
 weightedCorr(t[indices,x],t$penetrance\_lqt2[indices], method="Spearman", weights = t$weight[indices]), R=1000)  
   
   
 FP.data[i,"Spearman"]<-foo$t0  
 FP.data[i,"Spearman\_low"]<-quantile(foo$t,c(0.025,0.975), na.rm = T)[1][[1]]  
 FP.data[i,"Spearman\_high"]<-quantile(foo$t,c(0.025,0.975), na.rm = T)[2][[1]]  
 FP.data[i,"n"]<-length(t[,x])   
}  
  
#5   
#trellis.device(device = "quartz", height = 25, width = 40, color = TRUE)  
#library(lattice)  
#trellis.device(device = "RStudioGD", height = 25, width = 40, color = TRUE)  
  
forestplot(FP.data$Feature,FP.data$Spearman,FP.data$Spearman\_low,FP.data$Spearman\_high, cex = 0.5, boxsize = 0.8)



#added additional lqt2 columns but now the plot is too large to display. Created a plot with only these columns later.   
  
rm(FP.data)  
  
tmp<-herg.combined.data[!is.na(herg.combined.data$provean\_score) & !is.na(herg.combined.data$revel\_score) & !is.na(herg.combined.data$pamscore),]  
  
  
# Weighted R2 between observed LQT2 penetrance and post-test probability  
foo <- boot(tmp, function(data,indices)  
 weightedCorr(tmp$p\_mean\_w[indices],tmp$penetrance\_lqt2[indices], method="pearson", weights = tmp$weight[indices])^2, R=1000)  
print("EM estimated LQT2 diagnosis probability versus observed cohort LQT2 diagnosis probability")

## [1] "EM estimated LQT2 diagnosis probability versus observed cohort LQT2 diagnosis probability"

foo$t0

## [1] 0.9104879

quantile(foo$t,c(0.025,0.975), na.rm = T)

## 2.5% 97.5%   
## 0.8902618 0.9290539

lqt2\_dist <- na.omit(herg.combined.data$lqt2\_dist)  
  
model<- lm(penetrance\_lqt2~lqt2\_dist, data = tmp, weights = weight)  
quantile(foo$t,c(0.025,0.975), na.rm = TRUE) #added na.rm =TRUE

## 2.5% 97.5%   
## 0.8902618 0.9290539

model <- lm(penetrance\_lqt2~lqt2\_dist, data = tmp, weights = weight)  
mod<-data.frame(model$fitted.values,model$model$penetrance\_lqt2,model$weights)  
foo <- boot(mod, function(data,indices)  
 weightedCorr(mod$model.fitted.values[indices],mod$model.model.penetrance\_lqt2[indices], method="pearson", weights = mod$model.weights[indices])^2, R=1000)  
print("LQT2 diagnosis probability density versus observed cohort LQT2 diagnosis probability")

## [1] "LQT2 diagnosis probability density versus observed cohort LQT2 diagnosis probability"

foo$t0

## [1] 0.4804367

quantile(foo$t,c(0.025,0.975))

## 2.5% 97.5%   
## 0.4054793 0.5544454

model <- lm(penetrance\_lqt2~revel\_score, data = tmp, weights = weight)   
mod<-data.frame(model$fitted.values,model$model$penetrance\_lqt2,model$weights)  
foo <- boot(mod, function(data,indices)  
 weightedCorr(mod$model.fitted.values[indices],mod$model.model.penetrance\_lqt2[indices], method="pearson", weights = mod$model.weights[indices])^2, R=1000)  
print("REVEL score versus observed cohort LQT2 diagnosis probability")

## [1] "REVEL score versus observed cohort LQT2 diagnosis probability"

foo$t0

## [1] 0.3968236

quantile(foo$t,c(0.025,0.975))

## 2.5% 97.5%   
## 0.3375976 0.4535098

rm(tmp)  
  
#6  
  
# Evaluate only variants with Heterozygous peak tail current measured.  
  
results.2.data<-herg.combined.data[!is.na(herg.combined.data$ht\_tailPeak),]  
  
resultTable<-calcAllPvals(yList, xList, 1000, 'weight', results.2.data[results.2.data$total\_carriers>0,])

## x y nPerms weightedCorr pValue n  
## 1 hm\_ssPeak penetrance\_lqt2 1000 0.42810160 0.234 12  
## 2 hm\_tailPeak penetrance\_lqt2 1000 -0.46813545 0.000 73  
## 3 hm\_vhalfact penetrance\_lqt2 1000 -0.10806278 0.570 37  
## 4 hm\_vhalfinact penetrance\_lqt2 1000 0.45412143 0.142 18  
## 5 hm\_recovfrominact penetrance\_lqt2 1000 -0.03342009 0.919 4  
## 6 hm\_taudeact\_fast penetrance\_lqt2 1000 -0.15720614 0.555 17  
## 7 ht\_ssPeak penetrance\_lqt2 1000 -0.47026995 0.019 34  
## 8 ht\_tailPeak penetrance\_lqt2 1000 -0.61893165 0.000 94  
## 9 ht\_vhalfact penetrance\_lqt2 1000 0.03143114 0.863 45  
## 10 ht\_vhalfinact penetrance\_lqt2 1000 -0.29075675 0.150 29  
## 11 ht\_recovfrominact penetrance\_lqt2 1000 -0.42143857 0.381 7  
## 12 ht\_taudeact\_fast penetrance\_lqt2 1000 0.07950543 0.736 19  
## 13 pph2\_prob penetrance\_lqt2 1000 0.34631206 0.007 91  
## 14 provean\_score penetrance\_lqt2 1000 -0.45174753 0.000 91  
## 15 blast\_pssm penetrance\_lqt2 1000 0.07507438 0.528 91  
## 16 pamscore penetrance\_lqt2 1000 0.07224788 0.552 93  
## 17 aasimilaritymat penetrance\_lqt2 1000 0.05287112 0.680 93  
## 18 lqt2\_dist penetrance\_lqt2 1000 0.68842117 0.000 94  
## 19 lqt2\_dist\_sequence penetrance\_lqt2 1000 0.66167723 0.000 94  
## 20 lqt2\_dist\_observed penetrance\_lqt2 1000 0.48102653 0.000 72  
## 21 lqt2\_dist\_max penetrance\_lqt2 1000 0.73400917 0.000 94  
## 22 lqt2\_dist\_obs\_mean penetrance\_lqt2 1000 0.50647705 0.000 72  
## 23 lqt2\_dist\_obs\_max penetrance\_lqt2 1000 0.56209253 0.000 72  
## 24 lqt2\_dist\_obs\_mean penetrance\_lqt2 1000 0.50647705 0.000 72  
## 25 revel\_score penetrance\_lqt2 1000 0.62793302 0.000 93  
## 26 p\_mean\_w penetrance\_lqt2 1000 0.92600393 0.000 94

foo <- boot(results.2.data, function(data,indices)  
 weightedCorr(results.2.data$p\_mean\_w[indices],results.2.data$penetrance\_lqt2[indices], method="pearson", weights = results.2.data$weight[indices])^2, R=1000)  
print("EM estimated LQT2 diagnosis probability versus observed cohort LQT2 diagnosis probability")

## [1] "EM estimated LQT2 diagnosis probability versus observed cohort LQT2 diagnosis probability"

foo$t0

## [1] 0.9403061

quantile(foo$t,c(0.025,0.975), na.rm = T)

## 2.5% 97.5%   
## 0.9018079 0.9686441

model <- lm(penetrance\_lqt2~ht\_tailPeak, data = results.2.data, weights = weight)  
mod<-data.frame(model$fitted.values,model$model$penetrance\_lqt2,model$weights)  
foo <- boot(mod, function(data,indices)  
 weightedCorr(mod$model.fitted.values[indices],mod$model.model.penetrance\_lqt2[indices], method="pearson", weights = mod$model.weights[indices])^2, R=1000)  
print("Heterozygously measured peak tail current versus observed cohort LQT2 diagnosis probability")

## [1] "Heterozygously measured peak tail current versus observed cohort LQT2 diagnosis probability"

foo$t0

## [1] 0.3470365

quantile(foo$t,c(0.025,0.975))

## 2.5% 97.5%   
## 0.1909125 0.5329507

model<-lm(penetrance\_lqt2~lqt2\_dist, data = results.2.data, weights = weight)  
mod<-data.frame(model$fitted.values,model$model$penetrance\_lqt2,model$weights)  
foo <- boot(mod, function(data,indices)  
 weightedCorr(mod$model.fitted.values[indices],mod$model.model.penetrance\_lqt2[indices], method="pearson", weights = mod$model.weights[indices])^2, R=1000)  
print("LQT2 probability density versus observed cohort LQT2 diagnosis probability")

## [1] "LQT2 probability density versus observed cohort LQT2 diagnosis probability"

foo$t0

## [1] 0.5205787

quantile(foo$t,c(0.025,0.975))

## 2.5% 97.5%   
## 0.3315306 0.6935842

model <- lm(penetrance\_lqt2~revel\_score, data = results.2.data, weights = weight)  
mod<-data.frame(model$fitted.values,model$model$penetrance\_lqt2,model$weights)  
foo <- boot(mod, function(data,indices)  
 weightedCorr(mod$model.fitted.values[indices],mod$model.model.penetrance\_lqt2[indices], method="pearson", weights = mod$model.weights[indices])^2, R=1000)  
print("REVEL score versus observed cohort LQT2 diagnosis probability")

## [1] "REVEL score versus observed cohort LQT2 diagnosis probability"

foo$t0

## [1] 0.454215

quantile(foo$t,c(0.025,0.975))

## 2.5% 97.5%   
## 0.2881039 0.6212077

rm(results.2.data)

## Variance explained from literature dataset

tmp2<-herg.combined.data[!is.na(herg.combined.data$provean\_score) & !is.na(herg.combined.data$penetrance\_lqt2) & !is.na(herg.combined.data$lqt2\_dist) & !is.na(herg.combined.data$revel\_score),]  
foo <- boot(tmp2, function(data,indices)  
 weightedCorr(tmp2$p\_mean\_w[indices],tmp2$penetrance\_lqt2[indices], method="pearson", weights = tmp2$weight[indices])^2, R=1000)  
print("EM estimated LQT2 diagnosis probability versus observed literature LQT2 diagnosis probability")

## [1] "EM estimated LQT2 diagnosis probability versus observed literature LQT2 diagnosis probability"

foo$t0

## [1] 0.9088678

quantile(foo$t,c(0.025,0.975))

## 2.5% 97.5%   
## 0.8903271 0.9265508

model <- lm(penetrance\_lqt2~lqt2\_dist, data = tmp2, weights = weight)  
mod<-data.frame(model$fitted.values,model$model$penetrance\_lqt2,model$weights)  
foo <- boot(mod, function(data,indices)  
 weightedCorr(mod$model.fitted.values[indices],mod$model.model.penetrance\_lqt2[indices], method="pearson", weights = mod$model.weights[indices])^2, R=1000)  
print("LQT2 probability density versus observed literature LQT2 diagnosis probability")

## [1] "LQT2 probability density versus observed literature LQT2 diagnosis probability"

foo$t0

## [1] 0.4787778

quantile(foo$t,c(0.025,0.975))

## 2.5% 97.5%   
## 0.4057900 0.5559474

model <- lm(penetrance\_lqt2~revel\_score, data = tmp2, weights = weight)  
mod<-data.frame(model$fitted.values,model$model$penetrance\_lqt2,model$weights)  
foo <- boot(mod, function(data,indices)  
 weightedCorr(mod$model.fitted.values[indices],mod$model.model.penetrance\_lqt2[indices], method="pearson", weights = mod$model.weights[indices])^2, R=1000)  
print("REVEL versus observed literature LQT2 diagnosis probability")

## [1] "REVEL versus observed literature LQT2 diagnosis probability"

foo$t0

## [1] 0.3971701

quantile(foo$t,c(0.025,0.975))

## 2.5% 97.5%   
## 0.3345960 0.4565936

# Evaluate only variants with Heterozygous peak tail current measured.  
tmp3<-herg.combined.data[!is.na(herg.combined.data$ht\_tailPeak) & !is.na(herg.combined.data$provean\_score) & !is.na(herg.combined.data$penetrance\_lqt2) & !is.nan(herg.combined.data$penetrance\_lqt2) & !is.na(herg.combined.data$lqt2\_dist),]  
foo <- boot(tmp3, function(data,indices)  
 weightedCorr(tmp3$p\_mean\_w[indices],tmp3$penetrance\_lqt2[indices], method="pearson", weights = tmp3$weight[indices])^2, R=1000)  
print("EM estimated LQT2 diagnosis probability versus observed literature LQT2 diagnosis probability")

## [1] "EM estimated LQT2 diagnosis probability versus observed literature LQT2 diagnosis probability"

foo$t0

## [1] 0.9360721

quantile(foo$t,c(0.025,0.975))

## 2.5% 97.5%   
## 0.8883768 0.9666846

lqt2\_dist <-   
 model <- lm(penetrance\_lqt2~lqt2\_dist, data = tmp3, weights = weight)  
mod<-data.frame(model$fitted.values,model$model$penetrance\_lqt2,model$weights)  
foo <- boot(mod, function(data,indices)  
 weightedCorr(mod$model.fitted.values[indices],mod$model.model.penetrance\_lqt2[indices], method="pearson", weights = mod$model.weights[indices])^2, R=1000)  
print("LQT2 probability density versus observed literature LQT2 diagnosis probability")

## [1] "LQT2 probability density versus observed literature LQT2 diagnosis probability"

foo$t0

## [1] 0.4860049

quantile(foo$t,c(0.025,0.975))

## 2.5% 97.5%   
## 0.2862489 0.6788551

model <- lm(penetrance\_lqt2~ht\_tailPeak, data = tmp3, weights = weight)  
mod<-data.frame(model$fitted.values,model$model$penetrance\_lqt2,model$weights)  
foo <- boot(mod, function(data,indices)  
 weightedCorr(mod$model.fitted.values[indices],mod$model.model.penetrance\_lqt2[indices], method="pearson", weights = mod$model.weights[indices])^2, R=1000)  
print("Heterozygous peak tail current versus observed literature LQT2 diagnosis probability")

## [1] "Heterozygous peak tail current versus observed literature LQT2 diagnosis probability"

foo$t0

## [1] 0.354972

quantile(foo$t,c(0.025,0.975))

## 2.5% 97.5%   
## 0.1881273 0.5632725

model <- lm(penetrance\_lqt2~revel\_score, data = tmp3, weights = weight)  
mod<-data.frame(model$fitted.values,model$model$penetrance\_lqt2,model$weights)  
foo <- boot(mod, function(data,indices)  
 weightedCorr(mod$model.fitted.values[indices],mod$model.model.penetrance\_lqt2[indices], method="pearson", weights = mod$model.weights[indices])^2, R=1000)  
print("REVEL versus observed literature LQT2 diagnosis probability")

## [1] "REVEL versus observed literature LQT2 diagnosis probability"

foo$t0

## [1] 0.432538

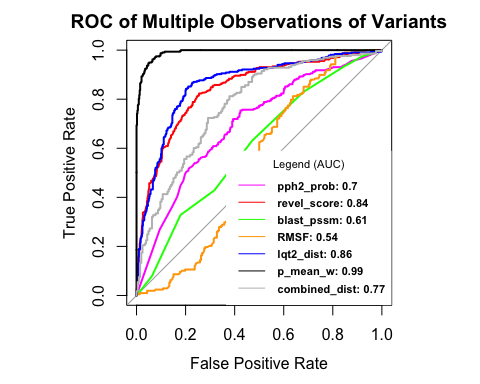
quantile(foo$t,c(0.025,0.975))

## 2.5% 97.5%   
## 0.2680968 0.6080763

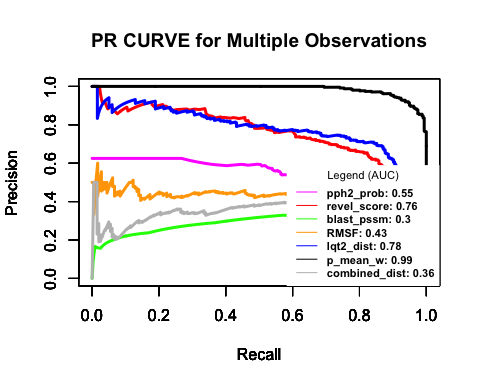
#Plot with only lqt2\_dist data (would not fit on large plot)

# Part 4: ROC and AUC plots

## ROC’s of observed cohort LQT2 diagnosis probability with 0.5 as cutoff all variants.

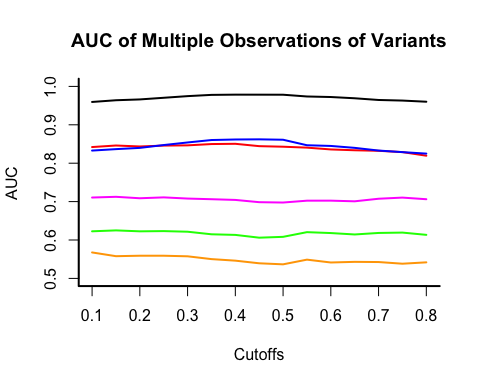


###Precision-Recall Curve



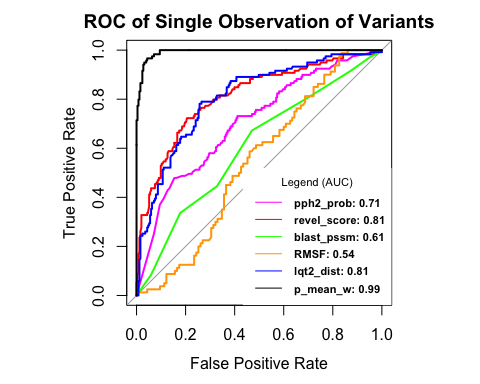
#### AUC’s from ROC’s observed literature LQT2 diagnosis probability at multiple cutoffs

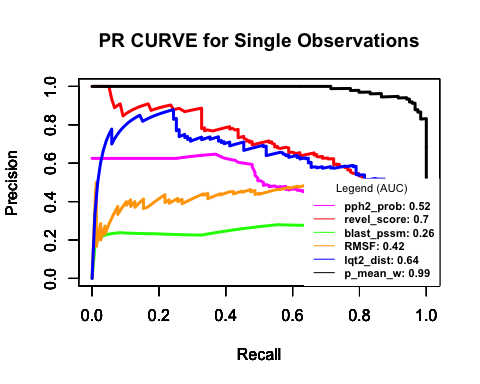
## [1] "858 369"  
## [1] "858 360"  
## [1] "858 355"  
## [1] "858 350"  
## [1] "858 342"  
## [1] "858 325"  
## [1] "858 324"  
## [1] "858 318"  
## [1] "858 317"  
## [1] "858 283"  
## [1] "858 280"  
## [1] "858 273"  
## [1] "858 253"  
## [1] "858 250"  
## [1] "858 237"



## ROC’s of observed cohort LQT2 diagnosis probability with single observation variants.

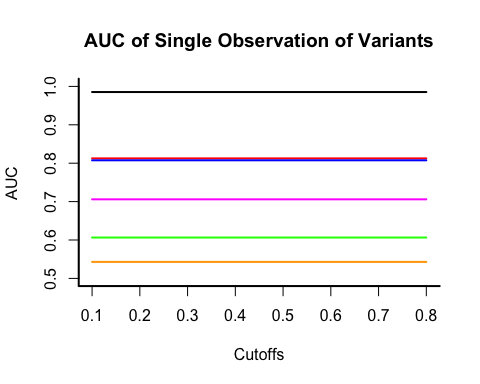
## Area under the curve: 0.8125  
## Area under the curve: 0.6064  
## Area under the curve: 0.543  
## Area under the curve: 0.8075  
## Area under the curve: 0.9931

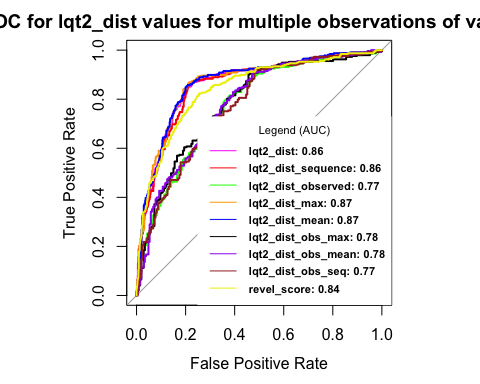


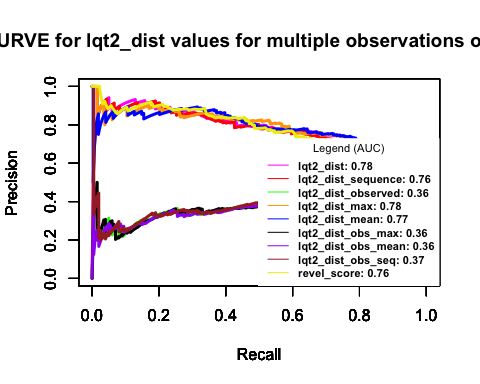
###PR Curve 

#### AUC for N = 1 variants from the literature.

## [1] "372 119"  
## [1] "372 119"  
## [1] "372 119"  
## [1] "372 119"  
## [1] "372 119"  
## [1] "372 119"  
## [1] "372 119"  
## [1] "372 119"  
## [1] "372 119"  
## [1] "372 119"  
## [1] "372 119"  
## [1] "372 119"  
## [1] "372 119"  
## [1] "372 119"  
## [1] "372 119"

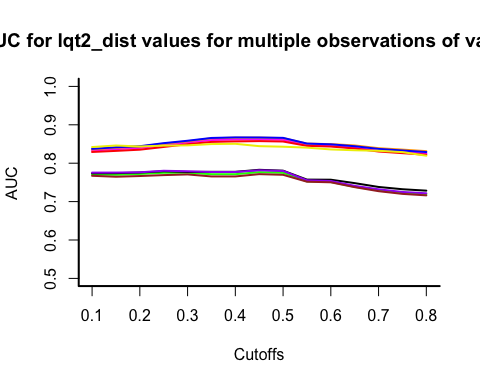


##ROC For All lqt2\_dist predictions  ###Precision-Recall Curve



####AUC For All lqt2\_dist predictions

## [1] "858 369"  
## [1] "858 360"  
## [1] "858 355"  
## [1] "858 350"  
## [1] "858 342"  
## [1] "858 325"  
## [1] "858 324"  
## [1] "858 318"  
## [1] "858 317"  
## [1] "858 283"  
## [1] "858 280"  
## [1] "858 273"  
## [1] "858 253"  
## [1] "858 250"  
## [1] "858 237"



#Code for func dist on penetrance\_lqt2 This has alredy been run and the output saved as an RData to be loaded elsewhere