

R Notebook

Functions

```
# function to plot ROC curve using ROCR library
plot_ROC <- function(true_outliers,
                     pvals,
                     outlier_cutoff = 1,
                     clr = TRUE,
                     ref_line = TRUE,
                     ...) {

  out_labels <- ifelse(rowSums(true_outliers == 0) >= outlier_cutoff, 0, 1)
  pred_val <- prediction(pvals, out_labels)
  perf_val <- performance(pred_val, "tpr", "fpr")
  plot(perf_val,
       colorize = clr,
       xlim=c(0,1),
       ylim=c(0,1),
       ...)

  if (ref_line == TRUE) {
    abline(a=0, b= 1)
  }

}
```

```
# function to put ROC curves on same plot
ROC_compare_plot <- function(true_outlier,
                             mamba_ppr,
                             prp,
                             outlier_cutoff = 1) {

  pvals <- list(mamba_ppr, prp)

  out_labels <- ifelse(rowSums(true_outlier == 0) >= outlier_cutoff, 0, 1)
  out_labels <- rep(list(out_labels), length(pvals))

  # ROCR can look at multiple ROC curves at same time if they are together as a list
  pred_val <- prediction(pvals, out_labels)
  perf_val <- performance(pred_val, "tpr", "fpr")

  plot(perf_val,
       col = as.list(1:length(pvals)),
```

```

      xlim=c(0,1),
      ylim=c(0,1))

  legend(x = "bottomright",
         legend = c("MAMBA PPR", "PRP"),
         fill = 1:length(pvals))
}

```

```

# AUC function
ROC_AUC <- function(true_outlier,
                    mamba_ppr,
                    prp,
                    outlier_cutoff = 1) {

  pvals <- list(mamba_ppr, prp)

  out_labels <- ifelse(rowSums(true_outlier == 0) >= outlier_cutoff, 0, 1)
  out_labels <- rep(list(out_labels), length(pvals))

  # ROCR can look at multiple ROC curves at same time if they are together as a list
  pred_val <- prediction(pvals, out_labels)
  auc_val <- performance(pred_val, measure = "auc")

  return(auc_val)
}

```

0.975 Nonoutlier Rate

Here, we look at the case where 0.975 is “the proportion of non-replicable SNPs which are well behaved, or nonoutliers” (quoted from the MAMBA documentation).

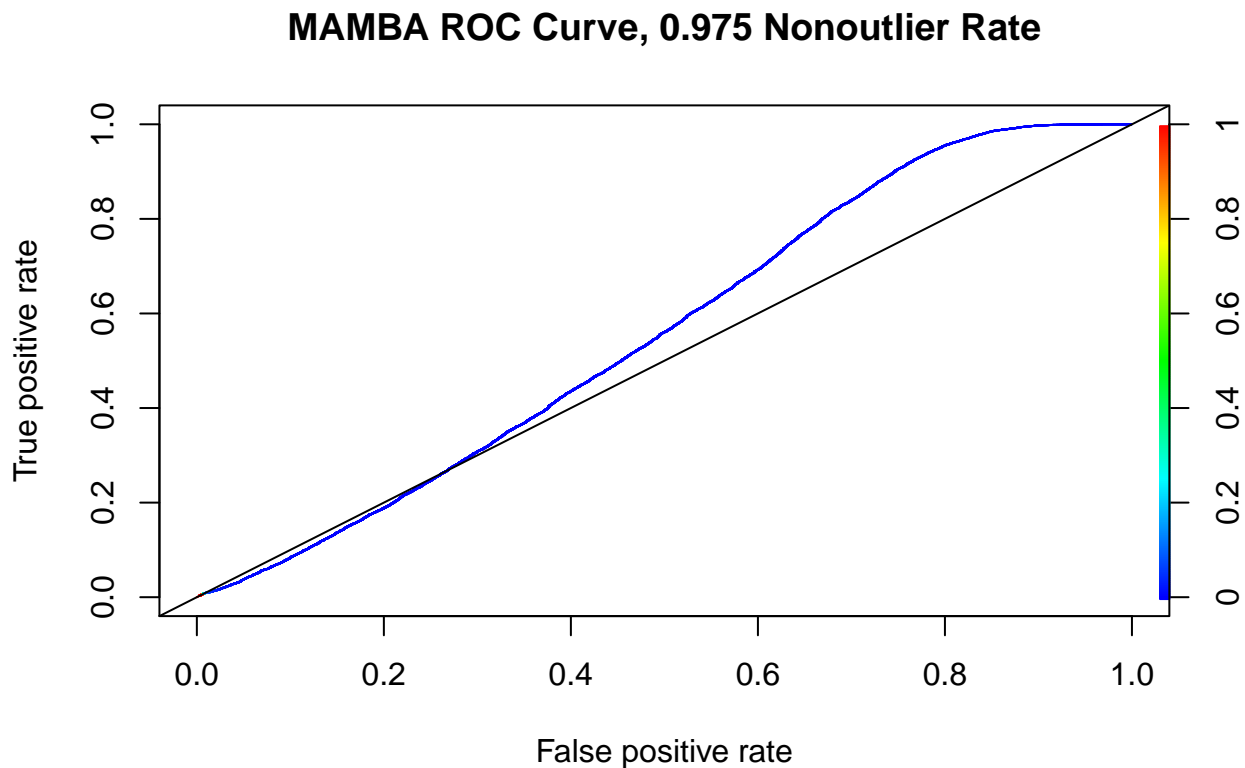
```
load(file = "data/prp_data/post_prp_data_pval_p975.rda")
load(file = "data/mamba_data/sim_mamba_mod_p975.rda")
load(file = "data/mamba_data/mamba_data_p975.rda")
```

```
prp_975 <- post_prp_data_pval
sim_mod_p975 <- sim_mod
mamba_data_p975 <- mamba_data
```

```
outliers_p975 <- mamba_data_p975$0jk
mamba_ppr_p975 <- sim_mod_p975$ppr
```

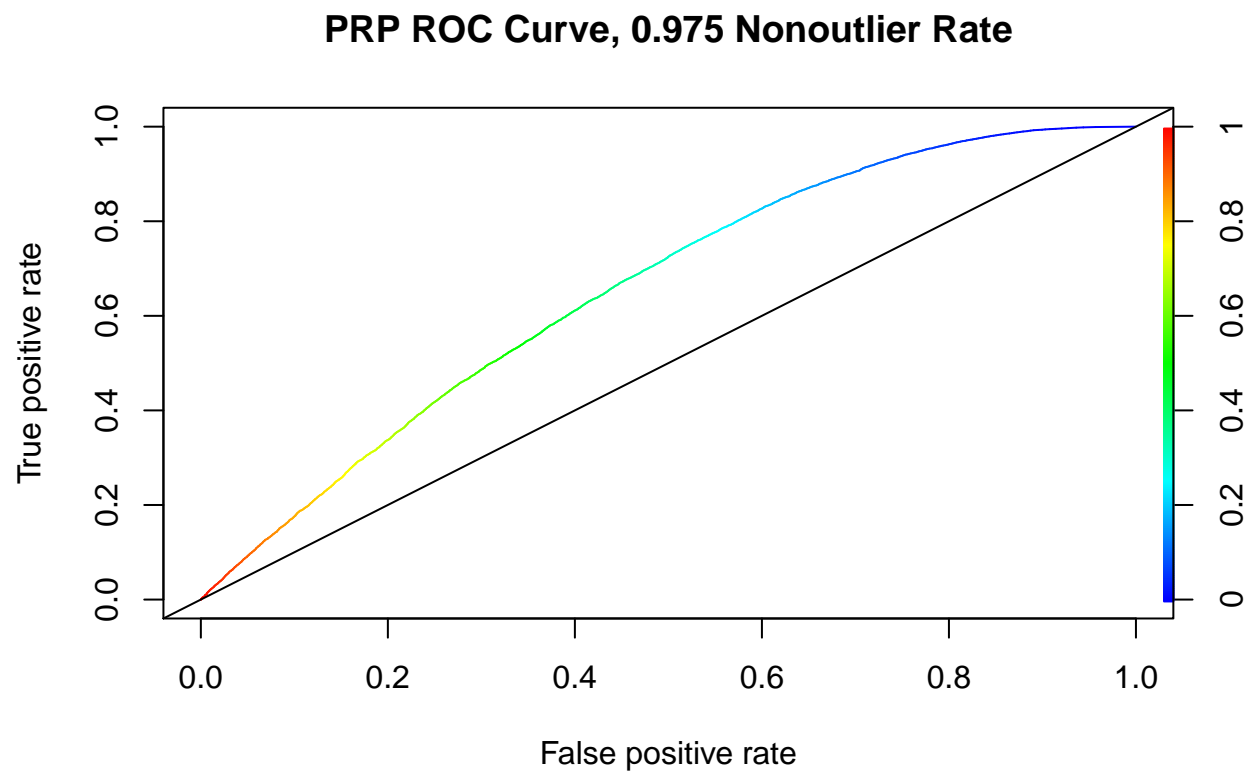
First we look at the ROC curve for the MAMBA values at a 0.975 nonoutlier study rate.

```
plot_ROC(outliers_p975,
          mamba_ppr_p975,
          main = "MAMBA ROC Curve, 0.975 Nonoutlier Rate")
```



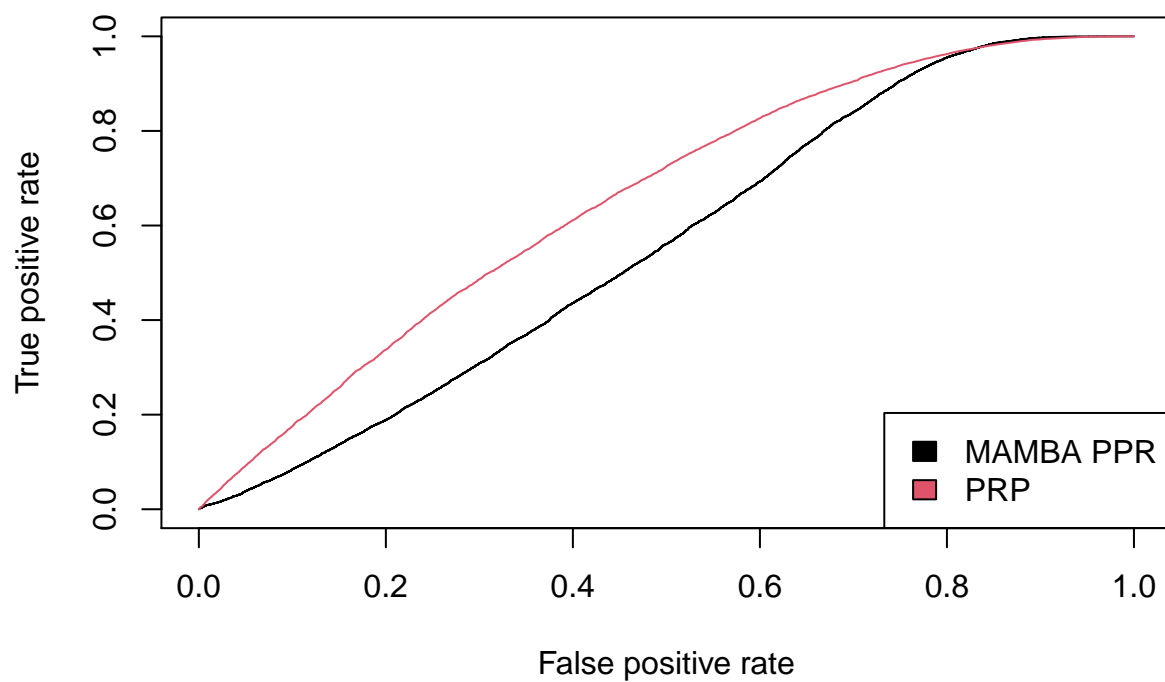
The PRP curve for the same nonoutlier rate is given below.

```
plot_ROC(outliers_p975,
         prp_975,
         main = "PRP ROC Curve, 0.975 Nonoutlier Rate")
```



Overlaying the plots, we can see that PRP works slightly better than MAMBA.

```
ROC_compare_plot(true_outlier = outliers_p975,
                 mamba_ppr = mamba_ppr_p975,
                 prp = prp_975)
```



```
AUC_p975 <- ROC_AUC(true_outlier = outliers_p975,  
                    mamba_ppr = mamba_ppr_p975,  
                    prp = prp_975)
```

The MAMBA AUC is 0.5571214, while the PRP AUC is 0.6542936.

0.9 Nonoutlier Rate

Here, we look at the case where 0.9 is “the proportion of non-replicable SNPs which are well behaved, or nonoutliers.”

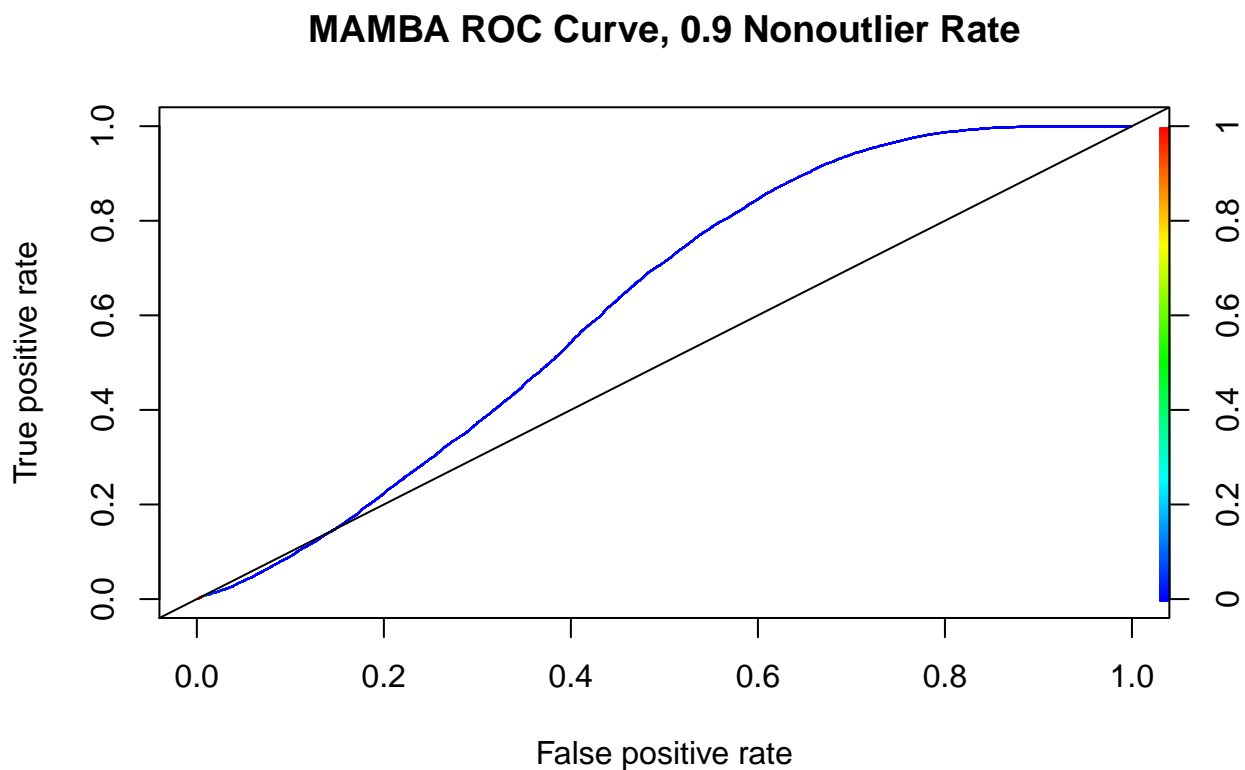
```
load(file = "data/prp_data/post_prp_data_pval_p90.rda")
load(file = "data/mamba_data/sim_mamba_mod_p90.rda")
load(file = "data/mamba_data/mamba_data_p90.rda")
```

```
prp_90 <- post_prp_data_pval
sim_mod_p90 <- sim_mod
mamba_data_p90 <- mamba_data

outliers_p90 <- mamba_data_p90$0jk
mamba_ppr_p90 <- sim_mod_p90$ppr
```

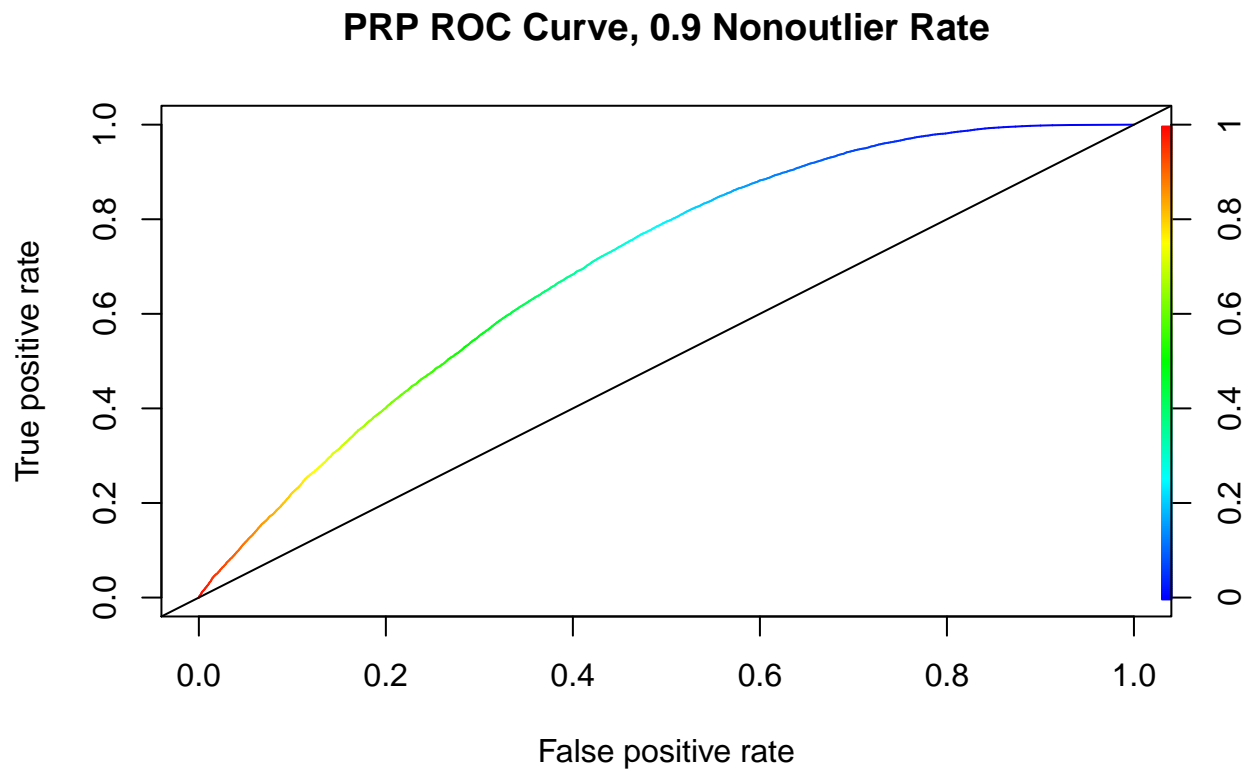
First we look at the ROC curve for the MAMBA values at a 0.9 nonoutlier study rate.

```
plot_ROC(outliers_p90,
          mamba_ppr_p90,
          main = "MAMBA ROC Curve, 0.9 Nonoutlier Rate")
```



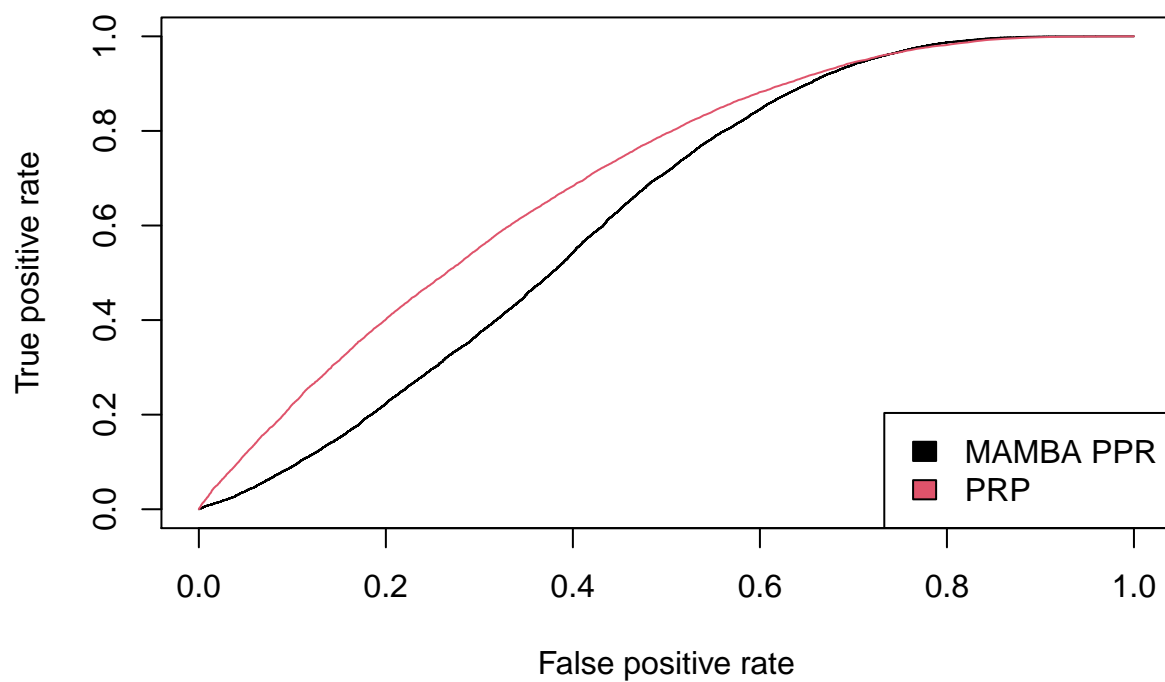
The PRP curve for the same nonoutlier rate is given below.

```
plot_ROC(outliers_p90,
        prp_90,
        main = "PRP ROC Curve, 0.9 Nonoutlier Rate")
```



Overlaying the plots, we can see that PRP still has a better curve than MAMBA when the rate of nonoutlier studies is 0.9.

```
ROC_compare_plot(true_outlier = outliers_p90,
                mamba_ppr = mamba_ppr_p90,
                prp = prp_90)
```



```
AUC_p90 <- ROC_AUC(true_outlier = outliers_p90,  
                    mamba_ppr = mamba_ppr_p90,  
                    prp = prp_90)
```

The MAMBA AUC is 0.6219448, while the PRP AUC is 0.6982682.

0.75 Nonoutlier Rate

Here, we look at the case where 0.75 is “the proportion of non-replicable SNPs which are well behaved, or nonoutliers.”

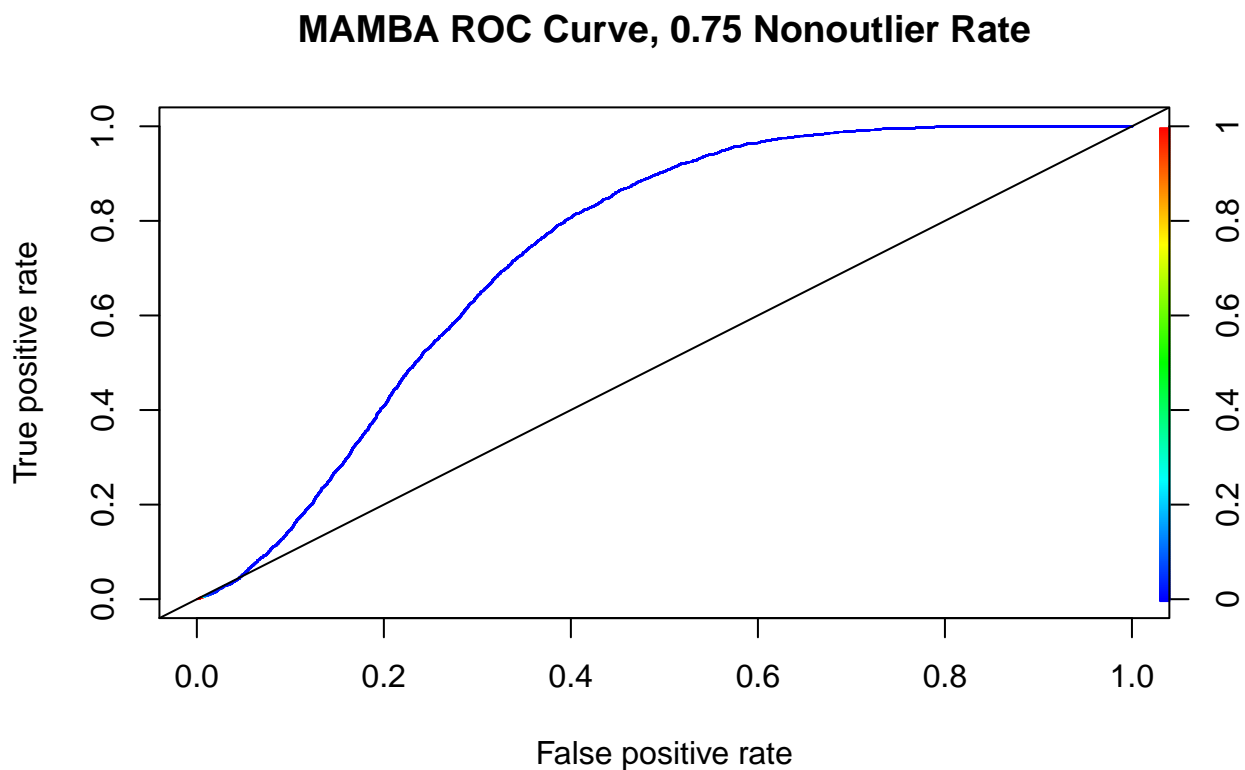
```
load(file = "data/prp_data/post_prp_data_pval_p75.rda")
load(file = "data/mamba_data/sim_mamba_mod_p75.rda")
load(file = "data/mamba_data/mamba_data_p75.rda")
```

```
prp_75 <- post_prp_data_pval
sim_mod_p75 <- sim_mod
mamba_data_p75 <- mamba_data
```

```
outliers_p75 <- mamba_data_p75$0jk
mamba_ppr_p75 <- sim_mod_p75$ppr
```

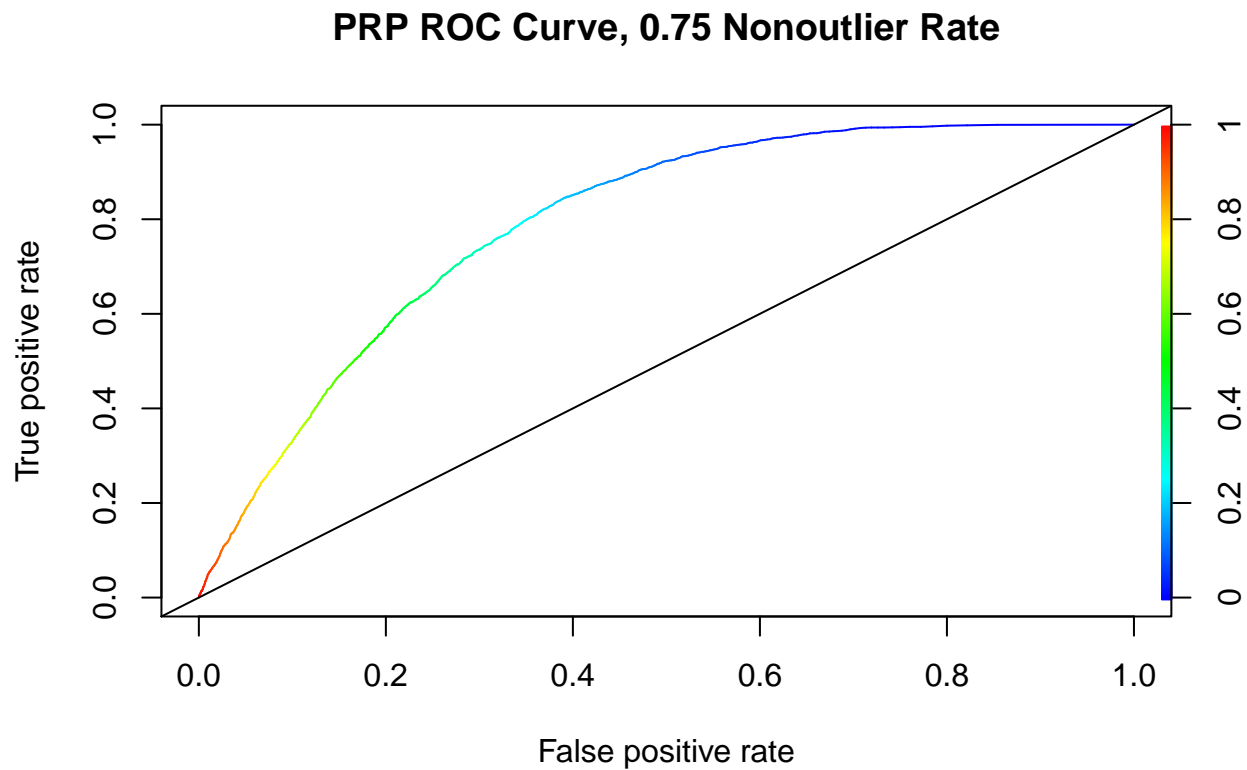
First we look at the ROC curve for the MAMBA values at a 0.75 nonoutlier study rate.

```
plot_ROC(outliers_p75,
          mamba_ppr_p75,
          main = "MAMBA ROC Curve, 0.75 Nonoutlier Rate")
```



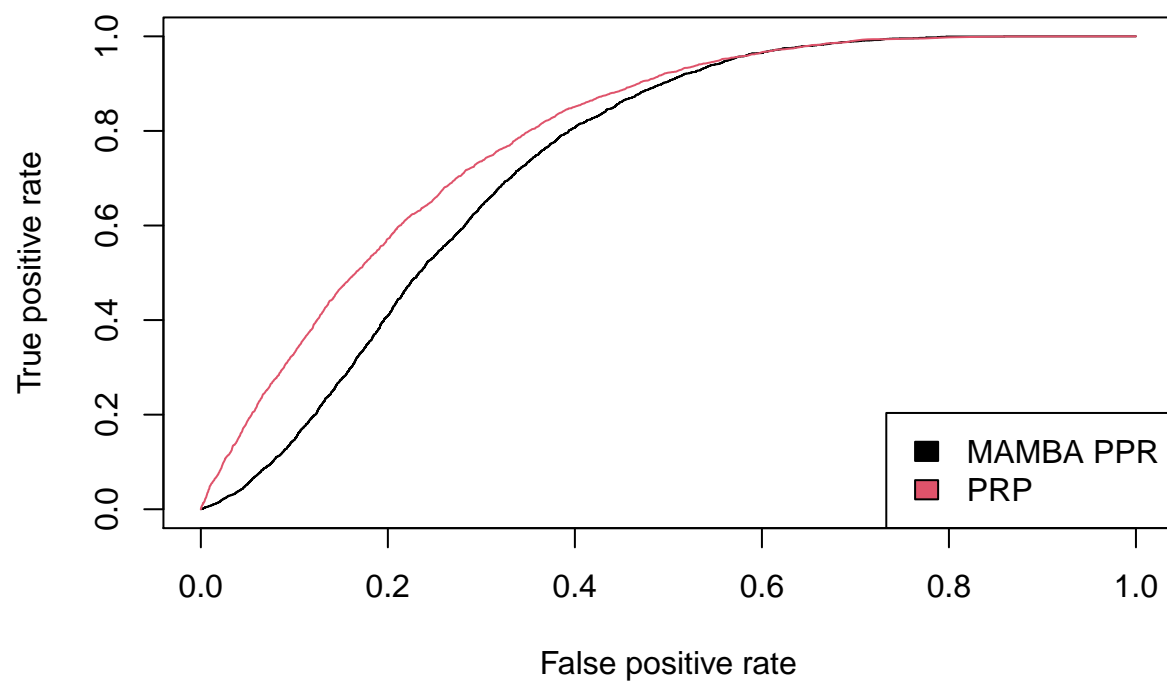
The PRP curve for the same nonoutlier rate is given below.

```
plot_ROC(outliers_p75,
         prp_75,
         main = "PRP ROC Curve, 0.75 Nonoutlier Rate")
```



Overlaying the plots, we can see that PRP has a slightly better curve than MAMBA when the rate of nonoutlier studies is 0.75.

```
ROC_compare_plot(true_outlier = outliers_p75,
                 mamba_ppr = mamba_ppr_p75,
                 prp = prp_75)
```



```
AUC_p75 <- ROC_AUC(true_outlier = outliers_p75,  
                    mamba_ppr = mamba_ppr_p75,  
                    prp = prp_75)
```

The MAMBA AUC is 0.7371868, while the PRP AUC is 0.7902613.

0.5 Nonoutlier Rate

We look at the case where 0.5 is “the proportion of non-replicable SNPs which are well behaved, or nonoutliers.”

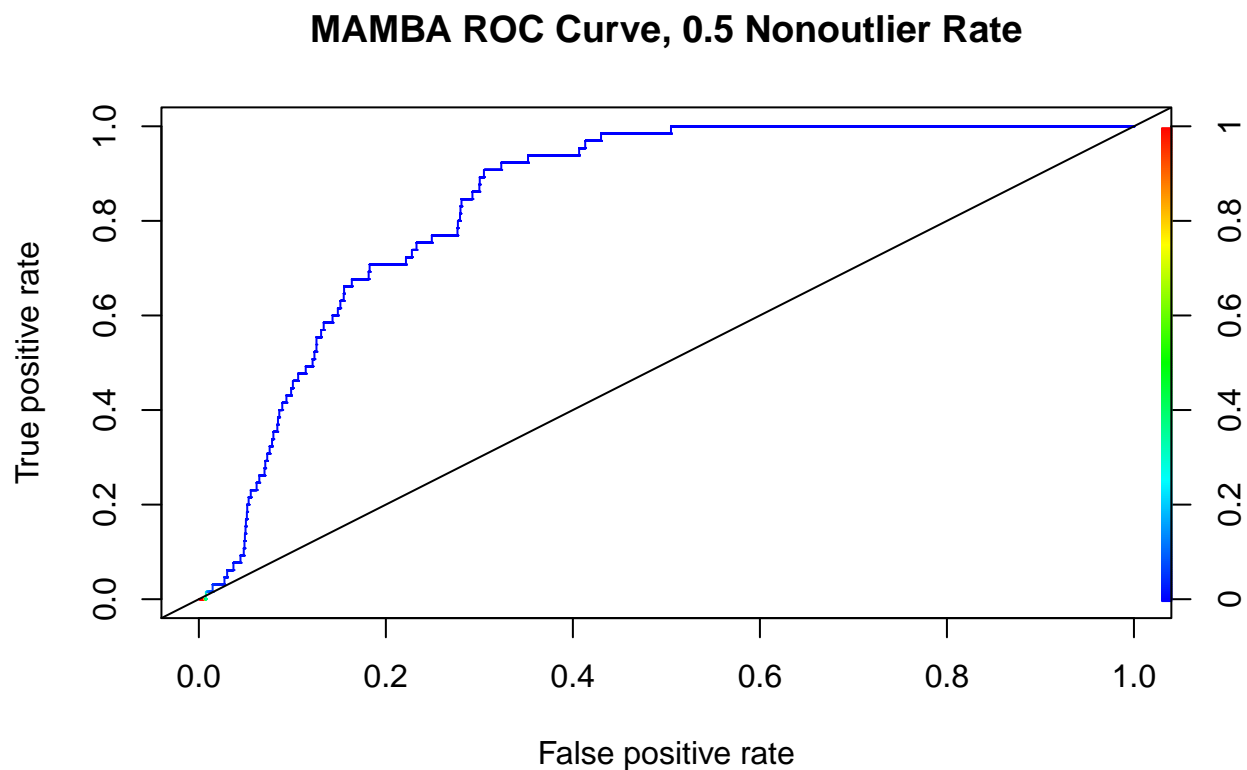
```
load(file = "data/prp_data/post_prp_data_pval_p50.rda")
load(file = "data/mamba_data/sim_mamba_mod_p50.rda")
load(file = "data/mamba_data/mamba_data_p50.rda")
```

```
prp_50 <- post_prp_data_pval
sim_mod_p50 <- sim_mod
mamba_data_p50 <- mamba_data
```

```
outliers_p50 <- mamba_data_p50$0jk
mamba_ppr_p50 <- sim_mod_p50$ppr
```

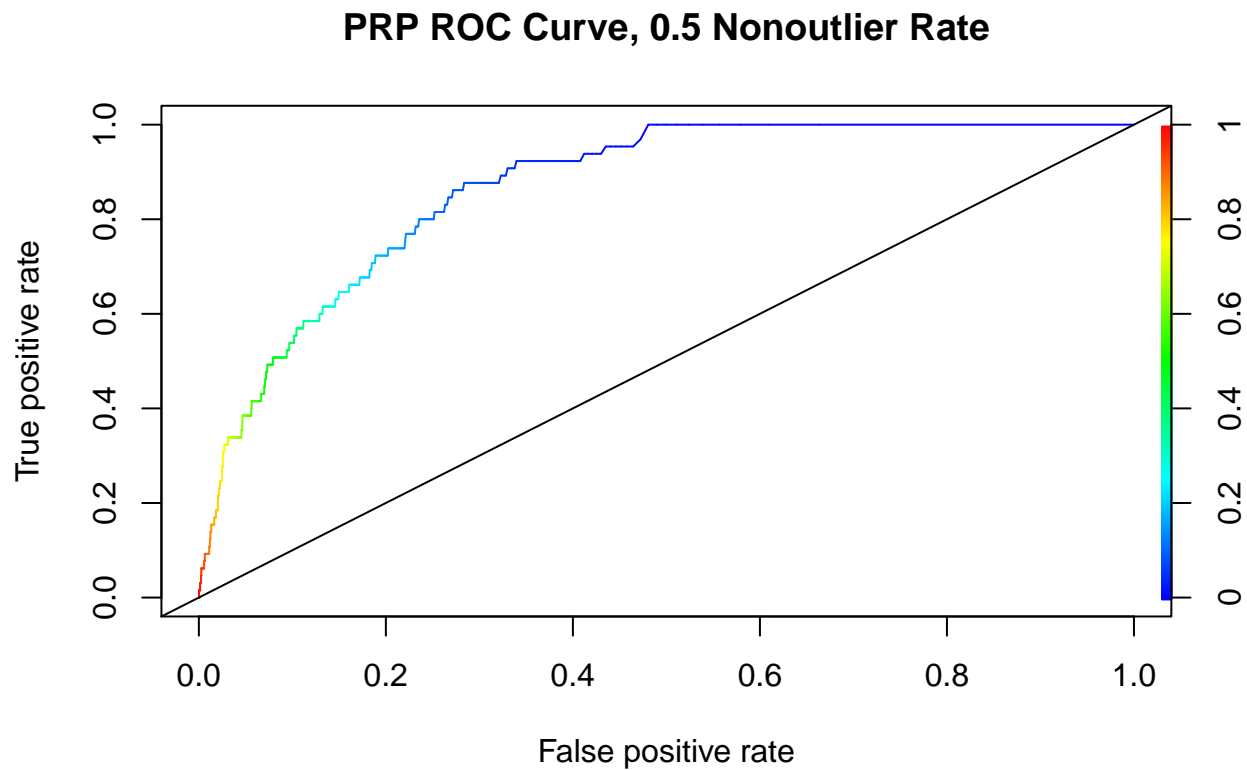
First we look at the ROC curve for the MAMBA values at a 0.5 nonoutlier study rate.

```
plot_ROC(outliers_p50,
          mamba_ppr_p50,
          main = "MAMBA ROC Curve, 0.5 Nonoutlier Rate")
```



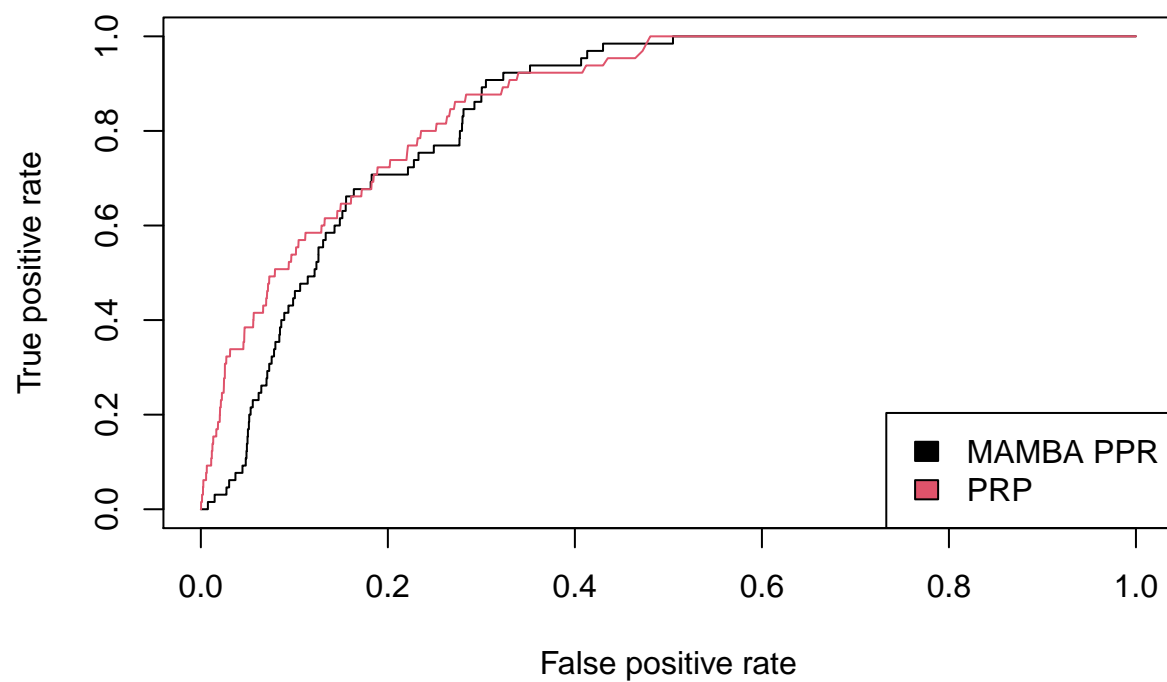
The PRP curve for the same nonoutlier rate is given below.

```
plot_ROC(outliers_p50,
         prp_50,
         main = "PRP ROC Curve, 0.5 Nonoutlier Rate")
```



Overlaying the plots, we can see that the MAMBA PPR curve and the PRP curve are starting to become more similar when the rate of nonoutlier studies is 0.50.

```
ROC_compare_plot(true_outlier = outliers_p50,
                 mamba_ppr = mamba_ppr_p50,
                 prp = prp_50)
```



```
AUC_p50 <- ROC_AUC(true_outlier = outliers_p50,  
                    mamba_ppr = mamba_ppr_p50,  
                    prp = prp_50)
```

The MAMBA AUC is 0.8454397, while the PRP AUC is 0.8664545.

0.25 Nonoutlier Rate

Finally, we look at the case where 0.25 is “the proportion of non-replicable SNPs which are well behaved, or nonoutliers.” In this specific case, because every single SNP has at least one outlier study, we need change the cutoff for the number of outlier studies for the SNP to be considered not replicable.

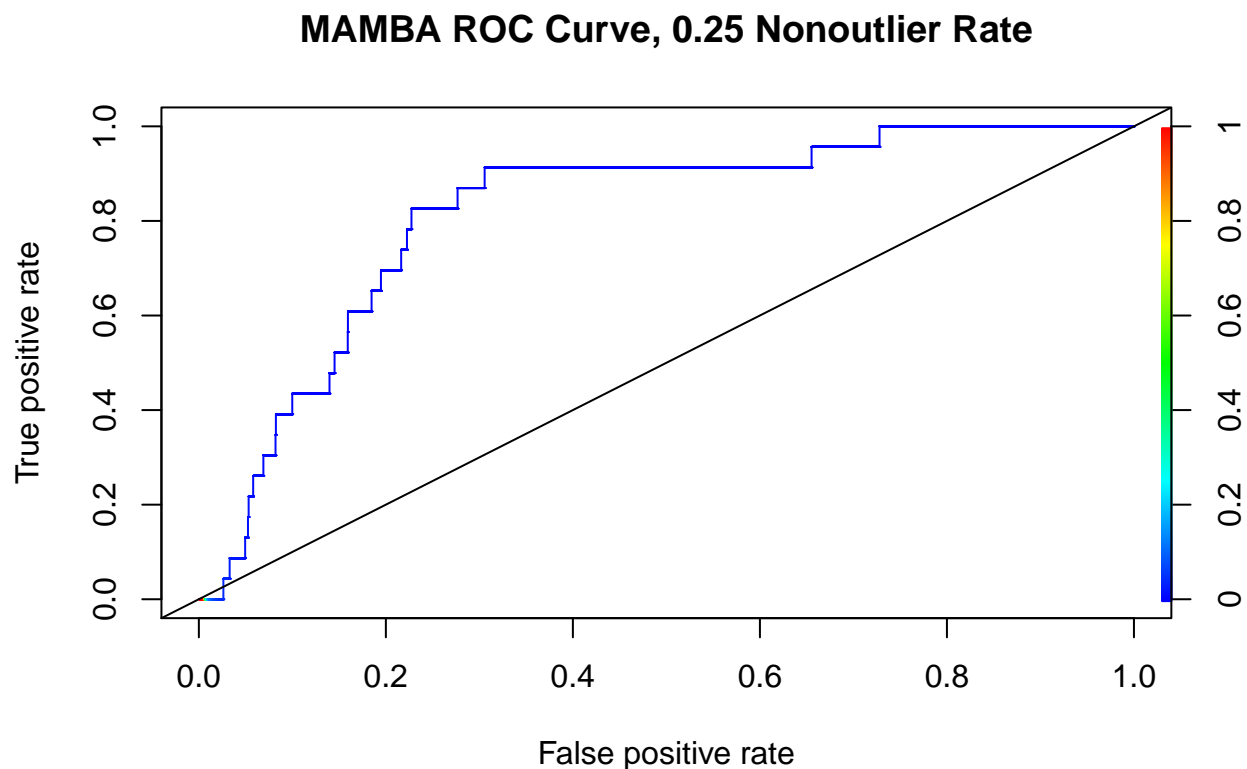
```
load(file = "data/prp_data/post_prp_data_pval_p25.rda")
load(file = "data/mamba_data/sim_mamba_mod_p25.rda")
load(file = "data/mamba_data/mamba_data_p25.rda")

prp_25 <- post_prp_data_pval
sim_mod_p25 <- sim_mod
mamba_data_p25 <- mamba_data

outliers_p25 <- mamba_data_p25$0jk
mamba_ppr_p25 <- sim_mod_p25$ppr
```

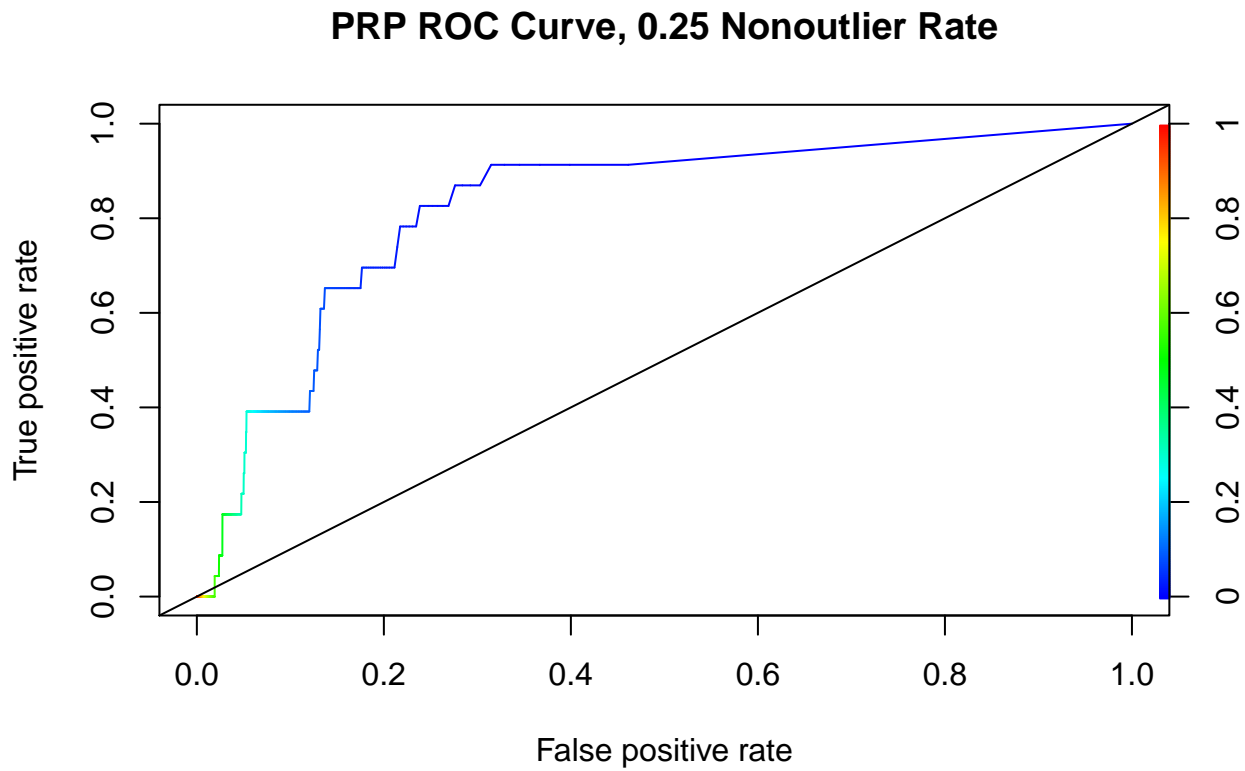
First we look at the ROC curve for the MAMBA values at a 0.25 nonoutlier study rate. Here, we changed the number of outlier studies cutoff for a SNP to be considered not replicable from one study to three studies (out of a total of ten studies in the simulated data).

```
plot_ROC(outliers_p25,
          mamba_ppr_p25,
          outlier_cutoff = 3,
          main = "MAMBA ROC Curve, 0.25 Nonoutlier Rate")
```



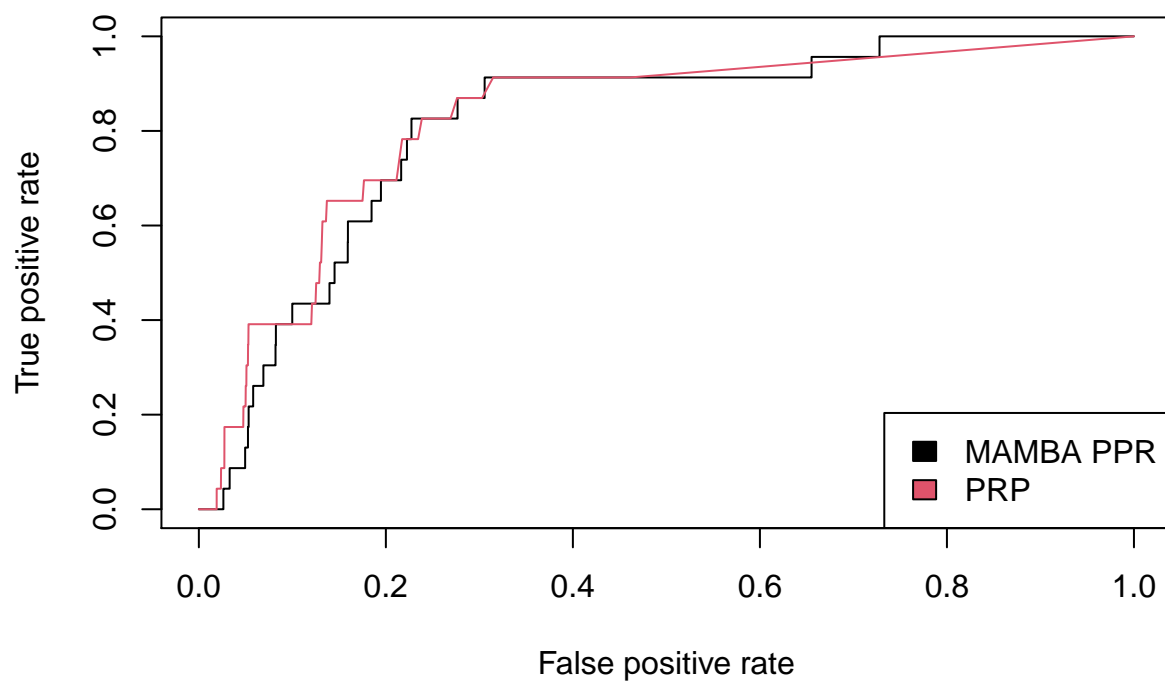
The PRP curve for the same nonoutlier rate is given below.

```
plot_ROC(outliers_p25,  
         prp_25,  
         outlier_cutoff = 3,  
         main = "PRP ROC Curve, 0.25 Nonoutlier Rate")
```



Overlaying the plots, we can see the curves are extremely similar at a 0.25 nonoutlier rate (with our cutoff set at 3 outlier studies).

```
ROC_compare_plot(true_outlier = outliers_p25,  
                 mamba_ppr = mamba_ppr_p25,  
                 prp = prp_25,  
                 outlier_cutoff = 3)
```

```
# AUC
AUC_p25 <- ROC_AUC(outliers_p25,
  mamba_ppr = mamba_ppr_p25,
  prp = prp_25,
  outlier_cutoff = 3)
```

The MAMBA AUC is 0.8164956, while the PRP AUC is 0.8256668.