

Model comparison between Chi-Squared and Beta Model

```
library(doMC)

## Loading required package: foreach
## Loading required package: iterators
## Loading required package: parallel
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
## 
##     filter, lag
## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union
library(stringr)
library(tidyr)
library(ggplot2)
library(readr)
library(eulerr)
library(qqman)

##
## For example usage please run: vignette('qqman')
##
## Citation appreciated but not required:
## Turner, (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. Journ
## 

registerDoMC(4) # Set number of cores here for parallel processing

donors <- c('CTRL-NEUHE723FGT-02545-G', 'CASE-NEUFV237VCZ-01369-G', 'CTRL-NEUHE723FGT-02545-G', 'CASE-NC
rbps = c("24a-hNIL-control-tdp", "24a-hNIL-c9-tdp", "24a-hNIP-control-tdp", "24a-hNIP-c9-tdp")

input_files <- paste0("../data/", rbps, "_input_allelic.out")
ip_files <- paste0("../data/", rbps, "_ip_allelic.out")
filtered_input_files <- str_replace(input_files, "out$", "10readsInput_filtered_epsilon0.3_sharedhet.txt")

sample_index <- 'CTRL-NEUHE723FGT-02545-G'
rbp = "24a-hNIL-control-tdp"
epsilon = 0.3
input_file <- input_files[1]
```

```

ip_file <- ip_files[1]
filtered_input_file <- str_replace(input_file, "out$", "10readsInput_filtered_epsilon0.3_sharedhet.txt")
input <- read_tsv(input_file)

## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Rows: 2791142 Columns: 12

## -- Column specification -----
## Delimiter: "\t"
## chr (4): contig, variantID, refAllele, altAllele
## dbl (8): position, refCount, altCount, totalCount, lowMAPQDepth, lowBaseQDep...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

bqtlis_beta_files <- paste0("../results/",
                           rbps,
                           '_beta_10readsInput_filtered_epsilon0.3_ipreads30_allhet_struct_with_peaks.tsv'
                           #'_beta_filtered_epsilon0.3_ipreads30_struct_with_peaks.tsv.gz')
bqtlis_betas <- lapply(bqtlis_beta_files, read_tsv)

## Rows: 3788 Columns: 28
## -- Column specification -----
## Delimiter: "\t"
## chr (4): chrom, variantID, refAllele, altAllele
## dbl (24): position, refCount_input, altCount_input, totalCount_input, pred_r...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 3224 Columns: 28
## -- Column specification -----
## Delimiter: "\t"
## chr (4): chrom, variantID, refAllele, altAllele
## dbl (24): position, refCount_input, altCount_input, totalCount_input, pred_r...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 2860 Columns: 28
## -- Column specification -----
## Delimiter: "\t"
## chr (4): chrom, variantID, refAllele, altAllele
## dbl (24): position, refCount_input, altCount_input, totalCount_input, pred_r...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 3011 Columns: 28
## -- Column specification -----
## Delimiter: "\t"
## chr (4): chrom, variantID, refAllele, altAllele
## dbl (24): position, refCount_input, altCount_input, totalCount_input, pred_r...
##

```

```
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Run chi-squared tests to test if the distribution of alt and ref counts are the same between the two groups (IP and input) or different

```
chisq_files <- paste0("../results/", rbps, '_chisq_results.txt')
all_chisq <- lapply(1:length(rbps), function(j) {
  chisq = foreach(i = 1:nrow(bqtls_betas[[j]]), .combine = bind_rows) %dopar% {
    here = bqtls_betas[[j]][i,]
    input_ref_count <- here$RefCount_input
    input_alt_count <- here$AltCount_input
    IP_ref_count <- here$RefCount_IP
    IP_alt_count <- here$AltCount_IP

    contingency_table <- matrix(c(input_ref_count, input_alt_count, IP_ref_count, IP_alt_count),
                                  nrow = 2,
                                  byrow = TRUE,
                                  dimnames = list(c("Input", "IP"), c("Ref", "Alt")))

    # Perform chi-squared test
    chi_square_result <- chisq.test(contingency_table)

    # Calculate proportions in each group
    prop_input <- input_ref_count / (input_ref_count + input_alt_count)
    prop_IP <- IP_ref_count / (IP_ref_count + IP_alt_count)

    # Calculate log ratio of proportions
    log_ratio_prop <- log10(prop_IP) - log10(prop_input)

    with(chi_square_result, tibble(statistic, parameter, p.value, lor = log_ratio_prop))
  }

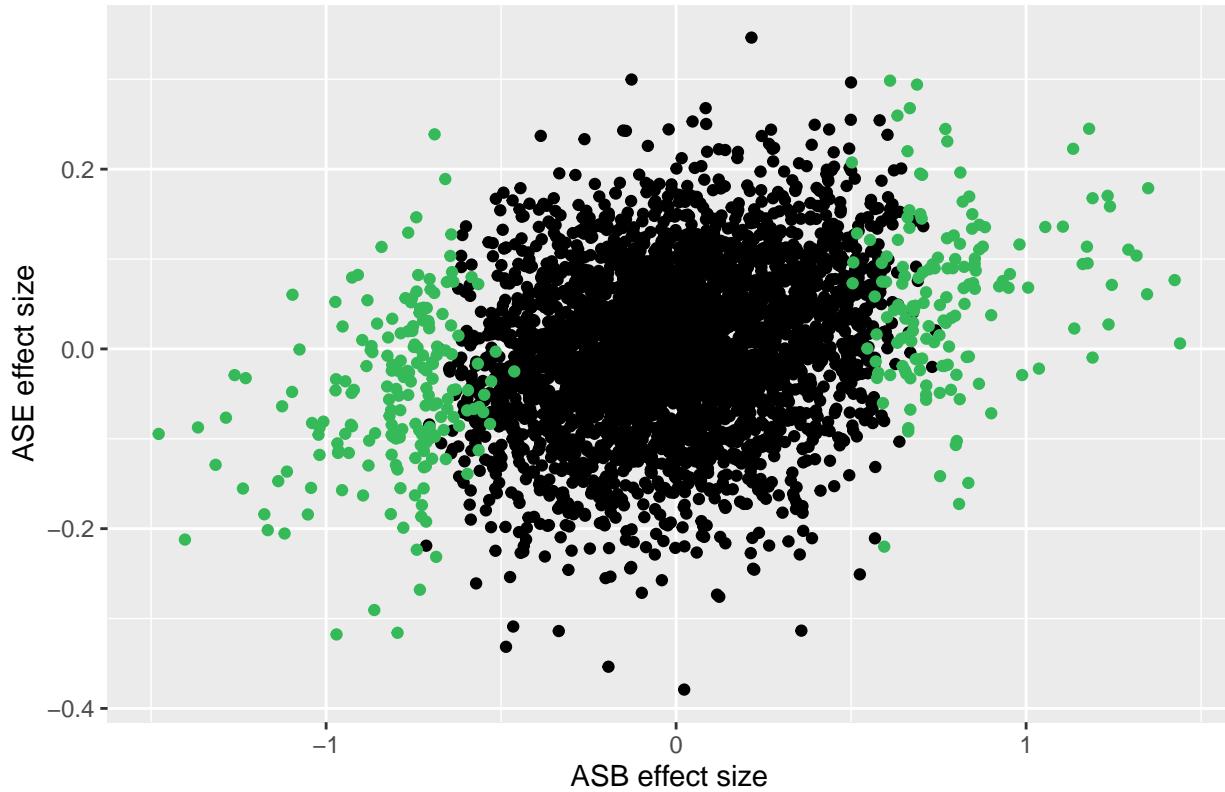
  chisq$q.value = p.adjust(chisq$p.value, method = 'fdr')
  write_tsv(chisq, chisq_files[j])
  return(chisq)
})
```

Visualizing significant hits from each model

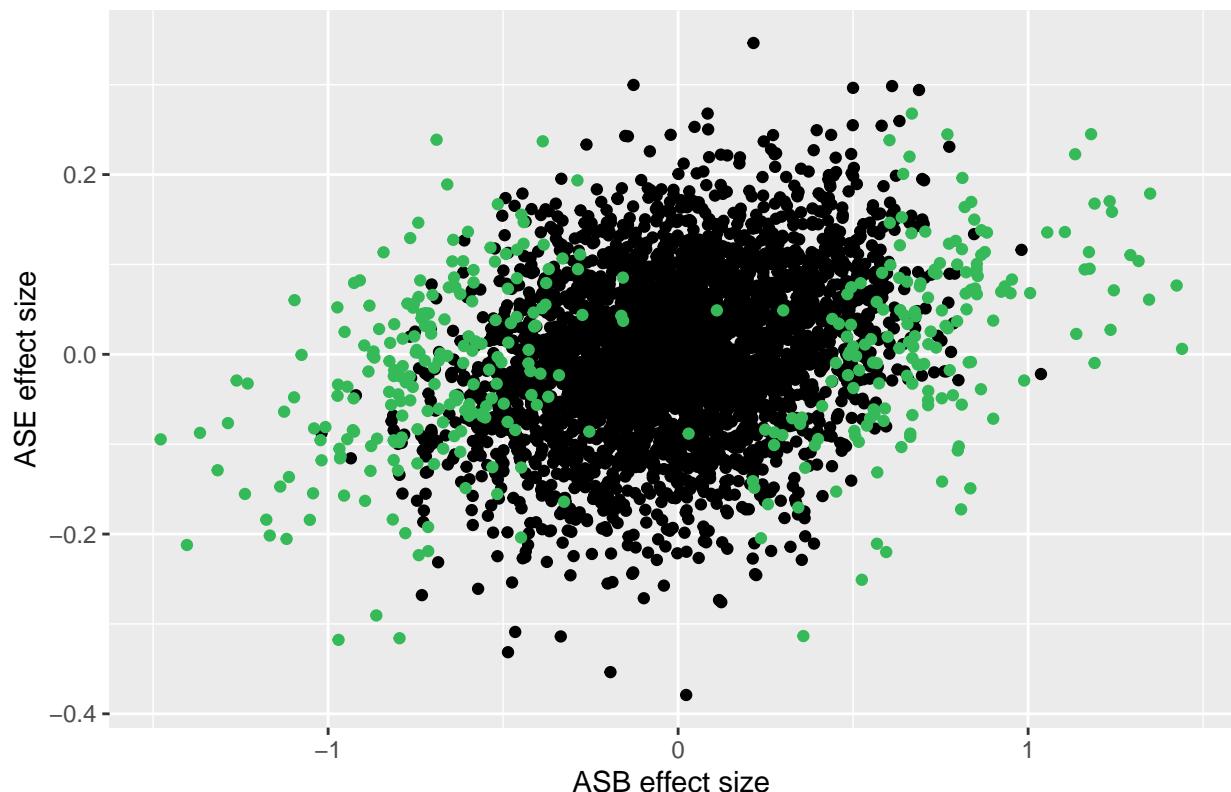
```
for (j in 1:length(rbps)) {
  bqtls_beta <- bqtls_betas[[j]]
  chisq <- all_chisq[[j]]
  for (qthresh in c(0.05)) {
    p1 <- ggplot(bqtls_beta, aes(x = asb_loc, y = ase_loc)) +
      geom_point(data=filter(bqtls_beta, asb_q >= qthresh), color='black')+
      geom_point(data=filter(bqtls_beta, asb_q < qthresh), color='#36ba59') +
      labs(x = "ASB effect size", y = "ASE effect size", title = str_glue("{sum(bqtls_beta$asb_q >= {qthresh})}"))
    p2 <- ggplot(bqtls_beta, aes(x = asb_loc, y = ase_loc)) +
      geom_point(data=bqtls_beta[chisq$q.value >= qthresh,], color='black')+
      geom_point(data=bqtls_beta[chisq$q.value < qthresh,], color='#36ba59') +
      labs(x = "ASB effect size", y = "ASE effect size", title = str_glue("{sum(chisq$q.value < {qthresh})}"))
    print(p1)
```

```
        print(p2)
    }
}
```

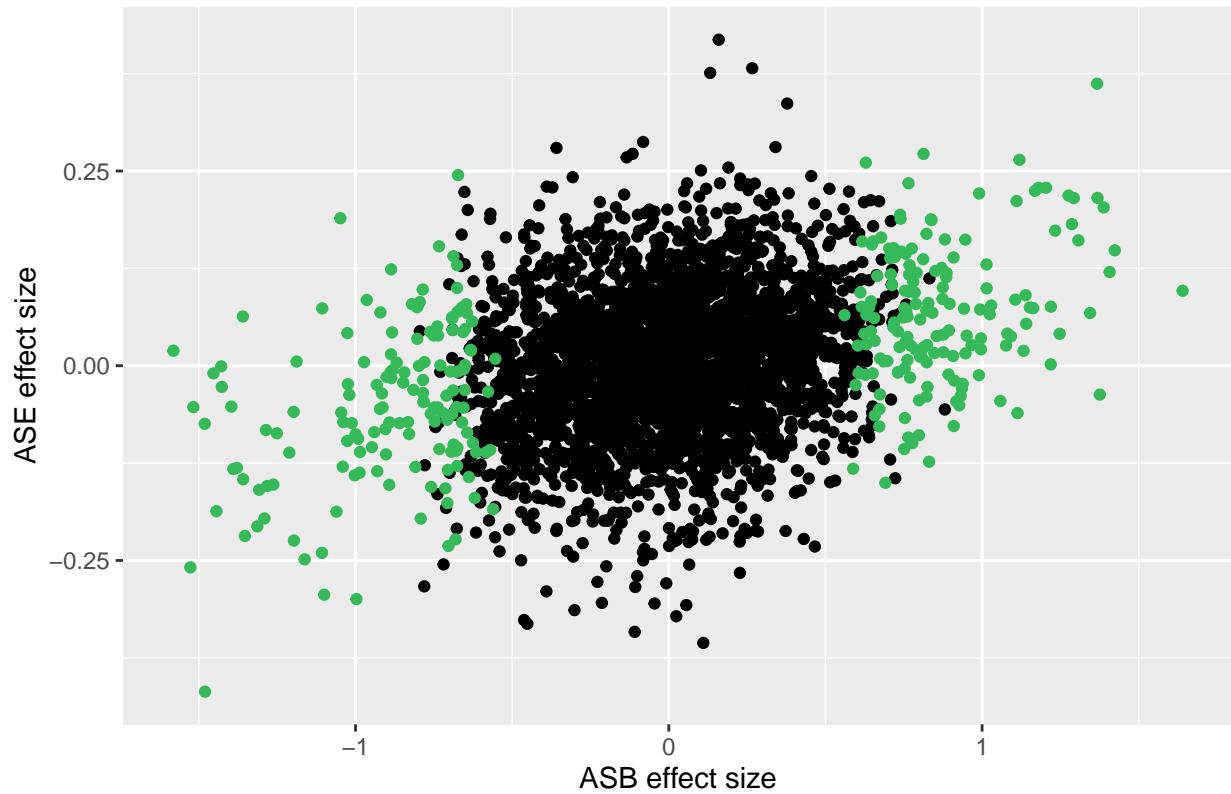
345 SNPs with $q < 0.05$ | beta model for 24a-hNIL-control-tdp



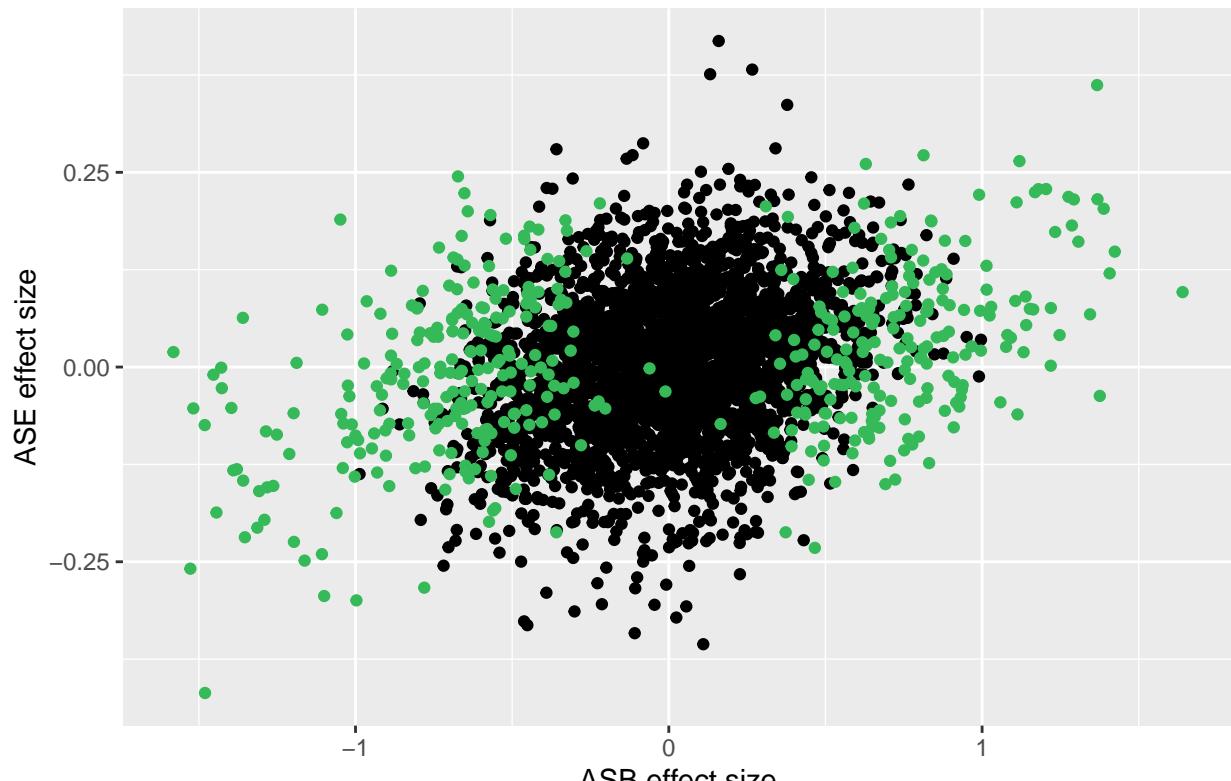
378 SNPs with $q < 0.05$ | chi squared model for 24a–hNIL–control–tdp



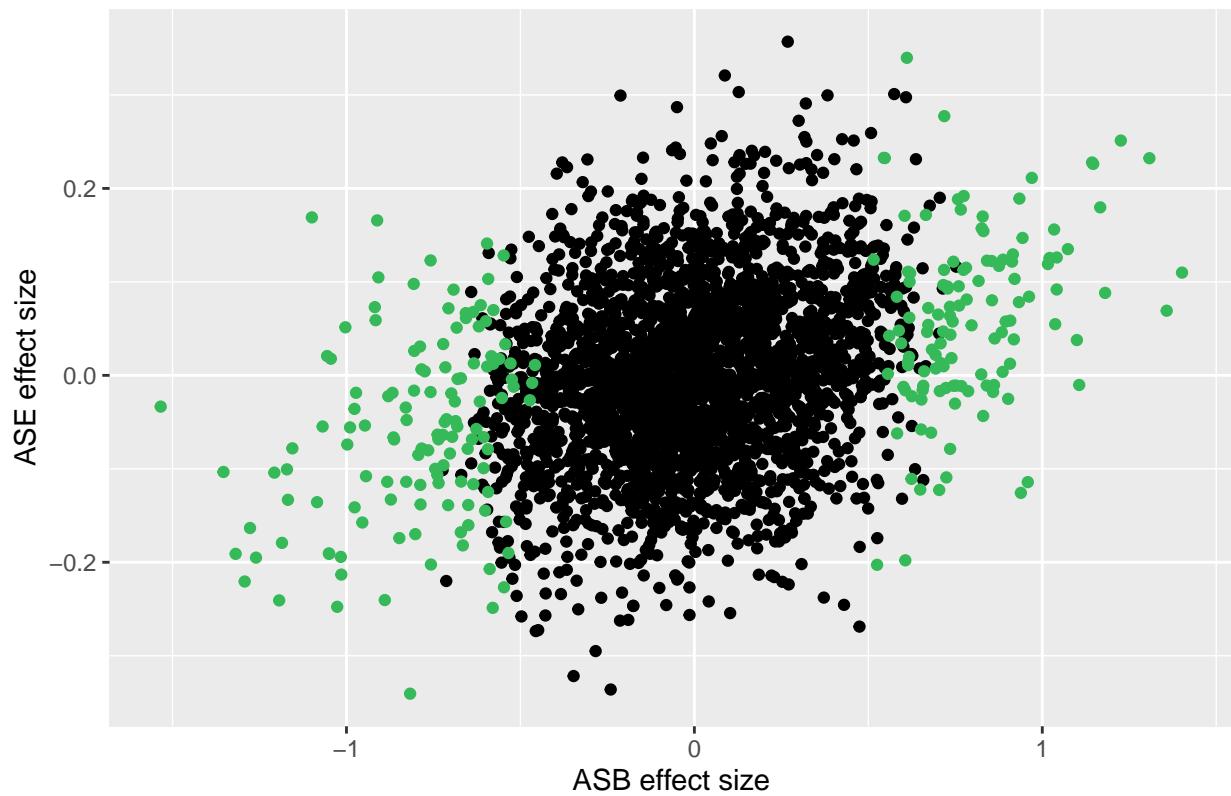
316 SNPs with $q < 0.05$ | beta model for 24a–hNIL–c9–tdp



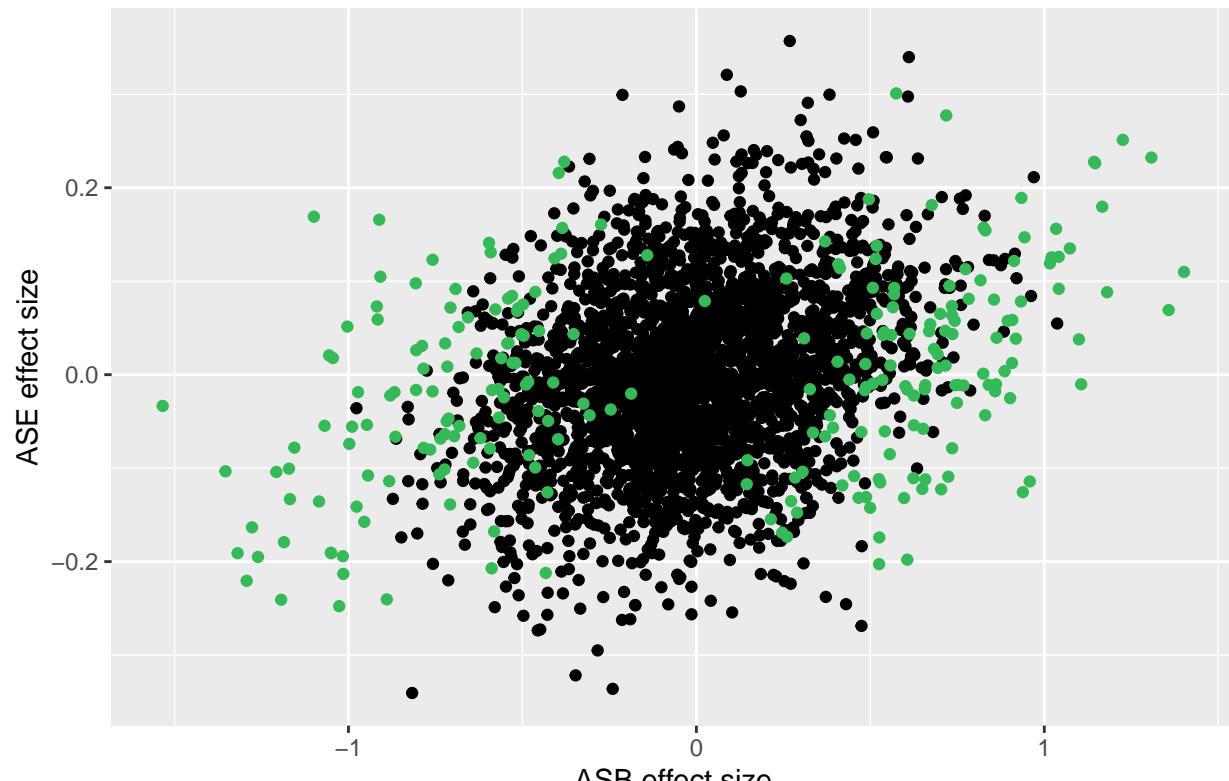
449 SNPs with $q < 0.05$ | chi squared model for 24a–hNIL–c9–tdp



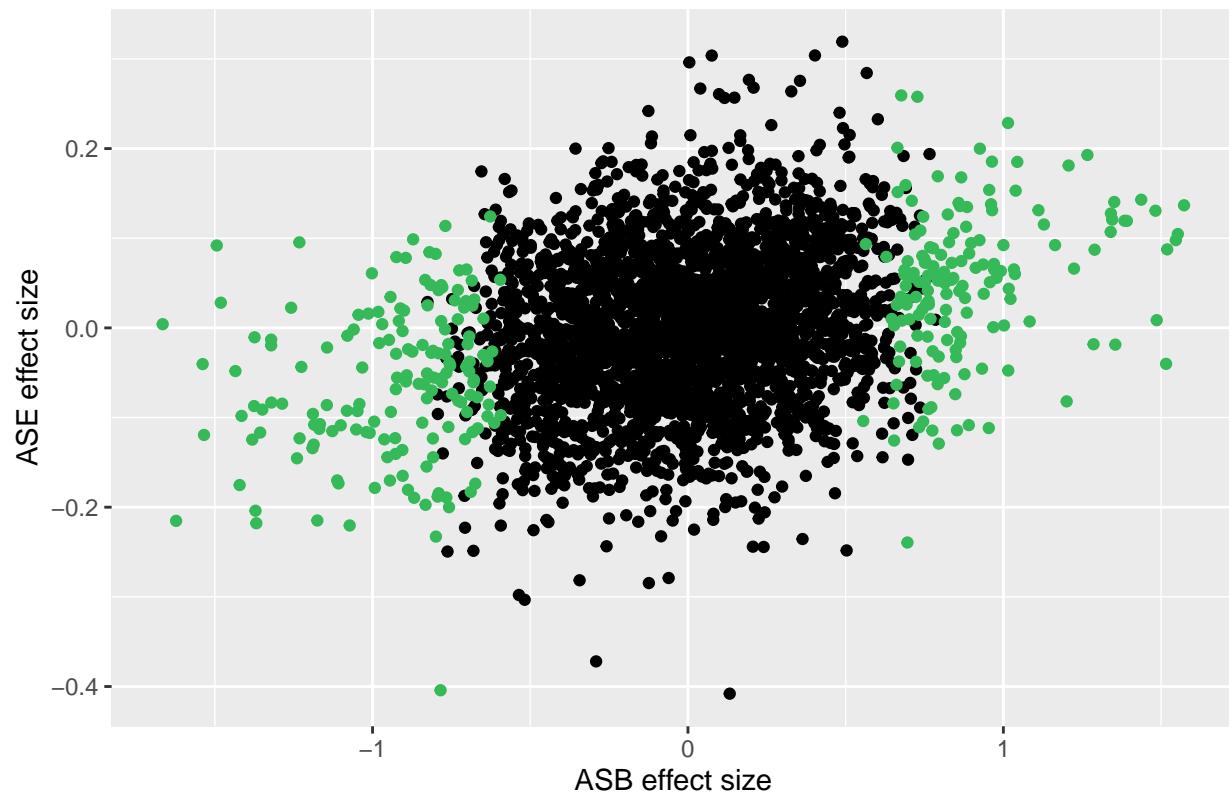
250 SNPs with $q < 0.05$ | beta model for 24a–hNIP–control–tdp



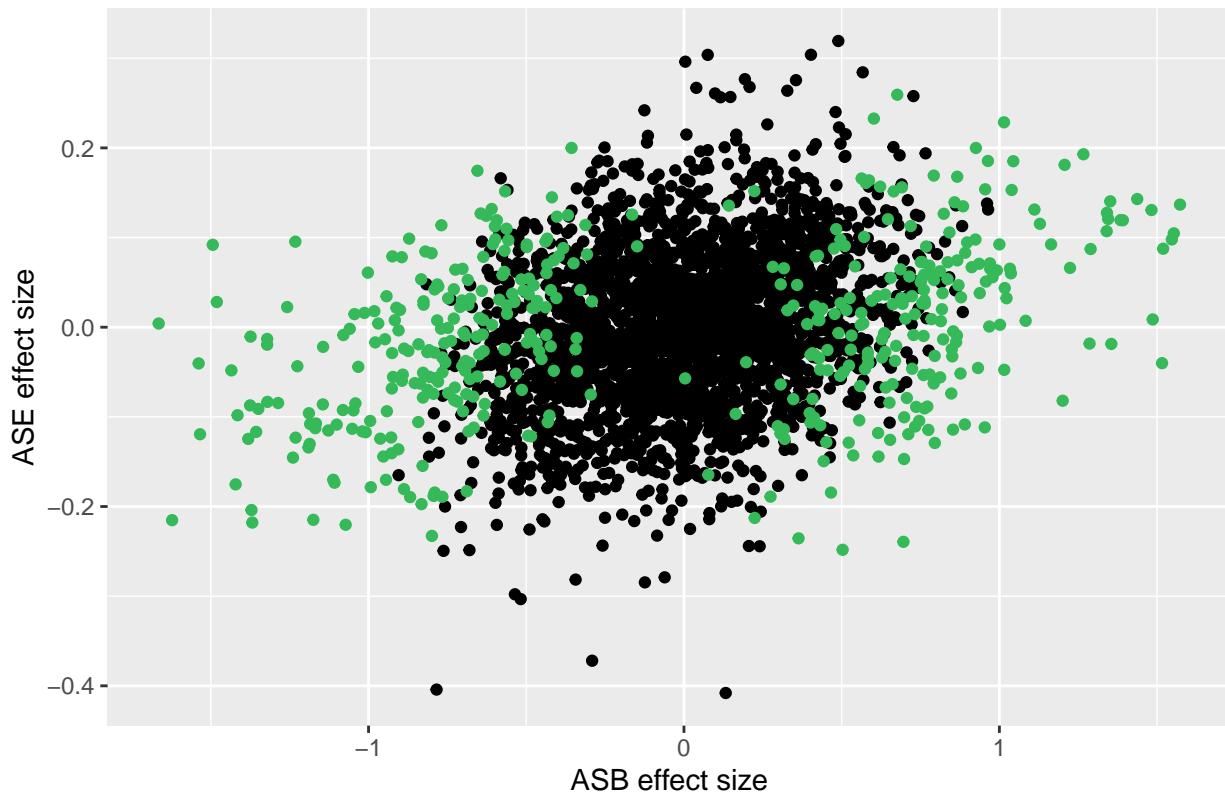
240 SNPs with $q < 0.05$ | chi squared model for 24a–hNIP–control–tdp



315 SNPs with $q < 0.05$ | beta model for 24a–hNIP–c9–tdp



448 SNPs with $q < 0.05$ | chi squared model for 24a–hNIP–c9–tdp



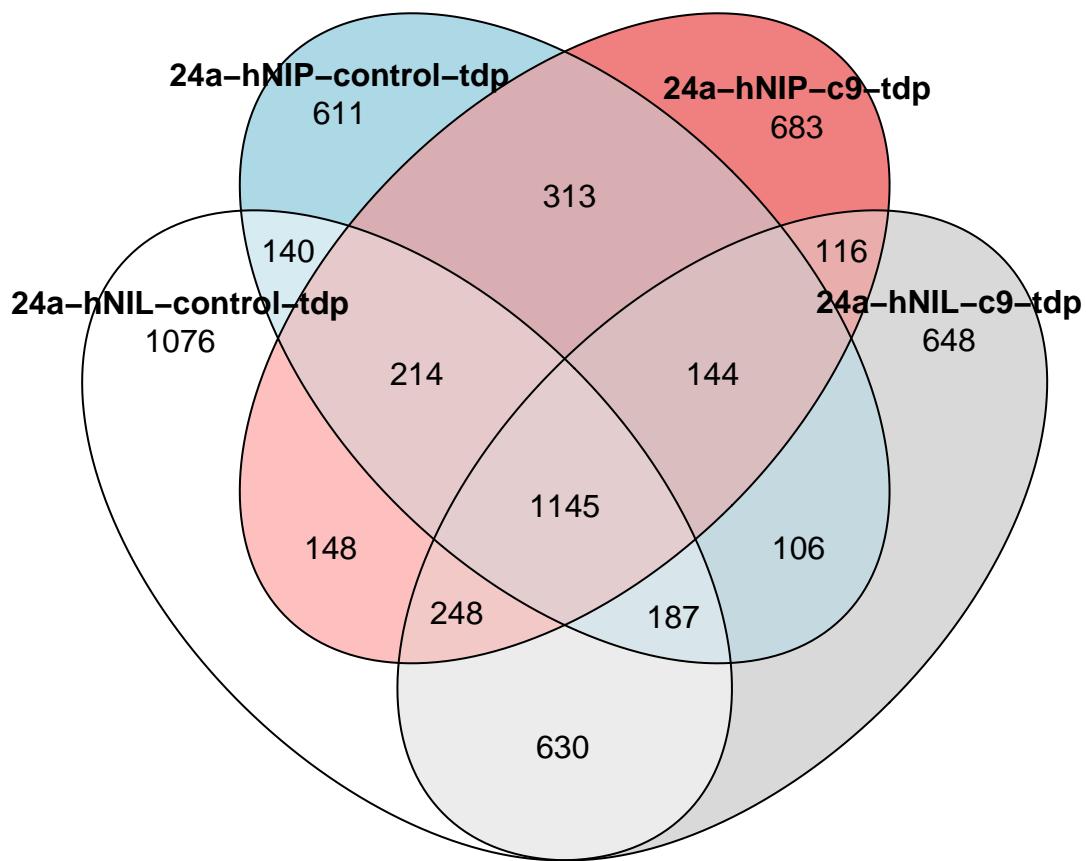
How many significant hits are in peaks?

```
qthresh = 0.05
x = bqtls_beta[bqtls_beta$asb_q < qthresh,]
y = bqtls_beta[chisq$q.value < qthresh,]
x_idx = x$in_peak == 1
y_idx = y$in_peak == 1
cat(sum(x_idx)%>%paste0(" out of ", nrow(x), " ", round(sum(x_idx)/nrow(x)*100,3), "% of variants in pe
## 14 out of 315 4.444% of variants in peaks under chi-squared model
cat(sum(y_idx)%>%paste0(" out of ", nrow(y), " ", round(sum(y_idx)/nrow(y)*100,3), "% of variants in pe
## 28 out of 448 6.25% of variants in peaks under beta model
cat(paste0(
  length(intersect(x$variantID, y$variantID)), " ",
  length(union(x$variantID, y$variantID)), " ",
  round(length(intersect(x$variantID, y$variantID))/length(union(x$variantID, y$variantID))*100,
  "% of variants in common between the two models"
), '\n')
## 260 503 51.69% of variants in common between the two models
```

Shared rbQTLs Venn-diagram (looking across all 4 samples)

```
bqtls_betas%>%lapply(function(x) x$variantID)%>%
`names<-`(`rbps`)%>%
```

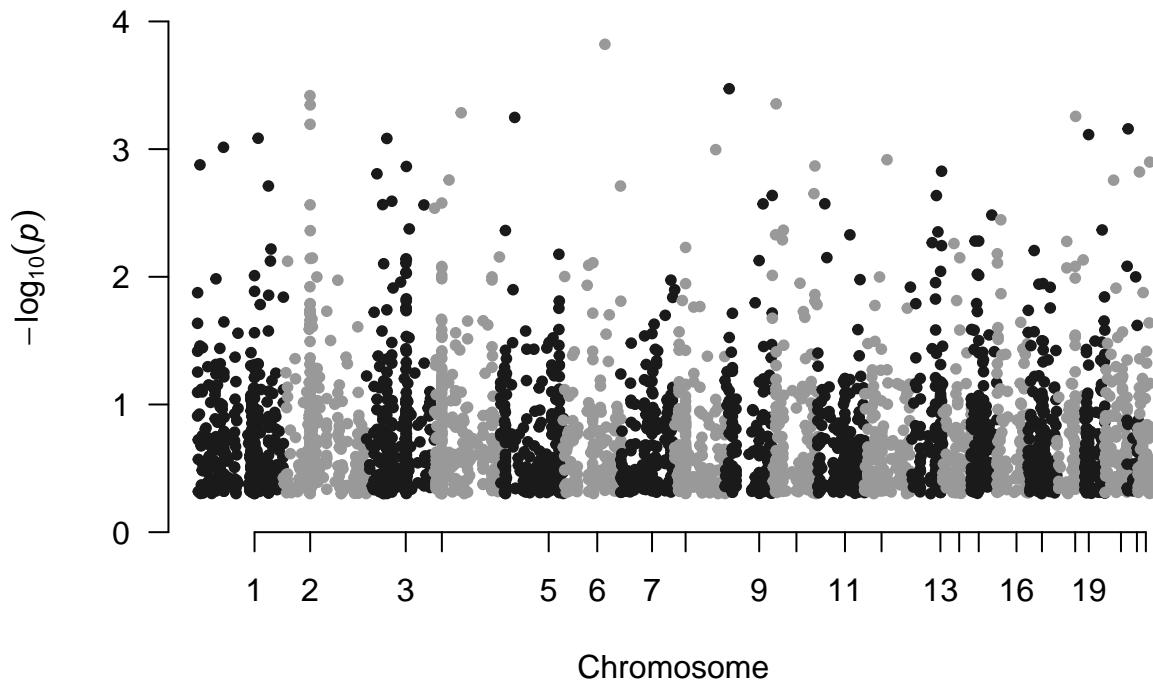
```
venn%>%
plot
```



Manhattan plots across all 4 conditions for both models

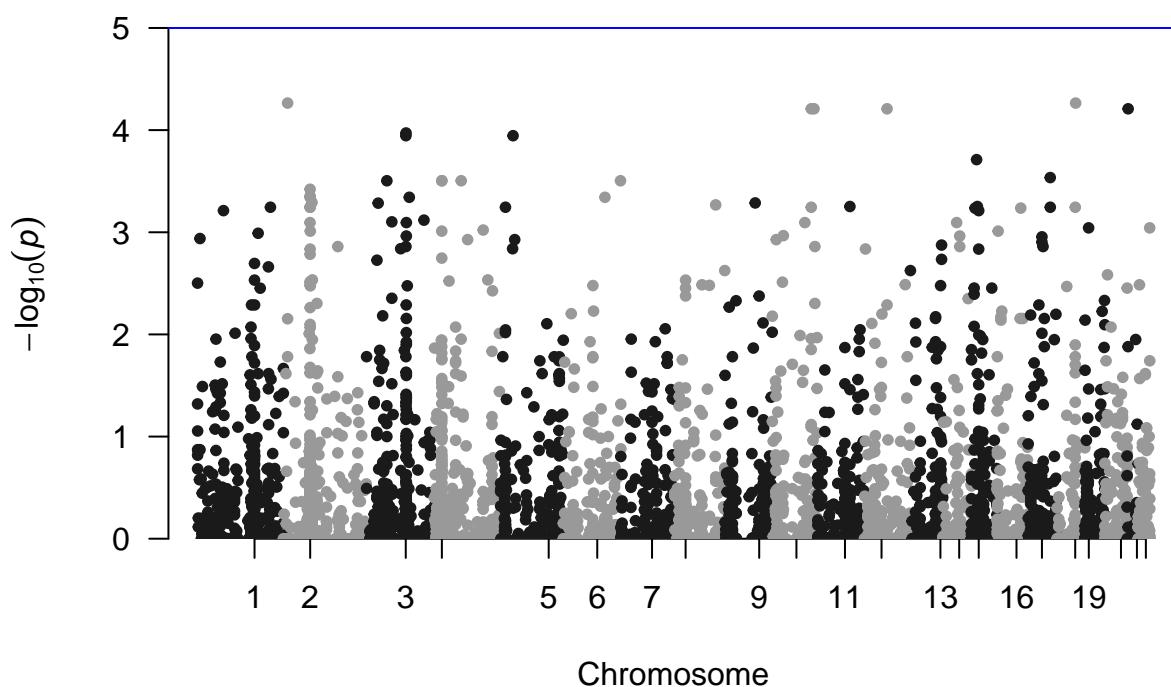
```
for (j in 1:length(bqtls_betas)) {
  bqtls_beta <- bqtls_betas[[j]]
  manhattan_plot <- manhattan(data.frame(CHR = bqtls_beta$chrom%>%str_extract("[0-9]+")%>%as.integer,
                                         BP = bqtls_beta$position,
                                         SNP = bqtls_beta$variantID,
                                         P = bqtls_beta$asb_q), main = paste0(rbps[j], "beta model"))
  print(manhattan_plot)
  manhattan_plot <- manhattan(data.frame(CHR = bqtls_beta$chrom%>%str_extract("[0-9]+")%>%as.integer,
                                         BP = bqtls_beta$position,
                                         SNP = bqtls_beta$variantID,
                                         P = all_chisq[[j]]$q.value), main = paste0(rbps[j], "Chi-squared"))
  print(manhattan_plot)
}
```

24a-hNIL-control-tdpbeta model



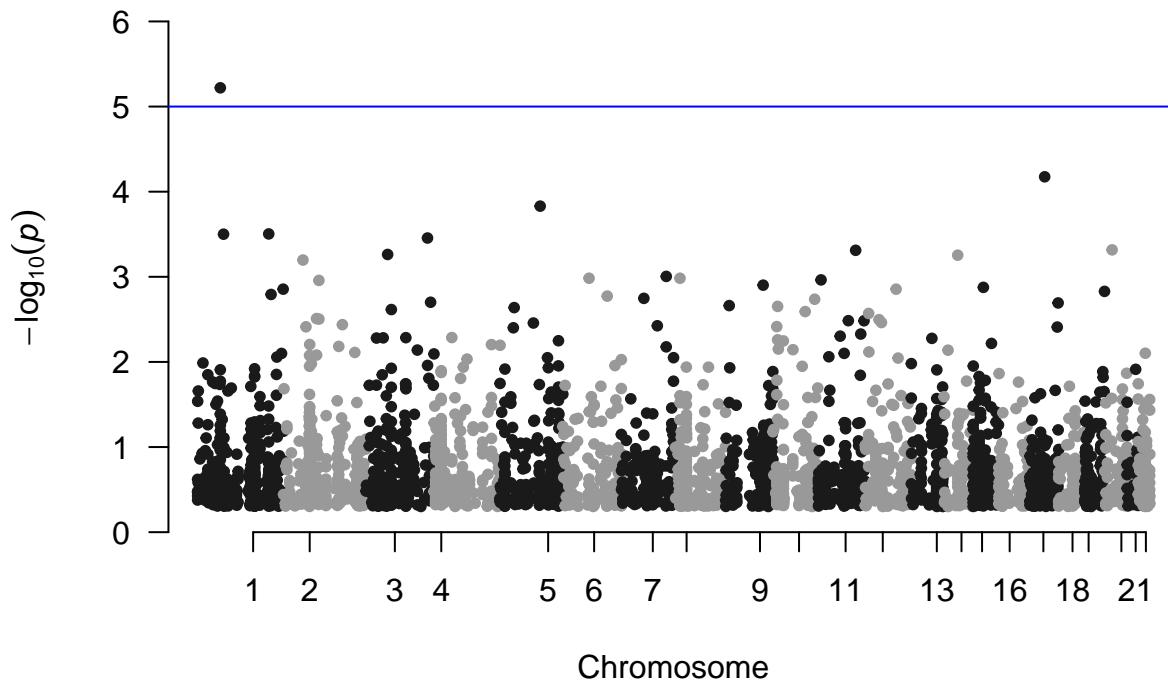
```
## $xpd  
## [1] FALSE
```

24a-hNIL-control-tdpChi-squared model



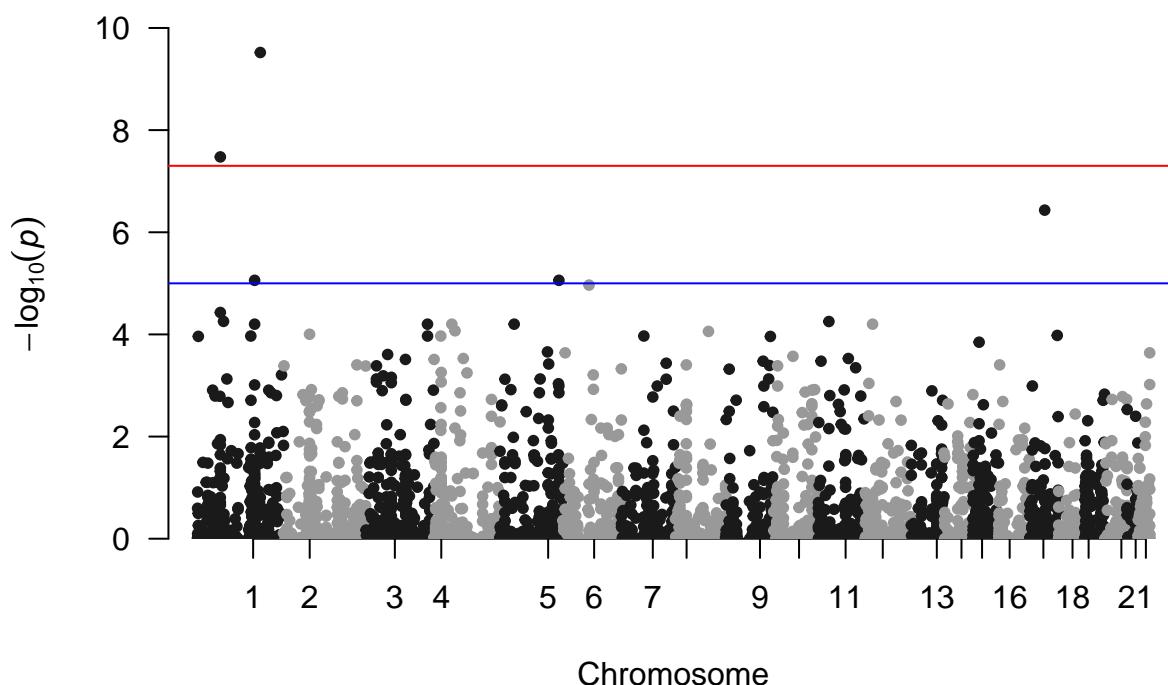
```
## $xpd  
## [1] FALSE
```

24a-hNIL-c9-tdpbeta model



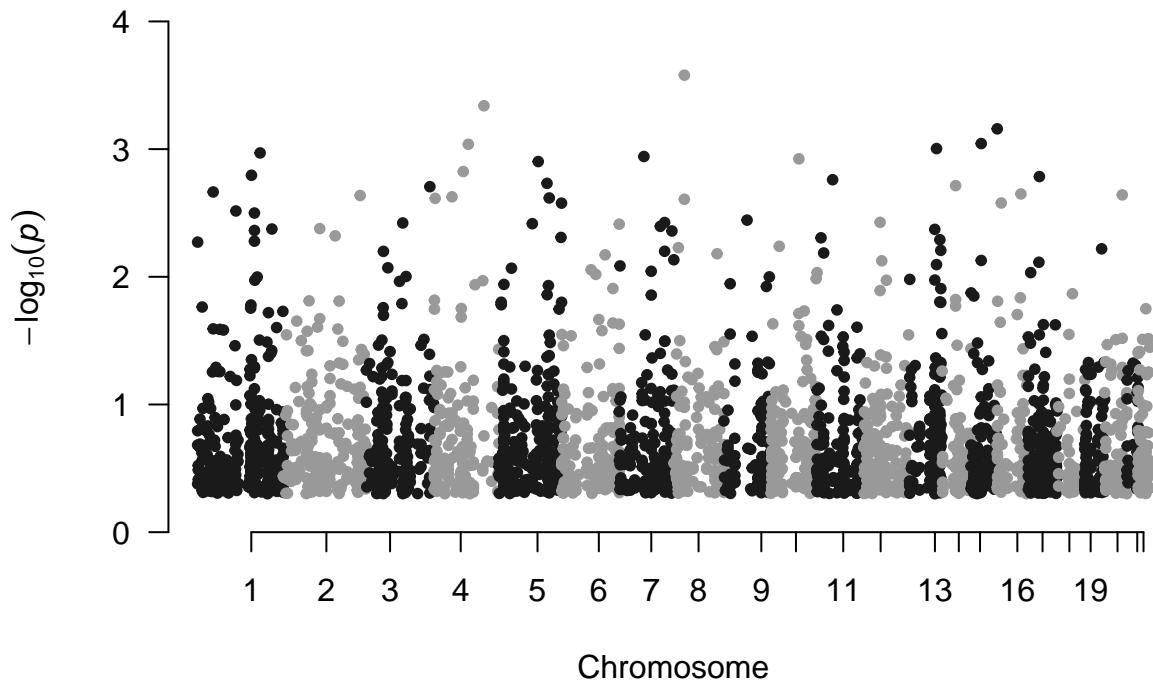
```
## $xpd  
## [1] FALSE
```

24a-hNIL-c9-tdpChi-squared model



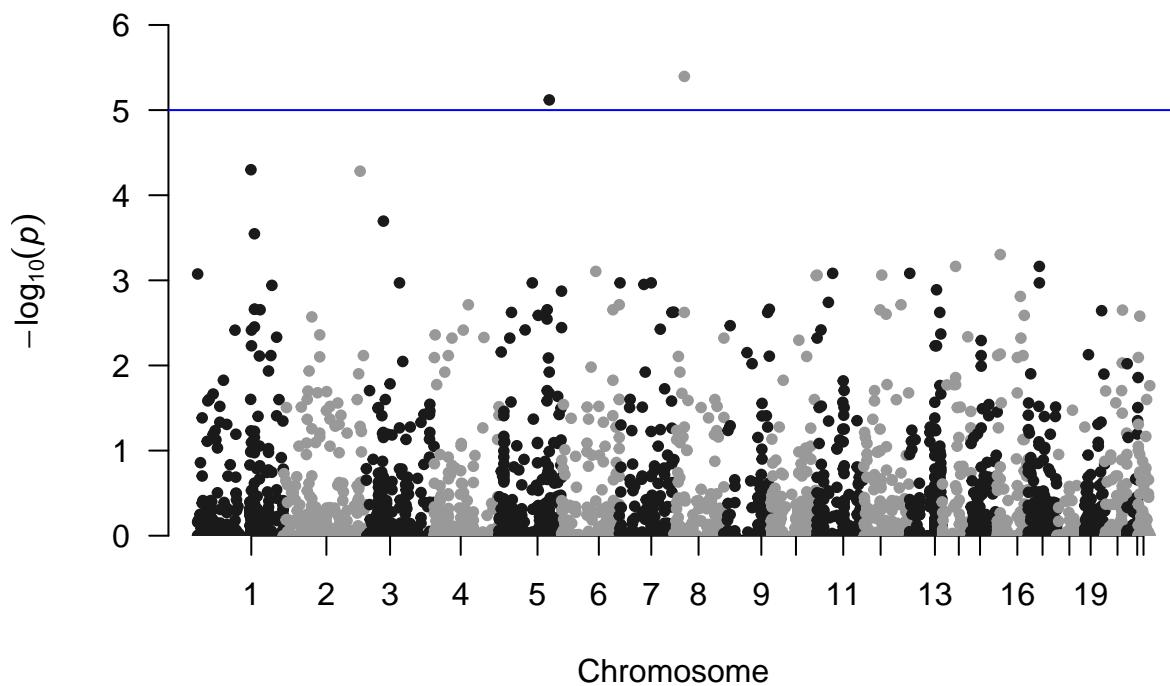
```
## $xpd  
## [1] FALSE
```

24a-hNIP-control-tdpbeta model



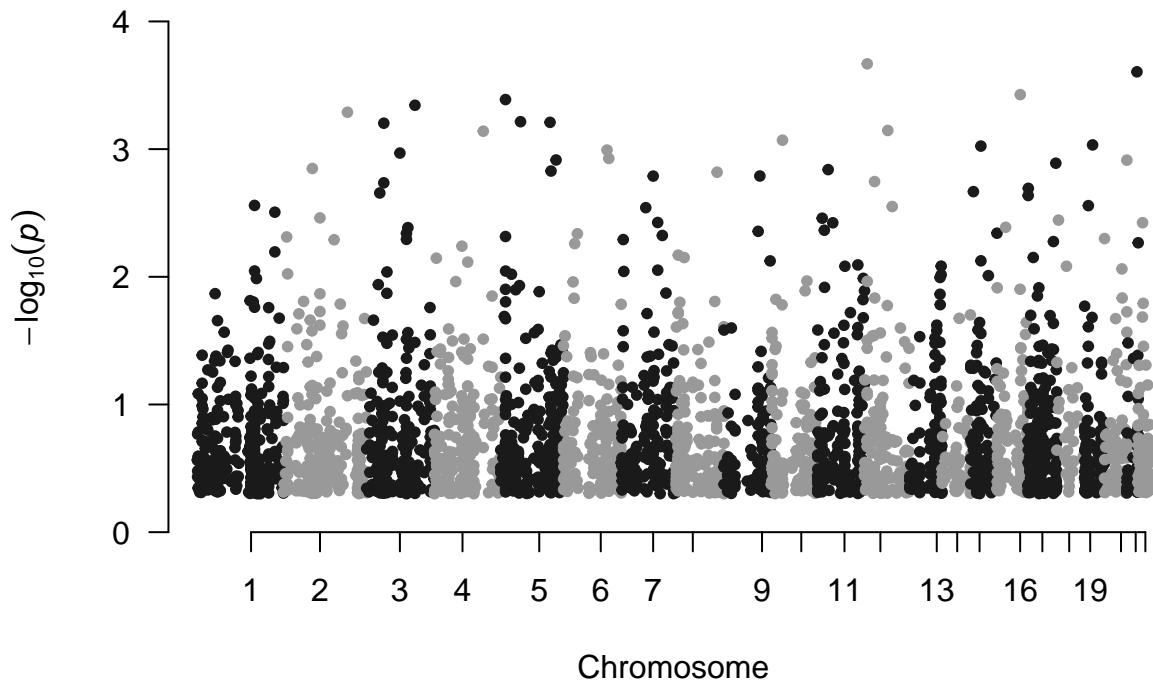
```
## $xpd  
## [1] FALSE
```

24a-hNIP-control-tdpChi-squared model



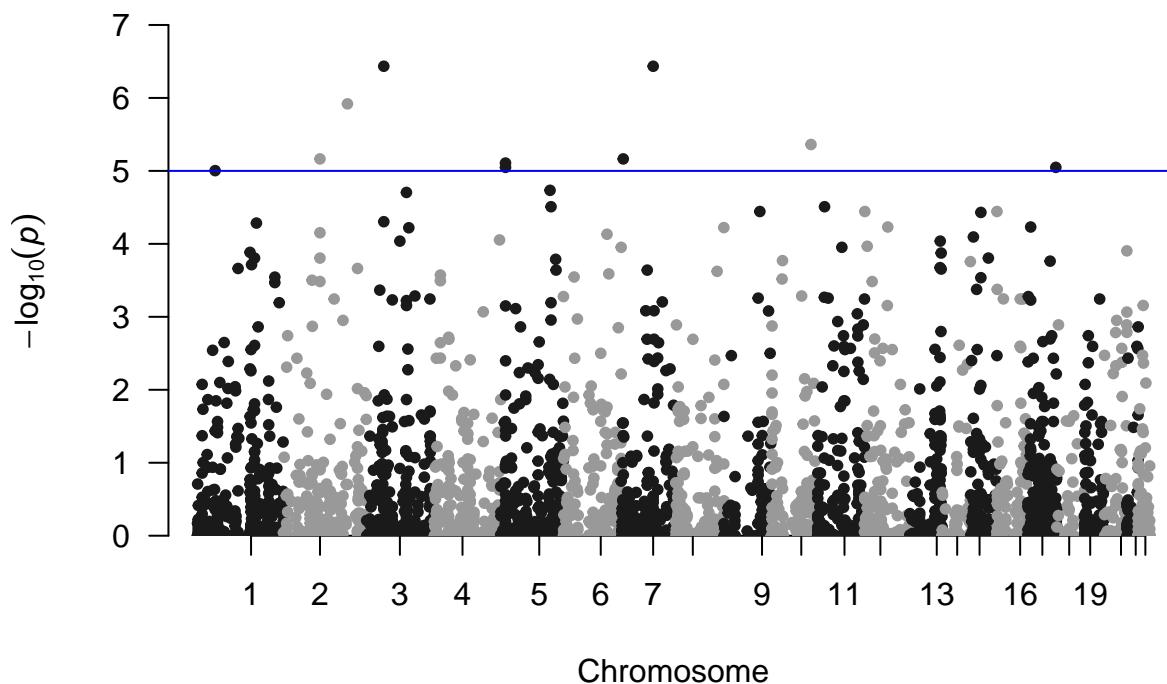
```
## $xpd  
## [1] FALSE
```

24a-hNIP-c9-tdpbeta model



```
## $xpd  
## [1] FALSE
```

24a-hNIP-c9-tdpChi-squared model

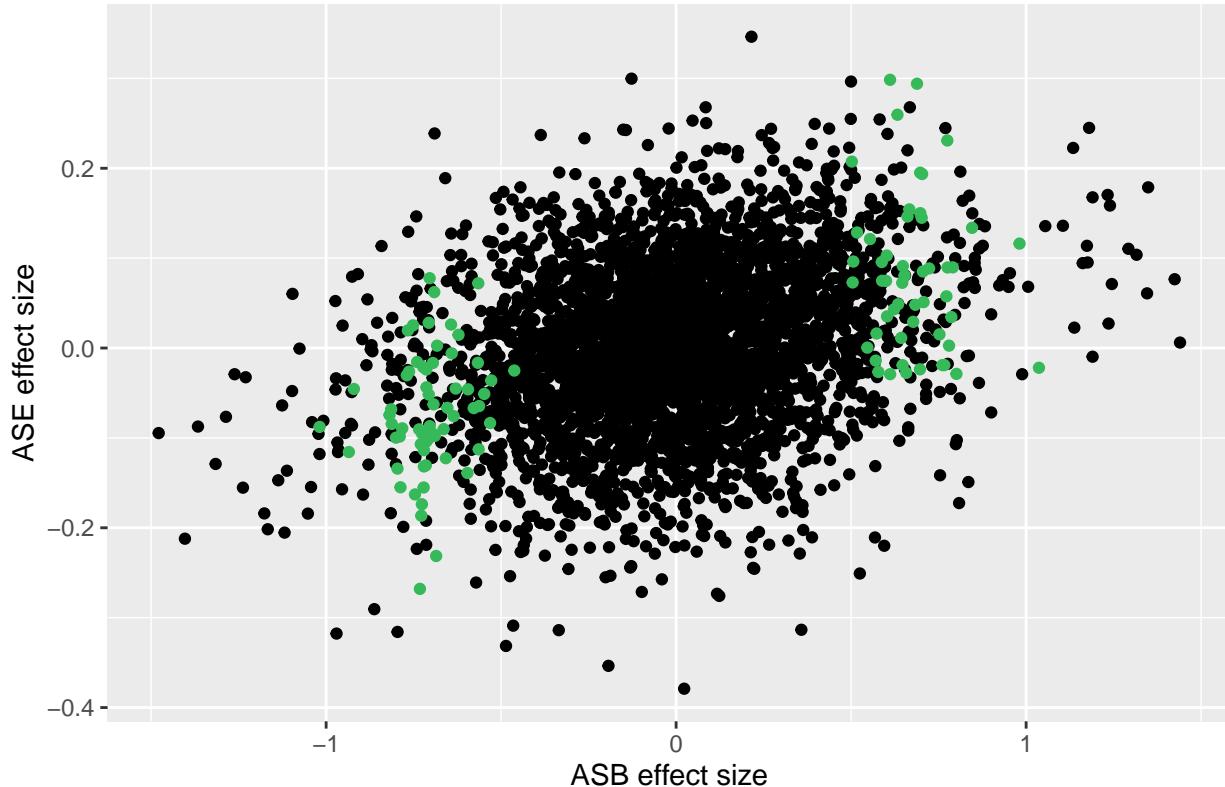


```
## $xpd  
## [1] FALSE
```

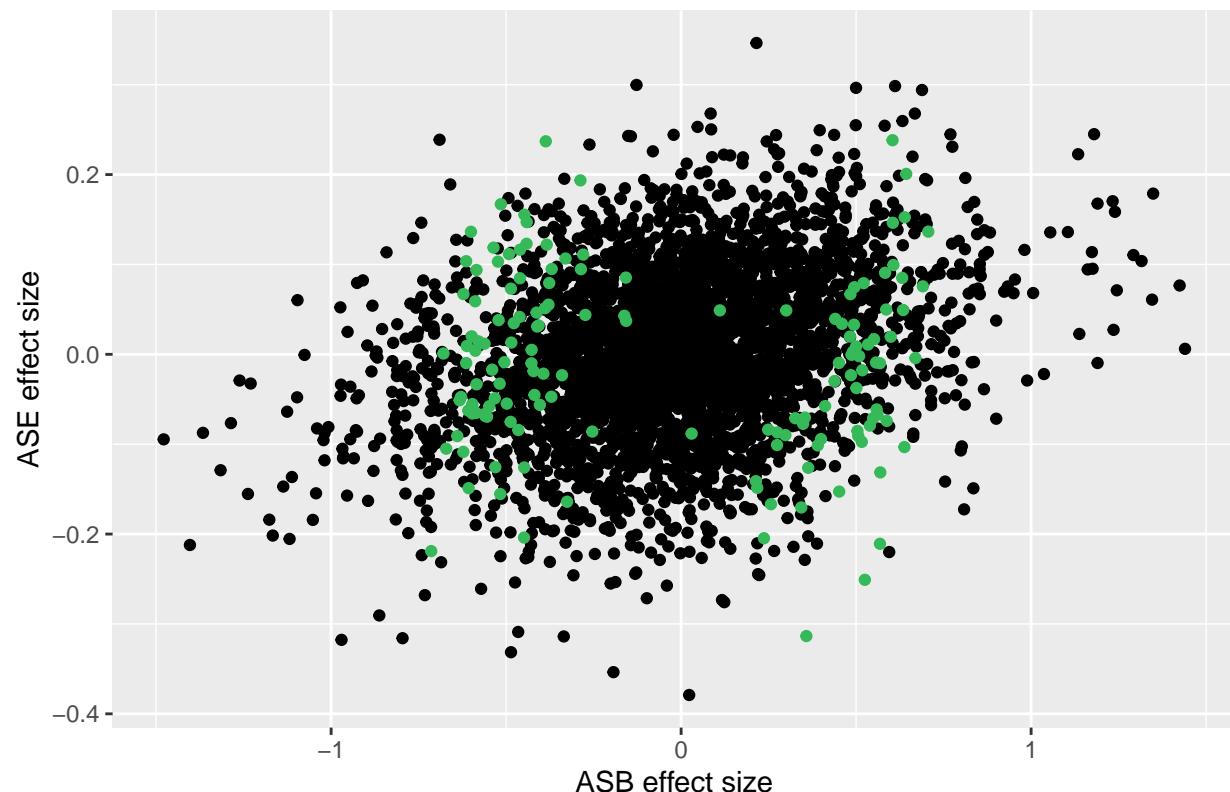
Further investigation into where significant hits between two models vary

```
for (j in 1:length(bqtls_betas)) {  
  bqtls_beta <- bqtls_betas[[j]]  
  chisq <- all_chisq[[j]]  
  qthresh = 0.05  
  in_chi_sq <- chisq$q.value < qthresh  
  in_beta <- bqtls_beta$asb_q < qthresh  
  idx <- !in_chi_sq & in_beta  
  p1 = ggplot(bqtls_beta, aes(x = asb_loc, y = ase_loc, text = variantID)) +  
    geom_point(data=bqtls_beta[!idx,], color='black') +  
    geom_point(data=bqtls_beta[idx,], color='#36ba59') +  
    labs(x = "ASB effect size", y = "ASE effect size", title = str_glue("{sum(idx)} SNPs with q < {qthresh}"))  
  idx <- in_chi_sq & !in_beta  
  p2 = ggplot(bqtls_beta, aes(x = asb_loc, y = ase_loc, text = variantID)) +  
    geom_point(data=bqtls_beta[!idx,], color='black') +  
    geom_point(data=bqtls_beta[idx,], color='#36ba59') +  
    labs(x = "ASB effect size", y = "ASE effect size", title = str_glue("{sum(idx)} SNPs with q < {qthresh}"))  
  print(p1)  
  print(p2)  
}
```

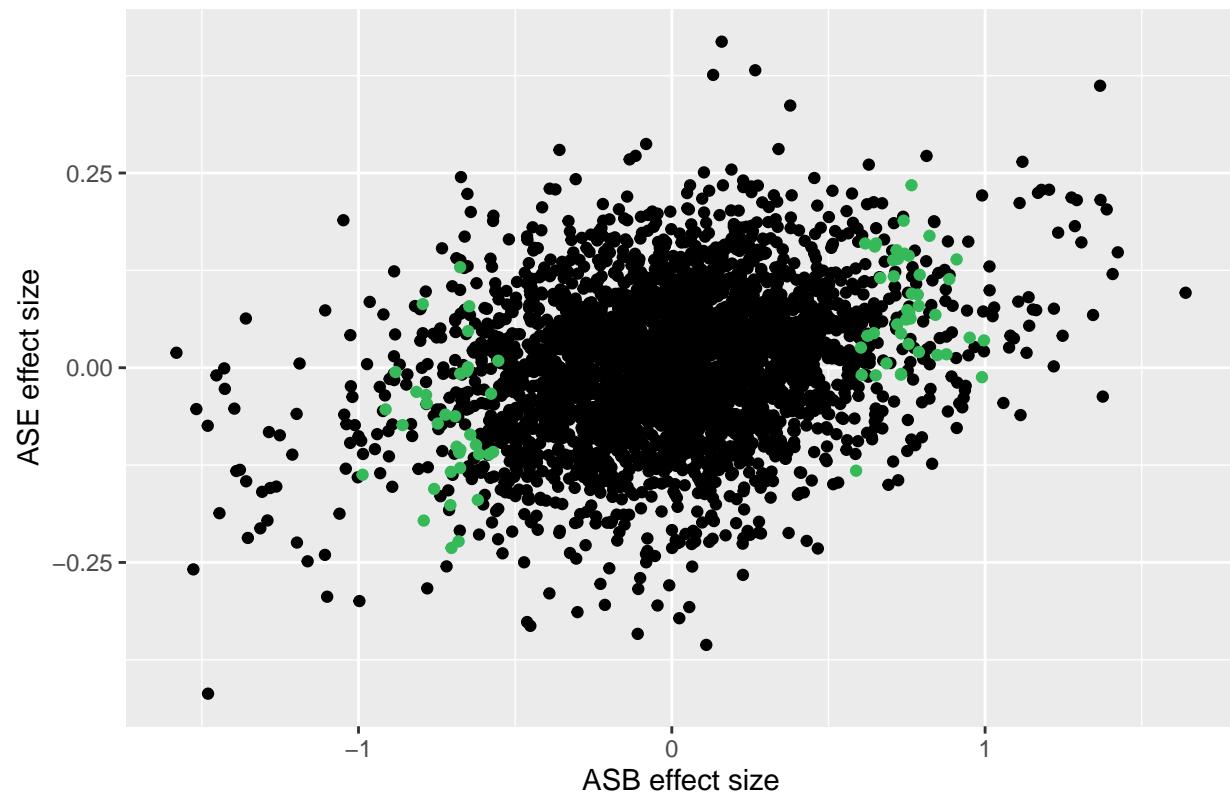
113 SNPs with $q < 0.05$ in beta but not in chi-sq model $q < 0.05$



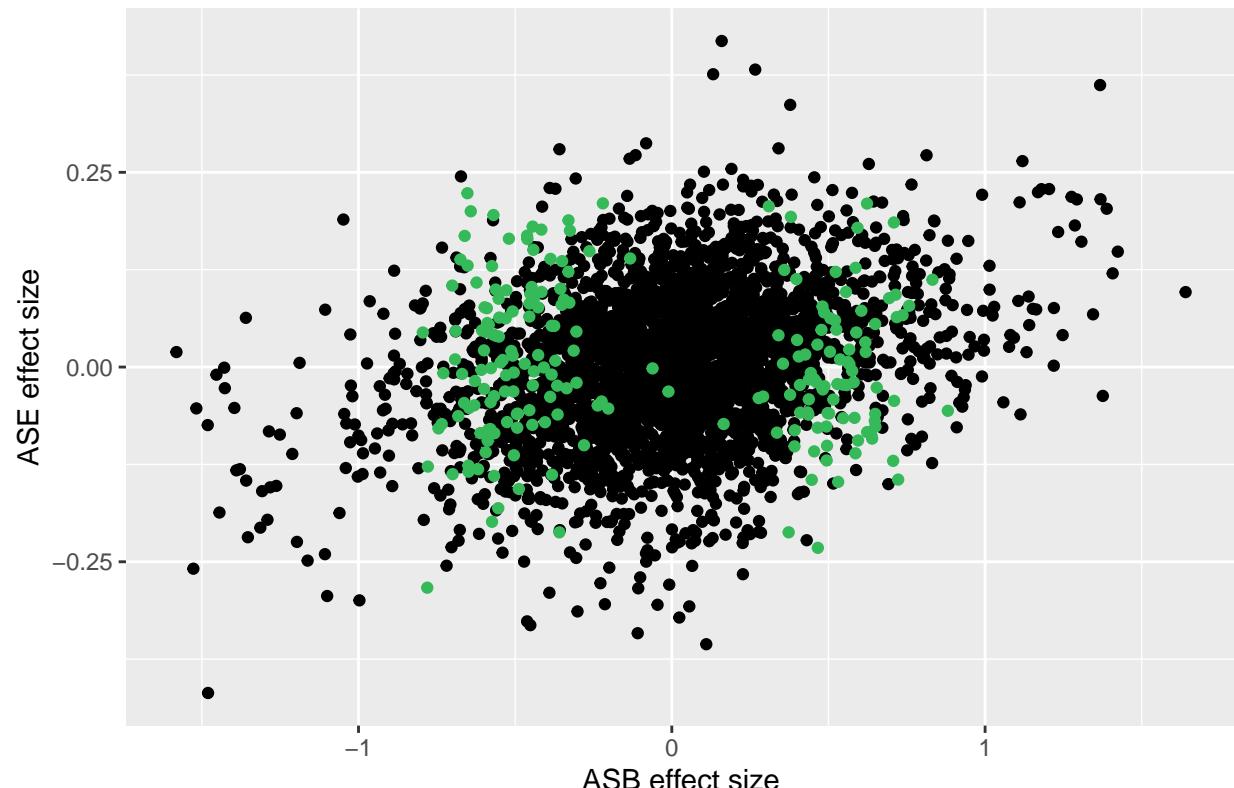
146 SNPs with $q < 0.05$ in chi squared model but not in beta model $q < 0.0$



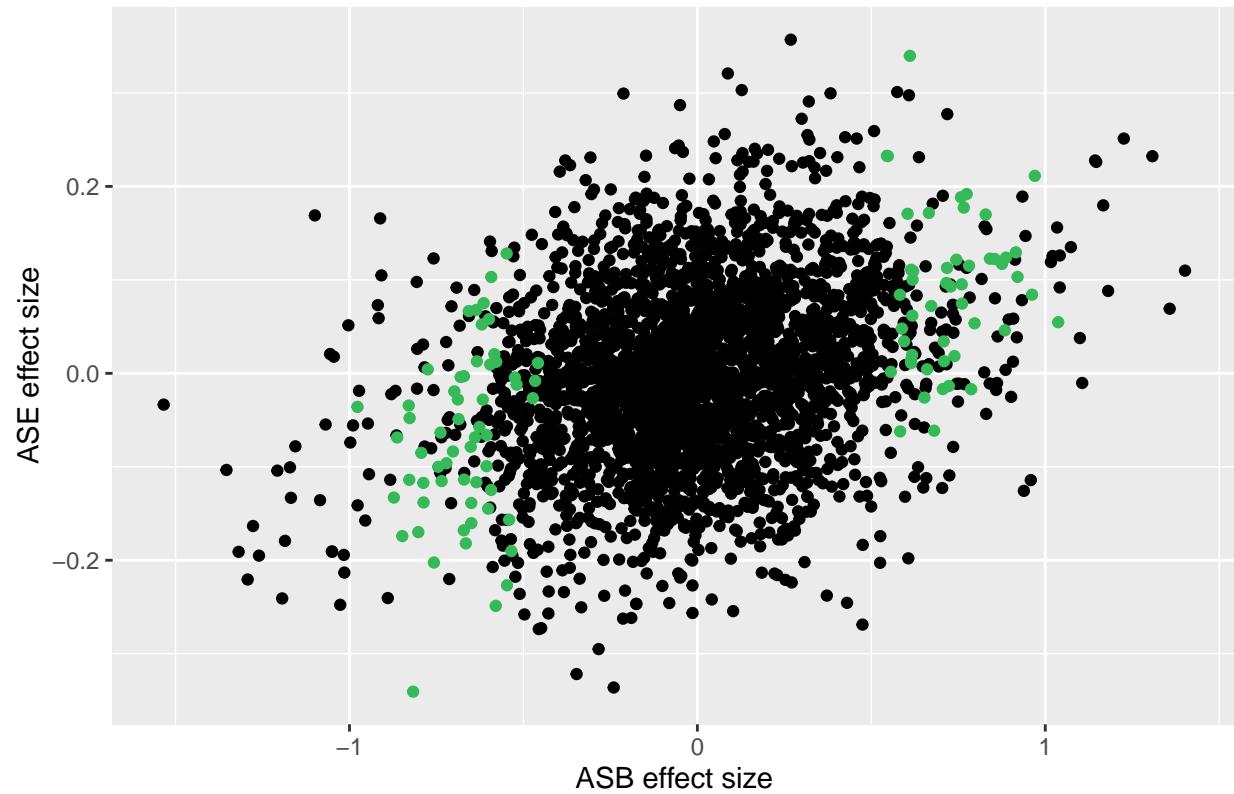
80 SNPs with $q < 0.05$ in beta but not in chi-sq model $q < 0.05$



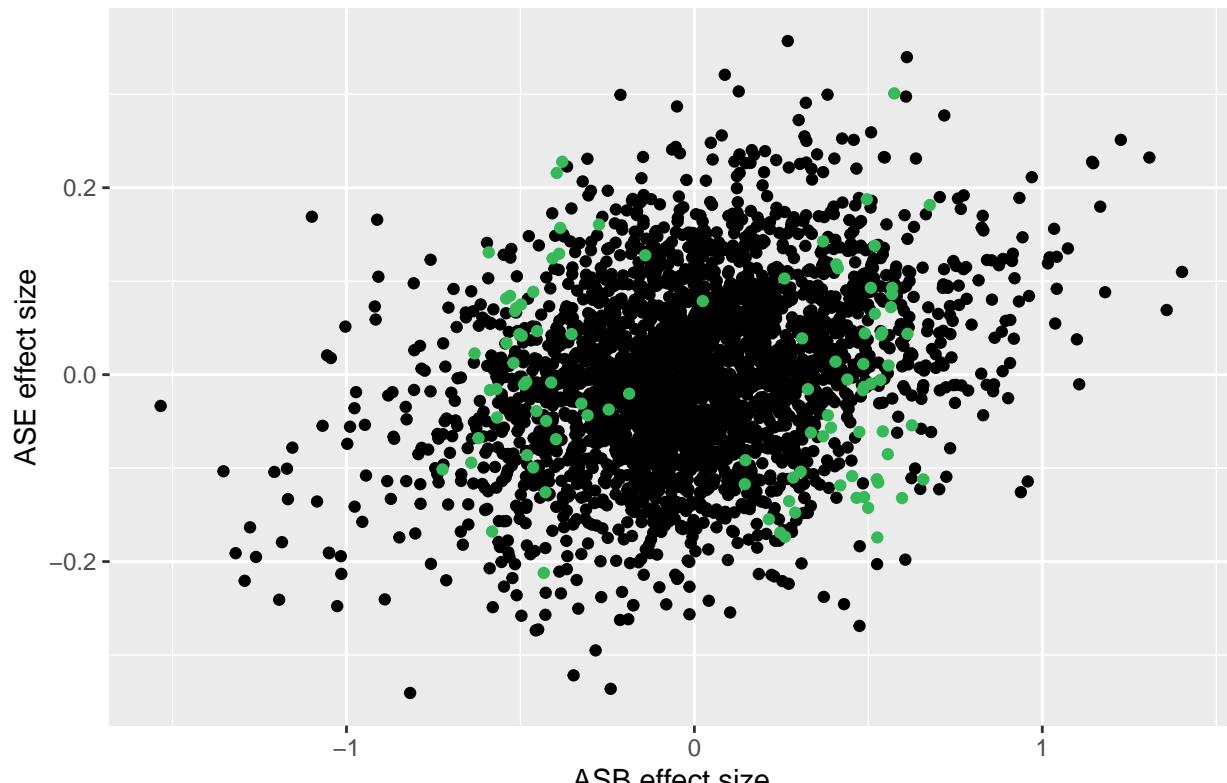
213 SNPs with $q < 0.05$ in chi squared model but not in beta model $q < 0.05$



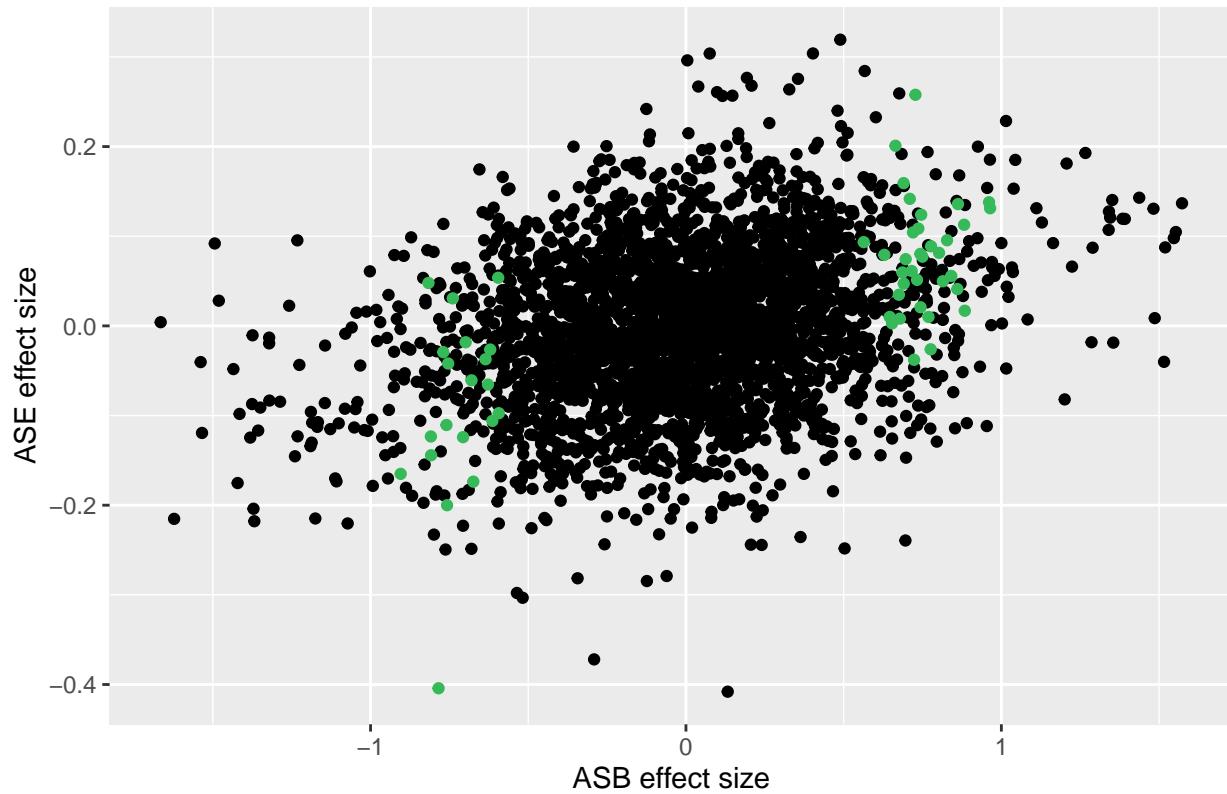
107 SNPs with $q < 0.05$ in beta but not in chi-sq model $q < 0.05$



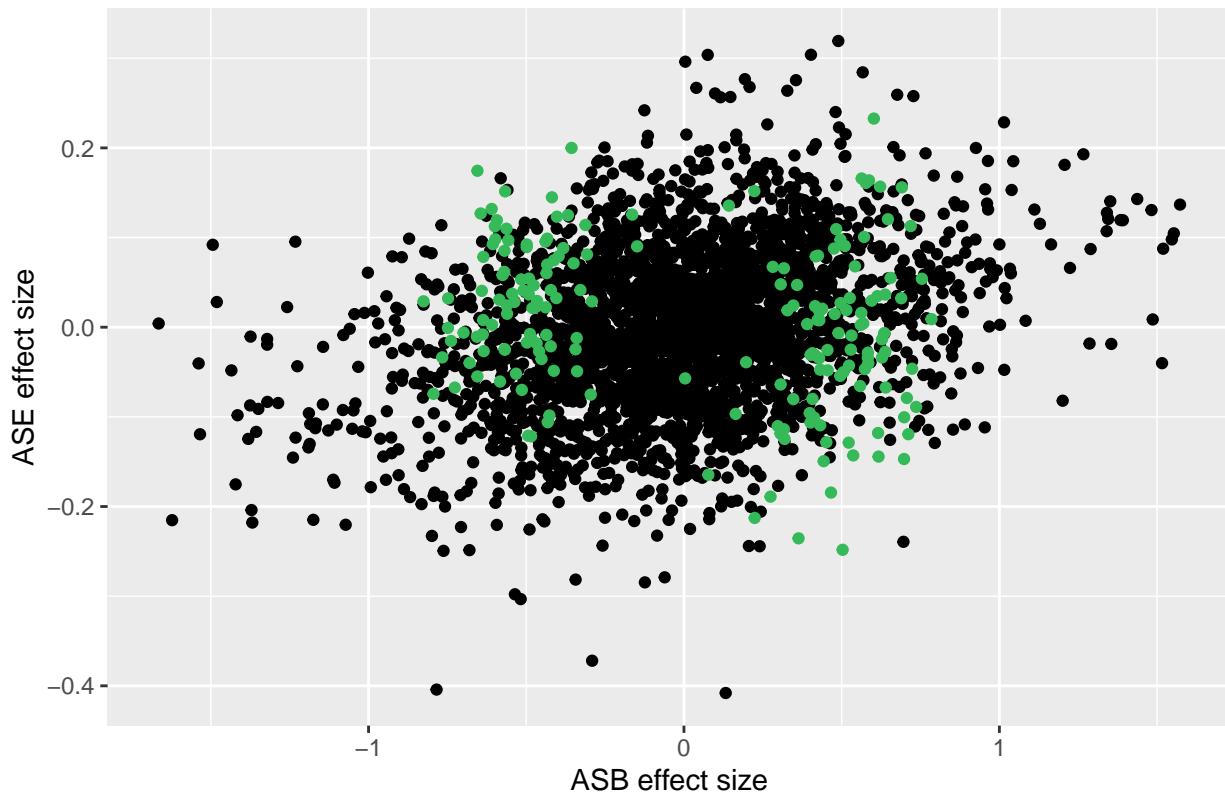
97 SNPs with $q < 0.05$ in chi squared model but not in beta model $q < 0.05$



55 SNPs with $q < 0.05$ in beta but not in chi-sq model $q < 0.05$



188 SNPs with $q < 0.05$ in chi squared model but not in beta model $q < 0.0$



```

for (j in 1:length(bqtls_betas)) {
  bqtls_beta <- bqtls_betas[[j]]
  chisq <- all_chisq[[j]]
  in_chi_sq <- chisq$q.value < 0.05
  in_beta <- bqtls_beta$asb_q < 0.05
  idx <- !in_chi_sq & in_beta
  cat(rbps[j], '\n')
  cat(sum(idx), ' in beta and not in chi-squared\n')
  idx <- in_chi_sq & !in_beta
  cat(sum(idx), ' in chi-squared and not in beta\n')
  cat("\n-----\n")
}

## 24a-hNIL-control-tdp
## 113  in beta and not in chi-squared
## 146  in chi-squared and not in beta
##
## -----
## 24a-hNIL-c9-tdp
## 80   in beta and not in chi-squared
## 213  in chi-squared and not in beta
##
## -----
## 24a-hNIP-control-tdp
## 107  in beta and not in chi-squared
## 97   in chi-squared and not in beta
##

```

```

## -----
## 24a-hNIP-c9-tdp
## 55  in beta and not in chi-squared
## 188  in chi-squared and not in beta
##
## -----

```

In 24a-hNIP-c9-tdp, rs33943686 looks like it's right before a junction and upstream of a peak. This was found by chi-squared but not by the beta model.

In 24a-hNIP-c9-tdp, looks like it's right before a junction and upstream of a peak

High-effect-size hits that align between two models:

```

for (j in 1:length(bqtls_betas)) {
  bqtls_beta <- bqtls_betas[[j]]
  chisq <- all_chisq[[j]]
  in_chi_sq <- chisq$q.value < 0.01
  in_beta <- bqtls_beta$asb_q < 0.01
  idx <- in_chi_sq & in_beta & abs(bqtls_beta$asb_loc) > 1.3
  cat(rbps[j], '\n')
  cat(paste(bqtls_beta$variantID[idx], collapse = "\n"))
  cat("\n-----\n")
}

## 24a-hNIL-control-tdp
## rs12759741
## rs13008829
## rs596904
## rs4150099
## rs508468
## rs1894603
## rs1343706
## rs911783
## rs4822790
##
## -----
## 24a-hNIL-c9-tdp
## rs1536115
## rs1061337
## rs7570707
## rs10177491
## rs11683792
## rs4688233
## rs2937667
## rs10937003
## rs12629148
## rs28729471
## rs11748087
## rs7720760
## rs6902058
## rs2068626
## rs36177855
## rs36183797
## rs13251050
## rs10082462

```

```
## rs6578918
## rs490507
## rs6588940
## rs2708333
## rs36036395
## rs8077024
## rs6035850
## -----
## 24a-hNIP-control-tdp
## rs13146448
## rs10275799
## rs3779639
## rs8929
## rs7324866
## rs399535
## -----
## 24a-hNIP-c9-tdp
## rs3008620
## rs1053316
## rs333234
## rs1992898
## rs2230534
## rs11719486
## rs5004095
## rs3804772
## rs27758
## rs28010
## rs831640
## rs12514851
## rs62385377
## rs10058
## rs9374
## rs28927678
## rs1061731
## rs7845483
## rs17741842
## rs2771040
## rs2488319
## rs2708361
## rs11177879
## rs73374491
## rs34303822
## rs12926574
## rs9962322
## rs11557092
## rs10419448
## rs10485816
## rs73905782
## rs11701571
## rs9625874
## -----
```

From 24a-hNIL-control-tdp rs56264956, found by beta model and not by chi-squared model is somewhat near a peak

```
bqtls_betas[[1]]%>%filter(variantID == "rs56264956")  
  
## # A tibble: 1 x 28  
##   chrom position variantID  refAllele altAllele refCount_input altCount_input  
##   <chr>     <dbl> <chr>      <chr>      <chr>          <dbl>          <dbl>  
## 1 chr18  52461733 rs56264956 C         T             38             21  
## # i 21 more variables: totalCount_input <dbl>, pred_ratio <dbl>,  
## #   refCount_IP <dbl>, altCount_IP <dbl>, totalCount_IP <dbl>,  
## #   shrunk_input_logratio <dbl>, ase_loc <dbl>, ase_sd <dbl>, ase_q <dbl>,  
## #   shrunk_IP_logratio <dbl>, asb_loc <dbl>, asb_sd <dbl>, asb_q <dbl>,  
## #   in_peak_pos <dbl>, in_peak_neg <dbl>, in_peak <dbl>, near_peak_100k <dbl>,  
## #   in_exon <dbl>, in_transcript <dbl>, in_gene <dbl>, in_utr <dbl>
```

How many SNPs are in peaks?

Look at overlap of these between the models and between the samples

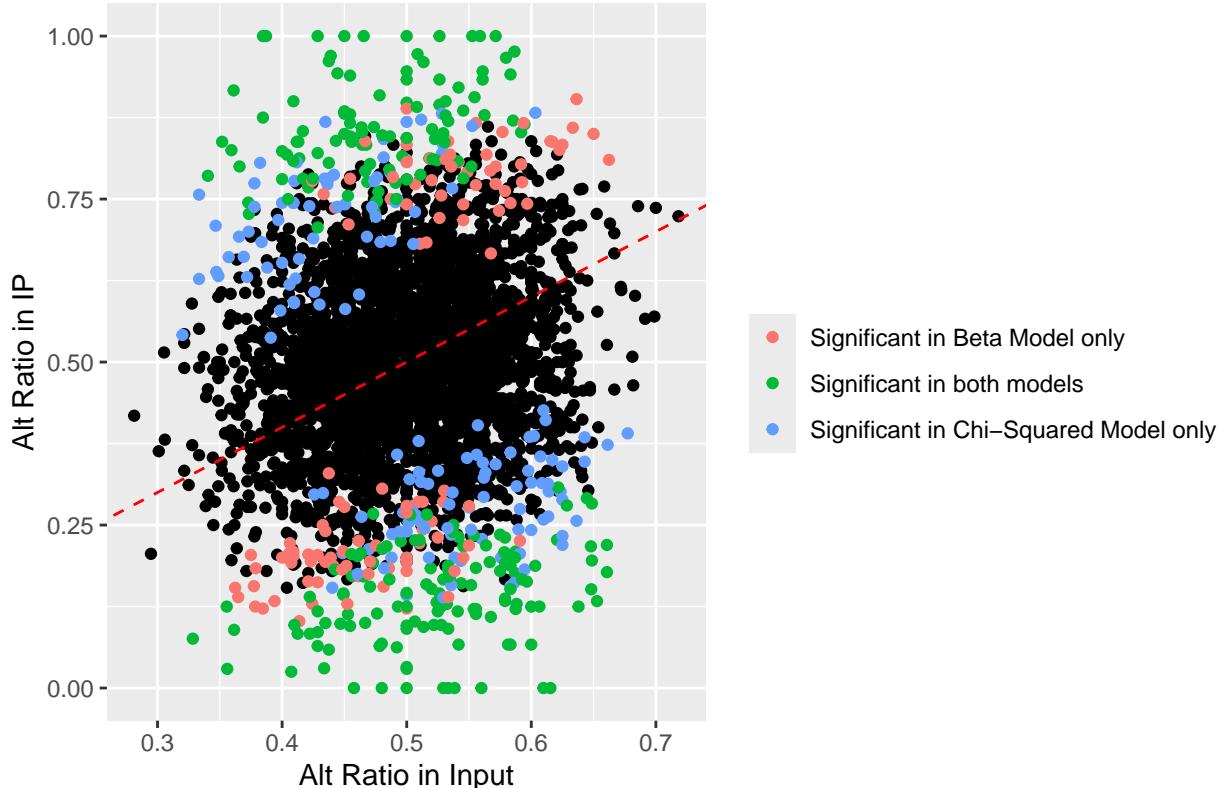
```
for (j in 1:length(bqtls_betas)) {  
  bqtls_beta <- bqtls_betas[[j]]  
  chisq <- all_chisq[[j]]  
  in_chi_sq <- chisq$q.value < 0.01  
  in_beta <- bqtls_beta$asb_q < 0.01  
  cat(rbps[j], '\n')  
  idx <- in_chi_sq & !in_beta & bqtls_beta$in_peak  
  cat(sum(idx), ' in peaks chi sq only\n')  
  idx <- in_beta & !in_chi_sq & bqtls_beta$in_peak  
  cat(sum(idx), ' in peaks beta only\n')  
  idx <- in_beta & in_chi_sq & bqtls_beta$in_peak  
  cat(sum(idx), ' in peaks both\n')  
  cat("\n-----\n")  
}  
  
## 24a-hNIL-control-tdp  
## 8  in peaks chi sq only  
## 1  in peaks beta only  
## 2  in peaks both  
##  
## -----  
## 24a-hNIL-c9-tdp  
## 17  in peaks chi sq only  
## 1  in peaks beta only  
## 1  in peaks both  
##  
## -----  
## 24a-hNIP-control-tdp  
## 5  in peaks chi sq only  
## 0  in peaks beta only  
## 0  in peaks both  
##  
## -----  
## 24a-hNIP-c9-tdp
```

```

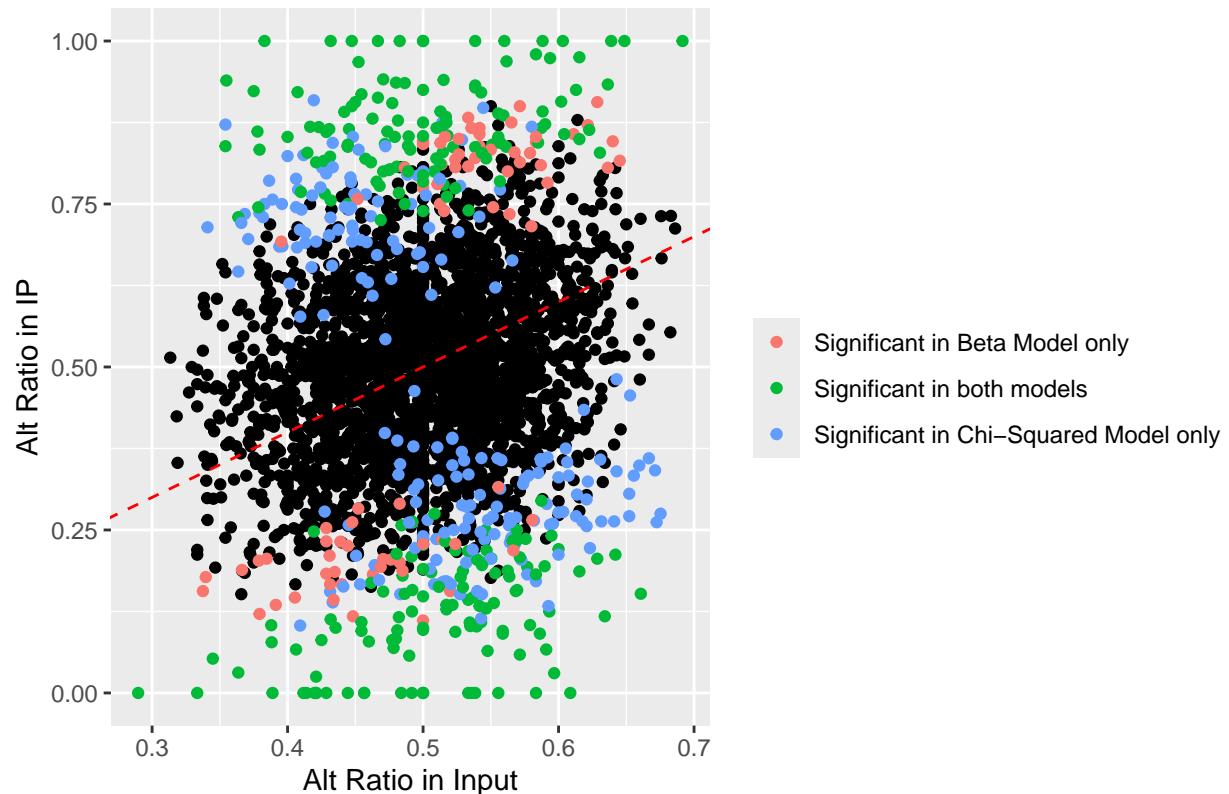
## 8 in peaks chi sq only
## 0 in peaks beta only
## 2 in peaks both
##
## -----
for (j in 1:length(bqtls_betas)) {
  bqtls_beta <- bqtls_betas[[j]]
  chisq <- all_chisq[[j]]
  in_chi_sq <- chisq$q.value < 0.05
  in_beta <- bqtls_beta$asb_q < 0.05
  group <- rep('', nrow(bqtls_beta))
  group[!in_chi_sq & !in_beta] <- "Not significant"
  group[in_chi_sq & !in_beta] <- "Significant in Chi-Squared Model only"
  group[!in_chi_sq & in_beta] <- "Significant in Beta Model only"
  group[in_chi_sq & in_beta] <- "Significant in both models"
  # Plotting input ratio vs IP ratio and seeing if it's called a bqtl by one model, another, or both
  p <- ggplot(bqtls_beta[!(in_chi_sq & in_beta),], aes(x=altCount_input/totalCount_input, y = altCount_ip/totalCount_ip))
    geom_point(aes(color = 'black'), data = bqtls_beta[group == "Not significant",], color="black")+
    geom_point()+
    #lims(x = c(0,1), y = c(0,1))+ 
  geom_abline(slope = 1, intercept = 0, col="red", lty="dashed")+
  labs(x = "Alt Ratio in Input", y = "Alt Ratio in IP", color = "", title = paste0(rbps[j], " | Significant"))
  print(p)
}

```

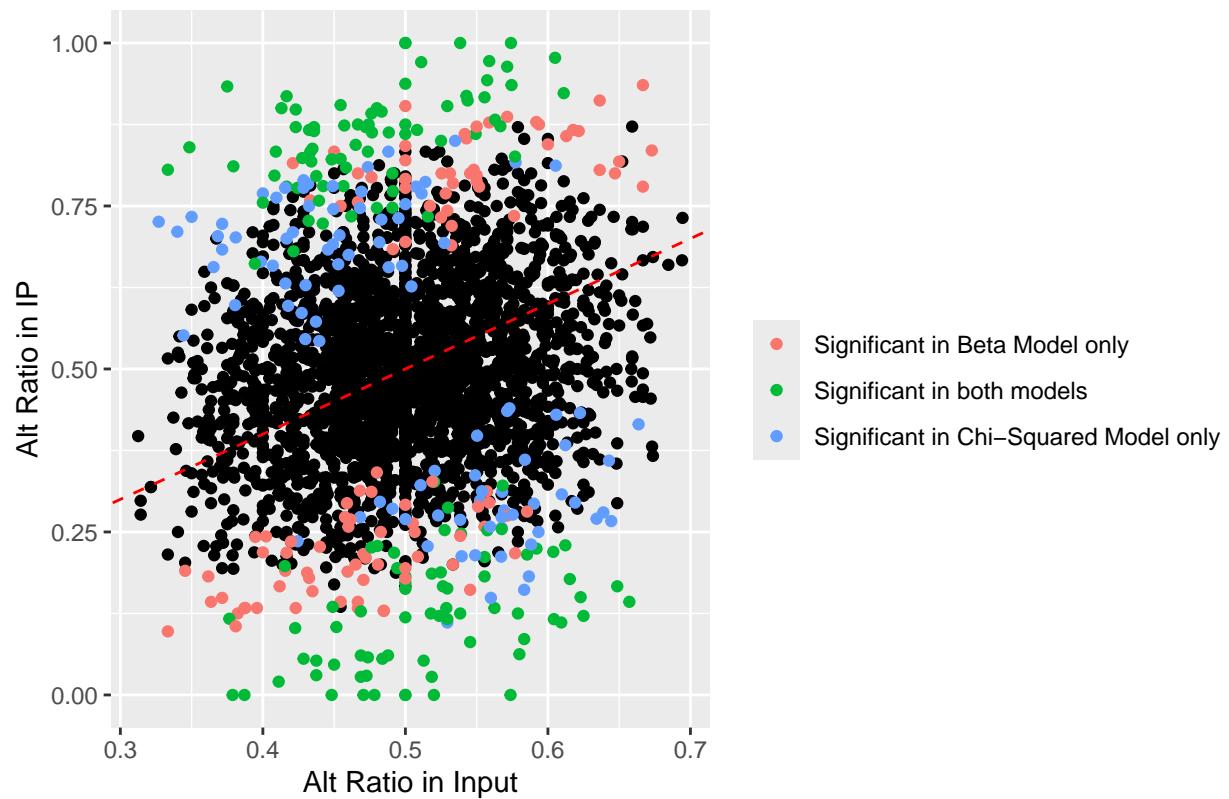
24a-hNIL-control-tdp | Significant hits of two models both at $q < .05$



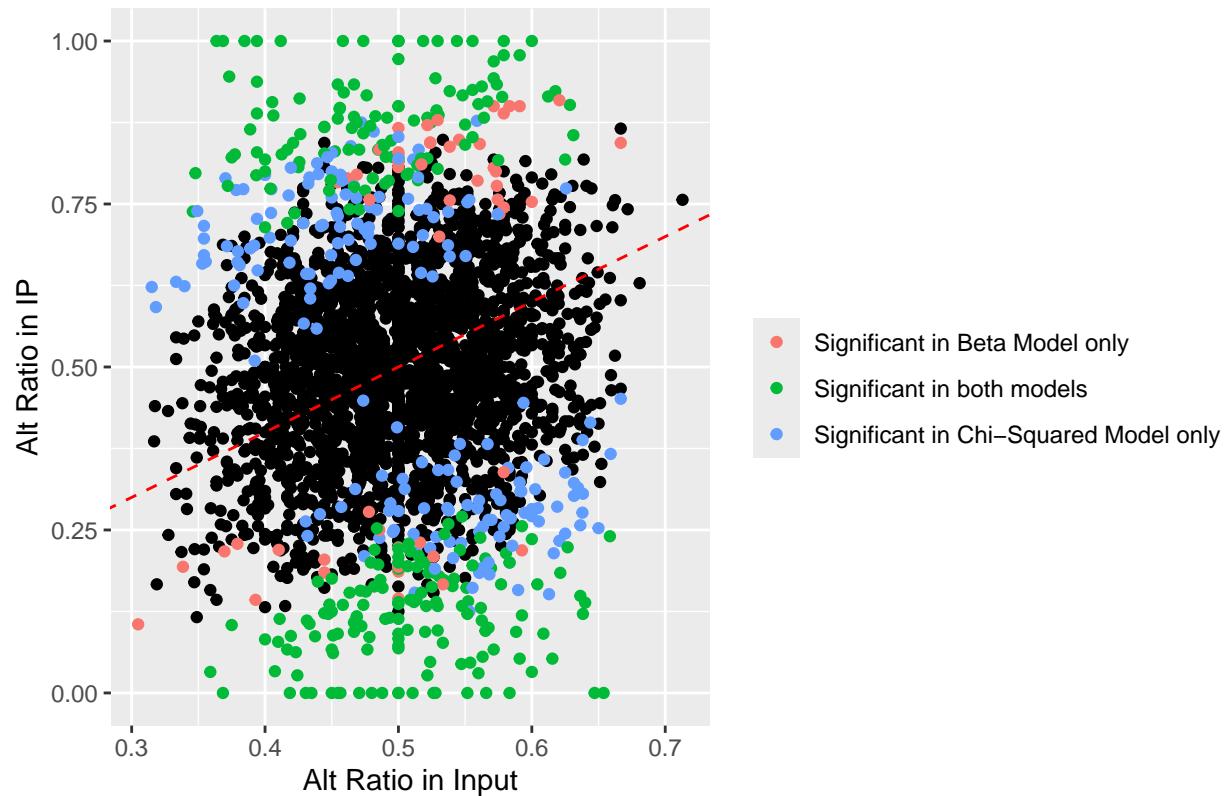
24a–hNIL–c9–tdp | Significant hits of two models both at $q < .05$



24a–hNIP–control–tdp | Significant hits of two models both at $q < .05$



24a-hNIP-c9-tdp | Significant hits of two models both at $q < .05$



Looking at UNC13A locus (a particular locus interesting to ALS biology)

```
# UNC13A
for (j in 1:length(bqtls_betas)) {
  bqtls_beta <- bqtls_betas[[j]]
  chisq <- all_chisq[[j]]
  in_chi_sq <- chisq$q.value < 0.05
  in_beta <- bqtls_beta$asb_q < 0.05
  in_unc13a = with(bqtls_beta, (position >= 17601336) & (position <= 17688365))
  cat(rbps[j], '\n')
  cat(sum(in_unc13a), "SNPs in UNC13A made it past filtering\n")
  cat(sum(in_unc13a & in_chi_sq), "SNPs in UNC13A significant by chi-sq model\n")
  cat(sum(in_unc13a & in_beta), "SNPs in UNC13A significant by beta model\n")
  cat("\n-----\n")
}

## 24a-hNIL-control-tdp
## 2 SNPs in UNC13A made it past filtering
## 0 SNPs in UNC13A significant by chi-sq model
## 0 SNPs in UNC13A significant by beta model
##
## -----
## 24a-hNIL-c9-tdp
## 3 SNPs in UNC13A made it past filtering
## 0 SNPs in UNC13A significant by chi-sq model
## 0 SNPs in UNC13A significant by beta model
##
```

```

## -----
## 24a-hNIP-control-tdp
## 3 SNPs in UNC13A made it past filtering
## 0 SNPs in UNC13A significant by chi-sq model
## 0 SNPs in UNC13A significant by beta model
##
## -----
## 24a-hNIP-c9-tdp
## 3 SNPs in UNC13A made it past filtering
## 0 SNPs in UNC13A significant by chi-sq model
## 0 SNPs in UNC13A significant by beta model
##
## -----

```

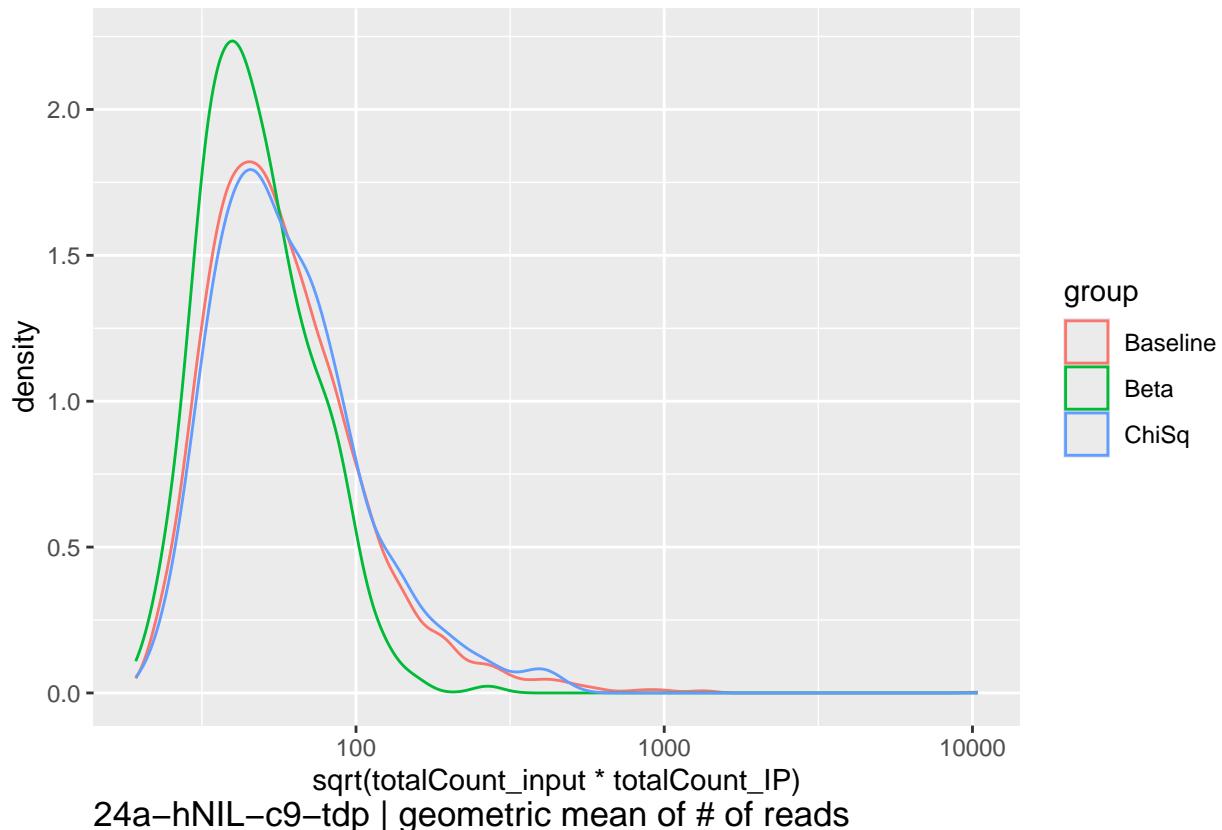
Looking at read counts for significant hits – replacing x axis with geometric mean of input and IP totalCount

```

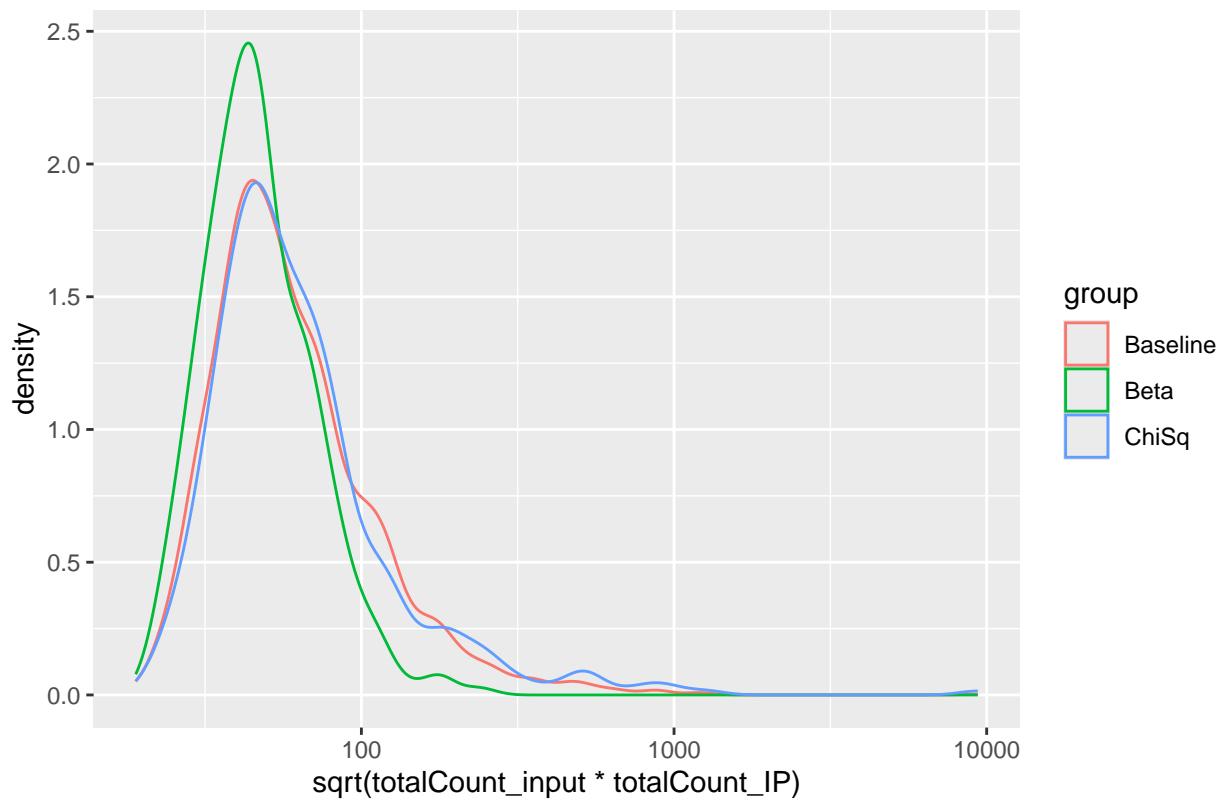
for (j in 1:length(bqtls_betas)) {
  bqtls_beta <- bqtls_betas[[j]]
  chisq <- all_chisq[[j]]
  in_chi_sq <- chisq$q.value < 0.05
  in_beta <- bqtls_beta$asb_q < 0.05
  x=bqtls_beta$totalCount_input
  y=bqtls_beta$totalCount_IP
  neither <-
  #dat <- data.table(totalCount_input = c(x[in_chi_sq], x[in_beta], x[!in_beta & !in_chi_sq]),
  #                   totalCount_IP = c(y[in_chi_sq], y[in_beta], y[!in_beta & !in_chi_sq]),
  #                   group = rep(c("ChiSq", "Beta", "Neither"), c(sum(in_chi_sq), sum(in_beta), sum(!in_beta &
  dat <- data.frame(totalCount_input = c(x[in_chi_sq], x[in_beta], x),
                     totalCount_IP = c(y[in_chi_sq], y[in_beta], y),
                     group = rep(c("ChiSq", "Beta", "Baseline"), c(sum(in_chi_sq), sum(in_beta), nrow(bqtls_beta))
  p1 <- ggplot(dat,aes(x = totalCount_input, color=group))+ 
    geom_density()+
    scale_x_log10()+
    labs(title = paste(rbps[j],"| # of reads input"))
  p2 <- ggplot(dat,aes(x = totalCount_IP, color=group))+ 
    geom_density()+
    scale_x_log10()+
    labs(title = paste(rbps[j],"| # of reads IP"))
  p3 <- ggplot(dat,aes(x = sqrt(totalCount_input*totalCount_IP), color=group))+ 
    geom_density()+
    scale_x_log10()+
    labs(title = paste(rbps[j],"| geometric mean of # of reads"))
  #print(p1)
  print(p3)
}

```

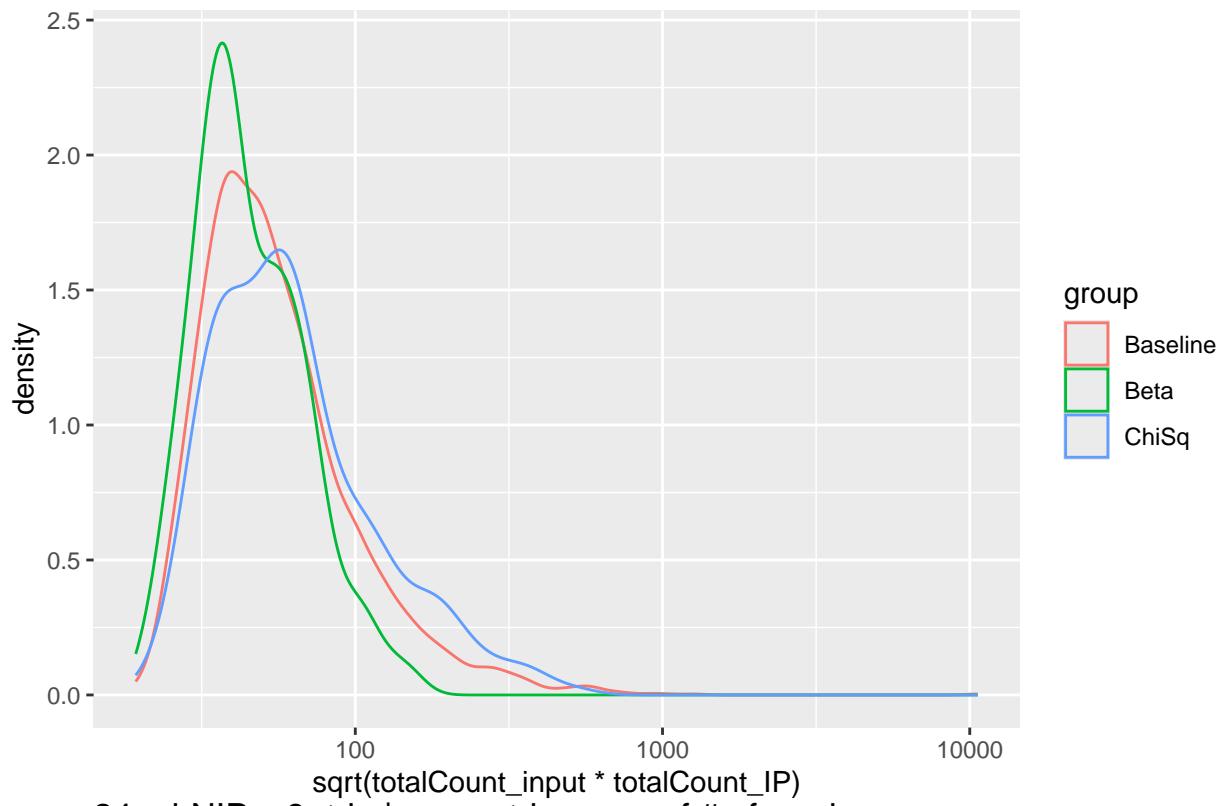
24a-hNIL-control-tdp | geometric mean of # of reads



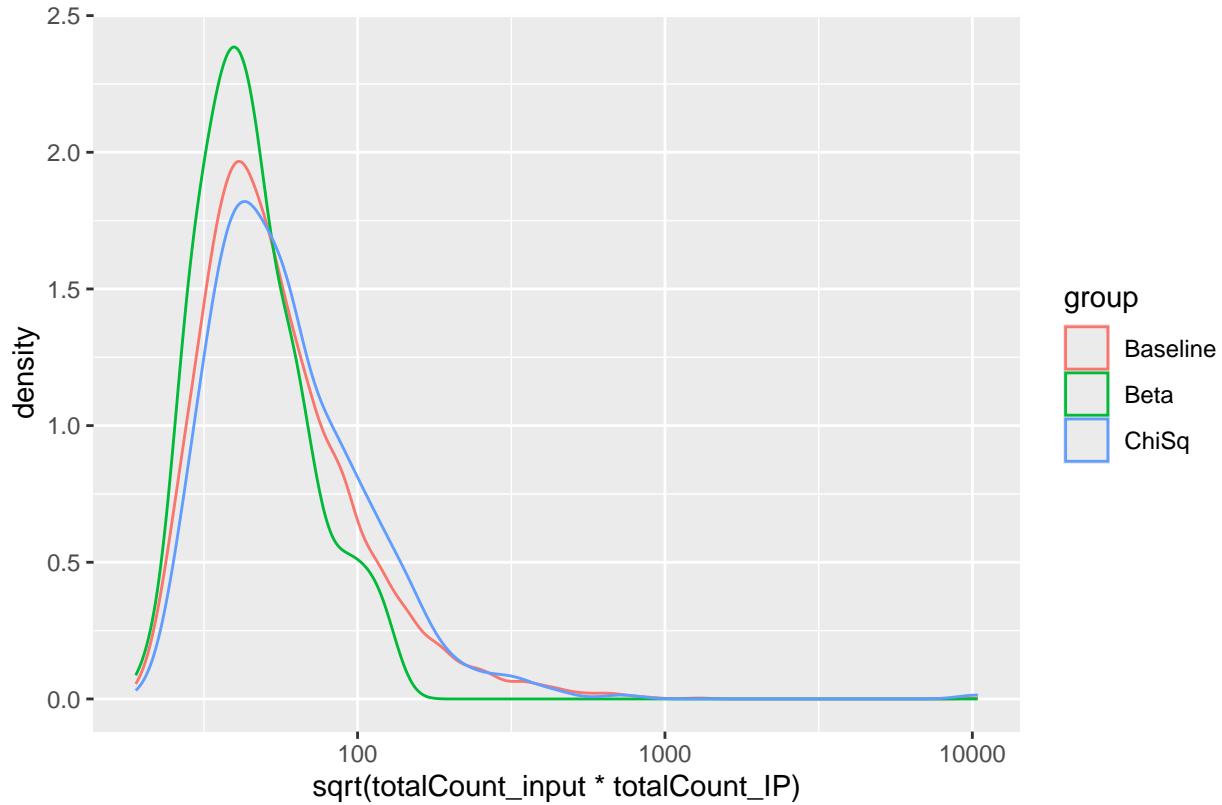
24a-hNIL-c9-tdp | geometric mean of # of reads



24a–hNIP–control–tdp | geometric mean of # of reads



24a–hNIP–c9–tdp | geometric mean of # of reads



```

lapply(1:length(bqtls_betas), function(j) {
  bqtls_beta <- bqtls_betas[[j]]
  chisq <- all_chisq[[j]]
  in_chi_sq <- chisq$q.value < 0.05
  in_beta <- bqtls_beta$asb_q < 0.05
  c(sum(in_chi_sq), sum(in_beta), sum(!in_beta & !in_chi_sq))
}) -> res
x = as.data.frame(Reduce(res, f=rbind))
names(x) = c("ChiSq", "Beta", "Neither")
x

##      ChiSq Beta Neither
## init    378  345   3297
## X       449  316   2695
## X.1     240  250   2513
## X.2     448  315   2508

```

Top hits for beta model

```

lapply(1:length(bqtls_betas), function(j) {
  bqtls_beta <- bqtls_betas[[j]]
  chisq <- all_chisq[[j]]
  in_chi_sq <- chisq$q.value < 0.05
  in_beta <- bqtls_beta$asb_q < 0.05
  bqtls_beta[in_beta,]
}) -> res

res[[1]] %>% arrange(asb_q)

## # A tibble: 345 x 28
##   chrom position variantID refAllele altAllele refCount_input altCount_input
##   <chr>   <dbl> <chr>      <chr>      <chr>      <dbl>      <dbl>
## 1 chr6    122441902 rs508468 T         G             17          17
## 2 chr9    14610593  rs1343706 T         C             12          16
## 3 chr2    79716851  rs13008829 G         A             11          14
## 4 chr10   14052489  rs10906580 A         G              8          10
## 5 chr2    80294216  rs2862286 C         T             38          31
## 6 chr4    77166953  rs4150099 G         T             32          27
## 7 chr18   53532971  rs2292044 C         G             27          21
## 8 chr5    40831825  rs11748087 A         G             20          16
## 9 chr2    79973291  rs472725  G         A             29          40
## 10 chr21   17791390  rs7277763 C         T             16          25
## # i 335 more rows
## # i 21 more variables: totalCount_input <dbl>, pred_ratio <dbl>,
## #   refCount_IP <dbl>, altCount_IP <dbl>, totalCount_IP <dbl>,
## #   shrunk_input_logratio <dbl>, ase_loc <dbl>, ase_sd <dbl>, ase_q <dbl>,
## #   shrunk_IP_logratio <dbl>, asb_loc <dbl>, asb_sd <dbl>, asb_q <dbl>,
## #   in_peak_pos <dbl>, in_peak_neg <dbl>, in_peak <dbl>, near_peak_100k <dbl>,
## #   in_exon <dbl>, in_transcript <dbl>, in_gene <dbl>, in_utr <dbl>

```

Top hits for chi-squared model

```

lapply(1:length(bqtl_beta), function(j) {
  bqtl_beta <- bqtl_beta[[j]]
  chisq <- all_chisq[[j]]
  in_chi_sq <- chisq$q.value < 0.05
  in_beta <- bqtl_beta$asb_q < 0.05
  cbind(bqtl_beta, chisq)[in_chi_sq,] %>% as_tibble
}) -> res

res[[1]] %>% arrange(q.value)

## # A tibble: 378 x 33
##   chrom position variantID refAllele altAllele refCount_input altCount_input
##   <chr>    <dbl> <chr>      <chr>      <chr>      <dbl>      <dbl>
## 1 chr2     15607269 rs10929378 C          T          61         86
## 2 chr18    53532971 rs2292044 C          G          27         21
## 3 chr10    114436017 rs1057139 C          G          85         86
## 4 chr10    122247593 rs911783 A          T          24         15
## 5 chr12    65964567 rs1042725 C          T          77         86
## 6 chr21    17791390 rs7277763 C          T          16         25
## 7 chr3     116705036 rs1518335 T          G          15         24
## 8 chr3     116323558 rs1920191 A          G          32         42
## 9 chr5     36103803 rs17343598 G          A          19         12
## 10 chr15    41305239 rs1132639 T          A          19         37
## # i 368 more rows
## # i 26 more variables: totalCount_input <dbl>, pred_ratio <dbl>,
## #   refCount_IP <dbl>, altCount_IP <dbl>, totalCount_IP <dbl>,
## #   shrunk_input_logratio <dbl>, ase_loc <dbl>, ase_sd <dbl>, ase_q <dbl>,
## #   shrunk_IP_logratio <dbl>, asb_loc <dbl>, asb_sd <dbl>, asb_q <dbl>,
## #   in_peak_pos <dbl>, in_peak_neg <dbl>, in_peak <dbl>, near_peak_100k <dbl>,
## #   in_exon <dbl>, in_transcript <dbl>, in_gene <dbl>, in_utr <dbl>, ...

lapply(1:length(bqtl_beta), function(j) {
  bqtl_beta <- bqtl_beta[[j]]
  chisq <- all_chisq[[j]]
  in_chi_sq <- chisq$q.value < 0.05
  in_beta <- bqtl_beta$asb_q < 0.05
  c(rbps[j],
    sum(in_chi_sq),
    sum(in_beta),
    nrow(bqtl_beta),
    #sum(!in_beta & !in_chi_sq),
    sum(in_chi_sq & bqtl_beta$in_peak),
    sum(in_beta & bqtl_beta$in_peak),
    #sum(!in_beta & !in_chi_sq & bqtl_beta$in_peak),
    sum(bqtl_beta$in_peak),
    sum(in_chi_sq & bqtl_beta$near_peak_100k),
    sum(in_beta & bqtl_beta$near_peak_100k),
    sum(bqtl_beta$near_peak_100k)
  )
}) -> res

x = as.data.frame(Reduce(res, f=rbind))
names(x) = c('rbp', "ChiSq", "Beta", "Everything", "ChiSq In Peak", "Beta In Peak", "Everything In Peak")
x
##                                rbp ChiSq Beta Everything ChiSq In Peak Beta In Peak Everything In Peak

```

```

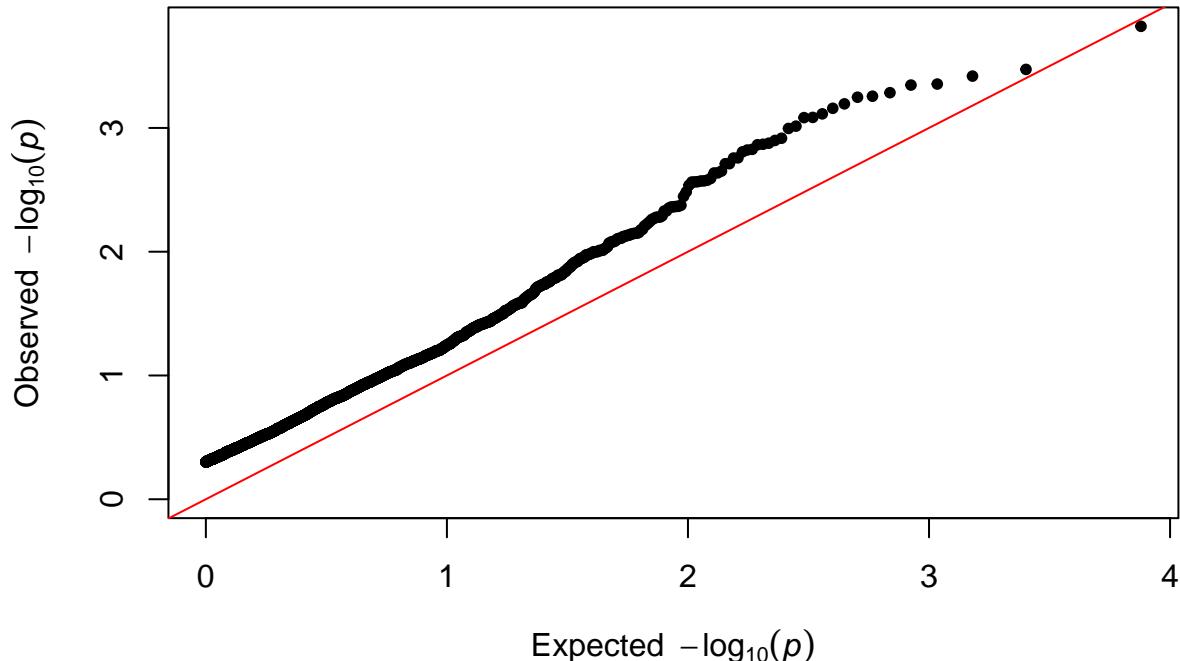
## init 24a-hNIL-control-tdp    378   345      3788      20       10
## X      24a-hNIL-c9-tdp     449   316      3224      33       11
## X.1   24a-hNIP-control-tdp  240   250      2860      16       9
## X.2   24a-hNIP-c9-tdp     448   315      3011      28      14
##          Everything In Peak ChiSq Near Peak Beta Near Peak Everything Near Peak
## init            380           58        45           907
## X              330           72        40           736
## X.1             274           40        35           627
## X.2             258           60        29           586

```

qq-plots

```
bqtls_betas[[1]]$asb_q %>% qq(main = "Beta ASB q-values") %>% print
```

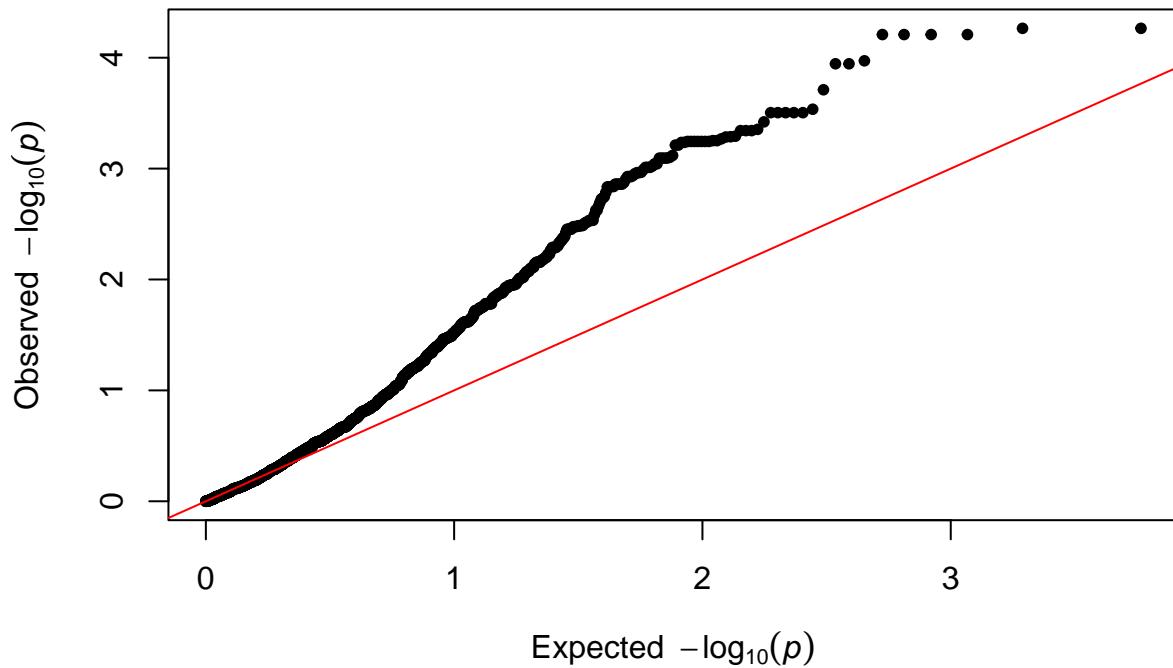
Beta ASB q-values



```
## NULL
```

```
all_chisq[[1]]$p.value %>% p.adjust(method = 'fdr') %>% qq(main = "Chi-Sq q-values") %>% print
```

Chi-Sq q-values



```
## NULL
```

Venn diagram of overlap between significant hits of two models in hNIL-control

```
j = 1
bqtls_beta <- bqtls_betas[[j]]
chisq <- all_chisq[[j]]
in_chi_sq <- chisq$q.value < 0.05
in_beta <- bqtls_beta$asb_q < 0.05

list(chiSq = bqtls_beta$variantID[in_chi_sq],
     beta = bqtls_beta$variantID[in_beta])%>%
venn%>%
plot
```

