Inspection of Popoolation Data from Onion BSA RNASEQ Experiment

Try to retrieve using Rcurl

```
# library('RCurl', lib.loc='C:/Program Files/R/R-3.0.2/library') all <-
# scp('genome3.pfr.co.nz','/workspace/genome_analysis/plant/Allium/cepa/NXD_sample_align/35
# Files (x86)/WinSCP/PuTTY/craptop2013.ppk')),binary=FALSE )
```

Gave up and scp-ed from /workspace/genome_analysis/plant/Allium/cepa/NXD_sample_align/35.cmh_test_a

```
column_names <- c("BoltA1_RG.bam", "BoltA2_RG.bam", "BoltA3_RG.bam", "BoltB1_RG.bam",
    "BoltB2_RG.bam", "BoltB3_RG.bam", "NonA1_RG.bam", "NonA2_RG.bam", "NonA3_RG.bam",
    "NonB1_RG.bam", "NonB2_RG.bam", "NonB3_RG.bam")

all_bolt <- read.table("All_bolt_assoc.cmh", col.names = c("chrom", "pos", "ref_base",
    column_names, "pval"))
acp267_LD <- read.table("ACP267_LD.cmh.bz2", col.names = c("chrom", "pos", "ref_base",
    column_names, "pval"))
Bolt_notACP267 <- read.table("Bolt_notACP267_LD_assoc.cmh.bz2", , col.names = c("chrom",
    "pos", "ref_base", column_names, "pval"))
```

Bind together

```
combined <- cbind(all_bolt, acp267_LD$pval, Bolt_notACP267$pval)
```

Now plot the pvalues

```
library(ggplot2)
hist_all_pval <- ggplot(all_bolt, aes(x = pval)) + geom_histogram() + ggtitle("all")
hist_ACP_pval <- ggplot(acp267_LD, aes(x = pval)) + geom_histogram() + ggtitle("ACP267")
hist_NotACP_pval <- ggplot(Bolt_notACP267, aes(x = pval)) + geom_histogram() +
    ggtitle("NotACP267")

hist_all_pval
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```

```
hist_ACP_pval
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```
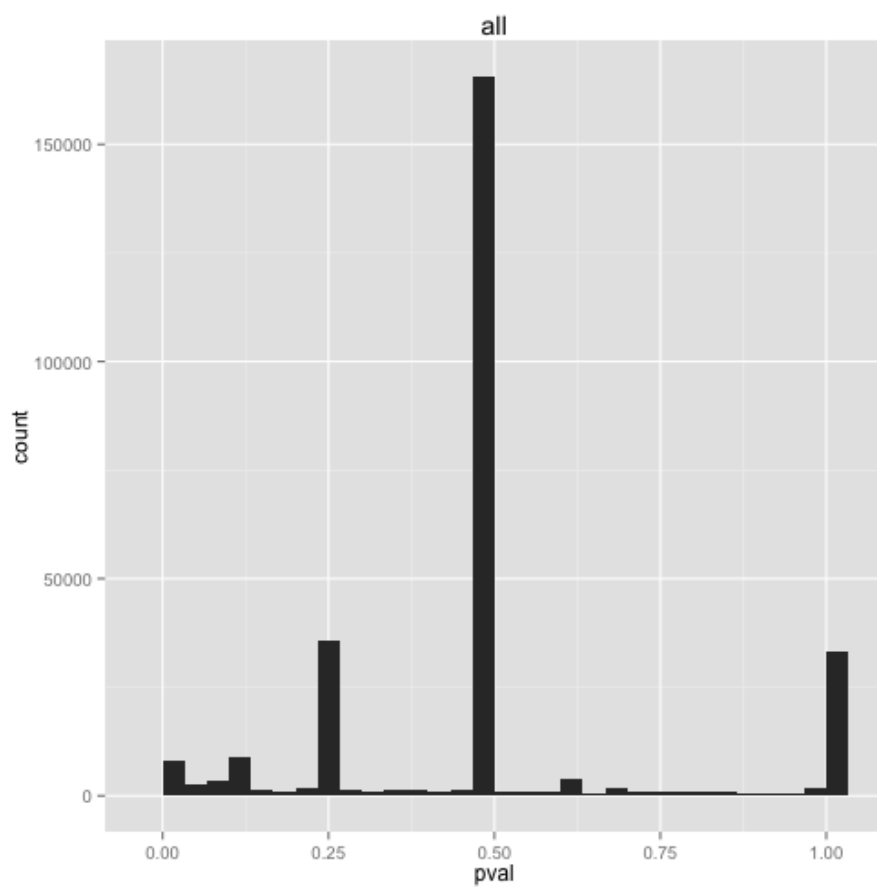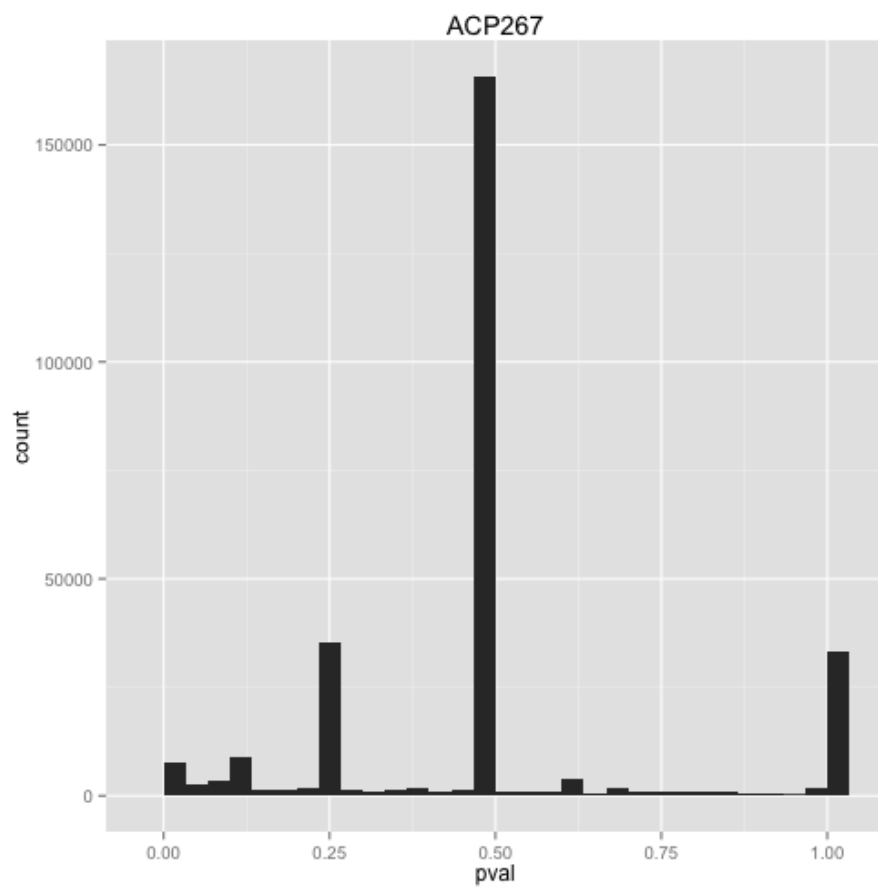
Figure 1: plot of chunk unnamed-chunk-4
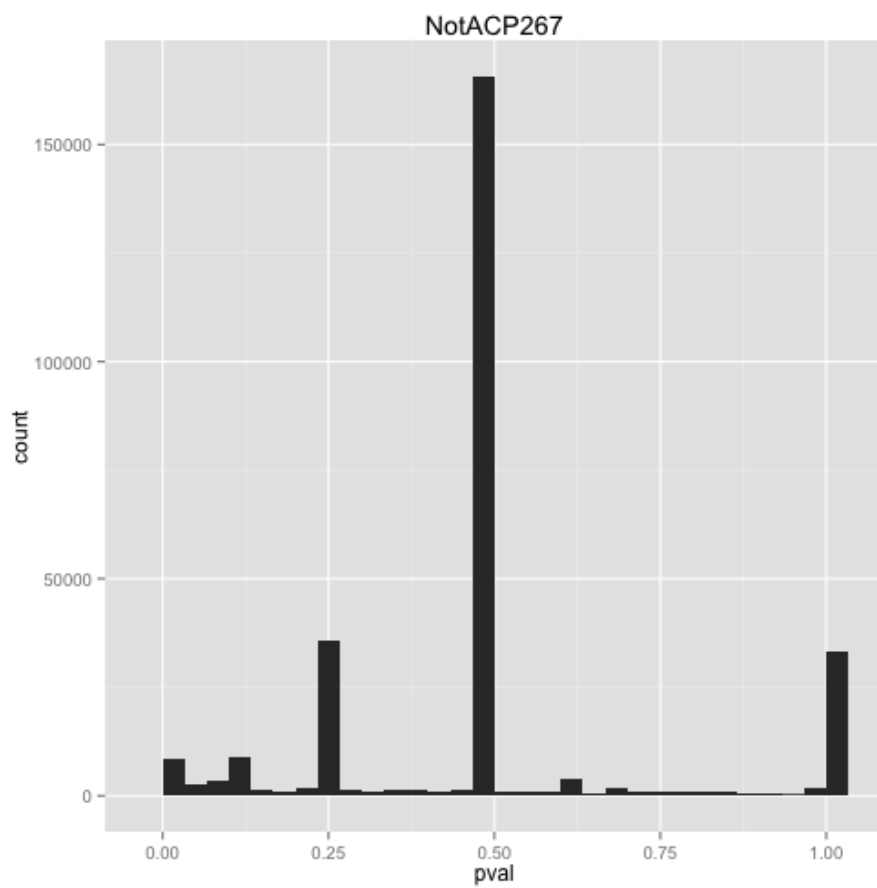
Figure 2: plot of chunk unnamed-chunk-4

Figure 3: plot of chunk unnamed-chunk-4

```
hist_NotACP_pval
```

`## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.`

Now estimate FDR

```
library("qvalue", lib.loc = "/Library/Frameworks/R.framework/Versions/3.0/Resources/library"
```

```
##
## Attaching package: 'qvalue'
##
## The following object is masked from 'package:ggplot2':
##
##     qplot
```

```
all_qobj <- qvalue(all_bolt$pval)
acp_qobj <- qvalue(acp267_LD$pval)
nonACP_qobj <- qvalue(Bolt_notACP267$pval)
qplot(all_qobj)
```

Estimate p value cutoff for FDR= 0.01

```
FDR <- 0.01
max(all_qobj$pvalues[all_qobj$qvalues <= FDR])
```

`## [1] 8.997e-05`

```
max(acp_qobj$pvalues[acp_qobj$qvalues <= FDR])
```

`## [1] 7.034e-05`

```
max(nonACP_qobj$pvalues[nonACP_qobj$qvalues <= FDR])
```

`## [1] 8.973e-05`

Use minimum of these as cutoff All are about 7e-05

```
pval_cutoff <- max(acp_qobj$pvalues[acp_qobj$qvalues <= FDR])
```
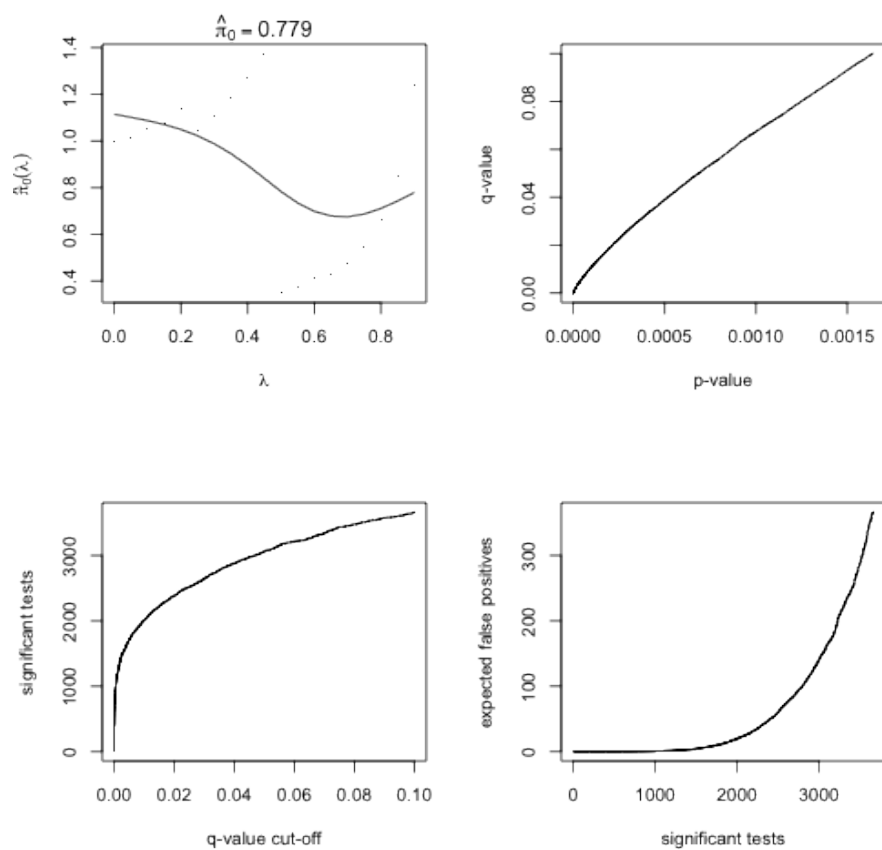
Get the **minimum** pvalues by contig

Figure 4: plot of chunk unnamed-chunk-5

```
by_contig <- aggregate(data = combined, cbind(pval, acp267_LD$pval, Bolt_notACP267$pval) ~
    chrom, min)
colnames(by_contig) <- c("chrom", "all_pval", "ACP267_pval", "NotACP267_pval")
```

Now form a Venn diagram as described at http://www.ats.ucla.edu/stat/r/faq/venn.htm

Form summary of booleans-contigs with minimum p-value < FDR cutoff

```
by_contig$sig_all <- by_contig$all_pval < max(all_qobj$pvalues[all_qobj$qvalues <=
    FDR])
by_contig$sig_acp267 <- by_contig$ACP267_pval < max(acp_qobj$pvalues[acp_qobj$qvalues <=
    FDR])
by_contig$sig_nonACP <- by_contig$NotACP267_pval < max(nonACP_qobj$pvalues[nonACP_qobj$qval
    FDR])

library("limma", lib.loc = "/Library/Frameworks/R.framework/Versions/3.0/Resources/library")
a <- vennCounts(by_contig[5:7])
vennDiagram(a)
```

Make a scatterplot

```
ggplot(by_contig, aes(x = -log10(ACP267_pval))) + geom_point(aes(alpha = -log10(all_pval),
    y = -log10(NotACP267_pval))) + ggtitle("Contig minimum CMH test pvalue") +
    coord_fixed()
```

Producing pdf output using pandoc

```
pandoc popoolation2_pvalues.md -o popoolation2_pvalues.pdf
```
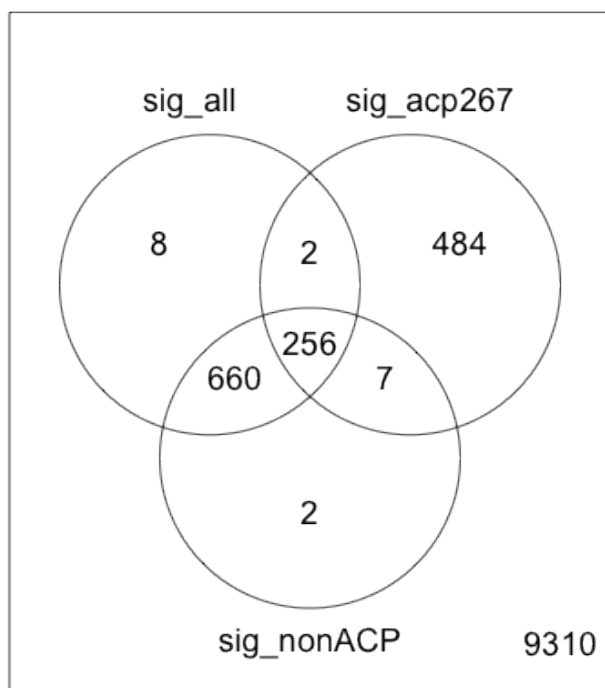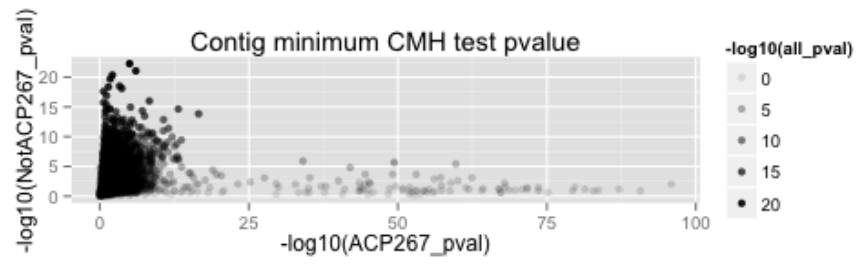
Figure 5: plot of chunk Venn

Figure 6: plot of chunk Scatterplot of Contig Minimum Pvalues