NGS DATA ANALYSIS

Priyanshu Sharma

2023-08-12

```r
setwd("D:/R/Pan_NGS_data/Rna_seq/Results")
# STEP 2 Define the samples with their read IDs (SRA run accessions) and calculate RPKM values...
samples_name <- c("AMB + LF rep 1", "AMB + LF rep 2", "AMB + LF rep 3", "control 1", "control 2", "control 3")
names(samples_name) <- c( "SRR3396388", "SRR3396389", "SRR3396390", "SRR3396391", "SRR3396392", "SRR3396392")

# Construct the file path -
files <- c()

for (s in samples_name) {
  fp <- paste0(s, ".bam")
  if (file.exists(fp)) files <- c(files, fp)
}

# Load in the BAM files
bams_files <- BamFileList(files, yieldSize = 100000, asMates=TRUE)
# Overlap BAM reads and genes
overlap <- summarizeOverlaps(exons, bams_files, mode="Union", singleEnd=FALSE, ignore.strand=TRUE, fragments=TRUE)
overlap

## class: RangedSummarizedExperiment
## dim: 7127 6
## metadata(0):
## assays(1): counts
## rownames(7127): ETS1-1 ETS1-2 ... tY(GUA)O tY(GUA)Q
## rowData names(0):
## colnames(6): AMB + LF rep 1.bam AMB + LF rep 2.bam ... control 2.bam
##   control 3.bam
## colData names(0):

# Extract the raw-reads per gene
geneRawReads <- assay(overlap)
geneRawReads[1:10,]

##         AMB + LF rep 1.bam AMB + LF rep 2.bam AMB + LF rep 3.bam control 1.bam
## ETS1-1                 306               262               159
      286
## ETS1-2                 311               237               161
```

```
        243
## ETS2-1                  96              43              37
         22
## ETS2-2                  83              45              34
         31
## HRA1                     1               1               1
          1
## ICR1                   104              95              78
        132
## IRT1                   144             103              94
        129
## ITS1-1                  21              25              18
         15
## ITS1-2                  41              52              31
         36
## ITS2-1                  58              69              47
         63
##         control 2.bam control 3.bam
## ETS1-1            447           363
## ETS1-2            487           407
## ETS2-1             28            17
## ETS2-2             27            19
## HRA1               0            1
## ICR1              104           174
## IRT1              116           152
## ITS1-1             19            32
## ITS1-2             50            45
## ITS2-1             62            64
```

```r
colnames(geneRawReads) <- gsub(".bam", "", colnames(geneRawReads))
write.table(geneRawReads, "gene_raw_reads.txt", sep = "\t", quote = FAL
SE)

# Calculate the RPKM values per gene
# RPKM = (10^9 * C)/(N * L)
# C = Number of reads mapped to a gene
# N = Total mapped reads in the sample
# L = gene length in base-pairs for a gene

# Get the total number of reads per samples
totalMappedReads <- apply(geneRawReads, 2, sum)

# Loop through all genes, compute RPKM
geneIdx <- 1
geneRPKM <- t(apply(geneRawReads, 1, function(C){
  geneLength <- as.numeric(gene.lengths[geneIdx])
  RPKM <- (10 ^ 9 * C) / (totalMappedReads * geneLength)
  geneIdx <<- geneIdx + 1
  return(round(RPKM, d = 1))
}))
```
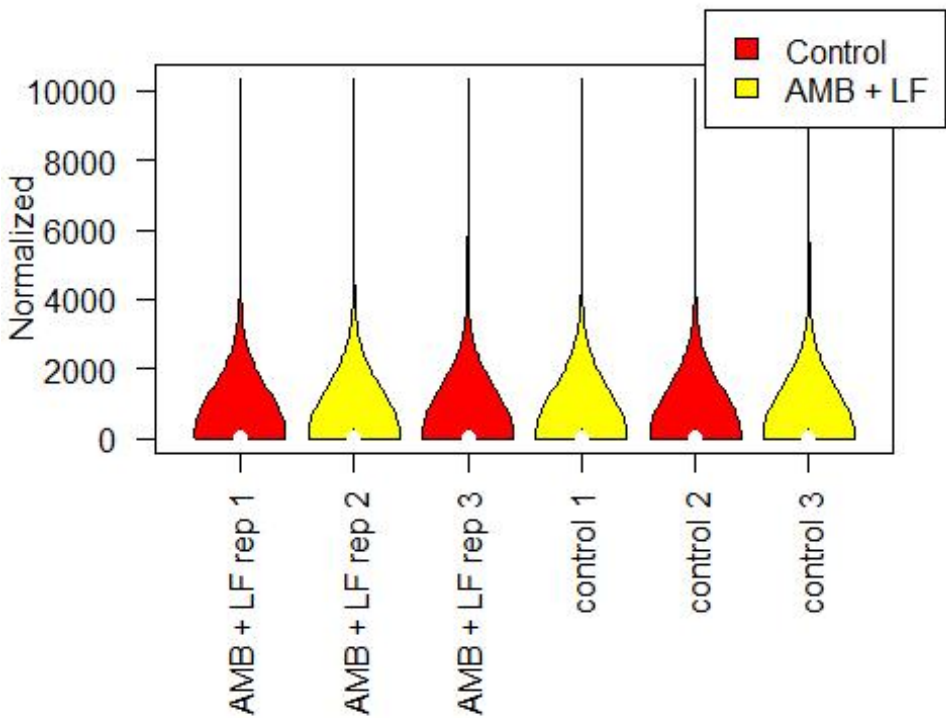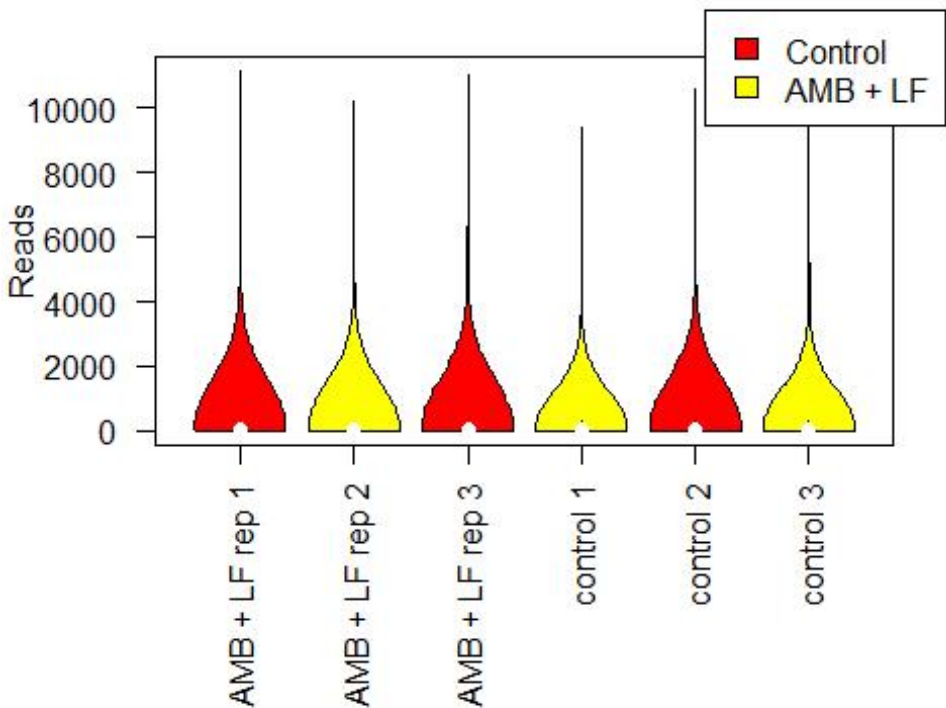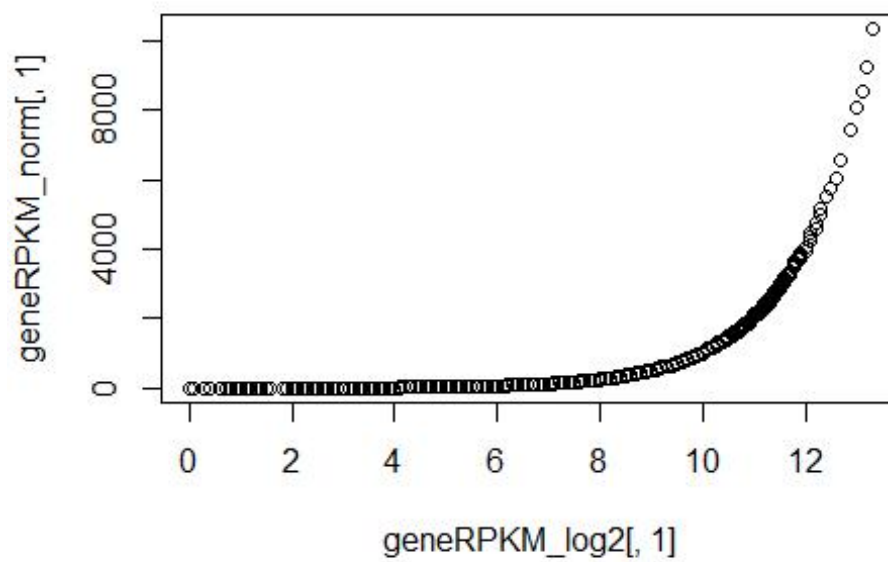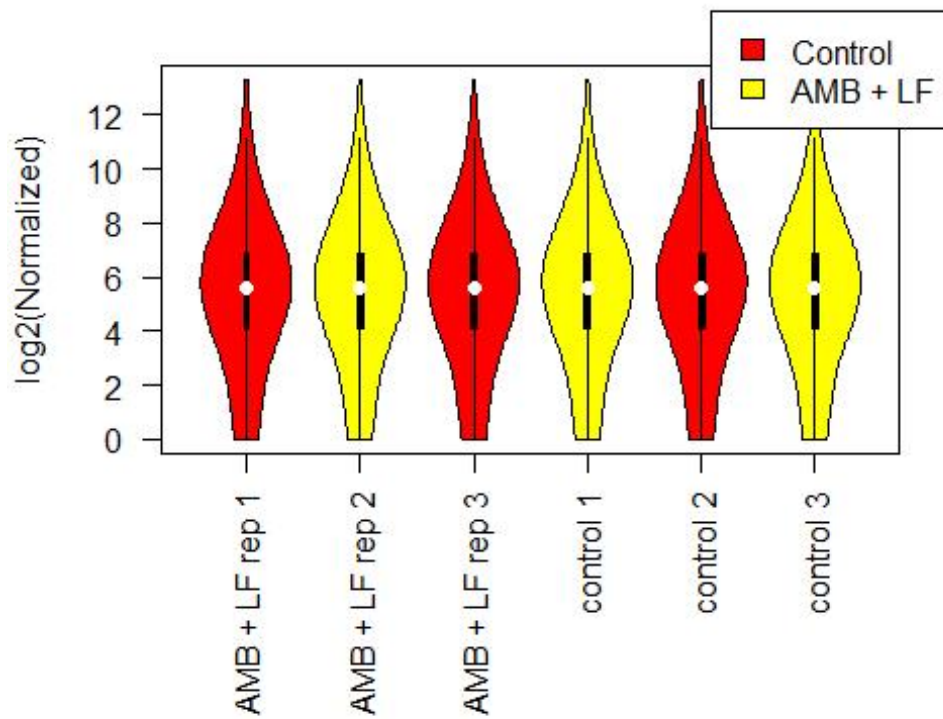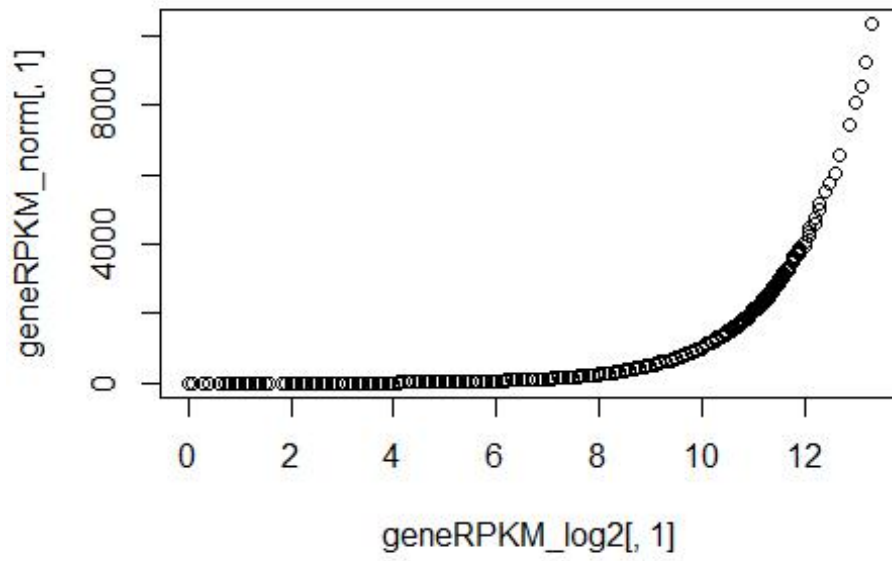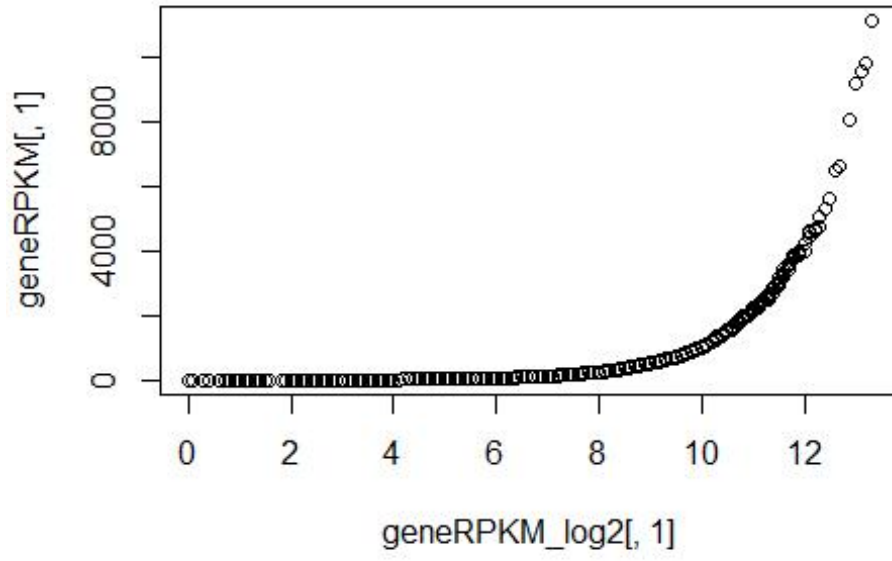
```
geneRPKM[1:10,]
```

```
##              AMB+LF       AMB+LF       AMB+LF control 1 control 2
## ETS1-1        59.4         59.5         35.5      61.0      96.2
## ETS1-2        60.3         53.8         36.0      51.8     104.8
## ETS2-1        61.8         32.4         27.4      15.6      20.0
## ETS2-2        53.4         33.9         25.2      21.9      19.3
## HRA1           0.2          0.3          0.3       0.3       0.0
## ICR1           4.4          4.7          3.8       6.2       4.9
## IRT1          13.1         11.0          9.9      12.9      11.7
## ITS1-1         7.9         11.0          7.8       6.2       7.9
## ITS1-2        15.4         22.9         13.4      14.9      20.9
## ITS2-1        33.9         47.3         31.7      40.6      40.2
## control 3
## ETS1-1      66.4
## ETS1-2      74.5
## ETS2-1      10.3
## ETS2-2      11.5
## HRA1         0.2
## ICR1         7.0
## IRT1        13.1
## ITS1-1      11.4
## ITS1-2      16.0
## ITS2-1      35.3
```

Including Plots

Histogram of geneRPKM_log2[, 1]

Histogram of pvals

Histogram of fc

Volcano plot
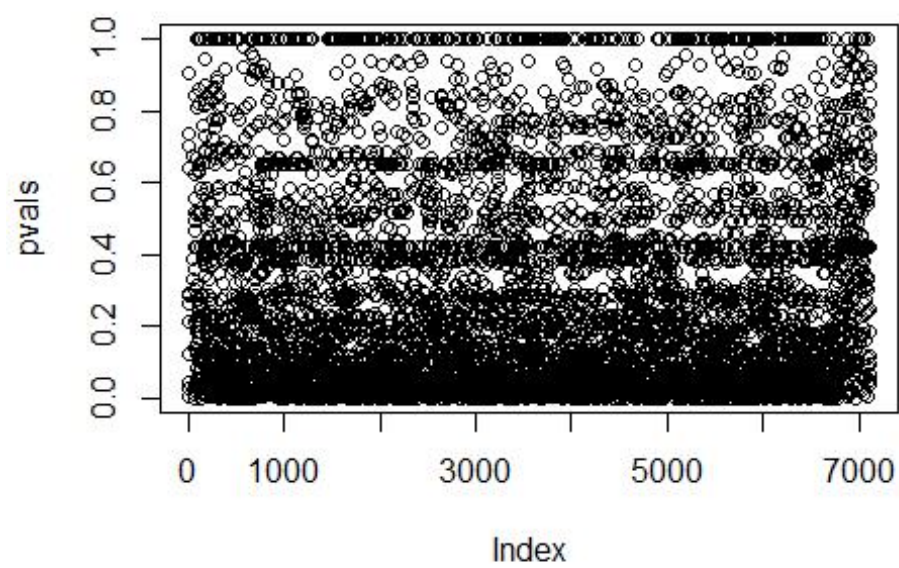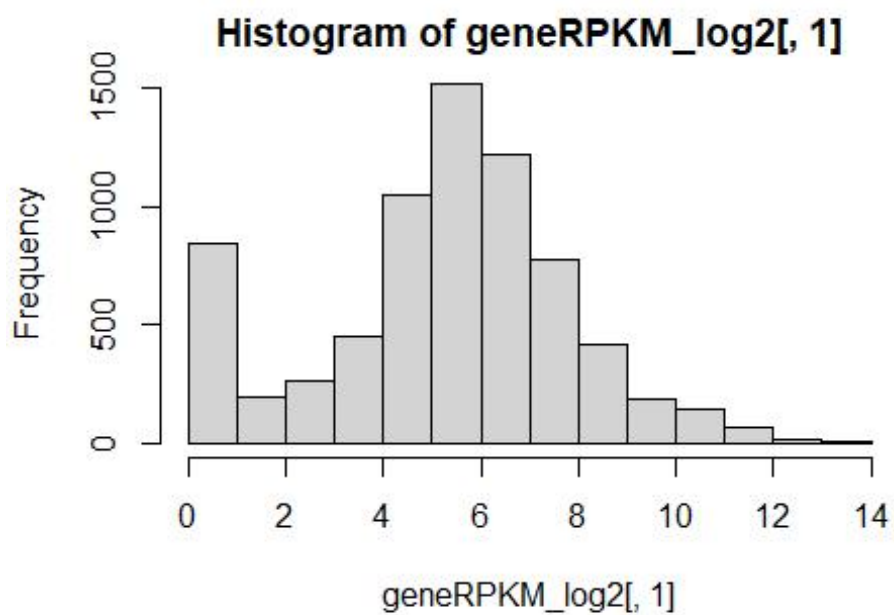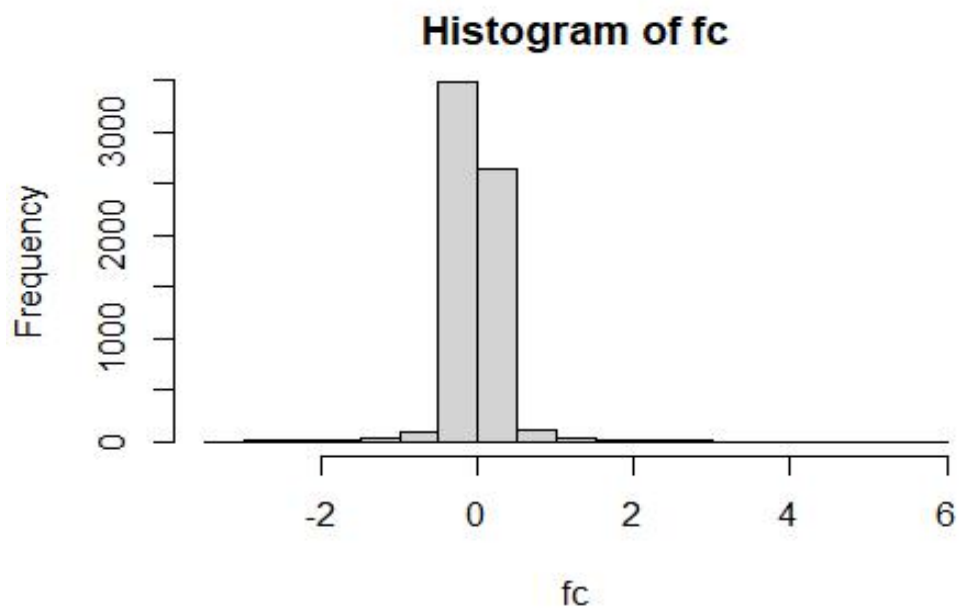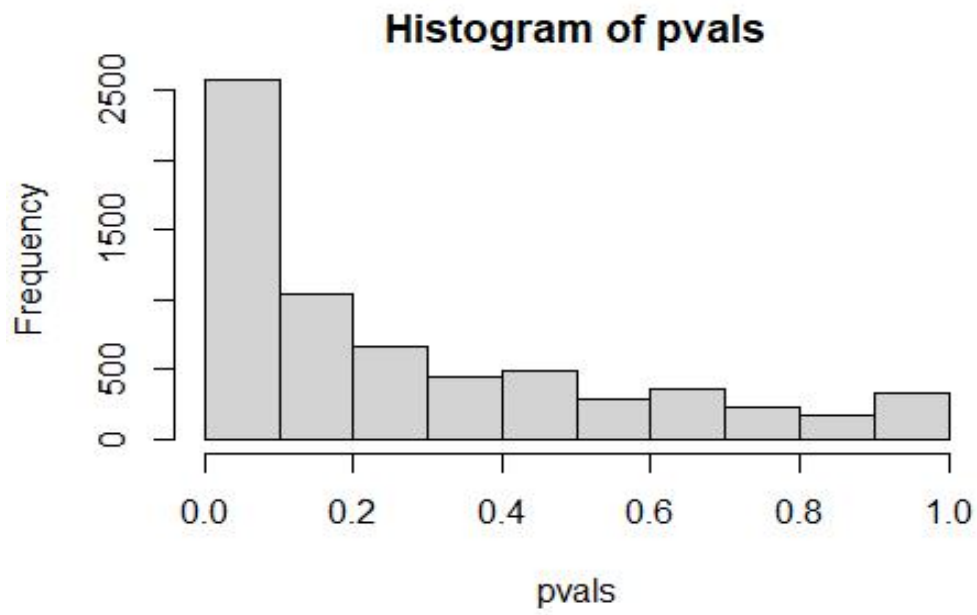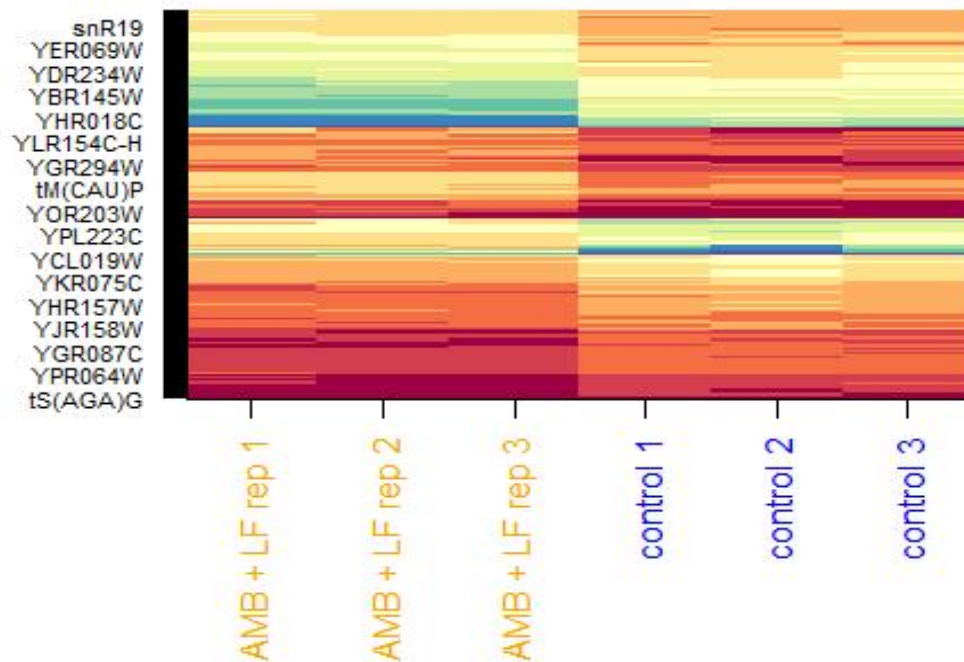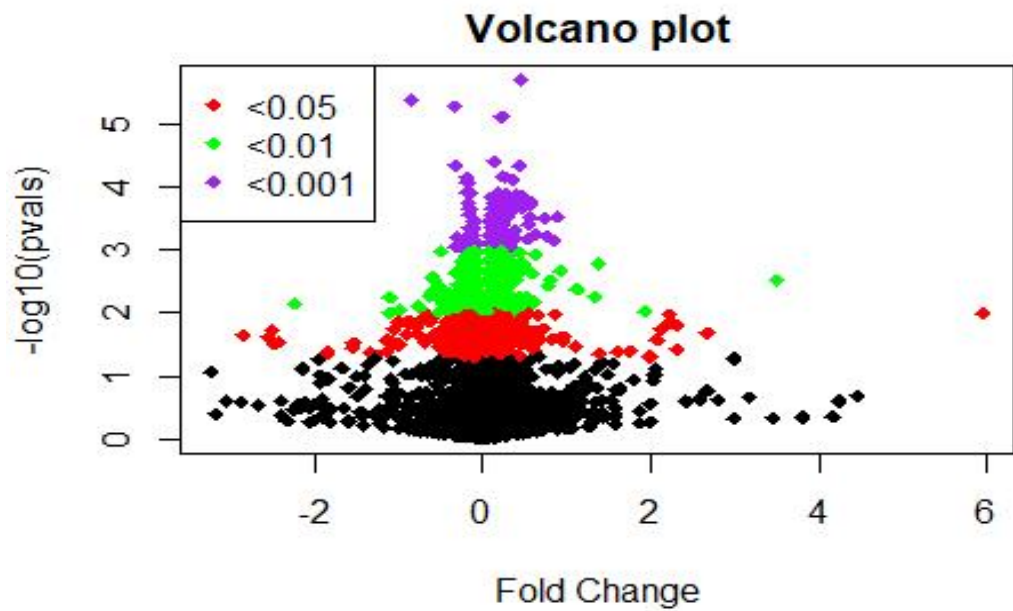

```
##          Control Control(SD) AMB + LF AMB + LF(SD)   FC        P
## ETS1-1     6.1        0.4       5.7         0.4    -0.1 0.291413
## ETS1-2     6.2        0.6       5.7         0.3    -0.1 0.261761
## ETS2-1     3.9        0.5       5.3         0.6     0.5 0.029178
```

```
## ETS2-2      4.1          0.5      5.2         0.5  0.4 0.043726
## HRA1        0.0          0.0      0.0         0.0  NaN      NaN
## ICR1        2.6          0.4      2.1         0.1 -0.3 0.121094
## IRT1        3.6          0.1      3.5         0.1  0.0 0.287864
## ITS1-1      3.0          0.4      3.1         0.4  0.1 0.701983
## ITS1-2      4.1          0.3      4.1         0.3  0.0 0.906643
## ITS2-1      5.2          0.1      5.3         0.3  0.0 0.735499

##           GeneName  Chr    Start      End Control Control(SD) AMB + LF
## tS(AGA)G           VII   561662   561743     0.8         0.2      0.0
## YMR322C       SNO4 XIII  918366   919079     1.2         0.3      0.2
## YNL109W            XIV   418964   419509     1.2         0.3      0.0
## YIR021W-A           IX   398514   398726     1.4         0.5      0.2
## tD(GUC)J1            X   204735   204806     1.4         0.5      0.0
## YNR073C       MAN2 XIV   774792   776300     2.0         0.4      0.4
## YFL058W       THI5  VI    12929    13951     1.7         0.5      0.3
## YPR007C       REC8 XVI   569336   571378     1.7         0.1      0.3
## YDR442W            IV  1345647  1346039     1.9         0.5      0.8
## YBR180W       DTR1  II   589741   591459     1.5         0.1      0.9
##           AMB + LF(SD)    FC        P
## tS(AGA)G           0.0 -Inf 0.011015
## YMR322C            0.3 -2.5 0.024172
## YNL109W            0.0 -Inf 0.024384
## YIR021W-A          0.3 -2.8 0.022259
## tD(GUC)J1          0.0 -Inf 0.044467
## YNR073C            0.6 -2.4 0.028682
## YFL058W            0.4 -2.5 0.018640
## YPR007C            0.4 -2.5 0.030434
## YDR442W            0.5 -1.3 0.040943
## YBR180W            0.2 -0.7 0.010154
```

Conclusion

This study provides valuable insights into the complex mechanism of synergy between AMB and LF in S. cerevisiae. The unexpected involvement of oxidative stress response pathways and the pivotal role of transcription factors Aft1p and Zap1p highlight the multifaceted nature of the synergistic interaction. The implications of our findings extend beyond S. cerevisiae, suggesting novel avenues for the development of enhanced antifungal therapies with broader applicability in the fight against invasive fungal infections.