

RNA Analysis

Michael Callahan

2024-04-03

Load data

```
library(here)
library(dplyr)

current_project_path <- here()

cols_dir_path <- file.path(current_project_path, "Data", "colData.Rdata")
counts_dir_path <- file.path(current_project_path, "Data", "countData.Rdata")
# tpm_dir_path <- file.path(current_project_path, "Data", "tpm.Rdata")

load(cols_dir_path)
load(counts_dir_path)

countData <- countData %>% select(-length)
```

Differential Expression Analysis - CMP/CFUE setup

CMP is denom

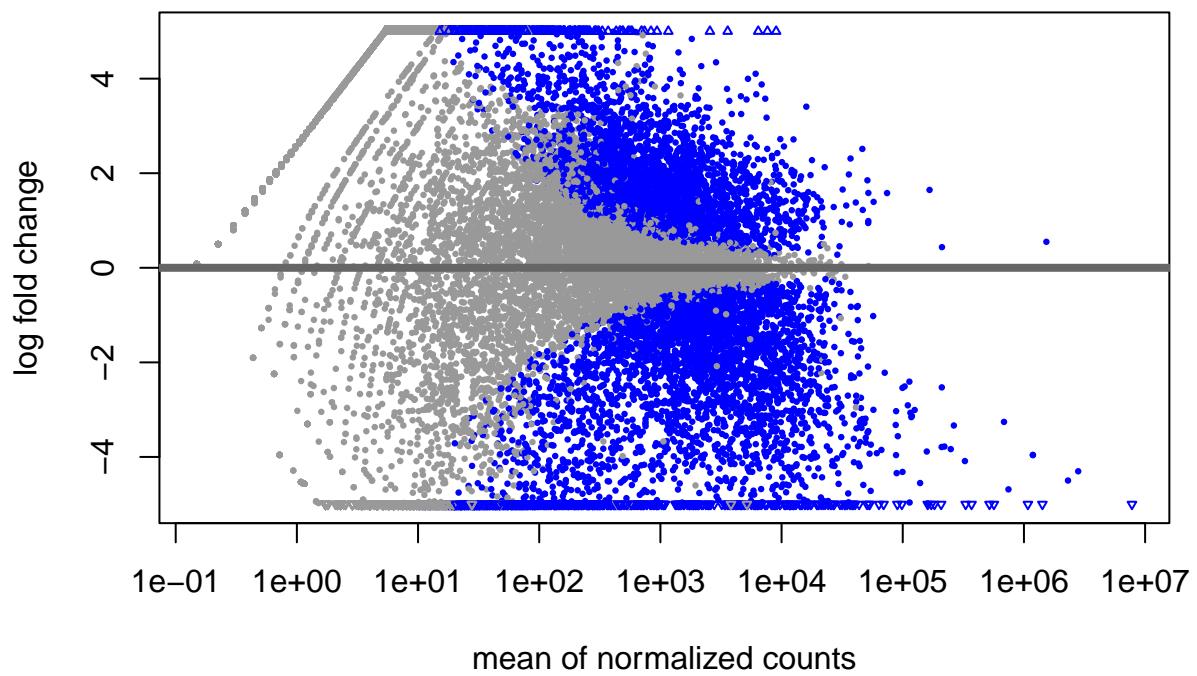
```
helper_dir_path <- file.path(current_project_path, "analysis", "Helper_functions.R")
source(helper_dir_path)

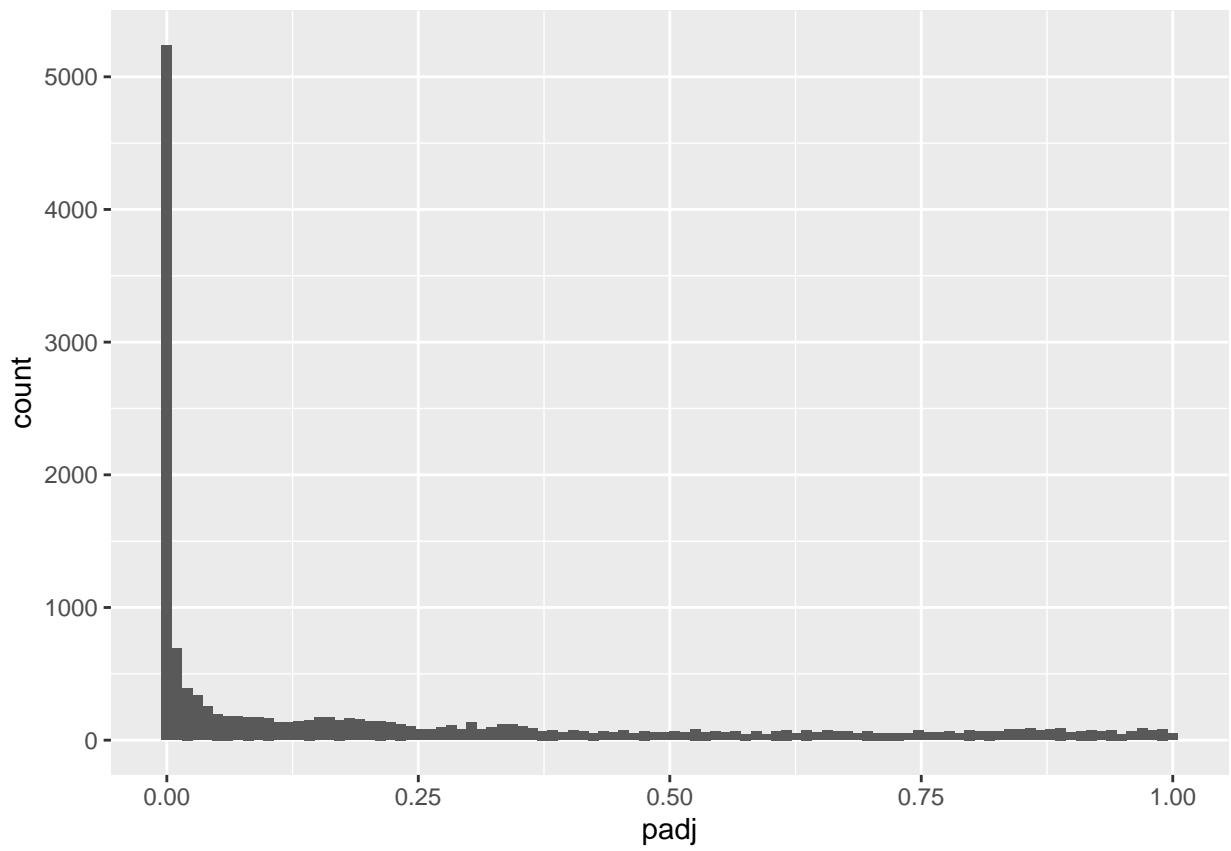
selected_lines = c('CFUE', 'CMP')

selected_samples <- colData %>%
  filter(group %in% selected_lines) %>%
  mutate(group = factor(group, levels = selected_lines)) %>%
  arrange(group) %>%
  pull(source_name)
```

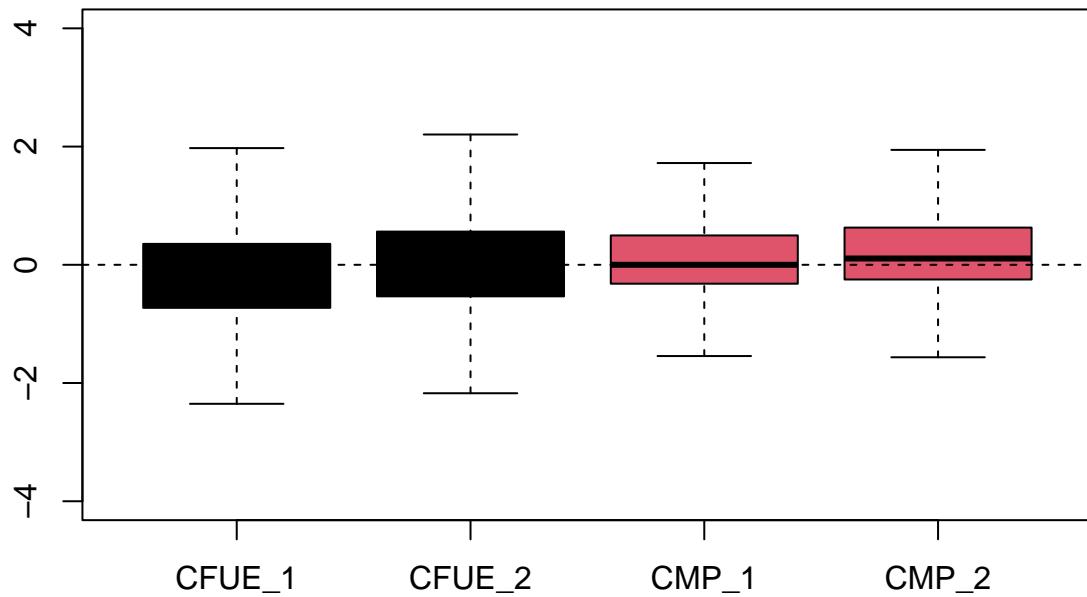
DESeq2

```
DEresults = get_DE_results(selected_lines, colData, countData)
```

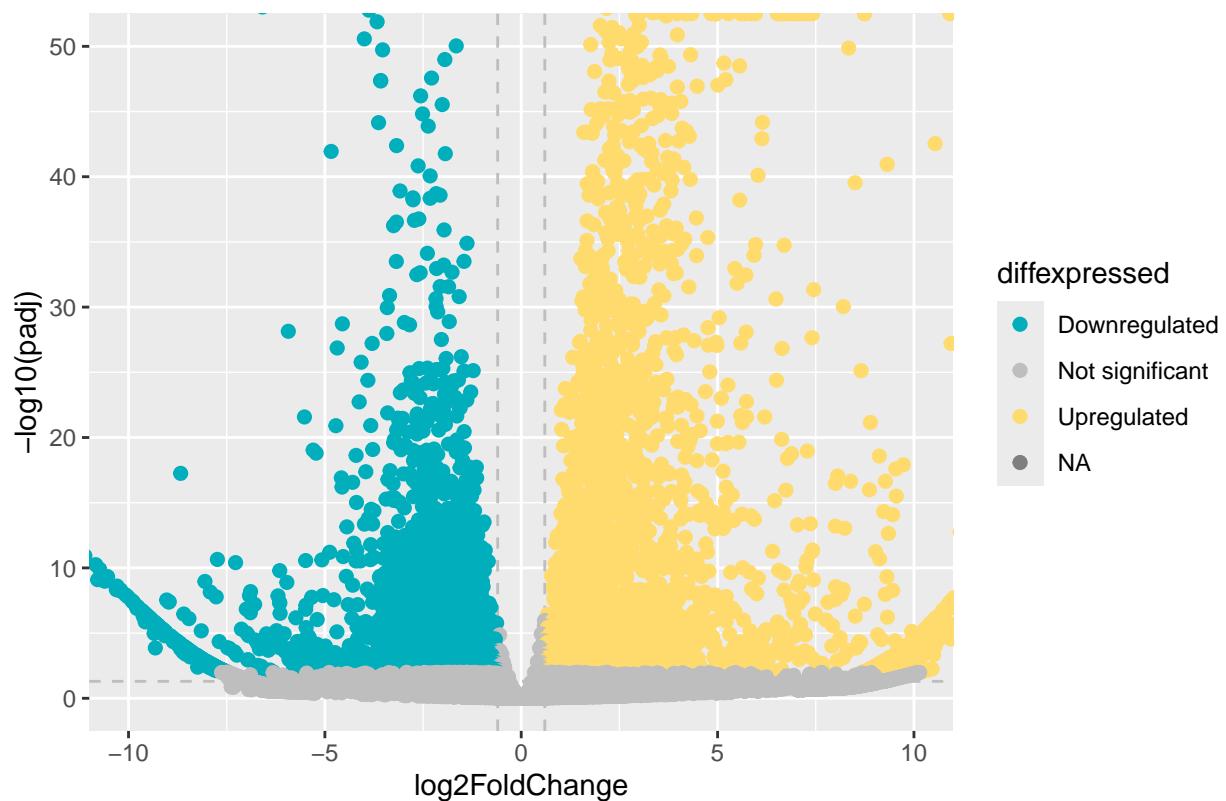




Normalized Counts



Volcano Plot: CFUE vs. CMP



```
DEresults_print <- DEresults[1:10,]
kable(DEresults_print)
```

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
ENSMUSG00000001763.10	24259.49	6.523522	0.1361695	47.90738	0	0
ENSMUSG00000004319.11	35665.11	4.958592	0.1279849	38.74359	0	0
ENSMUSG00000019838.10	22510.73	4.934590	0.1295893	38.07869	0	0
ENSMUSG00000020300.10	21231.66	5.966079	0.1321296	45.15324	0	0
ENSMUSG00000022797.11	113887.24	4.963075	0.1133247	43.79518	0	0
ENSMUSG00000024588.8	22891.49	5.583420	0.1205164	46.32913	0	0
ENSMUSG00000025270.9	39183.89	13.488823	0.2976389	45.31942	0	0
ENSMUSG00000026532.7	160687.69	6.815698	0.1391425	48.98359	0	0
ENSMUSG00000026605.10	89029.35	4.341960	0.1025293	42.34849	0	0
ENSMUSG00000026880.10	26185.87	4.180457	0.1112150	37.58897	0	0

Limma-VOOM

```
debug(get_lv_results)
lv_results = get_lv_results(selected_lines, colData, countData)
```

```
## debugging in: get_lv_results(selected_lines, colData, countData)
```

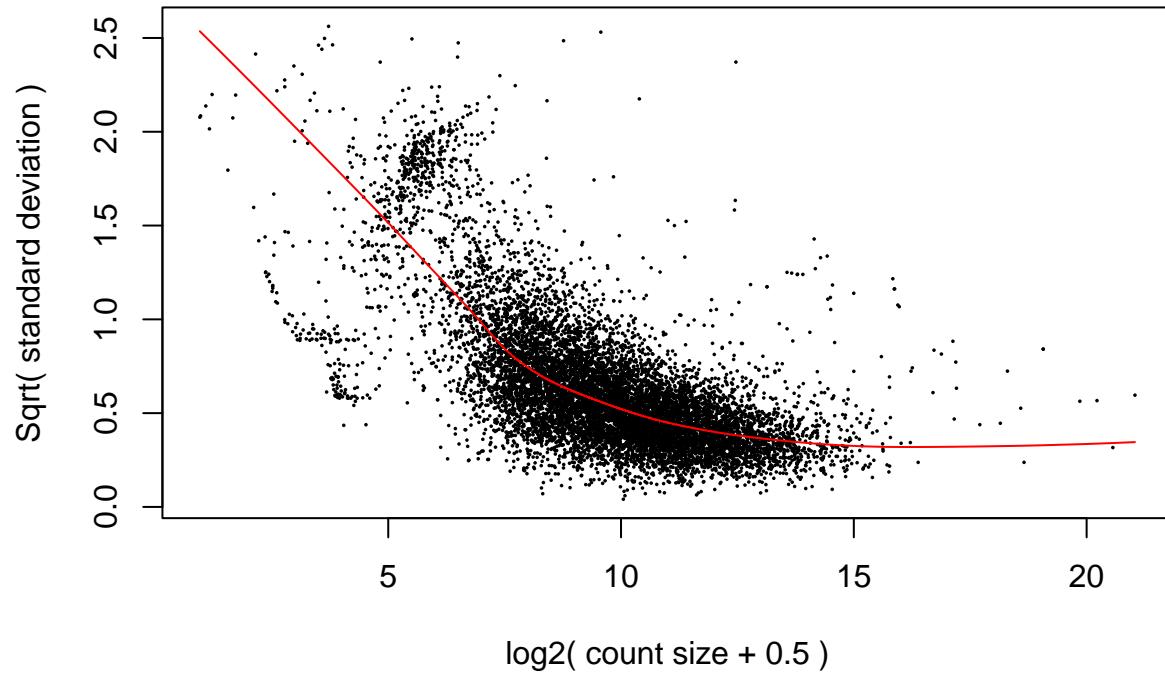
```

## debug: {
##   d0 <- DGEList(countData[, selected_samples])
##   d0 <- calcNormFactors(d0)
##   cutoff <- 5
##   drop <- which(apply(cpm(d0), 1, max) < cutoff)
##   d <- d0[-drop, ]
##   dim(d)
##   group = as.character(colData[match(selected_samples, colData$source_name),
##     ]$group)
##   mm <- model.matrix(~group)
##   colnames(mm) <- gsub("group", "", colnames(mm))
##   y <- voom(d, mm, plot = T)
##   fit <- lmFit(y, mm)
##   tmp <- eBayes(fit)
##   top.table <- topTable(tmp, sort.by = "P", n = Inf)
##   pval_plot <- ggplot(data = top.table, aes(x = adj.P.Val)) +
##     geom_histogram(bins = 100)
##   plot(pval_plot)
##   return(top.table)
## }

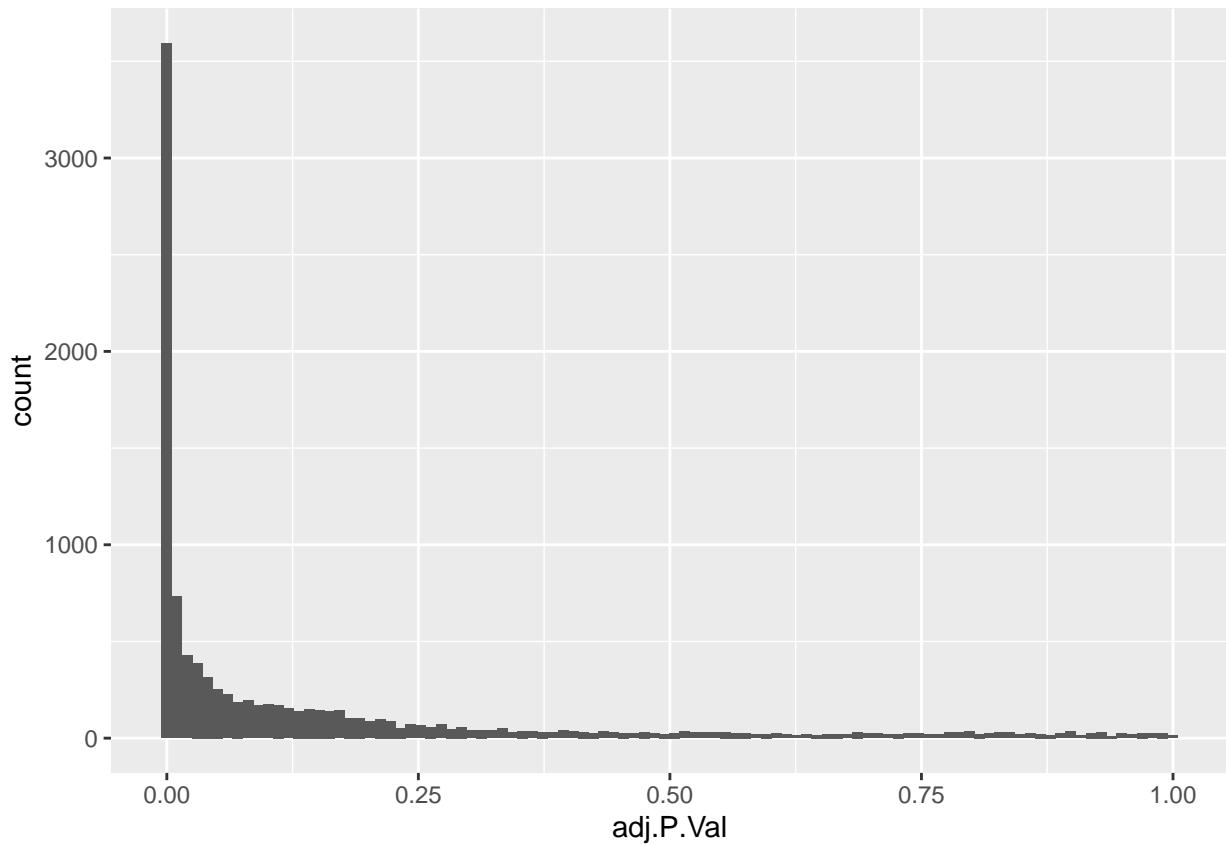
## debug: d0 <- DGEList(countData[, selected_samples])
## debug: d0 <- calcNormFactors(d0)
## debug: cutoff <- 5
## debug: drop <- which(apply(cpm(d0), 1, max) < cutoff)
## debug: d <- d0[-drop, ]
## debug: dim(d)
## debug: group = as.character(colData[match(selected_samples, colData$source_name),
##   ]$group)
## debug: mm <- model.matrix(~group)
## debug: colnames(mm) <- gsub("group", "", colnames(mm))
## debug: y <- voom(d, mm, plot = T)

```

voom: Mean–variance trend



```
## debug: fit <- lmFit(y, mm)
## debug: tmp <- eBayes(fit)
## debug: top.table <- topTable(tmp, sort.by = "P", n = Inf)
## debug: pval_plot <- ggplot(data = top.table, aes(x = adj.P.Val)) + geom_histogram(bins = 100)
## debug: plot(pval_plot)
```



```
## debug: return(top.table)
## exiting from: get_lv_results(selected_lines, colData, countData)
```

```
lv_results_print <- lv_results[1:10,]
kable(lv_results_print)
```

	logFC	AveExpr	t	P.Value	adj.P.Val	B
ENSMUSG00000026532.7	-6.784311	9.319293	-57.26033	0	0	21.65967
ENSMUSG00000031543.14	-6.794380	8.558492	-51.39757	0	0	20.60699
ENSMUSG00000022797.11	-4.903382	9.717955	-49.47529	0	0	20.62134
ENSMUSG00000021061.11	-7.773284	8.937458	-45.43674	0	0	19.57512
ENSMUSG00000069917.7	-11.281857	8.864213	-44.95223	0	0	18.56586
ENSMUSG00000026605.10	-4.259348	9.649689	-44.82236	0	0	19.74675
ENSMUSG00000034248.7	-7.093337	8.307174	-43.64757	0	0	19.18844
ENSMUSG0000004552.12	-6.117913	8.053612	-42.46552	0	0	19.02481
ENSMUSG00000038664.12	-3.882415	9.803481	-42.45527	0	0	19.25500
ENSMUSG00000028644.12	-4.625642	8.764018	-41.37637	0	0	18.96422

Comparison

```
compare_de_lv(DEresults, lv_results)
```

GO term analysis

```
go_results = get_GO_results(DEresults)

kable(go_results$up_reg[1:10,c(7:11)])
```

precision	recall	term_id	source	term_name
0.6250000	0.1720841	GO:0005515	GO:MF	protein binding
0.6466780	0.1641726	GO:0005737	GO:CC	cytoplasm
0.4013135	0.1762028	GO:0048518	GO:BP	positive regulation of biological process
0.8121528	0.1428135	GO:0005488	GO:MF	binding
0.8293015	0.1397887	GO:0005622	GO:CC	intracellular anatomical structure
0.3691670	0.1757157	GO:0048522	GO:BP	positive regulation of cellular process
0.7570698	0.1403132	GO:0043229	GO:CC	intracellular organelle
0.7001704	0.1423623	GO:0043231	GO:CC	intracellular membrane-bounded organelle
0.7148211	0.1404848	GO:0043227	GO:CC	membrane-bounded organelle
0.7642249	0.1387222	GO:0043226	GO:CC	organelle

```
kable(go_results$down_reg[1:10,c(7:11)])
```

precision	recall	term_id	source	term_name
0.7038184	0.1769743	GO:0005737	GO:CC	cytoplasm
0.3296590	0.2029692	GO:0044271	GO:BP	cellular nitrogen compound biosynthetic process
0.8654971	0.1444980	GO:0005622	GO:CC	intracellular anatomical structure
0.3661729	0.1850627	GO:1901564	GO:BP	organonitrogen compound metabolic process
0.3876560	0.1787370	GO:0036094	GO:MF	small molecule binding
0.7884417	0.1447335	GO:0043229	GO:CC	intracellular organelle
0.2948056	0.2087698	GO:0005829	GO:CC	cytosol
0.7481940	0.1456408	GO:0043227	GO:CC	membrane-bounded organelle
0.7997936	0.1437937	GO:0043226	GO:CC	organelle
0.3751734	0.1797940	GO:0043167	GO:MF	ion binding

Differential Expression Analysis - HSC/CMP setup

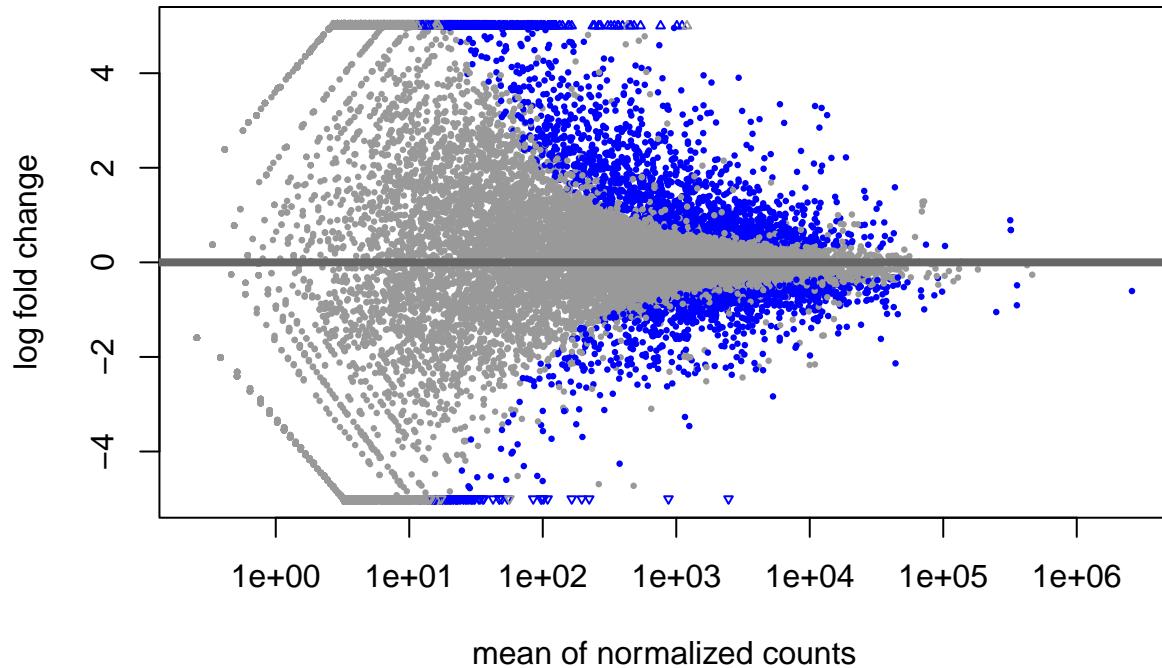
HSC is denominator.

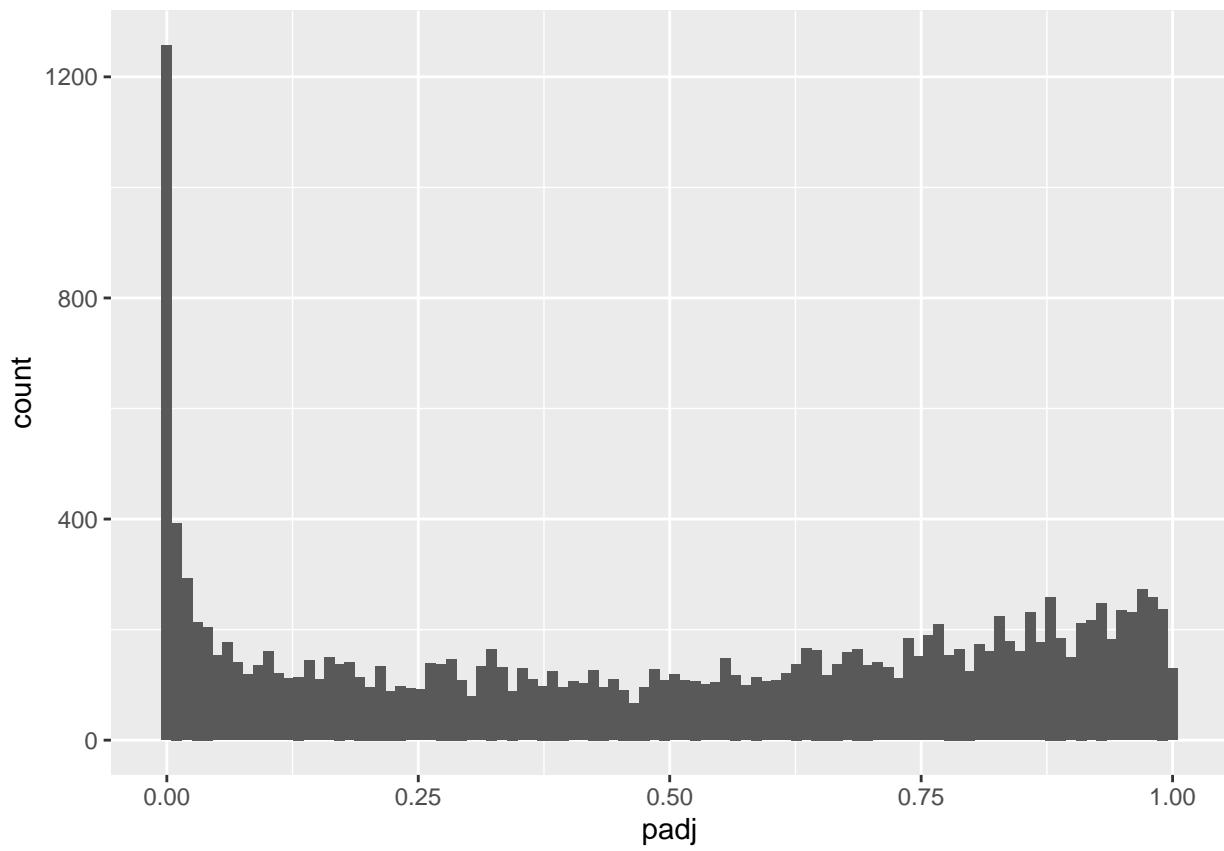
```
selected_lines = c('CMP', 'HSC')

selected_samples <- colData %>%
  filter(group %in% selected_lines) %>%
  mutate(group = factor(group, levels = selected_lines)) %>%
  arrange(group) %>%
  pull(source_name)
```

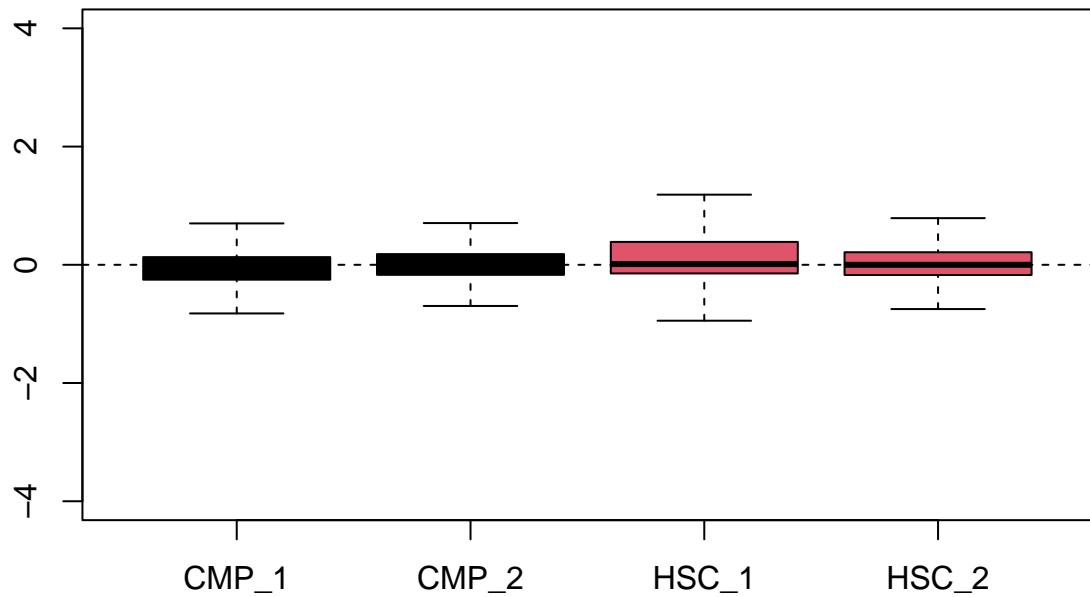
DESeq2

```
DEresults = get_DE_results(selected_lines, colData, countData)
```

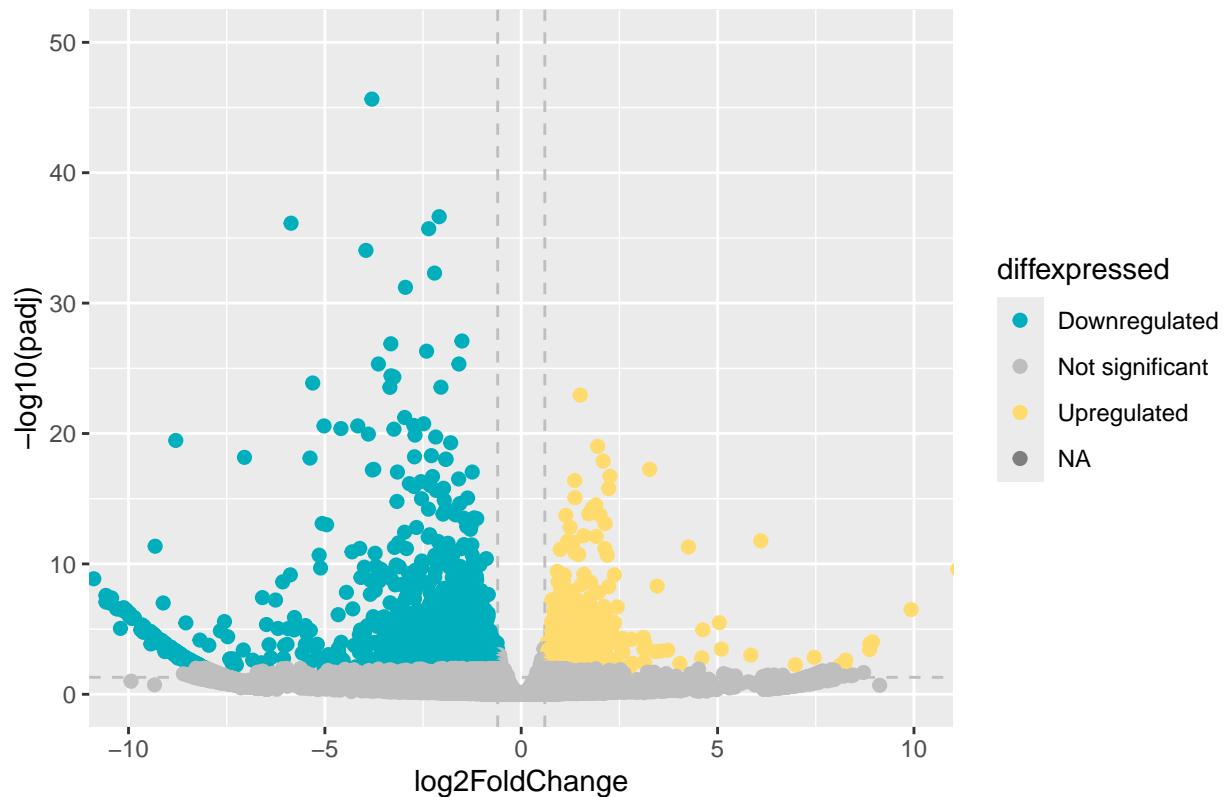




Normalized Counts



Volcano Plot: CMP vs. HSC



```
DEresults_print <- DEresults[1:10,]
kable(DEresults_print)
```

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
ENSMUSG00000027684.12	2932.769	-3.903796	0.2261537	-17.26170	0	0
ENSMUSG00000064262.5	1836.765	-3.802355	0.2550230	-14.90985	0	0
ENSMUSG00000094248.1	2462.930	11.216976	0.8032987	13.96364	0	0
ENSMUSG00000021360.11	10741.815	-2.088976	0.1558726	-13.40181	0	0
ENSMUSG00000040254.7	1104.910	-5.861761	0.4407612	-13.29918	0	0
ENSMUSG00000047867.7	8599.804	-2.356164	0.1783246	-13.21278	0	0
ENSMUSG00000005800.2	1601.630	-3.954594	0.3063241	-12.90984	0	0
ENSMUSG00000078606.4	14533.286	-2.208328	0.1754509	-12.58659	0	0
ENSMUSG00000003949.12	6429.452	-2.948988	0.2382572	-12.37733	0	0
ENSMUSG00000045868.8	25572.804	-1.509827	0.1303529	-11.58261	0	0

GO term analysis

```
go_results = get_GO_results(DEresults)
kable(go_results$up_reg[1:10,c(7:11)])
```

precision	recall	term_id	source	term_name
0.0960000	0.3333333	GO:0030527	GO:MF	structural constituent of chromatin
0.0930851	0.2464789	GO:0000786	GO:CC	nucleosome
0.0986667	0.0984043	GO:0046982	GO:MF	protein heterodimerization activity
0.0513514	0.3877551	GO:0071459	GO:BP	protein localization to chromosome, centromeric region
0.0540541	0.2739726	GO:0071168	GO:BP	protein localization to chromatin
0.0486486	0.6206897	GO:0061644	GO:BP	protein localization to CENP-A containing chromatin
0.0478723	0.6000000	GO:0061638	GO:CC	CENP-A containing chromatin
0.0478723	0.6000000	GO:0043505	GO:CC	CENP-A containing nucleosome
0.0478723	0.5806452	GO:0034506	GO:CC	chromosome, centromeric core domain
0.0567568	0.1544118	GO:0034502	GO:BP	protein localization to chromosome

```
kable(go_results$down_reg[1:10,c(7:11)])
```

precision	recall	term_id	source	term_name
0.2752066	0.1181270	GO:0002376	GO:BP	immune system process
0.2115702	0.1316195	GO:0006955	GO:BP	immune response
0.2809917	0.1074250	GO:0051239	GO:BP	regulation of multicellular organismal process
0.1785124	0.1371429	GO:0002682	GO:BP	regulation of immune system process
0.2454545	0.1073753	GO:0007166	GO:BP	cell surface receptor signaling pathway
0.1818182	0.1214128	GO:0006952	GO:BP	defense response
0.2553719	0.1015779	GO:0009605	GO:BP	response to external stimulus
0.3016529	0.0927337	GO:0006950	GO:BP	response to stress
0.2090909	0.1072488	GO:0048584	GO:BP	positive regulation of response to stimulus
0.1752066	0.1175166	GO:0051240	GO:BP	positive regulation of multicellular organismal process

Differential Expression Analysis - CMP/EBLST setup

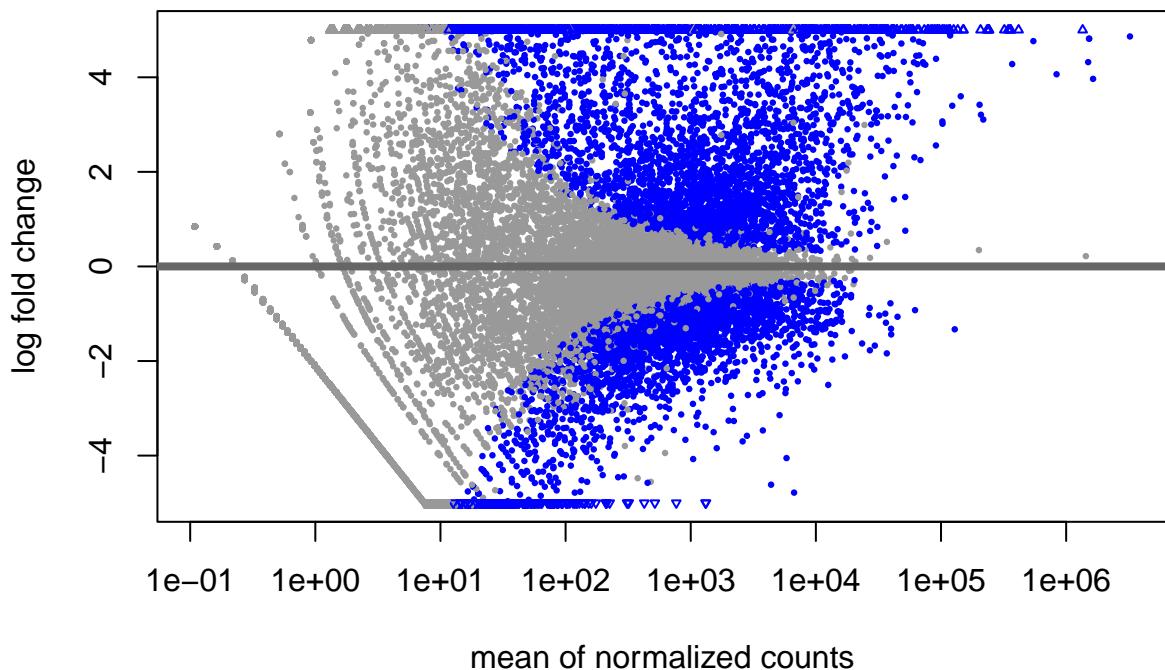
CMP is denom

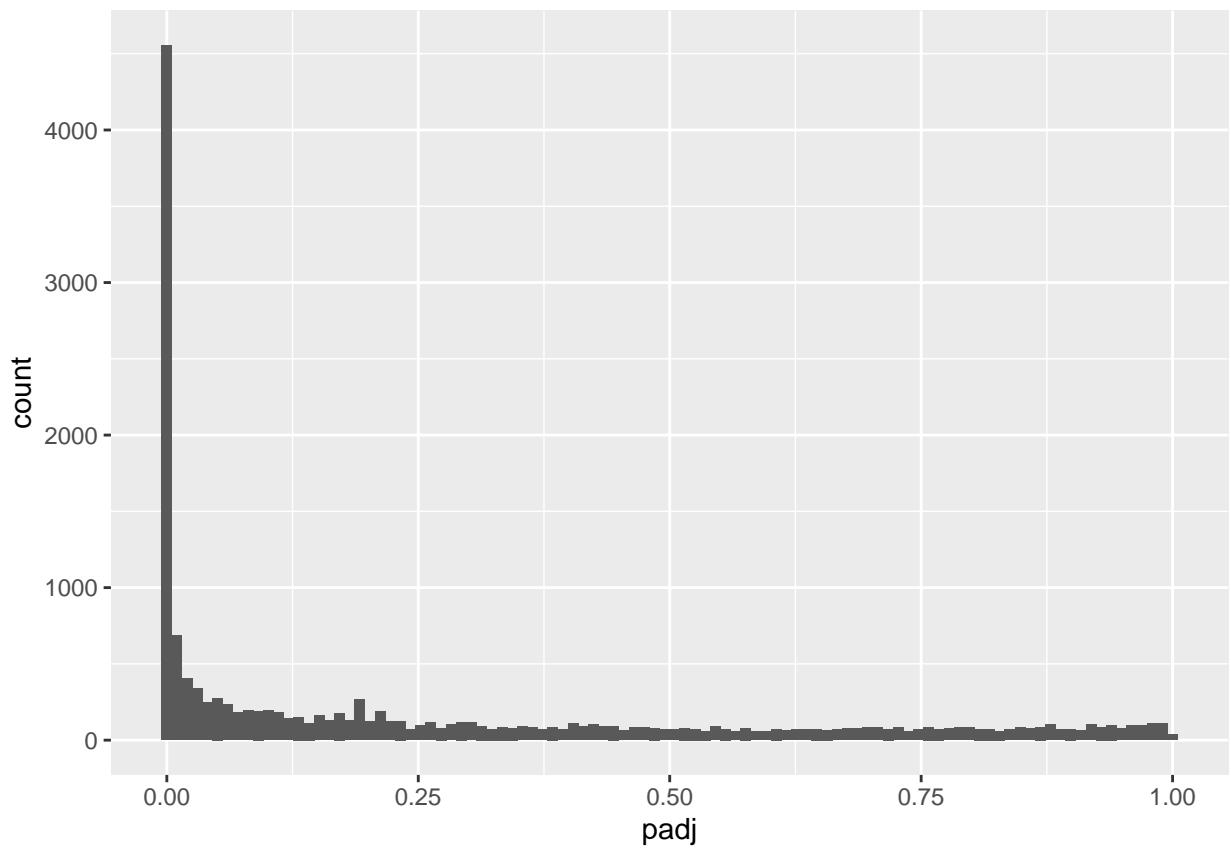
```
selected_lines = c('EBLST', 'CMP')

selected_samples <- colData %>%
  filter(group %in% selected_lines) %>%
  mutate(group = factor(group, levels = selected_lines)) %>%
  arrange(group) %>%
  pull(source_name)
```

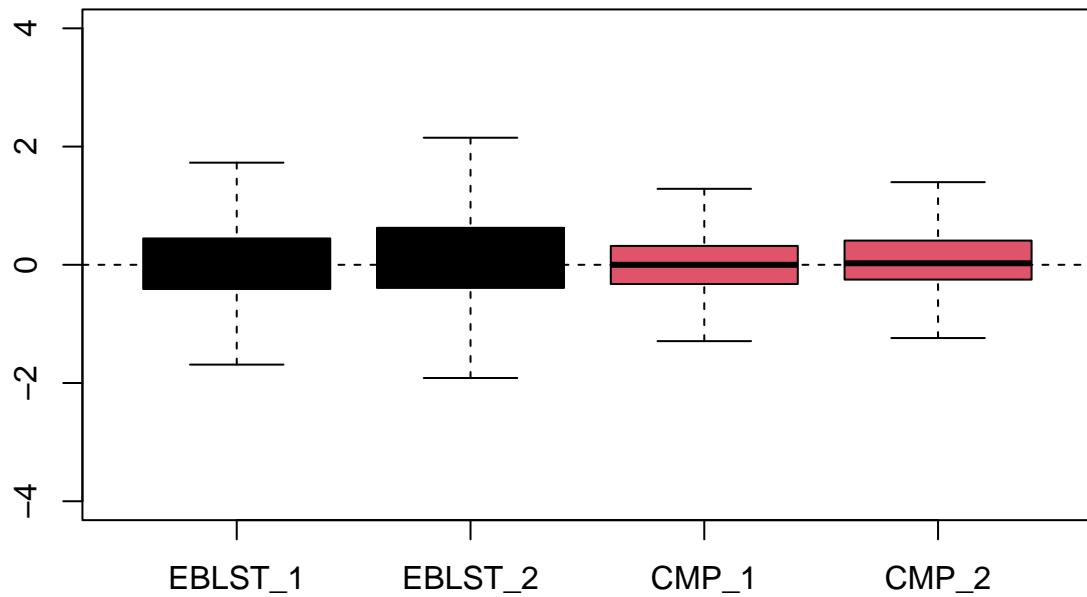
DESeq2

```
DEresults = get_DE_results(selected_lines, colData, countData)
```

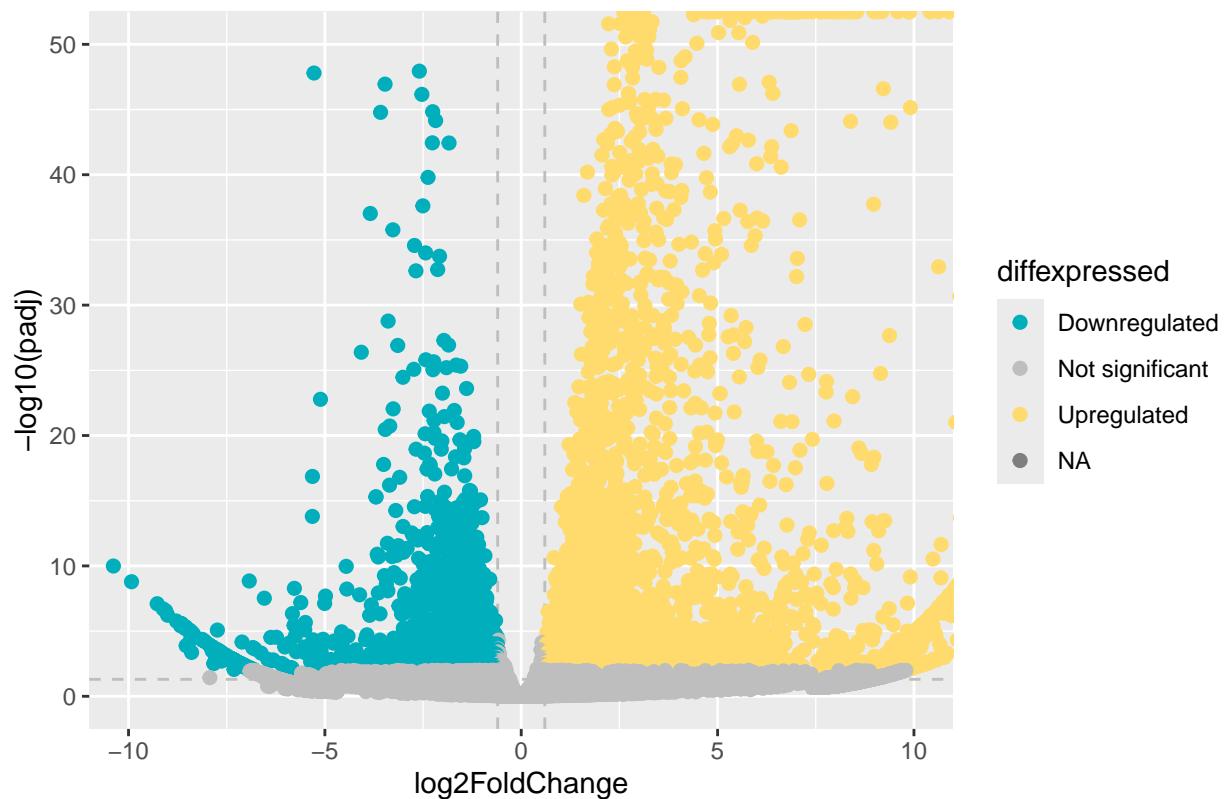




Normalized Counts



Volcano Plot: EBLST vs. CMP



```
DEresults_print <- DEresults[1:10,]
kable(DEresults_print)
```

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
ENSMUSG00000000804.10	37810.517	6.054046	0.1475706	41.02473	0	0
ENSMUSG00000001156.7	27425.312	6.888723	0.1459194	47.20909	0	0
ENSMUSG00000001763.10	15698.982	6.348545	0.1470736	43.16577	0	0
ENSMUSG00000004319.11	92964.401	6.827780	0.1300334	52.50788	0	0
ENSMUSG00000007379.11	9750.684	6.979139	0.1840664	37.91642	0	0
ENSMUSG00000007659.14	24299.334	8.405108	0.1775463	47.34037	0	0
ENSMUSG00000008730.13	66675.148	4.752642	0.1257476	37.79509	0	0
ENSMUSG00000011831.12	33542.644	6.972051	0.1751505	39.80607	0	0
ENSMUSG00000018909.11	37895.842	4.927321	0.1302236	37.83740	0	0
ENSMUSG00000019302.12	17869.222	6.189180	0.1357426	45.59496	0	0

GO term analysis

```
go_results = get_GO_results(DEresults)
kable(go_results$up_reg[1:10,c(7:11)])
```

precision	recall	term_id	source	term_name
0.5987718	0.1677820	GO:0005515	GO:MF	protein binding
0.6182670	0.1598478	GO:0005737	GO:CC	cytoplasm
0.3834788	0.1733192	GO:0048518	GO:BP	positive regulation of biological process
0.7823269	0.1400049	GO:0005488	GO:MF	binding
0.3277367	0.1819538	GO:0051179	GO:BP	localization
0.3408328	0.1767061	GO:1901564	GO:BP	organonitrogen compound metabolic process
0.3502351	0.1716025	GO:0048522	GO:BP	positive regulation of cellular process
0.6286098	0.1383899	GO:0065007	GO:BP	biological regulation
0.7731683	0.1327246	GO:0005622	GO:CC	intracellular anatomical structure
0.6111484	0.1384662	GO:0050789	GO:BP	regulation of biological process

```
kable(go_results$down_reg[1:10,c(7:11)])
```

precision	recall	term_id	source	term_name
0.3530864	0.1516437	GO: 0044271	GO:BP	cellular nitrogen compound biosynthetic process
0.6918519	0.1211833	GO: 0005737	GO: CC	cytoplasm
0.4205052	0.1223872	GO: 0097159	GO: MF	organic cyclic compound binding
0.2844444	0.1403167	GO: 0005829	GO: CC	cytosol
0.2750617	0.1405501	GO: 0018130	GO:BP	heterocycle biosynthetic process
0.2824691	0.1384653	GO: 1901362	GO:BP	organic cyclic compound biosynthetic process
0.2750617	0.1401963	GO: 0019438	GO:BP	aromatic compound biosynthetic process
0.3779099	0.1219824	GO: 0036094	GO: MF	small molecule binding
0.3665181	0.1229644	GO: 0043167	GO: MF	ion binding
0.2686420	0.1399177	GO: 0034654	GO:BP	nucleobase-containing compound biosynthetic process