

实验记录-2021.7.19

using GEO data to explore DEGs between autistic spectrum disorder patients and normal people

GSE15402

data source	GSE15402
title	Gene expression profiling differentiates autism case-controls and phenotypic variants of autism spectrum disorders
	Homo sapiens
experiment type	Expression profiling by array
status	Public on Dec 22, 2009

use GEO2R to analyze

以下为参数记录

p值调整方法: Benjamini & Hochberg

log转换方法: auto-detected

limma precision weights (vooma) : 不使用

强制归一化处理: 使用

组别及说明:

组别	说明
control-nonautistic	对照组 (非自闭症组)
sever-language-autistic	自闭症组-亚组 (严重程度: 高)
mild-autistic	自闭症组-亚组 (严重程度: 中)
savant-autistic	自闭症组-亚组 (严重程度: 低)


```

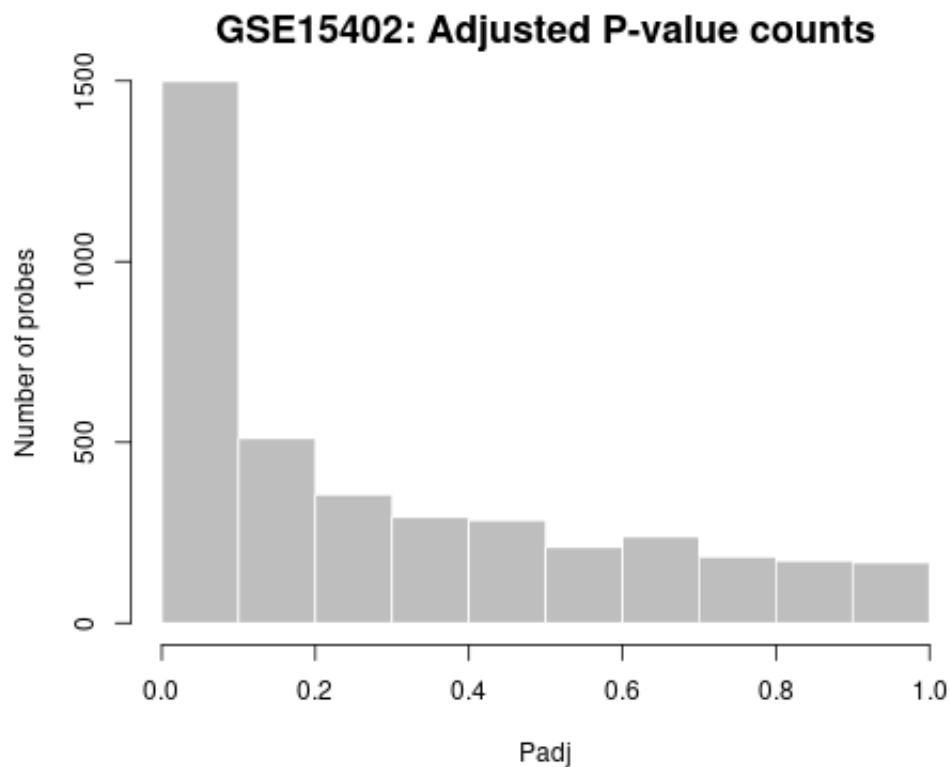
# set up contrasts of interest and recalculate model coefficients
cts <- c(paste(groups[1],"-",groups[2],sep=""), paste(groups[1],"-
",groups[3],sep=""), paste(groups[1],"-",groups[4],sep=""))
cont.matrix <- makeContrasts(contrasts=cts, levels=design)
fit2 <- contrasts.fit(fit, cont.matrix)

# compute statistics and table of top significant genes
fit2 <- eBayes(fit2, 0.01)
tT <- topTable(fit2, adjust="fdr", sort.by="B", number=250)

tT <- subset(tT,
select=c("ID", "adj.P.Val", "P.Value", "F", "Gene.symbol", "Gene.title"))
write.table(tT, file=stdout(), row.names=F, sep="\t")

# Visualize and quality control test results.
# Build histogram of P-values for all genes. Normal test
# assumption is that most genes are not differentially expressed.
tT2 <- topTable(fit2, adjust="fdr", sort.by="B", number=Inf)
hist(tT2$adj.P.Val, col = "grey", border = "white", xlab = "P-adj",
ylab = "Number of genes", main = "P-adj value distribution")

```



```

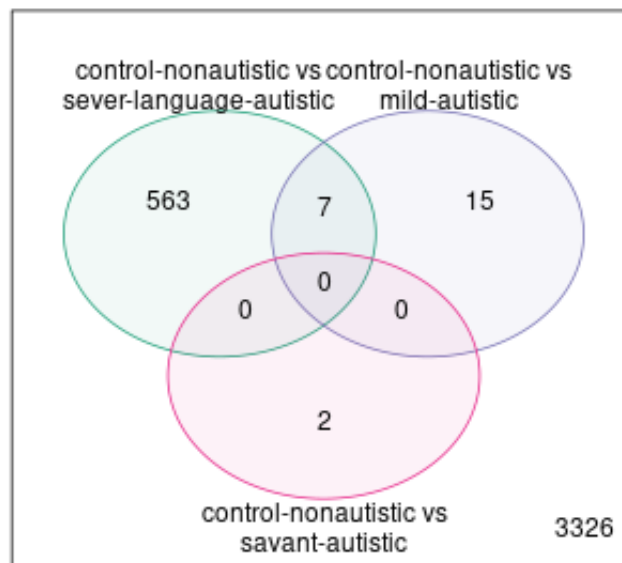
# summarize test results as "up", "down" or "not expressed"
dT <- decideTests(fit2, adjust.method="fdr", p.value=0.05)

# Venn diagram of results
vennDiagram(dT, circle.col=palette())

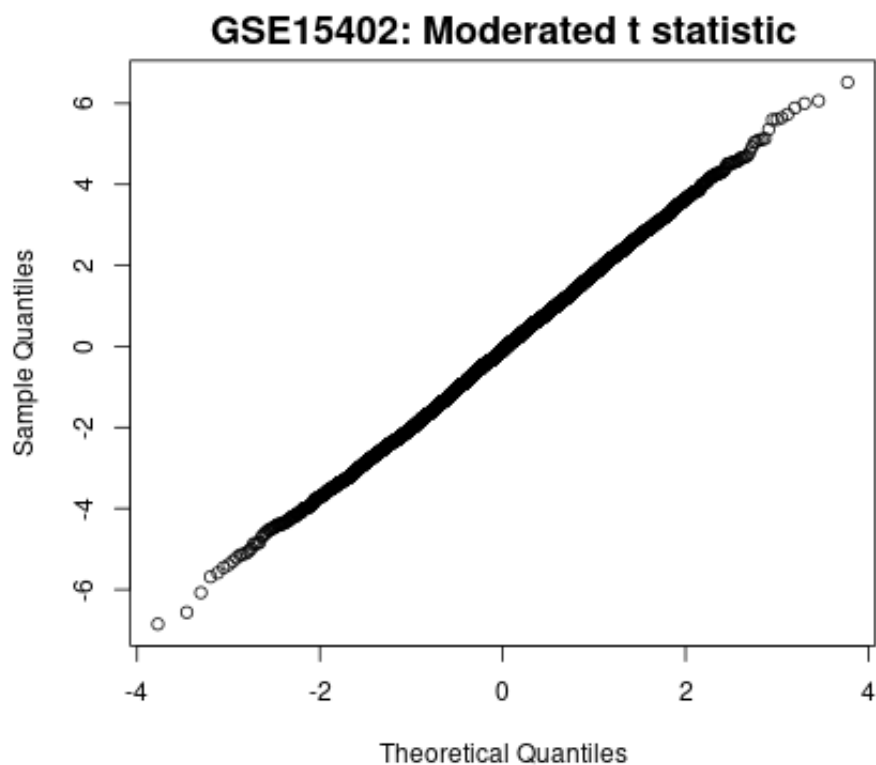
```

Venn Diagram

GSE15402: limma, Padj<0.05



```
# create Q-Q plot for t-statistic
t.good <- which(!is.na(fit2$F)) # filter out bad probes
qqt(fit2$t[t.good], fit2$df.total[t.good], main="Moderated t statistic")
```

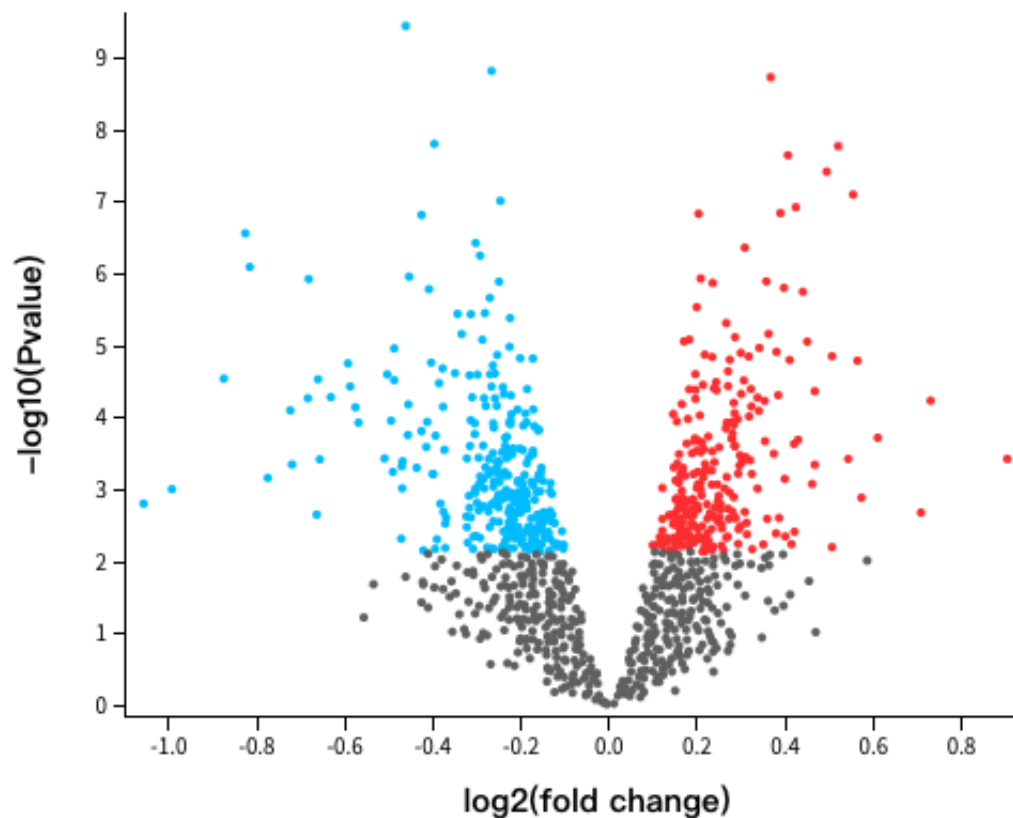


```
# volcano plot (log P-value vs log fold change)
colnames(fit2) # list contrast names
ct <- 1       # choose contrast of interest
volcanoplot(fit2, coef=ct, main=colnames(fit2)[ct], pch=20,
  highlight=length(which(dT[,ct]!=0)), names=rep('+', nrow(fit2)))
```

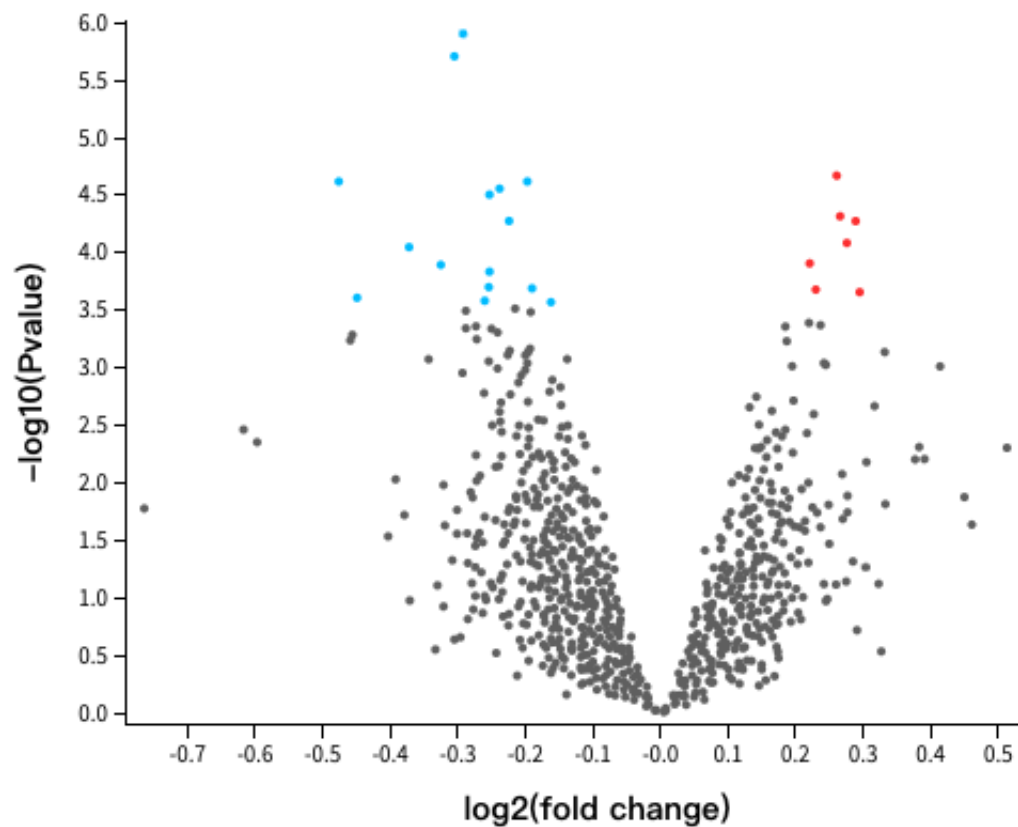
Volcano plot

GSE15402: Gene expression profiling differentiates autism...

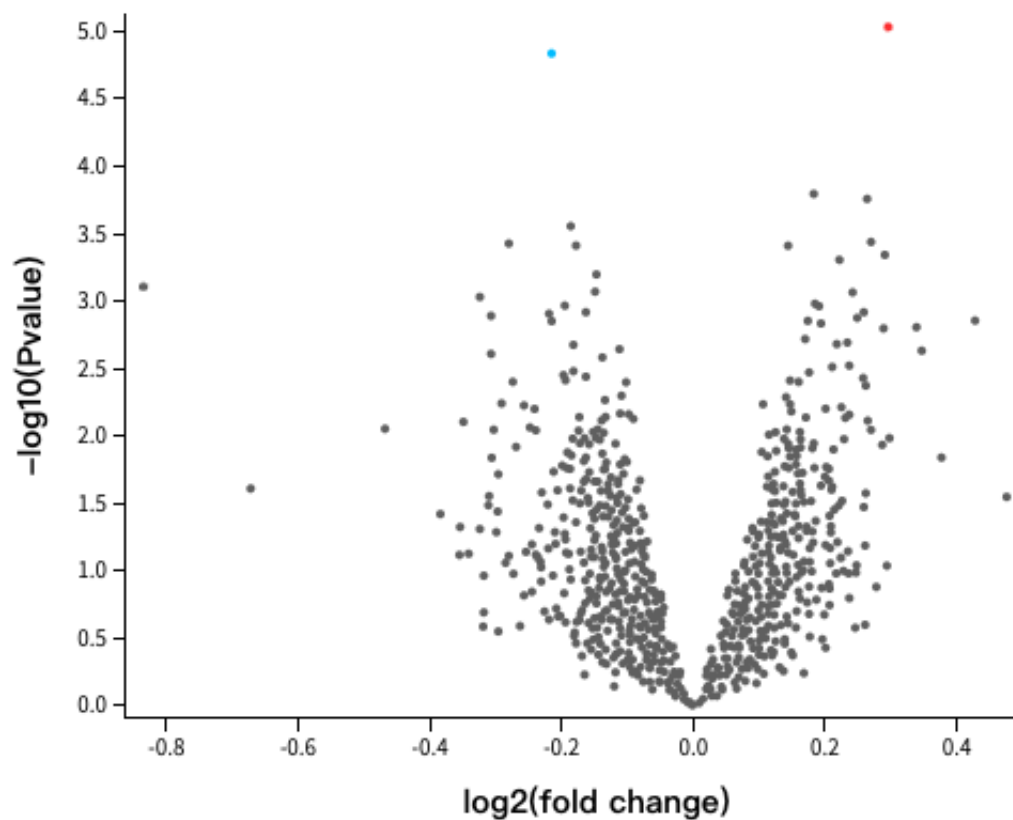
control-nonautistic vs sever-language-autistic,
Padj<0.05



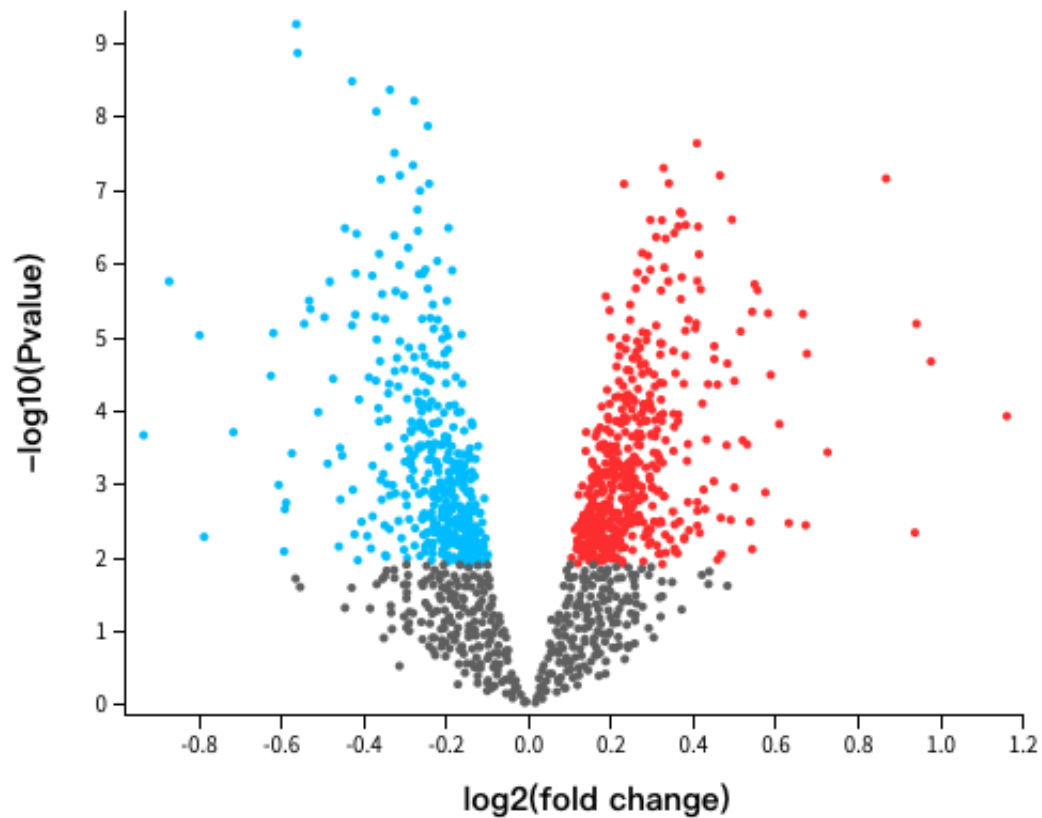
Volcano plot
GSE15402: Gene expression profiling
differentiates autism...
control-nonautistic vs mild-autistic, $P_{adj} < 0.05$



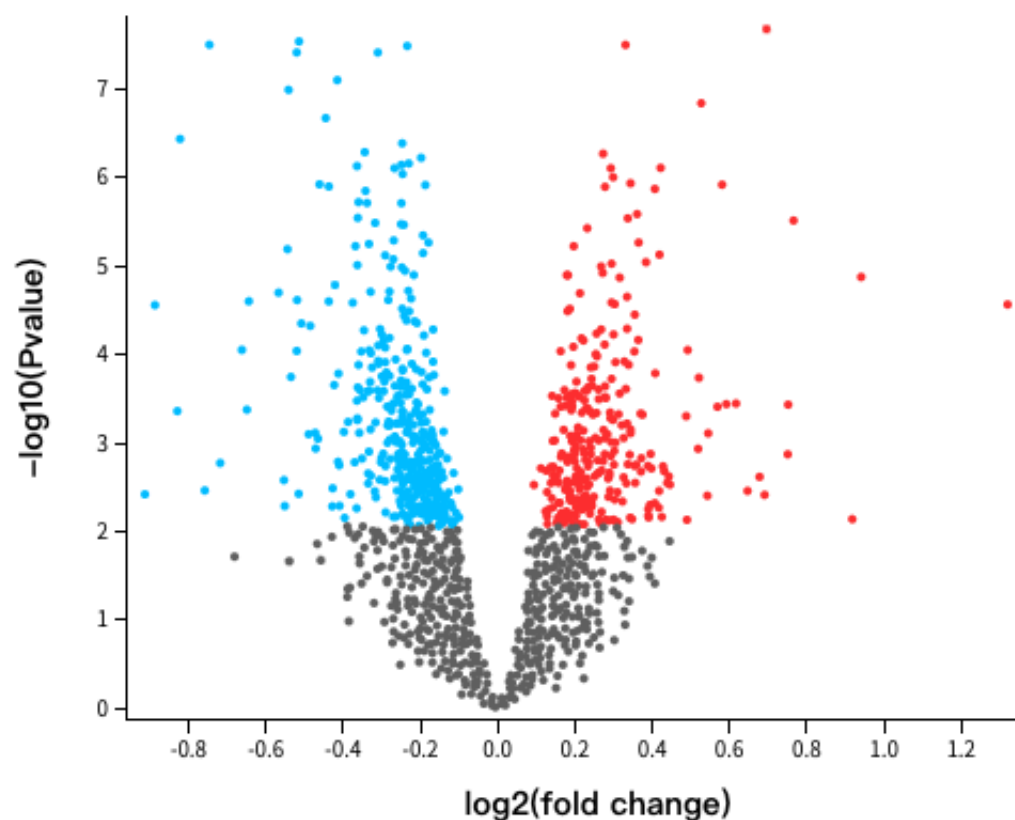
Volcano plot
GSE15402: Gene expression profiling
differentiates autism...
control–nonautistic vs savant–autistic, $P_{adj} < 0.05$



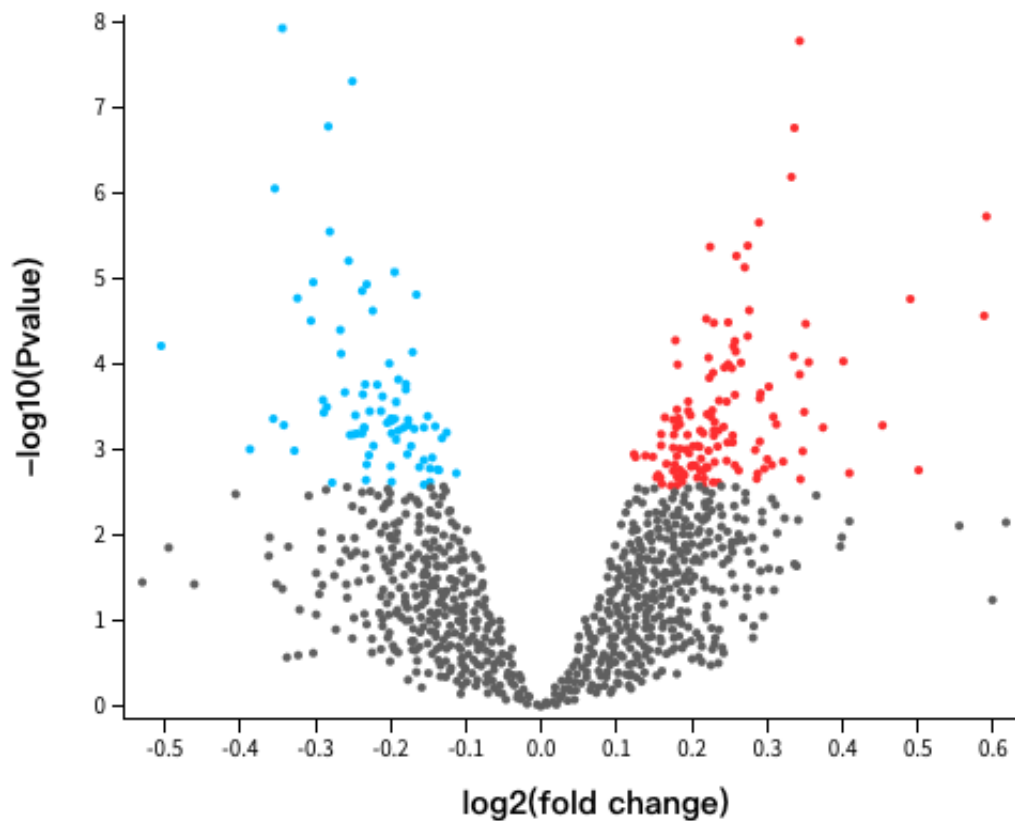
Volcano plot
GSE15402: Gene expression profiling
differentiates autism...
sever-language-autistic vs savant-autistic, $P_{adj} < 0.05$



Volcano plot
GSE15402: Gene expression profiling
differentiates autism...
sever-language-autistic vs mild-autistic, $P_{adj} < 0.05$

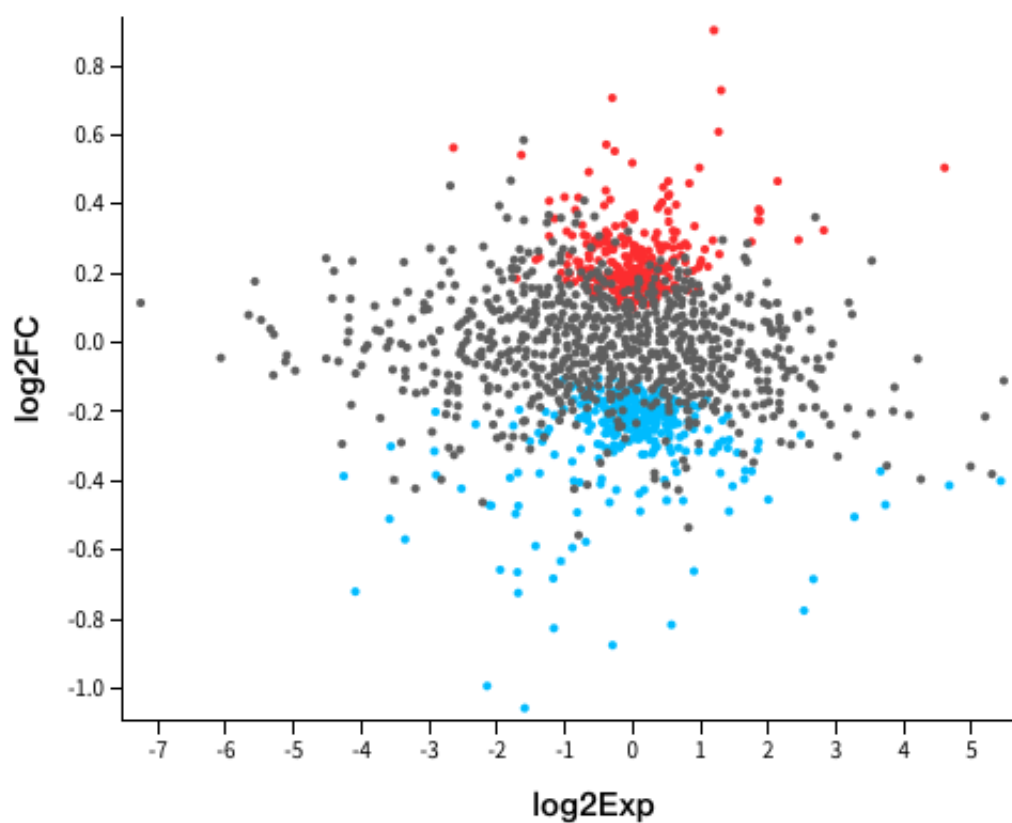


Volcano plot
GSE15402: Gene expression profiling
differentiates autism...
mild-autistic vs savant-autistic, $P_{adj} < 0.05$

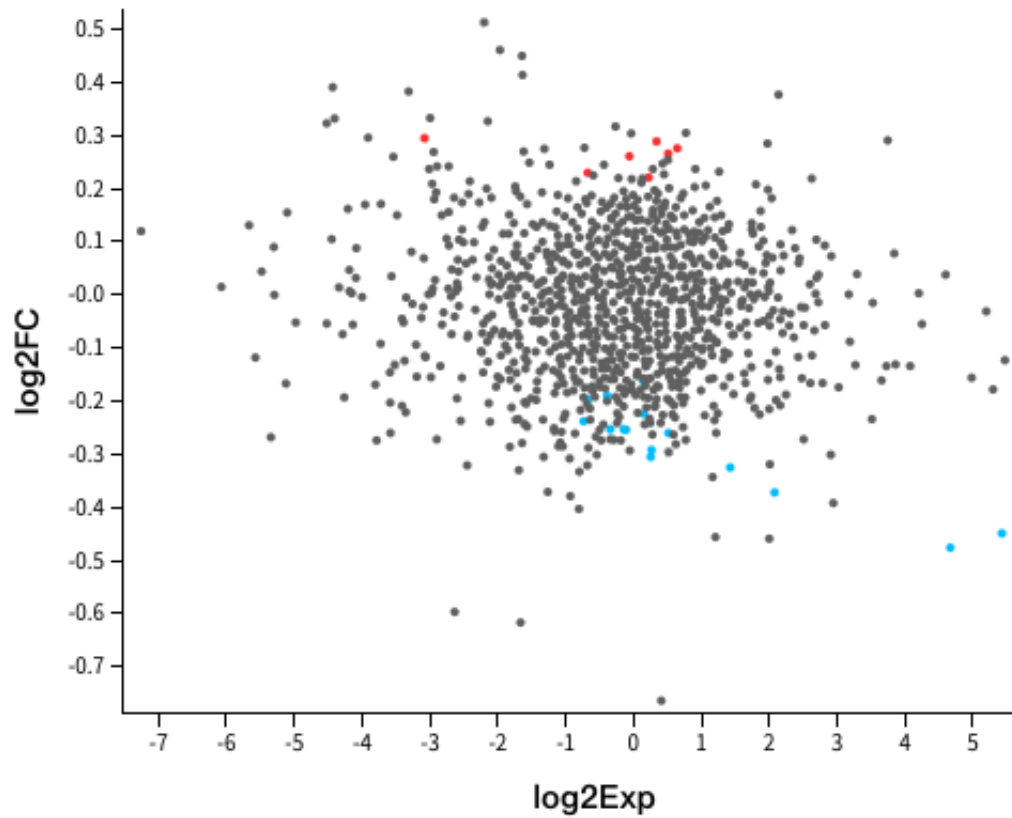


```
# MD plot (log fold change vs mean log expression)
# highlight statistically significant (p-adj < 0.05) probes
plotMD(fit2, column=ct, status=dT[,ct], legend=F, pch=20, cex=1)
abline(h=0)
```

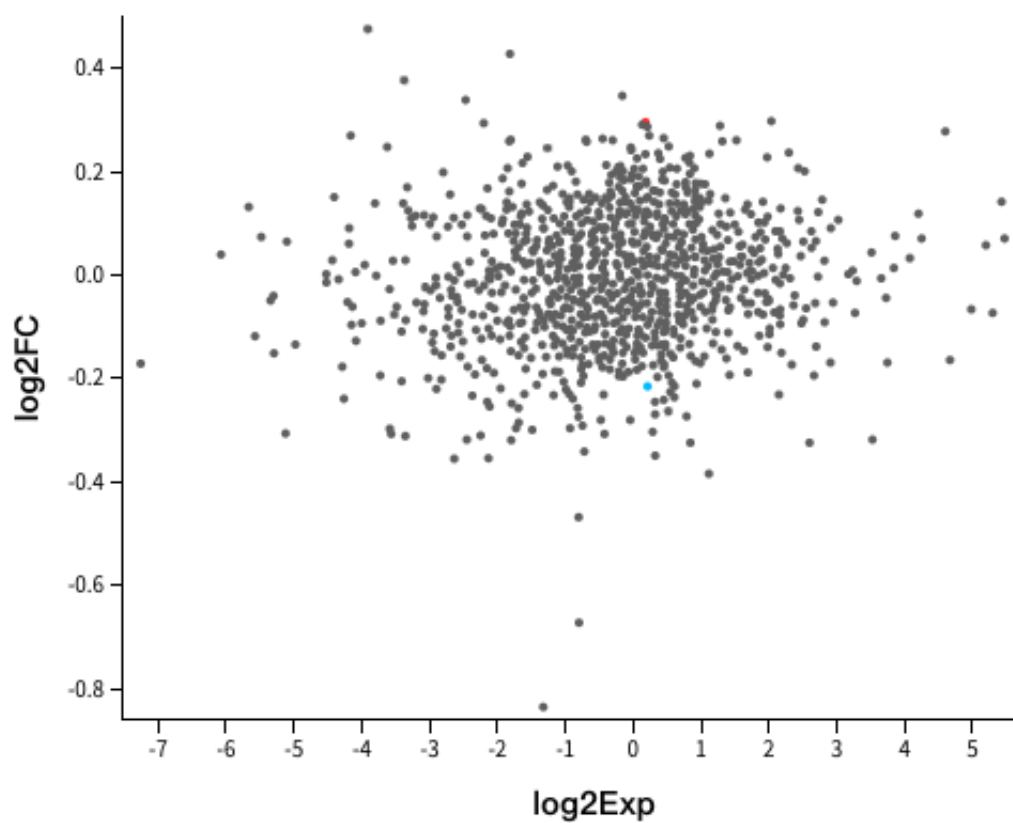
Meandiff plot
GSE15402: Gene expression profiling
differentiates autism...
control–nonautistic vs sever–language–autistic,
 $P_{adj} < 0.05$



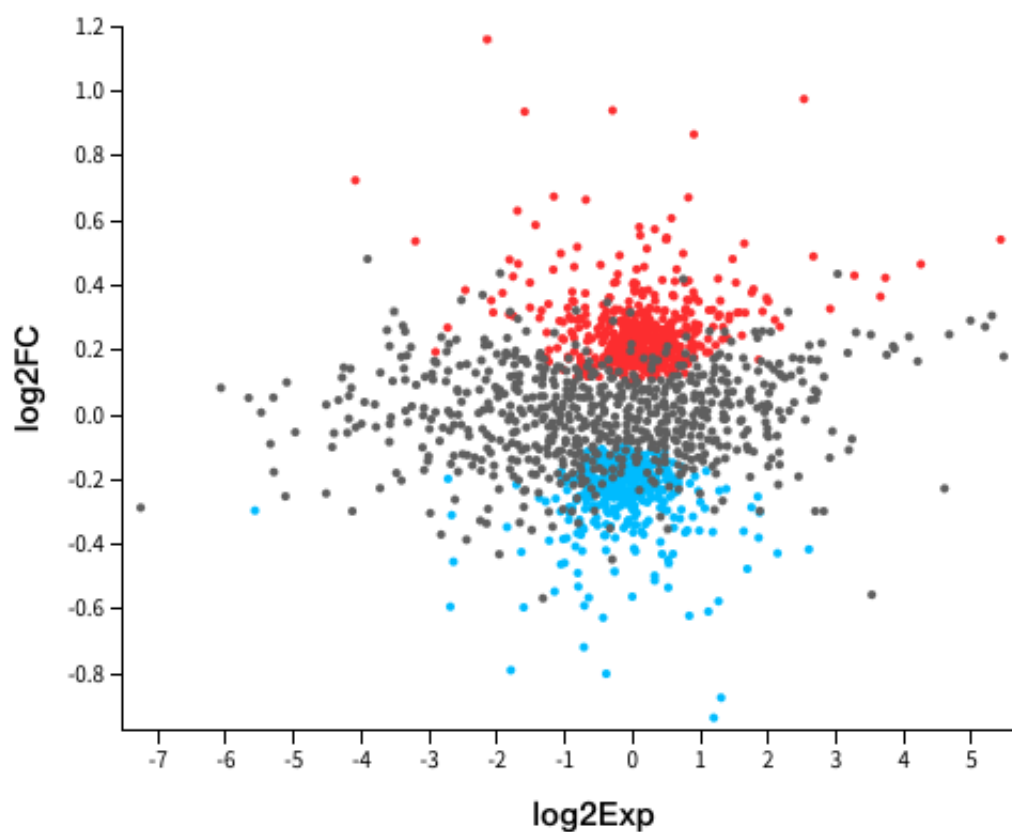
Meandiff plot
GSE15402: Gene expression profiling
differentiates autism...
control-nonautistic vs mild-autistic, $P_{adj} < 0.05$



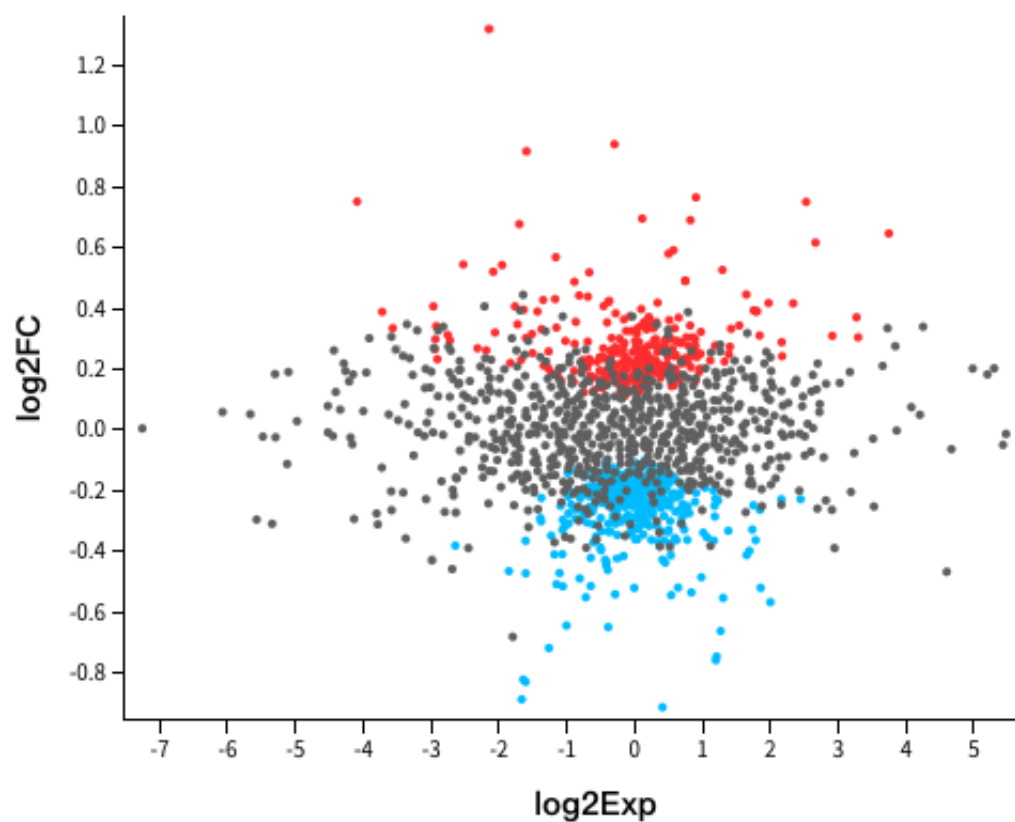
Meandiff plot
GSE15402: Gene expression profiling
differentiates autism...
control–nonautistic vs savant–autistic, Padj<0.05



Meandiff plot
GSE15402: Gene expression profiling
differentiates autism...
sever-language-autistic vs savant-autistic, $P_{adj} < 0.05$



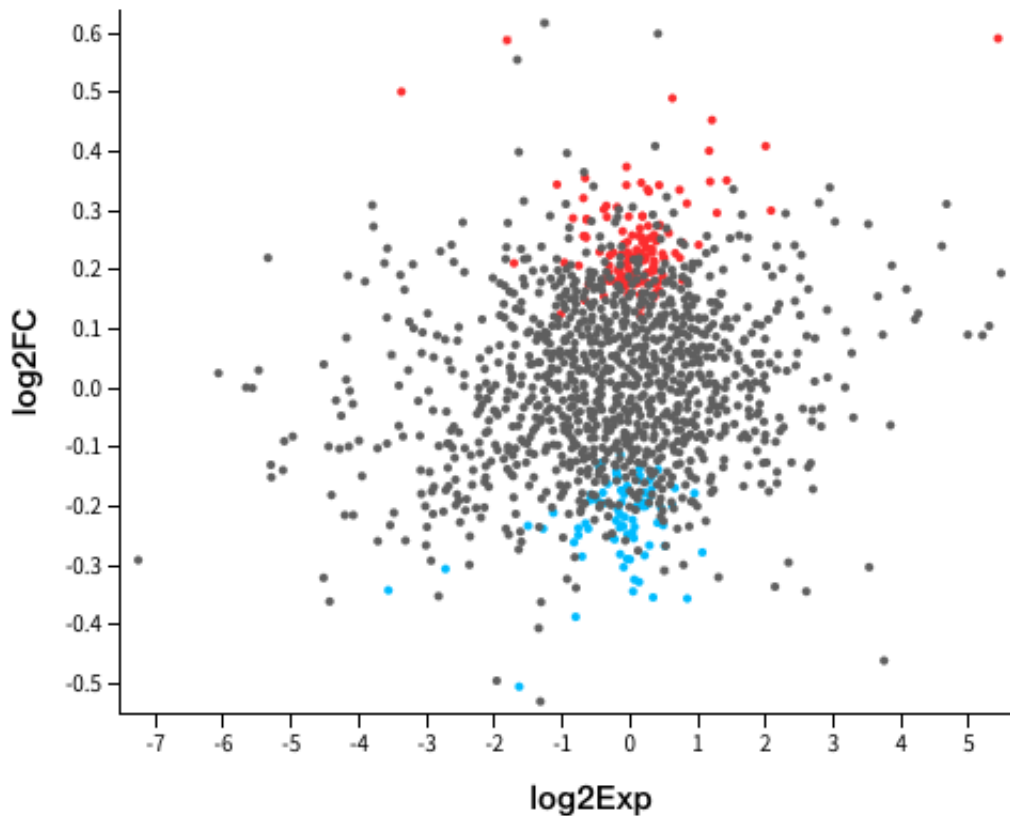
Meandiff plot
GSE15402: Gene expression profiling
differentiates autism...
sever-language-autistic vs mild-autistic, $P_{adj} < 0.05$



Meandiff plot

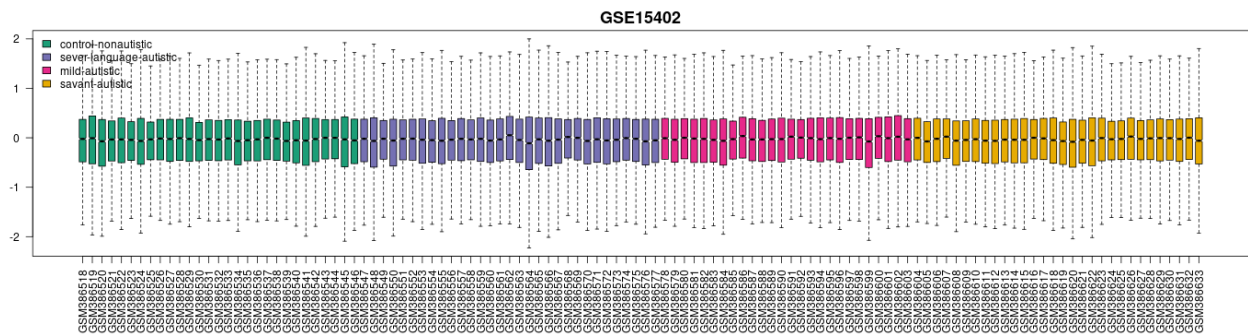
GSE15402: Gene expression profiling differentiates autism...

mild-autistic vs savant-autistic, Padj<0.05

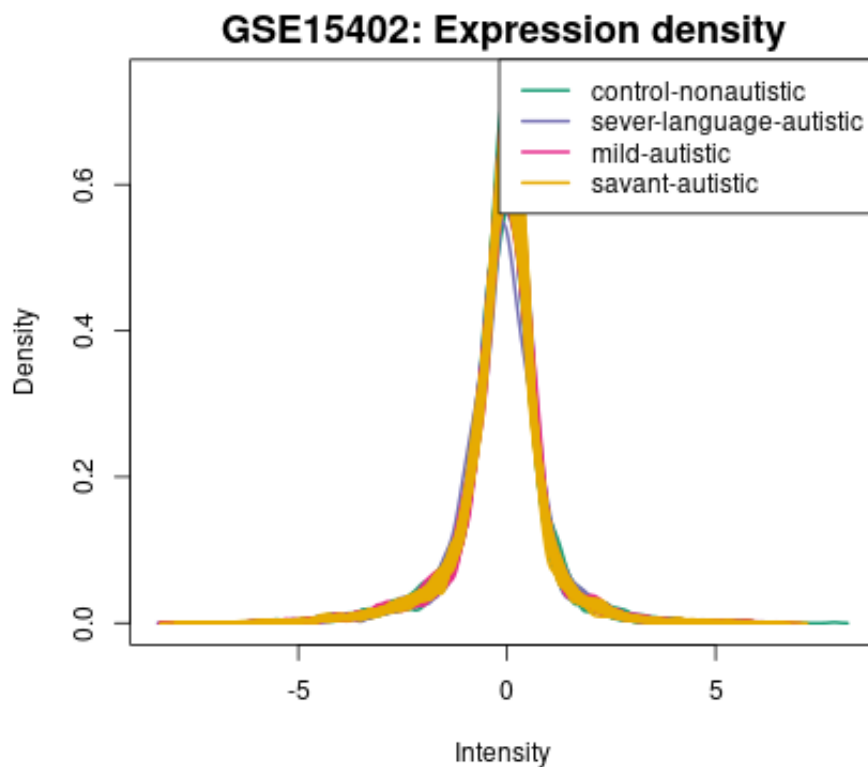


```
#####
# General expression data analysis
ex <- exprs(gset)

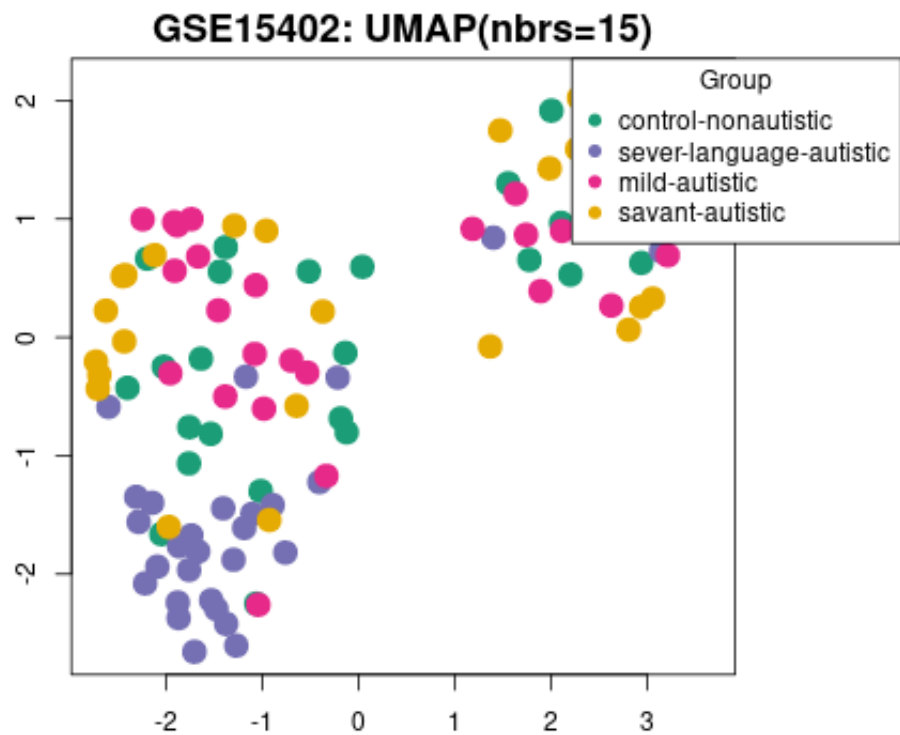
# box-and-whisker plot
dev.new(width=3+ncol(gset)/6, height=5)
ord <- order(gs) # order samples by group
palette(c("#1B9E77", "#7570B3", "#E7298A", "#E6AB02", "#D95F02",
          "#66A61E", "#A6761D", "#B32424", "#B324B3", "#666666"))
par(mar=c(7,4,2,1))
title <- paste ("GSE15402", "/", annotation(gset), sep = "")
boxplot(ex[,ord], boxwex=0.6, notch=T, main=title, outline=FALSE, las=2,
col=gs[ord])
legend("topleft", groups, fill=palette(), bty="n")
dev.off()
```

```
# expression value distribution
par(mar=c(4,4,2,1))
title <- paste ("GSE15402", "/", annotation(gset), " value distribution", sep
="")
plotDensities(ex, group=gs, main=title, legend ="topright")
```



```
# UMAP plot (dimensionality reduction)
ex <- na.omit(ex) # eliminate rows with NAs
ex <- ex[!duplicated(ex), ] # remove duplicates
ump <- umap(t(ex), n_neighbors = 15, random_state = 123)
par(mar=c(3,3,2,6), xpd=TRUE)
plot(ump$layout, main="UMAP plot, nbrs=15", xlab="", ylab="", col=gs, pch=20,
cex=1.5)
legend("topright", inset=c(-0.15,0), legend=levels(gs), pch=20,
col=1:nlevels(gs), title="Group", pt.cex=1.5)
library("maptools") # point labels without overlaps
pointLabel(ump$layout, labels = rownames(ump$layout), method="SANN", cex=0.6)
```



```
# mean-variance trend, helps to see if precision weights are needed
plotSA(fit2, main="Mean variance trend, GSE15402")
```

