Reporte

Armando_Trapaga

2025-02-01

Contents

Impact of Lack of Glutamatergic Activation on Molecular Heterogeneity in Spiral Ganglion Neurons in the Mouse Cochlea

Data processing

Libraries

```
library("recount3")
library("edgeR")
library("ggplot2")
library("limma")
library("pheatmap")
library("RColorBrewer")
```

Creating RangedSummarizedExperiment object

The data was recovered from the study *"Single-cell RNA-seq of spiral ganglion neurons from wildtype and Vglut3-/- mice"* in recount3. The study investigates the molecular heterogeneity of spiral ganglion neurons in the mouse cochlea, comparing two genetic backgrounds: **wildtype and Vglut3-/-**. Individual neurons were analyzed using the Smart-seq2 approach and the NextSeq platform, with bioinformatic analysis conducted in R.

```
data <- recount3::create_rse_manual(
     project = "SRP149148",
     project_home = "data_sources/sra",
     organism = "mouse",
     annotation = "gencode_v23",
     type = "gene"
)
assay(data, "counts") <- compute_read_counts(data)
```

Data wrangling

Before analyzing the data, it is necessary to restructure it, correct any incorrect values, and transform variables. These steps must be completed prior to starting the analysis.

```
data <- expand_sra_attributes(data)

colData(data)[
   ,
   grepl("^sra_attribute", colnames(colData(data)))
 ]
```

```
## DataFrame with 226 rows and 4 columns
##            sra_attribute.age sra_attribute.genotype sra_attribute.source_name
##                  <character>            <character>               <character>
## SRR7227668               P26               Wildtype    Spiral ganglion neuron
## SRR7227677               P27               Wildtype    Spiral ganglion neuron
## SRR7227679               P27               Wildtype    Spiral ganglion neuron
## SRR7227681               P27               Wildtype    Spiral ganglion neuron
## SRR7227682               P27               Wildtype    Spiral ganglion neuron
## ...                      ...                    ...                       ...
## SRR7227771               P26               Wildtype    Spiral ganglion neuron
## SRR7227772               P26               Wildtype    Spiral ganglion neuron
## SRR7227773               P26               Wildtype    Spiral ganglion neuron
## SRR7227774               P26               Wildtype    Spiral ganglion neuron
## SRR7227775               P27              Vglut3-/-    Spiral ganglion neuron
##            sra_attribute.tonotopic_location
```

```
##                                 <character>
## SRR7227668                             Base
## SRR7227677                             Apex
## SRR7227679                             Apex
## SRR7227681                             Apex
## SRR7227682                             Apex
## ...                                     ...
## SRR7227771                             Base
## SRR7227772                             Base
## SRR7227773                             Base
## SRR7227774                             Base
## SRR7227775                           Middle
```

```
data$sra_attribute.age  <- as.factor(data$sra_attribute.age)

data$sra_attribute.genotype <- as.factor(data$sra_attribute.genotype)

data$sra_attribute.tonotopic_location <- as.factor(data$sra_attribute.tonotopic_location)
```

```
summary(as.data.frame(colData(data)[
    ,
    grepl("^sra_attribute.[age|genotype|tonotopic_location]", colnames(colData(data)))
]))
```

```
##  sra_attribute.age sra_attribute.genotype sra_attribute.tonotopic_location
##  P25: 14           Vglut3-/-: 40          Apex  : 56
##  P26:109           Wildtype :186          Base  : 42
##  P27:103                                  Middle:128
```

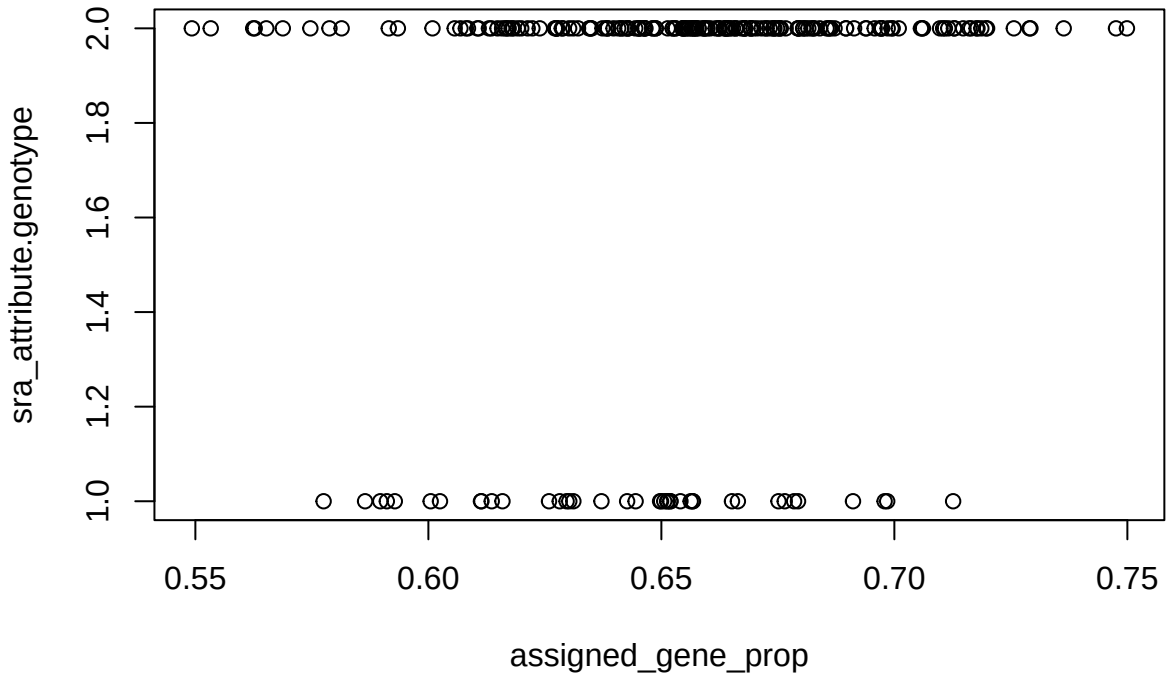
```
data$assigned_gene_prop <- data$recount_qc.gene_fc_count_all.assigned / data$recount_qc.gene_fc_count_al

summary(data$assigned_gene_prop)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.5492  0.6314  0.6567  0.6555  0.6794  0.7499
```

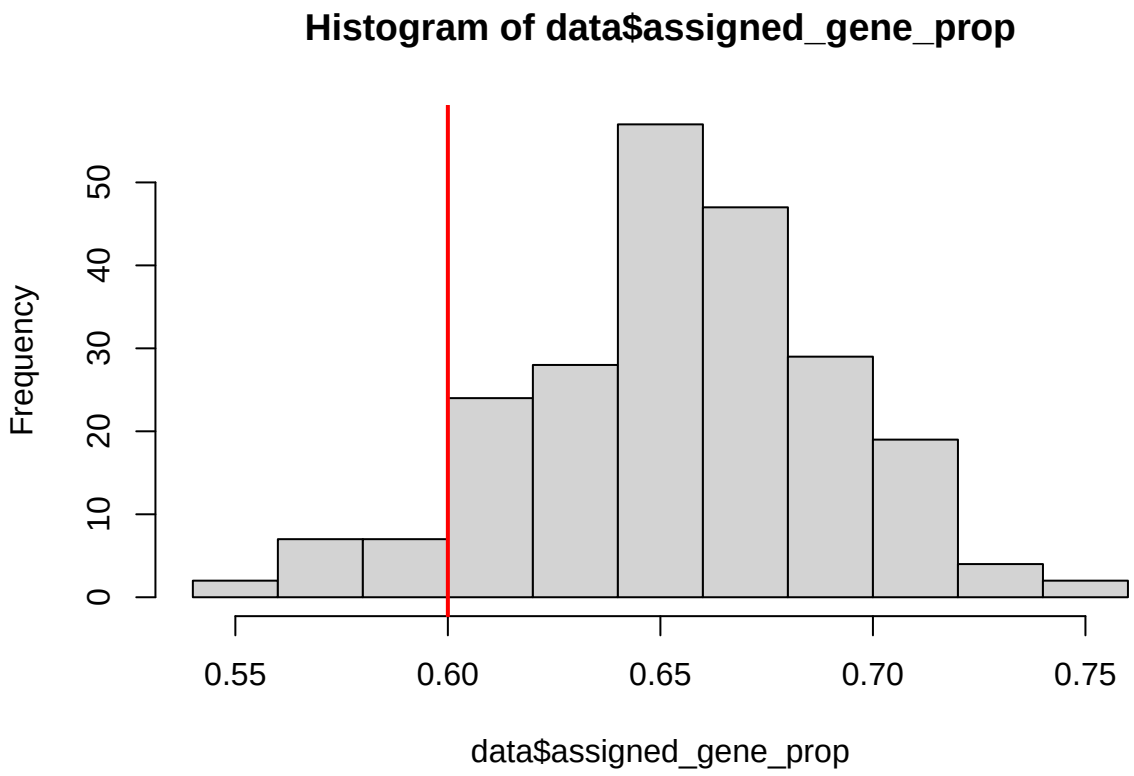
Visualizing Transformed Data

To understand the nature of the data, we need to visualize it to determine if filtering is necessary.

```
with(colData(data), plot(assigned_gene_prop, sra_attribute.genotype))
```




```
hist(data$assigned_gene_prop)
abline(v=0.6, col="red", lwd=2)
```




```
table(data$assigned_gene_prop < 0.6)
```

```
##
## FALSE  TRUE
##   210    16
```

```
table(data$assigned_gene_prop < 0.6)
```

```
##
## FALSE  TRUE
##   210    16
```

Data filtering

After visualizing the data, we observed a set of genes with poor quality compared to others. Therefore, we need to filter these genes.

```
unfiltered_data <- data

data <- data[,data$assigned_gene_prop > 0.6]

data <- data[edgeR::filterByExpr(assay(data,"counts")), ]
```

```
## No group or design set. Assuming all samples belong to one group.
```

After filter we have:

```
## [1] 7269  210
```

```
## [1] 13.12
```

Data Normalize

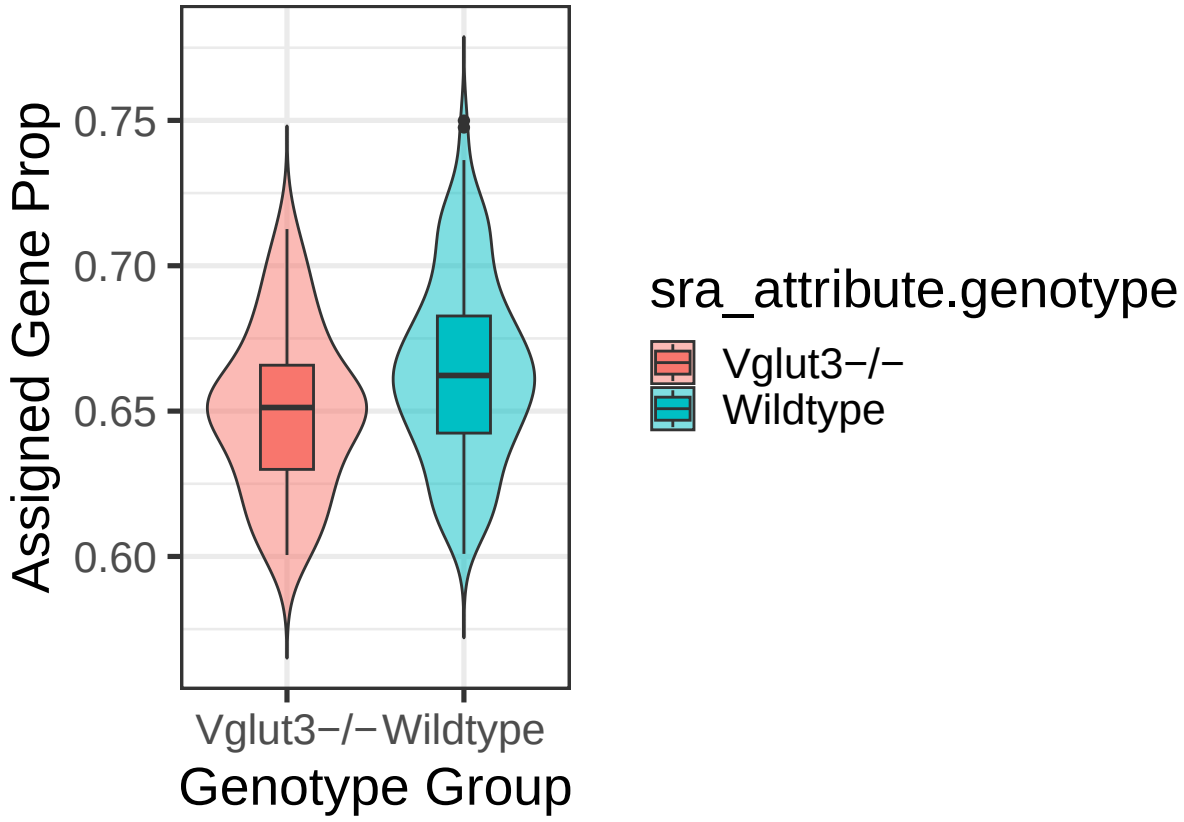
Normalizing RNAseq data is essential for obtaining accurate and comparable results. During sequencing, technical variations, such as differences in sequencing depth between samples or batch effects, can introduce **biases**. Normalization corrects these technical variations, allowing observed differences in gene expression to reflect true biological variations.

```
dge <- DGEList(
    counts = assay(data, "counts"),
    genes = rowData(data)
)
dge <- calcNormFactors(dge)
```

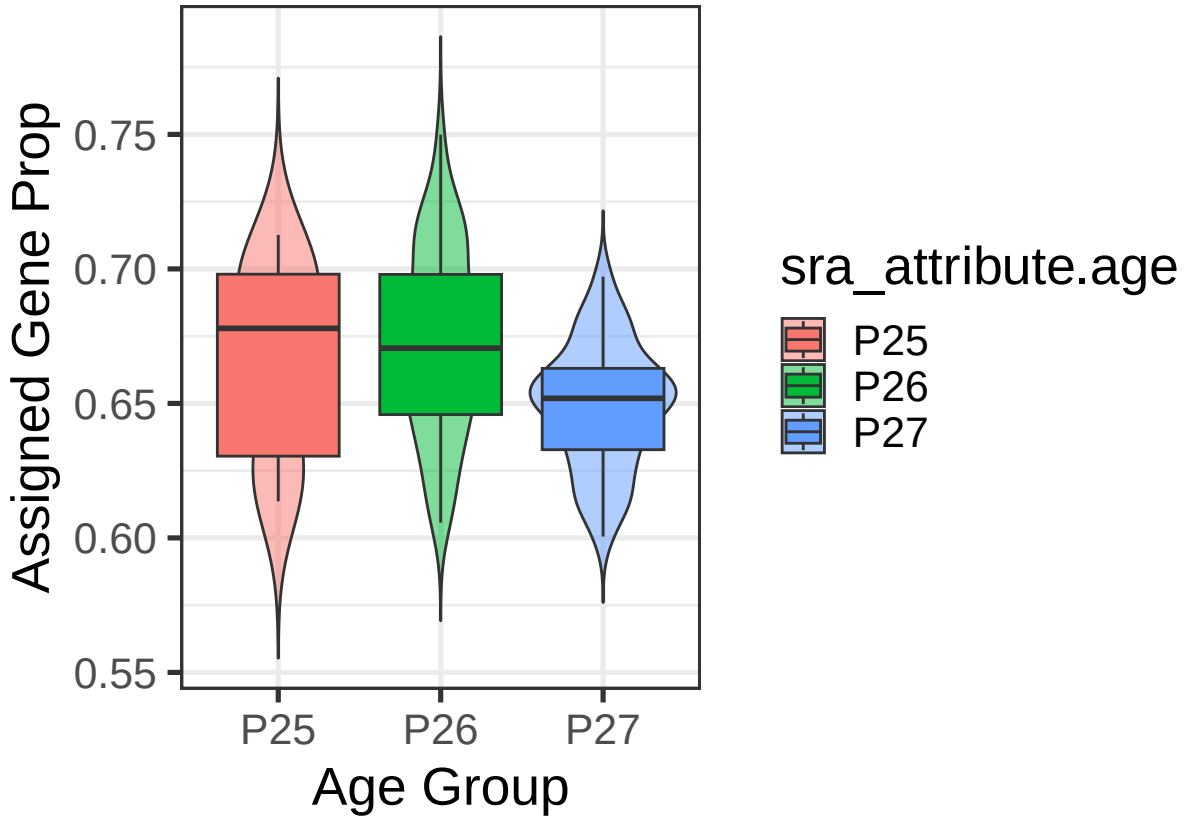
Expression data analysis

Data distribution

```
ggplot(as.data.frame(colData(data)), aes(y = assigned_gene_prop, x = sra_attribute.genotype, fill=sra_a
    geom_violin(trim = FALSE, alpha= 0.5 ) +
    geom_boxplot(width = 0.3) +
    theme_bw(base_size = 20) +
    ylab("Assigned Gene Prop") +
    xlab("Genotype Group")
```




```
ggplot(as.data.frame(colData(data)), aes(y = assigned_gene_prop, x = sra_attribute.age, fill=sra_attribu
    geom_violin(trim = FALSE, alpha= 0.5 ) +
    geom_boxplot() +
    theme_bw(base_size = 20) +
    ylab("Assigned Gene Prop") +
    xlab("Age Group")
```



Statistical model

To continue its necesary to create a statistical model

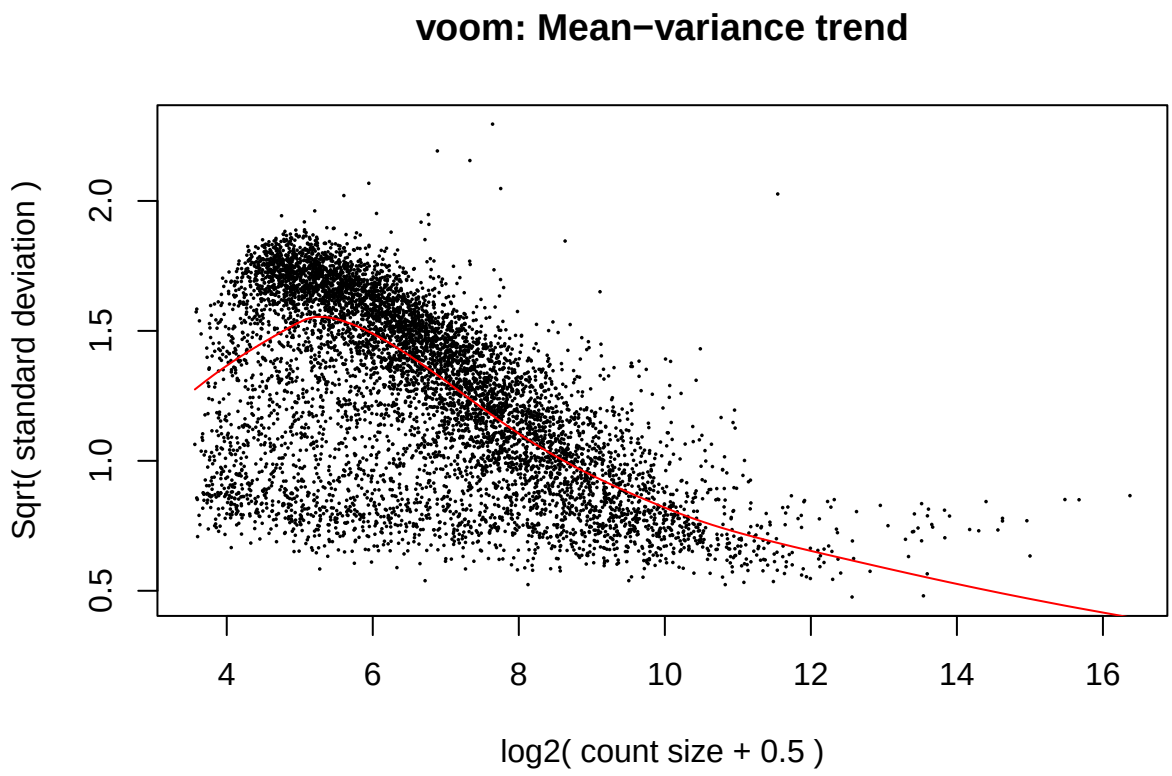
```
mod <- model.matrix(~ sra_attribute.genotype + sra_attribute.age + sra_attribute.tonotopic_location + as
    data = colData(data)
)
```

```
## [1] "(Intercept)"
## [2] "sra_attribute.genotypeWildtype"
## [3] "sra_attribute.ageP26"
## [4] "sra_attribute.ageP27"
## [5] "sra_attribute.tonotopic_locationBase"
## [6] "sra_attribute.tonotopic_locationMiddle"
## [7] "assigned_gene_prop"
```

Performing expression data analysis

When we have the model, its time to perform an empirical Bayesian analysis after our gene expression analysis to construct a full statiscal analysis of the results, we use *"sra_attribute.genotypeWildtype"* as our coeficient for the analysis.

```
vGene <- voom(dge, mod, plot = TRUE)
```

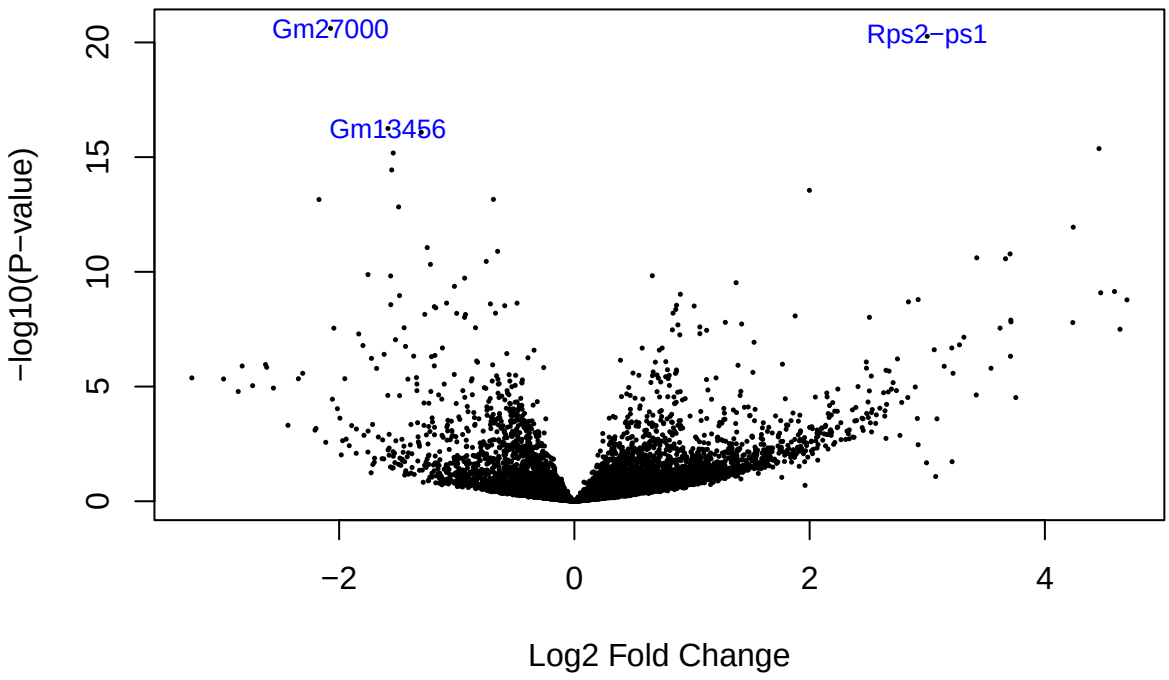



```
eb_results <- eBayes(lmFit(vGene))

results <- topTable(
    eb_results,
    coef = 2,
    number = nrow(data),
    sort.by = "none"
)
dim(results)
```

```
## [1] 7269   17
```

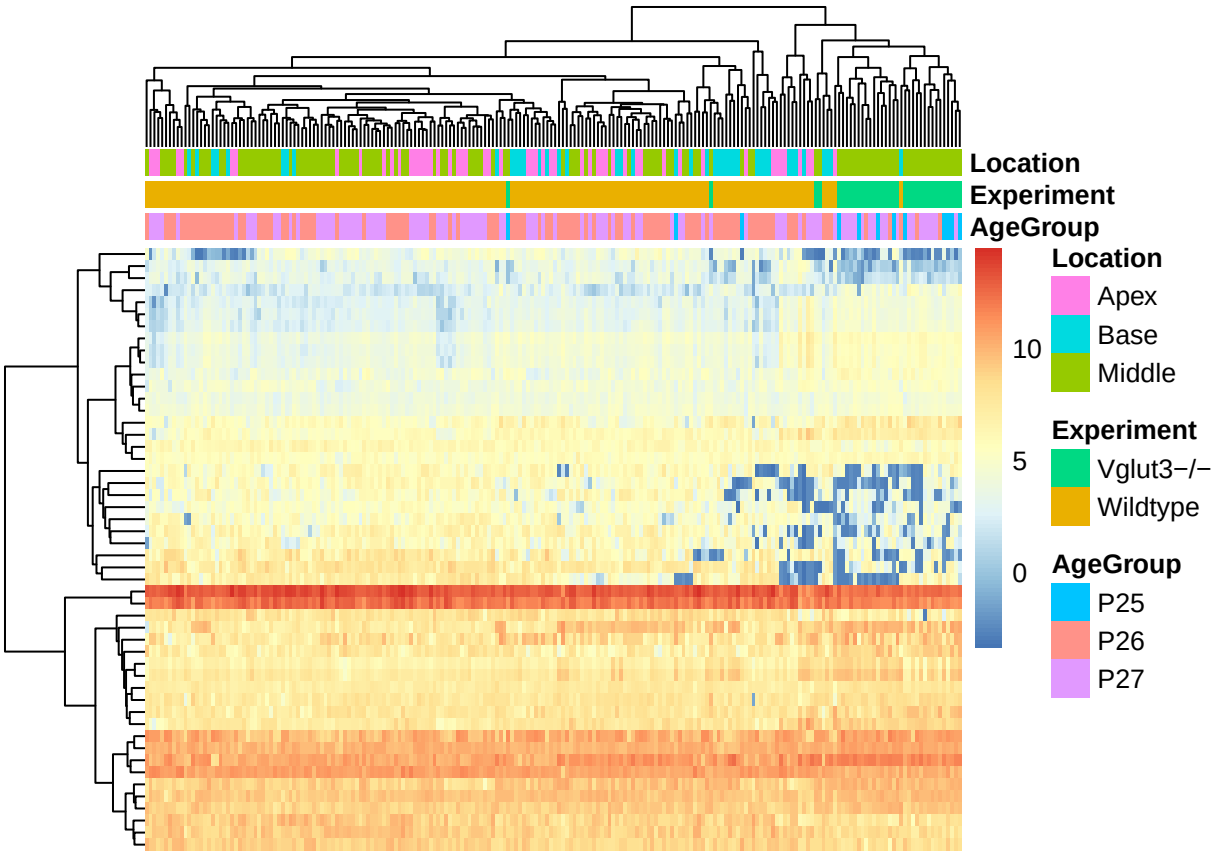
Visualizing expression data

```
volcanoplot(eb_results, coef = 2, highlight = 3, names = results$gene_name)
```



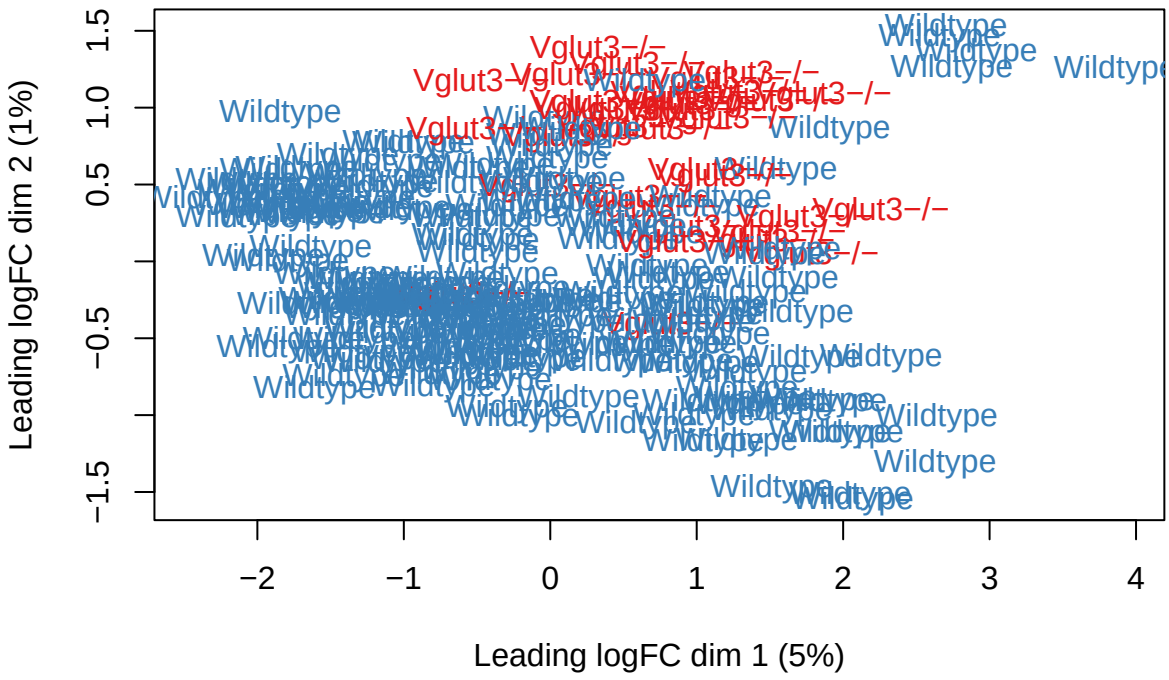
Visualizing Data Clusters

```
exprs_heatmap <- vGene$E[rank(results$adj.P.Val) <= 50, ]

df <- as.data.frame(colData(data)[, c("sra_attribute.age", "sra_attribute.genotype", "sra_attribute.tone
colnames(df) <- c("AgeGroup", "Experiment", "Location")

pheatmap(
    exprs_heatmap,
    cluster_rows = TRUE,
    cluster_cols = TRUE,
    show_rownames = FALSE,
    show_colnames = FALSE,
    annotation_col = df
)
```

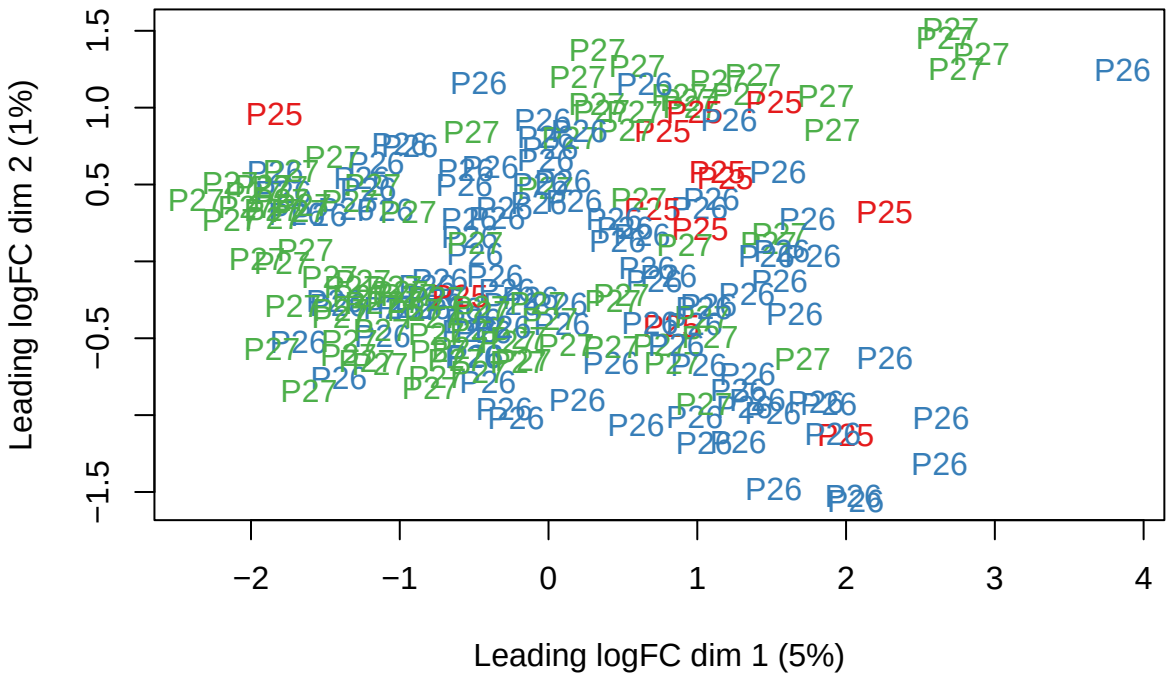


Visualizing multidimensional scaling

```
col.group <- df$Experiment
levels(col.group) <- brewer.pal(nlevels(col.group), "Set1")

col.group <- as.character(col.group)


plotMDS(vGene$E, labels = df$Experiment, col = col.group)
```




```
col.group <- df$AgeGroup
levels(col.group) <- brewer.pal(nlevels(col.group), "Set1")

col.group <- as.character(col.group)

plotMDS(vGene$E, labels = df$AgeGroup, col = col.group)
```



Results

In this study two genotypes were compared, the wildtype and Vglut3-/-, which lacks the vesicular glutamate transporter 3 (VGLUT3), essential for excitatory neurotransmission in these neurons. The results indicate that the absence of Vglut3 causes significant changes in the gene expression of spiral ganglion neurons, affecting their activity and possibly their auditory function(Shrestha et al. 2018). Specifically, the genes Gm27000, Rps2-ps1, and Gm13456 showed differential expression:

- Gm27000 and Gm13456 may be involved in mechanisms of synaptic plasticity and adaptive response to the lack of excitatory neurotransmission or maybe be realted to synaptic stability or neuronal differentiation(Di Fruscio et al. 1998).
- Rps2-ps1, a pseudogene related to ribosomal protein synthesis, may be altering protein translation and neuronal homeostasis(Harris, Jolivet, and Attwell 2012).

Therefore, we can conclude that these changes were primarily observed as a result of the genotype and not due to age or location.

These findings suggest that Vglut3 is not only crucial for neurotransmission in the spiral ganglion but also regulates the expression of key genes for neuronal function. Its absence could affect synaptic connectivity and the ability of these neurons to process auditory signals, which would have implications for sound perception and neuronal plasticity in the cochlea.

References

Di Fruscio, M, C A Gilchrist, R T Baker, and D A Gray. 1998. "Genomic Structure of Unp, a Murine Gene Encoding a Ubiquitin-Specific Protease." *Biochimica Et Biophysica Acta* 1398 (1): 9–17. https://doi.org/10.1016/s0167-4781(98)00035-9.

Harris, Julia J, Renaud Jolivet, and David Attwell. 2012. "Synaptic Energy Use and Supply." *Neuron* 75 (5): 762–77. https://doi.org/10.1016/j.neuron.2012.08.019.

Shrestha, Brikha R, Chester Chia, Lorna Wu, Sharon G Kujawa, M Charles Liberman, and Lisa V Goodrich. 2018. "Sensory Neuron Diversity in the Inner Ear Is Shaped by Activity." *Cell* 174 (5): 1229–1246.e17. https://doi.org/10.1016/j.cell.2018.07.007.