Differential Gene Expression

Ioana-Beatrice Neamtu

ANALYSIS OF MEDIA vs LPS (Lipopolysaccharide) TREATED MONOCYTES using the Bioconductor *DESeq2* Vignette (Love *et al.,* 2024, link (https://www.bioconductor.org/packages/release/bioc/vignettes/DESeq2/inst/doc/DESeq2.html))

Building DESeqDataSet Object and Count Normalization

Bioconductor **DESeq2** package (Love *et al.,* 2014) was used to estimate the relationship between **variance** and **mean** in the high-throughput sequencing experimental count data, and to assess **differential expression** using a model based on the **negative binomial distribution**.

The **DESeqDataSet** object **dds** stored read counts and intermediate calculations for differential expression analysis.

- **DESeqDataSetFromMatrix** : Created a **DESeqDataSet object** from a matrix (pre-loaded data) of non-negative integer count values for differential expression analysis.
- **estimateSizeFactors**: **Size factors** ensured that **variations** in gene expression between samples reflected true biological differences rather than technical issues, such as varying sequencing depths.

LPS vs Media - Size Factors
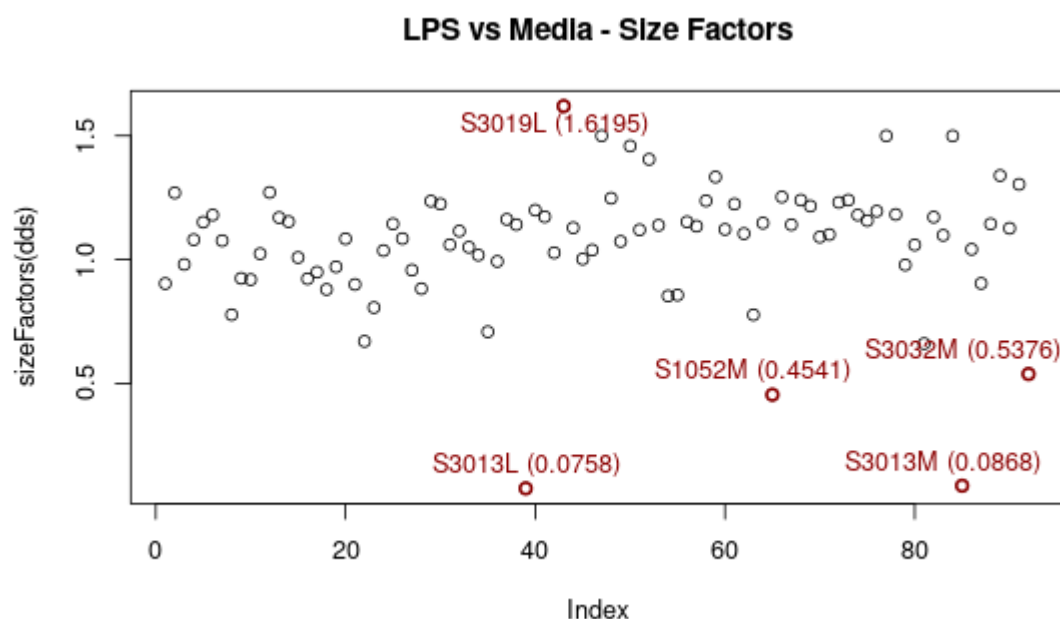
- Size factors fell around 1, with exception of S3013L (0.07576956), S1052M (0.45409687), S3013M (0.08675980), S3032M (0.53764182) and S3019L (1.61950669).
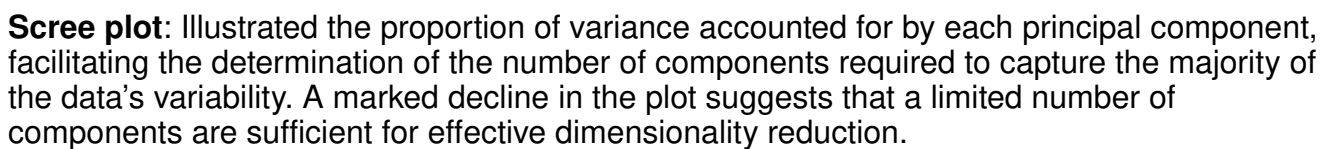
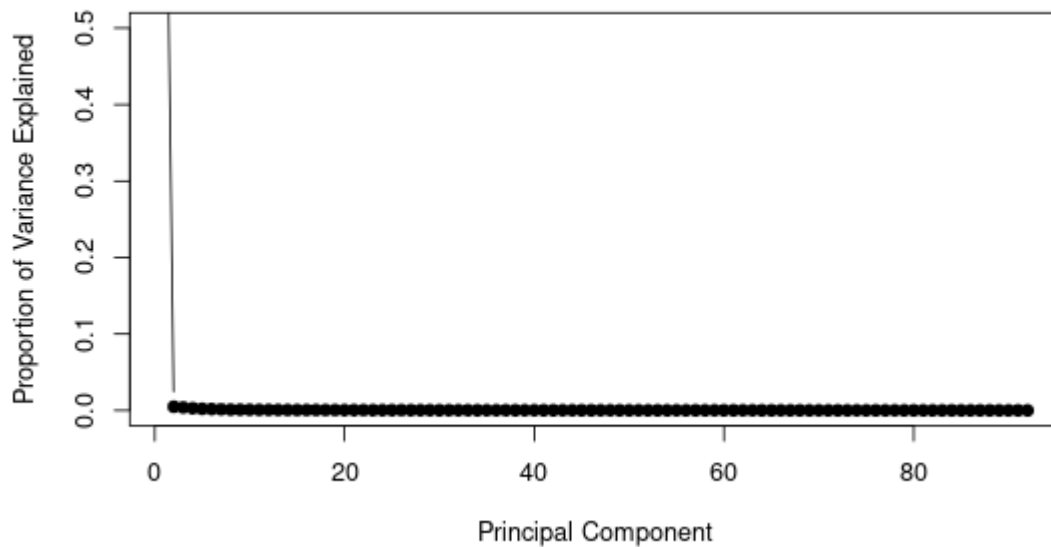**DEseq2** used **non-normalized** gene count data and adjusted for library size differences internally.

Quality Control Assessment

Variance Exploration

Before **clustering**, it was necessary to remove **variance** dependence on the mean. *DESeq2* offered **rlog()** and **vst()** functions, with vst() being preferred for its speed. Both methods ignore the design formula by default, but if large count differences are expected, the blind parameter should be set to FALSE. Here, blind=TRUE: unbiased by the experimental design or known sources of variation (McDonnell Genome Institute - Washington University: Genomic Visualization and Interpretations, link (https://genviz.org/module-04-expression/0004/02/01/DifferentialExpression/)).

- **Principal Component Analysis (PCA)** statistical tool was used for the exploration data stratification without relying on strong initial assumptions (Love *et al.,* 2014, link (https://bioconductor.org/packages/release/bioc/manuals/DESeq2/man/DESeq2.pdf); Giuliani, 2017).

- **Loadings** are the coefficients that indicated how much each original variable contributed to a principal component in PCA.

- ***ggplot2*** (Wickham, 2016) package of the ***tidyverse*** (Wickham *et al.,* 2019) library and **CRAN** library ***ggrepel*** (Slowikowski, 2014) were used for visualization of PCA data derived through *DESeq2* library function **plotPCA**.

LPS vs Media: PCA Plot

Scree plot: Illustrated the proportion of variance accounted for by each principal component, facilitating the determination of the number of components required to capture the majority of the data's variability. A marked decline in the plot suggests that a limited number of components are sufficient for effective dimensionality reduction.

LPS vs Media: Variance Scree Plot

Plotting **Sparsity** using *DESeq2* library function **plotSparsity**: proportion of zero counts across all genes or transcripts in the dataset.

Concentration of counts over total sum of counts
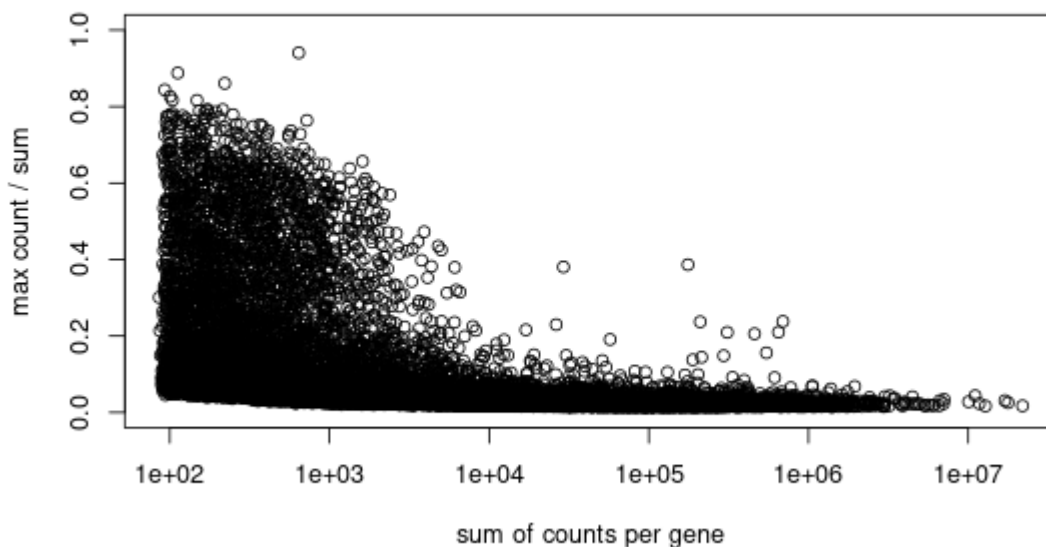
Visual Heatmaps of Hierarchical Clustering at different levels

pheatmap library (Kolde, 2019) : heatmap illustrating high-dimensioal matrix values through gradients of colors and dendogram to show the relationships and patterns between rows and columns of data. Used to help identify clusters, trends, and correlations among variables.

Heatmap of Correlated Samples using Matrix Data Values - LPS vs Media
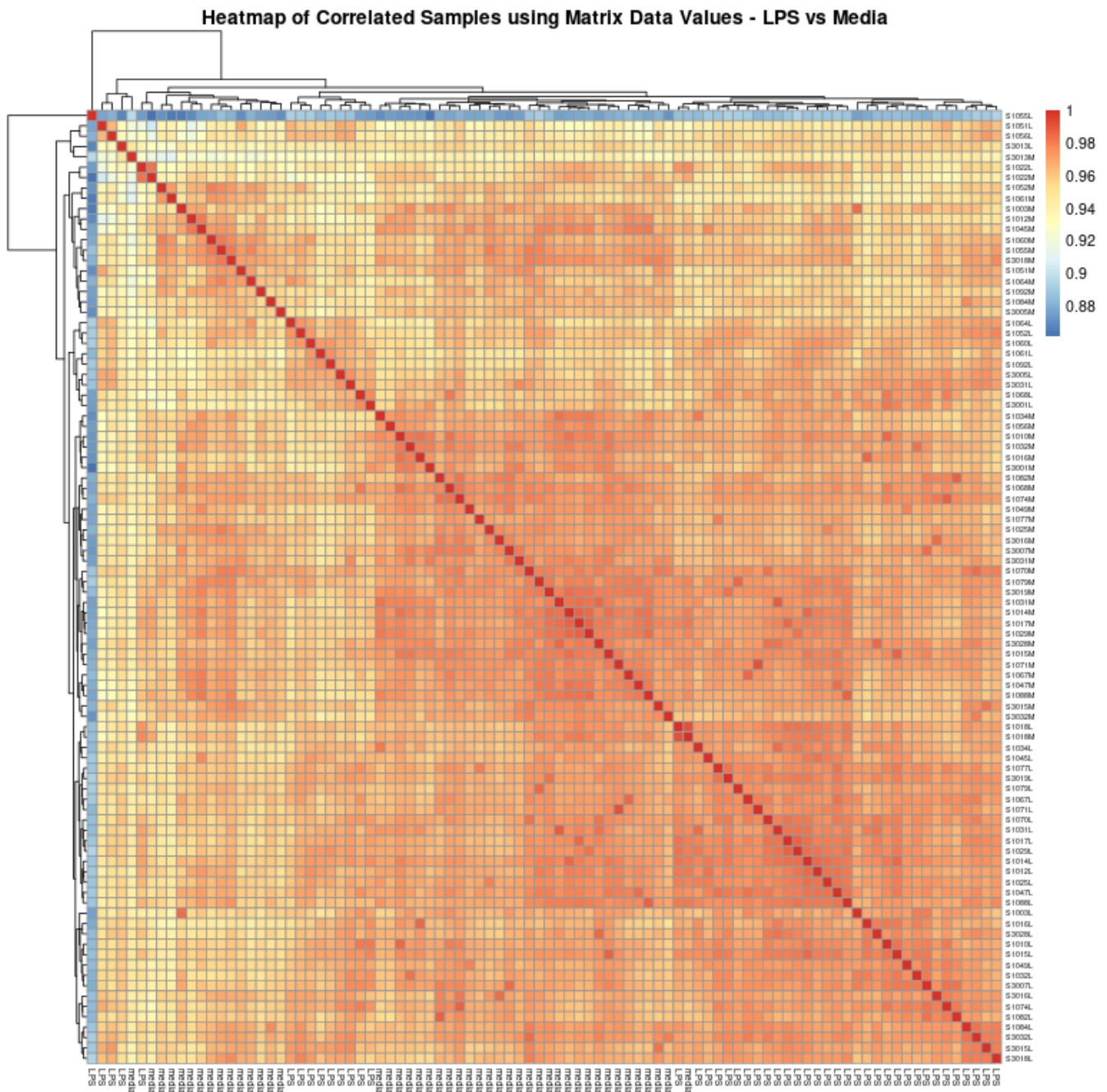
DIFFERENTIAL EXPRESSION

DESeq with Wald Test

Negative Binomial General Linear Model fitting and Wald statistics: employed to test for significance of coefficients, using sizeFactors and dispersion estimates.

Gamma-Poisson distribution formula as per *DESeq2* documentation (Love *et al.,* 2014):

$$K_{ij} \sim \mathrm{NB}(\mu_{ij}, \alpha_i)$$

$$\mu_{ij} = s_j q_{ij}$$

$$\log_2(q_{ij}) = x_{j.}\beta_i$$

Counts K_{ij} for gene i and sample j are modeled with a Negative Binomial distribution, having a mean μ_{ij} and gene-specific dispersion α_i. The mean μ_{ij} is the product of a sample-specific size factor s_j and a parameter q_{ij}, which is proportional to the expected concentration of fragments for sample j. The coefficients β_i represent the log2 fold changes for gene i in each column of the model matrix X (Love, 2014).

Wald test nbinomWaldTest() function evaluated whether the $\log_2\mathrm{FC}$ of gene expression between conditions is significantly different from zero.
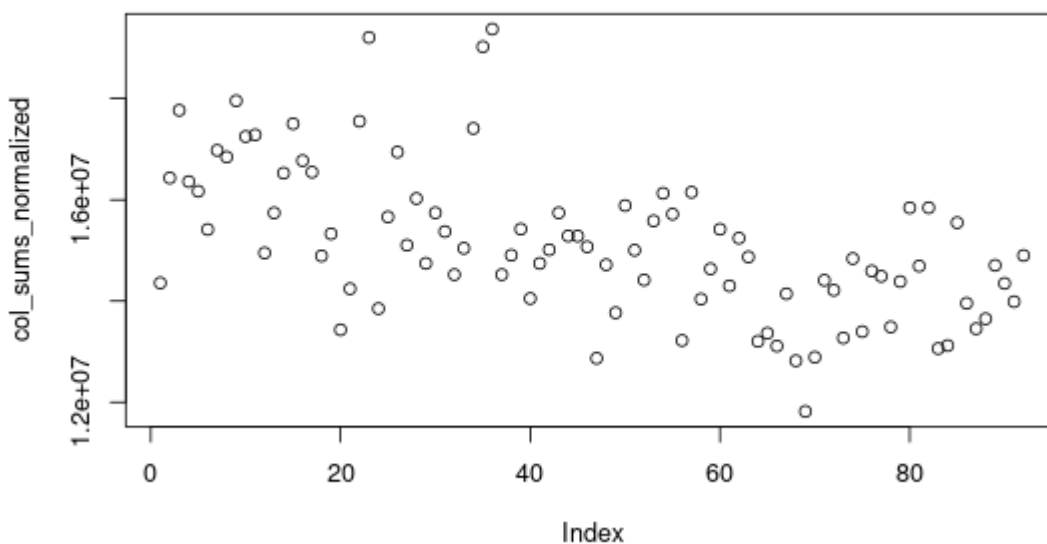
```
out of 20576 with nonzero total read count
adjusted p-value < 0.1
LFC > 0 (up)       : 3755, 18%
LFC < 0 (down)     : 3170, 15%
outliers [1]       : 2133, 10%
low counts [2]     : 0, 0%
(mean count < 0)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results
```

Raw and Normalized counts investigation:



Total Raw Counts per Sample - LPS vs Media
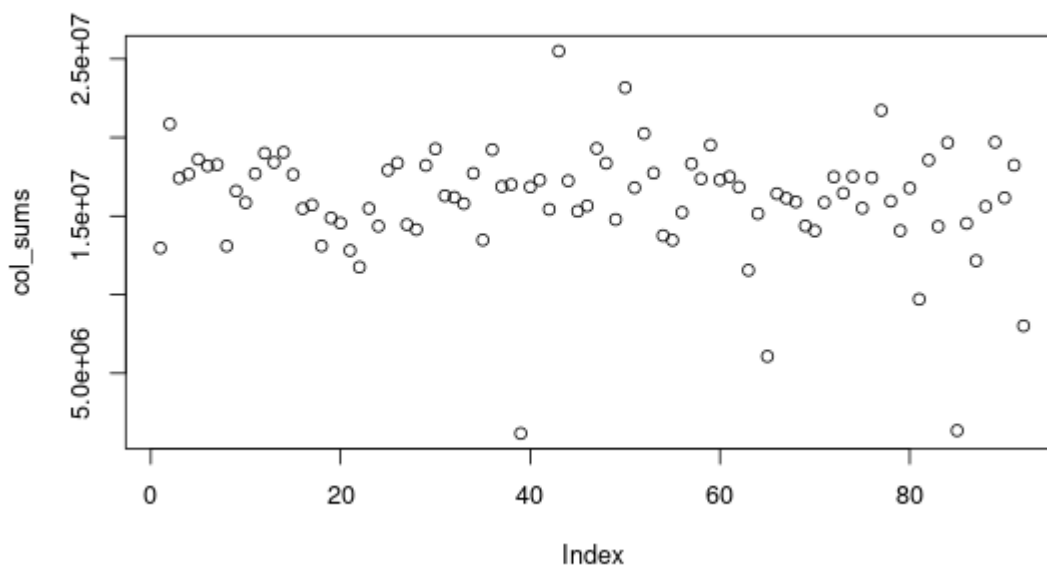
Total Normalized Counts per Sample - LPS vs Media

Dispersion Estimators of Alternative $\log_2 \text{FC}$ Shrinkage

dds additionally retained the design formula used for estimating **dispersion** and $\log_2 \text{FC}$ in the model to reflect the variance of the distribution, using the The Cox-Reid Adjusted Profile Likelihood (APL) method (McCarthy *et al.,* 2012; Landau and Liu, 2013).

Types Normal, Ashr and Apeglm
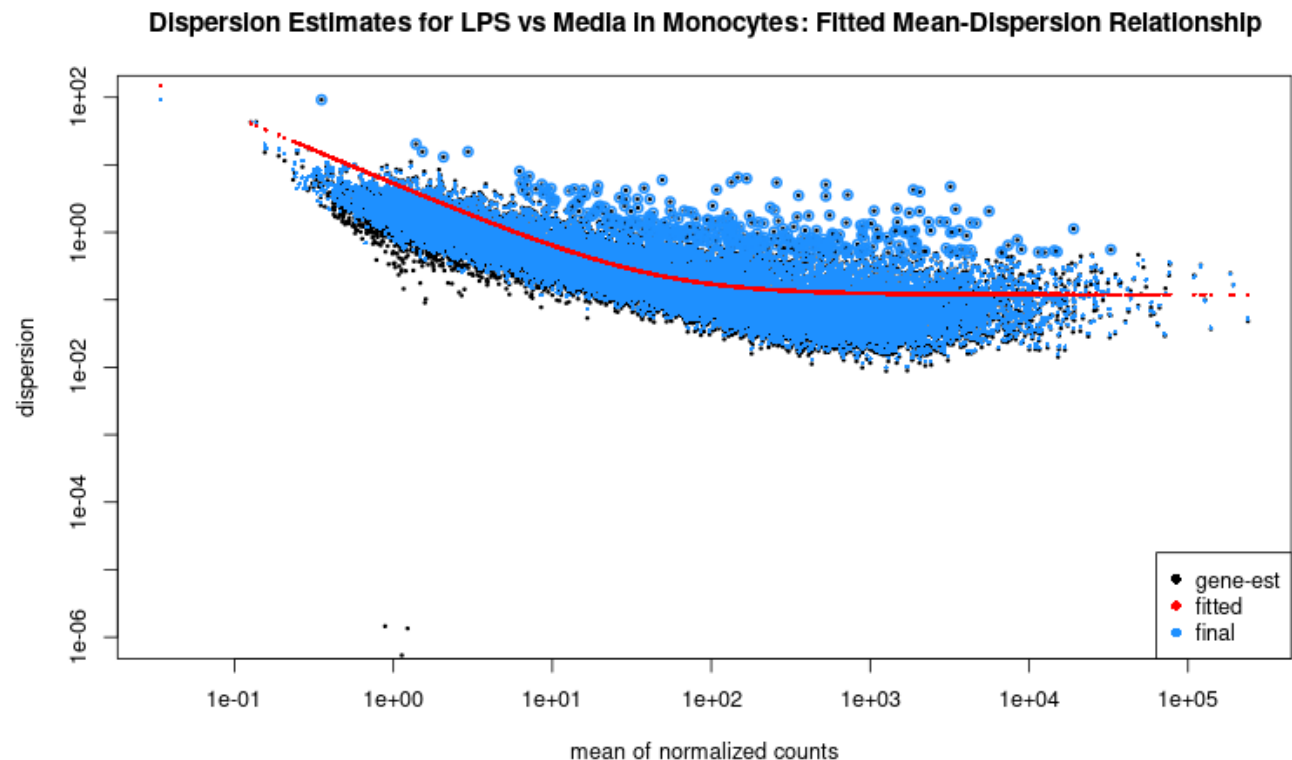
DESeq2 package **lfcShrink** function shrinks $\log_2 \text{FC}$ estimates to improve reliability, particularly for genes with low counts or high variability.

- **Normal** shrinkage is basic and often less accurate; uses the posterior mean from the Maximum Likelihood Estimate; *DESeq2* default method (Love *et al.,* 2014); introduces more bias than **ashr** and **apeglm** shrinkage estimators (Zhu *et al.,* 2018).
- **Ashr** adaptive shrinkage based on gene-specific information; more flexible in estimates uncertainty (Stephens *et al.,* 2023).
- **Apeglm** (approximate posterior estimation for generalized linear models) most precise method, applying Bayesian shrinkage; typically recommended for robust differential expression results (Zhu *et al.,* 2018).

MA plots (M log ratio versus A mean average) for each shrinkage method (**normal, ashr, apeglm**) showed $\log_2 \text{FC}$ versus mean expression, helping to visualize and compare differential expression with stabilized estimates. Plotted scatter plots of $\log_2 \text{FC}$s (on the y-axis) versus the mean of normalized counts (on the x-axis) using *DESeq2* function **plotMA()**.
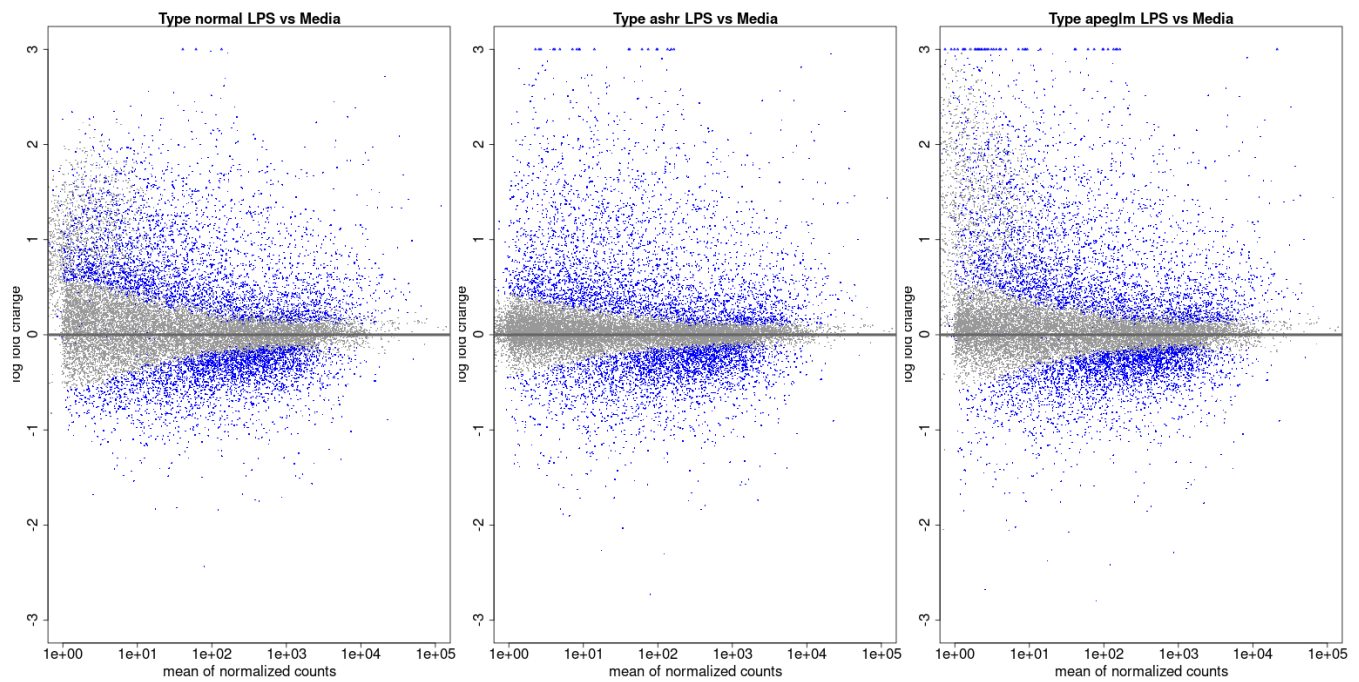
Further results explorations revealed comparable results between shrinkage estimation methods.

Hide

```
summary(DGE_table_condition_normal)
```

```
out of 20576 with nonzero total read count
adjusted p-value < 0.1
LFC > 0 (up)       : 3896, 19%
LFC < 0 (down)     : 3137, 15%
outliers [1]       : 0, 0%
low counts [2]     : 0, 0%
(mean count < 0)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results
```
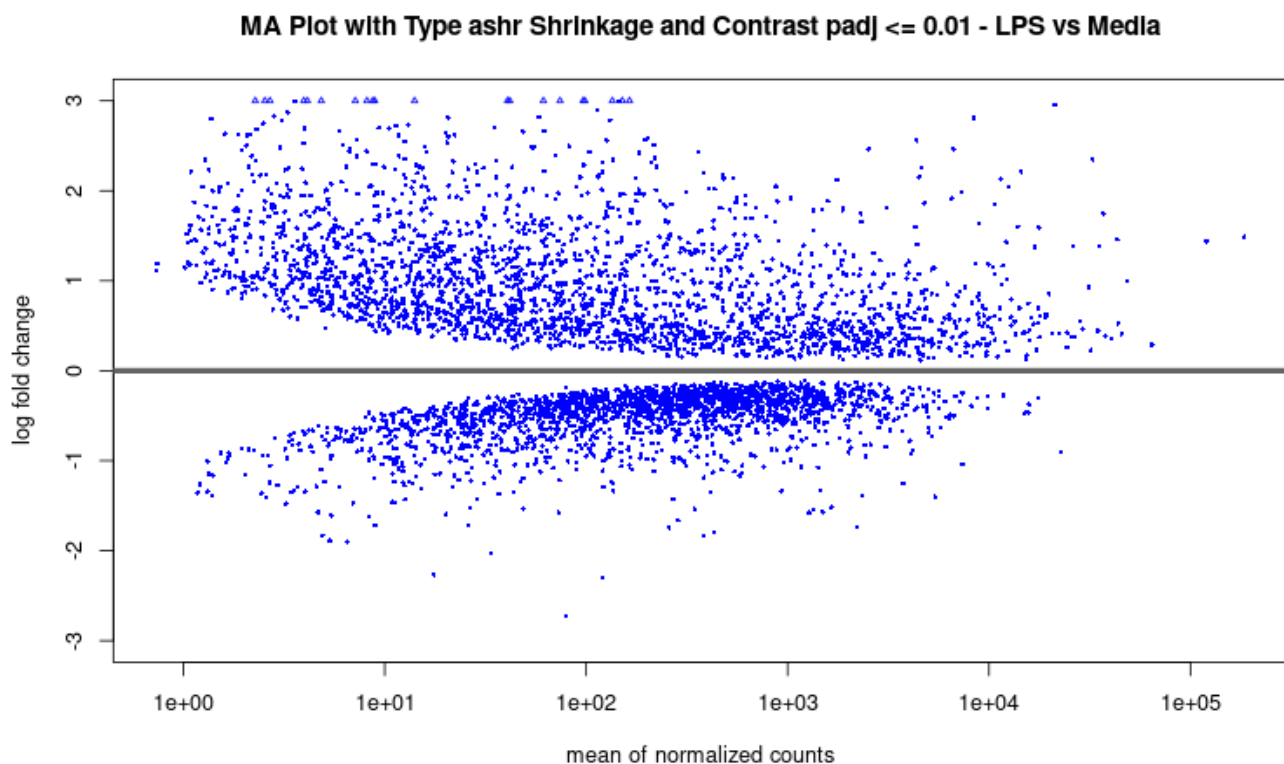
Hide

```
summary(DGE_table_condition_ashr)
```

```
out of 20576 with nonzero total read count
adjusted p-value < 0.1
LFC > 0 (up)       : 3878, 19%
LFC < 0 (down)     : 3155, 15%
outliers [1]       : 0, 0%
low counts [2]     : 0, 0%
(mean count < 0)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results
```

Hide

```
summary(DGE_table_condition_apeglm)
```

```
out of 20576 with nonzero total read count
adjusted p-value < 0.1
LFC > 0 (up)       : 3896, 19%
LFC < 0 (down)     : 3137, 15%
outliers [1]       : 0, 0%
low counts [2]     : 0, 0%
(mean count < 0)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results
```


MA Plot with Type ashr Shrinkage and Contrast padj <= 0.01 - LPS vs Media

Summary statistics for filtered, significant results (where src = shrunken results with contrast; *padj* ≤ 0.01)

```
out of 20576 with nonzero total read count
adjusted p-value < 0.1
LFC > 0 (up)       : 3878, 19%
LFC < 0 (down)     : 3155, 15%
outliers [1]       : 0, 0%
low counts [2]     : 0, 0%
(mean count < 0)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results


    symbol              baseMean        log2FoldChange        lfcSE              pv
alue              padj
 Length:4490        Min.   :      0.74   Min.   :-2.7301   Min.   :0.03204   Min.
:0.000e+00   Min.   :0.000e+00
 Class :character   1st Qu.:     24.39   1st Qu.:-0.3983   1st Qu.:0.08713   1st Q
u.:3.000e-10   1st Qu.:6.000e-09
 Mode  :character   Median :    132.80   Median : 0.3167   Median :0.13233   Median
:2.456e-06   Median :2.250e-05
                    Mean   :   1037.10   Mean   : 0.3081   Mean   :0.17152   Mean
:2.385e-04   Mean   :1.217e-03
                    3rd Qu.:    601.57   3rd Qu.: 0.9172   3rd Qu.:0.21802   3rd Q
u.:1.951e-04   3rd Qu.:1.192e-03
                    Max.   :184669.41   Max.   : 5.4206   Max.   :0.88743   Max.
:2.175e-03   Max.   :9.966e-03
```

Enhanced Volcano Plot Visualization

The resulting **Volcano** plot visualized the significance and direction (up or downregulated) of gene expression changes between conditions, helping identify significant genes with large fold changes. To achieve this, the Bioconductor *EnhancedVolcano* package (Blighe *et al.*, 2023, link (https://bioconductor.org/packages/devel/bioc/vignettes/EnhancedVolcano/inst/doc/EnhancedVolcano.html)) was employed.

LPS vs Media: Differentially Expressed Genes
EnhancedVolcano

total = 20576 variables
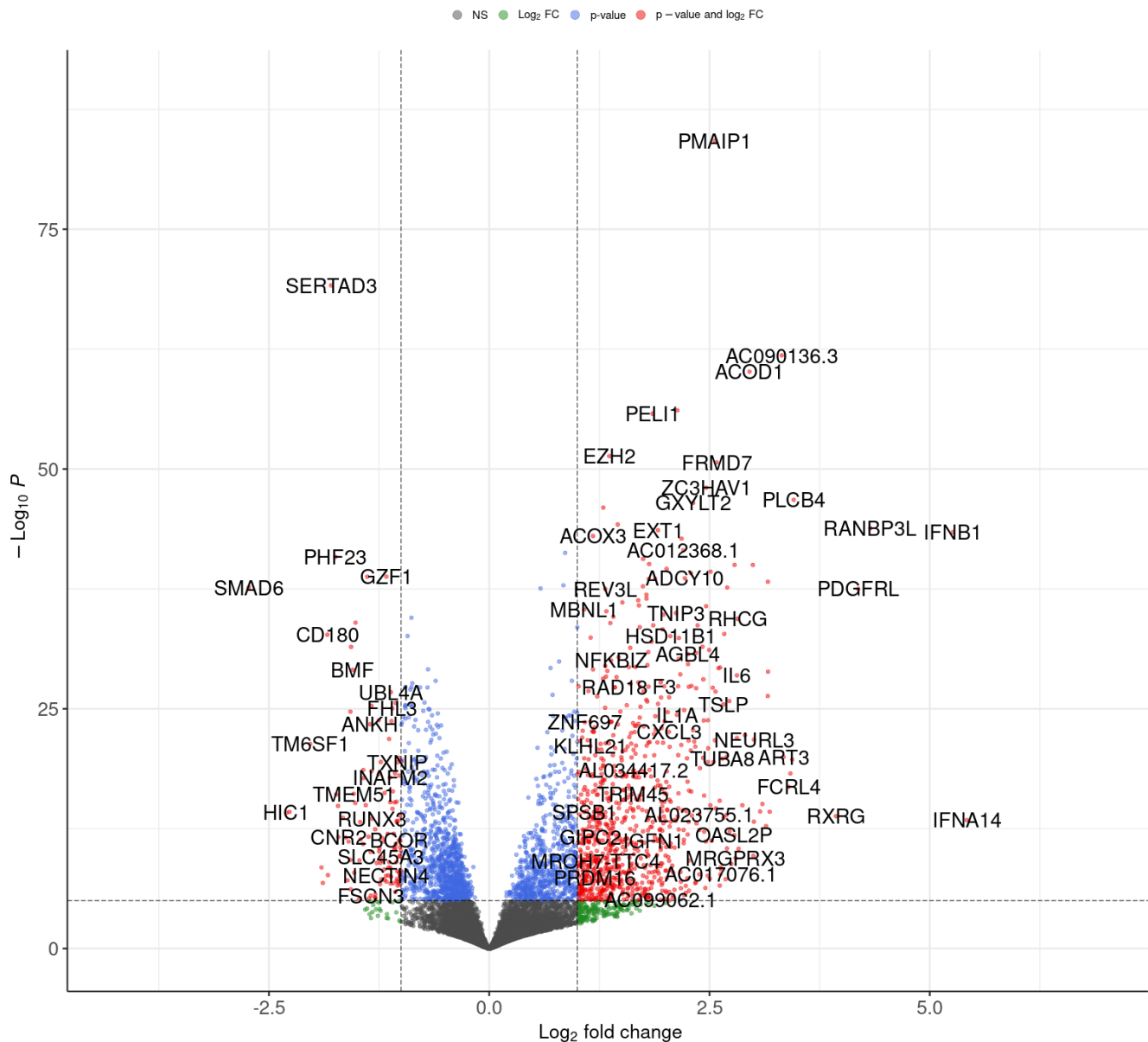
PATHWAY ANALYSIS

Annotation

Using ***AnnotationDbi*** (Pagès *et al.,* 2023, link (https://jorainer.github.io/ensembldb/reference/EnsDb-AnnotationDbi.html)) virtual base class function **mapIds** to map gene symbols to **ENSEMBL IDs** (Harrison *et al.,* 2024, link (https://www.ensembl.org/index.html)), for integrating and managing **gene annotations** effectively for further analysis.

- **Mapping Gene Symbols to ENSEMBL IDs**: Employing the ***org.Hs.eg.db*** (Carlson, 2023) package to convert **gene symbols** to ENSEMBL IDs, storing these IDs in a new column.
- **Adding Gene Symbols**: Creating a column in the data frame to store the original gene symbols for reference.
- **Handling Multiple ENSEMBL IDs**: Concatenating any multiple ENSEMBL IDs associated with a gene symbol into a single, comma-separated string.
- **Filtering for Reliability**: Retained only rows with valid gene symbols and ENSEMBL IDs, and filtered

on **baseMean > 50** to exclude low-expression noise, ensuring biological relevance in downstream analyses.

```
    symbol              baseMean         log2FoldChange      lfcSE              p
value            padj             ensgene
 Length:2616       Min.   :    50.01   Min.   :-2.73014   Min.   :0.03204   Min.
:0.000e+00   Min.   :0.000e+00   Length:2616
 Class :character   1st Qu.:   164.39   1st Qu.:-0.40123   1st Qu.:0.07337   1st Q
u.:0.000e+00   1st Qu.:1.000e-09   Class :character
 Mode  :character   Median :   436.97   Median :-0.22686   Median :0.09448   Media
n :7.505e-07   Median :7.571e-06   Mode  :character
                   Mean   :  1720.26   Mean   : 0.07031   Mean   :0.10730   Mean
:2.005e-04   Mean   :1.028e-03
                   3rd Qu.:  1284.10   3rd Qu.: 0.48170   3rd Qu.:0.12524   3rd Q
u.:1.256e-04   3rd Qu.:8.052e-04
                   Max.   :184669.41   Max.   : 5.26195   Max.   :0.39345   Max.
:2.164e-03   Max.   :9.923e-03
```

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) (Szklarczyk *et al.,* 2023, link (https://string-db.org/)): web database that consolidates known and predicted protein-protein interactions (PPIs), enabling the study and visualization of protein networks across organisms.

- **Exporting Differentially Expressed Genes**: Segregated and sorted upregulated ($\log_2\mathrm{FC} > 0$) and downregulated ($\log_2\mathrm{FC} < 0$) genes by effect size, then exported full tables and individual gene symbol and ENSEMBL ID lists for use in downstream applications in STRING network analysis.

  - Selected the top 100 DEGs per condition to balance biological relevance and network interpretability, applying consistent criteria across all STRING analyses.

```
  [1] "IFNB1"       "RANBP3L"     "PLCB4"       "LINC02015"   "NEURL3"      "ACOD1
"     "LINC02068"   "RHCG"        "IL6"         "OSR2"        "LAMA3"       "FRM
D7"       "PMAIP1"      "IL36G"       "SLC25A48"
 [16] "LMNB1-DT"    "ZC3HAV1"     "MIR155HG"    "CSF3"        "INHBA-AS1"   "BCL2L
14"      "PTGS2"       "LAMP3"       "GXYLT2"      "IL2RA"       "CCL20"       "ADC
Y10"      "CCL4L2"      "BATF2"       "DNAAF1"
 [31] "PTCD2"       "OASL"        "LINC02426"   "STEAP1B"     "NOCT"        "IL1A"
"TNIP3"       "IFIT5"       "CXCL3"       "SLC44A3-AS1" "CD69"        "SMTNL1"
"EIF1B-AS1"   "AIM2"        "F3"
 [46] "LINC01093"   "C1QTNF1"     "RTP4"        "LINC02056"   "HERC6"       "HMGA2
"     "SYNPO2"      "EXT1"        "NAV2"        "MACC1"       "ITGB8"       "ABC
A6"       "HERC5"       "PNPT1"       "BAALC"
 [61] "PELI1"       "USP18"       "CCR7"        "MSANTD3"     "EBF1"        "ACO1"
"INHBA"       "CCL3L1"      "IFIH1"       "GCKR"        "DDX60"       "ISG20"
"STAT4"       "E2F7"        "GCNT4"
 [76] "CCL4"        "NLRP3"       "CARD18"      "CMAHP"       "C2orf92"     "TRIM5
5"       "MIR3945HG"   "GSDME"       "ZFPM2"       "IL23A"       "MIR3681HG"   "PPM
1K"       "RHOH"        "RSAD2"       "IL6R-AS1"
 [91] "CACNA1A"     "SERPINB2"    "RASGRP3"     "TRIQK"       "LINC00649"   "KCNH8
"     "MREG"        "EREG"        "LINC02324"   "TNF"
```

Permalink STRING Output LPS vs Media Upregulated link (https://version-12-0.string-db.org/cgi/network?networkId=bh5XmeUb8ZLd)

```
  [1] "SMAD6"      "HIC1"       "CD180"       "SERTAD3"     "PHF23"        "THBD"
"OLIG1"       "CCL2"       "ID1"         "EEPD1"       "BMF"          "TMEM51"      "S1P
R3"       "DUSP7"      "EGR2"        "LRRC25"
 [17] "ANKH"       "SH2D3C"     "VWF"         "RUNX3"       "SLC25A19"     "HMOX1"
"SDS"       "IL21R"      "ARL4C"       "GZF1"        "ARRDC2"       "INAFM2"      "ID3
"        "CD9"        "UBL4A"       "C5orf24"
 [33] "FHL3"       "GPR157"     "PHF13"       "GPR35"       "SMAD7"        "P2RY2"
"ZNF331"     "HHEX"       "ICOSLG"      "B3GNT8"      "TXNIP"        "CXXC5"       "BCO
R"        "KCTD6"      "LSR"         "CHIT1"
 [49] "TNFRSF12A"  "VASH1"      "RARG"        "CEBPA"       "SLC25A25"     "COL23A1"
"MRFAP1L1"   "TLR1"       "FAM114A1"    "ZNF34"       "ZKSCAN8"      "PRR5L"       "TME
M52B"     "CLP1"       "CNEP1R1"     "STX1A"
 [65] "CRTAM"      "ZBTB3"      "IRF2BP1"     "LINC00324"   "HPS6"         "CTTNBP2NL"
"ZBTB45"     "LYL1"       "CTBP2"       "TOB2"        "SERAC1"       "CASS4"       "PCS
K6"       "CYTIP"      "SNX24"       "SH3BP4"
 [81] "ENC1"       "ZNF211"     "TGFBR1"      "BTG2"        "ST6GALNAC4"   "ARMC7"
"CHAMP1"     "S1PR1"      "MFSD5"       "ZNF615"      "CSTF1"        "CISH"        "RAP
2B"       "GXYLT1"     "MAP1S"       "ZNF740"
 [97] "KLF11"      "TFEB"       "ZNF764"      "NUAK2"
```
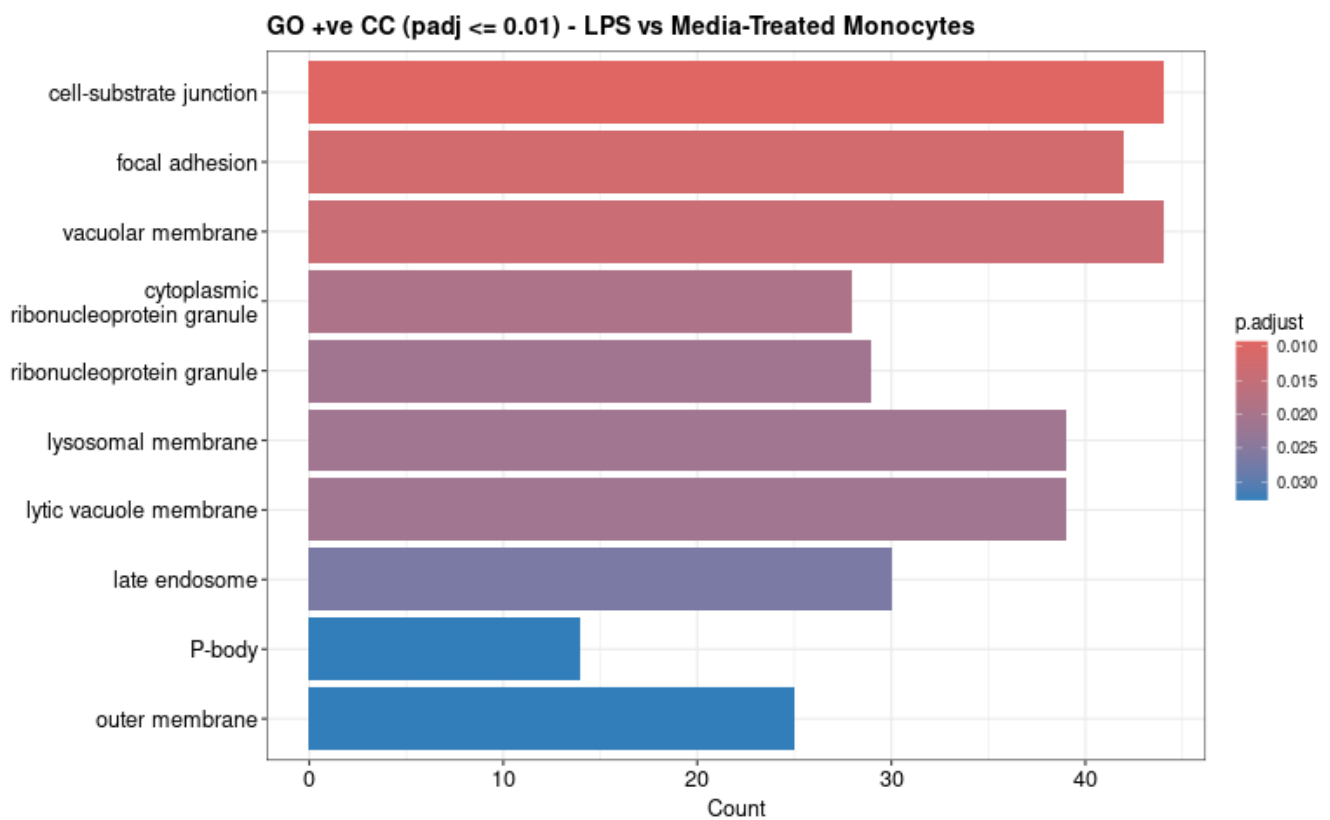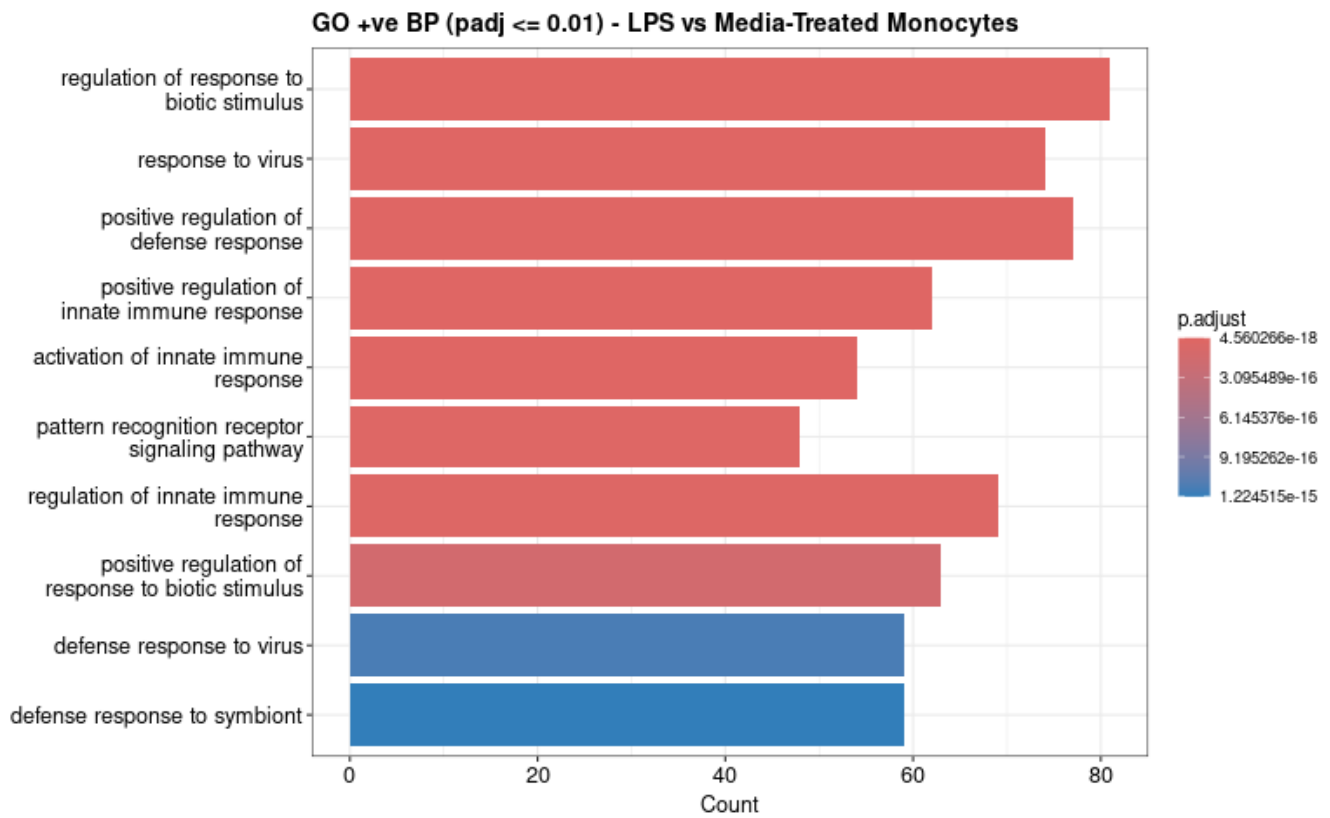
Permalink STRING Output LPS vs Media Downregulated link (https://version-12-0.string-db.org/cgi/network?networkId=bqKQz1zwmH2w)
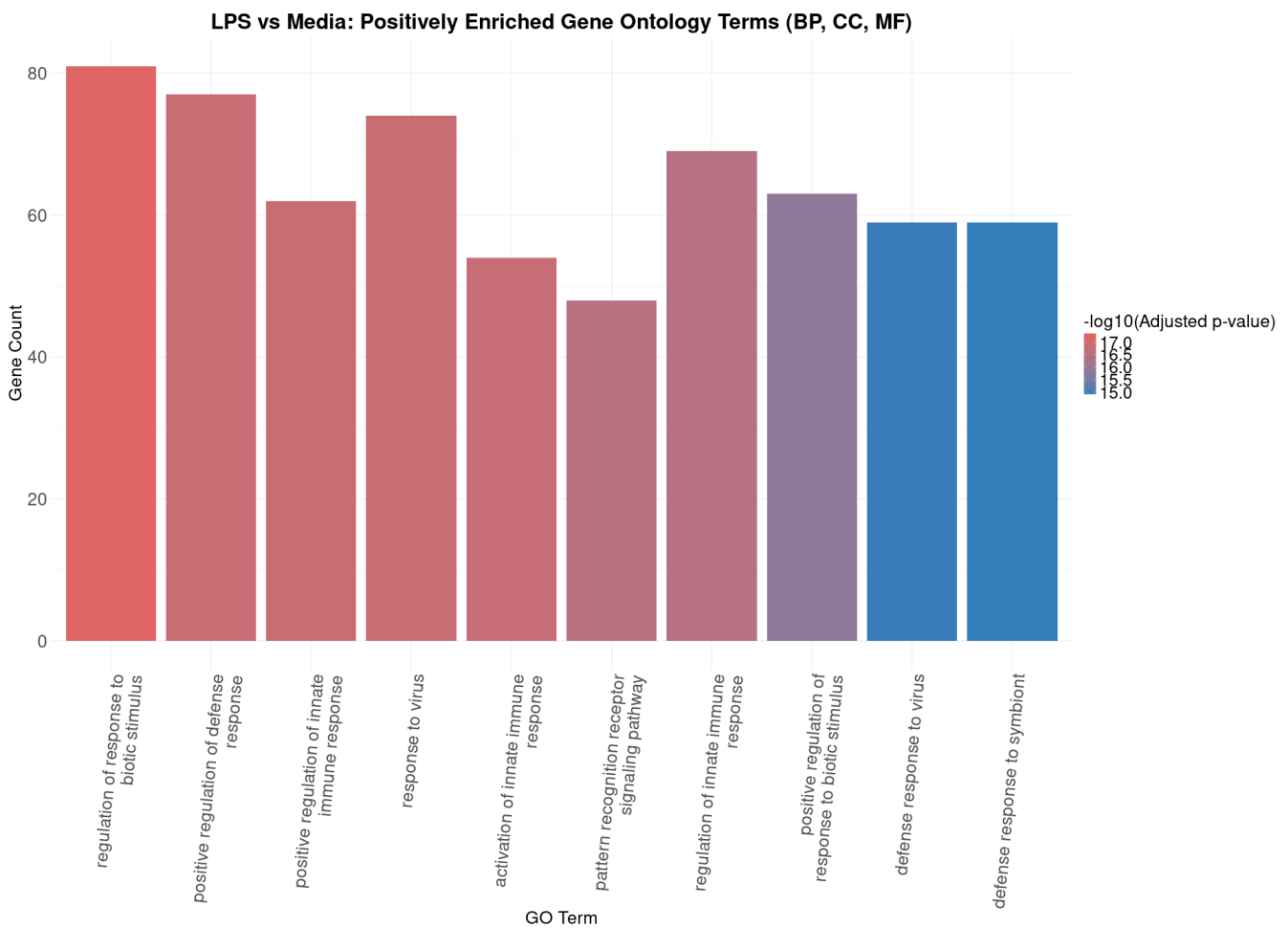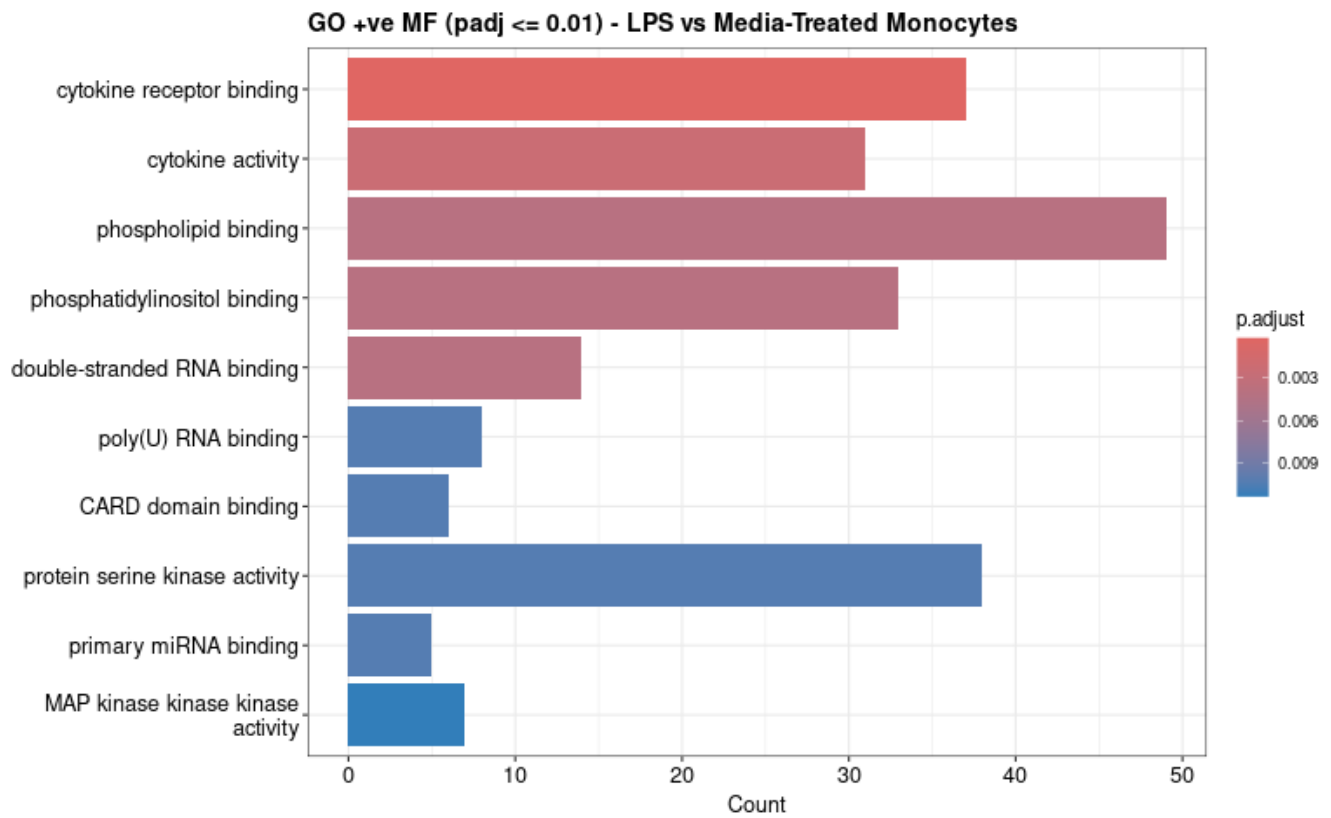
GO Enrichment Analysis

GO (Gene Ontology) Enrichment Analysis (Harris *et al.,* 2004): ***clusterProfiler*** package (Yu *et al.,* 2012; Wu *et al.,* 2021; Xu *et al.,* 2024, link (https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html)), for statistical analysis and visualization of functional profiles for genes and gene clusters, function **enrichGO** takes a gene vector, using gene symbols as keys, and returns the enriched GO categories, adjusted for false discovery rate (**FDR**).

- **OrgDb** = "org.Hs.eg.db" specifies the organism (human genes) database to use for annotations:

   - **Biological Process (BP)**: Describes biological objectives/ processes that a gene/ gene product is involved in
   - **Cellular Component (CC)**: Indicates locations within cells where a gene product is active or located
   - **Molecular Function (MF)**: Describes specific biochemical activities/functions of a gene product at the molecular level
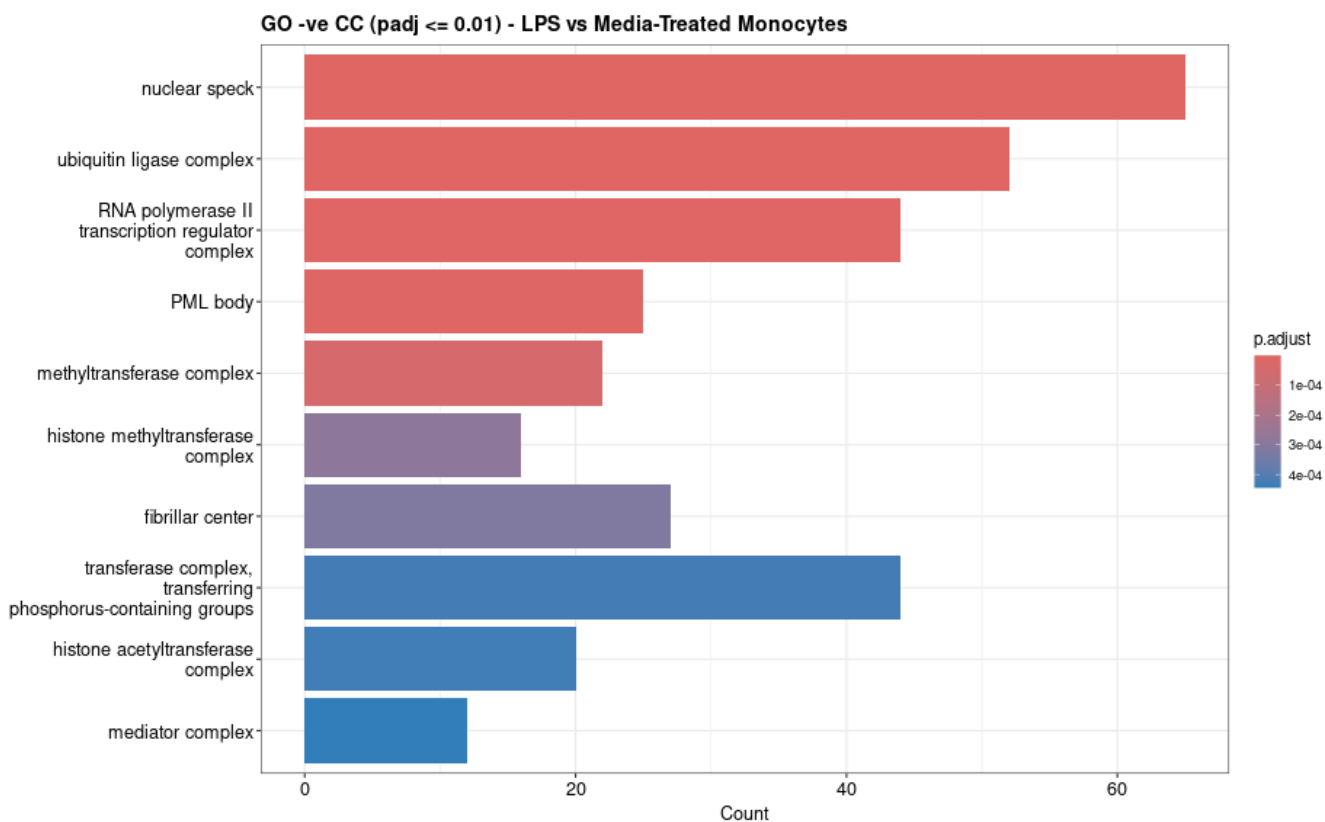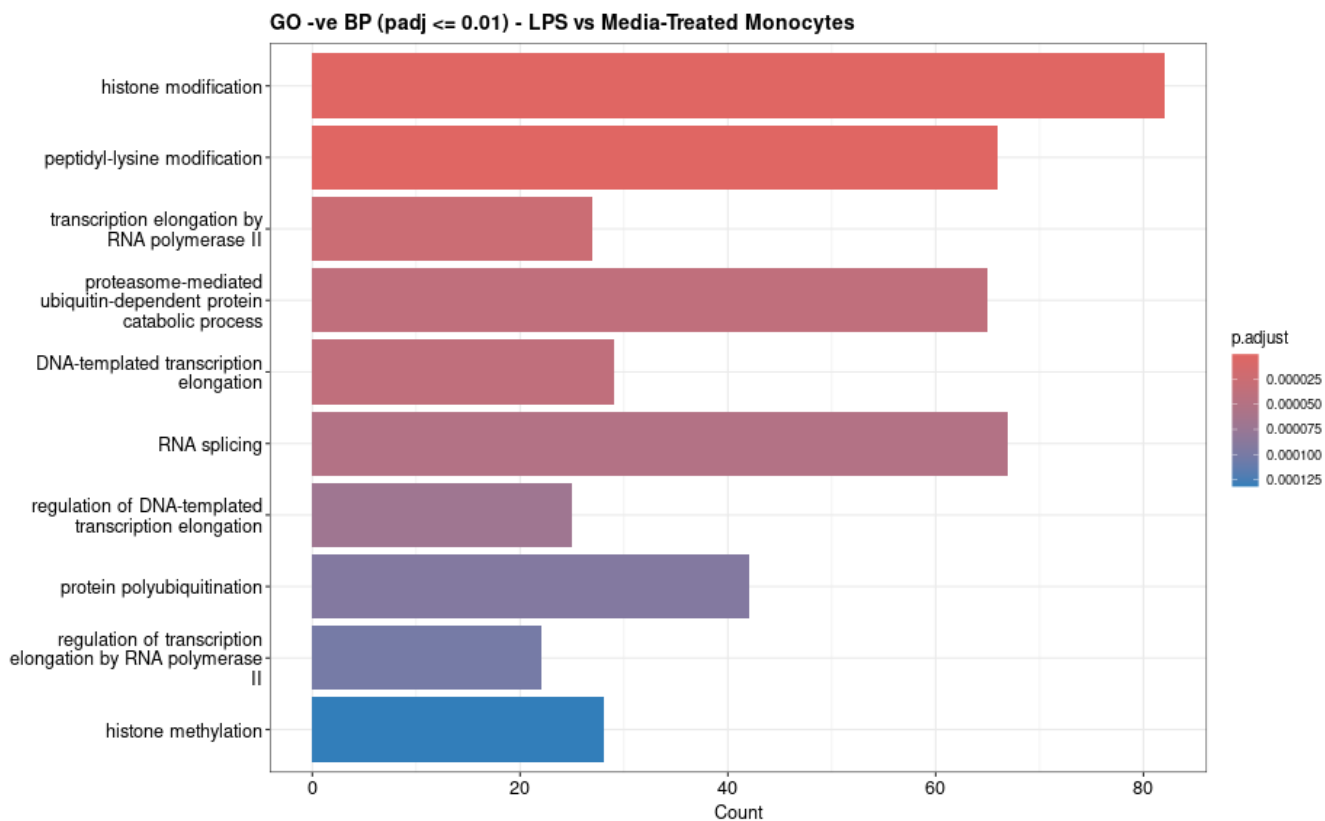
Visualized the top 10 **enriched GO terms** for each ontology to understand which **biological processes**, **cellular components**, and **molecular functions** are most significantly associated with the gene set. Results visualized with Bar Plots were further corroborated during **Gene Set Enrichment Analysis (GSEA)** (Mootha *et al.,* 2003; Subramanian *et al.,* 2005).
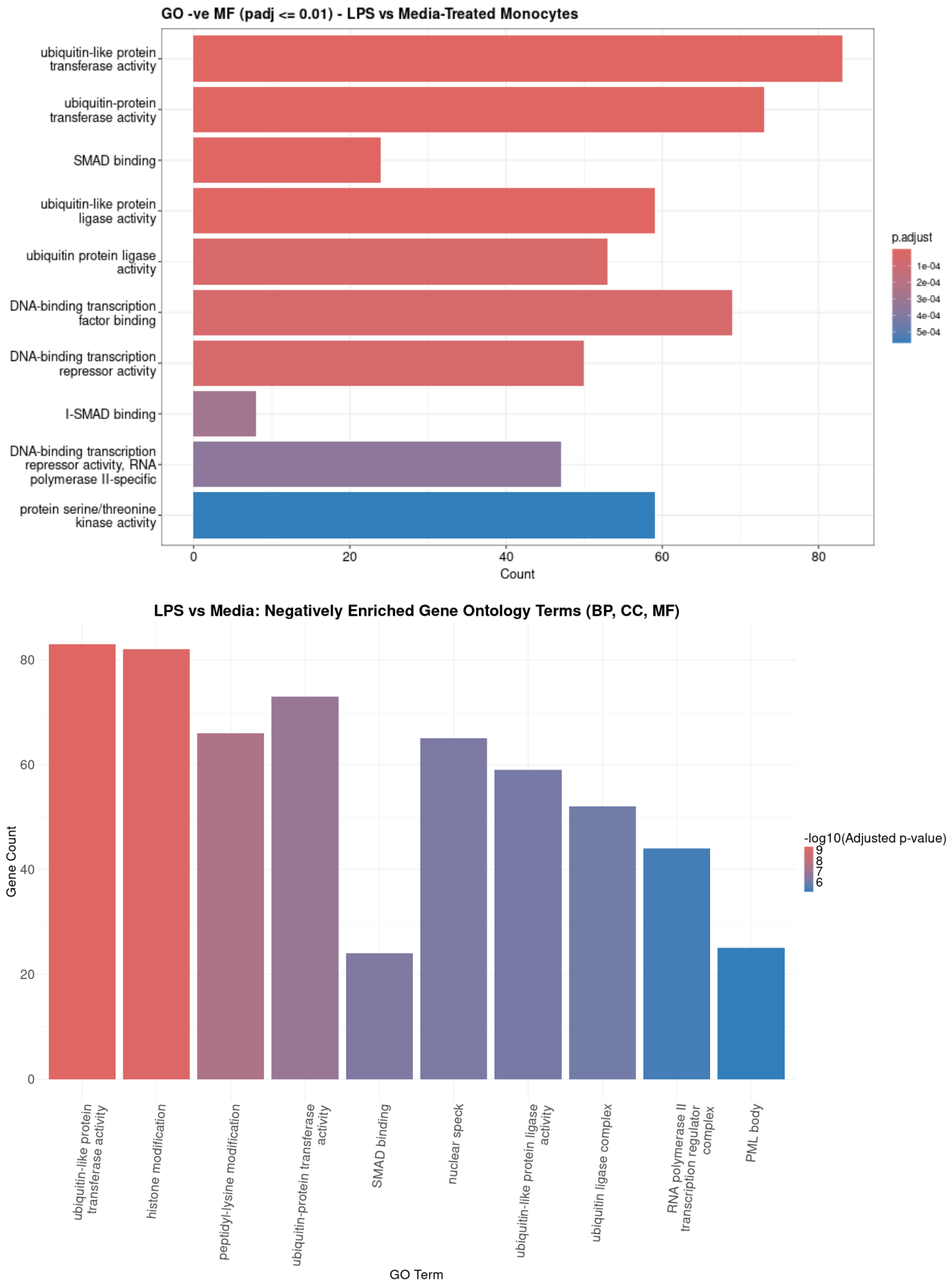
Upregulated genes

GO +ve BP (padj <= 0.01) - LPS vs Media-Treated Monocytes

GO +ve CC (padj <= 0.01) - LPS vs Media-Treated Monocytes

Downregulated genes

GO -ve BP (padj <= 0.01) - LPS vs Media-Treated Monocytes

GO -ve CC (padj <= 0.01) - LPS vs Media-Treated Monocytes

GO -ve MF (padj <= 0.01) - LPS vs Media-Treated Monocytes

LPS vs Media: Negatively Enriched Gene Ontology Terms (BP, CC, MF)

ANALYSIS OF Tumor Necrosis Factor- *alpha*

(TNF-α) secretion in LPS TREATED MONOCYTES

Following the above workflow pipeline, a dataset of 38 out of 92 initial samples was statistically manipulated for downstream analysis. These samples represented the LPS-treated monocytes, where **cytokine release**, specifically **TNF-α** secretion capacity, was recorded. The metadata identified three distinct populations, of **high**, **low** and **intermediate** responders to *ex vivo Escherichia coli*-LPS stimulation, respectively. One sample, displaying intermediate response, was removed from the dataset. This was done to prevent skewing the analysis and ensure more accurate comparisons between responders.

The scope of studying this secondary dataset was to investigate **immune tolerance** in endotoxin-challenged acute stage monocytes showing **impaired production of the pro-inflammatory cytokine TNF-α**.

The **DESeqDataSet** was built with **Low Response** as the **reference level**, **sex** as a **covariate**, and all **factor levels** retained for unbiased exploration prior to normalization.
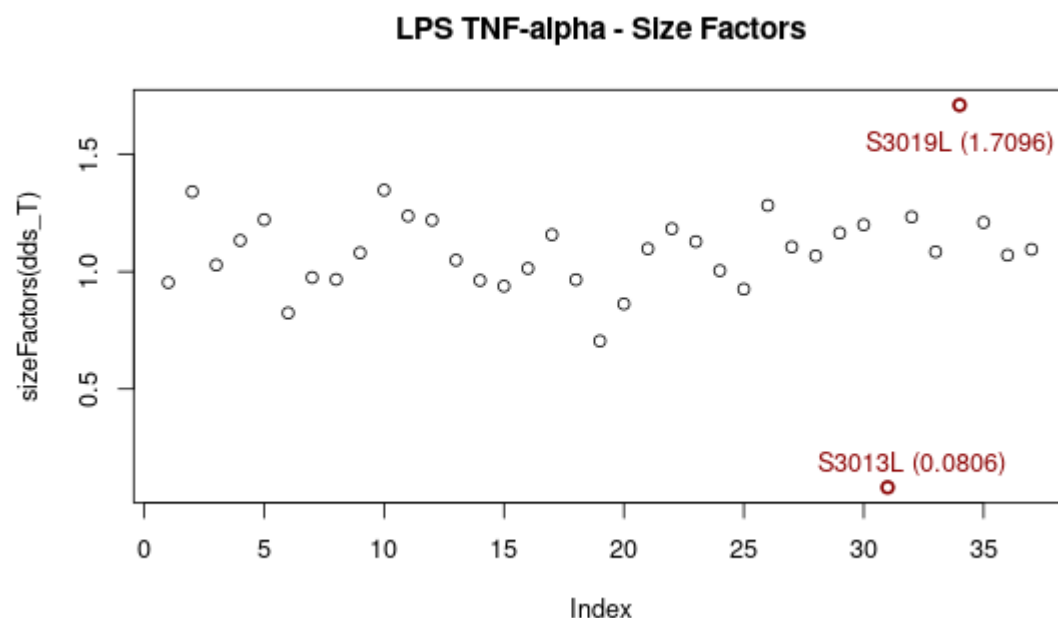
- Only two size factors notably deviated from 1 and may indicate outliers: sample 3013L, with a size factor of 0.08056695, which was consistent with earlier observations of the primary count matrix in the LPS vs. Media comparison (approximately 0.08), and sample S3019L, with a size factor of 1.70958503, which aligned with its previous value of 1.61950669.


Quality Control Assessment

Variance Exploration and PCA

PCA Plot LPS TNF-alpha Responses - PC1 vs PC2

- The PCA plot revealed two distinct populations along PC1, which explained 20.34% of the total variance in the dataset, indicating the presence of two major groups stratified by TNF-α response. Additionally, further stratification along PC2, which accounts for 12.55% of the variance, shows sex-stratified subpopulations exhibiting additional differentiation. While PC1 captured the primary clustering of samples, PC2 offered further insight into the complexity and relationships within those groups.

TNF-α Scree Plot of Variance Stabilized Data
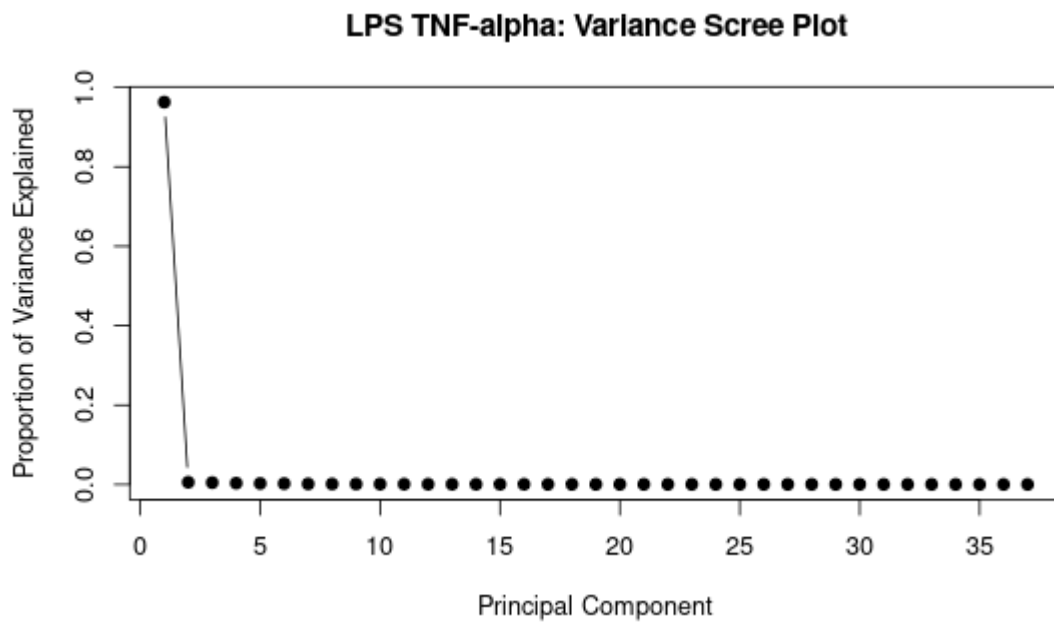
Visual Heatmaps of Hierarchical Clustering at different levels
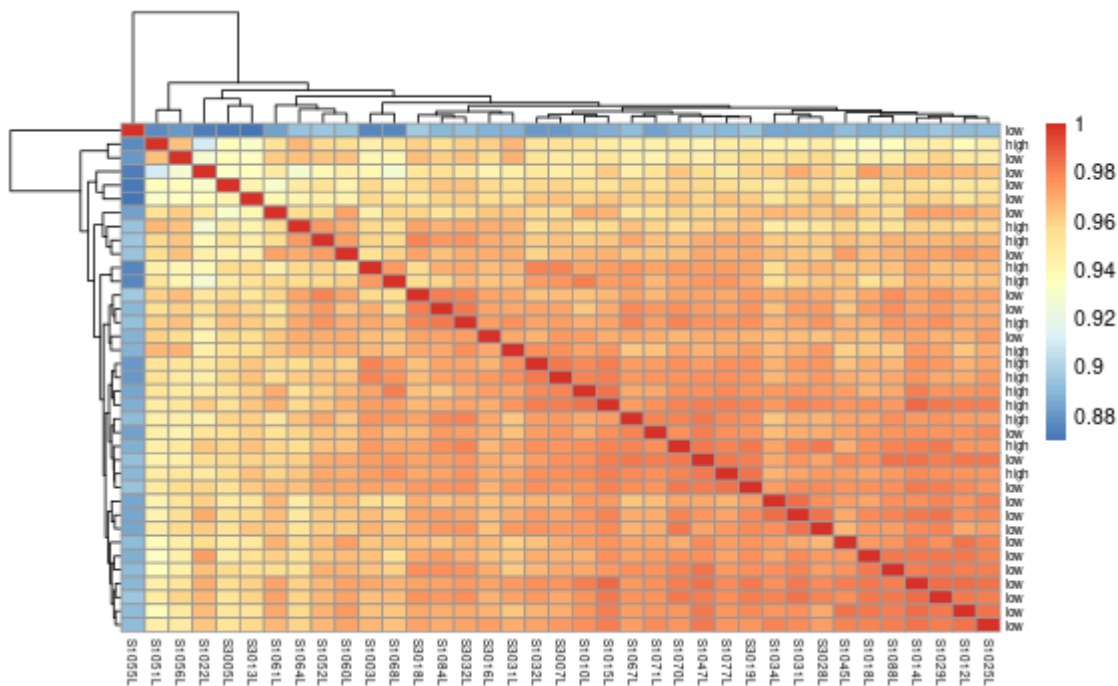
DIFFERENTIAL EXPRESSION

DESeq with Wald Test and $\mathrm{Log_2FC}$ *Shrinkage*

```
log2 fold change (MLE): sex Male vs Female
Wald test p-value: sex Male vs Female
DataFrame with 18195 rows and 6 columns
             baseMean log2FoldChange     lfcSE      stat    pvalue      padj
            <numeric>      <numeric> <numeric> <numeric> <numeric> <numeric>
AL139246.3   12.91226     -0.1690741  0.319468 -0.529237  0.596641  0.910014
PRDM16       38.65416      0.5269346  0.429557  1.226692  0.219938  0.776851
PEX10        80.66117     -0.0483698  0.126892 -0.381189  0.703063  0.933831
AP006222.1    8.58979     -0.0570738  0.321206 -0.177686  0.858970  0.975430
PEX14       265.94142      0.1422488  0.123712  1.149839  0.250210  0.793959
...               ...            ...       ...       ...       ...       ...
MT-TP       14.1266399      0.3724925  0.341273  1.091478  0.275063  0.802967
RNA5-8S4     0.0249268     -1.1449528  2.982399 -0.383903        NA        NA
AL592183.1 68.2486193     -0.0559103  0.407786 -0.137107  0.890946  0.981628
AC004556.1  7.1945880      0.6027209  0.740895  0.813503  0.415929  0.848757
AC240274.1 24.2628279      0.1131247  0.233587  0.484294  0.628177  0.915287


out of 18195 with nonzero total read count
adjusted p-value < 0.1
LFC > 0 (up)        : 19, 0.1%
LFC < 0 (down)      : 18, 0.099%
outliers [1]        : 1821, 10%
low counts [2]      : 0, 0%
(mean count < 0)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results
```
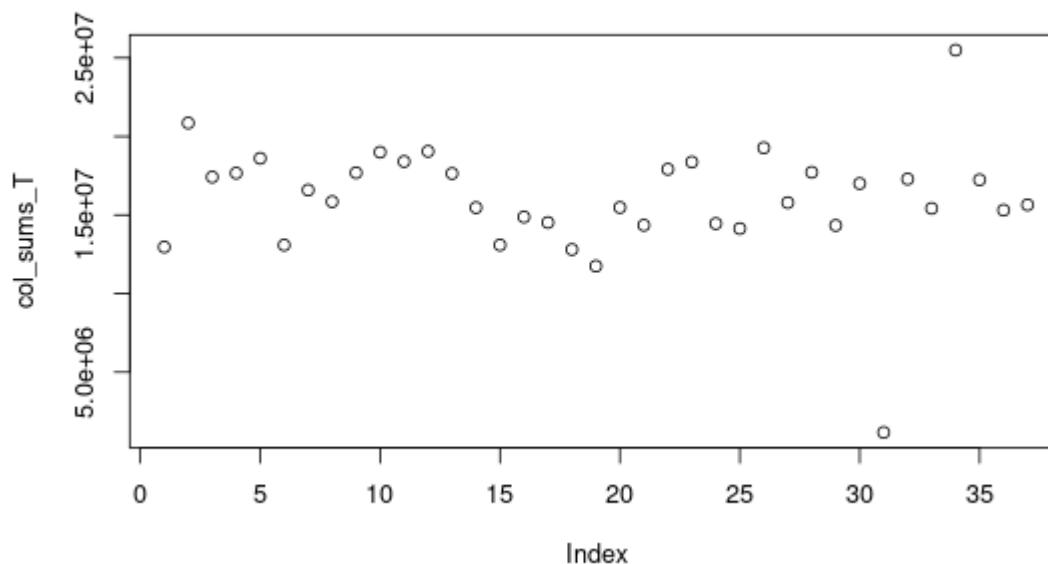

Total Raw Counts per Sample - LPS TNF-alpha

- This was consistent with the incipient investigation between LPS and media treated samples. While sample 3013L appeared to be an outlier potentially skewing the data, it was not removed from the dataset, to avoid introducing bias and because plotting normalized results identified no counts bellow the expected threshold.
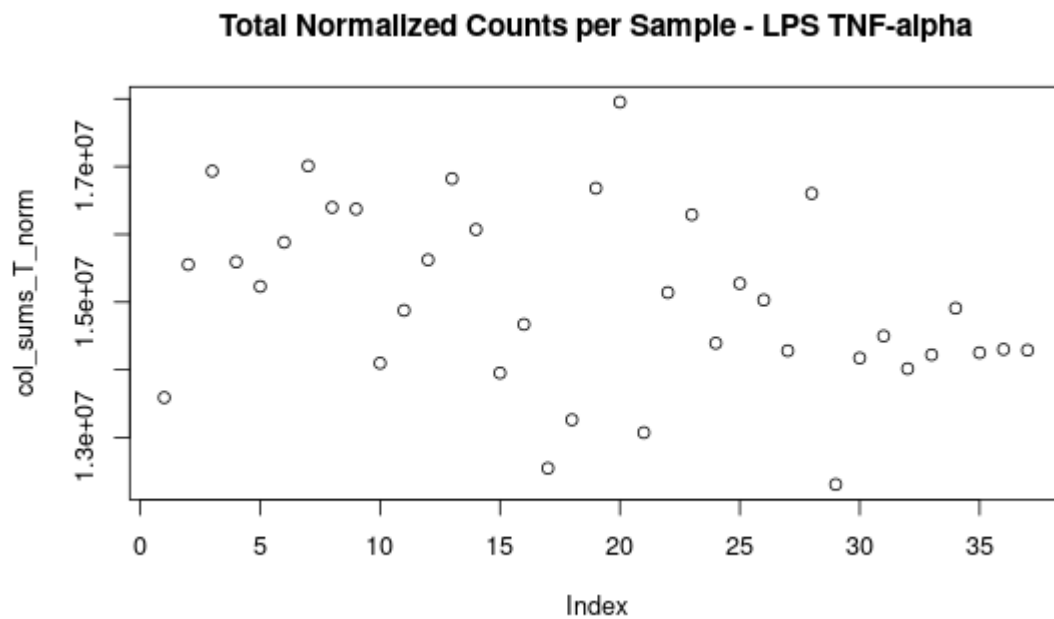
Dispersion Estimation Plot
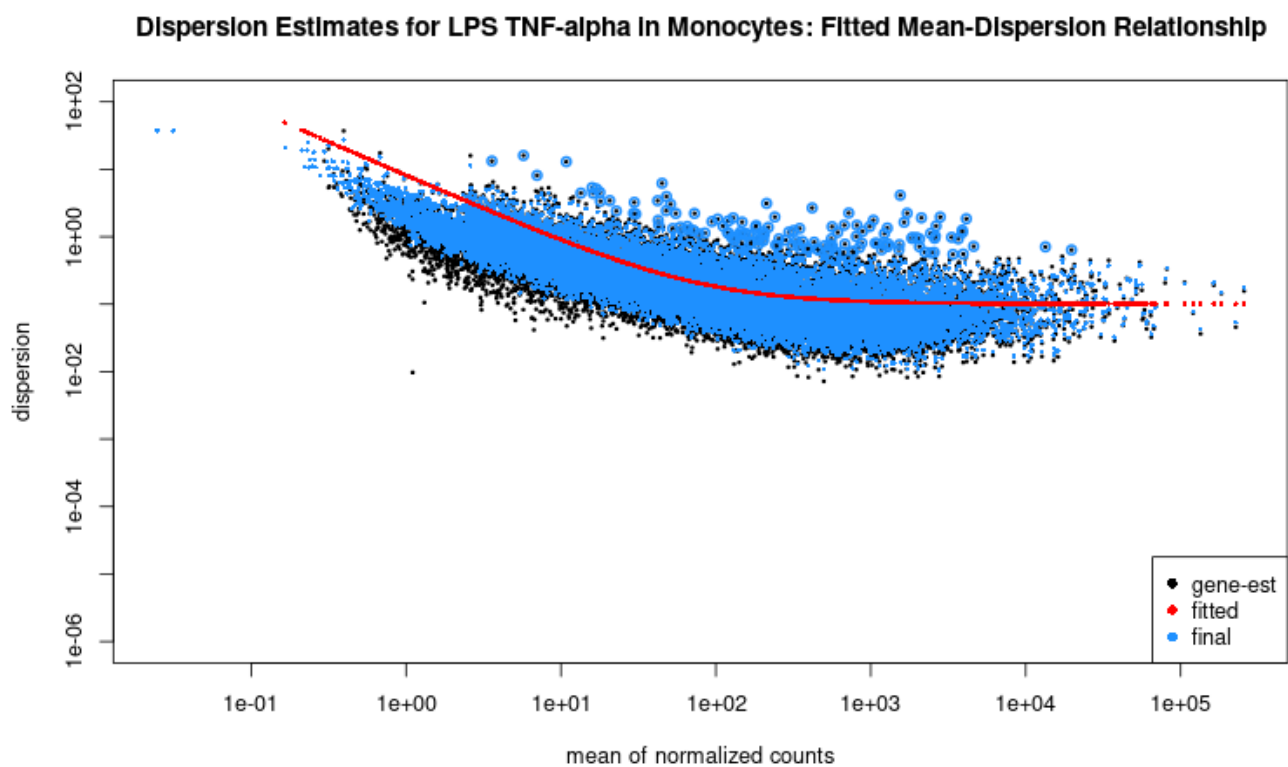
MA Plots

Shrunken $\log_2 \mathrm{FC}$ values were computed using the ***ashr*** method for the LPS TNF-α contrast. An initial **MA plot** visualized the overall shrinkage effect across all genes, while a second, filtered **MA plot** (padj ≤ 0.01) highlighted only significantly DEGs for clearer interpretation and downstream analysis. The accompanying **summary statistics** provided an overall snapshot of gene regulation of signifficant changes (p < 0.1 and padj ≤ 0.01, respectivelly).

Type ashr LPS TNF-alpha
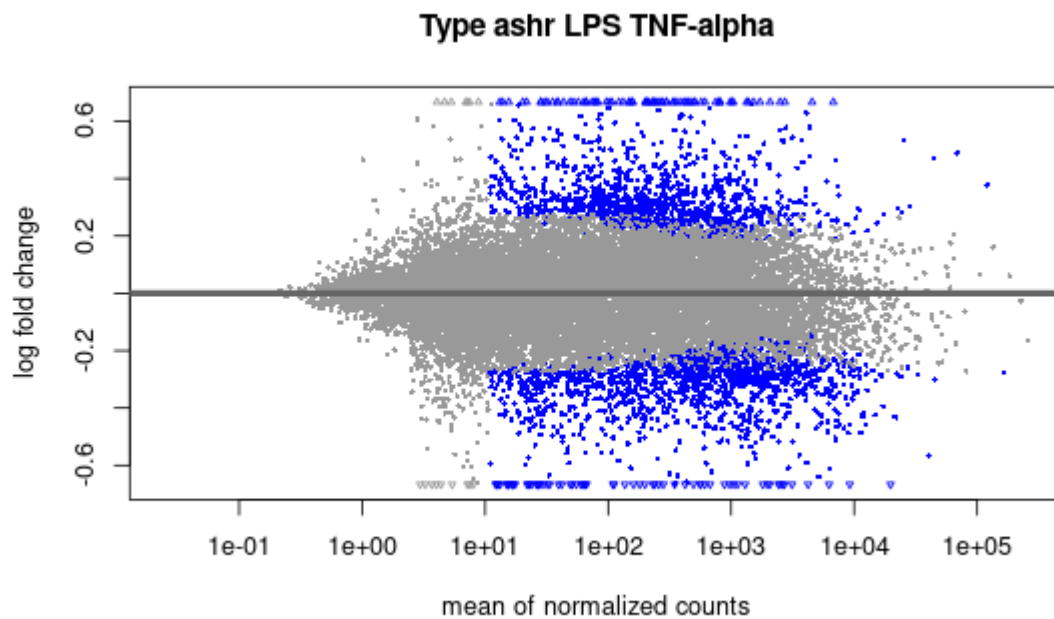

```
out of 18195 with nonzero total read count
adjusted p-value < 0.1
LFC > 0 (up)       : 1020, 5.6%
LFC < 0 (down)     : 1318, 7.2%
outliers [1]       : 36, 0.2%
low counts [2]     : 4914, 27%
(mean count < 11)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results
```

MA Plot with Type ashr Shrinkage and Contrast With Sex Covariate (padj <= 0.1) - LPS TNF-alpha
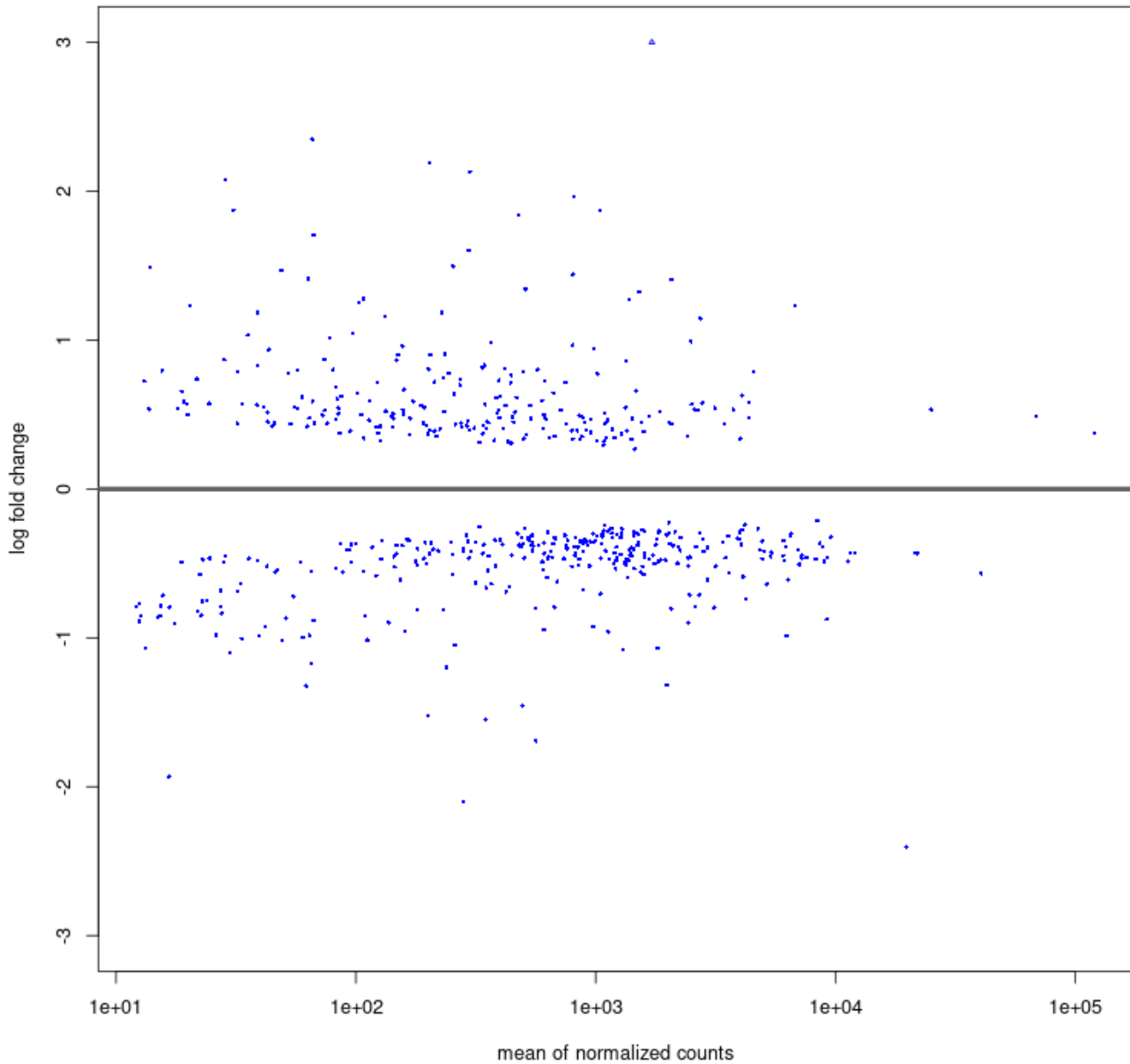

```
out of 555 with nonzero total read count
adjusted p-value < 0.1
LFC > 0 (up)       : 228, 41%
LFC < 0 (down)     : 327, 59%
outliers [1]       : 0, 0%
low counts [2]     : 0, 0%
(mean count < 11)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results
```
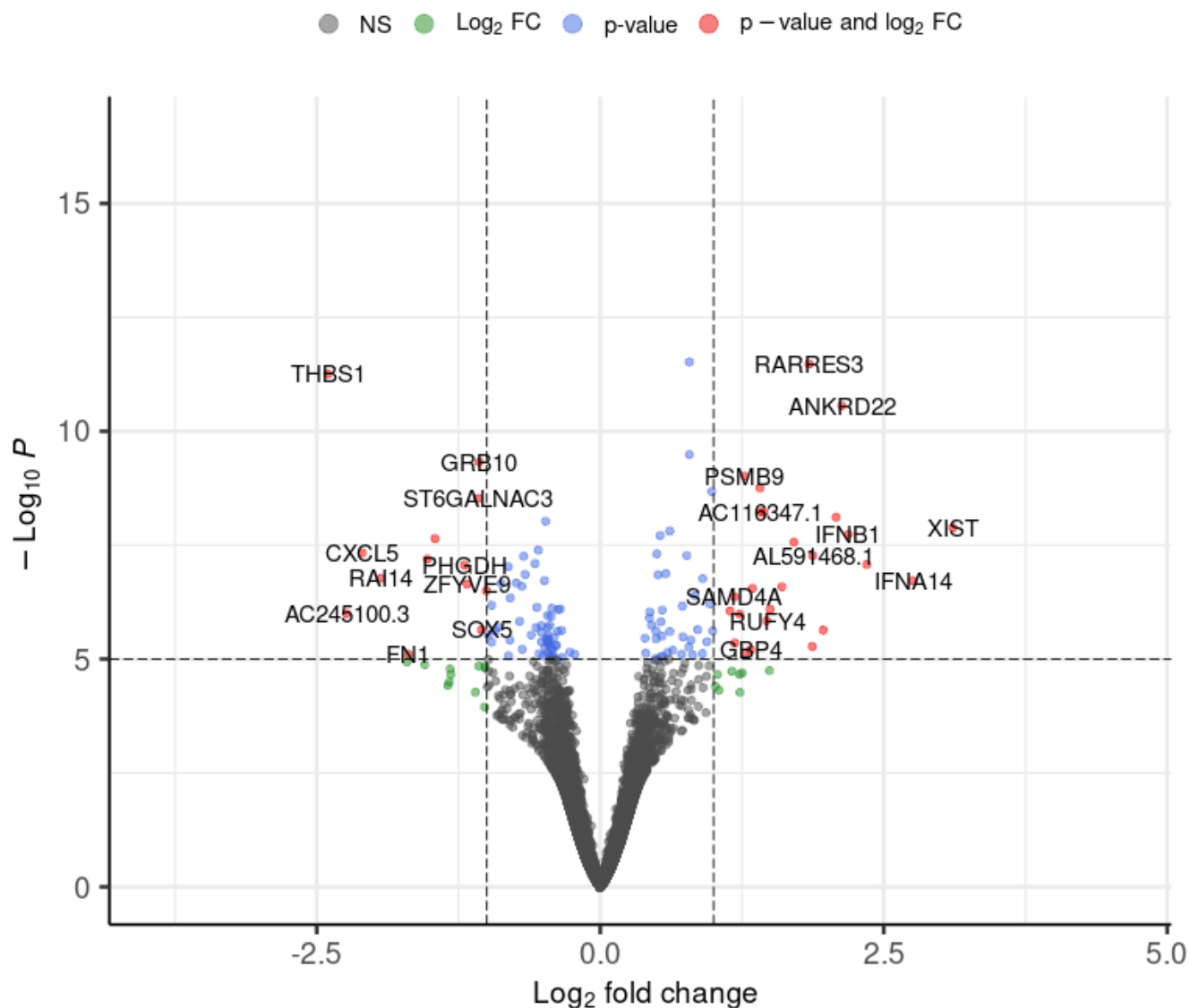
Volcano Plot

PATHWAY ANALYSIS

Annotation

Gene **annotation** was performed by mapping gene symbols to ENSEMBL IDs using *AnnotationDbi* and *org.Hs.eg.db*, followed by filtering for valid, high-expression entries (baseMean > 50) to ensure reliability in downstream analyses.

```
     symbol              baseMean          log2FoldChange          lfcSE              p
value              padj              ensgene
 Length:451         Min.   :    51.01   Min.   :-2.40341   Min.   :0.05229   Min.
:0.000e+00   Min.   :2.200e-08   Length:451
 Class :character   1st Qu.:   266.12   1st Qu.:-0.44958   1st Qu.:0.11040   1st Q
u.:1.018e-05   1st Qu.:1.007e-03   Class :character
 Mode  :character   Median :   749.00   Median :-0.32349   Median :0.14585   Media
n :6.223e-05   Median :3.030e-03   Mode  :character
                    Mean   :  2091.89   Mean   :-0.03981   Mean   :0.20091   Mean
:1.118e-04   Mean   :3.711e-03
                    3rd Qu.:  1547.26   3rd Qu.: 0.45766   3rd Qu.:0.23590   3rd Q
u.:1.890e-04   3rd Qu.:6.098e-03
                    Max.   :120511.00   Max.   : 3.10942   Max.   :0.97665   Max.
:4.141e-04   Max.   :9.883e-03
```

PPI analysis using STRING

```
  [1] "XIST"      "HAPLN3"   "IFNB1"    "ANKRD22"  "SERPING1" "LGALS3BP" "TGM2"
"APOL3"     "PSME2"    "APOL1"    "GBP4"     "NEURL3"   "PSMB9"    "KLHDC7B" "HLA-
DPA1" "SAMD4A"    "ZBP1"     "HLA-DPB1" "BCL2L14"
 [20] "HLA-DOA"   "UBE2L6"   "ST3GAL5"  "VAMP5"    "IL27"     "GIMAP4"   "SLC6A12"
"OSR2"      "IGF2BP3"  "ETV7"     "HLA-DMB"  "APOBEC3G" "CD40"     "OTOF"     "TMEM
229B" "PARP3"    "NT5C3A"   "PSMB10"   "IRF8"
 [39] "POLR3K"    "MT2A"     "TNFSF10"  "NUB1"     "GDPD5"    "GPBAR1"   "APOL2"
"ARL4C"     "MYEOV"    "IL15"     "CD274"    "SDC3"     "ZMYM3"    "MYO7B"    "LAP3
"     "GBP5"     "FIRRE"    "CYSLTR2"  "RANBP3L"
 [58] "AIM2"      "PLA1A"    "DNPEP"    "IL12RB1"  "CHST12"   "PPA1"     "IGFLR1"
"PLAC8"     "PSME1"    "CNDP2"    "GOLM1"    "IL15RA"   "C15orf48" "EVL"      "CD72
"     "ING4"     "PSMA4"    "KLF4"     "ZFAND4"
 [77] "AMER1"     "HLA-DRA"  "SECTM1"   "FGD2"     "GNGT2"    "ATOX1"    "GSDMD"
"PSMB8"     "TMEM126B" "CD300A"   "POPDC2"   "ACP2"     "IFI35"    "RAB8A"    "RNF1
38"    "ISOC2"    "IL31RA"   "AGPAT5"   "TPMT"
 [96] "TTC39A"    "RNF114"   "B2M"      "COA4"     "ACSS1"
```
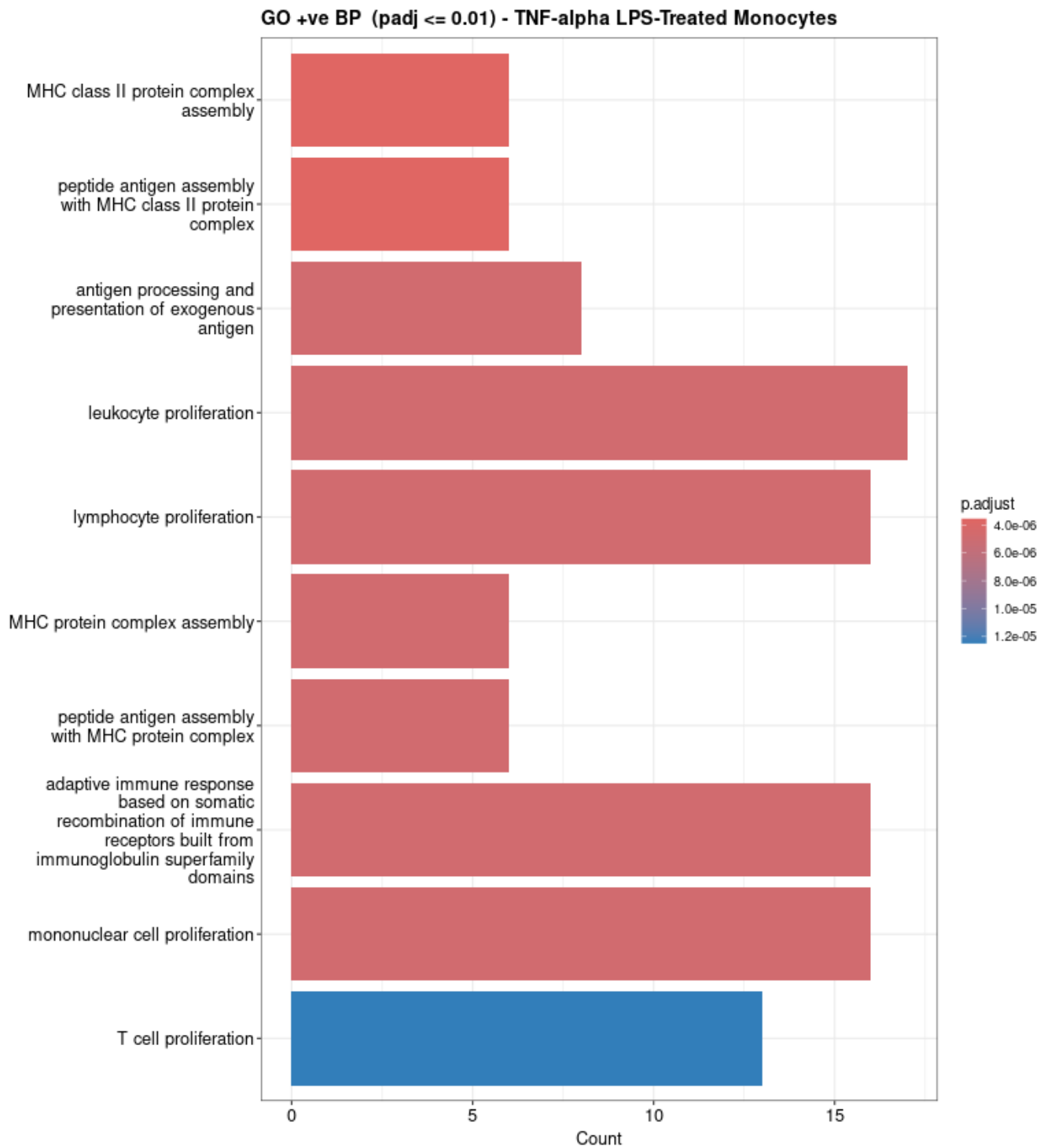
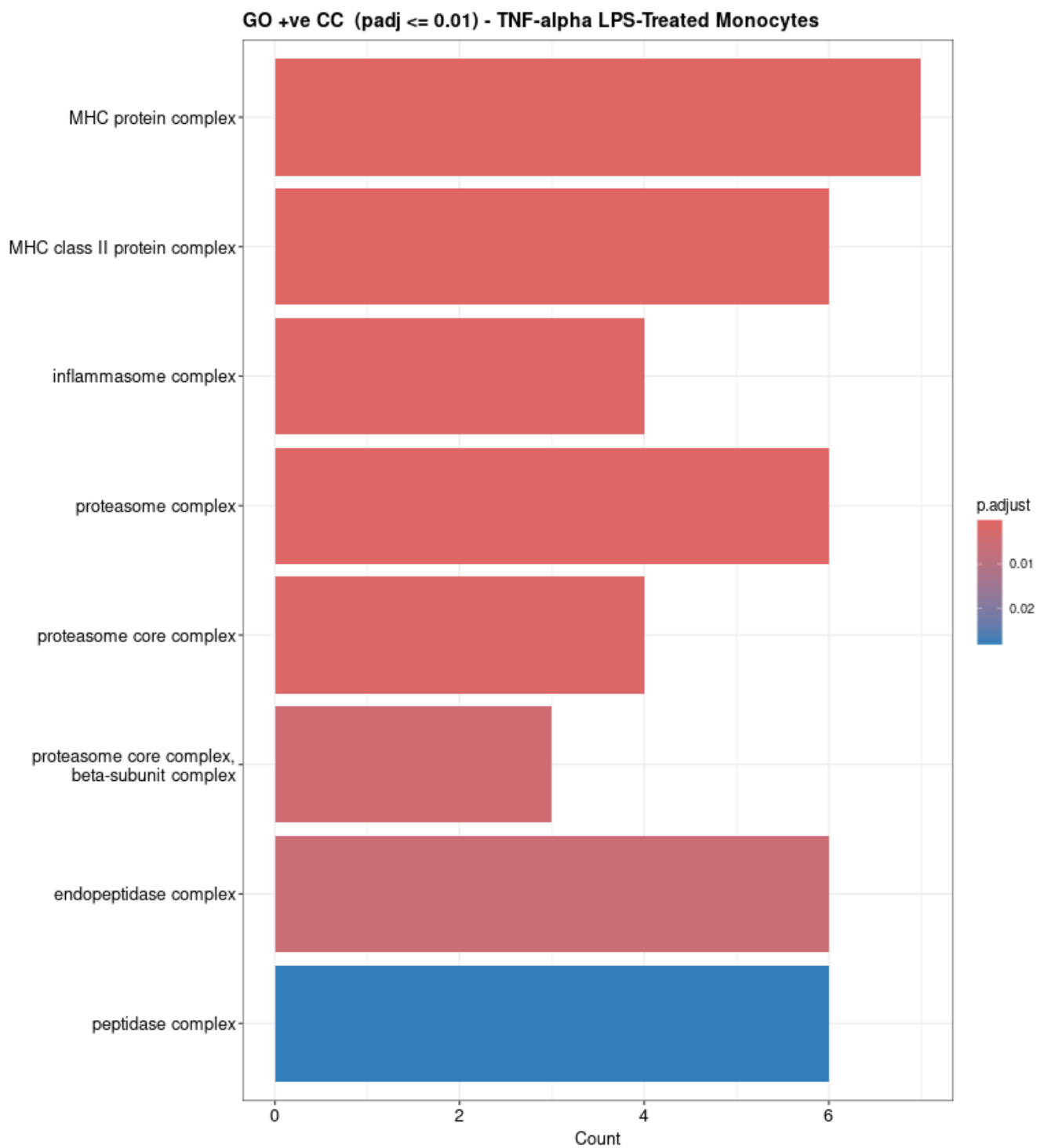Permalink STRING Output LPS TNF-α Upregulated link (https://version-12-0.string-db.org/cgi/network?
networkId=bGJ2fUAIpL2p)

```
  [1] "THBS1"        "CXCL5"       "FN1"        "MMP9"        "PSAT1"        "CHAC1
"        "NAV3"        "IL1R2"       "PHGDH"       "ZFYVE9"      "ST6GALNAC3"  "GRB
10"      "SOX5"         "SLC7A11-AS1" "HPGD"
 [16] "KANK1"        "C22orf42"    "ME1"        "KIFC3"       "KCNJ11"       "MMP19
"        "GJB2"        "GPR68"       "ADM2"        "DDIT4"       "CD177"       "LAC
C1"       "SOWAHC"       "ADAMTS2"     "AMPH"
 [31] "SESN2"        "SLC38A1"     "CXCL1"      "SLC7A1"      "ANKRD18A"     "SEMA6
B"        "IL7R"        "KLC1"        "CEBPG"       "INPP1"       "GPRC5A"      "SGP
P2"       "EBF1"         "MET"         "SLC26A2"
 [46] "SLC7A11"      "BMP6"        "RIT1"       "MMP14"       "GLIS3"        "CLIC4
"        "RASSF8"       "PXK"         "FGFR1"       "ACSL1"       "SERPINB8"    "ACV
R1"       "JRK"          "ZFYVE1"      "VEGFA"
 [61] "TPST1"        "IRS1"        "EIF1B"      "LMO4"        "RRAGC"        "BRPF3
"        "FMN1"        "CCL20"       "RFFL"        "LBX2-AS1"    "PPARD"       "FLT
1"        "NUP58"        "REST"        "NRIP3"
 [76] "BCAT1"        "MAEA"        "ULBP1"      "SETD5"       "BASP1"        "ATP13
A3"       "SNUPN"        "TXNRD1"      "ZXDC"        "GCLM"        "PNPLA1"      "LYR
M4"       "KIAA0232"     "HSPA13"      "SLC1A5"
 [91] "GOT1"         "CYP1B1"      "ASPH"       "SIPA1L2"     "KLHL9"        "PRDM4
"        "SSH2"         "KEAP1"       "PLD1"        "STRN3"
```
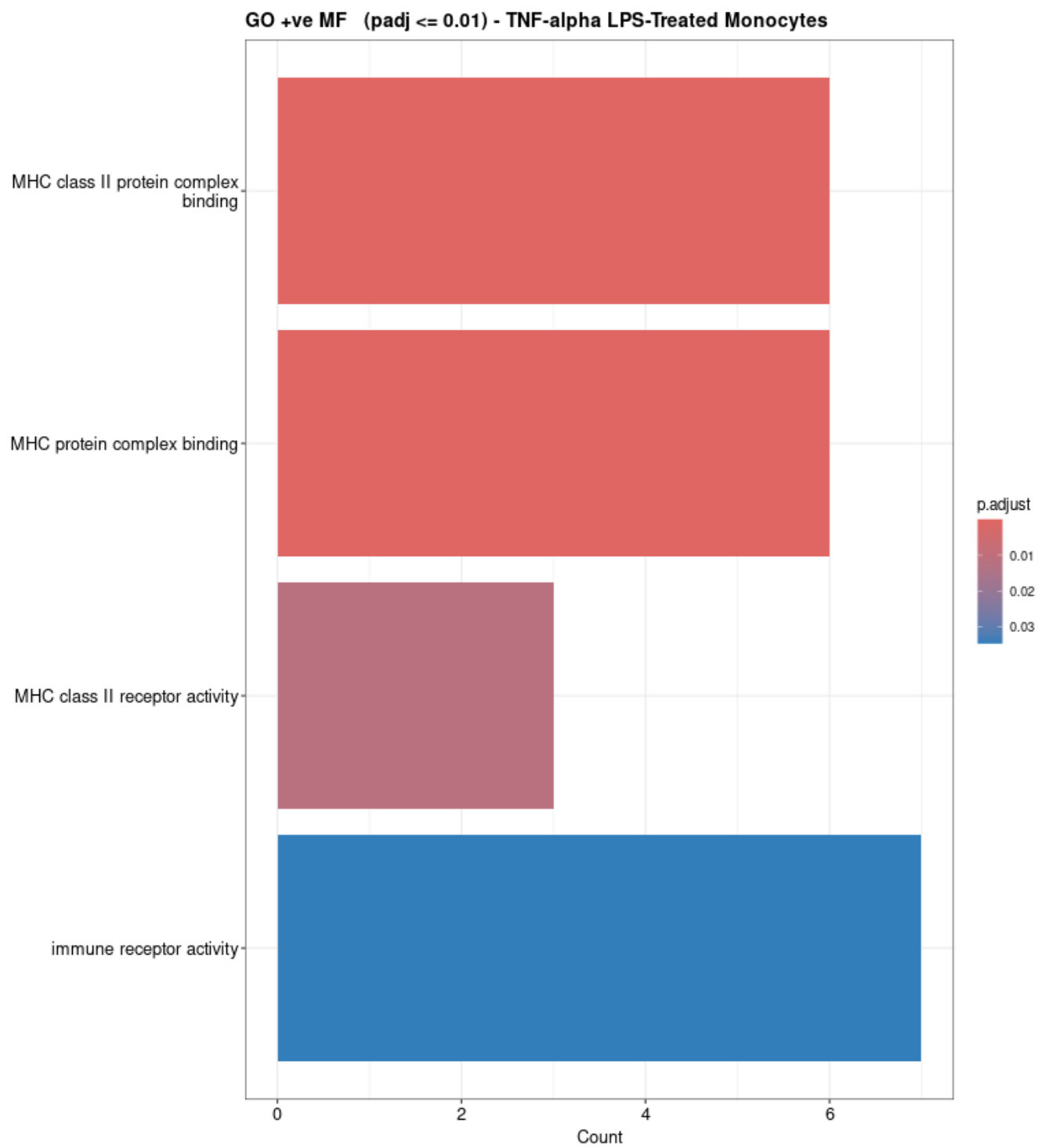
Permalink STRING Output LPS TNF-α Downregulated link (https://version-12-0.string-db.org/cgi/network?networkId=bhzs9lGfv73z)

GO Enrichment Analysis

Upregulated genes

GO +ve BP (padj <= 0.01) - TNF-alpha LPS-Treated Monocytes

GO +ve CC (padj <= 0.01) - TNF-alpha LPS-Treated Monocytes

GO +ve MF   (padj <= 0.01) - TNF-alpha LPS-Treated Monocytes

TNF-alpha LPS-Treated Monocytes: Positively Enriched Gene Ontology Terms (BP, CC, MF)

Downregulated genes

GO -ve BP (padj <= 0.01) - TNF-alpha LPS-Treated Monocytes

GO -ve CC (padj <= 0.01) - TNF-alpha LPS-Treated Monocytes

TNF-alpha LPS-Treated Monocytes: Negatively Enriched Gene Ontology Terms (BP, CC, MF)