

Pathway analysis

```
suppressMessages({
  library(limma)
  library(stringr)
  library(DOSE)
  library(clusterProfiler)
  library(org.Mm.eg.db)
  library(AnnotationDbi)
  library(enrichplot)
  library(scater)
  library(gridExtra)
  library(umap)
  library(Seurat)
})
```

Background

This report is an exploration of DE genes and enriched pathways from single-nucleus RNA-sequencing expression profiles of neurons sampled from the mid thoracic spinal cord of mice after spinal cord injury. There are 17 injured 3 uninjured samples and seven time points: control and one day, four days, seven days, 14 days, 1 month, and 2 months post injury. GEO accession number: GSE234774. This analysis explores differentially expressed genes and pathways that change in a time dependent manner across the time points.

Data overview

Below is a table showing the number of cells for each time point and replicate. There are also UMAPs of neuronal cells from the mid thoracic spinal cord. The first UMAP is coloured by condition (injured and uninjured) while the second is coloured by time point.

Table of replicates and cells

```
# Read in data
data <- readRDS("/verona/nobackup/dkim/projects/proj_degeneration/data/samples/injury/GSE234774_Tabulae-
data$condition <- factor(data$condition, levels=c("uninjured", "injured"))
data$timepoint <- factor(data$timepoint, levels=c("uninjured", "1d", "4d", "7d", "14d", "1m", "2m"))

data$replicate_clean <- factor(data$replicate_clean, levels=c("Uninjured-1", "Uninjured-2", "Uninjured-3",
"1d-1", "1d-2", "1d-3", "4d-1", "4d-2", "4d-3",
"7d-1", "7d-2", "7d-3",
"14d-1", "14d-2", "14d-3",
"1m-1", "1m-2", "1m-3",
"2m-1", "2m-2", "2m-3"))

table(data$timepoint, data$replicate_clean)
```

```
##
```

```
##          Uninjured-1 Uninjured-2 Uninjured-3 1d-1 1d-2 1d-3 4d-1 4d-2 4d-3
```

##	uninjured	877	2217	1263	0	0	0	0	0	0		
##	1d	0	0	0	551	2	894	0	0	0		
##	4d	0	0	0	0	0	0	22	17	705		
##	7d	0	0	0	0	0	0	0	0	0		
##	14d	0	0	0	0	0	0	0	0	0		
##	1m	0	0	0	0	0	0	0	0	0		
##	2m	0	0	0	0	0	0	0	0	0		
##												
##		7d-2	7d-3	14d-1	14d-2	14d-3	1m-1	1m-2	1m-3	2m-1	2m-2	2m-3
##	uninjured	0	0	0	0	0	0	0	0	0	0	0
##	1d	0	0	0	0	0	0	0	0	0	0	0
##	4d	0	0	0	0	0	0	0	0	0	0	0
##	7d	1309	753	0	0	0	0	0	0	0	0	0
##	14d	0	0	850	874	956	0	0	0	0	0	0
##	1m	0	0	0	0	0	612	999	1580	0	0	0
##	2m	0	0	0	0	0	0	0	0	689	1058	1112

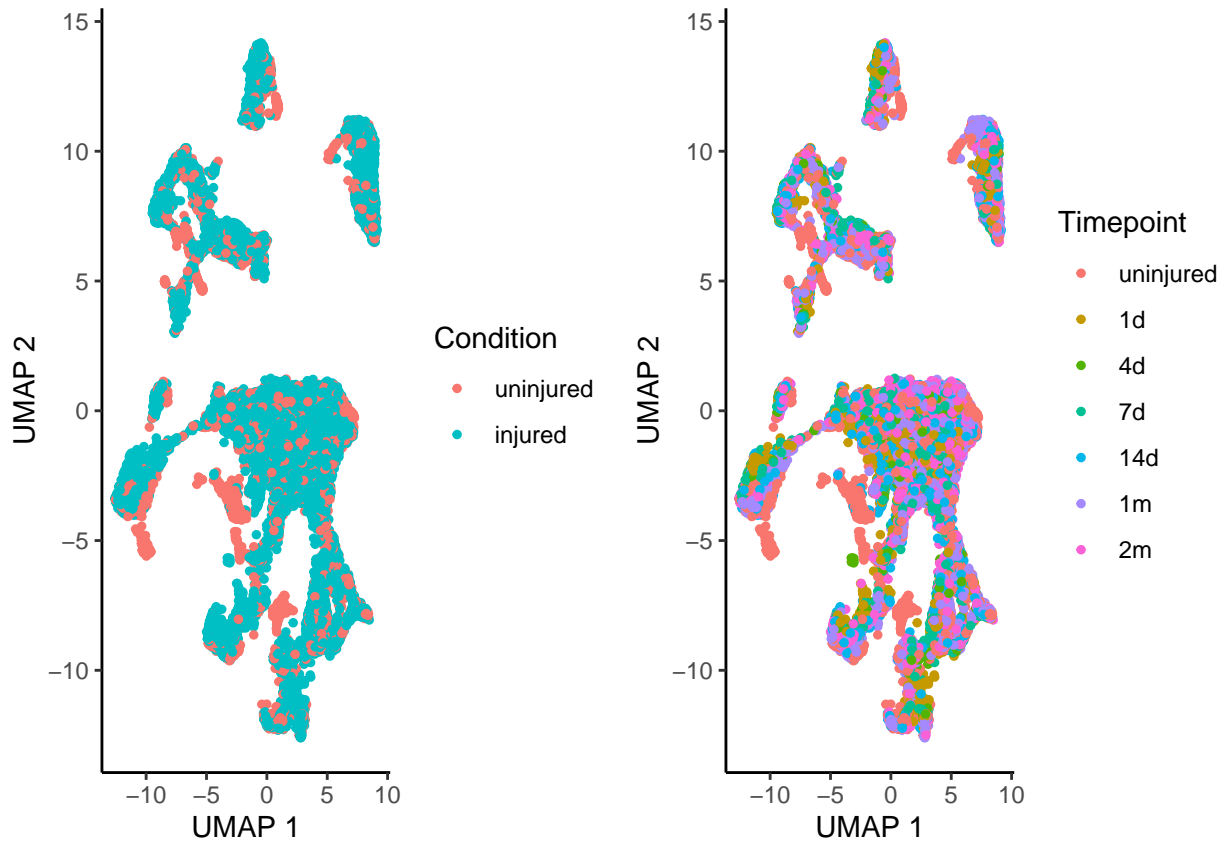
UMAP

```
# Extract UMAP embedding
umap <- data.frame(data@reductions$umap@cell.embeddings)

# Plot UMAP where cells are coloured by condition
umap$condition <- factor(data$condition, levels=c("uninjured", "injured"))
umap$timepoint <- factor(data$timepoint, levels=c("uninjured", "1d", "4d", "7d", "14d", "1m", "2m"))
p1 <- ggplot(umap, aes(x=umap_1, y=umap_2, color=condition)) +
  geom_point(size=1) +
  theme(panel.background=element_blank(),
        axis.line=element_line(color="black")) +
  labs(x="UMAP 1", y="UMAP 2", color="Condition")

# Plot UMAP where cells are coloured by timepoint
p2 <- ggplot(umap, aes(x=umap_1, y=umap_2, color=timepoint)) +
  geom_point(size=1) +
  theme(panel.background=element_blank(),
        axis.line=element_line(color="black")) +
  labs(x="UMAP 1", y="UMAP 2", color="Timepoint")

grid.arrange(p1, p2, ncol=2)
```



Injured vs uninjured across time

Differential expression analysis

Differential expression analysis to identify genes that are differentially expressed in a time dependent manner. Below are the top 10 DE genes after running Limma (adjusted p-value < 0.05). Tests were corrected for multiple testing using Benjamini Hochberg correction and no logFC threshold was applied. Genes with a positive logFC are those that linearly increase in expression over time.

```
design <- model.matrix(~as.numeric(factor(data$timepoint)))
lm <- lmFit(data@assays$originalexp$data, design=design)
lm <- eBayes(lm)
tt <- topTable(lm, coef=2, n=Inf, sort.by="p")
tmp_tt <- head(tt, n=10)
tmp_tt <- round(tmp_tt, digits=4)
tmp_tt
```

##	logFC	AveExpr	t	P.Value	adj.P.Val	B
## Psap	-0.0779	0.3034	-42.6497	0	0	853.4761
## Ctsb	-0.0468	0.0869	-46.5056	0	0	1007.6735
## Hexa	-0.0338	0.0672	-43.8723	0	0	901.2161
## Itm2b	-0.0530	0.1332	-38.1867	0	0	688.6532
## Sdcbp	-0.0366	0.1354	-35.7452	0	0	604.9949
## Pabpc1	-0.0282	0.0791	-35.5705	0	0	599.1897
## Serinc3	-0.0249	0.0651	-33.7086	0	0	538.8692
## Rrbp1	-0.0362	0.1425	-32.9142	0	0	513.9979
## Tpp1	-0.0199	0.0460	-32.4035	0	0	498.2829

```
## Cts1      -0.0190  0.0333 -32.1858      0      0  491.6505
```

Pathway enrichment analysis

Enriched GO terms (biological pathways) for both up- and down-regulated DE genes that change linearly post spinal cord injury. Tests were corrected for multiple testing using Benjamini Hochberg correction (adjusted p-value < 0.05).

```
up_genes <- rownames(tt)[tt$adj.P.Val<0.05 & tt$logFC>0]

up_ego <- enrichGO(gene=up_genes,
  keyType="SYMBOL",
  OrgDb=org.Mm.eg.db,
  ont="BP",
  pAdjustMethod="BH",
  pvalueCutoff=0.01,
  qvalueCutoff=0.05,
  readable=TRUE)

down_genes <- rownames(tt)[tt$adj.P.Val<0.05 & tt$logFC<0]

down_ego <- enrichGO(gene=down_genes,
  keyType="SYMBOL",
  OrgDb=org.Mm.eg.db,
  ont="BP",
  pAdjustMethod="BH",
  pvalueCutoff=0.01,
  qvalueCutoff=0.05,
  readable=TRUE)

up_ego@result$p.adjust <- signif(up_ego@result$p.adjust , digits=2)
down_ego@result$p.adjust <- signif(down_ego@result$p.adjust , digits=2)

p1 <- dotplot(up_ego, showCategory=5, font.size=8) + ggtitle("Upregulated")
p2 <- dotplot(down_ego, showCategory=5, font.size=8) + ggtitle("Downregulated")

grid.arrange(p1, p2, ncol=2)
```

