# Lab 4

## Knowledge Based Systems

Rasmus Hammar

```
library(csVisuNet)
```

# Task 1

1)

```
autcon <- readRDS("autism_control.RDS")
str(autcon)
```

```
'data.frame':   146 obs. of  36 variables:
 $ MAP7     : num  2.36 2.52 2.6 2.37 2.03 ...
 $ PTGS2    : num  3.76 3.74 3.76 3.76 3.76 ...
 $ NCKAP5L  : num  1.97 2.02 1.89 2.07 1.97 ...
 $ ZSCAN18  : num  2.42 2.32 2.49 2.38 2.33 ...
 $ RHPN1    : num  2.63 2.51 2.67 2.7 2.65 ...
 $ PPOX     : num  2.05 2.07 2.03 2.17 2.09 ...
 $ NPR2     : num  2.61 2.58 2.54 2.54 2.55 ...
 $ NCS1     : num  2.29 2.4 2.28 2.47 2.42 ...
 $ PSMG4    : num  2.34 2.42 2.46 2.51 2.45 ...
 $ SCIN     : num  1.66 1.67 1.63 1.7 1.53 ...
 $ CSTB     : num  2.39 2.39 2.26 2.46 2.41 ...
 $ TSPOAP1  : num  2.67 2.63 2.64 2.56 2.63 ...
 $ TCP11L1  : num  2.26 2.45 2.28 2.24 2.3 ...
 $ 234817_at: num  1.61 1.63 1.56 1.75 1.67 ...
 $ TMLHE    : num  1.94 2.01 1.89 2.19 1.83 ...
 $ PSMD4    : num  3.34 3.33 3.34 3.4 3.33 ...
 $ ZFP36L2  : num  3.16 3.05 3.05 2.97 2.8 ...
 $ B3GNT7   : num  2.51 2.55 2.57 2.35 2.54 ...
 $ MSI2     : num  2.26 2.19 2.41 1.72 1.85 ...
 $ CAPS2    : num  1.25 1.14 1.31 1.26 1.33 ...
 $ MIR646HG : num  1.59 1.63 1.65 1.6 1.75 ...
 $ CLDN17   : num  2.33 2.41 2.41 2.41 2.44 ...
 $ BAHD1    : num  2.93 2.95 3.01 2.89 2.93 ...
 $ OR51B5   : num  1.95 1.87 1.86 1.9 1.87 ...
 $ C11orf95 : num  2.78 2.76 2.79 2.55 2.8 ...
 $ ATXN8OS  : num  2.27 2.26 2.21 2.36 2.23 ...
```

```
 $ NRG2     : num  2.28 2.18 2.4 2.37 2.41 ...
 $ LOC400655: num  1.44 1.59 1.28 1.52 1.42 ...
 $ GJA9     : num  2.09 2.14 2.01 2.16 2.09 ...
 $ VPS8     : num  1.84 1.81 1.82 1.84 1.82 ...
 $ FLRT2    : num  1.59 1.51 1.61 1.56 1.43 ...
 $ C1QTNF7  : num  1.32 1.42 1.39 1.28 1.54 ...
 $ KLF8     : num  2.23 2.17 2.23 2.22 2.29 ...
 $ CWF19L2  : num  1.24 1.3 1.34 1.4 1.4 ...
 $ DEPDC1   : num  1.62 1.77 1.73 1.75 1.76 ...
 $ decision : Factor w/ 2 levels "autism","control": 2 2 2 2 2 2 2 2 2 2 ...
```

2)

```
ros <- rosetta(autcon)
```

With 193 rules in total, the quality looks fine.

3)

```
library(RCy3)
cytoscapePing()
vis <- visunetcyto(ros$main)
```

```
vis <- readRDS("vis.RDS")
```

Minimum meanAccuracy = 0.88

Minimum meanSupp = 13.0

Minimum meanDecisionCoverage = 0.245

Difference between the subnetworks is mainly that control has many more nodes and main/ biggest nodes are different genes.

Strongest connected nodes:

- Autism: PSMG4
- Control: MAP7

RHPN1 seems to be a cancer gene and there is no obvious connection to autism.

`vis` is a list with information on the networks, such as nodes and edges.

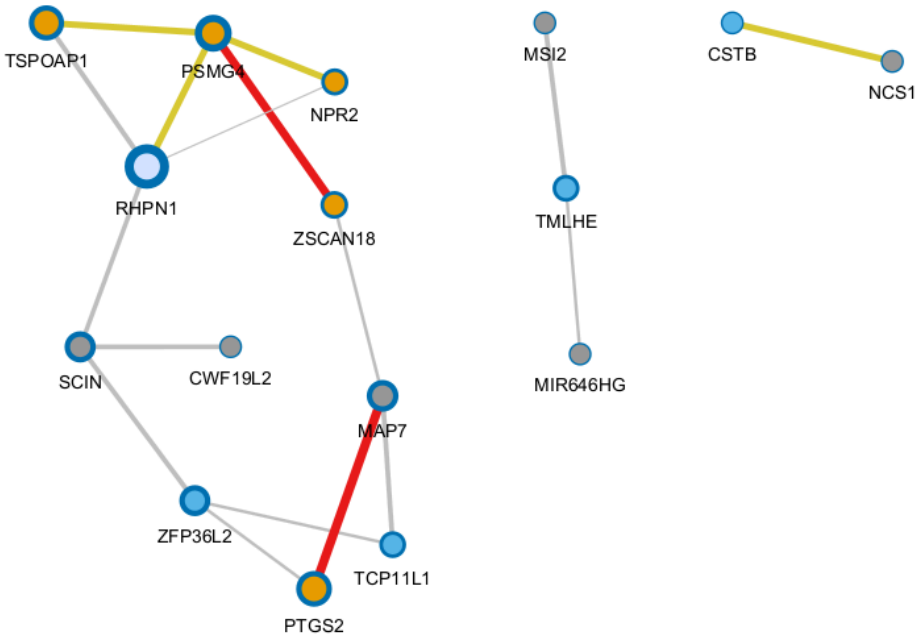


Figure 1: Autism subnetwork

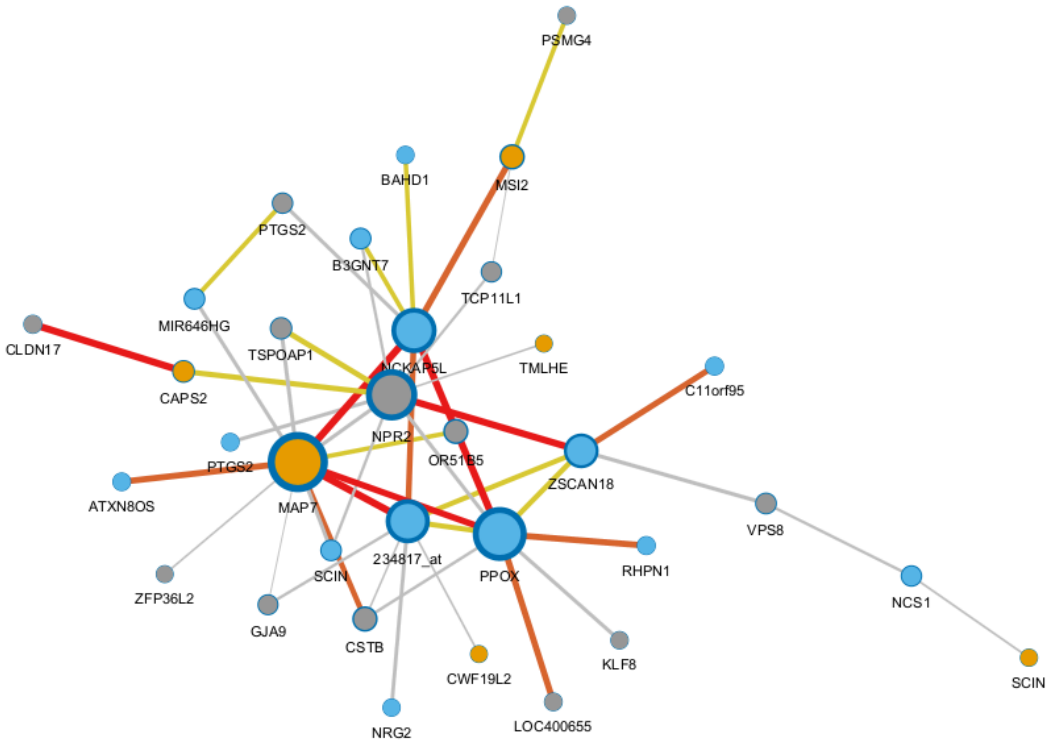


Figure 2: Control subnetwork

4)

```r
SFARI <- readRDS("SFARI_Genes.RDS")
vis_genes <- vis$all$nodes$label

overlapped_genes <- SFARI[which(SFARI$gene.symbol %in% vis_genes), ]
overlapped_genes[, c(2, 1, 3:4)] |>
  knitr::kable(row.names = FALSE)
```

| gene.symbol | status | gene.name | chromosome |
|---|---|---|---|
| NCKAP5L | 9 | NCK-associated protein 5-like | 12 |
| PTGS2 | 9 | prostaglandin-endoperoxide synthase 2 | 1 |
| TMLHE | 9 | trimethyllysine hydroxylase, epsilon | X |
| TSPOAP1 | 9 | TSPO associated protein 1 | 17 |

```r
overlapped_genes[, c(2, 5:8)] |>
  knitr::kable(row.names = FALSE)
```

| gene.symbol | genetic.category | gene.score | syndromic | number.of.reports |
|---|---|---|---|---|
| NCKAP5L | Rare Single Gene Mutation | 5 | 0 | 2 |
| PTGS2 | Genetic Association, Functional | 4 | 0 | 6 |
| TMLHE | Rare Single Gene Mutation, Genetic Association | 3 | 0 | 5 |
| TSPOAP1 | Rare Single Gene Mutation | 4 | 0 | 5 |

5)

a)

```r
library(org.Hs.eg.db)
library(clusterProfiler)
```

b)

```r
gene_entrez <- bitr(
  unique(as.character(vis_genes)),
  fromType = "SYMBOL",
  toType = "ENTREZID",
  OrgDb = org.Hs.eg.db
)
```

```
'select()' returned 1:1 mapping between keys and columns
```

```
Warning in bitr(unique(as.character(vis_genes)), fromType = "SYMBOL", toType =
"ENTREZID", : 9.68% of input gene IDs are fail to map...
```

c)

```
genes_GO <- groupGO(
  gene = unique(gene_entrez$ENTREZID),
  OrgDb = org.Hs.eg.db,
  ont = "MF",
  level = 5,
  readable = T
)
```

```
most_freq <- as.data.frame(genes_GO) |>
  dplyr::filter(Count >= max(Count)) |>
  dplyr::select(-ID) |>
  t()
rownames(most_freq) <- c("Description", "Count", "GeneRatio", "geneID")
knitr::kable(most_freq)
```

| | GO:0043169 |
|---|---|
| Description | cation binding |
| Count | 9 |
| GeneRatio | 9/28 |
| geneID | ZSCAN18/CAPS2/PTGS2/VPS8/SCIN/NCS1/KLF8/TMLHE/ZFP36L2 |

d)

```
x1 <- as.data.frame(genes_GO)[
  which(as.data.frame(genes_GO)$Count > 1),
  c('ID', 'Count')
]
x <- as.numeric(x1[, 2])
names(x) <- x1$ID
barplot(x, col = rainbow(20), las = 2)
```

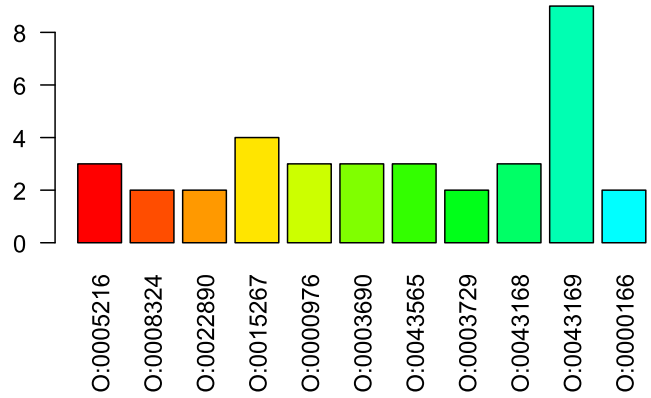

Figure 3: GO entries with count > 0.

Cation binding (GO:0043169) is an integral part of how cells, but specifically neurons, work. If this process is being regulated differently that may contribute to phenotypic differences observed as autism.