04: Over-representated GO terms in nominally significant proteins

Eneko Villanueva,Rayner Queiroz, Manasa Ramakrishna, Tom Smith

November 12, 2019

Contents

1. Introduction

Here, we will look at the over-representated GO terms in the "top" proteins according to their estimated changes in RNA binding. We detect 225 proteins different in cells treated with 100uM NaAs2 and 184 proteins in those cells treated with 400uM NaAs2. We want to see if there is a functional theme to either set of proteins that can shed light on how the cells react to sodium arsenite. We know they trigger heat-shock response and stress granule formation but can we detect any of these proteins that are also RNA-binding proteins.

```
# GO terms
human_go <- readRDS("../shared_files/h_sapiens_go_full.rds")

# Proteomics data
Ctrl_vs_100uM <- read.delim("../results/Ctrl-vs-100uM-Arsenite-Treated-BLOCK-rawp-le-0.05.csv")
Ctrl_vs_400uM <- read.delim("../results/Ctrl-vs-400uM-Arsenite-Treated-BLOCK-rawp-le-0.05.csv")
As_100_vs_400uM <- read.delim("../results/100-vs-400uM-Arsenite-Treated-BLOCK-rawp-le-0.05.csv")

# To perform the GO over-representation correctly,
# we need to adjust for the abundance of the
# proteins in the cell lysate.
total_as_protein_quant <- readRDS("../results/total_as_res_pro_agg_norm")
mean_as_expression <- as.data.frame(apply(2^exprs(total_as_protein_quant),
    1, mean))
colnames(mean_as_expression) <- "mean_expression"
```

```
tmp_100uM <- Ctrl_vs_100uM %>% merge(mean_as_expression,
    by.x = "uniprot_id", by.y = "row.names") %>% mutate(sig = P.Value <
    0.05, sig_up = (P.Value < 0.05 & logFC > 0), sig_dw = (P.Value <
    0.05 & logFC < 0)) %>% arrange(adj.P.Val)

tmp_400uM <- Ctrl_vs_400uM %>% merge(mean_as_expression,
    by.x = "uniprot_id", by.y = "row.names") %>% mutate(sig = P.Value <
    0.05, sig_up = (P.Value < 0.05 & logFC > 0), sig_dw = (P.Value <
    0.05 & logFC < 0)) %>% arrange(adj.P.Val)

tmp_100_400uM <- As_100_vs_400uM %>% merge(mean_as_expression,
    by.x = "uniprot_id", by.y = "row.names") %>% mutate(sig = P.Value <
    0.05, sig_up = (P.Value < 0.05 & logFC > 0), sig_dw = (P.Value <
    0.05 & logFC < 0)) %>% arrange(adj.P.Val)
```

```
p1 = plotSigTerms(tmp_100uM, "sig")
```

```
## [1] 211
## No terms were significantly enriched! Return NULL objectsNULL
## NULL
```

```
p2 = plotSigTerms(tmp_100uM, "sig_up")
```

```
## [1] 139
## No terms were significantly enriched! Return NULL objectsNULL
## NULL
```

```
p3 = plotSigTerms(tmp_100uM, "sig_dw")
```

```
## [1] 72
## No terms were significantly enriched! Return NULL objectsNULL
## NULL
```

```
p4 = plotSigTerms(tmp_400uM, "sig")
```

```
## [1] 176
## No terms were significantly enriched! Return NULL objectsNULL
## NULL
```

```
p5 = plotSigTerms(tmp_400uM, "sig_up")
```

```
## [1] 121
## No terms were significantly enriched! Return NULL objectsNULL
## NULL
```

```
p6 = plotSigTerms(tmp_400uM, "sig_dw")
```

```
## [1] 55
## No terms were significantly enriched! Return NULL objectsNULL
## NULL
```

```
p7 = plotSigTerms(tmp_100_400uM, "sig")
```

```
## [1] 49
## No terms were significantly enriched! Return NULL objectsNULL
## NULL
```

```
p8 = plotSigTerms(tmp_100_400uM, "sig_up")
```

```
## [1] 18
## No terms were significantly enriched! Return NULL objectsNULL
## NULL
```

```
p9 = plotSigTerms(tmp_100_400uM, "sig_dw")
```

```
## [1] 31
## No terms were significantly enriched! Return NULL objectsNULL
## NULL
```

2. Conclusions

From what we can see above, none of the lists of genes have any significant functional themes surrounding them. This is telling us that in this instance, treating cells with NaAs2 at both 100uM and 400uM dosages hasn't had a huge biological impact on the cells.