A network motif-enrichment approach to mapping genetic dysregulation in ALS

SAMUEL SLEDZIESKI, MALEK KABANI

What is Amyotrophic Lateral Sclerosis (ALS)?

- Neurodegenerative/neuromuscular disease
- Affects motor neurons in the spinal cord and throughout the body leading to a loss of mobility
- 30,000 people are estimated to have ALS (5,000 new cases each year)
- 10% of cases are known as familial ALS and are linked to known genetic mutations
- 90% of cases are sporadic ALS, which is not linked to any known genetic mutation
- **Goal: identify dysregulated disease genes in ALS to elucidate mechanism, identify possible therapeutic targets**

https://www.als.org/understanding-als/what-is-als, https://www.umich-als.org

ALS amyotrophic lateral sclerosis

Motor neurons in the brain and spinal cord degenerate in ALS, which in turn leads to loss of muscle function.

normal nerve (muscle contracts)
diseased nerve (muscle is unable to contract)

There is no cure. ALS is always fatal.

Patients to loose the ability to:
Speak **Grab Objects**
Swallow **Move**
Walk **Breathe**
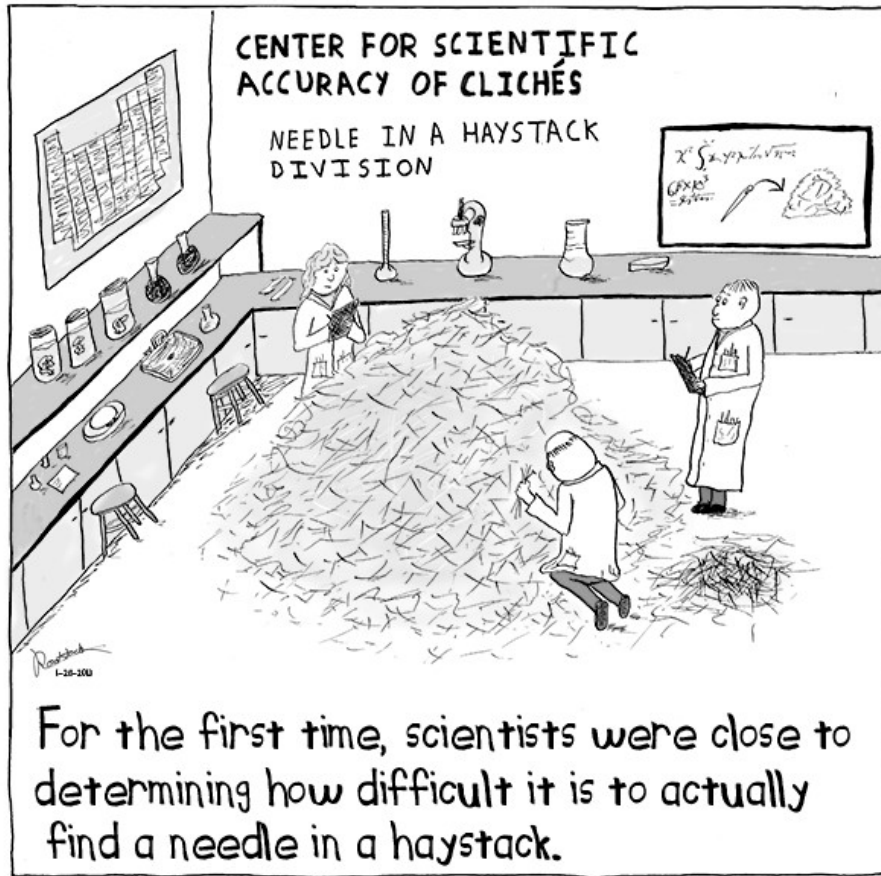
By 2040, ALS incidence will increase worldwide by 70%.

every **90 min** someone is diagnosed and passes from ALS

life expectancy is **2 to 4 years**

Disease genes are needles in a haystack

Center for Scientific Accuracy of Clichés — Needle in a Haystack Division. "For the first time, scientists were close to determining how difficult it is to actually find a needle in a haystack."
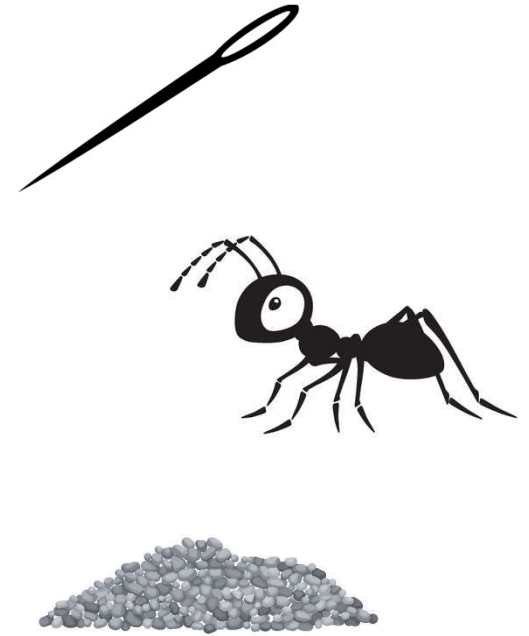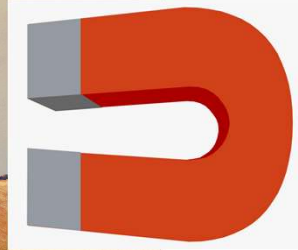

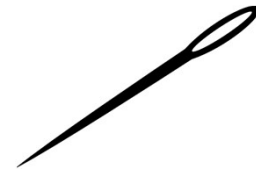https://ivebecomemyparents.com/2015/03/22/cliches-hot-tub-parties-and-puberty-a-post-in-pictures/

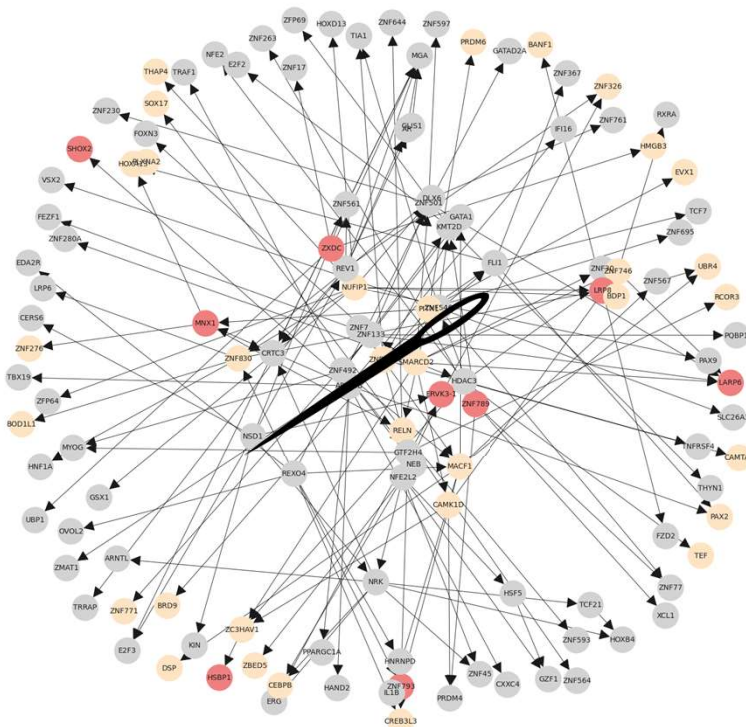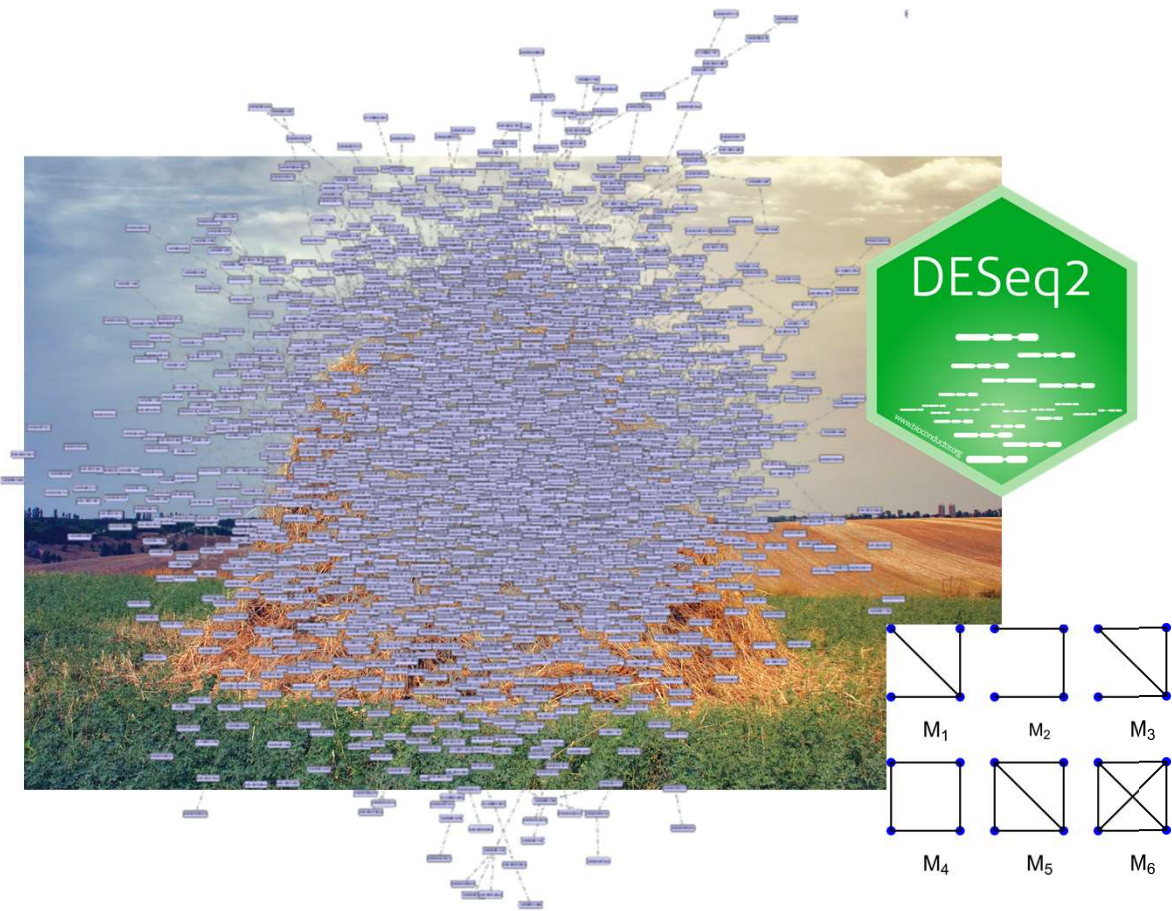# Disease genes are needles in a haystack

Disease genes are needles in a haystack

# Disease genes are needles in a haystack

# Combining filters: differential expression and network structure

Differential Gene Expression

- Differential gene expression analysis was performed using the **DESeq2 [1]** package
- Identified 4,875 genes differentially expressed at adjusted p-value <= 0.05 (iPSC)
- Identified 392 DEG at adjusted p-value <= 1e-5 (iPSC)
- Shown here – top 30 by adjusted p-value, sorted by log-fold change (LFC)
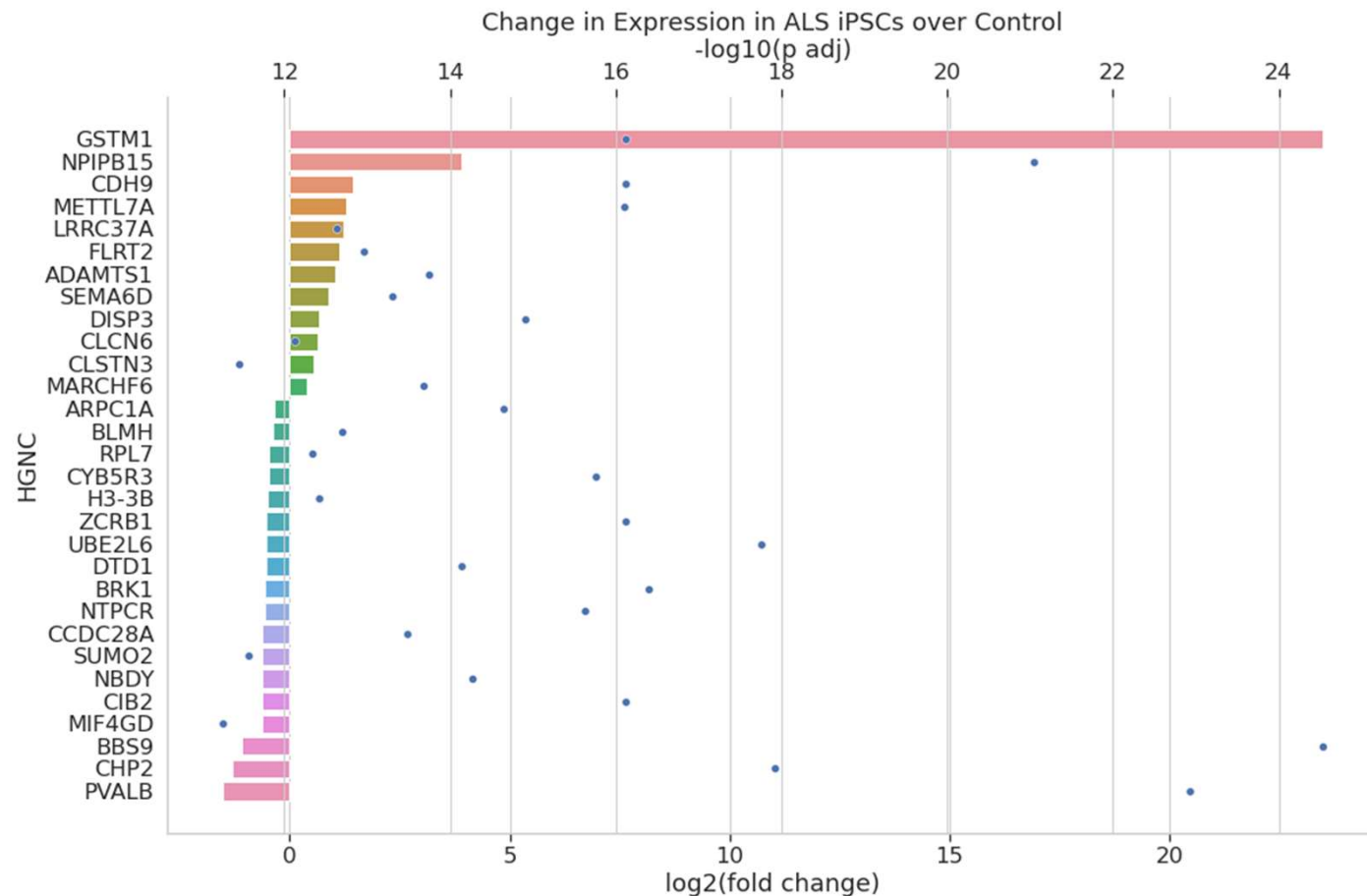
Reconstruction of four regulatory networks

- Regulatory networks were inferred using **GENIE3 [2]**
- Strength of regulatory relationship inferred for all pairs of genes
- Filtered nodes by known transcription factors
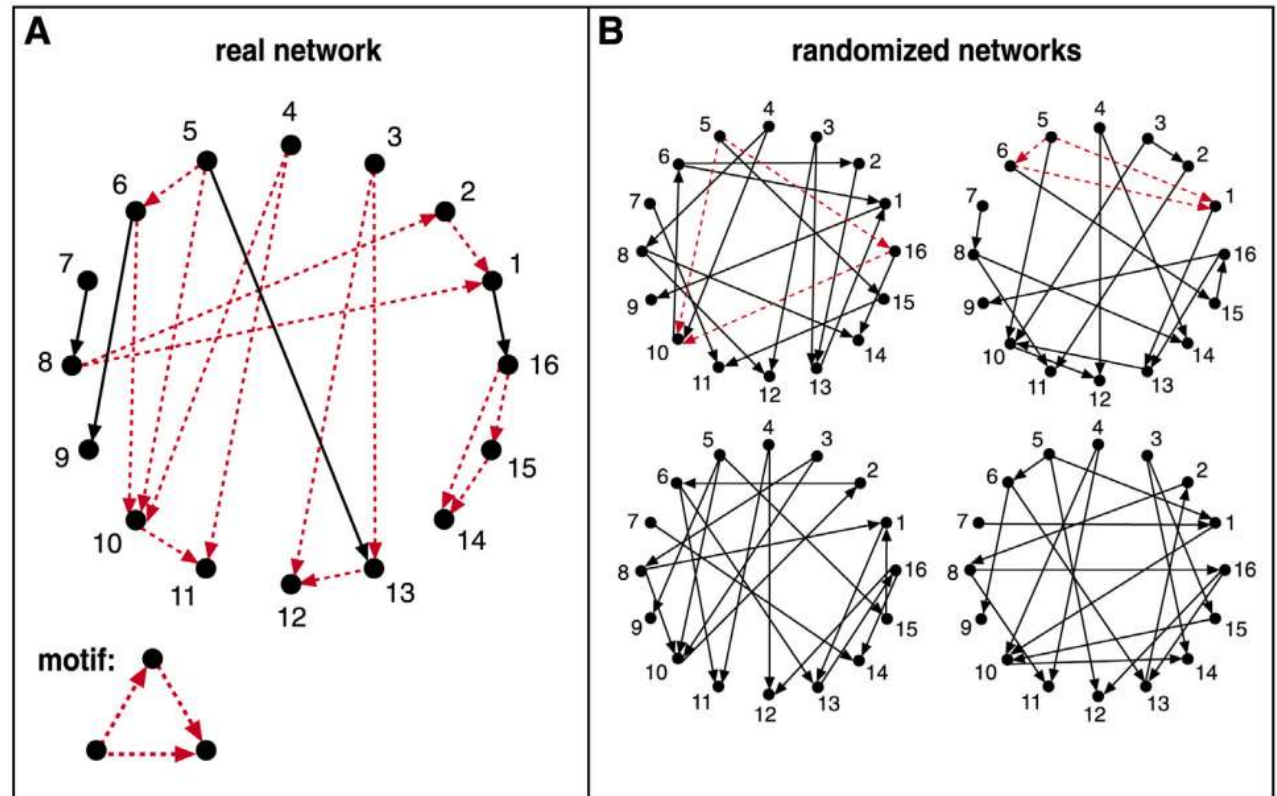- Filtered edges at 99.9th percentile of scores
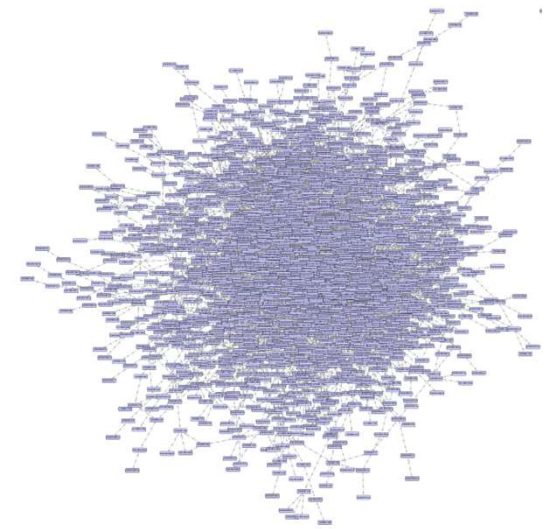
Network	Nodes	Edges	Diameter	Components	Clustering Coefficient	Density
iPSC, Control	2171	3651	10	53	0.013	0.001
iPSC, ALS	2258	4400	12	45	0.024	0.001
iMN, Control	2262	3833	10	42	0.006	0.001
iMN, ALS	1957	3463	11	81	0.012	0.001

Mining networks for enriched motifs

- Motif enrichment was computed using **mFinder [3,4]**
  - Compares occurrence of motifs in input network with randomly generated networks to compute p-value
- Searched for motifs of size **3,4,5**
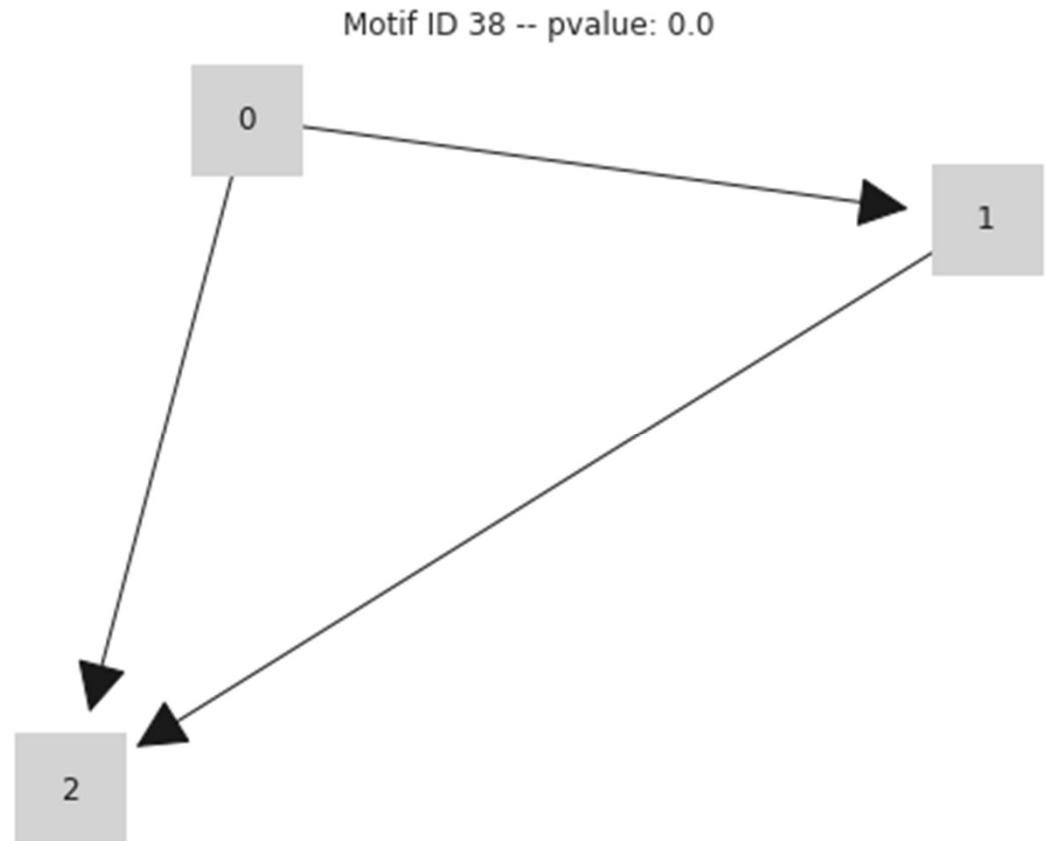
Motif Size	Total Possible Directed Motifs
3	13
4	199
5	9,364
6	1,530,843

http://oeis.org/A003085, [5],
https://www.weizmann.ac.il/mcb/UriAlon/sites/mcb.UriAlon/files/uploads/NetworkMotifsSW/mfinder/mfindermanual.pdf

# Feedforward motifs are enriched in biological networks

- Highly enriched in **both** control network and ALS network
- Sanity check of motif discovery method

FEED FORWARD LOOPS

Motif ID 38 -- pvalue: 0.0
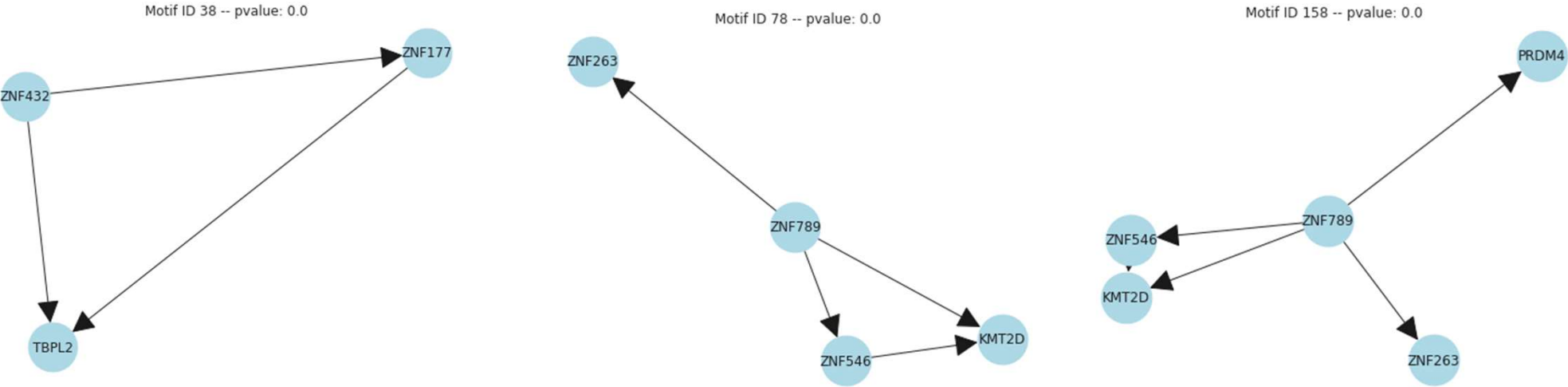
Motifs enriched in ALS

Motif Size	Control	ALS	ALS Only		ALS (Instances)	ALS Only (Instances)
3	1	2	1		89	24
4	16	16	0		7,721	0
5	165	183	29		90,375	331

Motif ID 98 -- pvalue: 0.0

Motif ID 101904 -- pvalue: 0.0

Motif ID 541342 -- pvalue: 0.0

Motifs enriched in ALS

Motif Size	Control	ALS	ALS Only		ALS (Instances)	ALS Only (Instances)
3	1	2	1		89	24
4	16	16	0		7,721	0
5	165	183	29		90,375	331

Motif ID 38 -- pvalue: 0.0
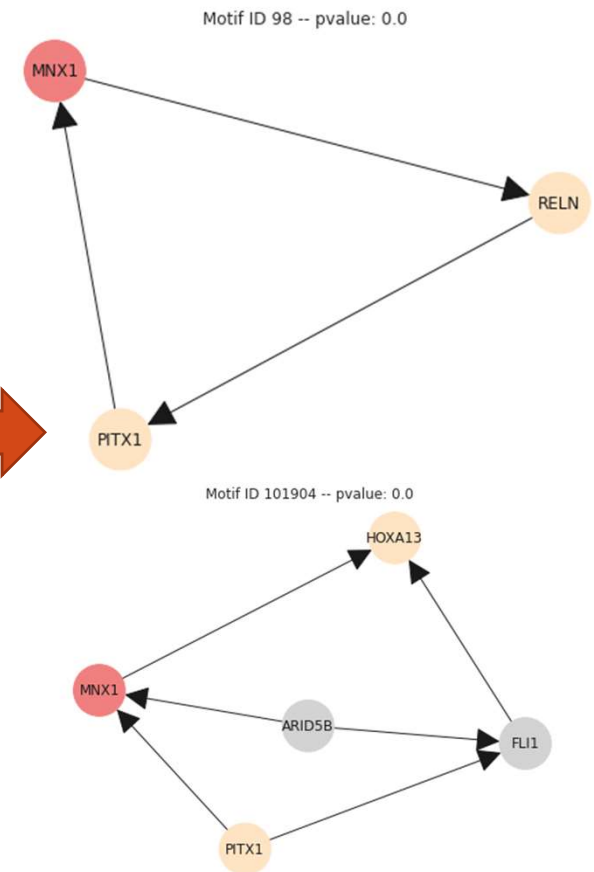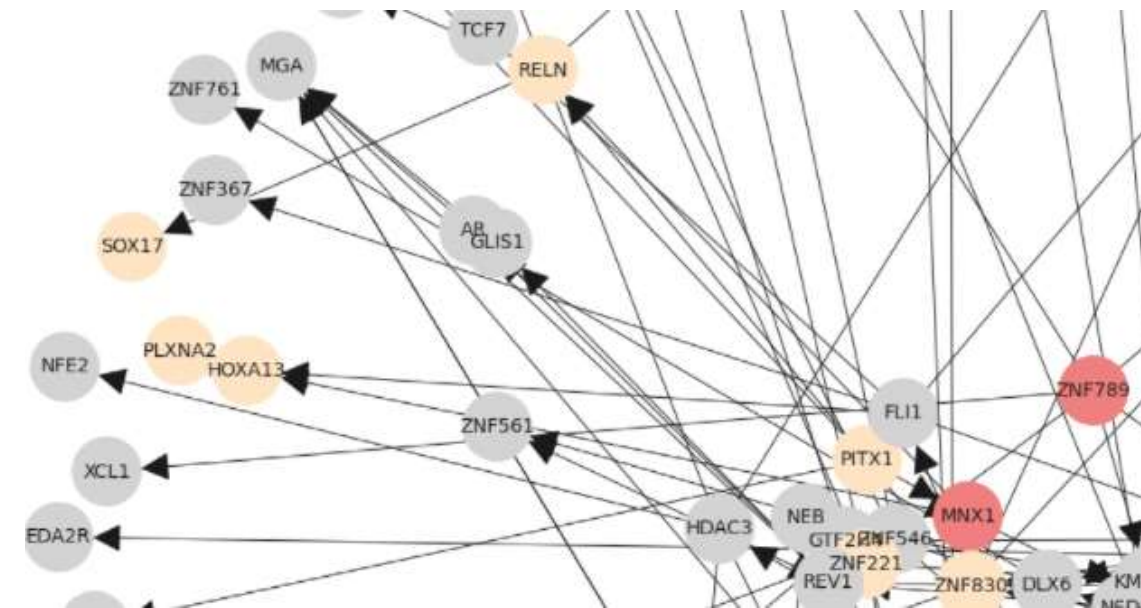
Motif ID 78 -- pvalue: 0.0

Motif ID 158 -- pvalue: 0.0
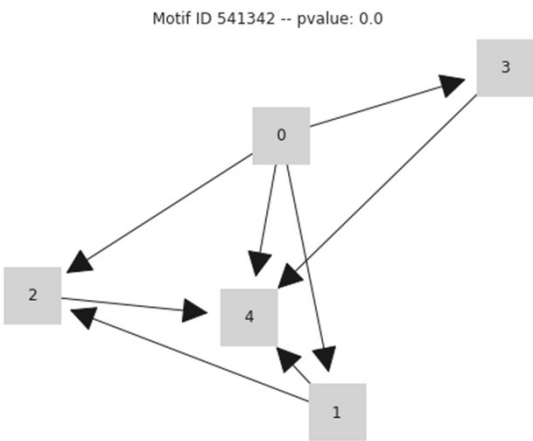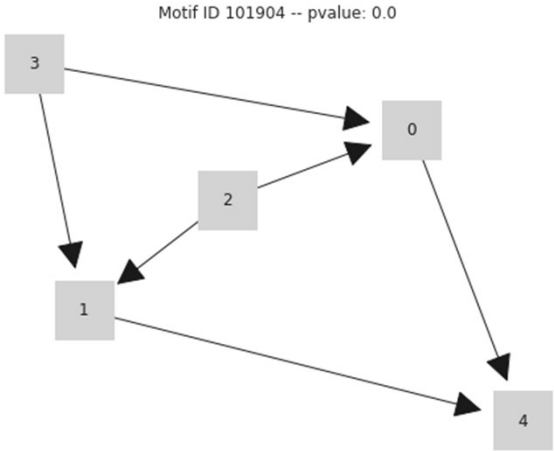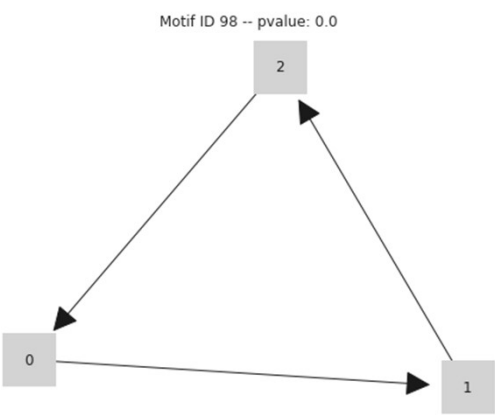
Finding the needle – convergence of DEG and Motif Analysis

- 10 motifs differentially present in ALS and containing a **highly significant** (p <= 1e-5) differentially expressed gene.

- 126 genes in one of the 10 motifs or one step away in the full graph

  - 9 **highly significant** DEG
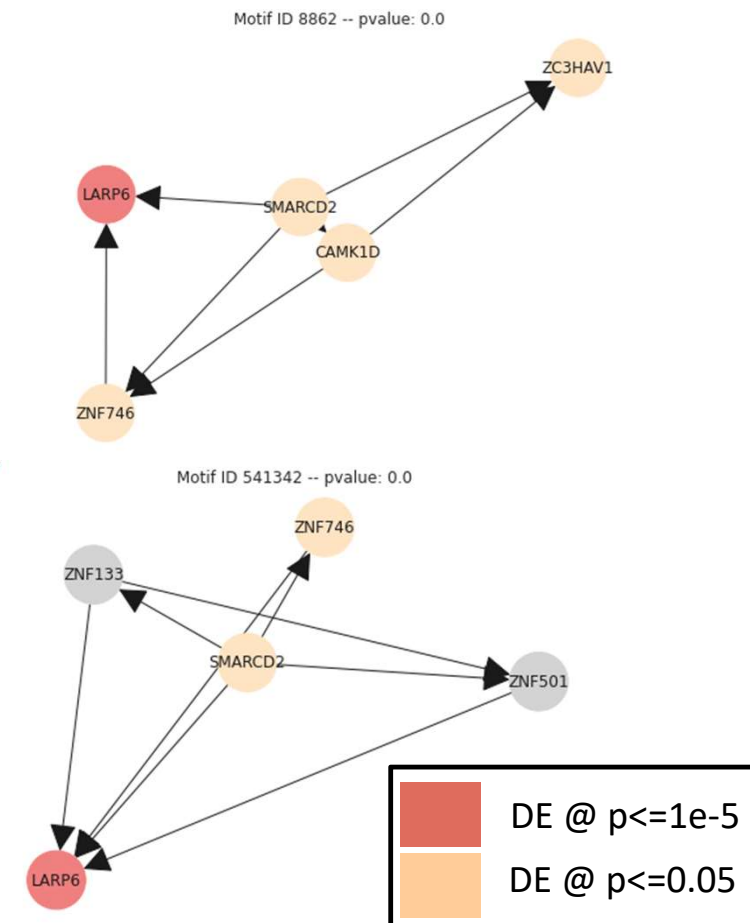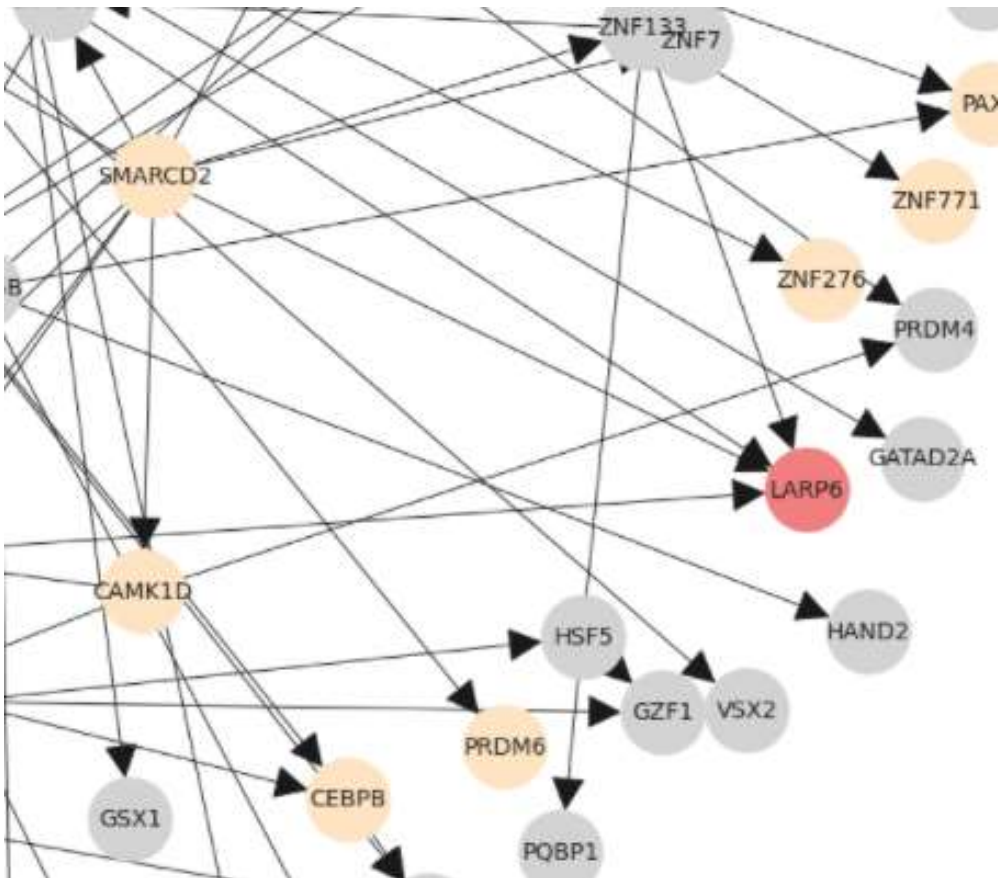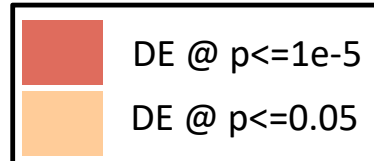
  - 33 **significant** DEG (p <= 0.05)
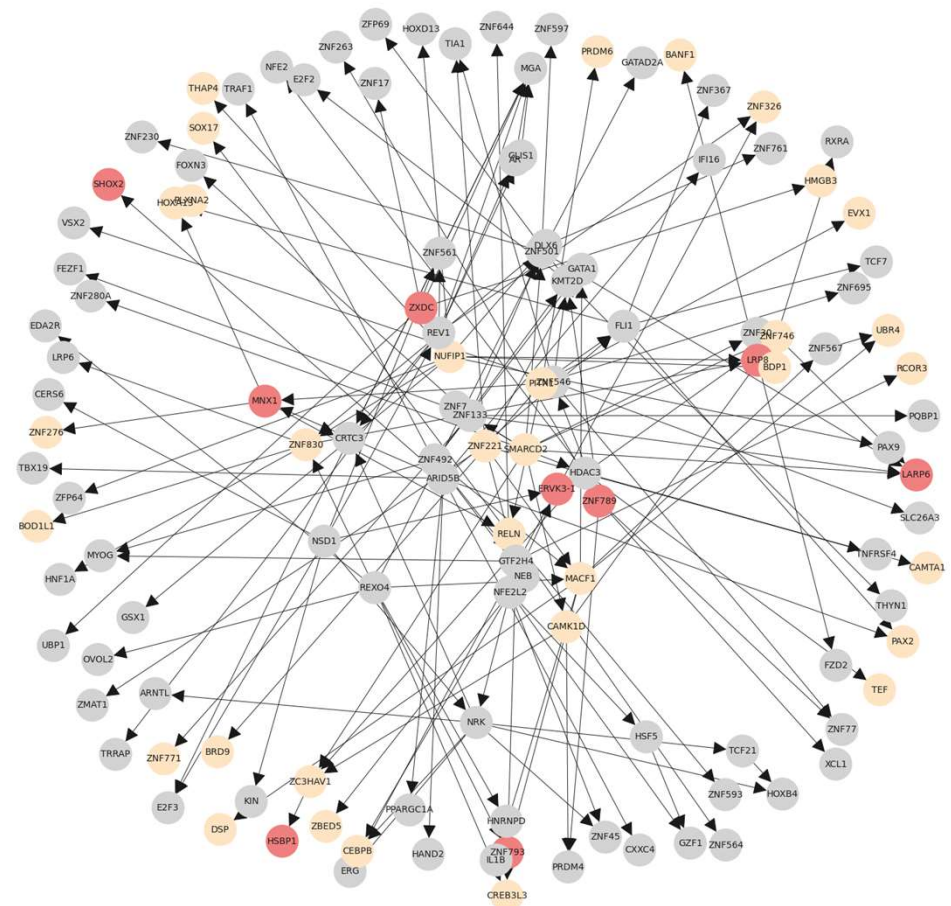
DE @ p<=1e-5
DE @ p<=0.05

Finding the needle – motifs in the focus subnetwork

DE @ p<=1e-5

DE @ p<=0.05

Motif ID 98 -- pvalue: 0.0

Motif ID 101904 -- pvalue: 0.0

Finding the needle – motifs in the focus subnetwork

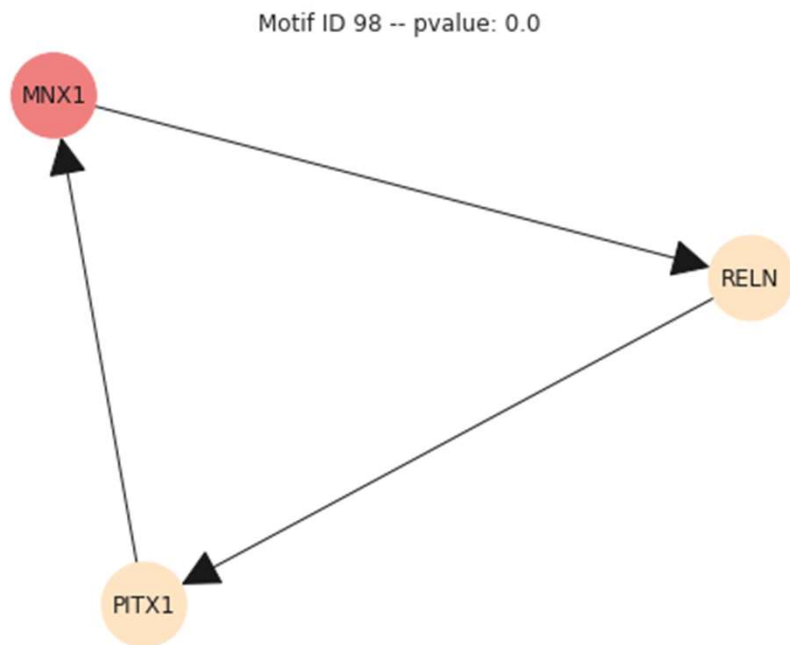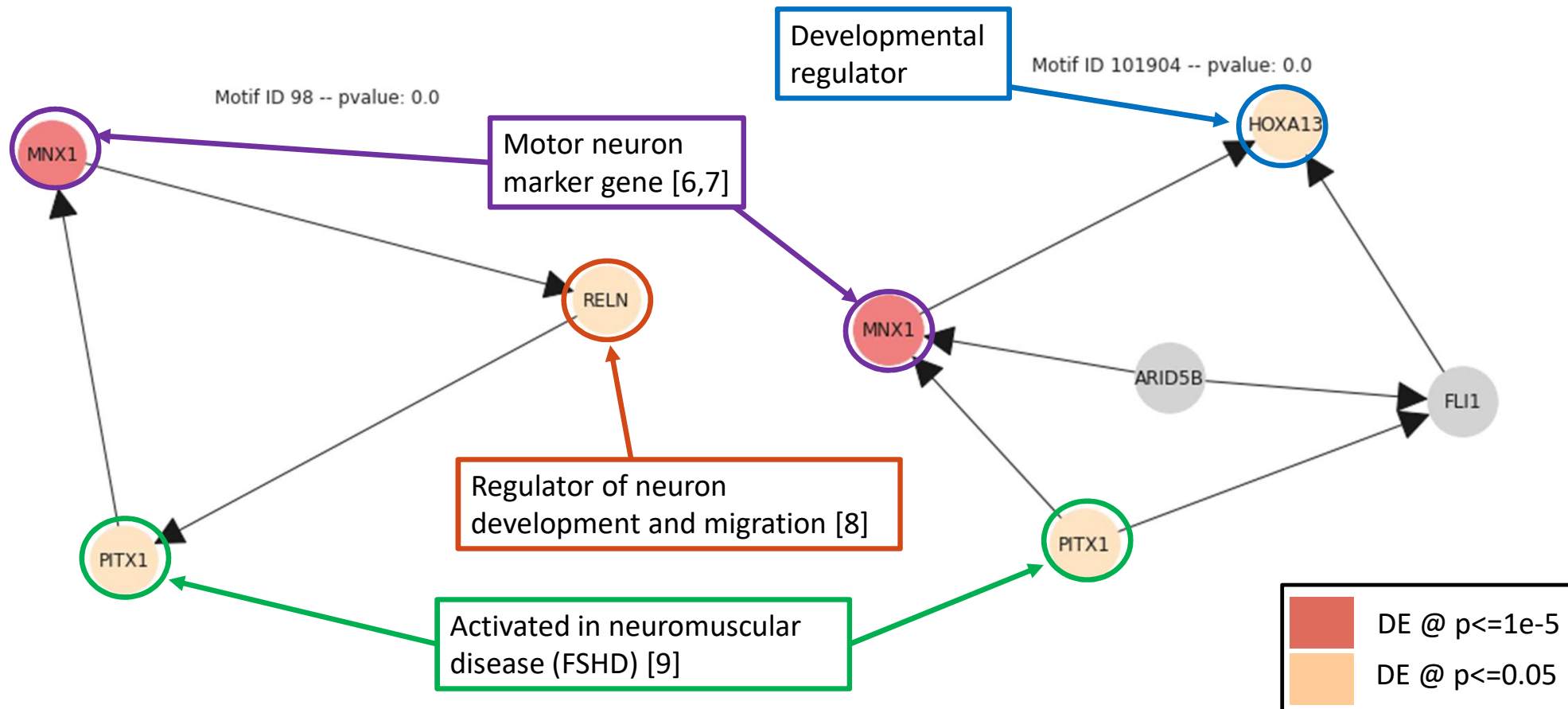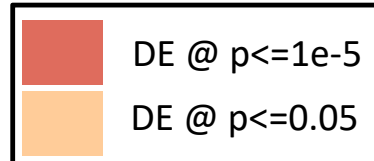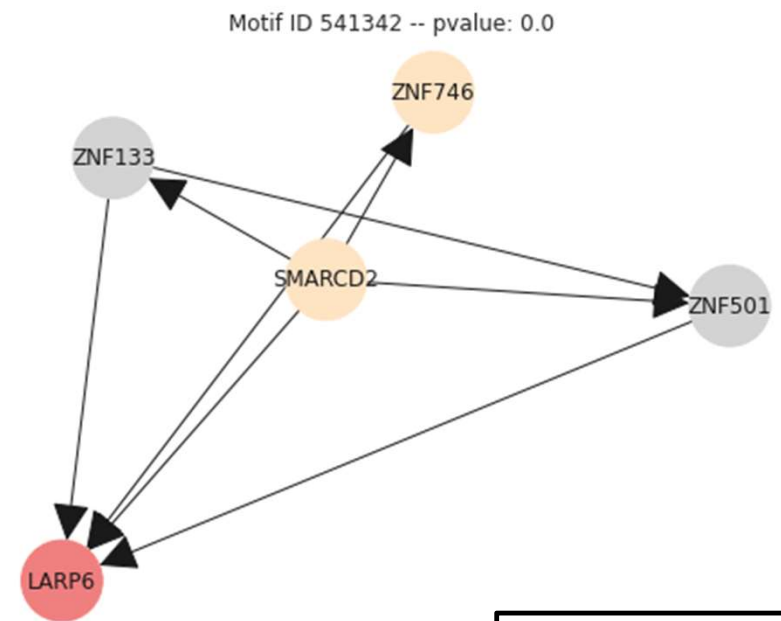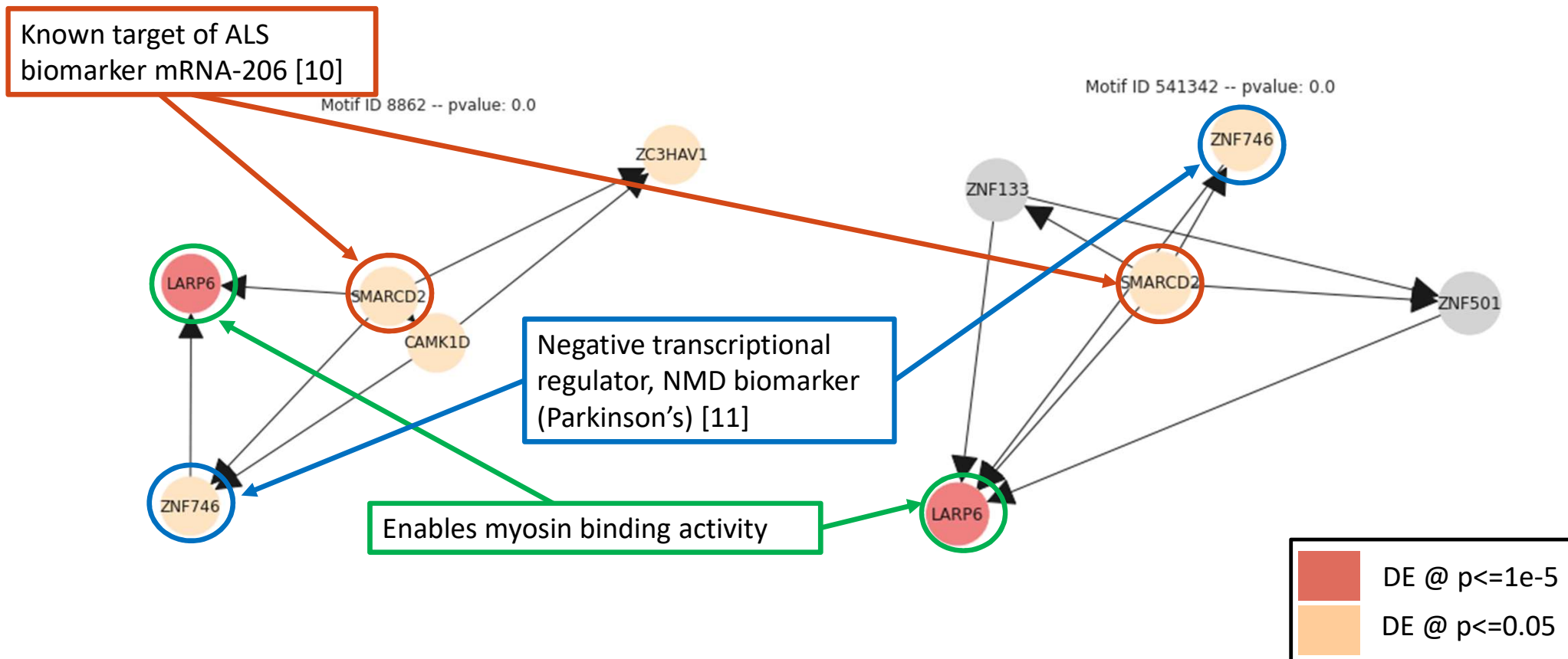# Finding the needle – MNX1 motifs
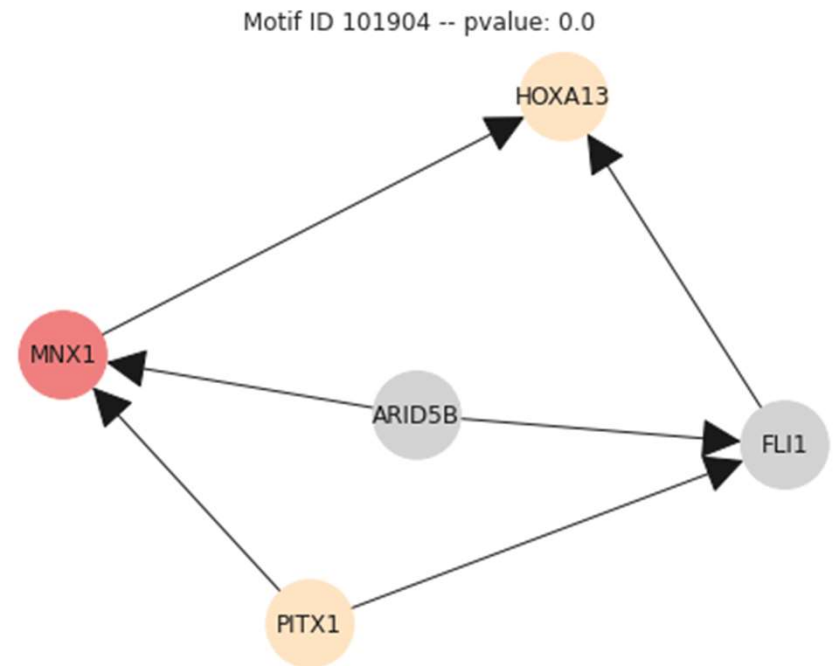
Finding the needle – MNX1 motifs

Finding the needle – LARP6 motifs

Motif ID 8862 -- pvalue: 0.0

Motif ID 541342 -- pvalue: 0.0

DE @ p<=1e-5

DE @ p<=0.05
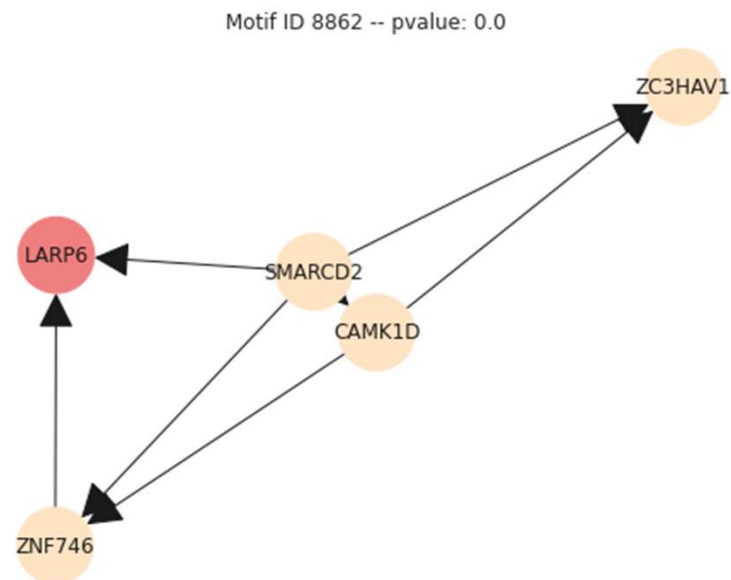
Finding the needle – LARP6 motifs

Summary & Conclusions

- Combine orthogonal filtering methods to identify disease genes in ALS
- Identified pathways associated with myosin binding and motor neuron activity
- Limitations:
  - Statistical power – limited tissue samples from each group means that GSEA and DEG analysis is limited in power and may miss important genes or gene sets
  - Noise – network inference on transcriptomic data can be noisy, which may lead to inconsistent motif inference. Additionally, motif enrichment inference is itself potentially noisy
  - False negatives – hyper-limiting the search is good to find genes to focus on, but may end up ignoring or overlooking important factors
- Future work:
  - Go back and re-evaluate edges as activating or repressing edges
  - More extensive comparison of iPSCs vs. iMNs
  - Multi-omic analysis – integrate proteomic and epigenomic assays

References

[1] Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology, 15(12), 1-21.

[2] Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., & Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. PloS one, 5(9), e12776.

[3] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: simple building blocks of complex networks. Science, 298(5594), 824-827.

[4] Kashtan, N., Itzkovitz, S., Milo, R., & Alon, U. (2004). Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. Bioinformatics, 20(11), 1746-1758.

[5] F. Harary and E. M. Palmer, Graphical Enumeration, Academic Press, NY, 1973, pp. 124 and 241.

[6] Sun, M. A., Ralls, S., Wu, W., Demmerle, J., Jiang, J., Miller, C., ... & Macfarlan, T. S. (2021). Homeobox transcription factor MNX1 is crucial for restraining the expression of pan-neuronal genes in motor neurons. bioRxiv.

[7] Madill, M., McDonagh, K., Ma, J., Vajda, A., McLoughlin, P., O'Brien, T., ... & Shen, S. (2017). Amyotrophic lateral sclerosis patient iPSC-derived astrocytes impair autophagy via non-cell autonomous mechanisms. Molecular brain, 10(1), 1-12.

[8] Ibi, D., Nakasai, G., Koide, N., Sawahata, M., Kohno, T., Takaba, R., ... & Hiramatsu, M. (2020). Reelin supplementation into the hippocampus rescues abnormal behavior in a mouse model of neurodevelopmental disorders. Frontiers in cellular neuroscience, 285.

[9] Dixit, M., Ansseau, E., Tassin, A., Winokur, S., Shi, R., Qian, H., ... & Chen, Y. W. (2007). DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. Proceedings of the National Academy of Sciences, 104(46), 18157-18162.

[10] Toivonen, J. M., Manzano, R., Oliván, S., Zaragoza, P., García-Redondo, A., & Osta, R. (2014). MicroRNA-206: a potential circulating biomarker candidate for amyotrophic lateral sclerosis. PLoS One, 9(2), e89065.

[11] Alieva, A. K., Filatova, E. V., Karabanov, A. V., Illarioshkin, S. N., Slominsky, P. A., & Shadrina, M. I. (2015). Potential biomarkers of the earliest clinical stages of Parkinson's disease. Parkinson's Disease, 2015.

[12] Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... & Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences, 102(43), 15545-15550.
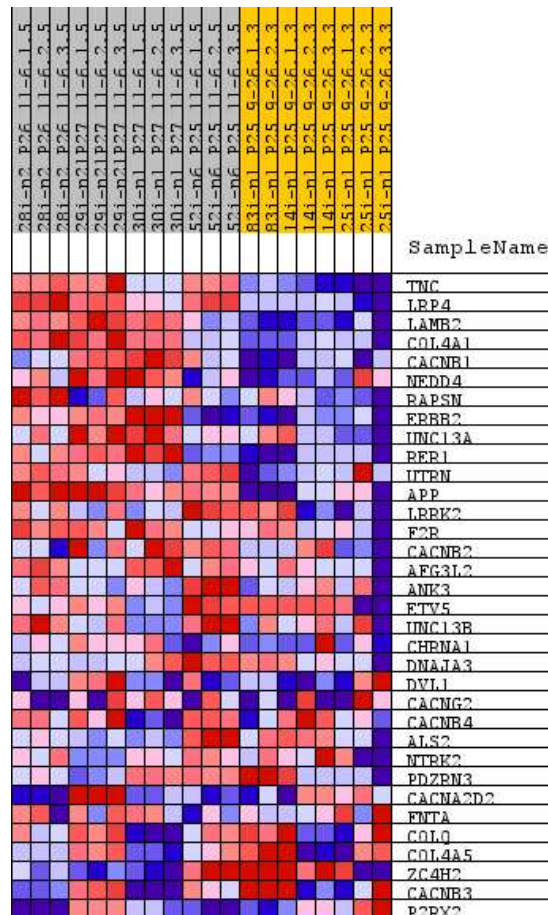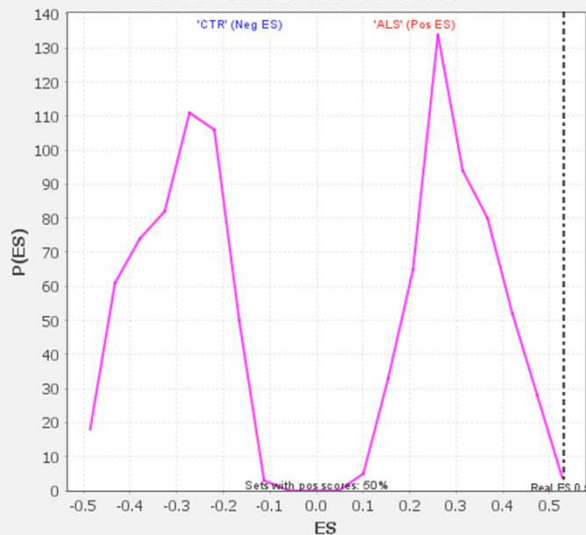
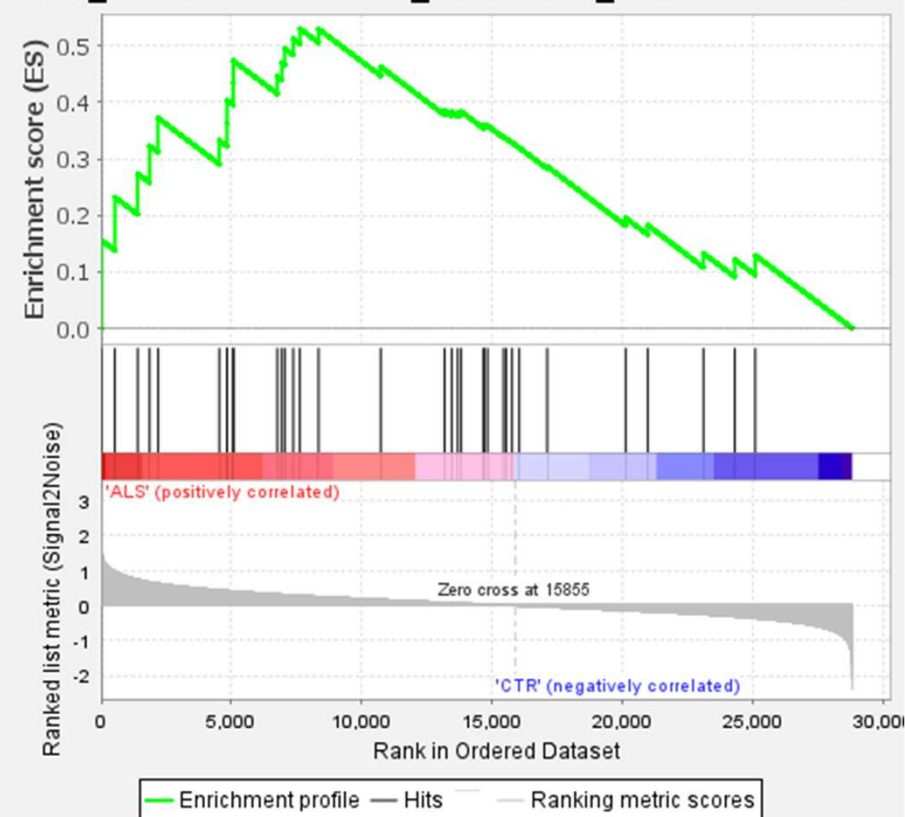# Most genes are not differentially expressed, regardless of counts

Mean Counts vs. (ALS/CTR) Fold Change for 31098 genes in iPSC
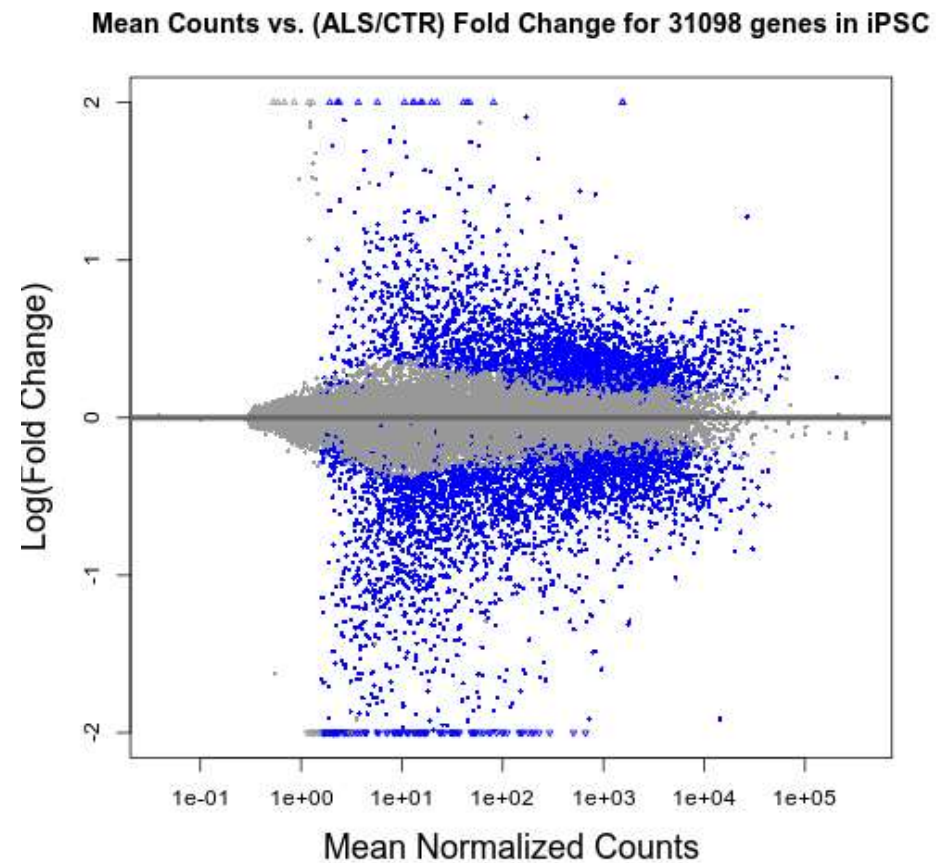
Gene Set Enrichment Analysis in ALS

- Enrichment for Neuromuscular junction development using **GSEA [12]**

Network edge overlap (at top end) is minimal