LabBook_11_11_2016

Claire Green

Thursday

After some drifting with my project, I have got some direction and I think I know what the next steps will involve. Today I set about investigating how VarElect could help me find that direction. VarElect is an online resource provided by the Weizmann Institute of Science which also provides services such as GeneCards, MalaCards and PathCards. The input into VarElect is a list of gene names and a list of phenotype descriptors. For my analysis I used my DEGS (filename "My TDP-43 Genes.txt") and the phenotype descriptors:

- neurodegeneration
- ALS
- "amyotrophic lateral sclerosis"
- "TDP-43 pathology"

Results showed 92 directly related genes and 86 indirectly related genes. Directly related genes are those which have been directly associated to the phrases in question in GeneCards. Indirectly related genes are those which can be linked to the phenotype through another gene. I.e., they may be involved in the same pathway.

The results can be found in the file "VarElect_Results_cmgreen1-sheffield-ac-uk-20161110-093524738.xlsx" in the folder VarElectDEGs.

I decided to put the top 10 genes into GeneMania to discover their relationship. There wasn't much to see (Figure 1) but as I started to add the most correlated genes (using physical interaction and coexpression) the genes seemed to divide off into groups (Figures 2 and 3).

What's interesting is there seems to be a cluster formed with TARDBP and PABPC1, and through the gene AHNAK to DMD. TARDBP as we know encodes TDP-43 which is involved in RNA transcription, modification and transportation. PABPC1 has a very similar role to TDP-43: it shuttles in and out of the nucleus into the cytoplasm, binds mRNA and is involved in its splicing and stability during transportation.

In fact, in Friebaum et al (2010)'s PPI investigation of TDP-43 (which I have used in my previous analysis) they identified PABPC1 as the link between the two distinct protein interaction networks created from TDP-43 PPI partners. Further analysis showed this protein wasn't affected by either the A315T or M337V mutation of TARDBP, though this analysis was performed in HEK cells. PABPC1 is also found in the stress granules with TDP-43, and is used as a marker of stress granules.

As I was reading about PABPC1 on GeneCards, I noticed the section for transcription factors that bind within its promotor inluded TARDBP. This suggests that the expression of TDP-43 should have an effect on PABPC1, possibly meaning its presence in the DEG list is downstream of TDP-43 aggregation. However, if mutation of TARDBP doesn't affect the levels of PABPC1, why would it be in the DEG list if it were downstream? I then looked at the list of transcription factors for both PABPC1 and TARDBP. They share largely the same transcription factors, namely:

[1] "RBBP5" "HDAC1" "PHF8" "KDM5B" "SAP30" "SIN3A" "SIX5" "TAF1" "USF1" "EGR1"
[11] "PML" "POU2F2" "SPI1" "RUNX3" "SP1" "TAF7" "BHLHE40" "ZBTB7A" "E2F6" "MAZ"
[21] "NRF1" "ZNF143" "GTF2F1" "E2F1" "MAFF" "ATF1" "CCNT2" "HMGN3" "HDAC8" "RNF2"
[31] "FOXK2" "ID3" "SMAD5" "ZHX2" "KLF13" "KLF1" "TARDBP" "CREM"

Though none of these genes are DEGs except TARDBP in the 6500 list, though

Figure 1: Top 10 genes from VarElect

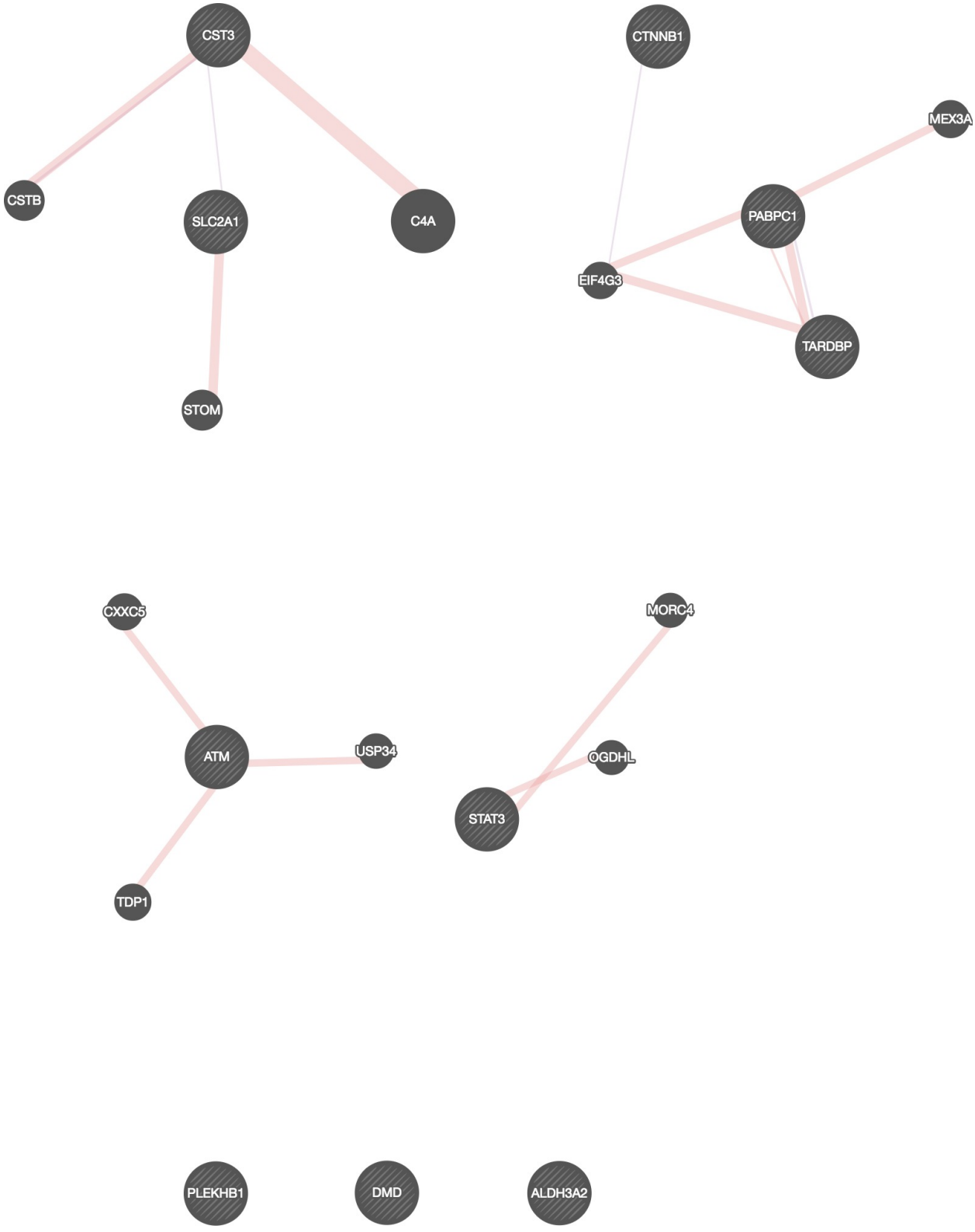

Figure 2: Top 10 genes + 10

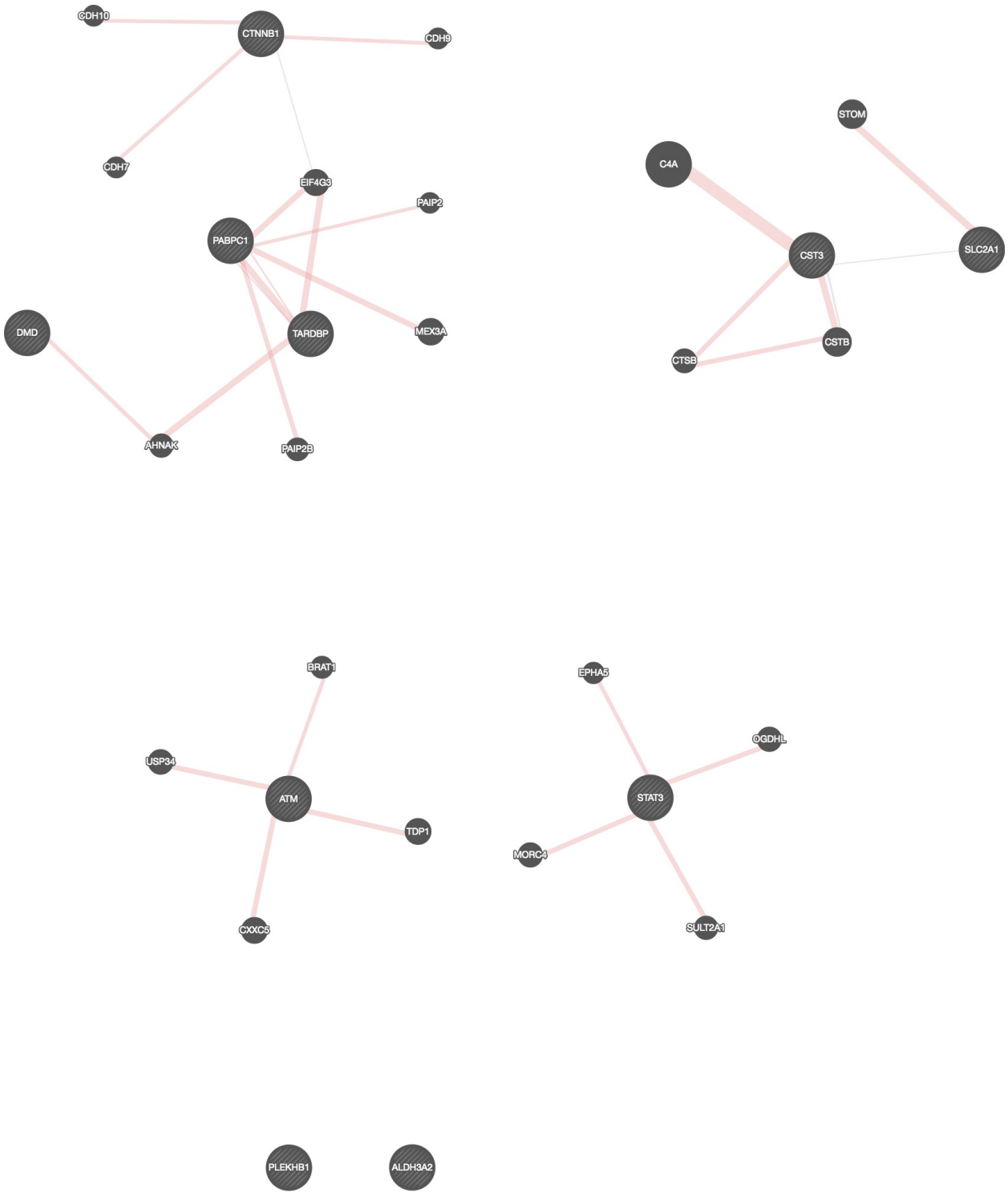

Figure 3: Top 10 genes + 20