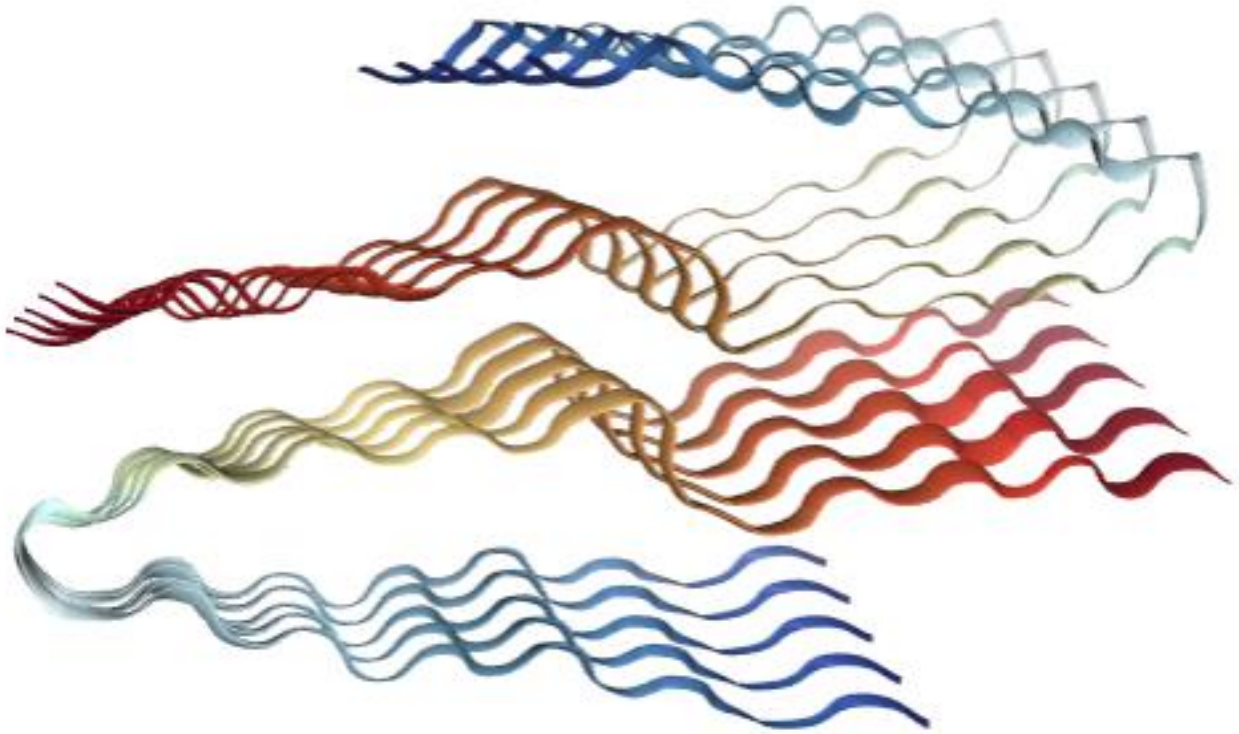
Amyotrophic Lateral Sclerosis (ALS) and the TDP-43 Protein

Aaron Boussina
Bioinformatics
25-Nov-2019



ALS Overview

ALS is a fatal neuromuscular disease that results in neuron death and loss of motor function.

ALS impacts approximately 2 in 100,000 people worldwide annually making it the third most common neurodegenerative disease (Chiò, 2013).

There is no cure for ALS, and existing treatments only serve to temporarily alleviate symptoms. Unfortunately, despite existing intervention, the average survival rate for ALS is only between 2-4 years (Hobson, 2016).

Since ALS is a complex disease resulting from both genetic and environmental factors, there is no existing genetic or biochemical diagnostic test, and diagnosis relies on observation of the physical symptoms (Hardiman, 2017).

Mutations in over 180 genes have been associated with familial and, to a lesser degree, sporadic ALS (Wang, 2019).

Background & Significance – Associated Genes

ALS-Associated Gene	Phenotype Function	Mutation Frequency in Familial ALS
SOD1	Superoxide Dismutase	4%
C9orf72	Endosomal Transport	40% – 50%
FUS	RNA-Binding Protein	4%
CCNF	Protein Kinase Activator	4%
TIA1	RNA-Binding Protein	2.2%
TBK1	Serine Kinase	1%
TARDBP	DNA-binding Transcriptional Repressor	5%

Proteins encoded by ALS-associated genes encompass a wide breadth of biochemical functions. Despite this diversity, however, the formation of abnormal aggregates of protein TDP-43 in neuron and glial cell membranes has been observed in approximately 97% of ALS cases (Nguyen, 2018).

TDP-43 is encoded by the TARDP gene which has a relatively rare association with ALS (~5%).

Initial Literature Review

A no-limits query of “TDP-43 ALS” was entered into NCBI PubMed. There were 1445 results, of which two recent results (within the last two months) were selected for review. This query was also entered into the Online Mendelian Inheritance in Man (OMIM) database and an additional result selected for review.

Biochemical Role of TDP-43 (Mitra, 2019)

TDP-43 is a critical protein in non-homologous end joining repair for double strand DNA breaks within neurons. The inhibition of TDP-43 protein activity, therefore, results in genome instability which could account for the downstream symptoms observed in ALS.

Properties of TDP-43 Aggregates (Neumann, 2006)

The observed TDP-43 aggregates found in the neuron cell membranes in ALS are hyperphosphorylated, ubiquitinated, and cleaved. Specifically, there appear to be cleavage products of 25kD and 45kD.

Prion-Like Propagation of ALS (McAlary, 2019)

The pathological mechanism of ALS appears to be Prion-Like in nature. For TDP-43, there is a low complexity structural region of the protein that is rich in glutamine and asparagine and can be considered prion-like. One theory is that cleavage of this region can result in the creation of a prion-like protein that induces aggregation of misfolded TDP-43 proteins within neuron and glial cells.

FASTA Sequence:

(from Uniprot)

>sp|Q13148|TADBP_HUMAN TAR DNA-binding protein 43 OS=Homo sapiens OX=9606 GN=TARDBP PE=1 SV=1
MSEYIRVTEDENDEPIEIPSEDDGTVLLSTVTAQFPGACGLRYRNPVSQ
CMRGVRLVEGILHAPDAGWGNLVYVVNYPKDNKRKMDETDASSAVK
VKRAVQKTSDLIVLGLPWKTTEQDLKEYFSTFGEVLMVQVKKDLKTG
HSKGFGFVRFTEYETQVKVMSQRHMIDGRWCDCKLPNSKQSQDEPL
RSRKVFVGRCTEDMTEDELREFFSQYGDVMDVFIPKPFRAFAFVTFAD
DQIAQSLCGEDLIIKGISVHISNAEPKHNSNRQLERSGRFGGNPGGFG
NQGGFGNSRGGGAGLGNNQGSNMGGGMNFGAFSINPAMMAAAQA
ALQSSWGMMGMLASQQNQSGPSGNNQNQGNMQREPNQAFGSGNN
SYSGSNSGAAIGWGSASNAGSGSGFNGGFGSSMDSKSSGWGM

BLAST:

To identify similar sequences, NCBI's protein BLAST search was utilized with the BLOSUM62 scoring matrix. The results demonstrate the presence of similar TAR DNA-binding proteins in a multitude of species, the most similar being the orangutan (99.76% match), the treeshrew (97.83% match), and the cow (97.58% match).
The house mouse, the most prevalent model organism for ALS research, has a match of 96.14%.

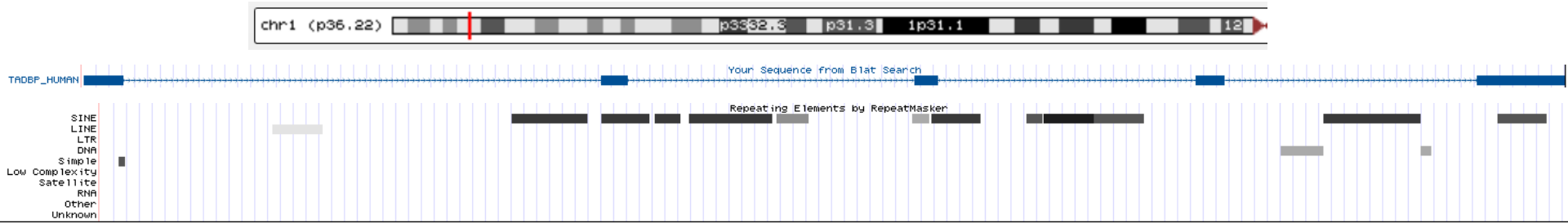
Sequence Searches – UCSC Genome Browser

The FASTA peptide sequence for TDP-43 was entered into the UC Santa Cruz Genome Browser BLAT search. The browser identified the location of gene TARDBP at Chromosome 1 (p36, 22).

The gene has 4 intron regions and 5 exon regions as well as 13 short interspersed nuclear element (SINE) repeat regions scattered throughout.

A view of the gene expression plots from GTEx shows the Cerebrellar Hemisphere as the tissue with the highest expression and the Cerebellum as the tissue with the second highest expression level. This is an expected result due to the protein's involvement in neuron cell activity.

The expanded SNP view reveals that the vast majority of SNPs are at the intron locations. A view of the flagged SNPs, however, shows that nearly every clinically significant SNP occurs at the last exon on the 3' end. This region also corresponds to the C-terminal prion-like section of the TDP-43 protein (McAlary, 2019).



SNPs & Primer Design

```
PRODUCT SIZE: 225, PAIR ANY_TH COMPL: 9.54, PAIR 3'_TH COMPL: 5.92

    1 TTTGGGAAACAATCAAGGTAGTAATATGGGTGGTGGGATGAACTTTGGTGCGTTCAGCAT
                                        >>>>>>>>>>>>>>>>>>>>

   61 TAATCCAGCCATGATGGCTGCCGCCCAGGCAGCACTACAGAGCAGTTGGGGTATGATGGG

  121 CATGTTAGCCAGCCAGCAGAACCAGTCAGGCCCATCGGGTAATAACCAAAACCAAGGCAA

  181 CATGCAGAGGGAGCCAAACCAGGCCTTCGGTTCTGGAAATAACTCTTATAGTGGCTCTAA
                                                                <<<<

  241 TTCTGGTGCAACAATTGGTTGGGGATCAGCATCCAATGCAGGGTCGGGCAGTGGTTTTAA
      <<<<<<<<<<<<<<<<<<

  301 TGGAGGCTTTGGCTCAAGCATGGATTCTAAGTCTTCTGGCTGGGGAATGTAGACAGTGGG

  361 GTTGTGGTTGGTTGGTATAGAATGGTGGGAATTCAAATTTTTCTAAACTCATGGTAAGTA

  421 TATTGTAAAATACATATGTACTAAGAATTTTCAAAATTGGTTTGTTCAGTGTGGAGTATA

  481 TTCAGCAGTATTTTTGACATT
```

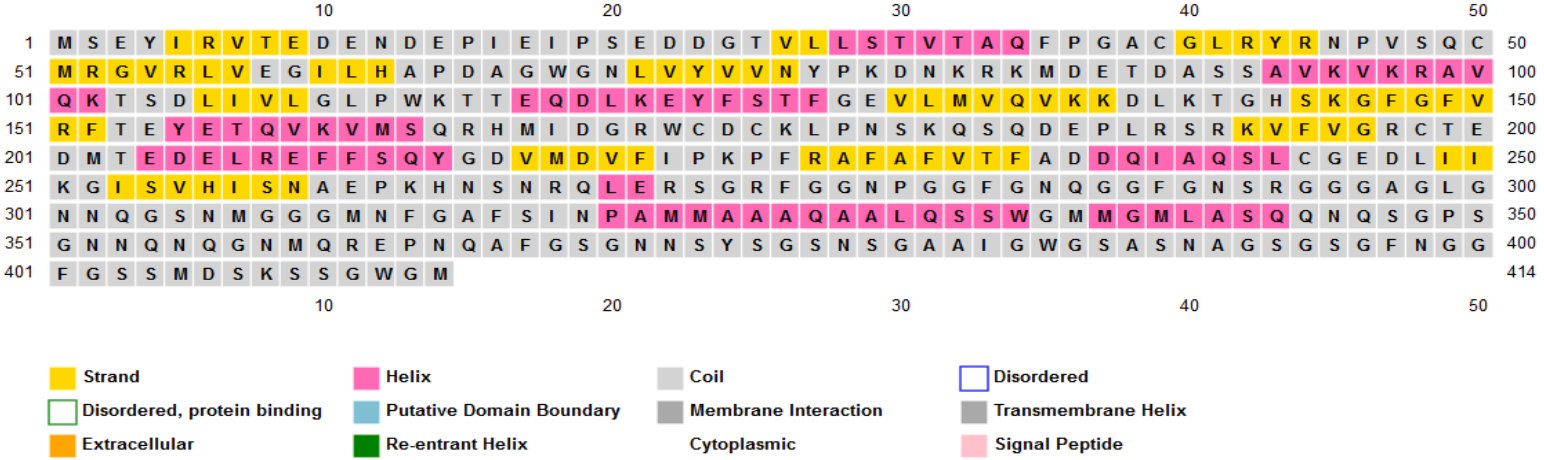
One specific SNP identified from the UCSC Genome Browser at the final 3' end exon, rs367543041, was entered into dbSNP for further analysis.

dbSNP identified this SNP as a missense variant with a global frequency of 0.00003. dbSNP also linked to ClinVar Accession RCV000020663.10 describing this SNP as pathogenic for ALS Type 10.

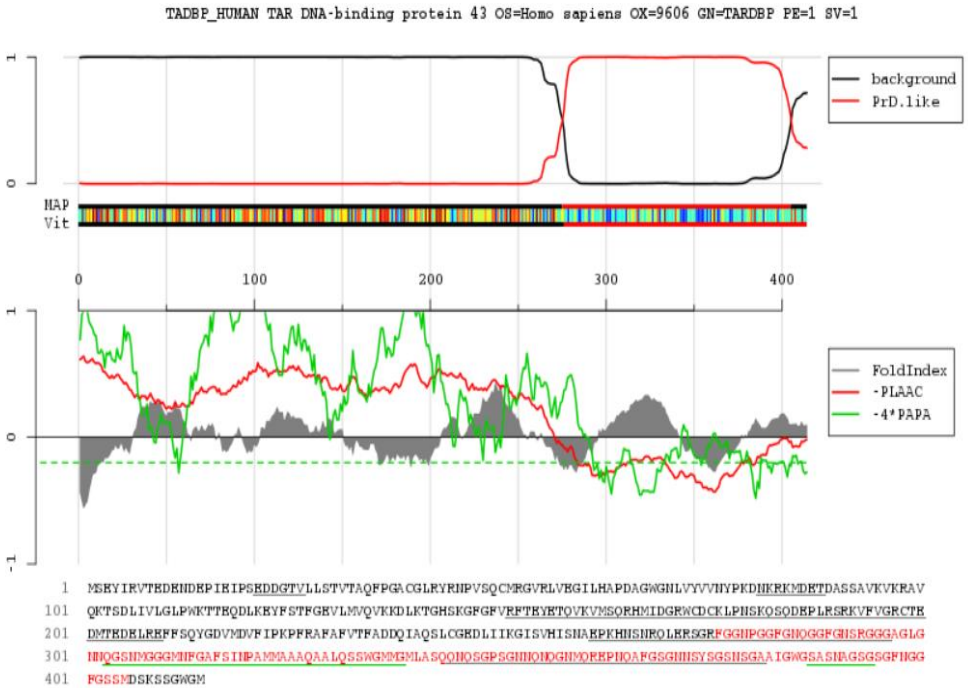
The linked OMIM entry (605078.0013) was then used to determine that this SNP causes an Alanine to Threonine substitution. The linked study (Corrado, 2009) describes that this SNP has a relatively high frequency in patients with ALS and results in instability of the mutant TDP-43.

Primer3Web from University of Massachusetts was used to identify PCR primers that could be utilized for a genetic assay of rs367543041

Protein Structure & Properties – PSIPRED 4.0



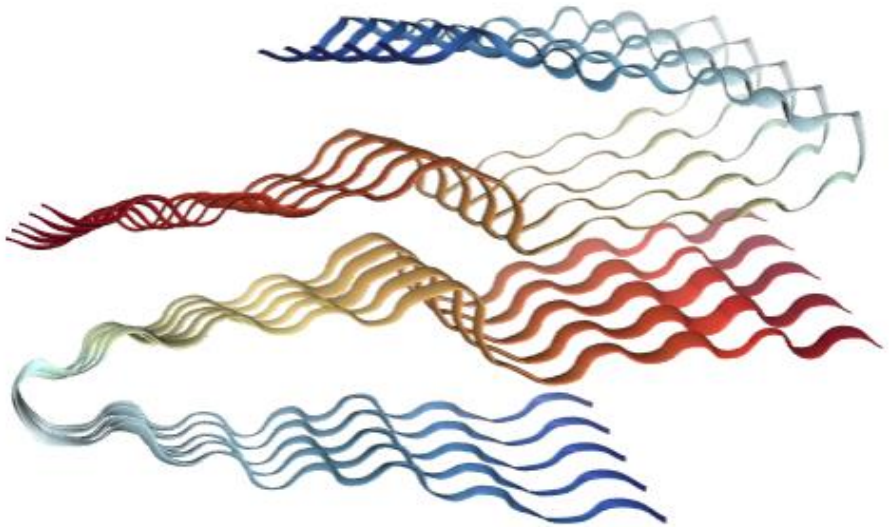
The major secondary structural features of TDP-43 are strands and helixes. The C-terminal end of the protein from position 344 to position 414 is a region of low structural complexity consisting of only coils. This region also contains a high ratio of asparagine (N) and glutamine (Q) (22.5%). These two properties have been found to be associated with prion forming domains (Sabate, 2015).



The TPD-43 FASTA sequence was entered into the Prion Like Amino Acid Composition tool (PLAAC) developed by MIT which uses a hidden markov model to predict prions based on experimentally determined prion domains from yeast (Alberti, 2009).

The C-terminal region of low structural complexity and high N+Q content was determined to be prion-like in nature. This supports the conclusions from the initial literature review that cleaved products of the low complexity region in TDP-43 could cause misfolding in normal TDP-43 proteins resulting in protein aggregation and symptoms observed in ALS.

A query of the RCBS Protein Data Bank (PDB) was performed to retrieve a 3D visualization of this low-complexity prion-like region

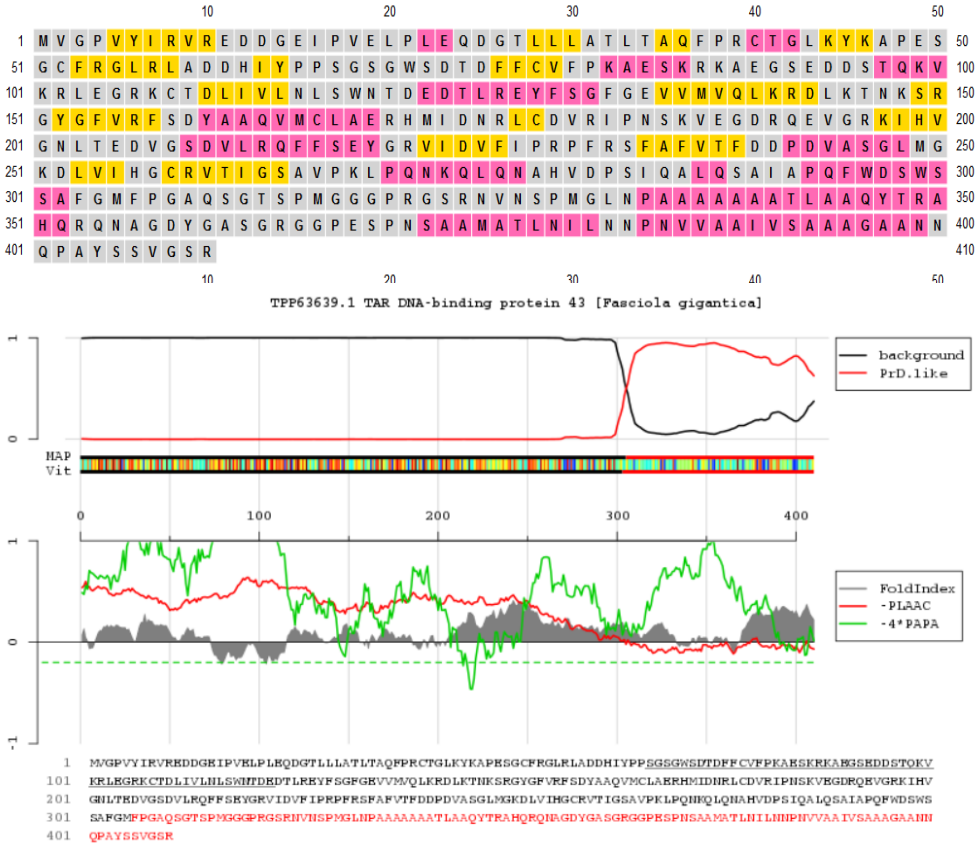


Multiple Sequence Alignment – NCBI COBALT

To identify potential conserved regions of the TDP-43 protein, a multiple sequence alignment was performed using NCBI COBALT.

The center region of the TDP-43 protein is highly conserved while the low-complexity region at the C-terminal end is not well conserved. This might be because the structural and prion-like characteristics of this region are more important than the primary structure. This is consistent with the observation that prion sequences are better predicted by Q+N content and low structural complexity than by primary structure (Sabate, 2015).

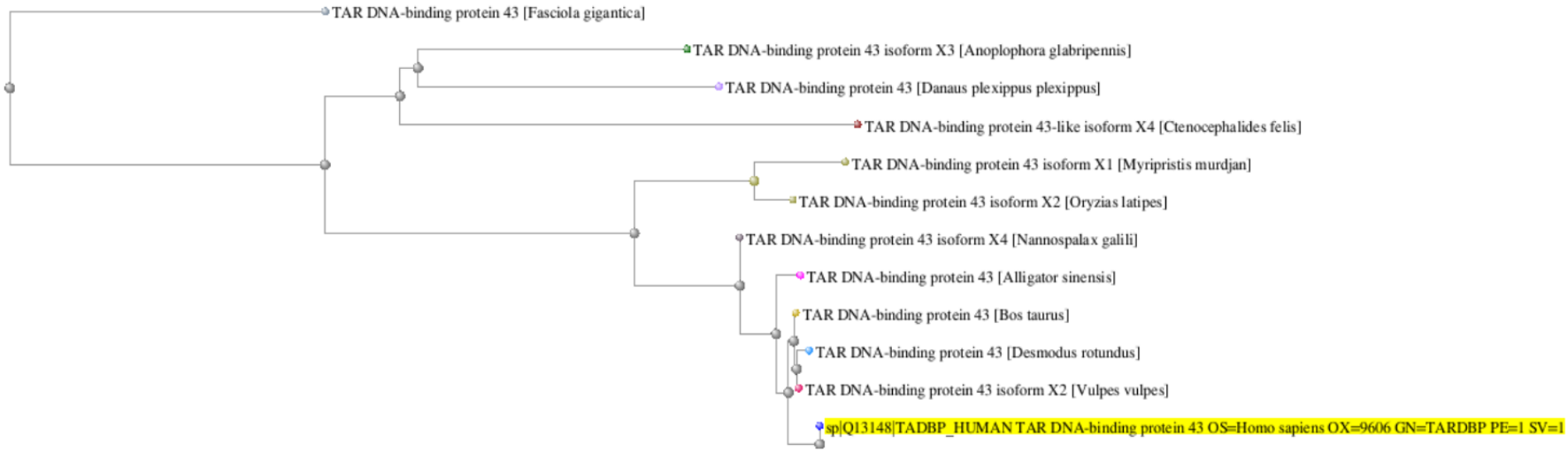
Multiple Sequence Alignment – Structure of Most Dissimilar Sequence



To test whether TDP-43 secondary structural elements are conserved, PSIPRED & PLAAC structural analyses were performed on the most dissimilar sequence from the Multiple Sequence Alignment – the TDP-43 of Fasciola gigantica (flatworm).

While the flatworm TDP-43 secondary structure at the C-terminal end differs from human TDP-43, it still contains stretches of low complexity that are prion-like in nature. This supports the conclusion that the prionogenic C-terminal end of the TDP-43 protein is an important structurally conserved element despite low conservation of the primary sequence.

Multiple Sequence Alignment – Phylogenetic Tree



To further analyze sequence relationships across the selected organisms, a phylogenetic tree was generated from the COBALT multiple sequence alignment utilizing the Fast Minimum Evolution method with Grishin evolutionary distance

Conclusion

A variety of bioinformatics tools and methods were used to conduct an exploratory analysis of the TDP-43 protein and its role in the neurodegenerative disorder ALS.

An initial literature review identified a hypothesis that prion-like regions in TDP-43 could become cleaved and cause misfolding of wild-type TDP-43 resulting in loss of protein function and genomic instability of neuron cells.

The theory that a loss of TDP-43 function could result in ALS symptoms is supported by the SNP meta-analysis of rs367543041 whereby mutant TDP-43 expressed in ALS patients was found to be relatively unstable.

The theory that the TDP-43 protein contains a prion-like region was confirmed by analysis of the protein's secondary structure. Further, through multiple sequence alignment it was identified that this low complexity prion-like region is conserved at the secondary structure level, despite being unconserved in primary structure. This suggests that the prion-like nature of this region is important for TDP-43 function.

Taken together, these analyses support the theory that cleavage of TDP-43 is a crucial causative factor in the formation of neural membrane protein aggregates and the development of ALS.

References

Chiò, A., Logroscino, G., Traynor, B. J., Collins, J., Simeone, J. C., Goldstein, L. A., & White, L. A. (2013). Global epidemiology of amyotrophic lateral sclerosis: a systematic review of the published literature. Neuroepidemiology, 41(2), 118-130.

Hobson, E. V., & McDermott, C. J. (2016). Supportive and symptomatic management of amyotrophic lateral sclerosis. Nature Reviews Neurology, 12(9), 526.
Byrne, S., Walsh, C., Lynch, C., Bede, P., Elamin, M., Kenna, K., ... & Hardiman, O. (2011). Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis. Journal of Neurology, Neurosurgery & Psychiatry, 82(6), 623-627.

Hardiman, O., Al-Chalabi, A., Chio, A., Corr, E. M., Logroscino, G., Robberecht, W., ... & Van Den Berg, L. H. (2017). Amyotrophic lateral sclerosis. Nature Reviews Disease Primers, 3, 17071.

Wang, J. (2019). Novel Aspects on Motor Neuron Disease: The Recent Genetic Studies on ALS. In Novel Aspects on Motor Neuron Disease. IntechOpen.

Nguyen, H. P., Van Broeckhoven, C., & van der Zee, J. (2018). ALS genes in the genomic era and their implications for FTD. Trends in Genetics, 34(6), 404-423.

Mitra, J., & Hegde, M. L. (2019). A Commentary on TDP-43 and DNA Damage Response in Amyotrophic Lateral Sclerosis. Journal of experimental neuroscience, 13, 1179069519880166.

Su, Y., Ming, G. L., & Song, H. (2015). DNA damage and repair regulate neuronal gene expression. Cell research, 25(9), 993.

McAlary, L., Plotkin, S. S., Yerbury, J. J., & Cashman, N. (2019). Prion-Like Propagation of Protein Misfolding and Aggregation in Amyotrophic Lateral Sclerosis. Frontiers in Molecular Neuroscience, 12, 262.

Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., ... & McCluskey, L. F. (2006). Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science, 314(5796), 130-133.

Corrado, L., Ratti, A., Gellera, C., Buratti, E., Castellotti, B., Carlomagno, Y., ... & Baralle, F. E. (2009). High frequency of TARDBP gene mutations in Italian patients with amyotrophic lateral sclerosis. Human mutation, 30(4), 688-694.

Sabate, R., Rousseau, F., Schymkowitz, J., & Ventura, S. (2015). What makes a protein sequence a prion?. PLoS computational biology, 11(1), e1004013.

Alberti, S., Halfmann, R., King, O., Kapila, A., & Lindquist, S. (2009). A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. Cell, 137(1), 146-158.