Do TDP-43 cryptic splicing events happen near neurodegenerative disease risk genes?

Hannah Whiteley[1], Anna-Leigh Brown[2] , Pietro Fratta[2]
[1]Birkbeck University, [2]Department of Neuromuscular Diseases, UCL Institute of Neurology, London

Introduction

RNA binding proteins (RBPs) tightly regulate splicing in cells. Loss of function of **TDP-43**, an RBP, results in **cryptic splicing** leading to inclusion of cryptic exons. TDP-43 proteinopathy is a well-known trait of **neurodegenerative diseases** such as Amyotrophic lateral sclerosis (ALS) and Frontotemporal dementia (FTD). This project is trying to determine if there are more **SNPs** within **neurodegenerative disease risk genes** and are **they affecting inclusion of cryptic exons?**

Cryptic splicing

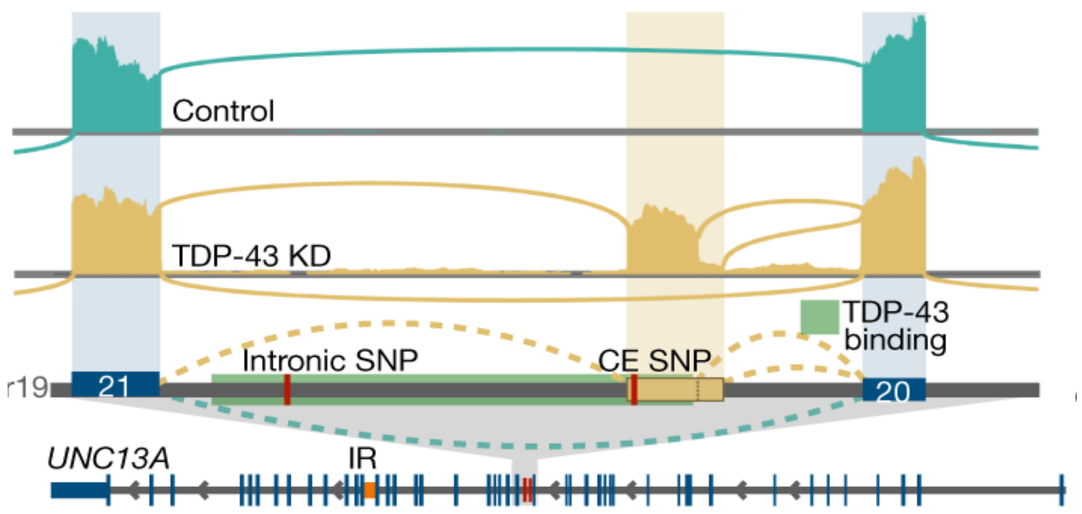
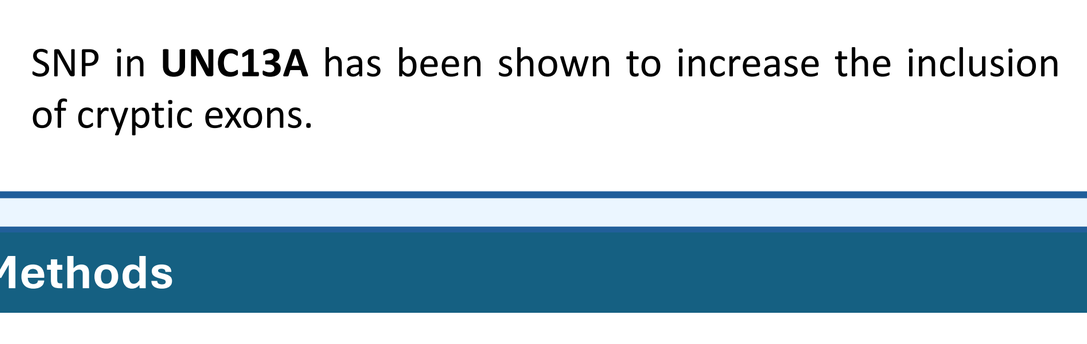


Figure 1. Adapted from (Brown et al., 2022). A SNP was found to increase the inclusion of a cryptic exon in UNC13A.

SNP in **UNC13A** has been shown to increase the inclusion of cryptic exons.

Methods

Data from the **NHGRI-EBI GWAS Catalog,** (Sollis E, et al.) was filtered to include **only neurodegenerative diseases**. SNP information was **overlapped with cryptic regions**. Sequences around the SNPs were made into **heptamers** including both the **healthy** and **risk alleles**. **Z-scores** for various **RBPs** were compared for each heptamer, (Ray et al., 2013).



Results - *Initial data analysis of publicly available GWAS data*

169 Studies with **1336 reported genes** across **82** neurodegenerative diseases and **3,189 SNPs.** Initial analysis found **rs429358** to be the **most frequently reported SNP**. This SNP is located on the **ApoE gene** which is **highly associated** with a **risk of Alzheimer's disease.** SNPs appear across **657 genomic regions**.

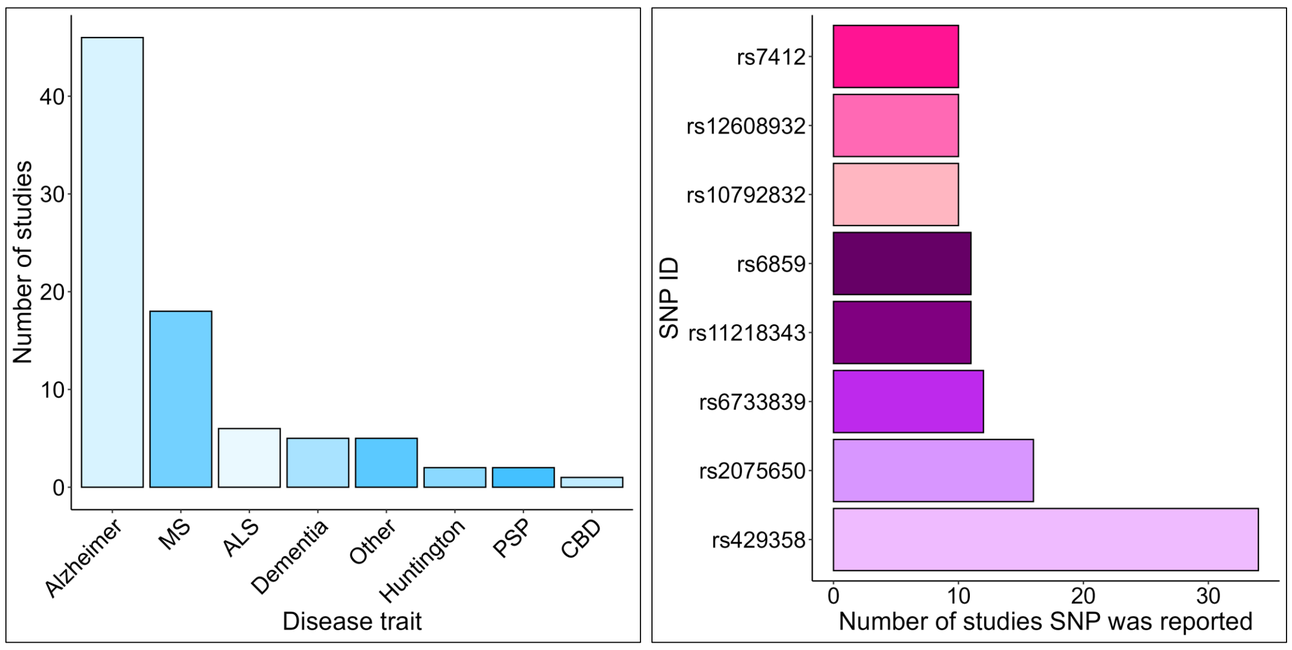


Figure 3. a) Showing neurodegenerative disease traits from GWAS. Multiple Sclerosis (MS), Amyotrophic lateral sclerosis (ALS), Progressive supranuclear palsy (PSP), Corticobasal degeneration (CBD). b) Ten highest reported SNPs across studies.

Results - *Preliminary analysis of GWAS and splice junction intersections*

We found **7 unique SNPs** within the **original range** which were **in 2 cryptic splicing junctions**, **10 SNPs in an extended range of 10 kb** and **21 SNPs in an extended 20kb range**. **rs55970842** appeared in **5 cryptic splice junctions**, **rs113020870** in **4 junctions** and **rs906175 in 3 junctions**.

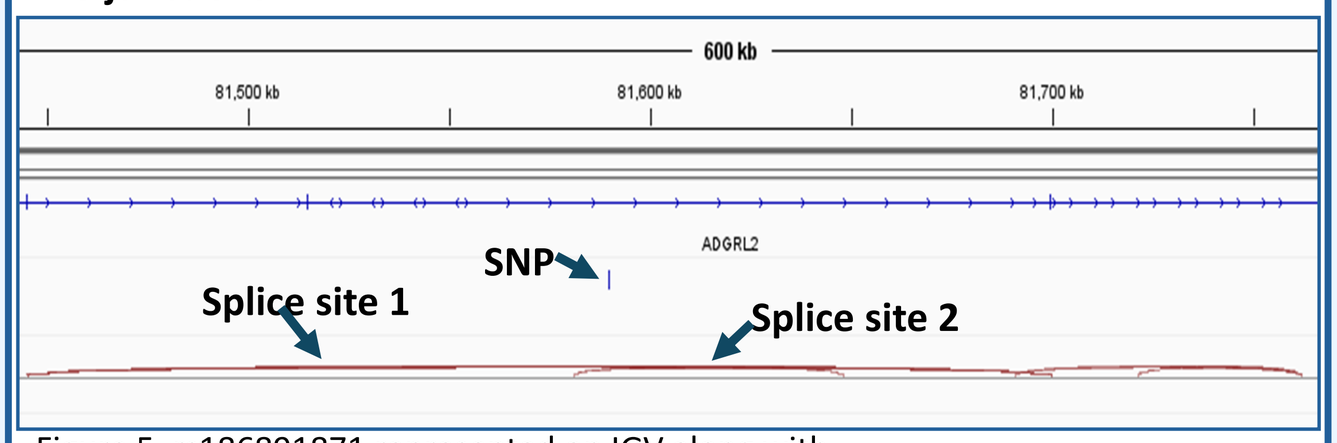


Figure 5. rs186891871 represented on IGV along with

Multiple SNP's were present in **multiple cryptic splicing sites**. Figure 5 shows **rs186891871** appearing within **two cryptic junctions** in the **ADGRL2** gene, a gene associated with **Alzheimer's**.

SNP ID	Disease trait	Mapped Gene
rs186891871	Late- onset Alzheimer's disease	ADGRL2
rs8112449	Multiple sclerosis	CDC37
rs393152	Corticobasal degeneration	LINC02210, LINC02210-CRHR1

Table 1. Three SNPs which are present in multiple cryptic splicing sites, the neurodegenerative disease trait and mapped gene associated with the SNP.

Citations

Brown, A.-L. et al. (2022) TDP-43 loss and ALS-risk SNPs drive mis-splicing and depletion of UNC13A. Nature 603, 131–137

Ray, D., et al., 2013. A compendium of RNA-binding motifs for decoding gene regulation. Nature 499, 172–177. https://doi.org/10.1038/nature12311

Sollis E, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource.

Future work

Filter SNPs that **appear** in **3 or more TDP-43 knockdown experiments** and for each SNP analyse the **Z-scores** for various RBPs associated with each **heptamer** surrounding the SNP.