Are ATXN2 variants modifying our understanding about neural pathogenesis, phenotypes, and diagnostic?

Jose Miguel Laffita-Mesa*, Martin Paucar, Per Svenningsson

ATXN2 one gene with multiple phenotype effects: ATXN2 gene encodes a cytosolic protein (ataxin-2) with pleiotropic functions (see below). This gene contains a number of exonic Cytosine-Adenine-Guanine (CAG)-repeats which encodes a polyglutamine tract (polyQ) in the N-terminal intrinsically disordered region (IDR) of the protein. ATXN2 CAG repeats are interrupted by CAA codons which is relevant only for DNA and RNA but not for protein since CAA also encodes glutamine (Q).

The ataxin-2 polyQ lengths vary in the healthy population, with 22Q and 23Qs as the most frequent variants. ATXN2 polyQ length variation is associated with smaller volume of some brain structures (putamen, thalamus, amygdala and globus pallidus) potentially affecting cognition in old age. Genome Wide Association Studies have shown that genetic variants in ATXN2 and its surrogate ATXN2L are associated with cognition and educational attainment (reviewed in Laffita-Mesa et al., 2021a).

When polyQ length surpasses certain thresholds, it becomes toxic to specific neuronal populations (main target Purkinje cells) reverberating in multiple neurological diseases. For instance, the presence of at least one allele of > 22polyQ lowers the Age at Onset (AO) in patients with transthyretin familial amyloid polyneuropathy with the Val30Met variant (Santos et al., 2019).

Intermediate-length ATXN2 polyQ (29-to-31Q) modulate AO in SCA3/MJD and is a proven disease modifier for C9ORF72-ALS (van Blitterswijk et al., 2014; de Mattos et al., 2019). Hence, intermediate-length ATXN2 polyQ confers an increased risk for a broad number of neurodegenerative diseases including amyotrophic lateral sclerosis (ALS), Parkinson’s disease, progressive supranuclear palsy, Alzheimer’s disease, frontotemporal dementia and multiple system atrophy (Figure 1A and B) (Lee et al., 2011; Ross et al., 2011).

PolyQ length ≥ 32Qs is the monogenic cause for spinocerebellar ataxia Type 2 (SCA2) (Imbert et al., 1996; Pulst et al., 1996). Furthermore, the polyQ repeat ATXN2 variation explains ~40–60% of Age of disease Onset (AO) and others SCA2 phenotype markers (Figure 1B). In SCA2 CAG-repeats are unstable in both mitotic and germline tissues showing a marked bias toward larger repeats expansions in paternal transmissions which underlies the genetic anticipation phenomenon (Figure 1C).

Therefore, the disruption of ataxin-2 functions caused by polyQ expansion, may affect multiple brain structure/circuit involved in motor, emotions, cognitive functions and may also affect metabolic pathways.

Is the ataxin-2 polyQ tract the only quilt for different neural phenotypes? PolyQ expansions are present in the coding regions of other eight genes causing other neurodegenerative diseases such as Ataxin-1/SCA1, Ataxin-3/SCA3, CACNA1A, SCA6, Ataxin-7/SCA7, FUS/TLS (Fused in Sarcoma/Translocated in Liposarcoma) and Ataxin-2-like protein.

Emerging ATXN2 variants in exon-11/ intrinsically disordered region (IDR) and neural phenotypes: Current pathophysiological mechanisms are built on models with either hyperexpanded polyQ repeats (gain-of-function) or ablation of the whole gene (loss-of-function). Other studies have focused in ataxin-2’s Like-Sm and Like-Sm Associated domains. The SCA2 field has been CAG/polyQ centric and apart the interspersed CAA interruptions within the CAG repeats and the two single nucleotide polymorphisms (SNPs) rs695871 and rs695872 no other variants were shown as relevant contributors to neural phenotypes. A recent work examined ATXN2 variants as modulators of phenotype variability in different neurodegenerative disease cohorts (SCAs, C9ORF72-ALS, SCA3, and Parkinson’s Disease) (Laffita-Mesa et al., 2021b). That work identified a 9bp duplication (c.109_117dupCGGAGCCGGG) in the ATXN2-S/ATXN2-AS region as potential modulator for C9ORF72-ALS and SCA3 which are otherwise modified by intermediate ATXN2-CAG repeats. The structural change causes for the $R_1^2$;$S_1^2$-$G_1^2$, resulting in $R_1$-$R_2$=$\Sigma_1$-$S_1$-$G_1$-$R_3$, located in the ataxin-2 N-terminus (Figures 1D-F, 2A and 2B). The variant was in trans with intermediate alleles, suggestive for pathological allelic interaction (lower section in Figure 2D).

The current heterologous expression systems are not capable to express this large protein. In silico 3D predictions are consistent with an intrinsically disordered protein (Figure 2A). There are multiple mouse models that, together with fly and yeast, demonstrate that ataxin-2 is a cytosolic protein crucial for RNA metabolism. Ataxin-2 binds to polysomes while activating and stabilizing targets transcriptomes by polyadenylation. Ataxin-2 is also constitutive factor for RNA and stress granules. RNA granules are very dynamic higher order structures regulating the timing, trafficking, and transcriptomic expression. The main interacting proteins for these functions are polyadenylate-binding protein 1, TAR DNA-binding protein 43 (TDP-43), tia1 cytoplasmic granule-associated RNA-binding protein-like 1, Eukaryotic initiation factors, Staufen, fragile X mental retardation 1, Ataxin-2-binding protein 1 (A2BP1 also known as RBF0X1), RNA-binding protein FUS/TLS (Fused in Sarcoma/Translocated in Liposarcoma) and Ataxin-2-like protein. Furthermore, ataxin-2 functions as activator of mRNA that controls circadian rhythm, but also repress CamKII translation influencing synaptic plasticity and memory (Lim and Allada, 2013). Yeast ataxin-2 proteins were observed as cloud-like structures surrounding the mitochondria functioning as redox sensor, coupling autophagy to the metabolic state of mitochondria (Lin et al., 2020). Moreover, ataxin-2 is involved in endocytic processes (CIN85 and endophilins-1/2 as interactors) and lipid metabolism (Figure 1E) (reviewed in Scoles and Pulst, 2018).

Ax1atxin-2 is a cytosolic protein with pleiotropic functions: Human ataxin-2 is a large 140 kD cytosolic protein with ubiquitous expression, high expression in neural tissues, and the cerebellum having one of the highest levels (https://gtexportal.org/home/gene/ATXN2). This protein is mainly localized in the trans-Golgi network (main functional domains are shown in Figure 1D and E). It has been difficult to elucidate its 3D structure and physiological functions as it is not possible to obtain crystal structures.

Perspective
In C9ORF72-ALS is co-expressed with CAG/CAA sequence mosaic consisting of 8CAG–1CAG–4CAG–1CAG–8CAG and 7CAG–2CAA–4CAG–1CAA–8CAG. This may have potential toxic consequences at DNA/RNA levels. McGurk et al. (2021) showed that mechanisms associated with the purity of the sequence of the ataxin-2 CAG repeats tract contribute to disease (McGurk et al., 2021).

Droplet digital PCR in retrotranscribed mRNA supported 9bp-dup expression consistent with use of the ATG at 495bp upstream the CAG repeat (Transcript IDs in Gtex: ENST00000608853.5, ENST00000483311.5, ENST00000616825.4, ENST00000550104.5, ENST00000377617.7). However, there is another ATG 15 bp upstream to the CAG repetitions encoding for (ENST00000535949.5, ENST0000389153.8, ENST000000549455.1, ENST000000542287.6) which may not include the variant and produce a smaller ataxin-2. As shown in Figure 1F, the longest splice variants ENST00000608853.5 and ENST00000535949.5 are widely expressed in brain tissues and spinal cord in healthy donors (https://gtexportal.org/home/gene/ATXN2).

The locus for the 9-bp duplication (12–111599394, GRch38) is variable in different European populations at GnomAD and different protein alterations are predicted. For instance, the variant rs776634254, 12–111599058-C-A GRch38 may introduce stop codons (p. Glu153Ter). Hence, the different variants may modify the expressivity of other neurological diseases. Our ongoing studies are aimed for the above and other variants in the close vicinity of the CAG repeats (star colored in orange at ATXN2-S/AS locus with diverse genetic variants located in the close vicinity of the CAG repeats (star colored in orange (Laffita-Mesa et al., 2021a) (Figure 2D). Nonetheless, there is another ATG 15 bp upstream to the CAG repetitions encoding for ATXN2 with 22CAG repeats in Figures 2D–ALS is co-expressed with CAG/CAA sequence mosaic consisting of 8CAG–1CAG–4CAG–1CAG–8CAG and 7CAG–2CAA–4CAG–1CAA–8CAG. This may have potential toxic consequences at DNA/RNA levels. McGurk et al. (2021) showed that mechanisms associated with the purity of the sequence of the ataxin-2 CAG repeats tract contribute to disease (McGurk et al., 2021).

We have provided evidence that the small 9-bp duplication could be a phenotype modifier in neural diseases (SCA3 and C9ORF72-ALS (Figure 2E)). For this we ruled out other modifiers and found that cases with the duplication develop SCA3 and ALS earlier than others with the same mutation (Laffita-Mesa et al., 2021b). Nonetheless, larger cohort studies are necessary to better control for effects of the environment and to define the contribution to the overall disease variance in SCA2, SCA3 or other diseases connected with ATXN2. This limitation may be sorted out through international collaborations as SCAs, and genetically determined ALS are rare diseases.

The transcriptomic profile for the Natural Antisense Transcript ATXN2-AS in carriers shows differences compared to controls (Laffita-Mesa et al., 2021a) (Figure 2F). Ongoing studies are aimed to better define sequence differences for these ATXN2-AS splice variants throughout brain tissues targeted in SCA2 and ALS.

All these genetic alterations are in the intrinsically disordered domain (Figures 1E and 2A–C). These regions are crucial for liquid-liquid phase transitions, for hydrogel formation and RNA granules formation. Liquid-liquid phase transitions are determined by weak interactions among intrinsically disordered regions of proteins with mRNAs resulting in condensed proteinaceous membrane-less organelles containing halted transcriptomes bound to ribosomal units, translation factors and RNA binding proteins (RBP). The vast majority are cytosolic transcriptomes which are condensed with RNA binding proteins and are transported bound to motor proteins along axons to their destination at the synaptic terminals and dendrites for further processing, catalysis and/or trafficking. IDR are also prone to assemble into amyloid like fibers and is increased when RBP are mutated (Figure 2H).

Genetic alterations in the ataxin-2 N-terminal IDR could be in cis or trans with the polyQ repetitions with different interacting allelic effects (Figure 2D). At the protein level this is translated into hybrids repetitive stretches with and/or without non-expanded polyQ. These polypeptides may aggregate, disturbing TDP43 localization, and harmfully sensitizing cells at the same extent of full CAG expansions. For instance, the combination of the 21bp duplication c.432_452dup, Pro145-Ala151 in cis with 22CAG repeats in ATXN2 was presented in an ataxic individual with no other causative mutation (Inada et al., 2021) (Figure 2C and I).

Finally, the oxidation of ataxin-2 C-terminal IDR may also disrupts the ataxin-2 function in coupling mitochondrial redox state and autophagy (Lin et al., 2020) (Figure 2J).

Future perspectives Abnormal expansions of CAG, CUG and polyQ tracts located in the ATXN2 remain as the main causative and modifying factor for several diseases. Notwithstanding the above, other genetic changes in the same region deserve attention as our knowledge about ATXN2 pathogenesis and pleiotropism is still limited. Structural
Figure 2 | Functional consequences and genetic mechanisms of ataxin-2 variants in the W-terminal IDR.
(A) 3D Alpha Fold prediction for the human ataxin-2. The polyQ (Q166-Q188) is indicated as an alpha helix in light blue. The LSM motifs harbor residues from Q254-S345, the LSM-AD F353-Q475. (B) 3D structure prediction for the region R37-338-G39 where 9bp duplication is located. (C) Prediction for the region P145-A151 with 21bp duplication. The model confidence for regions dark blue colored is very high (predicted Local Distance Difference Test, pLDDT > 90), light blue confident (80 > pLDDT > 70), yellow low (70 > pLDDT > 50) and very low (pLDDT < 50) for orange colored. The regions below 50 pLDDT may be unstructured in isolation. (D) ATXN2 genetic variations and allelic interaction, polyQ and hybrids repeats. (E) Phenotype modifier effect of the 9bp dup, R37-338-G39-R40-G42-R. The mutation is in the IDR located in the N-termius and it was found to decrease the age at onset for SCA3 and C9orf72 ALS. (F) For the ATXN2-AS it was found aberrant splicing in carriers of the 9bp duplication. (G) Several nucleotide changes are in the primer binding site for the most common used primer set limiting the accuracy in the genetic testing services. (H) Liquid-liquid transition for hydrogen formation aided by IDR in ataxin-2. (J) Observed effect of the 21-bp dup in cell viability. Inada et al., 2021 found an ataxic case with normal ataxin-2 polyQ expansions. The genotype was 22/22CAG, but one allele had a 21-bp duplication doubling the P145-A151 tract close to the polyQ tail in ataxin-2. This makes one allele with 43 repeats units of hybrid stretch, polyQ and a double P145-A151 tract and the other allele with 22 units only. In a set of functional experiments, they transfected cells separately with wild type and the 21-bp dup followed by treatment with increasing amounts of H2O2. Cells with the duplication showed decreased survival, TDP43 mislocalization and increased susceptibility to cellular stress. (J) Other studies shown that the oxidation of ataxin-2 IDR may disrupt the ataxin-2 function in coupling mitochondrial redox state and autophagy (Lin et al., 2020).

We apologize to authors whose work were either not referenced in this perspective or cited as part of review articles due to space limitations.

Parkinsonfonden and ALF support from Region Stockholm. PS was a Wallenberg Parkinsonfonden and ALF support from Region Stockholm.