
Bioinformatic Analyses of the Ataxin-2 Family Since Algae Emphasize Its Small Isoforms, Large Chimerisms, and the Importance of Human Exon 1B as Target of Therapies to Prevent Neurodegeneration

[Georg W. J. Auburger](#) ^{*}, [Jana Key](#), [Suzana Gispert](#), [Isabel Lastres-Becker](#), Luis-Enrique Almaguer-Mederos, [Carole Bassa](#), Antonius Auburger, Georg Auburger, Aleksandar Arsovic, Thomas Deller, [Nesli-Ece Sen](#) ^{*}

Posted Date: 12 January 2026

doi: 10.20944/preprints202601.0750.v1

Keywords: evolution; endosymbiosis; terrestrialization; gene duplication; mitochondrial targeting sequence; polyglutamine neurotoxicity; antisense oligonucleotides; polyAla repeat; LSm-fold; unfolded protein response



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Bioinformatic Analyses of the Ataxin-2 Family Since Algae Emphasize Its Small Isoforms, Large Chimerisms, and the Importance of Human Exon 1B as Target of Therapies to Prevent Neurodegeneration

Georg W. J. Auburger ^{1,*}, Jana Key ², Suzana Gispert ^{2,3}, Isabel Lastres-Becker ^{2,4}, Luis-Enrique Almaguer-Mederos ², Carole Bassa ⁵, Antonius Auburger ⁶, Georg Auburger ⁷, Aleksandar Arsovic ², Thomas Deller ¹ and Nesli-Ece Sen ^{2,8,*}

¹ Institute for Clinical Neuroanatomy, and

² Experimental Neurology, Clinic of Neurology, Faculty of Medicine, Goethe University, Frankfurt, Germany

³ Institute of Neurophysiology, Neuroscience Center, Goethe University, Frankfurt am Main 60590, Germany

⁴ Department of Biochemistry, Medical College, Autonomous University of Madrid (UAM), Madrid, Spain; Instituto de Investigación Sanitaria La Paz (IdiPaz), Madrid, Spain; Instituto de Investigaciones Biomédicas Sols-Morreale (CSIC-UAM), Madrid, Spain; Centro de Investigación Biomédica en Red, Área Enfermedades Neurodegenerativas, CIBERNED, Instituto de Salud Carlos III, 28029, Madrid, Spain

⁵ genOway, F-69362, Lyon, France

⁶ Clarity Labs, Sydney, Australia

⁷ AKTORmed, Neutraubling, Germany

⁸ Department of Molecular and Cellular Biology, Faculty of Sciences III, University of Geneva, Geneva, Switzerland.

* Correspondence: auburger@em.uni-frankfurt.de (G.W.J.A.); nesliecesen@gmail.com (N.-E.S.)

Abstract

Polyglutamine expansion in Ataxin-2 (ATXN2) is responsible for rare, dominantly inherited Spinocerebellar Ataxia type 2 (SCA2). Together with its paralog Ataxin-2-like (ATXN2L), both proteins received much interest since deletion of their yeast and fly orthologs alleviates TDP-43-triggered neurotoxicity in Amyotrophic Lateral Sclerosis models. Their typical structure across evolution combines LSm with LSm-Associated Domains and a PAM2 motif. To understand the physiological regulation and functions of Ataxin-2 homologs, the phylogenesis of sequences was analyzed. Human ATXN2 harbors multiple alternative start codons, e.g. from an intrinsically disordered sequence (IDR) present since armadillo, or from the polyQ sequence that arose since amphibians, or from the LSm domain since primitive eukaryotes. Multiple smaller isoforms also exist across the C-terminus. Therapeutic knockdown of polyQ expansions in human ATXN2 should selectively target exon 1B. PolyQ repeats developed repeatedly, usually framed and often interrupted by (poly)Pro, originally near PAM2. The LSmAD sequence appeared in algae as the characteristic Ataxin-2 feature with strong conservation. Frequently, Ataxin-2 has added domains, likely due to transcriptional readthrough of neighbor genes during cell stress. These chimerisms show enrichment of rRNA processing; nutrient store mobilization; membrane strengthening via lipid, protein, and glycosylated components; and cell protrusions. Thus, any mutation of Ataxin-2 has complex effects, also affecting membrane resilience.

Keywords: evolution; endosymbiosis; terrestrialization; gene duplication; mitochondrial targeting sequence; polyglutamine neurotoxicity; antisense oligonucleotides; polyAla repeat; LSm-fold; unfolded protein response

1. Introduction

Ataxin-2 was discovered as the disease protein responsible for Spinocerebellar Ataxia type 2 (SCA2) [1–3], and became widely known because its deletion mitigates the disease progression of TDP-43-triggered Amyotrophic Lateral Sclerosis (ALS) or Fronto-Temporal Dementia (FTD) [4,5], and perhaps also a Tau-triggered form of Parkinson plus syndrome that was named Progressive Supranuclear Palsy (PSP) [6–9]. SCA2 is a rare neurodegenerative process with autosomal dominant inheritance, starting in the spinal cord, brainstem, and cerebellum and leading to a typical combination of muscle cramps, deficient locomotor coordination, intention tremor, and slowed saccadic eye movements, so that patients over the course of 6-25 years become wheelchair-bound, bedridden, and die from swallowing problems [10–14]. The ages of clinical manifestation occur earlier in successive generations of a pedigree, and the molecular cause of this “anticipation of onset age” is the expansion of a polyglutamine (polyQ) repeat domain near the N-terminus of Ataxin-2 [15–17]. In most healthy individuals worldwide, this domain would contain 22Q, normally encoded by (CAG)₈-CAA-(CAG)₄-CAA-(CAG)₈ [18]. The CAA interruptions in this repeat stabilize the length during genome replication [19], and once they are lost, the DNA polymerase will have difficulty in copying the exact length, leading to somatic mosaicism, anticipation over generations, and eventually the appearance of a pure CAG-repeat of excessive size that shows neurotoxicity with accumulating properties [20–23]. At intermediate sizes of 31Q to 33Q, this neurotoxicity becomes relevant in a polygenic manner, affecting only the most vulnerable neurons (spinal and cortical motor neurons in ALS and FTD, respectively, or the nigrostriatal dopaminergic pathway in Parkinson’s disease and PSP) [4,24–31]. From the threshold of 34Q upward, the neurotoxicity is strong enough to act in a monogenic manner with increasing severity in the multi-system neurodegeneration known as SCA2, either with spinocerebellar deficits upon adult onset [32–34], or with developmental delay and seizures upon pediatric onset caused by very long polyQ domains [35]. Expansions where one CAA interruption is still present have been reported several times to be associated with a clinical picture of Parkinson’s disease or of ALS [36–40]. Preliminary data mainly from animal models also suggest that Ataxin-2 modifies the disease course of other spinocerebellar ataxias (SCA1 / SCA3), adult motor neuron disease (ALS/FTD), and tauopathy-associated Parkinsonism [4–7,9,16,28,41–47].

However, it is completely unclear how Ataxin-2 exerts this role as an important modifier gene for neurodegenerative processes. Ataxin-2 localizes normally to the cytoplasm. A minor quantity of Ataxin-2 associates with the Golgi apparatus [48–50], modulates endoplasmic reticulum (ER) dynamics [51], and associates with the endocytosis complex at the plasma membrane [52,53], where it mediates the fast recycling of the Notch receptor [54]. Thus, the role of Ataxin-2 for membrane dynamics was documented early on but was not investigated intensely, while an additional role of Ataxin-2 for RNA has caught most of the attention. The main quantity of Ataxin-2 is present at the rough ER, where translation of membrane proteins and secreted proteins occurs [55], and associates with polyribosomes [56]. Some Ataxin-2 was reported to interact with actin and fimbrin [57–59], and furthermore, some evidence in favor of microtubule association was published [8,60–63], so a role of Ataxin-2 for cell protrusions is apparent. The diffuse distribution of Ataxin-2 throughout the cytoplasm changes dramatically in periods of cell stress, when it relocalizes with damaged RNA to the cytoplasmic stress granules, where RNA surveillance and triage occur [64–66]. Particularly oxidative stress is known to cause the maximal recruitment of RNA and translation machinery to stress granules, and one report suggested that Ataxin-2 influences the fate of superfluous mitochondria as the main source of oxidative stress [67], a notion that is strengthened by evidence that Ataxin-2 controls the expression levels of PINK1 as the protein kinase which phosphorylates PARKIN to initiate the autophagic elimination of mitochondria, and that Ataxin-2 associates with PARKIN as the ubiquitin ligase which tags mitochondrial outer membrane proteins as targets of removal [68–70].

It is still unclear how these localizations reflect Ataxin-2 functions and interactions in specific pathways. As novel insights, this study analyzes Ataxin-2 orthologs to document isoforms and the evolution of different domains and regions, thus elucidating their regulation and functional context.

The discovery of extra domains that were integrated with Ataxin-2 into chimeric proteins during evolution, and their explanation as transcriptional readthrough across neighbor genes, triggers the notion that mutations in Ataxin-2 have complex effects on its isoforms, but also on the function of genomic neighbors.

The bioinformatic efforts in the current manuscript must take into account how multi-domain proteins such as Ataxin-2 became necessary, and how extra-domains are still getting added even in vertebrates. Recent studies in *Euglena gracilis* microalgae observed that cells with additional energy from endosymbionts can increase their cytoplasmic volume and show highly flexible cell shapes [71]. However, the diffusion of substrates, cofactors, and proteins becomes insufficient for efficient metabolism in such larger cells. This causes selection pressure that favors assembling multi-component complexes non-covalently, or even better, placing several domains from single pathways within one protein, rather than relying on diffusion. In this manner, the assembly of the LSm domain, LSmAD, PAM2 motif, and IDRs into one large protein would maximize the throughput in a chain of reactions, which were important for eukaryotic life. The classical domain combination that constitutes Ataxin-2 is clearly able to respond to oxidative stress and to manage some form of damage control, for RNAs and possibly other vulnerable compounds.

Importantly, diverse eukaryotic species live in different niches, so they had to adapt their metabolic processes further. Therefore, additional domains were appended to the core features of Ataxin-2 in a minority of organisms. The spectrum of variation between these additional domains and functions reflects changes in stressors, as well as alternative stress responses among eukaryotic phyla. In our view, available data on this “tinkering of nature” [72,73] can be exploited to understand the core and accessory mutation effects not only for Ataxin-2, but for any poorly understood protein. The compilation of such additional domains and the analysis of statistically enriched features among them might contribute to understanding how nature modifies and optimizes the core role of an enigmatic protein like Ataxin-2. This approach could elucidate how Ataxin-2 impacts several overlapping pathways. The wealth of sequencing data on species from all kingdoms of life, which is available in databases such as UniProt/UniParc, NCBI/GenBank and EMBL, can provide many novel insights into physiological connections between metabolic responses and the specific abiotic stressors in each habitat. This approach consumes enormous amount of time when performed manually, but should be amenable to automation, and is applicable to many scientific questions.

2. Results

2.1. Overall Structure of Ataxin-2 Family Members Across Evolution

The multi-domain protein Ataxin-2 family has three conserved sequence motifs in a defined order [74–76]. Their structures are dominated by the conspicuous LSm fold (see <https://alphafold.ebi.ac.uk/entry/A0A370Q1D4>, or https://www.rcsb.org/structure/AF_AFA0A044URM1F1, both last accessed on 1. December 2025), with the LSm-associated domain (LSmAD) also standing out in alpha-fold predictions, while most remaining Ataxin-2 is characterized by intrinsically disordered regions (IDRs), where the linear PAM2 motif is usually embedded within the C-terminal half. The three successive domains - surrounded by IDRs and in less than 5% of species by extra domains of known function - are shown schematically in Figure 1a and are meticulously assessed in text paragraphs below.

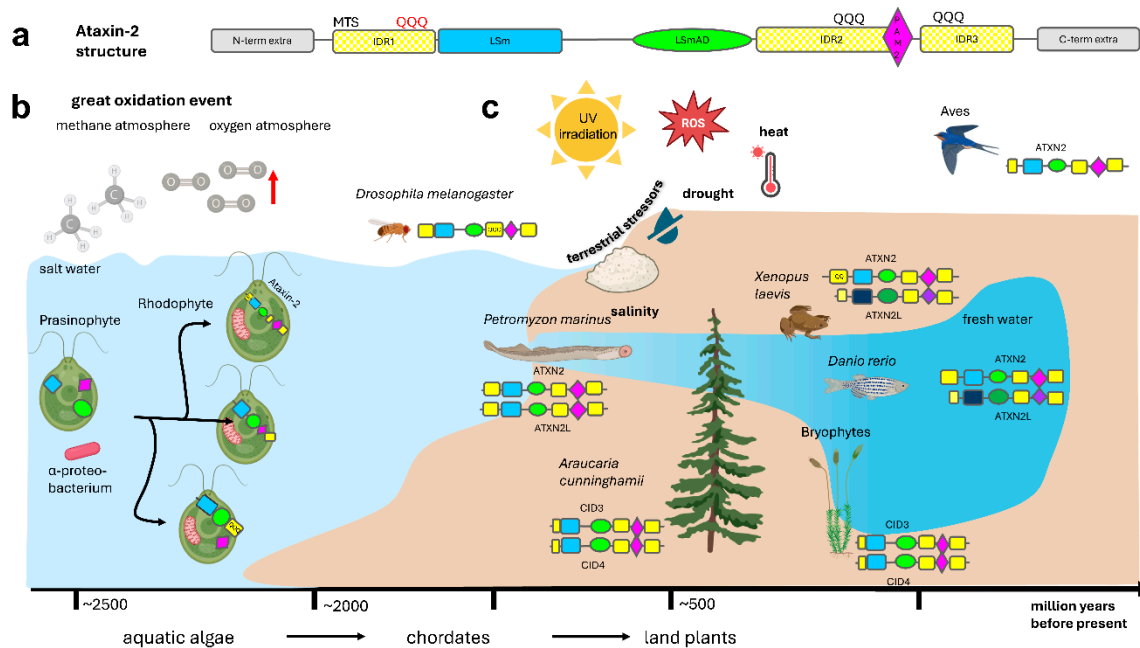


Figure 1. Schematic of the evolution of the Ataxin-2 protein family. **(a)** The structure of all members of the Ataxin-2 protein family contains an intrinsically disordered region 1 (IDR1, in yellow color) near the N-terminus, which contains a polyQ domain e.g. in primates, then an LSm domain with 5 beta-barrels flanked by an alpha-helix at each side (in blue color), after a spacer sequence followed by an LSm-Associated-Domain (LSmAD, in green color) with some alpha-helix structures, then IDR2 sometimes containing a polyQ domain e.g. in insects, a linear PAM2 motif (purple color), IDR3 sometimes containing a polyQ domain e.g. in mucorales fungi, and variant C-terminal sequences. The IDRs show poor sequence conservation across evolution (illustrated by hatched color). In less than 5% of Ataxin-2 orthologs, there are extra domains at the N-terminus or C-terminus (grey symbols) that are mostly functioning in membrane stress resistance or rRNA processing. **(b)** Ataxin-2 orthologs are not present in eubacteria, archeobacteria, or glaucophytes that evolved before the great oxidation event in the atmosphere. Ataxin-2 orthologs are documented in their typical and complete structure since rhodophytes that exploited rising oxygen levels via endosymbiotic mitochondria and chloroplasts, while defending against oxidative stress by carotenoid pigments. **(c)** Whereas one gene copy of Ataxin-2 is sufficient in insects and nematodes, and is again sufficient for flying birds, during the process of terrestrialization independent gene duplications were conserved from conifers to mosses and flowering plants (named as CID3/CID4) on the one hand, and from chordate fish to ray-finned fish, amphibia, reptiles until mammals (named as ATXN2/ATXN2L) on the other hand. PolyQ domains near the N-terminus or near the C-terminus existed already in algae, are found frequently upstream from the PAM2 motif in insects, and have expanded steadily in the unusually long N-terminus of ATXN2 since amphibians. Created in BioRender. Key, J. (<https://BioRender.com/8073e0f>) is licensed under CC BY 4.0.

Many functional insights are known for two of the three highly conserved domains, but the third least studied domain, was found to be characteristic of Ataxin-2:

- (1) Firstly, just after the ancestral N-terminus, the Like-SM (LSm) fold has a typical structure of small five-strand anti-parallel beta sheets (β_1 , β_2 , β_3 , β_4 , β_5), with SH3-type barrel tertiary structure. The ancestral quaternary structure is characterized by assembly into an LSm hexameric or heptameric ring, with U-rich RNA oligonucleotides binding inside the LSm torus lumen. Overall, the LSm domain in human ATXN2 and ATXN2L (synthetic full-length entries UniProt Q99700 and Q8WWM7) comprises 78 amino acid residues. LSm domains have descended from bacterial proteins like *Escherichia coli* Hfq and YlxS, as well as archaeal Sm1 / Sm2, which serve as RNA chaperones and in ribosomal pathways [77,78]. In eukaryotes, various Sm proteins in the nucleus form a heteroheptameric ring that is crucial for intron splicing [79].

Like-Sm proteins were subsequently also described in the cytoplasm, where LSM2-16 combine the RNA-binding sequence with a methyl-transferase domain [80,81], whereas Ataxin-2 combines the RNA-binding sequence with the LSmAD and the PAM2 motif [76,82]. Our datamining effort confirmed that the LSm sequence of Ataxin-2 has relevant similarity to the LSm domains of LSM2-16, so BlastP searches frequently confuse different families, and the sequence variability of the LSm domain is so strong in low organisms that current InterPro and Pfam algorithms fail to detect it in approximately one third of Ataxin-2 orthologs. The LSm/LSmAD region in Ataxin-2 binds to the RNA helicase DDX6 [64]. Together with the RNA helicase DDX6 and the LSm-containing factor LSM12, Ataxin-2 was found to influence circadian posttranscriptional regulation and olfactory habituation in fly neurons [83,84]. It is therefore important to note that an RNA helicase domain was chimerically added to Ataxin-2 orthologs in several species (see below for individual protein database entries).

- (2) Secondly, after a disordered bridge region of usually 50-60 amino acids, the LSmAD sequence stands out with a predicted alpha-fold structure. However, experimental analysis showed only a modest presence of α -helical structural elements, with a considerable degree of flexibility, devoid of tertiary structure and without RNA binding capacity [77]. Human Ataxin-2 LSmAD sequence (amino acids 409-477 in the synthetic full-length UniProt entry Q99700) contains a putative clathrin-mediated trans-Golgi signal (residues 414–416) and an ER exit signal (residues 426–428). Indeed, experimental analysis confirmed that deletion of 42 residues within LSmAD causes Golgi dispersion [50]. Overall, the ancient protein module comprising LSm and LSmAD with their connecting bridge sequences extends across some 250 amino acid residues in a very stable size across evolution, while most length variability of Ataxin-2 orthologs is due to IDR composition and length across the C-terminal half, and sometimes in short N-terminal regions. Here, it is important to note that our datamining effort found practically all LSmAD-containing sequences to represent Ataxin-2 orthologs, so the extremely well-conserved LSmAD domain is the unique characteristic feature of the Ataxin-2 family, and perfectly suited for the BlastP search for orthologs.
- (3) Thirdly, the 14 residues-short linear motif known as PAM2 was named as Poly(A)-binding protein interacting Motif 2. It connects to an MLLE sequence in the PABP C-terminus (also known as CTC, short for carboxy-terminal conserved domain), in dependence on nearby phosphorylation sites [62,78–81]. In plants, the PAM2 motif extends over 19 residues that contain a tandem duplicate of the core sequence [76]. Interestingly, its location is always outside globular domains [82], at approximately three-fifths of the protein length. It clearly functions to interact with mRNA 3' tails, and it exists as a component of over a dozen different eukaryotic proteins [82], several of which are known for their regulation of mRNA translation versus decay [79]. These protein families include PAIP1/2, LARP4, eRF3/GSPT1/2, TTC3, USP10, PAN3, GW182, Tob1/2, and other factors, so our datamining effort found its usefulness for Ataxin-2 ortholog searches to be limited, and particularly in low organisms the sequence variability of this motif makes its automated recognition by current InterPro and Pfam algorithms doubtful. The PAM2 motif was shown to prevent the phase separation of Ataxin-2 in cellular growth periods, while it localizes to the translation apparatus at the rough ER, promoting the relocation of Ataxin-2 to stress granules after cellular damage [55,56,83].

2.2. Compilation of Ataxin-2 Family Protein Sequences Until Excavata, Amoebozoa, and Algae

Conservation of Ataxin-2 sequences across evolution was used to judge the relevance of each domain and feature for physiology and pathology. Using the BlastP algorithm to search UniProt- UniParc, NCBI, and EMBL databases for ATXN2 homologs, the longest or otherwise relevant isoforms for many organisms were compiled in Table S1, but it was impossible in the current study to compile all isoforms and assess their genomic coding sequences. Separate tabs in this table show either an overview of the longest orthologs (Table S1a), or specifically the N-terminal intrinsically disordered domain 1 (IDR1) before the polyQ domain (Table S1b), sequences across the polyQ

domain with its flanking residues and interrupting residues (Table S1c), the IDR1 sequences around and after the polyQ domain (Table S1d), structured sequences around and after the LSm motif (Table S1e), structured sequences around the LSmAD motif (Table S1f), the intrinsically disordered domain 2 (IDR2, Table S1g), sequences across the PAM2 motif within IDR2 (Table S1h), ordered sequences before and after IDR3 with C-terminal variability (Table S1i). Some ATXN2L orthologs were compiled in Table S2. While this compilation cannot be exhaustive, it covers the main classes of eukaryotic organisms.

It is difficult to pinpoint the origin of Ataxin-2, given that precursors of Ataxin-2 in the databases cannot reliably be distinguished from fragmented Ataxin-2 due to sequencing problems, proteolysis, or isoforms. Furthermore, the absence of Ataxin-2 orthologs from species or organism classes may simply be due to incomplete sequence documentation, so only crude conclusions are possible. But as the prime result, Ataxin-2 orthologs were clearly identified in all eukaryotic kingdoms until amoebzoa, protozoa, and algae, but not in archaeobacteria or eubacteria. Details are provided in the following paragraphs of this chapter.

Figure 1b depicts the evolution and domain structure of Ataxin-2 proteins in all kingdoms of life. Among amorphea-amoebzoa, Ataxin-2 homologs as in UniProt entries A0A151ZGE7 and Q55DE7 with 900-1100 aa in *Dictyostelium* are already complete with N-terminal polyQ stretches, LSm, LSmAD, and a typical PAM2 motif recognized by the InterPro or Pfam algorithms.

Among protozoa, Ataxin-2 homologs like A0A1Z5KDA2 and A0A9N8E681 with 700-900 aa in stramenopiles (the most ancient among them have chloroplasts surrounded by four membranes as a testimony to secondary endosymbiosis [84]) have this multi-domain structure and C-terminal polyQ stretches. In phaeophytes (brown algae), as a photosynthetic class among stramenopiles, UniProt- UniParc entries A0A6H5L901, D7FVW6, and D7G1N4 in *Ectocarpus* species represent typical Ataxin-2 orthologs with short polyQ stretches upstream from LSm.

Interestingly, A0A0G4IZN6 with 400 aa in rhizaria contains a structured sequence not recognized as LSm by current algorithms, then LSmAD, IDRs with PAM2 embedded (PAM2 sequence takpkLNANAKTFT-MSAAAKEFV), but no polyQ stretch.

In naegleria from the excavata/discoba-schizopyrenida branch (that alternates between amoeboid and flagellate stages, having mitochondria with discoid cristae similar to *Euglenozoa* [85]), the UniProt entries A0A6A5CC11, A0AA88H5V6, and D2V0R5 with about 900 aa represent complete Ataxin-2 orthologs with N-terminal polyQ stretches.

In the closely related euglenozoan branch (the only photosynthetic branch among the excavata), smaller complete orthologs are documented. A0A0L1KIJ6 with >800 aa in *Perkinsella* shows a potential LSm coexisting with LSmAD and a primitive PAM2 motif (LNPNATAFLP) without polyQ stretches; A4H699 or A4HUM3 or Q4QHA3 in *Leishmania* (potential PAM2 precursor core sequence PNPSATPFVP), A0A1X0NQC5 in *Trypanosoma* (potential PAM2 core FNPAATPYTP), W6KMY5 in *Phytomonas* (potential PAM2 core PNPAAPFVP), A0A0M9FU89 in *Leptomonas* (potential PAM2 core PNPSATPFVP), all with >500 aa, show a LSm and LSmAD coexisting with a C-terminus that contains several polyQ stretches. Trypanosomal A0A061J3B7 with 279 aa, and A0A0S4IS39 with 213 amino acids, show LSm and LSmAD coexisting without further elements. Given that trypanosomes were reported to have exchanged genetic material with green algae and other archaeplastida [86,87], orthologs of Ataxin-2 were studied in detail among algae.

Ataxin-2 orthologs are also documented among alveolata (A0A813ABV5 with >1100 aa that contain an extra 2-hydroxyacid dehydrogenase domain, with homology to bacterial SerA protein that is essential for the biosynthesis of L-serine as precursor of phospholipids and sphingolipids). Among cryptista, one protein with significant LSmAD homology is documented (L1I7C3 with 522 aa).

In contrast, Ataxin-2 homologs appeared absent from the diplomonad and parabasalid-metamonad branches of excavata protists, which lack mitochondria, so these protists depend on nutrients and energy from hosts. In the monocercomonoides species from excavata, endosymbiont mitochondria were completely lost again [88], and an Ataxin-2 homolog could not be detected, while the microsporidia that lost mitochondria but retained mitosomes with FeS-metabolism-dependent

oxidative stress [89] did preserve the presence of Ataxin-2 orthologs (e.g., A0A4Q9L9Z0 and A0A086J3B3). These findings support the concept that Ataxin-2 is needed to protect cells from the oxidative stress emanating from endosymbionts.

Among algae, many entries for Ataxin-2 orthologs exist from chlorophytes (green algae) with well above 1000 aa (e.g., A0A8J4CZB7 or A0A383VPF7) that contain LSm, LSmAD, polyQ stretches shortly after LSmAD, and typical PAM2 motifs embedded in the long C-terminal IDRs, as well as a short N-terminal IDR, in close structural and length similarity to human ATXN2L and ATXN2. In the particularly interesting green algae *Cymbomonas tetramitiformis*, whose ancestors are thought to have played a crucial role in the primary endosymbiosis of plastids [90], A0AAE0L4G1 is documented as a typical Ataxin-2 ortholog of 668 aa without polyQ stretch. Similarly, among archaeplastida-rhodophytes (red algae), the three entries A0A1X6NXA5, A0A5J4YT85, and A0A2V3IVX2 (noteworthy for a C-terminal uninterrupted 30Q repeat) represent Ataxin-2 orthologs.

However, no Ataxin-2 homologs were detected in glaucophytes (green-blue algae, which have plastids surrounded by an inner membrane containing a remaining peptidoglycan layer, suggesting they are still synthesized by plastids as by ancestral cyanobacteria, having mitochondria with the typical eukaryotic flat cristae surrounded by a double membrane, and using asexual reproduction). They are considered the basal archaeplastida that engulfed cyanobacteria in an act of primary endosymbiosis, within a terrestrial freshwater ecosystem [91–95]. Therefore, Ataxin-2 orthologs apparently were not needed for the engulfment of mitochondria or plastids. Instead, the appearance of Ataxin-2 might be associated with the gained ability to synthesize robust inner membranes by eukaryotes rather than by plastids. Apparent absence of Ataxin-2 homologs was also observed in the mesostigma branch of green algae, which contains primitive plate-like chloroplasts, lacking a periplastidal ER [96,97], and was therefore placed at the basis of divergence between chlorophyta (green algae) and streptophyta (land plants). This finding agrees with the notion that Ataxin-2 presence correlates with eukaryotic synthesis of organellar membranes.

Overall, the appearance of Ataxin-2 ancestors occurred only after glaucophytes had developed the endosymbiosis of mitochondria and primary endosymbiosis of plastids (Figure 1b). Ataxin-2 ancestors are only documented since rhodophytes learnt to cope with the oxidative stress due to sun exposure and mitochondria/chloroplast presence [98]. Rhodophytes still have chloroplasts without external ER or unstacked (stroma) thylakoids, but developed α - and β -carotene (provitamin-A), lutein, and zeaxanthin as new antioxidative pigments [99], and reproduce sexually, while lacking centrioles and flagella. Carotenoids play a basic role as membrane reinforcers in view of their rigid conjugated double-bond backbone, as phylogenetic precursors of sterols [100,101]. When marine life forms subsequently moved to freshwater habitats and underwent terrestrialization, duplications of the Ataxin-2 gene occurred (Figure 1c).

2.3. Gene Duplication ATXN2 / ATXN2L in Animals, and CID3 / CID4 in Plants, Upon Entering Freshwater and Land

Datamining with STRING homology searches revealed that the gene duplication giving rise to land animal ATXN2 versus ATXN2L proteins is not only documented from mammals via *Xenopus laevis* to *Danio rerio* but is documented during earlier evolution until lampreys (UPI0014023775 vs UPI001403D160 in *Petromyzon marinus*, UPI002AB6496F vs UPI002AB6167E in *Lethenteron reissneri*) (Suppl. Material S1). Lampreys start life as larvae in freshwater, irrespective of whether they later live in coastal sea areas, rivers, or lakes. ATXN2 orthologs among fish species share more similarity with each other than with ATXN2L orthologs (Figure 2a), with specific residues being characteristic of ATXN2 versus ATXN2L (Figure 2b). The Pro vs Ser difference at the beginning of the L3 sequence within the LSm fold would cause a bend in the peptide chain and possibly result in a beta-turn only in ATXN2 orthologs, whereas the Ser/Ala vs Pro difference in the center of L5 may cause such a bend or beta-turn only in ATXN2L orthologs (Figure 2b). This suggests that different interactors and functions for ATXN2 versus ATXN2L have evolved. Thus, two copies of Ataxin-2 with differing point mutations are first detected in chordate lampreys as jawless fish that derive from the Devonian

period, some 360 million years ago [102]. The gene duplication without distinguishing sequence variants may also be present in other chordates, like hagfish. Indeed, three rounds of genome duplication events are known to have occurred during fish evolution, one of them during the Devonian [103]. With many genes duplicated, chordates were enabled to develop notochords and neural tubes with dorsal versus ventral specifications, which are established via signaling with retinoic acid (a carotenoid vitamin-A derivative) to control homeobox expression and somite formation [104–111], a pathway that establishes anterior-posterior gradients and is particularly pathway for the development of hindbrain, cerebellum and spinal cord that are affected most strongly in SCA2 and ALS [112].

The question of whether *ATXN2* or *ATXN2L* is the duplicated copy was assessed by genomics. The relevant human chromosome locus on 12q contains the gene order *PPP1CC* – *CCDC63* – *MYL2* – *VHRT* – *CUX2* – *PHETA1* – *SH2B3* – *ATXN2* – *U7* – *BRAP* – *ACAD10* – *ALDH2* – *MAPKAPK5* – *TMEM116* – *ERP29* – *NAA25* (underlined genes share the same transcriptional direction; *U7* refers to the 63 bp uridine-rich non-coding snRNA ENST00000607576.1, a potential binding target of Sm domains, which is positioned on the opposite strand downstream from polyQ-encoding exon 1, and upstream from exon 2 of human *ATXN2*, as the start of the LSm sequence encoding exon cluster), whereas the locus on 16p contains the gene order *SULT1A1* – *NPIP8* – *EIF3C* – *NPIP9* – *ATXN2L* – *TUFM* – *SH2B1* – *ATP2A1* – *RABEP2* – *CD19*, according to the human genome annotation in the UCSC genome browser. The synteny of *ATXN2* with *SH2B3*, and of *ATXN2L* with *SH2B1*, is conserved since *Nile tilapia* fish. Before the gene duplication event, in fish such as *Callorhinchus milii* (elephant shark) or *Oryzias latipes* (medaka), the genome contains only the *SH2B3* – *ATXN2* – *BRAP* locus. It is tempting to speculate about a functional interaction of *ATXN2* with *SH2B3* and *ATXN2L* with *SH2B1*, given that all these proteins contain proline-rich motifs and associate with GRB2 during PI3K/AKT growth signaling at neuronal membranes [53,113–116]. Such a potential competition for GRB2 would be a relatively recent development during evolution, given that *SH2B3* is no neighbor gene of *ATXN2* in the genomes of lancelet fish, *D. melanogaster* flies, or *C. elegans* nematodes, and is therefore not crucial for the ancient functions of Ataxin-2. However, it is conceivable that *SH2B3*, *ATXN2*, *BRAP* are jointly relevant for human diseases such as SCA2, ALS, cancer, and longevity, either via their gene variants in linkage disequilibrium, or via protein interactions [29,117,118]. Overall, *ATXN2* clearly evolved from ancient Ataxin-2 in its original genomic context, while *ATXN2L* represents the duplicated copy in a new genomic environment.

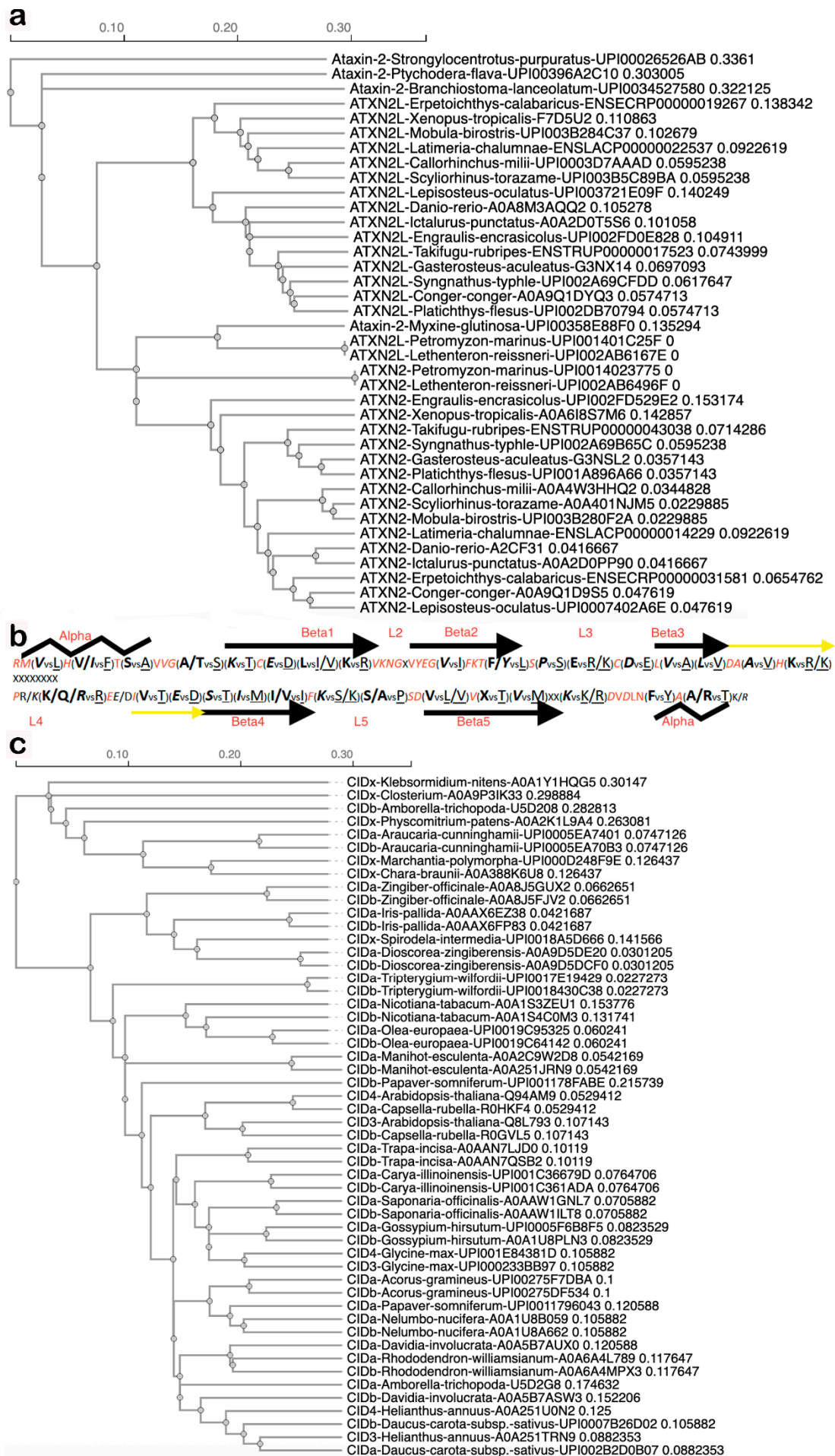


Figure 2. Duplication of the Ataxin-2 gene resulted in animal transcripts encoding ATXN2 versus ATXN2L, with characteristic differences of sequence and structure, while an independent duplication resulted in plant transcripts encoding CIDa versus CIDb, which co-evolved without consistent distinguishing features. **(a)** Guide tree of LSm domain evolution in fish generated by CLUSTAL Omega Multiple Sequence Alignment software shows single copy Ataxin-2 sequences (always with species name and database entry ID) in sea urchins, acorn worms, hagfish and early chordates (*Branchiostoma*), while other chordates (*Petromyzon marinus*, *Lethenteron reissneri*) already have two copies documented with point mutations, which remain typical and lead to divergence of ATXN2 from ATXN2L in all descended species into two separate groups. We use the term Ataxin-2 when an organism has only one copy or to refer to the gene family, while ATXN2 or ATXN2L are used to distinguish the two copies of a species. **(b)** Fish ATXN2 versus ATXN2L sequence specificities, in red color if invariable, in *italics* if conserved in ATXN2 until human, showing the invariable sequences to correspond to linker regions between predicted alpha and beta structures, whereas the 5 beta sheets typical for LSm folds and the surrounding alpha-helices exhibit characteristic signature residues that differentiate ATXN2 vs ATXN2L. **(c)** The guide tree of LSm sequences during plant terrestrialization shows single-copy Ataxin-2 orthologs named as CIDx in green algae, mosses, liverworts, duckweed, while double-copy Ataxin-2 orthologs (randomly dubbed CIDa or CIDb when there were no criteria to assign them to the CID3 group or the CID4 group) are documented since evergreen conifer trees and earliest flowering plants that still bear tracheids. While the two copies in each species usually localize together or nearby, a big exception was observed for *Amborella*, which had one primitive copy closely related to *Closterium* algae and one sophisticated copy that exhibited sequences akin to flowering *Rhododendron*.

ATXN2L is thought to be missing in birds. However, non-flying but feather-bearing kiwi birds in northern New Zealand have ATXN2 and ATXN2L entries in UniProt (*Apteryx mantelli* A0ABM4FEG3 and A0ABM4G1B9). Apparently, ATXN2L got lost before the evolution of flight, because ostriches, emu, and other kiwi species do not record ATXN2L sequences. It is known that large segmental deletions occurred when the macroevolution of avian genomes favored constrained size and became relatively static [119].

Among plants, two Ataxin-2 homologous gene copy translation products are not only documented in eudicot brassicales such as *Arabidopsis thaliana* (Q94AM9 vs Q8L793) and *Capsella rubella* (R0HKF4 vs R0GVL5) and other flowering land plants, but already in paleodicot / ANA-grade angiosperms such as amborellales (U5D2G8 vs U5D208 in *Amborella trichopoda*) and in gymnosperms such as conifers-araucariales (UPI0005EA7401 vs UPI0005EA70B3 in *Araucaria cunninghamii*) (Suppl. Material S2). The ancestral angiosperms, like amborellales as well as gymnosperms, underwent enormous diversification after two whole genome duplication events since 319 million years ago during the late carboniferous period [120], which was possibly the ancestral event where CID3/CID4 appeared. While the Ataxin-2 orthologs in conifers still have >800 amino acids, similar to chlorophyte Ataxin-2 orthologs, CID3 and CID4 in flowering plants show a length of around 600 residues, consistent with the notion that some of their stress resistance properties were lost during evolution to flowering plants due to unusually short IDRs at either end of the proteins.

Conifers were the first trees among land plants. They are still the dominating plant species, e.g., in the taiga of Siberia, where they show exceptional resilience to winters with temperatures below the freezing point, managing to stay evergreen. Their cold hardiness involves the sensing of low temperatures and a short photoperiod, to trigger molecular adaptations, such as (i) improved membrane fluidity, with the development of epicuticular wax (composed mainly of nonacosane, which is used as a component of paraffin, as 29-carbon straight-chain alkane that is derived from very-long-chain fatty acids [121]), (ii) the mobilization of solutes such as glycogen to lower the freezing point and minimize cold shock of vulnerable RNA, (iii) downregulation of photosynthesis and cell growth in parallel to upregulated expression of cold-induced proteins [122].

Importantly, there are no specific residues that are characteristic of either CID3 or CID4. Instead, both gene copies in each organism were more similar to each other than to the orthologs in other

Overall, events that duplicated the Ataxin-2 gene occurred twice during macroevolution while moving out of marine life, with both duplications becoming conserved with independent fates. On the one hand, one event among chordates of shallow waters was inherited by land animals, resulting in ATXN2 versus ATXN2L acquiring separate functions reflected by different protein lengths. On the other hand, another Ataxin-2 gene duplication occurred among conifers as early tree forms with unusual stress resilience, resulting in CID3 versus CID4, which lost some protein length and did not develop consistent distinguishing features. This is compatible with the notion that mechanical stress due to gravitational forces, more sunlight intensity, heat and cold exposure, and oxidative stress made it advantageous to have two gene copies when marine life forms underwent terrestrialization. The fact that flying birds lost the Ataxin-2 gene duplication might suggest that the higher weights of organismal bodies on land and the mechanical force of waves in shallow water require strengthened cell membranes. In contrast, life-forms that are suspended by buoyant forces in deep ocean water or in the air have less demand on membrane stability.

A comparison of the LSm and LSmAD sequences in human ATXN2 versus ATXN2L reveals very similar predicted structures, while pinpointing the putative crucial differences (Figure 3a/b). Additional comparison of the polyQ domain and the LSm domain in the human species versus their first appearance in the red alga *Gracilariopsis chorda* underlines the strong conservation of predicted structures, and presumably also their functions (Figure 3c/d).

2.4. Genomic Comparison of Exon-Intron Structure for Human and Murine Ataxin-2 Versus Ataxin-2-like

Mapping of the individual exons within genomic sequences (as documented at <https://www.ncbi.nlm.nih.gov/gene/6311>, last access on August 15, 2025) shows that human ATXN2L exons on chromosome 16 (Suppl. Material S3) stretch across 13,823 bp, with large blocks of non-coding sequences after exon 1, then (with the LSm domain encoded by exons 3-5) large non-coding blocks again after exon 6, then (with LSmAD encoded by exons 7-8) after exon 10, and again (with PAM2 motif encoded by exon 15) after the final exon 24B.

Human ATXN2 exons on chromosome 12 (previously studied in [124]) have a similar distribution overall (as shown at <https://www.ncbi.nlm.nih.gov/gene/11273>, last access on August 15, 2025), but they are found across a >10-fold larger genomic area with 146,713 bp (Suppl. Material S4), suggesting that much more sophisticated control over expression and splicing of exon clusters has become possible. Massive blocks of non-coding sequences are found after exon 1 (where the polyQ repeat is encoded), then (with the LSm domain encoded by exons 3B-5) again after exon 5, a smaller block after exon 6, then (with LSmAD encoded by exons 7-8) smaller blocks each after constitutive exon 9, alternative exon 10, constitutive exon 11, and again massive blocks after exon 14 (encoding IDR2, with the PAM2 motif encoded by exon 16), exon 18A, exon 18B, exon 20, exon 21, and the final exon 25.

The extraordinary similarity of translation starts, aligned exon boundaries (Suppl. Material S5, employing CLUSTAL Omega 1.2.4), and of the genomic clustering of several exons around the LSm, around the LSmAD, around the PAM2 motif, and the C-terminal intrinsically disordered sequences may reflect a joint regulation and separate functions in common for each cluster.

In addition to the ancient conserved and aligned exons, human ATXN2 has developed alternative splicing at least for exons 1A, 1B, 2, 3A, 10, 21, 23B, and 25 ([124–126], UCSC genome browser). Clearly, the regulation of human ATXN2 appears much more sophisticated than ATXN2L, given that it shows many more genomic non-coding control sequences between exons, at least three alternative translation starts ([127], UniProt, NCBI protein database) within the extended N-terminus containing the expanded polyQ stretch, many options for alternative splicing, longer sequences encoded by exons 11 and 13, and a much more variable C-terminus. Further analysis of murine *Atxn2* versus *Atxn2l* confirmed the strong adherence to exon boundaries but also documented several alternatively spliced exons for *Atxn2l* (Suppl. Material S5, [128,129]). The fact that Ataxin-2-like remained relatively unchanged during evolution, while Ataxin-2 underwent numerous adaptations to fine-tune its functions, indicates that human Ataxin-2 has acquired additional features and that

SCA2/ALS pathogenesis involves diverse mechanisms, which are not necessarily modelled in lower organisms.

Importantly, the human ATXN2 sequence shows two methionine residues as potential start codons at the initial LSm sequence, and database entries confirm that corresponding isoforms are translated. Similarly, the initial LSm sequence of ATXN2L orthologs contains one or two methionine residues. While such an Ataxin-2 translation start right at LSm was normal in ancient species like *Trypanosoma*, *Perkinsela*, and other protists, and was still producing the main isoforms among reptiles (Table S1a), a short intrinsically disordered N-terminal extension with IDR1 became usual already among protists, fungi, worms, and insects. An extension of the N-terminal IDR1 occurred progressively for ATXN2 homologs since fish, with one methionine at the LSm start being lost. However, there is database evidence that a minor isoform that starts from the LSm domain has always existed from zebrafish to primates (Table S1d).

2.5. C-Terminal Fragment Isoforms Are Prominent According to Exon Expression Analyses, and C-Terminal Epitopes Are the Target of Current Commercial Antibodies

Interestingly, RNAseq-based analyses of individual exon expression strength in central nervous tissue (as documented at <https://www.gtexportal.org/home/gene/ATXN2>, last accessed on October 8, 2025) reveal that the exons named 22, 23, 24A, and the C-term alternatively spliced final exon in Suppl. Material S6 has the highest score (in spinal cord 1.48 and 1.26, respectively, encoding C-terminal intrinsically disordered sequences with 7 Met residues). Only about half of these maximum expression levels are observed for exons 15 (containing 1 Met) with scores at 0.644 and exons 16 (encoding PAM2), 17 (3 Met), 18 (2 Met), 19 (3 Met) with scores above 0.8. Less than a quarter of these maximum expression levels are found for exons 5-12 (around the LSmAD sequence, including 1 Met in exon 5, 6 Met in exons 9-12), with scores above 0.407. Around a sixth of the maximum expression levels are apparent for exons 2-4 (encoding the LSm fold start, including two Met), with scores between 0.174 and 0.252. Particularly low expression at 3% of maximum levels is seen for exons 1A and 1B (encoding the N-terminal intrinsically disordered sequence and the polyQ domain, including 2 Met residues), with scores at 0.04.

Also, the GTEEx Gene Model based on ENSEMBL-documented transcripts of human ATXN2 (as documented at <https://www.gtexportal.org/home/gene/ATXN2>, last accessed on October 8, 2025) indicates the existence of many transcripts that comprise only the C-term disordered sequences, or begin at sequences downstream from PAM2, from LSmAD, and from LSm. Transcripts encoding only the LSm domain are always combined with LSmAD, but transcripts encoding only the LSmAD domain (entry ENST00000471866.5) or the sequences between LSmAD and PAM2 (ENST00000546483.1) are documented. Overall, only half of the transcripts, including all alternatively spliced transcripts, span the length of the ATXN2 gene, a tenth spans only the N-terminal parts, while some 40% span across C-terminal parts. The N-terminal parts seem less relevant than models of 3D-structure around the conspicuous LSm-fold would suggest, while the C-terminal intrinsically disordered sequences have a little-appreciated importance. Therefore, despite the notion of synthetic full-length Ataxin-2 being a multi-domain protein of >1300 residues, cells appear to have retained during evolution a great ability to use most of its components selectively.

These transcript findings suggest that short protein isoforms may exist with a Met start codon in the center or towards the C-terminal end of human ATXN2, which were previously considered as partial sequence fragments. Indeed, database entry Q99700.2 reflects a translation start from N-term disordered sequence, NP_002964.4 reflects a start from the polyQ-repeat, NP_001297052.1 reflects a start from the LSm domain, but there are also entries like XP_059732367.1 where the start lies upstream from LSmAD, or EAW97959.1 / EAW97960.1 / A0A2K6UG79 / A0A8C5XNX2 / A0A811ZS54 / A0A8C4TUK8 / A0A3S2PRM8 / A0A3Q3N8E0 / A0A3Q3FSG1 / A0A672RDQ6 where the start lies downstream from LSmAD, or A0A2K5CKD3 where that start occurs downstream from LSmAD but the PAM2 motif is selectively excluded, or A0A7J8EWQ7 / A0A7J7R7F2 where the start lies just upstream from PAM2, or A0A7J7S175 / A0A7J8BWL6 / A0A5F5Y0Q1 / A0A212D9N5 /

A0A5N4C9F7 / A0A5N4C9H8 / A0A401TLV8 where the start locates downstream from PAM2. A schematic depiction of the human ATXN2 protein sequence with potential methionine start codons, isoform candidates with UniProt-UniParc database entries, and the encoding exon structure is illustrated in Figure 4.

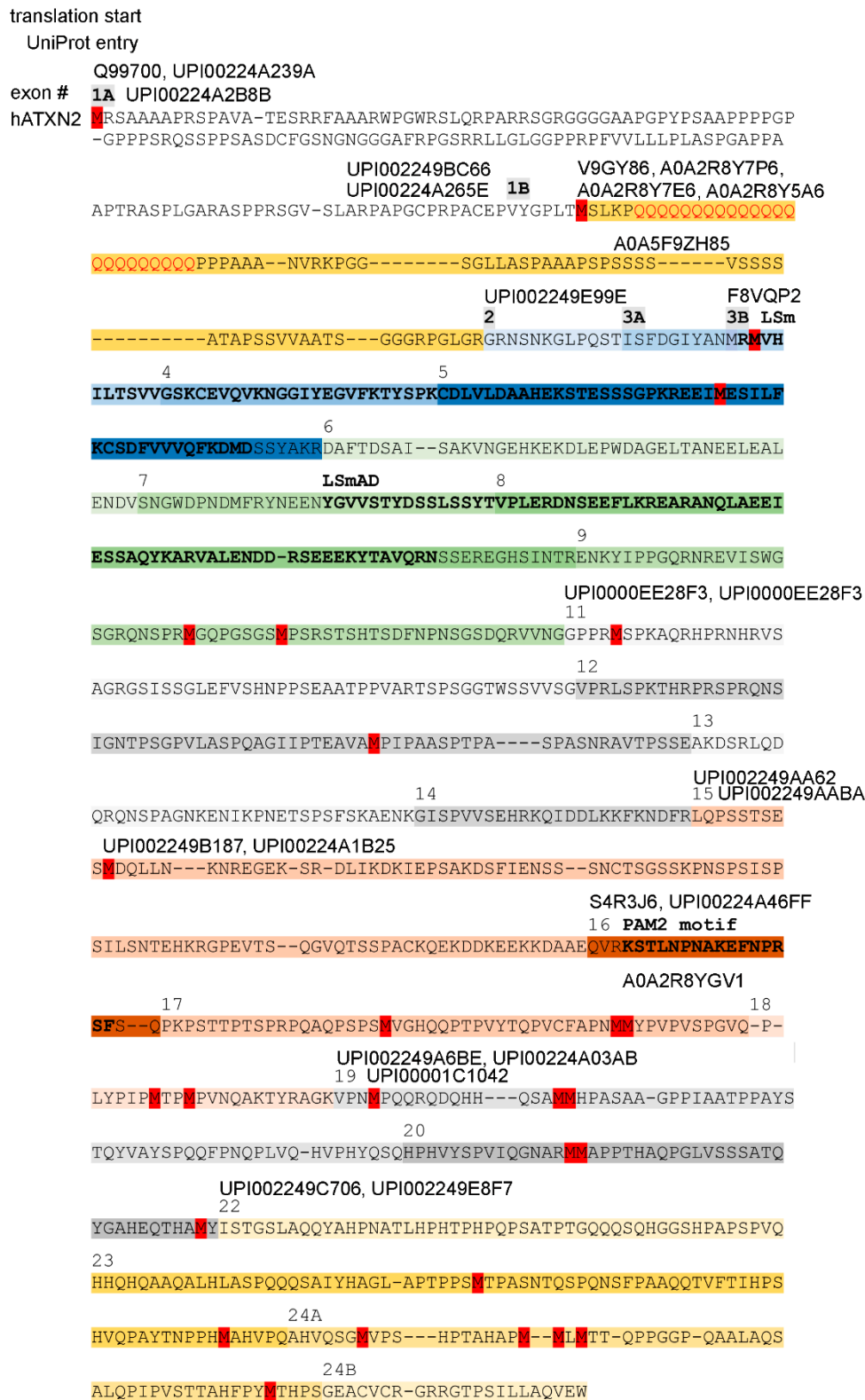


Figure 4. Multiple translation starts resulting in many long and short isoforms are documented in protein databases. The scheme shows a version of human ATXN2 without the sequences encoded by alternatively spliced exons 10 and 21, corresponding to the alignment of human ATXN2 versus ATXN2L and murine ATXN2

versus ATXN2L in Suppl. Material S5. The coding of amino acids by individual exons is illustrated by different background color hues, with the exon number indicated above the starting amino acid, highlighting the LSm and LSmAD domain as well as the PAM2 motif above the residues that stand out in bold letters. UniProt-UniParc database entry IDs are placed at different positions above the amino acid thread to indicate where each isoform of ATXN2 converges with the consensus sequence, while ignoring any isoform-specific start codon and downstream divergent amino acids that precede the constitutive ATXN2 fragments. The existence of multiple large and short isoforms of ATXN2 protein is supported by the analysis of mRNA isoforms (Figure S1) and by the molecular weight of bands detected in immunoblots by antibodies that target epitopes in the C-terminal half.

Commercial antibodies were often raised against full-length overexpressed Ataxin-2. However, in those cases where the companies used a fragment of Ataxin-2 as immunogen and revealed the epitope residues, a bias towards the more abundantly expressed epitopes downstream from LSmAD is obvious. Few commercial antibodies exist against the expanded polyQ domain, against IDR1 sequences, against the region across LSm or LSmAD, and against the sequences between LSm and LSmAD, while most targeted antibodies detect immunogens downstream from LSmAD (Figure S1 depicts the relevant immunogens relative to the exon structure of isoforms according to GTex database data). Reporting bias selects figures that represent the 120-130 kDa isoforms, while bands with smaller molecular weight are usually interpreted as unspecific cross-reactivity or are ignored. Thus, there is an important unmet need to study the C-terminal isoforms of ATXN2.

ATXN2 was reported to homo-dimerize or multimerize, and it is important to consider the fact that LSm-containing proteins assemble in homo-hexameric or homo-heptameric rings around RNA. Thus, these expression data would be compatible with a scenario where Ataxin-2 isoforms integrate into assemblies that are dominated by intrinsically disordered domains that promote RNP granule formation during microtubular transport. One sixth of the components would bear the LSm head that can be activated to bind a specific single-stranded AU-rich RNA sequence directly, and then mediate the association of a helicase to linearize the RNA, presumably in a stimulus-dependent manner at the end of transport, e.g., in a synapse before stimulus-dependent translation.

Given that the polyQ-encoding exon accounts for only 3% of maximum expression in the spinal cord, it seems plausible that there is a long temporal delay before polyQ-triggered accumulation and aggregation of N-terminal ATXN2 sequences become prominent in postmitotic neuronal tissue. In a mouse mutant with an ATXN2-Q100 expansion KnockIn, the levels of phosphopeptides from LSm to C-term in non-neuronal tissue are around 10% compared to WT [130], confirming that the expansion can lead to partial loss-of-function effects. However, in postmitotic neural tissue (the spinal cord) at 14 months of age selectively a phosphopeptide close to the polyQ-repeat before the LSm domain accumulates >110-fold relative to WT, and a phosphopeptide in the isoform that starts 160 residues even further upstream at the N-terminus accumulates >30-fold [131]. This suggests that either alternatively spliced expression of small isoforms or partial degradation via proteolysis is employed by cells, and the disease mechanisms in SCA2 and ALS do not necessarily concern the full-length protein. ATXN2-fragment-specific disease mechanisms also should be taken into account.

2.6. Most N-Terminal Start Codon with Subsequent Fragment Appears in Armadillo only for ATXN2

As mentioned above, the analysis of the ATXN2 start codons in humans and closely related primates shows at least 3 different translation starts across the N-terminus for isoforms in *Homo sapiens*, and 4 different isoforms in *Pongo abelii* as well as *Macaca nemestrina* (Table S1a). The most N-terminal alternatively spliced sequence starts with M(X)R(PX)S/T(XX)A... and is clearly conserved with quite constant length from primates until artiodactyla, and according to one database entry, even in armadillos (Table S1b), suggesting that it has functional relevance. This N-terminal extension is not observed for ATXN2L, so its function is specific to ATXN2. It starts with a sequence that has a strong likelihood to act as mitochondrial targeting signal (score of 0.9937 in MitoProt v1.101, with predicted cleavage site at residue 144; score of 0.75 at TPpred3 for residues 1-38), and then contains

three proline-rich motifs (PRMs) at amino acid residues 55-71, 117-123, 142-148 in human ATXN2 (Uni-Prot Q99700).

Sequences with significant homology (BlastP score $<5e-07$) to this fragment are not found in ATXN2 orthologs among paenungulata and marsupialia, but instead are contained in the SKI family transcriptional corepressor 1 (SKOR1), Actin nucleation-promoting factor (WAS), and Splicing Factor 3B subunit (SF3B4) from *Vombatus ursinus*, *Phasolarctos cinereus*, and *Monodelphis domestica*. In humans, close homologies exist with sequences within Transcription initiation factor TFIID subunit 4 (TAF4, $1.1e-9$) and the Splicing factor, proline-and glutamine-rich (SFPQ, $7.8e-8$), but also with Proline-rich protein 36 (PRR36, $4.6e-8$), Formin-2 (FMN2, $2.1e-7$), and Actin-binding protein WASF2 ($8.5e-7$).

Structural prediction software in InterPro characterizes this sequence as an intrinsically disordered region (IDR) with low complexity and proline-rich stretches, and previous literature has named all sequences upstream from the LSm domain as IDR1 of ATXN2 [62]. Thus, the above observations are in agreement with a genome-wide survey that concluded Pro-rich sequences to be involved in actin/cytoskeletal-associated functions, RNA splicing/turnover, DNA binding/transcription, and cell signaling [132,133]. At the N-terminus, this proline-rich sequence may influence the folding of nascent ATXN2. Its presence in recombinant full-length ATXN2 constructs for *in vitro* experiments, as well as the absence of such N-terminal sequences from ATXN2-null organisms, may have contributed to reports that ATXN2 is involved in actin-plastin pathways and in actin-endocytosis pathways [52,53,57,58,134]. However, the effects of Ataxin-2 on mitochondrial biogenesis, mitophagy, and oxidative stress responses cannot be attributed only to this IDR1, given that such an impact is already documented in the yeast ortholog Pbp1 [67–70,135–138].

2.7. The Usual Start Codon in Human ATXN2 / ATXN2L Is Followed by polyQ and a Repeat-Rich Fragment, Which Elongates Since Yeast / Insects

The second alternatively spliced exon in ATXN2 encodes the highly conserved translation start sequence MS(L/M)K(P)n(Q)n(P)n to produce an additional N-terminal fragment of up to 50-120 aa that contains several repeats, namely polyQ, polyPro, polyAla, polyGly and polySer, with one or several prolines usually at each side of the polyQ repeat. Within this sequence downstream from the polyQ repeat and before the polySer repeat, 17 amino acid residues enable N-terminal proteolysis of mammalian ATXN2 [139]. The entry A0A2K5X824 in *Macaca nemestrina* suggests that an alternative start sequence MGPHHVAEAP(A)_nT(A)_n can also be used, producing a fragment that contains a polyA repeat instead of the polyQ repeat. This fragment immediately before the LSm domain start does not show a relatively constant length, in contrast to the previous paragraph, but instead shows a massive extension across phylogenesis. A close inspection of the sequences compiled in Table S1d reveals the repeated appearance of polyQ repeats in testudines, again in monotremata, in proboscidea, and in primates. Instead of an N-terminal mammalian polyQ-repeat (MSLKPQQQQQQQQQQQQQQPPAAA) that is found in primate monkey (XP_054298120.1), birds and reptiles show polyAla (MSLKQAAAAAQAA in *Struthio camelus*, see entry XP_068766831.1), which seems to fulfil an analogous function [140] (Figure 5a). Both repeat types are known for their pathogenic expansion in patients [141,142]. PolyAla and polyGly repeats are found at each side of the fragment. In the center, a polySer repeat without or with interruptions has appeared with increasing size since monotremata. Around the first 100 amino acid residues of the human sequence are predicted to have an intrinsically disordered structure. The fact that this fragment extends in length since yeast / insects together with its poor conservation of individual amino acid residues may suggest that several repeats function as spacers between domains that interact, e.g., during autoregulation.

The N-terminal fragment before the LSm domain in ATXN2L, in comparison, has only doubled in length since *Danio rerio*, and contains only short repeats without much variation. It starts since *Xenopus* with the conserved sequence M(L/S)(K/M)(P/Q/K)(Q/P)(P/Q). About the first fifty amino

acids of human ATXN2L are predicted to show an intrinsically disordered structure. The total length of IDR1 sequences upstream from the LSm domain reaches 250 aa residues in human ATXN2L.

In contrast to previous assertions that long polyQ domains in ATXN2 are exclusively found in primates, this datamining effort observed ATXN2 orthologs with a Q>5 domain also in flying lemur, pika, rodents such as hamster / beaver / kangaroo rat, bats, carnivores, horses, unpaired and paired ungulates, armadillo, sloth, aardvark, duck-billed platypus, alligator, and leatherback sea turtle.

In human ATXN2 (UniProt Q99700), the InterPro sequence analysis shows the polyQ repeat at residues 166-188 to have a high alpha-fold confidence, and the aa residues 156-202 (YGPLTMSLKPQQQQQQQQQQQQQQQQQQQQQQPPPAANVRKPGGS), especially the residues downstream from the polyQ repeat, to display significant homology with the beta-sandwich domain in Sec23/24. The Sec23/24 heterodimer regulates the trafficking of membrane proteins, phospholipids, sphingolipids, and sterols via COPII-coated vesicles from the ER to the Golgi apparatus and to elongating cell protrusions [143]. This homology is not detected by the InterPro algorithm in other species. A fly dATX2 isoform indeed colocalizes with Sec23, but this was reported for isoform C, which starts from the LSm domain, similar to isoform A, but no Sec23 colocalization was found for isoform B, which starts 60 aa before the LSm domain begins [144]. While it remains unclear if Sec23 is relevant in this context, human ATXN2 indeed associates with the ER, disruption of Ataxin-2 in flies and nematodes disrupts ER dynamics, and the polyQ-expansion in human and mouse ATXN2, as well as the deletion of mouse ATXN2, leads to disturbed sterol and sphingolipid homeostasis [51,54,55,134,144–147]. To what degree N-terminal sequences of Ataxin-2 contribute to these effects remains to be elucidated.

The pathogenic expansion of the polyQ repeat within this translation start sequence to sizes $\geq 33Q$ is the monogenic cause of autosomal dominant SCA2, and at sizes $31 < (Q)n < 33$ contributes in a polygenic manner to the risk of sporadic ALS. It is interesting to note here that despite this long-term neurotoxicity of $>31Q$ stretches, long polyQ repeats appear among ATXN2 orthologs initially towards the C-terminus already in single-cell organisms like red algae (UniProt A0A2V3IVX2), amoebzoa (Q55DE7), dinoflagellates (A0A813ABV5, where an extremely long 73Q repeat is interrupted at three positions by single Met residues; or A0AA36IYW2, where a 41Q repeat is interrupted at 11 positions by single or double Ala residues), as well as insects like *Hymenoptera* (A0A158NW33, A0A836FT30, F4X2F6, A0A834JVP6, A0A834JR87), so they seem to play a relevant role for the functions of Ataxin-2, but are not documented for ATXN2L orthologs so far. Smaller polyQ repeat sizes $5 < (Q)n \leq 31$ are frequent among Ataxin-2 homologs towards the C-terminus (e.g., in chlorophytes such as A0A835VQM0, A0A9W6BKV0, nematodes such as MCP9264076.1, or insects such as Q8SWR8).

After this repeat-rich fragment but before the LSm domain, a connecting sequence is encoded by exons 2 and 3A in ATXN2 (for ATXN2L encoded by exon 2 and the first nine residues from exon 3), which appears in isoforms initiated from the first or second translation start codon and should therefore be regulated with exon 1B. However, their respective exons are separated genomically from exon 1, clustering instead with the subsequent exons that encode the LSm domain, possibly suggesting that this sequence may be regulated as an LSm modifier.

2.8. The Role of Proline Flanking Residues and Interrupting Residues for the polyQ Repeat

The observation that the pathogenic N-terminal polyQ repeat of ATXN2-IDR1 is adjacent to, or framed by, polyPro repeats or Pro residues (Figure 5b) is also documented for the pathogenic polyQ expansions in Huntington's disease [148]. Whereas PRMs are thought to require the presence of an Arg residue for their interactions with actin and endocytosis apparatus, the polyPro residues or Pro residues at the flanks or within polyQ repeats may play different roles. Pro residues are thought to modulate the poorly solvent polyQ stretch towards an α -helix secondary structure [149], and to mediate interactions e.g., with SH3 domains [150,151] and with membrane lipid bilayers [152]. The presence of polyPro helices in the intrinsically disordered N-terminal domain of human CPEB3 was shown to modulate solubility, acting as an amyloid breaker [153]. Importantly, proline- or glutamine-

rich activation domains are a common feature of transcription factors [154], so in the case of cytoplasmic Ataxin-2 orthologs, it is plausible that these sequences influence the binding to RNA.

Prolines not only serve a role in flanking the polyQ repeat. They are also frequently used to interrupt the polyQ domain. The interspersion of prolines or polyPro sequences within the polyQ repeat is observed not only in primates and some rodentia, but also in lagomorpha, pholidota, artiodactyla, marsupiala, and squamata (Table S1c). Detailed studies of Ataxin-2 orthologs with amino acid residues interrupting its polyQ domain indicate that Pro >>> His >> Ala > Met = Arg = Ser = Gly are most used to interrupt polyQ stretches. The flanking residue on the left side is most frequently Pro >> His > Met = Lys = Gly = Val = Ala = Arg, while the flanking residue on the right side is more often Pro >>> Met = Arg = His > Ala = Asp = Glu = Gly. Overall, the amino acid residues that flank and interrupt the polyQ domain are non-random. Flanking (poly)Pro or interrupting Pro was the most frequent feature throughout the evolution of Ataxin-2 orthologs.

a

SPECIES	DATABASE-ID	Sequence
Pongo pygmaeus (Bornean orangutan)	XP_054298120.1	M LK P Q Q Q Q Q Q Q Q Q Q Q Q Q Q P P A A A N V R K P G G G L L A P A A A P S S S V S S S A T A P S V A A A T G G G R P C L G R R N N K G L P Q T I F D G I Y A N
Oryctolagus cuniculus (European rabbit)	XP_051682770.2	M L K P Q Q Q Q P Q P Q P Q P P A A A A R K P G G G G G L L T P A A A P P P P P P S S S V S S S A A A A A G G G R P C L G R R N N K G L P Q T I F D G I Y A N
Marmota flaviventris (yellow-bellied marmot)	XP_034452669.2	M L K P Q Q Q Q Q Q P P A A A A R K P G G G G L L T P A A A P P S S S V S S S A A A A A A A G G G R P C L G R R N N K G L P Q T I F D G I Y A N
Meriones unguiculatus (Mongolian gerbil)	XP_060238432.1	M L K P Q P Q P P A P A G R K P G G G L L T P G A V P P V A T S S S T V P V P A A P V A S S A A A G G R P C L G R R N N K G L P Q T I F D G I Y A N
Myotis daubentonii (Daubenton's bat)	XP_059532310.1	M L K P Q Q Q Q Q P A A A A N R K P G G G L L T P A A A P P P V S S S T P A S S S S A A A A A A A G G R P C L G R R N N K G L P Q T I F D G I Y A N
Lutra lutra (Eurasian otter)	XP_047553287.1	M L K P Q Q Q Q Q Q Q Q Q Q Q Q P P P P P P A A A A N R K P G G G L L T P A A A P P S S S S A A A A A A A A A A A G G G R P C L G R R N N K G L P Q T I F D G I Y A N
Manis pentadactyla (Chinese pangolin)	XP_057347864.1	M L K P Q Q Q Q P P A A A A R K P G G G G L L T P A A A P P S S S V S S S A V A S S S S S A A A A A A A A A A A G G G R P C L G R R N N K G L P Q T I F D G I Y A N
Equus przewalskii (Przewalski's horse)	XP_070485015.1	M L K P Q Q Q Q Q Q Q P P A A A N R K P G G G G L L T P A A A P P S S S V S S S A V A S S S S S A A A A A A A A A A A G G G R P C L G R R N N K G L P Q T I F D G I Y A N
Balaenoptera musculus (blue whale)	XP_036679189.1	M L K P Q Q Q Q Q Q Q P P A A A N R K P G G G G G L L T P A A A P P S S S V S S S A V A S S S S S A A A A A A A A A A A G G G R P C L G R R N N K G L P Q T I F D G I Y A N
Bubalus bubalis (water buffalo)	XP_025122850.2	M L K P Q Q Q Q Q Q Q P Q P Q P P A A A N R K P G G G G G L L T P A A A P P S S S A A A S S S S V A A A A G G G R P C L G R R N N K G L P Q T I F D G I Y A N
Cholepus didactylus (sloth)	XP_037672752.1	M L K P Q Q Q Q Q Q Q Q Q Q Q T A A A N R K P G G G L L T P A A A P P S S A S S L P A A A S S S A A A A A G G G R P C L G R R N N K G L P Q T I F D G I Y A N
Elephas maximus indicus (Indian elephant)	XP_049722750.1	M L K P Q Q Q Q Q T A A A N R K P G G L L T A A A A P P S S S S A T A S S A A A S S A A A A A A A A A A G G G R P C L G R R N N K G L P Q T I F D G I Y A N
Sarcophilus harrisi (Tasmanian devil)	G3V222	M L K P P Q P Q P P Q P Q P Q P Q P P P P P A L P P A Q A Q P P A Q A Q P P A G A L T A A A A A G G A R K P G G L S S A S P A A G G G A G G R P C L G R R N N K G L P Q T I F D G I Y A N
Ornithorhynchus anatinus (duck-billed platypus)	XP_028914311.1	M L K P Q Q Q Q Q Q Q Q Q Q A A A G N R K P C L T S S S A S S A A A G G G G R Q L G R R S G K G L P Q T I F D G I Y A N
Grus monacha (hooded crane)	LA140182.1	M L K Q A A A A A Q A A G N R K P P G G G G G P L P A A G G G R Q N M R R R S G K G P P Q T I F D G I Y A N
Dermochelys coriacea (leatherback sea turtle)	XP_043353834.1	M L K Q Q Q Q Q Q Q X P Q A A A G N R K P P G G G G G P G P A A A G G G G G G G G G R Q N M R R R S G K G P P Q T I F D G I Y A N
Anolis carolinensis (green anole lizard)	XP_062814449.1	M L K Q Q Q Q Q Q Q P P P Q Q Q P A P P P P P P P P P P P Q G T A G G G G R K A P G G A G L G G G A G R A A L G R R N S G K G P L Q P I F Q G I Y A N
Ambystoma mexicanum (axolotl)	XP_069475627.1	M M K Q Q Q A G T N R K P C G G G G A P A P S G G G R Q N M R R R N S G K G P Q P I T E D G I Y A N
Xenopus laevis (African clawed frog)	XP_018099699.1	M M K Q Q A G N R K P C G G G V P A A A G G G G G R Q N M R R R N S G K G P P Q P V F D G I Y A N
Danio rerio (zebrafish)	A2CF31	M M K P Q A G G N R K P C G G N G A A A A A G T G G A G G G V G R Q N M R R R N N K A G P T A V A N G V Y A N

b

Ptilocolobus tephrosceles (Ugandan red colobus monkey)	XP_026312375.1	MSLK P Q Q Q Q Q Q Q P Q P P P P P P A A A N V R K
Saimiri boliviensis boliviensis (back-capped squirrel monkey)	XP_039327598.1	MSLK P Q Q Q Q P Q Q Q P Q Q Q Q Q Q P Q P P A A A N V R K
Aotus nancymaae (Nancy Ma's night monkey)	ENSANAP00000009056	LSLK P Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q P Q P P A A A N R K
Callithrix jacchus (common marmoset)	XP_035113537.1	MSLK P Q Q Q Q Q Q Q Q Q Q Q Q P Q P P A A A N V R K
Mus musculus (mouse)	NP_033151.3	MSLK P Q P P P A P A T G R K
Perognathus longimembris pacificus (Pacific pocket mouse)	XP_048198898.1	MSLK P Q Q Q P P Q Q P P P P P P A A A S A R K
Dipodomys merriami (Merriam's kangaroo rat)	XP_069897419.1	MSLK P Q Q Q Q P P P P P P Q Q Q P P P P P P P P P A A A S A R K
Dipodomys spectabilis (banner-tailed kangaroo rat)	XP_042524889.1	MSLK P Q Q Q Q Q P Q Q Q Q Q P Q Q Q Q P P P P P P P P P A A A S A R K
Mesocricetus auratus (Syrian hamster)	XP_040387936.1	MSLK P Q P Q P P P P S R K
Marmota flaviventris (rock chukker)	XP_034452669.2	MSLK P Q Q Q Q Q Q P P A A A T A R K
Eriaceus europaeus (hedgehog)	XP_060049065.1	MSLK P Q Q P Q P Q P P A V G A R K
Meles meles (European badger)	XP_045880703.1	MSLK P Q Q Q Q Q Q Q Q Q P Q P P P P A A A N A R K
Nannospalax galli (Middle East blind mole-rat)	XP_017651925.1	MSLK P Q Q P Q S P Q P P A A S T R K
Oryctolagus cuniculus (European rabbit)	XP_051682770.2	MSLK P Q Q Q P Q P Q P Q P P A A A S A R K
Canis lupus baileyi (Mexican wolf)	XP_072658620.1	MSLK P Q Q Q Q Q Q P Q P Q P P S A A S A R K
Enhydra lutris kenyoni (northern sea otter)	ADA2Y9KH12	MSLK P Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q P Q P P P P A A A N A R K
Bubalus bubalis (water buffalo)	XP_025122850.2	MSLK P Q Q Q Q Q Q Q Q P Q P P A A A N A R K
Nyctereutes procyonoides (common raccoon dog)	ADA811V488	MSLK P Q Q Q Q Q Q Q P P Q P Q P P S A A A S A R K
Canis lupus dingo (dingo, Papua singing dog)	XP_048957590.1	MSLK P Q Q Q Q Q Q P P Q P Q P P S A A A S A R K
Manis pentadactyla (Chinese pangolin)	XP_057347864.1	MSLK P Q Q Q P P A A S A R K
Tursiops truncatus (Atlantic bottlenose dolphin)	ENSTRP00000000897	MSLK P Q Q Q Q Q Q Q Q Q Q Q P P A H A R K
Ovis aries (sheep)	XDC63933.1	MSLK P Q Q Q Q Q Q P Q P P A A A N A R K
Sarcophilus harrisi (Tasmanian devil)	G3V222	MSLK P Q P Q P Q P P Q P P Q P P P P P P A L P P A Q A Q P P A Q A Q P P A G L T A A A A A G G A R K
Macrotis lagotis (greater bilby)	XP_074061826.1	MSLK P Q P Q P P Q P P Q P P Q P P P P A L P P A P A Q P P A G A P A A T A A G G A R K
Podarcis lilfordi (Lilford's wall lizard)	CAI5783228.1	MSLK P S P Q S P A A A A V T A A A V G A P P P P P L P P S P P Q P Q T T T T T T T T S Q A A P L P P P P Q G P A Q S H S S S S S S R K
Anolis sagrei (Cuban brown anole)	XP_060642029.2	MALK QSITTPASATGASAPSPSPPPPPQQQQQQQQAAPSPPPPPPPPPPPPGSGS=GGSGRRK
Struthio camelus (common ostrich)	XP_068766831.1	MSLK Q Q A A A A A Q A A G N R K
Grus monacha (hooded crane)	LA140182.1	MSLK Q A A A A A Q A A G N R K
Chrysemys picta bellii (painted turtle)	XP_065424825.1	MSLK Q Q Q Q Q A A A Q A A G N R K
Emydura macquarii macquarii (Macquarie river turtle)	XP_067395623.1	MSLK Q Q Q Q A P A Q A G N R K
Ambystoma mexicanum (axolotl)	XP_069475627.1	MSMK Q Q A G T N R K
Danio rerio (zebrafish)	A2CF31	MSMK P Q A G N R K

Figure 5. Evolution of the N-terminal IDR1 of ATXN2 (encoded by exons 1B, 2, and 3) with multiple repeat structures between ray-finned fish. **(a)** Noticeable length extension is documented, with the human sequence being too long to be included. While the sequences encoded by exons 2 and 3A were quite conserved, and the MS(M/L)K(P/Q) motif downstream from the translation start remained stable, a lengthening of polyQ, polyP, polyA, polyG, and polyS domains is apparent since turtles, reptiles, and marsupials (see also Table S1a). The polyQ repeat, whose expansion is pathogenic in the human disorders SCA2 and ALS13, has sizes of $n \geq 3$ since axolotl amphibia, $n \geq 5$ since turtles and other reptiles, $n \geq 10$ since *Platypus* and *Dasyurus* (armadillo) mammals, $n \geq 15$ since carnivores and bats, $n \geq 20$ since primates. **(b)** Non-random preference for Pro, Ala as interrupting residues or flanking residues across the polyQ domain. PolyA repeats arose next to polyQ in turtles, and substituted for polyQ in flying-birds, while polyP repeats were intermingled with polyQ and polyA in lizards. Pro residues came to flank the polyQ repeat and to replace residues within the polyQ repeat, since marsupials like the Tasmanian devil or the greater bilby. Thus, Pro and Ala residues appear to be exchangeable with Glu residues, with the physiological function of this IDR1 within ATXN2 remaining intact.

2.9. Ancient Start Codon Preceding LSm Domain as an Optional Third Start in Human ATXN2

The optional third translation start immediately upstream from LSm was in very frequent use during evolution until reptiles and is still documented in primates. The first of two methionine residues that may serve as a start codon in exon 3B was lost in muridae and cricetidae among rodentia, together with insectivora (VRMVHILTSVV... instead of MRMVHILTSVV...), but lagomorpha, dermoptera, and primates still have this first methionine (see Table S1e). Here, an isoform with abundant expression in spinocerebellar tissue can be produced that corresponds to Ataxin-2 from LSm until the C-terminus, without the appended N-terminus. This very ancient translation start codon produces an isoform with all core ATXN2 sequences (e.g., in XP_067991570.1). Deletion experiments in *Drosophila melanogaster* indicated that the LSm domain stimulates mRNA translation. At the same time, it antagonizes the assembly of ribonucleoprotein granules [83]. Via RNA binding, the LSm domain influences the assembly and stability of microtubules in *Drosophila melanogaster* neurites [63].

The crucial residues, with high conservation rate in the LSm domain of ATXN2 from mammals to insects, versus ATXN2L from mammals to amphibians, were compiled in Figure 2a/b. The LSm sequence variation in Ataxin-2 orthologs from nematodes to single-cell organisms is extensive, with insertions at various sites, so that multiple sequence alignments have difficulty defining core amino acids. It is worth noting again that the LSm sequence has limited usefulness in identifying Ataxin-2 orthologs, while the consensus LSmAD across evolution (and IDR2 / IDR3 within organism phyla), are much more powerful. Indeed, BlastP searches with the latter consensus sequences among fungi identify more than half of Ataxin-2 orthologs that have no recognized LSm domain, suggesting that the algorithms employed by InterPro and Pfam to identify LSm-folds need to be optimized. It is unlikely that the LSm-fold acts as a commodity rather than the necessity for Ataxin-2 functions, given that interference with this LSm domain is embryonically lethal.

The LSm domain, as an ancient RNA-binding motif with oligo(U)-specificity [155,156], is known to derive from Hfq protein in Bacteria / Archaea, where it assembles in homoheptameric rings and performs RNA chaperone function [157]. Hfq-binding RNAs were experimentally identified in *Salmonella typhimurium* bacteria, share a similar structure composed of three stem-loops, are enriched in the interacting sequence motif 5'-AAYAAYAA-3', and show Hfq-association at AU-rich single-strand sequences adjacent to the loops, or the polyU end of sRNAs [158–168]. Hfq interacts with regulatory short RNAs, modulates RNA decay, acts in ribosome biogenesis, and influences mRNA translation, showing localizations in the cytoplasm, in the nucleoid, and at the bacterial cell membrane [169–172]. Importantly, the loss of Hfq causes increased membrane permeability, more sensitivity to oxidative stress, and activates envelope stress responses via the sigma transcription factor RpoS for envelope stress [169,173–178], leading to the activation of outer membrane proteins, enzymes, and transporters, which are involved in amino acid uptake and biosynthesis, sugar uptake and metabolism, and cell energetics [169,177,179].

Hfq was adapted in eukaryotic nuclei as Sm modules in the heteroheptameric ring of splice proteins, and in eukaryotic cytoplasm in more than a dozen proteins implicated in RNA processing or decay, translation control, the telomerase RNA biogenesis, cell signaling, and gametogenesis [180–191]. Outside of ring structures, the LSm domain exists e.g., in LSM11 as a factor associated with the methylosome [192] or LSM12 as a possible methyltransferase and/or NAADP receptor [189,193]. Thus, the LSm-fold sequence consensus is not specific for Ataxin-2 orthologs. However, the LSm domain still associates preferentially with AU-rich sequences, usually in the 3' untranslated region of mRNAs [194].

2.10. RNA Processing Factors May Bind to the LSm-LSmAD Region, or Be Added to N/C-Termini of Ataxin-2 Orthologs, in Dependence on the Biochemical Needs of Different Ecological Niches

The region including the LSm and the LSmAD sequence contains around 250 residues, and in mammalian ATXN2, it was observed to mediate binding to the RNA helicase DDX6 [64,195], and to repress the illegitimate RAN translation of the CAG-repeat [196]. It is therefore interesting that

Coptotermes formosanus termites show a N-terminal stable addition of a DEAD box RNA helicase domain (DDX52 homologous) to their Ataxin-2 ortholog (UniProt entry A0A6L2P9Q7), *Trichogramma brassicae* wasps have a C-terminal addition of a DEAD box helicase (A0A6H5ISX2), and *Coemansia sp.* fungi have a C-terminal addition of an RNA recognition motif (A0A9W7Z4B9) (see Table S3). These observations support the concept that eukaryotic organisms have optimized multi-domain proteins in order to concentrate all cofactors and substrates of a pathway in one place, thus maximizing throughput efficiency.

In analogy, further RNA processing functions that complement Ataxin-2 roles may have been integrated into its multi-domain structure, to ensure the maximal efficiency needed for different environmental stresses. Observations of various domains added at the N- or C-terminus of Ataxin-2 orthologs were therefore compiled in Table S3.

A C-terminal addition of a sequence with homology to HhH-GPD domains, like NTH1, occurred in *Rhodotorula sp.* CCFEE 5036 fungus (A0A4U0VMY8). HhH-GPD, like NTH1, recognizes and repairs 8-oxoguanine and other oxidative damage in DNA [197,198].

N-terminal addition of a tetratricopeptide domain plus a SAM-dependent 23S rRNA methyltransferase domain in *Paramicrosporidium saccamoebae* fungus (A0A2H9TMQ1), only a N-terminal 23S rRNA methyltransferase domain in *Actinomortierella ambigua* fungus (KAF9164688.1), or C-terminal addition of a 28S rRNA (cytosine-C5)-methyltransferase domain in *Pseudozyma flocculosa* fungus (A0A5C3F1S9), and a C-terminal SAM-dependent RNA (cytosine-C5)-methyltransferase domain in *Wallemia ichthyophaga* fungus (A0A4T0IZQ6) to the Ataxin-2 sequence provide further support for the relevance of this pathway. Tetratricopeptide repeat contacts bind RNA and assemble the splice apparatus [199], while eukaryotic 5-methylcytosine (m⁵C) RNA methyltransferases are crucial for responses to oxidative stress [200,201].

Similarly, the N-terminal addition of Ribosomal protein L9 N-terminal and C-terminal domains to the Ataxin-2 sequence in *Coemansia interrupta* fungus (A0A9W8LDP4), *Coemansia spiralis* fungus (A0A9W8GBK0), *Coemansia bififormis* fungus (A0A9W8CWC1), and *Coemansia pectinata* fungus (A0A9W8H1X6) underlines this concept.

A ribonuclease inhibitor LRR repeat was added to the Ataxin-2 ortholog of *Rhizoctonia solani* fungi (A0A074SQD8) and of *Ceratobasidium sp.* fungi (A0A8H8MCZ2). The ribonuclease inhibitor serves to constrain the cytotoxicity of RNase A [202].

Regulation of the RNA polymerase was achieved by the addition of a CBF-B/NF-YA transcription factor domain in *Rhododendron giersonianum* shrub (A0AAV6I6Y5), or by the addition of an ACID domain of the MED25 complex in *Certonardoa semiregularis* sea stars (UPI001FC18B4E) to their Ataxin-2 orthologs. The MED25 complex regulates RNA polymerase via transcription modulators such as MYC [203].

The chimeric C-terminal addition of a RIO2 kinase domain in the bony fish *Anabarilius grahami* to the Ataxin-2 ortholog (ROL51429.1) points at rRNA processing, given that yeast RIO2 is required for the final endonucleolytic cleavage of 20S pre-rRNA at site D in the cytoplasm, converting it into the mature 18S rRNA [204]. The human ortholog RIOK2 also modulates neurotoxicity in mitochondrial complex I deficiency. In addition, it modulates ageing via telomere shortening [205,206]. It is important to note here that the *RIOK2* gene is localized next to the *ATXN2* gene since bony fish.

As further evidence that Ataxin-2 acts in RNA surveillance pathways, an N-terminal addition of a BRAP2 signature with an RNA-binding RRM motif, followed by a C3H2C3-type RING-H2 finger and a UBP-type zinc finger, occurred in the Ataxin-2 ortholog from *Pituophis catenifer* snake (XP_070807217.1). BRAP2 regulates cell survival via the antioxidant transcription factor NRF2 (SKN-1 in nematodes) and via the DNA repair factor BRCA1 [207–209]. Again it is noteworthy that the *BRAP* gene is the next neighbor of the *ATXN2* gene upstream from its translation start on chromosome 12q in the human genome.

Only a HECT-type E3 ubiquitin transferase domain was added to the N-terminus of Ataxin-2 in *Furculomyces boomerangus* fungus (A0A2T9YUP5), implicating this ortholog in protein degradation pathways.

Despite current concepts that Ataxin-2 family members impact RNA processing mostly for mRNA transcript turnover with a poly(A)-tail via the PAM2 motif, the chimeric extra domains added to Ataxin-2 are clearly enriched for rRNA processing functions like methylation, rather than for mRNA decay versus repair management. Thus, this protein family appears to contribute to adaptations of the translation machinery after oxidative stress, with the LSm domain providing direct binding to RNA, while the variation in LSm sequences between orthologs and paralogs may relate to differences in the target sequence.

However, the structured Ataxin-2 sequences from LSm to LSmAD are not only important for RNA processing and quality control. Observations in *Drosophila melanogaster* show that Ataxin-2, but not its intrinsically disordered sequences, is crucial for the pathfinding of neurites in vivo [210], and, interestingly, these roles of Ataxin-2 were also fine-tuned by chimeric extra domains in some species.

2.11. LSmAD Became the Hallmark of Ataxin-2 in Rhodophytes and Protists

The LSmAD sequence of ATXN2 in *Homo sapiens* is encoded by exons 7 and 8, with genomic clustering suggesting that exons 6 and 9 are co-regulated with LSmAD.

The LSmAD sequence is unique and is the hallmark for the Ataxin-2 protein family, in contrast to the LSm domain that also exists in the LSM1-16 protein family for each species, and in contrast to the PAM2 motif that also occurs in PAIP1/2, LARP4, eRF3/GSPT1/2, TTC3, USP10, PAN3, GW182, Tob1/2 and in other factors across the diverse species [211,212]. Since the presence of an LSmAD sequence defines the Ataxin-2 family, it is very unfortunate that there are no insights into the LSmAD functions or into the molecular interaction partners of LSmAD. Regarding the interaction of LSm with target mRNAs, the LSmAD sequence has an enhancing role in *Drosophila melanogaster* brains [213], and indeed it was shown to modulate LSm conformation [77]. Interestingly, it was also reported that LSmAD mediates the localization of ATXN2 to the Golgi apparatus [50]. Luckily, LSmAD is highly conserved in sequence and length across evolution, so it is uniquely helpful to identify orthologs of Ataxin-2. Using BlastP to search for homologs of yeast PBP1-LSmAD (FGVKSTFDEHLYTTKINKDDPNYSKRLQEAERIAKEESQGTSGNIHIAE-DRGIIIDDSGLDEEDLYSGVDRR) among single-cell organisms, nothing significant was detected in bacteria and archaea. The LSmAD sequence was also not detected in ciliate / flagellate parasites except *Trypanosoma* and *Plasmodium* (both adapted to oxidative stress environments like blood cells), and already in rhodophytes and chlorophytes. However, among eukaryotic protists, plants, and animals, many orthologs were found with high significance (BlastP score <1e-10) and compiled in Table S1f. Thus, the LSmAD sequence appeared only in species with multi-domain proteins involved in stress responses that evolved to counteract oxidative damage due to the endosymbiosis of mitochondria and chloroplasts, and due to exposure to sunlight.

Regarding crucial amino acids within this domain, it was reported that LSmAD contains a clathrin-mediated trans-Golgi signal (YDS, amino acid 414-416 in human ATXN2) and an endoplasmic reticulum exit signal (ERD, amino acid 426-428 in human ATXN2) [49]. However, the experimental evidence for their functional role was never unequivocal, with both not being well conserved in ATXN2, and ERD not being well conserved in ATXN2L either, despite the LSmAD sequence showing astonishing conservation since protists (Table S1f).

In view of this strong homology among LSmAD sequences throughout evolution (see Table S1f), we used them in extensive BlastP searches to define Ataxin-2 orthologs with unusual length (>1000 residues) that contain further protein domains of known function. The findings are reported in the last paragraphs of the Results text. It is relevant to know that many LSmAD-containing factors among chlorophytes were named CTC-interacting domain (CID) proteins, which refers to the presence of a PAM2 motif and its interaction with PAB1-binding protein (the MLLE domain of poly(A)-binding proteins was previously known as CTC domain) [214].

2.12. Sequence after LSmAD, Including the PAM2 Motif, Has a Polyampholytic Intrinsically Disordered Structure and Is Modified by Alternatively Spliced Exons Ante-10 and 10 in Human ATXN2

Downstream from the LSmAD border, the Ataxin-2 orthologous proteins are predicted to be mostly intrinsically disordered, with short interruptions at irregular positions. The total length of Ataxin-2 orthologs in protozoa and fungi varies between 650 and 1200 amino acids, with great variability and only a few consistent trends from algae to mammals. Practically all of the length variability occurs among the disordered sequences, much more towards the C-terminus in IDR2 before PAM2 (Table S1g), and IDR3 after PAM2, than at the N-terminus in IDR1. Thus, it is clear that the length and properties of these intrinsically disordered sequences are carefully controlled. The IDR2 shows clearly consensus patterns that differ between vertebrates, insects, fungi, and algae. Size variation and occurrence of primitive repeats in IDR2 are considerable.

RNA surveillance occurs in stress granules, and fat storage occurs in lipid droplets, both of which exist in liquid-liquid phase separation (LLPS). Thus, the quite variable sequences with polyampholytic intrinsically disordered structure [215] immediately from LSmAD to the C-terminus (extending up to 1200 amino acids in orthologs, like XP_061722979.1 in *Cydia pomonella*, where an unusual positive polyelectrolyte repeat follows the LSmAD sequence, or KAJ8263487.1 in *Conger conger*, or KAJ1504482.1 in *Coelomomyces lativittatus*), may play a role in localizing Ataxin-2 in the vicinity of these cytoplasmic compartments.

The fact that the human ATXN2 gene has the alternatively spliced exons ante-10 and 10, at the end of the LSmAD cluster of exons, represents more evidence that this ATXN2 region has importance and is regulated in dependence on functional context, like stress.

2.13. PAM2 Motif in Many Protists and Insects Lies Close to Long polyQ Stretches

PAM2, as a short linear motif, is immersed among sequences with intrinsically disordered structures (Table S1h), at about 3-4 fifths of the total protein length, in contrast to LSm and LSmAD, where InterPro pLDDT scores predict high alpha- (or beta-) fold confidence. Previous literature has named the intrinsically disordered sequence until PAM2 as IDR2, and the intrinsically disordered sequence after PAM2 as IDR3 [62]. PAM2 is crucial for the ability of ATXN2 to be incorporated into stress granules [62,216]. PAM2 sequences changed considerably during evolution, so the InterPro algorithms often do not detect them, particularly in lower species; therefore, such candidate PAM2 motifs were compiled in Table S1h. It was already reported for CID3/CID4 in land plants [214] that a part can be repeated in tandem (sekstLNPNAKEFk-LNPNAKSFTps), but among excavates, amoebzoa, and algae, there is only an imperfect core sequence with considerable variability for Ataxin-2 family members (pkssksKLNPNAPkFsLNinAKeF, see Table S1h).

Within the IDRs on each side of the PAM2 motif, polyQ repeat sequences are frequent from protists until insects, reaching up to 30Q without interruption (downstream from PAM2, for A0A2V3IVX2 in the rhodophyte *Gracilariopsis chorda*, or downstream from LSmAD for Q55DE7 in the amoebozoan *Dictyostelium discoideum*), at the extreme size of 73Q with 3 Met interruptions (A0A813ABV5 in *Symbiodinium necroappetens*), or later with interruptions by Pro residues (F4X2F6 in *Acromyrmex echinator*, A0A836FCM7 in *Acromyrmex insinuator*, A0A158NW33 in *Atta cephalotes*, A0A8J2HMF1 in *Cotesia congregata*, A0A834JR87 in *Vespula vulgaris*, A0AA40FF43 in *Melipona bicolor*). The 73Q domain in *Symbiodinium necroappetens* is translated from nucleotides (Met coding triplets in bold)

CAG(CAA)₂(CAG)₂CAA**ATG**(CAG)₂CAA(CAG)₂CAA(CAG)₂CAA(CAG)₂CAA(CAG)₄CAA(CAG)₂CAA(CAG)₆CAA(CAG)₃CAA**ATG**(CAG)₂CAA(CAG)₂(CAA)₂(CAG)₂CAA(CAG)₂CAA(CAG)₄(CAA)₂(CAG)₃(CAA)₃(CAG)₃CAA(CAG)₂CAA(CAG)₃**ATG**CAG, while the 30Q domain in *Gracilariopsis chorda* red algae gets translated from nucleotides (CAA)₃CAG(CAA)₂(CAG)₂(CAA)₂(CAG)₂(CAA)₅(CAG)₆CAA(CAG)₂(CAA)₄, and the 29Q domain in slime mold *Dictyostelium discoideum* is translated from nucleotides (CAA)₂₉. An uninterrupted 45Q domain in *Condyllostylus longicornis* is translated from (CAA)₂CAG(CAA)₇CAG(CAA)₈(CAG)₃(CAA)₄CAG(CAA)₅CAG(CAA)₃CAGCAA-(CAG)₃(CAA)₄,

see NCBI entries XP_055375020.1 and XM_055519045.1. Thus, there seems to be neither a codon preference nor a likely repeat structure underlying polyQ domains.

Downstream from LSmAD, the intrinsically disordered region 2 (IDR2) of human ATXN2 starts in the sequences encoded by exon 8 and extends across PAM2 into exon 17. The PAM2 motif in human ATXN2/ATXN2L is encoded by exons 16/15, respectively, which are genomically clustered with one preceding and two subsequent exons, suggesting joint regulation and function in this area. The subsequent constitutive *ATXN2* exons 22/23 still encode intrinsically disordered sequences (IDR3). Within the *ATXN2* exons 16 and 17, several methionine residues exist as candidate homologs of the yeast C-terminal methionine-rich region where metabolic redox sensing occurs to control phase separation of Ataxin-2 [217–219], but none of them is conserved in ATXN2L or in the single-copy Ataxin-2 orthologs, although the two paralogs and these orthologs show similar condensation dynamics.

2.14. Sequence Beyond LSmAD Without an Intrinsically Disordered Structure

In the C-terminal half of Ataxin-2 orthologs, the intrinsically disordered sequences are usually interrupted at least once, in primates up to 130 amino acids beyond the PAM2 motif, for some 40-200 residues (around position 500-880 for the typical Ataxin-2 ortholog of 1000 amino acid length) (Table S1i). For human ATXN2, this interrupting ordered sequence is encoded by exons 18/19/20, for human ATXN2L by exons 17/18/19. Interestingly, for human ATXN2, the alternatively spliced exon 18B evolved inside the start of this cluster, while the alternatively spliced exon 21 evolved at the end of this cluster, both encoding ordered sequences. This regulation effort is compatible with the notion that the functional relevance of this ordered area has to be modulated upon changes of stress versus stimulus.

2.15. C-Terminus Alternatives

The C-terminal end of Ataxin-2 orthologs can contain only an intrinsically disordered sequence, or first an intrinsically disordered part and then an ordered terminal end (Table S1i). Much alternative splicing is documented for human ATXN2 at the C-terminus, with constitutive strongly expressed exons 22/23 encoding intrinsically disordered ATXN2L-homologous sequences where some isoforms stop (e.g. Q99700-4). Furthermore, the alternatively spliced exons 24A/B encode ordered ATXN2L-homologous sequences including their joint stop (Suppl. Material S5). In ATXN2 exon 24A, there is another methionine-rich sequence as a potential similarity with the yeast PBP1-C-terminal region that acts as a sensor of metabolic redox status to control the phase separation of Ataxin-2 orthologs [217–219], but in the human species, these residues are situated in a structured region, and only one of these methionine residues is conserved in human ATXN2L. The translation products of the alternatively spliced exon 23B (encoding a sequence known since primates) and alternatively spliced exon 25 (encoding a Q(P/A/T)(H)₁₋₃Q(P/Q)(P/Q)L sequence before an alternative stop, followed by a long non-coding 3' untranslated region) are predicted to have ordered structure as well, mediating the modification of the C-terminus in periods of growth versus stress.

This C-terminal sequence, encoded by exons 22-24A in human ATXN2, shows the maximum expression among all regions of Ataxin-2 according to GTEx RNAseq quantifications in nervous tissue, as mentioned above. It shows clearly conserved residues from humans until early chordates (Table S1i). Regarding its function, it was shown that the C-terminal IDR enhances Ataxin-2 target mRNA interaction, influences the long-term habituation, ribonucleoprotein (RNP) granule formation, and neurodegeneration in *Drosophila melanogaster* retina [210,213,220]. Therefore, the maximal expression levels of the C-terminal exons may serve to encode Ataxin-2 isoforms that are crucial for RNP granule formation and motility during neurite transport, while the lower production of longer isoforms, including LSm-LSmAD, might serve the transition from RNP transport to active translation in stimulated synapses [221]. It may also be relevant that this C-terminus has the highest sequence homology (28.7%, $p = 3.5e-7$) in chordates with entry A0A8C4PYQ0, a regulator of trafficking from ER to Golgi with predicted localization to ER exit sites, thus being assumed to

influence the production of phospholipids and sterols, together with protein biosynthesis and trafficking, as well as calcium homeostasis.

The definition of highly conserved residues in this C-terminal sequence among insects (Table S1i) was employed to search for related proteins among nematodes and cnidaria, finding high homology with entry A7S107 (30.9%, $p = 3.8e-13$), again a regulator of trafficking from ER to Golgi with predicted localization to ER exit sites that is recognized as member of the SEC31 family [222,223].

2.16. Ataxin-2 Forms Chimeric Proteins with Very-Long-Chain Fatty Acid Synthases in Various Algae, Suggesting an Evolutionary Link and a Potential Role in Wax Biosynthesis in Plants

Several other motifs of known function were integrated with the core features of Ataxin-2 orthologs in chimeric proteins of several protozoa, algae, and plant species (Table S4). Given that plants have no mobility to run from danger and damage, stress response pathways are more prominent and more streamlined in plants than in animals. These additional domains may represent meaningful efforts to optimize the efficiency of Ataxin-2 within its biochemical context, given that the analysis of mice with mutant ATXN2 also documented prominent dysregulations in these processes. The data suggest Ataxin-2, together with co-translated neighbor genes, have a role in membrane stress responses, with strong conservation across evolution from algae to mammals.

Firstly, always a very-long-chain 3-ketoacyl-CoA synthase domain, and variably a chalcone/stilbene synthase domain, and a 3-oxoacyl-[acyl-carrier-protein] synthase were added to Ataxin-2 in *Chlorella sorokiniana* (A0A2P6U097), *Chlorella ohadii* (A0AAD5DQN1), *Auxenochlorella protothecoides* (V5NG01), and *Micractinium conductrix* (A0A2P6V2R2). The very-long-chain 3-ketoacyl-CoA synthases (abbreviated as KCS), also known as very-long-chain fatty acid (VLCFA) condensing enzymes, play a crucial role together with very-long-chain fatty acid elongases in the adaptation of lipid membranes to abiotic stress, prominently by the biosynthesis of (epi)cuticular wax [224–226].

Secondly, additional presence of a BRX domain is documented in *Lactuca saligna* (UniProt entries A0AA36E8E6 and A0AA36E8L8). BRX maintains brassinosteroid biosynthesis above a critical threshold [227].

Thirdly, a lecithin:cholesterol/phospholipid:diacylglycerol acyltransferase (LACT/PDAT) domain is present at the N-terminus of the Ataxin-2 ortholog in *Populus alba* (A0A4U5R0B4). Plant PDAT enzymes are involved in seed lipid droplet reserves by triacylglycerol formation via an acyl-CoA-independent pathway, and in membrane lipid remodeling [228,229]. Particularly in unicellular microalgae under stress conditions, the accumulation of triacylglycerol is accompanied by gradual chloroplast degradation with re-organization of membrane structures and remodeling of the lipidome [230].

Fourth, a formate/glycerate dehydrogenase catalytic domain with homology to the *Escherichia coli* SerA protein (crucial for serine biosynthesis) is added at the N-terminus of Ataxin-2 in *Symbiodinium necroappetens* (A0A813ABV5). L-serine is a precursor to sphingolipids and phospholipids [231], and formate/glycerate dehydrogenases have a role in the stress responses of plants (*Symbiodinium* carries a chloroplast) [232,233].

2.17. Ataxin-2 Is Chimeric with Oxysterol-Binding Proteins in Various Hexapods and with Myelin Biosynthesis Factors in Animals

This evidence from primitive plants is in good agreement with the integration of Ataxin-2 sequences into chimeric proteins among animals (Table S5).

In the arthropod-hexapod class of collembola, the Ataxin-2 domains are joined to oxysterol-binding protein sequences in *Folsomia candida* (A0A226E5R7), *Archesella cincta* (A0A1D2NII5), and *Allacma fusca* (A0A8J2NQ72, A0A8J2J4F7, A0A8J2NIC3). The role of sterols for plasma membrane dynamics is well known; oxysterols have an additional signaling function [234,235].

Furthermore, the addition of an N-terminal apolipoprotein-III precursor domain to Ataxin-2 occurred in the wax-producing honey bee *Apis cerana* (A0A2A3ESF8) and in the ant *Temnothorax longispinosus* (A0A4S2KN76). In insects, apolipoprotein-III assists in loading diacylglycerol, generated

from triacylglycerol stores in fat bodies via the adipokinetic hormone, onto hemolymph lipophorin, and it also modifies the stress responses to bacteria [236,237].

A C-terminal addition of a Ser/Thr hydratase domain to Ataxin-2 is documented in *Paralvoinea palmiformis* worms (A0AAD9K567). Such L-serine ammonia-lyases are central to the production of acetyl-CoA and lipids in other worms [238] and mediate stress responses against acid already in bacteria [239].

In *Hermetia illucens* flies (A0A7R8YXR3), a C-terminal C2HC-type zinc-finger transcription factor with homology to myelin transcription factor-1 (MYT1) is appended onto the Ataxin-2 ortholog. Here, it is important to know that animal myelin, sphingolipids, ceramides, and tear film meibum, similar to plant wax (in the Latin language, “cera” means wax), are produced from very-long-chain fatty acids [240]. MYT1 regulates, for example, the expression of the myelin protein PLP1 [241,242].

The *Cloeon dipterum* mayfly Ataxin-2 ortholog (A0A8S1C5C0, A0A8S1C908, A0A8S1C9E1) shows an N-terminal addition of a Ser/Thr protein kinase domain with homology to CAMKK. Human CAMKK controls adiposity and cholesterol homeostasis [243–245].

Among bony fishes, *Clarias magur* Ataxin-2 (KAF5904804.1) shows an N-terminal RALDH1 retinol dehydrogenase domain combined with a MAPKAP5 domain. The human RALDH1 ortholog ALDH1A1 protects against lipid peroxidation byproducts such as malonaldehyde, while MAPKAP5 (also known as PRAK or MK5) controls cytosolic phospholipase A2 (cPLA2) and mTORC1 among other substrates, protecting against oxidative stress in Parkinson’s disease [246–249]. With special relevance for SCA2, RALDH enzymes govern the metabolism of retinoic acid, its signaling impact, and therefore hindbrain differentiation [112]. Human *ALDH2*, encoding a major biosynthesis enzyme for retinoic acid in liver mitochondria [250], and *MAPKAP5* are neighbor genes of *ATXN2* on chromosome 12q24.

An Ataxin-2 ortholog in the naked mole rat *Heterocephalus glaber* (G5AKU3) shows C-terminal addition of a sesquipedalian family pleckstrin homology domain. Members of this family, such as yeast Boi1/Boi2 and human FAPP1, are required for the fusion of secretory vesicles with the plasma membrane at sites of polarised growth, acting by modulation of phosphatidylinositol 4-phosphate (PtdIns4P) levels [251,252]. Again, a sesquipedalian family member is a neighbor gene of *ATXN2*.

Overall, Ataxin-2 orthologs were combined with other sensors and effectors of membrane stress responses into longer multi-domain proteins, both in plants and animals.

2.18. Ataxin-2 Is Chimeric with an AMP Kinase Subunit and with a SCYL1 Ortholog in Some Fungi, Where Its Mutations May Impact Metabolic Reserves and Lipid Homeostasis

In fungi, domain additions to Ataxin-2 orthologs include (Table S6):

- An N-terminal KA1/Ssp2_C domain in *Neophaeococcomyces mojaviensis* (NCBI Genbank entry KAJ9652119.1). This C-terminal domain in the serine/threonine protein kinase AMPK catalytic (alpha) subunit Ssp2 acts as a sensor to regulate cellular responses to a variety of nutritional and environmental stresses [253].
- A longin-like domain exists at the C-terminus in *Saitoella complicata* fungus (UniProt entry A0A0E9NFC8). These sequences are involved in membrane dynamics regulation, occurring e.g., in GDP/GTP exchange factors for Rab7 / Rab32 / Rab38, such as the MON1/CCZ1-MC1 complex or the HPS1/HPS4-BLOC3 complex [254].
- An ARM-like fold together with a CIK-related domain is present at the C-terminus in *Aspergillus felis* fungus (A0A8H6QL38). This combination of armadillo-repeats in CIK- (Catalytically Inactive Kinase) related proteins is preserved in human SCYL1-3, which are involved in the regulation of traffic between the Golgi apparatus and the ER, or play a role between the trans-Golgi network and the endosomal system, but also contribute to tRNA export from the nucleus [255,256]. This observation seems relevant since SCYL1 mutations cause progressive spinocerebellar neurodegeneration in mouse and human (SCAR21 type), and motor neuron disease with mislocalization of TDP-43 in mouse, similar to ALS and to *ATXN2* mutation effects in humans [257–262]. SCYL1 mediates mTORC1 signals [263].

- An N-terminal addition of a phospholipid transporter P-type ATPase occurred in *Physocladia obscura* fungus (A0AAD5T663 and A0AAD5XG15).
- A C-terminal addition of a SEC6 domain is documented in *Candida oxycetoniae* fungus (A0AAI9WVV1) and in *Candidozyma haemuli* fungus (A0A2V1AQ46). The exocyst component SEC6 interacts with oxysterol-binding protein to regulate polarized vesicular membrane transport and growth in a sterol-dependent manner [264].
- Further, an N-terminal cytochrome c oxidase, subunit Va/VI domain exists in *Racocetra fulgida* (A0A9N9AHJ9). This observation suggests that Ataxin-2 can act on the intra-mitochondrial pathway of oxidative stress generation directly.
- A Calcineurin-like phosphoesterase domain was added in *Diversispora eburnea* (A0A9N8VJW7). Such a domain occurs, e.g. in nucleotidases and sphingomyelin phosphodiesterases [265].
- A Cysteine desulfurase was appended to the N-terminus, and at the same time, a cell cycle-regulated microtubule-associated protein (TPX2) domain was added at the C-terminus in *Bifiguratus adalaidae* (A0A261XY39). The mitochondrial cysteine desulfurase NifS / IscS / NFS1 acts in the core iron-sulfur cluster (ISC) assembly complex, where perturbations lead to the human disorder Friedreich's ataxia [266,267]. Cellular iron-sulfur (Fe/S) proteins are involved in lipid synthesis and steroid metabolism [268]. The TPX2 protein is an assembly factor for spindles, downstream from RanGTP [269].

Thus, Ataxin-2 orthologs (in diverse species from protists and algae to higher plants, fungi, and animals) are integrated into longer chimeric proteins that enable streamlined adaptation of lipids in membranes and fat bodies upon stress exposure, together with the production of (epi)cuticular wax or myelin from very-long-chain fatty acids, and the adaptation of chloroplast / mitochondrial homeostasis. This may suggest that mutations in Ataxin-2 orthologs play a role not only in RNA quality control but also in lipid adjustments after cell damage. Since LSmAD sequences are always adjacent to LSm domains and seem to function jointly with them, it is worth considering if the sensor of oxidative stress could be the LSm via binding to misfolded or oxidized RNA, while the Ataxin-2-specific effector of stress control could be LSmAD via modifications of lipid homeostasis.

3. Discussion

Current knowledge of the Ataxin-2 protein family is focused on the full-length isoforms. In particular, the roles of the LSm domain and the PAM2 motif and their impact on RNA processing have been investigated, although they are not unique to Ataxin-2. In contrast, very little knowledge exists about Ataxin-2 small isoforms and about the LSmAD component, its interactors, together with its functions. This focus has persisted, although small isoforms are more strongly expressed than the full-length isoform, and although LSmAD exists only in the context of Ataxin-2, constituting its defining feature. The fact that Ataxin-2 family members frequently show chimerisms with added domains at either terminus of the protein was previously unknown.

Here, this manuscript represents a datamining effort to analyze the evolution of Ataxin-2 structural components, their expression regulation, and the functional impact of their mutations. For a systematic evaluation, every effort was made to identify a maximum number of orthologs of Ataxin-2 in the various kingdoms of life, and to compile them in Table S1a. Overall, we hope to derive functional insights from lower organisms that can be applied in patients, where Ataxin-2 can be exploited as a crucial switch to mitigate the progression of neurodegenerative diseases [21].

Altogether, the data are compatible with a novel scenario where mutations in the Ataxin-2 protein orthologs since algae affect not so much mRNA, but mainly rRNA serving the biosynthesis of secreted and membrane proteins, and strongly impact membrane resilience in periods of cell stress (main datamining and experimental evidence is summarized in Figure 6). We speculate that strong expression of small isoforms of Ataxin-2 containing C-terminal IDR2 and IDR3 may alleviate the deleterious consequences of mutations in the N-terminal IDR1, so that human exon 1B seems the preferable target for knockdown therapies to prevent neurodegenerative diseases.

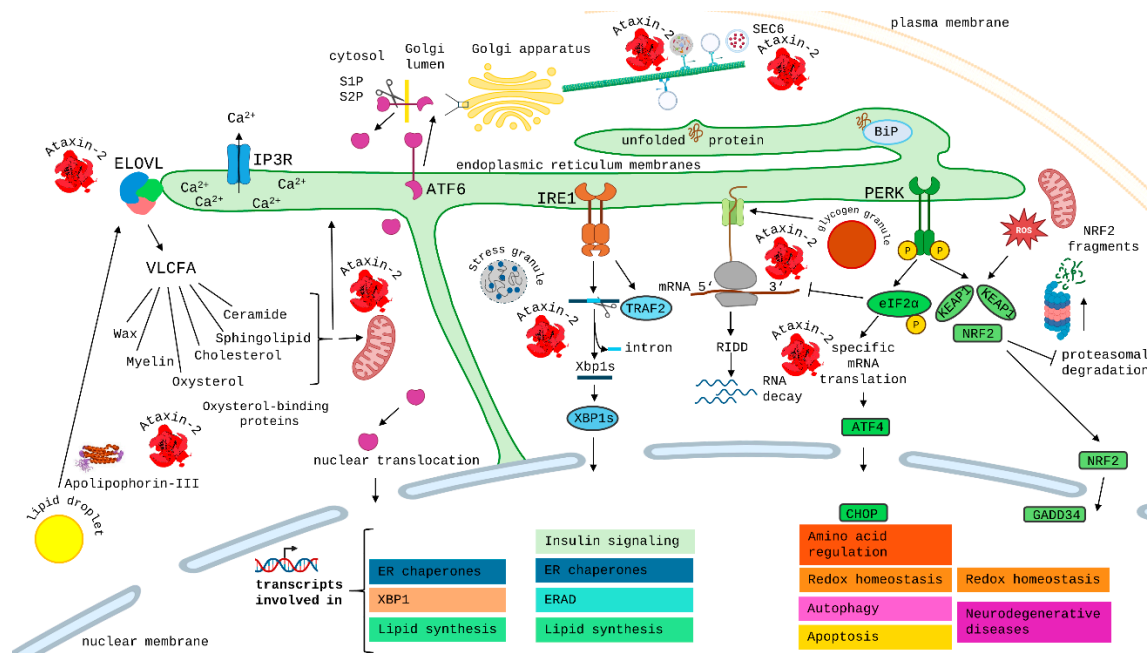


Figure 6. Synopsis of datamining and experimental evidence about mutation effects of Ataxin-2, confirming its relevance for rRNA and mRNA processing, and emphasizing the impact of the Ataxin-2 locus on membrane strengthening during stress periods. Fusion of additional domains onto Ataxin-2 with conservation across multiple species was documented with Apolipoprotein-III, Oxysterol-binding proteins, Elongases (ELOVL) for very-long-chain fatty acids (VLCFA), and SEC6 proteins as exocyst components. Further fusion events clarify its integration into RNA surveillance and translation modulation pathways, in stress granules, ribonucleoprotein granules during microtubular transport, XBP1 isoform splicing, regulated IRE1-dependent RNA decay (RIDD), and specific mRNA translation. A little-expressed isoform of Ataxin-2 with a mitochondrial targeting peptide underlines its function for endosymbiont management. Previous experimental data independently showed Ataxin-2 mutations to impact lipid droplets, VLCFA elongation, plant wax, oligodendrocyte myelin, oxysterols and cholesterol, sphingolipids and ceramide, subcutaneous fat, endoplasmic reticulum calcium homeostasis, chaperones such as BiP, insulin signaling and glycogen granules, amino acid regulation, autophagy, redox homeostasis, Golgi fragmentation, ribonucleoprotein trafficking in neurites, and exo-/endocytosis of vesicles. Although Ataxin-2 mutations therefore modulate multiple pathways in ER stress, it remains unclear if Ataxin-2 acts as a selective or generalized modifier of the three main processes in unfolded protein responses, which are under the control of ATF6 / IRE1 / PERK. Created in BioRender. Key, J. (2026) <https://BioRender.com/e9aey6>.

3.1. Ataxin-2 Chimerisms, Probably Due to Transcriptional Readthrough, Include Domains Enriched in Functions for rRNA Processing and Membrane Resilience

As a completely novel finding, large chimeric forms of Ataxin-2 with extra domains added to the N- or C-terminus were found to occur in approximately 5% of lower species, so this observation will be discussed first, in great detail. The lists of such domains added to Ataxin-2 in diverse species contain an enrichment of rRNA processing enzymes. If the chimerisms do not represent random errors of organismal protein synthesis or technical artifacts, then Ataxin-2 might contribute to defending and adapting the translation apparatus rRNAs in periods of cell stress, and to protect all vulnerable mRNAs with AU-rich elements and rapid turnover from oxidative damage, rather than acting on specific mRNAs. LSm and PAM2 perhaps make contact with RNA in an effort to sense oxidative damage, and only then act as modifiers of translation and repair. Under such oxidative stress, lipid peroxidation and membrane damage are bound to occur in parallel, and the associated LSmAD and IDR sequences may become crucial to protect the cell. A crucial question regarding LSmAD as a characteristic Ataxin-2 feature is whether it functions mainly in an auxiliary manner to modify the RNA-binding of the adjacent Lsm domain, RNA stability, and translational control – or if

it modulates membrane interactions and composition upon recognition of oxidized RNA by the LSm domain.

This novel concept takes into account the ability of eukaryotic organisms to create multi-domain proteins in order to maximize the efficiency within a metabolic pathway. Given that the order of genes within chromosomes and within genomes is dynamically reshuffled across evolution, there is a continuous sorting process if neighbor genes are well matched or should be assembled differently. It was recently shown that transcriptional readthrough across neighbor genes is frequent under stress conditions for GC-rich genes (such as Ataxin-2) [270,271], in particular upon calcium release from ER stores [272]. Also, healthy cells show this feature in genomic regions with high transcriptional activity [273,274]. Although most genes in a genome appear to have no functional connection to their neighbors, some genes indeed are clustered with functional interactors within the same locus (HLA genes, as a famous example). It is known in humans that the chromosomal locus of Ataxin-2 is part of an exceptionally large linkage disequilibrium block across 1 megabase [117,275], so there is a selective advantage in these neighbor genes to stay together. It is unclear according to what principles the positioning of new genes or duplicate gene-copies into a meaningful genomic context occurs, and how chromosomal rearrangements are adapted over time. Given that the genomic surrounding of Ataxin-2 has changed over time, we cannot know which among the observed chimeric proteins has a physiological added value. However, it is conceivable from the data above that mutations in Ataxin-2 during cellular stress may also affect the transcription of neighbor genes and the translation of their domains.

Of course, some annotated proteins in databases are based on preliminary documentation, and some might be artifacts due to technical problems. The limitations of our study mainly center on our inability to distinguish artificial protein fragmentation and chimerism from physiological short isoforms and extra-long readthrough multi-domain proteins, respectively. In the following, examples are provided that support transcriptional slippage rather than ribosomal slippage [276] to underlie the observed chimerisms: In the 2577 residues of the UniProt entry A0AAF0YF69 from *Vanrija pseudolonga* fungus, in the 2585 residues of A0AA48I7S8 from *Cutaneotrichosporon cavernicola* fungus, and in the 2553 residues of A0AAD3TXG0 from *Cutaneotrichosporon spelunceanum* fungus, Ataxin-2 orthologous sequences have an added cytidine deaminase domain together with immunoglobulin domains that are still enclosed within a chromosome 7 whole genome shotgun sequence, according to InterPro analysis. Such recombined fusion proteins could arise, e.g. during cDNA synthesis via erroneous annealing or may be the product of transcriptional readthrough, where intron retention due to cell stress is a known crucial feature [270,277].

Chimerisms of Ataxin-2 were frequently found among lower organisms, where cell stress is more frequent than in highly developed organisms. Interestingly, a chimerism is documented in the ATXN2 ortholog of the *Pituophis catenifer* snake (NCBI protein entry XP_070807217.1, see Table S3), where genomic information is already available. The N-terminal addition of an extra-domain with BRAP2 can indeed be explained as transcriptional readthrough, given that *BRAP2* is the closest neighbor gene upstream from *ATXN2* according to the UCSC genome browser from humans to *Alligator* reptiles and *Xenopus* amphibians, with the same direction of translation. Given that practically all genomes are unknown for the species with extra-domains added to Ataxin-2, it is probable that many or all the above chimerisms represent transcriptional readthrough events. Thus, the generation of readthrough transcripts and chimeric proteins appears to be a relevant feature of Ataxin-2, which gets activated in periods of cell stress.

Chimerism of ATXN2 among highly developed organisms, where the genome sequence is known for related species, is also documented in the *Heterocephalus glaber* naked mole rat (entry G5AKU3, see Table S5). A sesquipedalian family domain was added to its C-terminus. Again, this can be explained as transcriptional readthrough, given that the Sesquipedalian-1 protein is encoded by the *PHETA1* gene, the immediate neighbor of the *ATXN2* gene at its 3' end, with the same transcriptional orientation, as a genomic position conserved until human.

Further examples include the C-terminal addition of a RIO2 kinase domain to the ATXN2 ortholog of the bony fish *Anabarilius graham* (UniProt ROL51429.1), and indeed the UCSC genome browser of the bony fish *Danio rerio* shows the *RIOK2* gene to be localized at the 3' end of the *ATXN2* orthologous gene in the same transcriptional direction. A final example is the N-terminal addition of a RALDH1 retinol dehydrogenase domain and a MAPKAP5 domain to the ATXN2 ortholog in the bony fish *Clarias magur* (UniProt entry KAF5904804.1), and again the UCSC genome browser of the bony fish *Latimeria chalumnae* confirms the position of the *ALDH1A1* and *MAPKAPK5* genes at the 5' end of the *ATXN2* gene, but interestingly in the opposite transcriptional direction. Overall, these examples from species with known genome sequence indicate that the extra domains added onto Ataxin-2 are transcribed and translated from neighbor genes as a general rule.

Of course, it is important to be cautious and to ignore artifacts during datamining efforts, but further formal arguments and a host of experimental evidence led us to consider the above chimerisms as relevant, or at least worthwhile to be assessed further.

Firstly, technical artifacts should be distributed randomly across databases, but the Ataxin-2 orthologs with additional domains are frequent among low organisms but very rare among higher animals and plants.

Secondly, artifacts should implicate all pathways at random in Ataxin-2 biology, but the additional domains are preferentially involved in stress responses, rRNA processing, the mobilization of fat/glycogen stores, and the modulation of lipid membranes.

Thirdly, although aberrant fusion proteins may occur in nature as pathological events that cause pathology, e.g., cancer, their conservation across multiple related species points to a selective advantage of this fusion event in that particular ecological niche.

Fourth, previous experimental characterization of Ataxin-2 mutant organisms has revealed only small selective rather than striking or broad changes of RNA repair or mRNA translation in normal tissues and growing cells (in summary of many transcriptome profiling efforts from humans to mice and flies, [55,56,194,221,278–291]). In contrast, strong anomalies of lipid (and carbohydrate) stores and membrane homeostasis were readily demonstrable upon Ataxin-2 mutation not only in lower organisms such as flies, but in a conserved manner until mammals. Thus, a host of previously reported experimental evidence supports the concept that Ataxin-2 mutations affect membrane composition.

Irrespective of whether the extra-domains reflect only transcriptional readthrough events that are relevant selectively for stressed cells, or are part of a context of physiological functions of Ataxin-2, the pathways surrounding such extra-functions are affected by Ataxin-2 mutations as a consequence of the existence of chimeric proteins. So it seems relevant in the following paragraph to provide a summary of current experimental evidence that implicates Ataxin-2 mutations in membrane stress.

3.2. Experimental Confirmation that Mutations in Ataxin-2 Alter Membrane Resilience

In particular, the absence of the Ataxin-2 ortholog PBP1 in *Saccharomyces cerevisiae* yeast caused downregulation of several mitochondrial metabolism enzymes, namely citrate synthase, alpha-ketoglutarate dehydrogenase, glyoxylate reductase, succinyl-CoA ligase, and ATP synthase component ATP1 [138]. PBP1 is integrated in the phosphorylation cascade between AMPK and MTORC1 [292], sensing metabolic redox status via reversible oxidation of multiple methionines that control its condensation state [217–219]. Given that these methionine switches are not conserved during evolution, it is conceivable that oxidation damage is also sensed via lipid peroxidation or via oxidative RNA damage.

- In *Caenorhabditis elegans*, the Ataxin-2 ortholog regulates the fat content and size of cells, acting via MTORC1 signaling [293,294].
- In *Drosophila melanogaster* flies, Ataxin-2 accumulates in fat bodies together with Sec23 as a marker of ER exit sites [144]. The depletion of fly Ataxin-2 mitigates glycogen metabolism problems similar to mTORC1 repression rapamycin treatment [295].

- The absence of ATXN2 from *Mus musculus* tissue results in massive accumulation of lipid droplets and glycogen, together with reduced sterol degradation in cerebellum and elevated blood cholesterol [134]. Proteome profiling of these ATXN2-null mouse livers demonstrated significant deficits of enzymes in the fatty acid beta-oxidation and malonyl-CoA / methylmalonyl-CoA pathways [296]. Again, according to unbiased global proteome and metabolome profiling efforts, the ATXN2 polyQ expansion in mouse cerebellum has indeed its main impact on the very-long-chain fatty acid elongases such as ELOVL4 (residing at the ER in a multi-protein complex), the very-long-chain fatty acids VLCFA24-26 with precursors such as acetyl-CoA and N-acetylaspartate [130], and their derivative sphingolipids and ceramides [145]. This is accompanied by changes in inositol-tris-phosphate metabolism and calcium/calmodulin-dependent kinases [297]. Receptors for the inositol 1,4,5-trisphosphate lipids were also implicated in an SCA2 mouse model [298]. In the spinal cords from two different SCA2 mouse models with ATXN2 polyQ expansion, the suppression of enzymes for cholesterol biosynthesis from acetyl-CoA and squalene, together with cholesterol and oxysterol deficits, was the main finding [147,287,299]. The restoration of brain cholesterol turnover was reported to have therapeutic value in a SCA2 mouse model [300].
- For *Homo sapiens*, the brains of SCA2 patients show a reduction of C22/24-sphingomyelin and cholesterol levels [145], and the myelinated white matter of the brain is deficient [24,301–307], already at presymptomatic disease stages [31,308,309]. The subcutaneous fat deposits, the cheek fat body, the visceral fat, and the body weight of SCA2 patients decrease progressively, as well as the levels of testosterone as a cholesterol-derivative, in peripheral tissues [310–312]. In an epidemiological genome-wide association study of a Japanese population, variants of ATXN2 were found to underlie the susceptibility for dyslipidemia [313].

3.3. What Is the Role of Ataxin-2 for the Unfolded Protein Response and Retinoic Acid Signaling?

In bacteria, the RNA chaperone Hfq, as an ancestor of the LSm proteins, responds to envelope stress via the sigma transcription factor for envelope stress, RpoS. This sigma factor is regulated either at the post-transcriptional level via Hfq or at the proteolysis level via CLPP/CLPX-mediated unfolded protein responses [178]. Membrane stress responses via sigma transcription factors were not conserved in eukaryotes. Instead, stress granules and the Unfolded Protein Response (UPR) with the Sigma-1 Receptor (S1R) at the mitochondria-associated ER membrane (MAM) in interaction with the inositol 1,4,5-trisphosphate receptor (IP3R) and Inositol-Requiring Endoribonuclease 1 (IRE1) mediate the appropriate responses for ER stress in mammals [314–316]. For the UPR of bacterial and mitochondrial membranes, CLPP is crucial [317–323], and mutations in CLPP result in progressive ataxia and leukodystrophy in patients and mice [319,324,325].

The elongation to synthesize very-long-chain fatty acids occurs at the ER, and is the prerequisite for the biosynthesis of (glyco-)sphingolipids, ceramides and myelin / wax, when membranes need to be strengthened [326,327]. To integrate proteins into plasma membranes / the myelin sheath / the (epi)cuticular wax, their biosynthesis under stress conditions may require internal ribosomal entry site (IRES)-dependent translation from appropriately spliced mRNAs, then post-translational modification like glycosylation in the ER lumen, trafficking via the Golgi apparatus along microtubules to their destiny in a polarized cell, and finally their turn-over by the ER associated degradation (ERAD). Ataxin-2 localizes to all these compartments and may have evolved to modulate all these processes during stress periods, including the recruitment of lipids from cytoplasmic fat droplets and carbohydrates from glycogen granules [134], for the modification of membrane composition and complexity.

Ataxin-2 impacts not only ER membranes but also mitochondria. It was shown to control translation preferentially for mitochondrial factors [135], to modulate PINK1 levels [70,328], interact with PARK2 and sequester it into insoluble aggregates [68,69]. Thus, the crucial cytoplasmic factors that govern mitochondrial membranes regarding stress signaling, damage repair, and autophagic elimination are within the network of Ataxin-2 functions.

Retinoic-acid-triggered cell differentiation of mammals involves ER stress UPR via XBP1 [329–333]. IRE1 selectively acts on *XBP1* transcript, excising 26 nucleotides [334–336]. This generates an *XBP1* splice variant, which encodes a transcriptional activator to promote UPR target gene expression [334–336]. Very recently, *ATXN2* was also found to selectively target *XBP1* mRNA, stabilizing it and suppressing its translation, until Ataxin-2 is ubiquitinated by FBXO42 in a SKP-A/CULLIN-1 complex and gets degraded [291]. XBP1 is not only the central transcription activator in ER stress pathways via UPR, with an important role for neurodegenerative disorders [337–340], but is also crucial in physiology for secretory cell signaling to regulate lipid and insulin metabolism, as well as autophagy [341–343]. IRE1 and XBP1 influence the expression of MYC/MYCN and receive feedback control from them, within the pathway where retinoic acid governs differentiation, e.g., of neuroblastoma cells [344–349], with Ataxin-2 levels previously shown to depend on MYCN within this pathway [350]. During severe ER stress, regulated IRE1-dependent decay (RIDD) of mRNAs (to lessen ER load by reducing translation and post-translational ER translocation and secretion of proteins) generates single-strand mRNA fragments that lack 5'-caps or 3'-poly(A) tails, thus activating retinoic inducible gene-I (RIG-I) protein and mitochondria-dependent innate immunity [351]. The LSM domain of Ataxin-2 might recognize such RNA fragments, and Ataxin-2 might contribute to RIDD activity.

In mammals, the mutations in both *ATXN2* and *CLPP* cause ataxia phenotypes [1,21,324,325]. *ATXN2* mutations have a profound effect on IP3R1 expression [297,352], and mutations in mammalian IP3R1 also cause a phenotype of ataxia [353]. The relevance of the IRE1-XBP1 pathway among ER stress responses is well studied in ALS, but was also reported for Ataxia-Telangiectasia and Friedreich's Ataxia [354–361]. A crucial role of Ataxin-2 orthologs for vitamin A derivatives, cholesterol, and retinoid homeostasis (as modulators of membrane strength versus stress) was evident in lower organisms, and appears conserved until mammals, given that *Atxn2*-KO mice show altered retinoic acid signaling [352]. The receptors for retinoic acid and related compounds, RAR-RXR-RORalpha, are in close interaction [362–366]. RORalpha protects neurons against oxidative stress [367,368]. Abnormal retinoic acid signaling was reported in several forms of ataxia, namely SCA1 [369,370], SCA3 [371], SCA7 [372,373], FRDA [374], A-T [375], and the staggerer mouse mutant [376–379]. Several of these ataxias show disease course modification by Ataxin-2.

3.4. Are Chimerisms, Transcriptional Readthrough, and Neighbor Genes Relevant for Ataxin-2-like?

It remains unclear to what degree the roles and regulation of *ATXN2*, which remained in the ancient genomic context of Ataxin-2, are still conserved for *ATXN2L* after gene duplication into a different genomic context. Within its locus on human chromosome 16p, the same transcriptional orientation is shared among neighbor genes *EIF3C* – *NPIP9* – *ATXN2L* – *SH2B1* – *ATP2A1*. Chimeric proteins containing *ATXN2L* with extra domains from its neighbor proteins were not observed in our study, although *ATXN2L* expression is also activated in human neuroblastoma cells during stress periods [128].

The observation in Figure 2 that *ATXN2L* has acquired a distinct sequence pattern that is clearly separate from each *ATXN2* paralog throughout different species, from fish to human, is in good agreement with previous findings that the loss of one paralog does not trigger the compensatory induction of the other paralog [128]. Jointly, the data indicate separate roles in a cell.

Apart from their similar role for stress granules, *ATXN2L* deficiency in mice was recently shown to trigger multinucleated cells (perhaps due to deficits of nuclear membranes or their stability), selectively affecting nuclear envelope component SYNE2 at microtubule organizing centers (MTOC), the nuclear FMR1 interacting protein 2 (NUFIP1), and the nuclear alternative splicing pathway. It is also interesting to note that the nuclear homeobox transcription factor and pyramidal neuron marker *CUX1* is downregulated in *ATXN2L*-null mouse brain, in view of the *CUX2* gene being localized at the 3' end of the *ATXN2* gene in mouse and human, suggesting that a *CUX* expression enhancer acting in trans is part of the *ATXN2L* gene duplication segment and was affected by the mutation [59,128,129,380]. *ATXN2L* seems distributed more between the ER and the perinuclear space, while *ATXN2* appears associated more with the ER, Golgi, endosomes, and plasma membrane. Therefore,

we speculate that ATXN2L is specialized in the formation / maintenance / strengthening of nuclear membranes in eukaryotes, where microtubules are anchored, playing a less complex role that did not need to evolve much. In comparison, ATXN2 would be specialized for Golgi membranes, plasma membranes, and cell protrusions, a role that has become much more sophisticated for mammalian neurons / neurites / synapses than it was in protists.

3.4. Ataxin-2 Isoforms

Previous research mostly investigated the full-length isoforms of each species. Although several articles documented allelic splicing of Ataxin-2, and one article emphasized the existence of at least two start codons [124,126,127], any small bands in immunoblots got usually dismissed as apparent cross-reactivity. In particular, it was thought that the functions of Ataxin-2 depend on the joint presence of LSm, LsmAD, and PAM2 sequences. However, this datamining study identified many species where alternative isoforms either contain LSm and LSmAD or lack them. In addition, several database entries were encountered where the PAM2 motif was alternatively spliced out from the Ataxin-2 ortholog. Figure 4 documents six alternative isoform starts before the LSM sequence, another four starts before PAM2, and 3 additional starts after PAM2, but there is no functional understanding about the different contexts where each isoform is useful. It is clear that database entries may be contaminated by artifacts and fragments, but the vast majority of Ataxin-2 ortholog isoform entries start with a Met start codon, as expected for functional isoforms.

Together with the evidence that the C-terminal IDR3 sequences are most strongly expressed, and that multiple variations of C-terminal but not N-terminal short isoforms exist in spinocerebellar tissue, as seen in Figure S1, it is clear that the old concepts have to be revised and the short isoforms should be systematically investigated.

The smaller isoforms towards the C-terminus may be crucial for SCA2 and ALS patients. It is conceivable that the previous experiments targeting exon 1 of mouse *Atxn2* [134,381] were partially compensated by increased generation of shorter isoforms, some of which would contain all core domains. Such compensation efforts would explain why mice with *Atxn2*-knockout or *Atxn2*-knockin targeting exon 1 have only mild phenotypes of obesity, insulin resistance and dyslipidemia, or age-associated neurodegeneration, respectively [5,52,53,55,69,70,81,124,130,131,134,145,147,282,284,286,296,297,352,381,382]. These observations clearly diverge from *Atxn2l*-knockout mice with disruption at LSm / LSmAD sequences showing mid-gestation embryonic lethality with brain cortex lamination defects [59,128,129], from the knockout of complete fly Ataxin-2 or its LSm or LSmAD domains in *Drosophila melanogaster* that result in lethality [210], and from the knockdown of nematode Ataxin-2 in *C. elegans* that also causes embryonic lethality and germline defects [60,278,383].

3.5. How Are Ataxin-2 LSm and LSmAD Domains Essential for Eukaryotic Life?

The datamining about Ataxin-2 evolution confirmed that it became relevant when eukaryotic cells had already engulfed mitochondria and plastids, apparently when eukaryotic synthesis of organellar membranes started, and when organelles were distributed across the cell geometry. Ataxin-2 is not yet documented in glaucophyte, but appears in rhodophyte and chlorophyte algae and in other primitive protists. Upon terrestrialization of plants and animals, gene duplications were conserved, suggesting that more stress responses required higher dosage. However, it is unclear whether higher gravitational forces or increased exposure to sun light, heat, and drought are responsible. The fact that flying birds (although not flying bats) lost the duplicated gene might argue in favor of the importance of mechanical stress. In consequence, the notion that Ataxin-2 also has a role for the resilience of cellular membranes and cytoskeleton would be strengthened, in addition to its role for rRNA and mRNA quality control.

The essential parts of Ataxin-2 are evidently the beta-sheet structured LSm and the alpha-helix structured LSmAD, and perhaps also the linear PAM2 sequences. On the one hand, targeted interference with these domains causes lethality already at the embryonic stage, while interference

with the IDR sequences triggers mild phenotypes. On the other hand, only these three sequences and structures are strongly conserved from algae to humans, while IDRs have considerable variability.

3.6. How Should Future Investigations Be Re-Focused?

From the conclusions above, it is clear that samples from SCA2 and ALS13 patients, as well as their disease models and treatment interventions, should now be re-investigated. Urgent questions concern (i) the altered expression of neighbor genes due to the mutation or the therapeutic approach, (ii) in particular the altered expression and stability of the potentially active *U7* snRNA copy at the *ATXN2* gene locus (most *U7* copies are encoded on chromosome 12p13), as well as the processing of replication-dependent histone mRNA, together with the dynamics of U-bodies, Cajal bodies and spliceosomes [384–387], (iii) the altered expression of individual exons together with the stability of shorter Ataxin-2 isoforms, (iv) the altered resilience and dynamics of membranes and cytoskeleton of cells, (v) the oxidative damage and altered methylation of rRNA, in these samples.

This should be complemented by the generation of new mice with targeted conditional knockout of the *Atxn2* LSm domain and of the *Atxn2* LSmAD sequence, to first assess if this intervention causes lethal effects, and secondly to create tools where tissue-specific interactors of LSm and LSmAD can be studied.

Importantly, the interactor molecules of the LSmAD domain, as the characteristic feature of Ataxin-2 family members, should be documented to elucidate their functions. Subsequently, point mutations and deletions should target the LSmAD domain for a precise mapping of such interactions.

Furthermore, the impact of Ataxin-2 and its mutations on the different pathways of the unfolded protein response and the mechanical resilience / composition of membranes should be explored systematically, regarding the endoplasmic reticulum, mitochondria, plasma membrane endo/exocytosis, and nuclear membrane trafficking.

Most importantly, there is a consequence of our datamining effort for the design of preventive therapies against SCA2 and ALS13. In view of the complexity of Ataxin-2 isoforms and chimerisms, it seems crucial to selectively target the poorly expressed polyQ-surrounding sequences encoded by exon 1B of human *ATXN2* rather than the strongly expressed C-terminal sequences or the conserved domains where interference can trigger cell death. Selective targeting of these sequences before the LSm domain would target only a region of *ATXN2* that is non-essential and has evolved only since bony fish (Figure 4). Exon 1B encodes polyQ (as CAG repeat with CAA interruptions) and polyP (via CCC or CCG triplets), with both repeats jointly enhancing the GC content of the 5' end of *ATXN2* and thus increasing its mRNA stability but also its propensity for chimerisms in stress periods [270,388], while differentially modulating liquid-liquid phase-separation given that polyQ forms amyloid while polyP breaks it [153]. Such re-design of knockdown approaches to target exon 1B would avoid any impact on isoforms that are not full-length and would permit a partial compensation of functions by upregulated expression of smaller isoforms.

4. Materials and Methods

The publically available bioinformatics tools employed in this study comprise the BlastP algorithm at UniProt-UniParc and NCBI databases, with compilation of results in EXCEL as tables and in WORD as text files; additional homology searches via the STRING Heidelberg database; identification of additional orthologs by automated interrogation of these databases (harnessing the GenAI Claude Code tool to compile and share data via an iterative LSMAD-bootstrapping strategy, BlastP against NCBI nr database with taxonomy filters, ConservedDomain-Search validation with strict E-value thresholds, quality tier classification, and Python/BioPython implementation, followed by Suprabase data storage, and GitHub plus Docker software for data sharing); multi-protein sequence alignment via the CLUSTAL Omega algorithm; protein domain analyses by InterPro and Pfam algorithms as well as MitoProt and TPpred3 prediction tools of mitochondrial targeting sequences; structure analyses at ebi.ac.uk / rcsb.org databases and via PyMOL; expression analyses for mRNA isoforms at ENSEMBL and NCBI databases and for individual exons at GTEportal.org;

genomic analyses of exon-intron structure via the UCSC genome browser; visualization of alternatively spliced isoforms with Maya software; biomedical publications from the PubMed database. Regarding the protein structure 3D modelling, the AlphaFold predicted structures of human ATXN2 (AF-Q99700-F1-v6), ATXN2L (AF-Q8WWM7-F1-v6), and *G. chorda* ortholog PBPIP4 (AF-A0A2V3IVX2-F1-v6) were used in structural domain alignments. Structured regions corresponding to LSm, LSmAD and polyQ domains were extracted from the main predictions. Alignment was performed on the extracted molecules using the built-in “align” function with default parameters in Open-Source PyMOL, v3.1.0 (Schrödinger, LLC). Commercially licensed graphics softwares to illustrate results included Powerpoint, Adobe Photoshop, and BioRender.

5. Conclusions

Overall, the functions of Ataxin-2 in mRNA processing have been the main focus of attention so far, but its evolutionary data and genomic context suggest that its mutations also impact rRNA stress and membrane stress. Analysis of alternatively spliced exons and documented protein isoforms suggests low expression of N-terminal sequences, which could be targeted in neuroprotective efforts by knockdown strategies and may be compensated by strong expression of C-terminal isoforms, thus minimizing adverse effects.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Supplementary Figure S1: Schematic of documented transcript isoforms (ENSEMBL database IDs in the column at the left margin) with their exon structure (modified after <https://www.gtexportal.org/home/gene/ATXN2>, last accessed on 1. December 2025, with horizontal mirroring to have the N-terminal exon 1 at the left and the C-terminal exon 25 to the right - although the genomic DNA strand is copied in the opposite direction - and with adapted exon numbers, adhering to previous literature), highlighting the mRNAs with high expression in spinal cord and in cerebellar tissue in red color. The highly conserved domains of human ATXN2 are positioned relative to the exon scheme (MTS = mitochondrial targeting signal). At the right margin, the sum of amino acids encoded by each exon combination is provided for the spinocerebellar isoforms, to facilitate the comparison with experimental immunoblot data obtained by commercial antibodies with known epitope specificity, as provided by each manufacturer at the relevant websites. The position of immunogens used to generate each of these commercial antibodies is shown on top of this scheme, relative to the exon structure of isoforms under analysis.

Author Contributions: Conceptualization, G.W.J.A. and N.-E.S.; methodology, G.W.J.A, J.K., A.Au. and N.-E.S.; software, G.W.J.A, J.K., A.Au., G.A. and N.-E.S.; validation, G.W.J.A, S.G., J.K., I.L.-B., L.-E.A.-M., A.Ar., T.D., and N.-E.S.; formal analysis, G.W.J.A and C.B.; investigation, G.W.J.A; resources, G.A. and N.-E.S.; data curation, G.W.J.A and J.K.; writing—original draft preparation, G.W.J.A and N.-E.S.; writing—review and editing, G.W.J.A, J.K., T.D., and N.-E.S. with input from all authors; visualization, G.W.J.A, J.K. and N.-E.S.; supervision, N.-E.S.; project administration, G.W.J.A and N.-E.S.; funding acquisition, G.W.J.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deutsche Forschungsgemeinschaft, grant number AU96/21-1.

Data Availability Statement: Only data analyses were performed, no new primary data were generated. Upon request, the authors will provide additional details on the software employed.

Acknowledgments: The authors thank technical assistant Gabriele Köpf for invaluable help. During the preparation of this study, the authors used Claude Code to harness Python/BioPython programming software for the purposes of automated database searches of Ataxin-2 orthologs. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: Thomas Deller received honoraria from Novartis for lectures on human brain anatomy. The other authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

Ala	Alanine
ALS	Amyotrophic Lateral Sclerosis
AMPK	Protein Kinase AMP-Activated Catalytic Subunit Alpha 1
Arg	Arginine
ARM	Armadillo repeat
A-T	Ataxia Teleangiectasia
ATF6	Activating Transcription Factor 6,
ATP	Adenosine triphosphate
ATXN2	Ataxin-2
ATXN2L	Ataxin-2 like
BiP	Immunoglobulin Binding Protein
BLOC3	Biogenesis of Lysosome-Related Organelles Complex-3
BRAP2	BRCA1 Associated Protein 2
BRCA1	BR east CA nCer gene 1
BRX	Brevis radiX
C2HC	Cys2His2 Zinc Finger Protein Family
CAMKK	Calcium/calmodulin-Dependent Protein Kinase Kinase
CCZ1MC1	CCZ1B Vacuolar Protein Trafficking and Biogenesis Associated
CHOP	C/EBP Homologous Protein
CID3	Polyadenylate-binding Protein-interacting Protein 3
CIK	Catalytically Inactive Kinase
CLPP	ATP-dependent Clp Protease Proteolytic Subunit, mitochondrial
CoA	Coenzyme A
CPEB3	Cytoplasmic Polyadenylation Element Binding Protein 3
CTC	Carboxy-Terminal Conserved Domain
CUL-1	CULLIN-1
DDX	DEAD box RNA helicase domain
DEAD	Asp-Glu-Ala-Asp
DNA	Desoxyribonucleic Acid
ELOVL	Elongation of Very Long Chain Fatty Acids Protein
EMBL	European Molecular Biology Laboratory
ER	Endoplasmic Reticulum
ERAD	ER Associated Degradation
eRF3	Eukaryotic Release Factor 3
FAPP1	Four-Phosphate-Adaptor Protein 1
FBXO42	F-box Only Protein 42,
FeS	Iron-Sulfur-Cluster
FMN2	Formin-2
FTD	Fronto-Temporal Dementia
GADD34	Growth Arrest and DNA Damage-Inducible Protein
GDP	Guanosine Diphosphate
Glu	Glutamine
GSPT1	G1 to S Phase Transition 1 Protein
GTP	Guanosine Triphosphate
GW182	TNRC6 = Trinucleotide Repeat Containing 6A
HECT	Homologous to E6-AP C-Terminus
His	Histidine
HSP1/4	Heat Shock Protein 1/4
IDR	Intrinsically Disordered Sequence
IP3R1	Inositol 1,4,5-trisphosphate Receptor Type 1
IRE1	Iron Responsive Element
KAI1	CD82 = Cluster of Differentiation 82
KCS	3-ketoacyl-CoA Synthase
KO	Knockout
LACT	lecithin:cholesterol acyltransferase

LARP4	La-related Proteins
LLPS	Liquid-Liquid Phase Separation
LSm	like-Sm
LSM11/12	U7 snRNA-associated Sm-like protein 11/12
LSmAD	Lsm-associated Domain
MAM	Mitochondria-associated ER Membrane
MAPKAP5	MAP kinase-activated protein kinase 5
MED25	Mediator Complex Subunit 25
Met	Methionine
MK5	MAPK-activated Protein Kinase 5
MLLE	MLLEKITG, from French Mademoiselle
MON1	Vacuolar Fusion Protein Mon1
mRNA	Messenger RNA
MTOC	Microtubule Organizing Centers
MTORC1	Mechanistic Target of Rapamycin Complex 1
MYC	Myc Proto-Oncogene Protein, derived from myelocytomatosis-neuroblastoma
MYCN	N-Myc Proto-Oncogene Protein
MYT1	myelin transcription factor-1
NAADP	Nicotinic Acid Adenine Dinucleotide Phosphate
NRF2	Nuclear Factor Erythroid 2-Related Factor 2
NTH1	Nth Like DNA Glycosylase 1
PAIP1/2	Polyadenylate-binding Protein-interacting Protein 1/2
PAM2	PABP-interacting Motif 2
PAN3	Poly(A) Specific Ribonuclease Subunit 3
PARK2	PRKN = Parkin
PARKIN	Parkin RBR E3 Ubiquitin Protein Ligase
PBP1	Poly(A)-binding protein)
PDAT	phospholipid:diacylglycerol acyltransferase
PERK	Protein Kinase RNA-Like Endoplasmic Reticulum Kinase
PFAM	Protein Families Database
PINK1	PTEN Induced Kinase 1
polyQ	Polyglutamine repeat
PRAK	p38-Regulated/Activated Protein Kinase
PRM	Proline-rich Motifs
Pro	Proline
PRR36	Proline-rich Protein 36
PSP	Progressive Supranuclear Palsy
PtdIns4P	Phosphatidylinositol 4-phosphate
Q	Glutamine
Rab7/32/38	Ras-related Protein
RALDH1	Aldehyde Dehydrogenase 1 Family Member A1
RAN	Ras-related Nuclear Protein
RAR	Retinoid Acid Receptors
rFNA	Ribosomal RNA
RIDD	Regulated IRE1-dependent decay
RIOK2	Right Open Reading Frame Kinase
RNA	Ribonucleic acid
RNAseq	RNA sequencing
RNP	Ribonucleoprotein
ROR	RAR-related Orphan Receptors
ROS	Reactive Oxygen Species
RRM	RNA Recognition Motif
RXR	Retinoid X Receptors
SCA2	Spinocerebellar Ataxia Type 2
SCAR21	Spinocerebellar Ataxia, Autosomal Recessive 21
SCYL1	SCY1 Like Pseudokinase 1
SEC6	SNARE-binding Exocyst Subunit
Ser	Serine
SF3B4	Splicing Factor 3B subunit

SFPQ	Splicing Factor, Proline-and Glutamine-rich
shRNA	Short RNA
SKOR1	SKI Family Transcriptional Corepressor 1
SKP-A	S-phase Kinase Associated A Protein
TAF4	TFIID Subunit 4
TDP-43	TAR DNA-binding Protein 43
Tob1/2	Transducer Of ERBB2, 1
TORC2	Target of Rapamycin Complex 2
TPX2	Targeting Protein for Xklp2
TRAF2	TNF Receptor Associated Factor 2
TTC3	Tetratricopeptide Repeat Domain 3
UBP	Ubiquitin-Specific Protease
UPR	Unfolded Protein Response
USP10	Ubiquitin-Specific Peptidase 10
Val	Valine
VLCFA	Very-Long-Chain Fatty Acids
WASF2	WASP Family Member 2
WT	Wildtype
XBP1	X-box Binding Protein 1

References

1. Pulst, S.M., et al., *Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2*. Nat Genet, 1996. **14**(3): p. 269-76.
2. Sanpei, K., et al., *Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT*. Nat Genet, 1996. **14**(3): p. 277-84.
3. Imbert, G., et al., *Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats*. Nat Genet, 1996. **14**(3): p. 285-91.
4. Elden, A.C., et al., *Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS*. Nature, 2010. **466**(7310): p. 1069-75.
5. Becker, L.A., et al., *Therapeutic reduction of ataxin-2 extends lifespan and reduces pathology in TDP-43 mice*. Nature, 2017. **544**(7650): p. 367-371.
6. Shulman, J.M. and M.B. Feany, *Genetic modifiers of tauopathy in Drosophila*. Genetics, 2003. **165**(3): p. 1233-42.
7. Ross, O.A., et al., *Ataxin-2 repeat-length variation and neurodegeneration*. Hum Mol Genet, 2011. **20**(16): p. 3207-12.
8. Thompson, A.D., et al., *Analysis of the tau-associated proteome reveals that exchange of Hsp70 for Hsp90 is involved in tau degradation*. ACS Chem Biol, 2012. **7**(10): p. 1677-86.
9. Kim, A., S.H. Park, and B. Jeon, *An Autopsy Case of Progressive Supranuclear Palsy With Incidental ATXN2 Expansion*. JAMA Neurol, 2018. **75**(8): p. 1025-1027.
10. Auburger, G., et al., *Autosomal dominant ataxia: genetic evidence for locus heterogeneity from a Cuban founder-effect population*. Am J Hum Genet, 1990. **46**(6): p. 1163-77.
11. Orozco Diaz, G., et al., *Autosomal dominant cerebellar ataxia: clinical analysis of 263 patients from a homogeneous population in Holguin, Cuba*. Neurology, 1990. **40**(9): p. 1369-75.
12. Lastres-Becker, I., U. Rub, and G. Auburger, *Spinocerebellar ataxia 2 (SCA2)*. Cerebellum, 2008. **7**(2): p. 115-24.
13. Auburger, G.W., *Spinocerebellar ataxia type 2*. Handb Clin Neurol, 2012. **103**: p. 423-36.
14. Rub, U., et al., *Clinical features, neurogenetics and neuropathology of the polyglutamine spinocerebellar ataxias type 1, 2, 3, 6 and 7*. Prog Neurobiol, 2013. **104**: p. 38-66.
15. Lorenzetti, D., S. Bohlega, and H.Y. Zoghbi, *The expansion of the CAG repeat in ataxin-2 is a frequent cause of autosomal dominant spinocerebellar ataxia*. Neurology, 1997. **49**(4): p. 1009-13.
16. Lessing, D. and N.M. Bonini, *Polyglutamine genes interact to modulate the severity and progression of neurodegeneration in Drosophila*. PLoS Biol, 2008. **6**(2): p. e29.
17. Velazquez Perez, L., et al., *Molecular epidemiology of spinocerebellar ataxias in Cuba: insights into SCA2 founder effect in Holguin*. Neurosci Lett, 2009. **454**(2): p. 157-60.

18. Sequeiros, J., S. Seneca, and J. Martindale, *Consensus and controversies in best practices for molecular genetic testing of spinocerebellar ataxias*. Eur J Hum Genet, 2010. **18**(11): p. 1188-95.
19. Lyasota, O., et al., *Stability of the CAG Tract in the ATXN2 Gene Depends on the Localization of CAA Interruptions*. Biomedicines, 2024. **12**(8).
20. Matsuura, T., et al., *Mosaicism of unstable CAG repeats in the brain of spinocerebellar ataxia type 2*. J Neurol, 1999. **246**(9): p. 835-9.
21. Auburger, G., et al., *Efficient Prevention of Neurodegenerative Diseases by Depletion of Starvation Response Factor Ataxin-2*. Trends Neurosci, 2017. **40**(8): p. 507-516.
22. Kacher, R., et al., *CAG repeat mosaicism is gene specific in spinocerebellar ataxias*. Am J Hum Genet, 2024. **111**(5): p. 913-926.
23. Magana, J.J., L. Velazquez-Perez, and B. Cisneros, *Spinocerebellar ataxia type 2: clinical presentation, molecular mechanisms, and therapeutic perspectives*. Mol Neurobiol, 2013. **47**(1): p. 90-104.
24. Estrada, R., et al., *Spinocerebellar ataxia 2 (SCA2): morphometric analyses in 11 autopsies*. Acta Neuropathol, 1999. **97**(3): p. 306-10.
25. Varrone, A., et al., *Reduced striatal [123 I]FP-CIT binding in SCA2 patients without parkinsonism*. Ann Neurol, 2004. **55**(3): p. 426-30.
26. van de Warrenburg, B.P., et al., *Peripheral nerve involvement in spinocerebellar ataxias*. Arch Neurol, 2004. **61**(2): p. 257-61.
27. Lee, T., et al., *Ataxin-2 intermediate-length polyglutamine expansions in European ALS patients*. Hum Mol Genet, 2011. **20**(9): p. 1697-700.
28. Gispert, S., et al., *The modulation of Amyotrophic Lateral Sclerosis risk by ataxin-2 intermediate polyglutamine expansions is a specific effect*. Neurobiol Dis, 2012. **45**(1): p. 356-61.
29. Lahut, S., et al., *ATXN2 and its neighbouring gene SH2B3 are associated with increased ALS risk in the Turkish population*. PLoS One, 2012. **7**(8): p. e42956.
30. Velazquez-Perez, L., et al., *Early corticospinal tract damage in prodromal SCA2 revealed by EEG-EMG and EMG-EMG coherence*. Clin Neurophysiol, 2017. **128**(12): p. 2493-2502.
31. Velazquez-Perez, L., et al., *Progression of corticospinal tract dysfunction in pre-ataxic spinocerebellar ataxia type 2: A two-years follow-up TMS study*. Clin Neurophysiol, 2018. **129**(5): p. 895-900.
32. Velazquez-Perez, L., et al., *Progression of early features of spinocerebellar ataxia type 2 in individuals at risk: a longitudinal study*. Lancet Neurol, 2014. **13**(5): p. 482-9.
33. Jacobi, H., et al., *Long-term disease progression in spinocerebellar ataxia types 1, 2, 3, and 6: a longitudinal cohort study*. Lancet Neurol, 2015. **14**(11): p. 1101-8.
34. Velazquez-Perez, L., et al., *Hereditary Ataxias in Cuba: A Nationwide Epidemiological and Clinical Study in 1001 Patients*. Cerebellum, 2020. **19**(2): p. 252-264.
35. Rive Le Gouard, N., et al., *The Two Faces of Pediatric SCA2*. Eur J Neurol, 2025. **32**(8): p. e70314.
36. Charles, P., et al., *Are interrupted SCA2 CAG repeat expansions responsible for parkinsonism?* Neurology, 2007. **69**(21): p. 1970-5.
37. Kim, J.M., et al., *Importance of low-range CAG expansion and CAA interruption in SCA2 Parkinsonism*. Arch Neurol, 2007. **64**(10): p. 1510-8.
38. Yu, Z., et al., *PolyQ repeat expansions in ATXN2 associated with ALS are CAA interrupted repeats*. PLoS One, 2011. **6**(3): p. e17951.
39. Corrado, L., et al., *ATXN-2 CAG repeat expansions are interrupted in ALS patients*. Hum Genet, 2011. **130**(4): p. 575-80.
40. Wang, C., et al., *Linkage analysis and whole-exome sequencing exclude extra mutations responsible for the parkinsonian phenotype of spinocerebellar ataxia-2*. Neurobiol Aging, 2015. **36**(1): p. 545 e1-7.
41. Al-Ramahi, I., et al., *dAtaxin-2 mediates expanded Ataxin-1-induced neurodegeneration in a Drosophila model of SCA1*. PLoS Genet, 2007. **3**(12): p. e234.
42. Van Langenhove, T., et al., *Ataxin-2 polyQ expansions in FTL-ALS spectrum disorders in Flanders-Belgian cohorts*. Neurobiol Aging, 2012. **33**(5): p. 1004 e17-20.
43. Neuenschwander, A.G., et al., *Amyotrophic lateral sclerosis risk for spinocerebellar ataxia type 2 ATXN2 CAG repeat alleles: a meta-analysis*. JAMA Neurol, 2014. **71**(12): p. 1529-34.

44. Nobrega, C., et al., *Re-establishing ataxin-2 downregulates translation of mutant ataxin-3 and alleviates Machado-Joseph disease*. *Brain*, 2015. **138**(Pt 12): p. 3537-54.
45. Rubino, E., et al., *ATXN2 intermediate repeat expansions influence the clinical phenotype in frontotemporal dementia*. *Neurobiol Aging*, 2019. **73**: p. 231 e7-231 e9.
46. Glass, J.D., et al., *ATXN2 intermediate expansions in amyotrophic lateral sclerosis*. *Brain*, 2022. **145**(8): p. 2671-2676.
47. Mijajlovic, M., et al., *Transcranial sonography in spinocerebellar ataxia type 2*. *J Neurol*, 2008. **255**(8): p. 1164-7.
48. Shibata, H., D.P. Huynh, and S.M. Pulst, *A novel protein with RNA-binding motifs interacts with ataxin-2*. *Hum Mol Genet*, 2000. **9**(9): p. 1303-13.
49. Huynh, D.P., et al., *Expansion of the polyQ repeat in ataxin-2 alters its Golgi localization, disrupts the Golgi complex and causes cell death*. *Hum Mol Genet*, 2003. **12**(13): p. 1485-96.
50. Ng, H., S.M. Pulst, and D.P. Huynh, *Ataxin-2 mediated cell death is dependent on domains downstream of the polyQ repeat*. *Exp Neurol*, 2007. **208**(2): p. 207-15.
51. Del Castillo, U., et al., *Conserved role for Ataxin-2 in mediating endoplasmic reticulum dynamics*. *Traffic*, 2019. **20**(6): p. 436-447.
52. Nonis, D., et al., *Ataxin-2 associates with the endocytosis complex and affects EGF receptor trafficking*. *Cell Signal*, 2008. **20**(10): p. 1725-39.
53. Drost, J., et al., *Ataxin-2 modulates the levels of Grb2 and SRC but not ras signaling*. *J Mol Neurosci*, 2013. **51**(1): p. 68-81.
54. Bian, W., et al., *A spatially defined human Notch receptor interaction network reveals Notch intracellular storage and Ataxin-2-mediated fast recycling*. *Cell Rep*, 2023. **42**(7): p. 112819.
55. van de Loo, S., et al., *Ataxin-2 associates with rough endoplasmic reticulum*. *Exp Neurol*, 2009. **215**(1): p. 110-8.
56. Satterfield, T.F. and L.J. Pallanck, *Ataxin-2 and its Drosophila homolog, ATX2, physically assemble with polyribosomes*. *Hum Mol Genet*, 2006. **15**(16): p. 2523-32.
57. Satterfield, T.F., S.M. Jackson, and L.J. Pallanck, *A Drosophila homolog of the polyglutamine disease gene SCA2 is a dosage-sensitive regulator of actin filament formation*. *Genetics*, 2002. **162**(4): p. 1687-702.
58. Ralser, M., et al., *Ataxin-2 and huntingtin interact with endophilin-A complexes to function in plastin-associated pathways*. *Hum Mol Genet*, 2005. **14**(19): p. 2893-909.
59. Key, J., et al., *ATXN2L primarily interacts with NUFIP2, the absence of ATXN2L results in NUFIP2 depletion, and the ATXN2-polyQ expansion triggers NUFIP2 accumulation*. *Neurobiol Dis*, 2025. **209**: p. 106903.
60. Stubenvoll, M.D., et al., *Correction: ATX-2, The C. elegans Ortholog of Human Ataxin-2, Regulates Centrosome Size and Microtubule Dynamics*. *PLoS Genet*, 2016. **12**(12): p. e1006519.
61. Del Castillo, U., et al., *Ataxin-2 is essential for cytoskeletal dynamics and neurodevelopment in Drosophila*. *iScience*, 2022. **25**(1): p. 103536.
62. Boeynaems, S., et al., *Poly(A)-binding protein is an ataxin-2 chaperone that regulates biomolecular condensates*. *Mol Cell*, 2023. **83**(12): p. 2020-2034 e6.
63. Kim, S.K. and V.I. Gelfand, *PolyQ-Expansion of Ataxin-2 Disrupts Microtubule Stability and Impairs Axon Outgrowth*. *J Neurosci*, 2025. **45**(40).
64. Nonhoff, U., et al., *Ataxin-2 interacts with the DEAD/H-box RNA helicase DDX6 and interferes with P-bodies and stress granules*. *Mol Biol Cell*, 2007. **18**(4): p. 1385-96.
65. Swisher, K.D. and R. Parker, *Localization to, and effects of Pbp1, Pbp4, Lsm12, Dhh1, and Pab1 on stress granules in Saccharomyces cerevisiae*. *PLoS One*, 2010. **5**(4): p. e10006.
66. Yamagishi, R., et al., *Concerted action of ataxin-2 and PABPC1-bound mRNA poly(A) tail in the formation of stress granules*. *Nucleic Acids Res*, 2024. **52**(15): p. 9193-9209.
67. Palozzi, J.M., et al., *Mitochondrial DNA quality control in the female germline requires a unique programmed mitophagy*. *Cell Metab*, 2022. **34**(11): p. 1809-1823 e6.
68. Huynh, D.P., et al., *Parkin is an E3 ubiquitin-ligase for normal and mutant ataxin-2 and prevents ataxin-2-induced cell death*. *Exp Neurol*, 2007. **203**(2): p. 531-41.
69. Halbach, M.V., et al., *Both ubiquitin ligases FBXW8 and PARK2 are sequestered into insolubility by ATXN2 PolyQ expansions, but only FBXW8 expression is dysregulated*. *PLoS One*, 2015. **10**(3): p. e0121089.

70. Sen, N.E., et al., *Search for SCA2 blood RNA biomarkers highlights Ataxin-2 as strong modifier of the mitochondrial factor PINK1 levels*. *Neurobiol Dis*, 2016. **96**: p. 115-126.
71. O'Neill, E.C., et al., *The transcriptome of Euglena gracilis reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry*. *Mol Biosyst*, 2015. **11**(10): p. 2808-20.
72. Luisi, P.L., *Why are enzymes macromolecules?* *Naturwissenschaften*, 1979. **66**(10): p. 498-504.
73. Marijuan, P.C., R. del Moral, and J. Navarro, *On eukaryotic intelligence: signaling system's guidance in the evolution of multicellular organization*. *Biosystems*, 2013. **114**(1): p. 8-24.
74. Toro, I., et al., *RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex*. *EMBO J*, 2001. **20**(9): p. 2293-303.
75. Ralser, M., et al., *An integrative approach to gain insights into the cellular function of human ataxin-2*. *J Mol Biol*, 2005. **346**(1): p. 203-14.
76. Jimenez-Lopez, D. and P. Guzman, *Insights into the evolution and domain structure of Ataxin-2 proteins across eukaryotes*. *BMC Res Notes*, 2014. **7**: p. 453.
77. Zhang, S., et al., *The LSmAD Domain of Ataxin-2 Modulates the Structure and RNA Binding of Its Preceding LSm Domain*. *Cells*, 2025. **14**(5).
78. Kozlov, G., et al., *Structural basis of ligand recognition by PABC, a highly specific peptide-binding domain found in poly(A)-binding protein and a HECT ubiquitin ligase*. *EMBO J*, 2004. **23**(2): p. 272-81.
79. Huang, K.L., et al., *Phosphorylation at intrinsically disordered regions of PAM2 motif-containing proteins modulates their interactions with PABPC1 and influences mRNA fate*. *RNA*, 2013. **19**(3): p. 295-305.
80. Kozlov, G., et al., *Structural basis of binding of P-body-associated proteins GW182 and ataxin-2 by the Mlle domain of poly(A)-binding protein*. *J Biol Chem*, 2010. **285**(18): p. 13599-606.
81. Damrath, E., et al., *ATXN2-CAG42 sequesters PABPC1 into insolubility and induces FBXW8 in cerebellum of old ataxic knock-in mice*. *PLoS Genet*, 2012. **8**(8): p. e1002920.
82. Albrecht, M. and T. Lengauer, *Survey on the PABC recognition motif PAM2*. *Biochem Biophys Res Commun*, 2004. **316**(1): p. 129-38.
83. Petrauskas, A., et al., *Structured and disordered regions of Ataxin-2 contribute differently to the specificity and efficiency of mRNP granule formation*. *PLoS Genet*, 2024. **20**(5): p. e1011251.
84. Obornik, M. and J. Lukes, *Cell biology of chromerids: autotrophic relatives to apicomplexan parasites*. *Int Rev Cell Mol Biol*, 2013. **306**: p. 333-69.
85. Cavalier-Smith, T., *The excavate protozoan phyla Metamonada Grasse emend. (Anaeromonadea, Parabasalia, Carpediemonas, Eopharyngia) and Loukozoa emend. (Jakobea, Malawimonas): their evolutionary affinities and new higher taxa*. *Int J Syst Evol Microbiol*, 2003. **53**(Pt 6): p. 1741-58.
86. Hannaert, V., et al., *Plant-like traits associated with metabolism of Trypanosoma parasites*. *Proc Natl Acad Sci U S A*, 2003. **100**(3): p. 1067-71.
87. Whitfield, J., *Sleeping sickness bug swallowed a plant*. *Nature*, 2003.
88. Karnkowska, A., et al., *A Eukaryote without a Mitochondrial Organelle*. *Curr Biol*, 2016. **26**(10): p. 1274-84.
89. Burri, L., et al., *Microsporidian mitosomes retain elements of the general mitochondrial targeting system*. *Proc Natl Acad Sci U S A*, 2006. **103**(43): p. 15916-20.
90. Nowack, E.C.M. and A.P.M. Weber, *Genomics-Informed Insights into Endosymbiotic Organelle Evolution in Photosynthetic Eukaryotes*. *Annu Rev Plant Biol*, 2018. **69**: p. 51-84.
91. Kim, E. and L.E. Graham, *EEF2 analysis challenges the monophyly of Archaeplastida and Chromalveolata*. *PLoS One*, 2008. **3**(7): p. e2621.
92. Keeling, P.J., *The endosymbiotic origin, diversification and fate of plastids*. *Philos Trans R Soc Lond B Biol Sci*, 2010. **365**(1541): p. 729-48.
93. Keeling, P.J., *Diversity and evolutionary history of plastids and their hosts*. *Am J Bot*, 2004. **91**(10): p. 1481-93.
94. Ponce-Toledo, R.I., et al., *An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids*. *Curr Biol*, 2017. **27**(3): p. 386-391.
95. Figueroa-Martinez, F., C. Jackson, and A. Reyes-Prieto, *Plastid Genomes from Diverse Glaucophyte Genera Reveal a Largely Conserved Gene Content and Limited Architectural Diversity*. *Genome Biol Evol*, 2019. **11**(1): p. 174-188.

96. Lemieux, C., C. Otis, and M. Turmel, *Ancestral chloroplast genome in Mesostigma viride reveals an early branch of green plant evolution*. *Nature*, 2000. **403**(6770): p. 649-52.
97. Lee, R.E., *Phycology, chapter 5*. 2018, Cambridge University Press: New York.
98. Miyagishima, S.Y., *Taming the perils of photosynthesis by eukaryotes: constraints on endosymbiotic evolution in aquatic ecosystems*. *Commun Biol*, 2023. **6**(1): p. 1150.
99. Woelkerling, W.J., *An Introduction*. *Biology of the Red Algae*, ed. K.M. Cole and R.G. Sheath. 1990, Cambridge: Cambridge University Press.
100. Rohmer, M., P. Bouvier, and G. Ourisson, *Molecular evolution of biomembranes: structural equivalents and phylogenetic precursors of sterols*. *Proc Natl Acad Sci U S A*, 1979. **76**(2): p. 847-51.
101. Ourisson, G. and Y. Nakatani, *The terpenoid theory of the origin of cellular life: the evolution of terpenoids to cholesterol*. *Chem Biol*, 1994. **1**(1): p. 11-23.
102. Wu, F., P. Janvier, and C. Zhang, *The rise of predation in Jurassic lampreys*. *Nat Commun*, 2023. **14**(1): p. 6652.
103. Meyer, A. and M. Schartl, *Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions*. *Curr Opin Cell Biol*, 1999. **11**(6): p. 699-704.
104. Schilling, T.F. and R.D. Knight, *Origins of anteroposterior patterning and Hox gene regulation during chordate evolution*. *Philos Trans R Soc Lond B Biol Sci*, 2001. **356**(1414): p. 1599-613.
105. Moens, C.B. and V.E. Prince, *Constructing the hindbrain: insights from the zebrafish*. *Dev Dyn*, 2002. **224**(1): p. 1-17.
106. Wilson, L. and M. Maden, *The mechanisms of dorsoventral patterning in the vertebrate neural tube*. *Dev Biol*, 2005. **282**(1): p. 1-13.
107. Lupo, G., W.A. Harris, and K.E. Lewis, *Mechanisms of ventral patterning in the vertebrate nervous system*. *Nat Rev Neurosci*, 2006. **7**(2): p. 103-14.
108. Marletaz, F., et al., *Retinoic acid signaling and the evolution of chordates*. *Int J Biol Sci*, 2006. **2**(2): p. 38-47.
109. Kuraku, S., *Palaeophylogenomics of the vertebrate ancestor--impact of hidden paralogy on hagfish and lamprey gene phylogeny*. *Integr Comp Biol*, 2010. **50**(1): p. 124-9.
110. Campo-Paysaa, F., et al., *Evolution of retinoic acid receptors in chordates: insights from three lamprey species, Lampetra fluviatilis, Petromyzon marinus, and Lethenteron japonicum*. *Evodevo*, 2015. **6**: p. 18.
111. Handberg-Thorsager, M., et al., *The ancestral retinoic acid receptor was a low-affinity sensor triggering neuronal differentiation*. *Sci Adv*, 2018. **4**(2): p. eaao1261.
112. Bedois, A.M.H., H.J. Parker, and R. Krumlauf, *Retinoic Acid Signaling in Vertebrate Hindbrain Segmentation: Evolution and Diversification*. *Diversity*, 2021. **13**: p. 398.
113. Qian, X., et al., *Identification and characterization of novel substrates of Trk receptors in developing neurons*. *Neuron*, 1998. **21**(5): p. 1017-29.
114. Joe, R.M., et al., *Phosphorylation of the Unique C-Terminal Tail of the Alpha Isoform of the Scaffold Protein SH2B1 Controls the Ability of SH2B1alpha To Enhance Nerve Growth Factor Function*. *Mol Cell Biol*, 2018. **38**(6).
115. Wang, T.C., et al., *The adaptor protein SH2B3 (Lnk) negatively regulates neurite outgrowth of PC12 cells and cortical neurons*. *PLoS One*, 2011. **6**(10): p. e26433.
116. Lin, L., et al., *ATXN2L upregulated by epidermal growth factor promotes gastric cancer cell invasiveness and oxaliplatin resistance*. *Cell Death Dis*, 2019. **10**(3): p. 173.
117. Auburger, G., et al., *12q24 locus association with type 1 diabetes: SH2B3 or ATXN2?* *World J Diabetes*, 2014. **5**(3): p. 316-27.
118. Li, L., et al., *Ataxin-2: a powerful RNA-binding protein*. *Discov Oncol*, 2024. **15**(1): p. 298.
119. Zhang, G., et al., *Comparative genomics reveals insights into avian genome evolution and adaptation*. *Science*, 2014. **346**(6215): p. 1311-20.
120. Jiao, Y., et al., *Ancestral polyploidy in seed plants and angiosperms*. *Nature*, 2011. **473**(7345): p. 97-100.
121. Zhang, Z., D. Mistry, and R. Jetter, *Micromorphological and Chemical Characterization of *Drimys winteri* Leaf Surfaces: The Secondary Alcohols Forming Epicuticular Wax Crystals Are Accompanied by Alkanediol, Alkanetriol and Ketol Derivatives*. *Plant Cell Physiol*, 2024. **65**(8): p. 1245-1260.
122. Chang, C.Y., et al., *Champions of winter survival: cold acclimation and molecular regulation of cold hardiness in evergreen conifers*. *New Phytol*, 2021. **229**(2): p. 675-691.

123. Rice, D.W., et al., *Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm Amborella*. Science, 2013. **342**(6165): p. 1468-73.
124. Lastres-Becker, I., et al., *New alternative splicing variants of the ATXN2 transcript*. Neurol Res Pract, 2019. **1**: p. 22.
125. Nechiporuk, T., et al., *The mouse SCA2 gene: cDNA sequence, alternative splicing and protein expression*. Hum Mol Genet, 1998. **7**(8): p. 1301-9.
126. Affaitati, A., et al., *Identification of alternative splicing of spinocerebellar ataxia type 2 gene*. Gene, 2001. **267**(1): p. 89-93.
127. Scoles, D.R., et al., *ETS1 regulates the expression of ATXN2*. Hum Mol Genet, 2012. **21**(23): p. 5048-65.
128. Key, J., et al., *Mid-Gestation lethality of Atxn2l-Ablated Mice*. Int J Mol Sci, 2020. **21**(14).
129. Key, J., et al., *Conditional ATXN2L-Null in Adult Frontal Cortex CamK2a+ Neurons Does Not Cause Cell Death but Restricts Spontaneous Mobility and Affects the Alternative Splicing Pathway*. Cells, 2025. **14**(19).
130. Sen, N.E., et al., *Generation of an Atxn2-CAG100 knock-in mouse reveals N-acetylaspartate production deficit due to early Nat8l dysregulation*. Neurobiol Dis, 2019. **132**: p. 104559.
131. Almaguer-Mederos, L.E., et al., *Spinal Cord Phosphoproteome of SCA2 Mouse Model Reveals Alteration of ATXN2-N-Term PRM-SH3-Actin Interactome and of Autophagy*. Mol Cell Proteomics, 2025. **24**(11): p. 101072.
132. Morgan, A.A. and E. Rubenstein, *Proline: the distribution, frequency, positioning, and common functional roles of proline and polyproline sequences in the human proteome*. PLoS One, 2013. **8**(1): p. e53785.
133. Mandal, A., S. Mandal, and M.H. Park, *Genome-wide analyses and functional classification of proline repeat-rich proteins: potential role of eIF5A in eukaryotic evolution*. PLoS One, 2014. **9**(11): p. e111800.
134. Lastres-Becker, I., et al., *Insulin receptor and lipid metabolism pathology in ataxin-2 knock-out mice*. Hum Mol Genet, 2008. **17**(10): p. 1465-81.
135. van de Poll, F., et al., *Pbp1 associates with Puf3 and promotes translation of its target mRNAs involved in mitochondrial biogenesis*. PLoS Genet, 2023. **19**(5): p. e1010774.
136. Tuong Vi, D.T., et al., *Pbp1, the yeast ortholog of human Ataxin-2, functions in the cell growth on non-fermentable carbon sources*. PLoS One, 2021. **16**(5): p. e0251456.
137. Cornelius, N., et al., *Evidence of oxidative stress and mitochondrial dysfunction in spinocerebellar ataxia type 2 (SCA2) patient fibroblasts: Effect of coenzyme Q10 supplementation on these parameters*. Mitochondrion, 2017. **34**: p. 103-114.
138. Seidel, G., et al., *Quantitative Global Proteomics of Yeast PBP1 Deletion Mutants and Their Stress Responses Identifies Glucose Metabolism, Mitochondrial, and Stress Granule Changes*. J Proteome Res, 2017. **16**(2): p. 504-515.
139. Chitre, M. and P. Emery, *ATXN2 is a target of N-terminal proteolysis*. PLoS One, 2023. **18**(12): p. e0296085.
140. Leitgeb, B., et al., *Studying the structural properties of polyalanine and polyglutamine peptides*. J Mol Model, 2007. **13**(11): p. 1141-50.
141. Brown, L.Y. and S.A. Brown, *Alanine tracts: the expanding story of human illness and trinucleotide repeats*. Trends Genet, 2004. **20**(1): p. 51-8.
142. Kurokawa, R., et al., *Clinical and neuroimaging review of triplet repeat diseases*. Jpn J Radiol, 2023. **41**(2): p. 115-130.
143. Gillon, A.D., C.F. Latham, and E.A. Miller, *Vesicle-mediated ER export of proteins and lipids*. Biochim Biophys Acta, 2012. **1821**(8): p. 1040-9.
144. Vianna, M.C., et al., *Drosophila ataxin-2 gene encodes two differentially expressed isoforms and its function in larval fat body is crucial for development of peripheral tissues*. FEBS Open Bio, 2016. **6**(11): p. 1040-1053.
145. Sen, N.E., et al., *In Human and Mouse Spino-Cerebellar Tissue, Ataxin-2 Expansion Affects Ceramide-Sphingomyelin Metabolism*. Int J Mol Sci, 2019. **20**(23).
146. Watanabe, R., et al., *Intracellular dynamics of Ataxin-2 in the human brains with normal and frontotemporal lobar degeneration with TDP-43 inclusions*. Acta Neuropathol Commun, 2020. **8**(1): p. 176.
147. Canet-Pons, J., et al., *Atxn2-CAG100-KnockIn mouse spinal cord shows progressive TDP43 pathology associated with cholesterol biosynthesis suppression*. Neurobiol Dis, 2021. **152**: p. 105289.
148. Torricella, F., V. Tugarinov, and G.M. Clore, *Nucleation of Huntingtin Aggregation Proceeds via Conformational Conversion of Pre-Formed, Sparsely-Populated Tetramers*. Adv Sci (Weinh), 2024. **11**(24): p. e2309217.

149. Zhang, L., et al., *Possible Co-Evolution of Polyglutamine and Polyproline in Huntingtin Protein: Proline-Rich Domain as Transient Folding Chaperone*. J Phys Chem Lett, 2022. **13**(27): p. 6331-6341.
150. Darnell, G., et al., *Flanking polyproline sequences inhibit beta-sheet structure in polyglutamine segments by inducing PPII-like helix structure*. J Mol Biol, 2007. **374**(3): p. 688-704.
151. Ceccon, A., V. Tugarinov, and G.M. Clore, *Quantitative Exchange NMR-Based Analysis of Huntingtin-SH3 Interactions Suggests an Allosteric Mechanism of Inhibition of Huntingtin Aggregation*. J Am Chem Soc, 2021. **143**(25): p. 9672-9681.
152. Nagarajan, A., S. Jawahery, and S. Matysiak, *The effects of flanking sequences in the interaction of polyglutamine peptides with a membrane bilayer*. J Phys Chem B, 2014. **118**(24): p. 6368-79.
153. Ramirez de Mingo, D., et al., *Conformational dynamics in the disordered region of human CPEB3 linked to memory consolidation*. BMC Biol, 2022. **20**(1): p. 129.
154. Gerber, H.P., et al., *Transcriptional activation modulated by homopolymeric glutamine and proline stretches*. Science, 1994. **263**(5148): p. 808-11.
155. Achsel, T., H. Stark, and R. Luhrmann, *The Sm domain is an ancient RNA-binding motif with oligo(U) specificity*. Proc Natl Acad Sci U S A, 2001. **98**(7): p. 3685-9.
156. Sobti, M., et al., *Engineered rings of mixed yeast Lsm proteins show differential interactions with translation factors and U-rich RNA*. Biochemistry, 2010. **49**(11): p. 2335-45.
157. Schumacher, M.A., et al., *Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein*. EMBO J, 2002. **21**(13): p. 3546-56.
158. Zhang, A., et al., *Global analysis of small RNA and mRNA targets of Hfq*. Mol Microbiol, 2003. **50**(4): p. 1111-24.
159. Folichon, M., et al., *The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation*. Nucleic Acids Res, 2003. **31**(24): p. 7302-10.
160. Brennan, R.G. and T.M. Link, *Hfq structure, function and ligand binding*. Curr Opin Microbiol, 2007. **10**(2): p. 125-33.
161. Sittka, A., et al., *Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq*. PLoS Genet, 2008. **4**(8): p. e1000163.
162. Padalon-Brauch, G., et al., *Small RNAs encoded within genetic islands of Salmonella typhimurium show host-induced expression and role in virulence*. Nucleic Acids Res, 2008. **36**(6): p. 1913-27.
163. Link, T.M., P. Valentin-Hansen, and R.G. Brennan, *Structure of Escherichia coli Hfq bound to polyriboadenylate RNA*. Proc Natl Acad Sci U S A, 2009. **106**(46): p. 19292-7.
164. Lorenz, C., et al., *Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts*. Nucleic Acids Res, 2010. **38**(11): p. 3794-808.
165. Otaka, H., et al., *PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action*. Proc Natl Acad Sci U S A, 2011. **108**(32): p. 13059-64.
166. Horstmann, N., et al., *Structural mechanism of Staphylococcus aureus Hfq binding to an RNA A-tract*. Nucleic Acids Res, 2012. **40**(21): p. 11023-35.
167. Zeng, Q. and G.W. Sundin, *Genome-wide identification of Hfq-regulated small RNAs in the fire blight pathogen Erwinia amylovora discovered small RNAs with virulence regulatory function*. BMC Genomics, 2014. **15**(1): p. 414.
168. Reichelt, R., et al., *The archaeal Lsm protein from Pyrococcus furiosus binds co-transcriptionally to poly(U)-rich target RNAs*. Biol Chem, 2023. **404**(11-12): p. 1085-1100.
169. Guisbert, E., et al., *Hfq modulates the sigmaE-mediated envelope stress response and the sigma32-mediated cytoplasmic stress response in Escherichia coli*. J Bacteriol, 2007. **189**(5): p. 1963-73.
170. Diestra, E., et al., *Cellular electron microscopy imaging reveals the localization of the Hfq protein close to the bacterial membrane*. PLoS One, 2009. **4**(12): p. e8301.
171. Andrade, J.M., et al., *The RNA-binding protein Hfq is important for ribosome biogenesis and affects translation fidelity*. EMBO J, 2018. **37**(11).
172. Sharma, I.M., A. Korman, and S.A. Woodson, *The Hfq chaperone helps the ribosome mature*. EMBO J, 2018. **37**(11).

173. Figueroa-Bossi, N., et al., *Loss of Hfq activates the sigmaE-dependent envelope stress response in Salmonella enterica*. Mol Microbiol, 2006. **62**(3): p. 838-52.
174. Johansen, J., et al., *Conserved small non-coding RNAs that belong to the sigmaE regulon: role in down-regulation of outer membrane proteins*. J Mol Biol, 2006. **364**(1): p. 1-8.
175. Wang, M.C., et al., *The RNA chaperone Hfq is involved in stress tolerance and virulence in uropathogenic Proteus mirabilis*. PLoS One, 2014. **9**(1): p. e85626.
176. Vogt, S.L. and T.L. Raivio, *Hfq reduces envelope stress by controlling expression of envelope-localized proteins and protein complexes in enteropathogenic Escherichia coli*. Mol Microbiol, 2014. **92**(4): p. 681-97.
177. Deng, Y., et al., *The RNA Chaperone Hfq Is Involved in Colony Morphology, Nutrient Utilization and Oxidative and Envelope Stress Response in Vibrio alginolyticus*. PLoS One, 2016. **11**(9): p. e0163689.
178. Gottesman, S., *Trouble is coming: Signaling pathways that regulate general stress responses in bacteria*. J Biol Chem, 2019. **294**(31): p. 11685-11700.
179. Tobias, N.J., et al., *Photorhabdus-nematode symbiosis is dependent on hfq-mediated regulation of secondary metabolites*. Environ Microbiol, 2017. **19**(1): p. 119-129.
180. He, W. and R. Parker, *Functions of Lsm proteins in mRNA degradation and splicing*. Curr Opin Cell Biol, 2000. **12**(3): p. 346-50.
181. Friesen, W.J. and G. Dreyfuss, *Specific sequences of the Sm and Sm-like (Lsm) proteins mediate their interaction with the spinal muscular atrophy disease gene product (SMN)*. J Biol Chem, 2000. **275**(34): p. 26370-5.
182. Fromont-Racine, M., et al., *Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins*. Yeast, 2000. **17**(2): p. 95-110.
183. Tritschler, F., et al., *Similar modes of interaction enable Trailer Hitch and EDC3 to associate with DCP1 and Me31B in distinct protein complexes*. Mol Cell Biol, 2008. **28**(21): p. 6695-708.
184. Tang, W., et al., *Telomerase RNA biogenesis involves sequential binding by Sm and Lsm complexes*. Nature, 2012. **484**(7393): p. 260-4.
185. Mura, C., et al., *Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function*. RNA Biol, 2013. **10**(4): p. 636-51.
186. Reimer, K.A., et al., *The sole LSm complex in Cyanidioschyzon merolae associates with pre-mRNA splicing and mRNA degradation factors*. RNA, 2017. **23**(6): p. 952-967.
187. Brandmann, T., et al., *Molecular architecture of LSM14 interactions involved in the assembly of mRNA silencing complexes*. EMBO J, 2018. **37**(7).
188. Lekontseva, N.V., E.A. Stolboushkina, and A.D. Nikulin, *Diversity of LSM Family Proteins: Similarities and Differences*. Biochemistry (Mosc), 2021. **86**(Suppl 1): p. S38-S49.
189. Zhang, J., et al., *Lsm12 is an NAADP receptor and a two-pore channel regulatory protein required for calcium mobilization from acidic organelles*. Nat Commun, 2021. **12**(1): p. 4739.
190. Paya, G., et al., *Analysis of Lsm Protein-Mediated Regulation in the Haloarchaeon Haloferax mediterranei*. Int J Mol Sci, 2024. **25**(1).
191. Chen, Q., Y. Chen, and Q. Zheng, *The RNA-binding protein LSM family regulating reproductive development via different RNA metabolism*. Biochim Biophys Acta Mol Basis Dis, 2025. **1871**(5): p. 167808.
192. Yang, X.C., et al., *In vitro methylation of the U7 snRNP subunits Lsm11 and SmE by the PRMT5/MEP50/pICln methylosome*. RNA, 2023. **29**(11): p. 1673-1690.
193. Albrecht, M. and T. Lengauer, *Novel Sm-like proteins with long C-terminal tails and associated methyltransferases*. FEBS Lett, 2004. **569**(1-3): p. 18-26.
194. Yokoshi, M., et al., *Direct binding of Ataxin-2 to distinct elements in 3' UTRs promotes mRNA stability and protein expression*. Mol Cell, 2014. **55**(2): p. 186-98.
195. Wang, J.Y., et al., *PolyQ-expanded ataxin-2 aggregation impairs cellular processing-body homeostasis via sequestering the RNA helicase DDX6*. J Biol Chem, 2024. **300**(7): p. 107413.
196. Scoles, D.R., et al., *Repeat Associated Non-AUG Translation (RAN Translation) Dependent on Sequence Downstream of the ATXN2 CAG Repeat*. PLoS One, 2015. **10**(6): p. e0128769.
197. Bruner, S.D., D.P. Norman, and G.L. Verdine, *Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA*. Nature, 2000. **403**(6772): p. 859-66.

198. Nash, H.M., et al., *Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily*. *Curr Biol*, 1996. **6**(8): p. 968-80.
199. Wang, Q., et al., *The Clf1p splicing factor promotes spliceosome assembly through N-terminal tetratricopeptide repeat contacts*. *J Biol Chem*, 2003. **278**(10): p. 7875-83.
200. Yan, F., et al., *RlmN and Cfr are radical SAM enzymes involved in methylation of ribosomal RNA*. *J Am Chem Soc*, 2010. **132**(11): p. 3953-64.
201. Bohnsack, K.E., C. Hobartner, and M.T. Bohnsack, *Eukaryotic 5-methylcytosine (m(5)C) RNA Methyltransferases: Mechanisms, Cellular Functions, and Links to Disease*. *Genes (Basel)*, 2019. **10**(2).
202. Haigis, M.C., E.L. Kurten, and R.T. Raines, *Ribonuclease inhibitor as an intracellular sentry*. *Nucleic Acids Res*, 2003. **31**(3): p. 1024-32.
203. Kazan, K., *The Multitalented MEDIATOR25*. *Front Plant Sci*, 2017. **8**: p. 999.
204. Maurice, F., et al., *In vitro dimerization of human RIO2 kinase*. *RNA Biol*, 2019. **16**(11): p. 1633-1642.
205. Nithianandam, V., S. Sarkar, and M.B. Feany, *Pathways controlling neurotoxicity and proteostasis in mitochondrial complex I deficiency*. *Hum Mol Genet*, 2024. **33**(10): p. 860-871.
206. Ghosh, S., et al., *RIOK2 transcriptionally regulates TRiC and dyskerin complexes to prevent telomere shortening*. *Nat Commun*, 2024. **15**(1): p. 7138.
207. D'Amora, D.R., et al., *BRAP-2 promotes DNA damage induced germline apoptosis in C. elegans through the regulation of SKN-1 and AKT-1*. *Cell Death Differ*, 2018. **25**(7): p. 1276-1288.
208. Sakai, H., et al., *BRAP2 inhibits the Ras/Raf/MEK and PI3K/Akt pathways in leukemia cells, thereby inducing apoptosis and inhibiting cell growth*. *Exp Ther Med*, 2021. **21**(5): p. 463.
209. Li, S., et al., *Identification of a novel cytoplasmic protein that specifically binds to nuclear localization signal motifs*. *J Biol Chem*, 1998. **273**(11): p. 6183-9.
210. Bakthavachalu, B., et al., *RNP-Granule Assembly via Ataxin-2 Disordered Domains Is Required for Long-Term Memory and Neurodegeneration*. *Neuron*, 2018. **98**(4): p. 754-766 e4.
211. Bravo, J., et al., *Four distinct classes of proteins as interaction partners of the PABC domain of Arabidopsis thaliana Poly(A)-binding proteins*. *Mol Genet Genomics*, 2005. **272**(6): p. 651-65.
212. Xie, J., G. Kozlov, and K. Gehring, *The "tale" of poly(A) binding protein: the MLE domain and PAM2-containing proteins*. *Biochim Biophys Acta*, 2014. **1839**(11): p. 1062-8.
213. Singh, A., et al., *Antagonistic roles for Ataxin-2 structured and disordered domains in RNP condensation*. *Elife*, 2021. **10**.
214. Jimenez-Lopez, D., J. Bravo, and P. Guzman, *Evolutionary history exposes radical diversification among classes of interaction partners of the MLE domain of plant poly(A)-binding proteins*. *BMC Evol Biol*, 2015. **15**: p. 195.
215. McCarty, J., et al., *Complete Phase Diagram for Liquid-Liquid Phase Separation of Intrinsically Disordered Proteins*. *J Phys Chem Lett*, 2019. **10**(8): p. 1644-1652.
216. Scheidt, T., et al., *Stressing the role of a short linear motif in ataxin-2 condensation*. *Mol Cell*, 2023. **83**(12): p. 1961-1963.
217. Yang, Y.S., et al., *Yeast Ataxin-2 Forms an Intracellular Condensate Required for the Inhibition of TORC1 Signaling during Respiratory Growth*. *Cell*, 2019. **177**(3): p. 697-710 e17.
218. Kato, M., et al., *Redox State Controls Phase Separation of the Yeast Ataxin-2 Protein via Reversible Oxidation of Its Methionine-Rich Low-Complexity Domain*. *Cell*, 2019. **177**(3): p. 711-721 e8.
219. Prouteau, M. and R. Loewith, *TOR Signaling Is Going through a Phase*. *Cell Metab*, 2019. **29**(5): p. 1019-1021.
220. Huelsmeier, J., et al., *A C-terminal ataxin-2 disordered region promotes Huntingtin protein aggregation and neurodegeneration in Drosophila models of Huntington's disease*. *G3 (Bethesda)*, 2021. **11**(12).
221. Wijegunawardana, D., et al., *Ataxin-2 polyglutamine expansions aberrantly sequester TDP-43 ribonucleoprotein condensates disrupting mRNA transport and local translation in neurons*. *Dev Cell*, 2025. **60**(2): p. 253-269 e5.
222. Salama, N.R., J.S. Chuang, and R.W. Schekman, *Sec31 encodes an essential component of the COPII coat required for transport vesicle budding from the endoplasmic reticulum*. *Mol Biol Cell*, 1997. **8**(2): p. 205-17.
223. Tang, B.L., et al., *Mammalian homologues of yeast sec31p. An ubiquitously expressed form is localized to endoplasmic reticulum (ER) exit sites and is essential for ER-Golgi transport*. *J Biol Chem*, 2000. **275**(18): p. 13597-604.

224. De Bigault Du Granrut, A. and J.L. Cacas, *How Very-Long-Chain Fatty Acids Could Signal Stressful Conditions in Plants?* Front Plant Sci, 2016. 7: p. 1490.
225. Lewandowska, M., A. Keyl, and I. Feussner, *Wax biosynthesis in response to danger: its regulation upon abiotic and biotic stress.* New Phytol, 2020. 227(3): p. 698-713.
226. Khan, U.M., et al., *Comparative phylogenomic insights of KCS and ELO gene families in Brassica species indicate their role in seed development and stress responsiveness.* Sci Rep, 2023. 13(1): p. 3577.
227. Li, J., et al., *BREVIS RADIX is involved in cytokinin-mediated inhibition of lateral root initiation in Arabidopsis.* Planta, 2009. 229(3): p. 593-603.
228. Xu, Y., et al., *Properties and Biotechnological Applications of Acyl-CoA:diacylglycerol Acyltransferase and Phospholipid:diacylglycerol Acyltransferase from Terrestrial Plants and Microalgae.* Lipids, 2018. 53(7): p. 663-688.
229. Sah, S.K., et al., *Physiological Functions of Phospholipid:Diacylglycerol Acyltransferases.* Plant Cell Physiol, 2024. 65(6): p. 863-871.
230. Zienkiewicz, K., et al., *Stress-induced neutral lipid biosynthesis in microalgae - Molecular, cellular and physiological insights.* Biochim Biophys Acta, 2016. 1861(9 Pt B): p. 1269-1281.
231. Holecek, M., *Serine Metabolism in Health and Disease and as a Conditionally Essential Amino Acid.* Nutrients, 2022. 14(9).
232. Suzuki, K., et al., *Formate dehydrogenase, an enzyme of anaerobic metabolism, is induced by iron deficiency in barley roots.* Plant Physiol, 1998. 116(2): p. 725-32.
233. Igamberdiev, A.U. and L.A. Kleczkowski, *Corrigendum: The Glycerate and Phosphorylated Pathways of Serine Synthesis in Plants: The Branches of Plant Glycolysis Linking Carbon and Nitrogen Metabolism.* Front Plant Sci, 2018. 9: p. 984.
234. Fernandez-Bernal, A., et al., *Mission cholesterol: Uncovering its hidden role in ALS neurodegeneration.* Biochim Biophys Acta Mol Basis Dis, 2025. 1871(8): p. 168021.
235. Griffiths, W.J. and Y. Wang, *Oxysterol research: a brief review.* Biochem Soc Trans, 2019. 47(2): p. 517-526.
236. Kim, B.Y. and B.R. Jin, *Apolipoprotein III from honeybees (Apis cerana) exhibits antibacterial activity.* Comp Biochem Physiol B Biochem Mol Biol, 2015. 182: p. 6-13.
237. Maravilla, E., et al., *Apolipoprotein III interaction with phosphatidylglycerol and lipopolysaccharide: A potential mechanism for antimicrobial activity.* Chem Phys Lipids, 2020. 229: p. 104909.
238. Alencar, M.B., et al., *The role of l-serine and l-threonine in the energy metabolism and nutritional stress response of Trypanosoma cruzi.* mSphere, 2025. 10(3): p. e0098324.
239. Wiebe, M.A., et al., *Serine Deamination Is a New Acid Tolerance Mechanism Observed in Uropathogenic Escherichia coli.* mBio, 2022. 13(6): p. e0296322.
240. Sassa, T. and A. Kihara, *Metabolism of very long-chain Fatty acids: genes and pathophysiology.* Biomol Ther (Seoul), 2014. 22(2): p. 83-92.
241. Kim, J.G. and L.D. Hudson, *Novel member of the zinc finger superfamily: A C2-HC finger that recognizes a gliosis-specific gene.* Mol Cell Biol, 1992. 12(12): p. 5632-9.
242. Nielsen, J.A., et al., *Myelin transcription factor 1 (Myt1) modulates the proliferation and differentiation of oligodendrocyte lineage cells.* Mol Cell Neurosci, 2004. 25(1): p. 111-23.
243. Lin, F., T.J. Ribar, and A.R. Means, *The Ca²⁺/calmodulin-dependent protein kinase kinase, CaMKK2, inhibits preadipocyte differentiation.* Endocrinology, 2011. 152(10): p. 3668-79.
244. Kobayashi, T., et al., *HDL promotes adiponectin gene expression via the CAMKK/CAMKIV pathway.* J Mol Endocrinol, 2022. 68(2): p. 89-98.
245. Lin, C., et al., *Cholesterol metabolism regulated by CAMKK2-CREB signaling promotes castration-resistant prostate cancer.* Cell Rep, 2025. 44(6): p. 115792.
246. Zheng, M., et al., *Inactivation of Rheb by PRAK-mediated phosphorylation is essential for energy-depletion-induced suppression of mTORC1.* Nat Cell Biol, 2011. 13(3): p. 263-72.
247. Kostenko, S., et al., *Physiological roles of mitogen-activated-protein-kinase-activated p38-regulated/activated protein kinase.* World J Biol Chem, 2011. 2(5): p. 73-89.
248. Tang, J., et al., *PRAK interacts with DJ-1 and prevents oxidative stress-induced cell death.* Oxid Med Cell Longev, 2014. 2014: p. 735618.

249. Dancik, G.M., L. Varisli, and S.A. Vlahopoulos, *The Molecular Context of Oxidant Stress Response in Cancer Establishes ALDH1A1 as a Critical Target: What This Means for Acute Myeloid Leukemia*. Int J Mol Sci, 2023. **24**(11).
250. Hellgren, M., et al., *Alcohol dehydrogenase 2 is a major hepatic enzyme for human retinol metabolism*. Cell Mol Life Sci, 2007. **64**(4): p. 498-505.
251. Godi, A., et al., *FAPPs control Golgi-to-cell-surface membrane traffic by binding to ARF and PtdIns(4)P*. Nat Cell Biol, 2004. **6**(5): p. 393-404.
252. Masgrau, A., et al., *Distinct roles of the polarity factors Boi1 and Boi2 in the control of exocytosis and abscission in budding yeast*. Mol Biol Cell, 2017. **28**(22): p. 3082-3094.
253. Hardie, D.G., D. Carling, and M. Carlson, *The AMP-activated/SNF1 protein kinase subfamily: metabolic sensors of the eukaryotic cell?* Annu Rev Biochem, 1998. **67**: p. 821-55.
254. Sanchez-Pulido, L. and C.P. Ponting, *Hexa-Longin domain scaffolds for inter-Rab signalling*. Bioinformatics, 2020. **36**(4): p. 990-993.
255. Chafe, S.C. and D. Mangroo, *Scyl1 facilitates nuclear tRNA export in mammalian cells by acting at the nuclear pore complex*. Mol Biol Cell, 2010. **21**(14): p. 2483-99.
256. Pelletier, S., *SCYL pseudokinases in neuronal function and survival*. Neural Regen Res, 2016. **11**(1): p. 42-4.
257. Schmidt, W.M., et al., *Mutation in the Scyl1 gene encoding amino-terminal kinase-like protein causes a recessive form of spinocerebellar neurodegeneration*. EMBO Rep, 2007. **8**(7): p. 691-7.
258. Burman, J.L., et al., *Scyl1, mutated in a recessive form of spinocerebellar neurodegeneration, regulates COPI-mediated retrograde traffic*. J Biol Chem, 2008. **283**(33): p. 22774-86.
259. Pelletier, S., et al., *An early onset progressive motor neuron disorder in Scyl1-deficient mice is associated with mislocalization of TDP-43*. J Neurosci, 2012. **32**(47): p. 16560-73.
260. Schmidt, W.M., et al., *Disruptive SCYL1 Mutations Underlie a Syndrome Characterized by Recurrent Episodes of Liver Failure, Peripheral Neuropathy, Cerebellar Atrophy, and Ataxia*. Am J Hum Genet, 2015. **97**(6): p. 855-61.
261. Kuliyeve, E., et al., *Overlapping Role of SCYL1 and SCYL3 in Maintaining Motor Neuron Viability*. J Neurosci, 2018. **38**(10): p. 2615-2630.
262. Amano, G., et al., *SCYL1 arginine methylation by PRMT1 is essential for neurite outgrowth via Golgi morphogenesis*. Mol Biol Cell, 2020. **31**(18): p. 1963-1973.
263. Kaeser-Pebarnard, S., et al., *mTORC1 controls Golgi architecture and vesicle secretion by phosphorylation of SCYL1*. Nat Commun, 2022. **13**(1): p. 4685.
264. Alfaro, G., et al., *The sterol-binding protein Kes1/Osh4p is a regulator of polarized exocytosis*. Traffic, 2011. **12**(11): p. 1521-36.
265. Aravind, L. and E.V. Koonin, *Phosphoesterase domains associated with DNA polymerases of diverse origins*. Nucleic Acids Res, 1998. **26**(16): p. 3746-52.
266. Gakh, O., D.Y.t. Smith, and G. Isaya, *Assembly of the iron-binding protein frataxin in Saccharomyces cerevisiae responds to dynamic changes in mitochondrial iron influx and stress level*. J Biol Chem, 2008. **283**(46): p. 31500-10.
267. Gervason, S., et al., *Physiologically relevant reconstitution of iron-sulfur cluster biosynthesis uncovers persulfide-processing functions of ferredoxin-2 and frataxin*. Nat Commun, 2019. **10**(1): p. 3566.
268. Lill, R. and S.A. Freibert, *Mechanisms of Mitochondrial Iron-Sulfur Protein Biogenesis*. Annu Rev Biochem, 2020. **89**: p. 471-499.
269. Moss, D.K., A. Wilde, and J.D. Lane, *Dynamic release of nuclear RanGTP triggers TPX2-dependent microtubule assembly during the apoptotic execution phase*. J Cell Sci, 2009. **122**(Pt 5): p. 644-55.
270. Hadar, S., et al., *Stress-induced transcriptional readthrough into neighboring genes is linked to intron retention*. iScience, 2022. **25**(12): p. 105543.
271. Vilborg, A., et al., *Comparative analysis reveals genomic features of stress-induced transcriptional readthrough*. Proc Natl Acad Sci U S A, 2017. **114**(40): p. E8362-E8371.
272. Vilborg, A. and J.A. Steitz, *Readthrough transcription: How are DoGs made and what do they do?* RNA Biol, 2017. **14**(5): p. 632-636.
273. Caldas, P., et al., *Transcription readthrough is prevalent in healthy human tissues and associated with inherent genomic features*. Commun Biol, 2024. **7**(1): p. 100.

274. Mei, Y., et al., *Comprehensive resource for transcription readthrough events in healthy human tissues*. Sci Data, 2025. **12**(1): p. 1176.
275. Yu, F., et al., *Positive selection of a pre-expansion CAG repeat of the human SCA2 gene*. PLoS Genet, 2005. **1**(3): p. e41.
276. Chitale, G.G., S.R. Kulkarni, and S.A. Bapat, *Chimerism: A whole new perspective in gene regulation*. Biochim Biophys Acta Gen Subj, 2025. **1869**(3): p. 130767.
277. Shine, M., et al., *Co-transcriptional gene regulation in eukaryotes and prokaryotes*. Nat Rev Mol Cell Biol, 2024. **25**(7): p. 534-554.
278. Ciosk, R., M. DePalma, and J.R. Priess, *ATX-2, the C. elegans ortholog of ataxin 2, functions in translational regulation in the germline*. Development, 2004. **131**(19): p. 4831-41.
279. Lim, C. and R. Allada, *ATAXIN-2 activates PERIOD translation to sustain circadian rhythms in Drosophila*. Science, 2013. **340**(6134): p. 875-9.
280. Zhang, Y., et al., *A role for Drosophila ATX2 in activation of PER translation and circadian behavior*. Science, 2013. **340**(6134): p. 879-82.
281. Sudhakaran, I.P., et al., *FMRP and Ataxin-2 function together in long-term olfactory habituation and neuronal translational control*. Proc Natl Acad Sci U S A, 2014. **111**(1): p. E99-E108.
282. Fittschen, M., et al., *Genetic ablation of ataxin-2 increases several global translation factors in their transcript abundance but decreases translation rate*. Neurogenetics, 2015. **16**(3): p. 181-92.
283. Dansithong, W., et al., *Ataxin-2 regulates RGS8 translation in a new BAC-SCA2 transgenic mouse model*. PLoS Genet, 2015. **11**(4): p. e1005182.
284. Lastres-Becker, I., et al., *Mammalian ataxin-2 modulates translation control at the pre-initiation complex via PI3K/mTOR and is induced by starvation*. Biochim Biophys Acta, 2016. **1862**(9): p. 1558-69.
285. Lee, J., et al., *LSM12 and ME31B/DDX6 Define Distinct Modes of Posttranscriptional Regulation by ATAXIN-2 Protein Complex in Drosophila Circadian Pacemaker Neurons*. Mol Cell, 2017. **66**(1): p. 129-140 e7.
286. Hansen, M., et al., *The RNA-Binding Protein ATXN2 is Expressed during Megakaryopoiesis and May Control Timing of Gene Expression*. Int J Mol Sci, 2020. **21**(3).
287. Scoles, D.R., et al., *ALS-associated genes in SCA2 mouse spinal cord transcriptomes*. Hum Mol Genet, 2020. **29**(10): p. 1658-1672.
288. Inagaki, H., et al., *Direct evidence that Ataxin-2 is a translational activator mediating cytoplasmic polyadenylation*. J Biol Chem, 2020. **295**(47): p. 15810-15825.
289. Rounds, J.C., et al., *The disease-associated proteins Drosophila Nab2 and Ataxin-2 interact with shared RNAs and coregulate neuronal morphology*. Genetics, 2022. **220**(1).
290. Vieira de Sa, R., et al., *ATAXIN-2 intermediate-length polyglutamine expansions elicit ALS-associated metabolic and immune phenotypes*. Nat Commun, 2024. **15**(1): p. 7484.
291. Santos, C.C., et al., *Fbxo42 promotes the degradation of Ataxin-2 granules to trigger terminal Xbp1 signaling*. Nat Commun, 2025. **16**(1): p. 7523.
292. DeMille, D., et al., *PAS kinase is activated by direct SNF1-dependent phosphorylation and mediates inhibition of TORC1 through the phosphorylation and activation of Pbp1*. Mol Biol Cell, 2015. **26**(3): p. 569-82.
293. Bar, D.Z., et al., *Cell size and fat content of dietary-restricted Caenorhabditis elegans are regulated by ATX-2, an mTOR repressor*. Proc Natl Acad Sci U S A, 2016. **113**(32): p. E4620-9.
294. Bar, D.Z., C. Charar, and Y. Gruenbaum, *Small GTPases in C. elegans metabolism*. Small GTPases, 2018. **9**(5): p. 415-419.
295. Perlegos, A.E., et al., *TDP-43 impairs sleep in Drosophila through Ataxin-2-dependent metabolic disturbance*. Sci Adv, 2024. **10**(2): p. eadj4457.
296. Meierhofer, D., et al., *Ataxin-2 (Atxn2)-Knock-Out Mice Show Branched Chain Amino Acids and Fatty Acids Pathway Alterations*. Mol Cell Proteomics, 2016. **15**(5): p. 1728-39.
297. Arsovic, A., et al., *Mouse Ataxin-2 Expansion Downregulates CamKII and Other Calcium Signaling Factors, Impairing Granule-Purkinje Neuron Synaptic Strength*. Int J Mol Sci, 2020. **21**(18).
298. Bezprozvanny, I., *Role of inositol 1,4,5-trisphosphate receptors in pathogenesis of Huntington's disease and spinocerebellar ataxias*. Neurochem Res, 2011. **36**(7): p. 1186-97.

299. Sen, N.E., et al., *ATXN2 polyglutamine expansion impairs QKI-dependent alternative splicing and oligodendrocyte maintenance*. *BioRxiv*, 2025.
300. Nobrega, C., et al., *Restoring brain cholesterol turnover improves autophagy and has therapeutic potential in mouse models of spinocerebellar ataxia*. *Acta Neuropathol*, 2019. **138**(5): p. 837-858.
301. Armstrong, J., et al., *Spinocerebellar ataxia type 2 (SCA2) with white matter involvement*. *Neurosci Lett*, 2005. **381**(3): p. 247-51.
302. Rub, U., et al., *Consistent affection of the central somatosensory system in spinocerebellar ataxia type 2 and type 3 and its significance for clinical symptoms and rehabilitative therapy*. *Brain Res Rev*, 2007. **53**(2): p. 235-49.
303. Mercadillo, R.E., et al., *Parahippocampal gray matter alterations in Spinocerebellar Ataxia Type 2 identified by voxel based morphometry*. *J Neurol Sci*, 2014. **347**(1-2): p. 50-8.
304. Hernandez-Castillo, C.R., et al., *Extensive White Matter Alterations and Its Correlations with Ataxia Severity in SCA 2 Patients*. *PLoS One*, 2015. **10**(8): p. e0135449.
305. Stezin, A., et al., *In vivo microstructural white matter changes in early spinocerebellar ataxia 2*. *Acta Neurol Scand*, 2021. **143**(3): p. 326-332.
306. Tu, Y., et al., *Progressive white matter degeneration in patients with spinocerebellar ataxia type 2*. *Neuroradiology*, 2024. **66**(1): p. 101-108.
307. Sha, R., et al., *Global and Regional Brain Grey and White Matter Morphometry Alterations in Type 1, 2, and 3 Spinocerebellar Ataxias (SCAs) Patients*. *Cerebellum*, 2024. **24**(1): p. 4.
308. Velazquez-Perez, L., et al., *Central motor conduction time as prodromal biomarker in spinocerebellar ataxia type 2*. *Mov Disord*, 2016. **31**(4): p. 603-4.
309. Velazquez-Perez, L., et al., *Abnormal corticospinal tract function and motor cortex excitability in non-ataxic SCA2 mutation carriers: A TMS study*. *Clin Neurophysiol*, 2016. **127**(8): p. 2713-2719.
310. Diallo, A., et al., *Body Mass Index Decline Is Related to Spinocerebellar Ataxia Disease Progression*. *Mov Disord Clin Pract*, 2017. **4**(5): p. 689-697.
311. Rodriguez-Grana, T., et al., *Weight loss is correlated with disease severity in Spinocerebellar ataxia type 2: a cross-sectional cohort study*. *Nutr Neurosci*, 2022. **25**(8): p. 1747-1755.
312. Almaguer-Mederos, L.E., et al., *Testosterone Levels Are Decreased and Associated with Disease Duration in Male Spinocerebellar Ataxia Type 2 Patients*. *Cerebellum*, 2020. **19**(4): p. 597-604.
313. Yasukochi, Y., et al., *Identification of six novel susceptibility loci for dyslipidemia using longitudinal exome-wide association studies in a Japanese population*. *Genomics*, 2019. **111**(4): p. 520-533.
314. Sanchez-Alvarez, M., M.A. Del Pozo, and C. Bakal, *Publisher Correction: AKT-mTOR signaling modulates the dynamics of IRE1 RNase activity by regulating ER-mitochondria contacts*. *Sci Rep*, 2018. **8**(1): p. 6476.
315. Huang, S., Y. Xing, and Y. Liu, *Emerging roles for the ER stress sensor IRE1alpha in metabolic regulation and disease*. *J Biol Chem*, 2019. **294**(49): p. 18726-18741.
316. Lachance, V., et al., *Overview of Sigma-1R Subcellular Specific Biological Functions and Role in Neuroprotection*. *Int J Mol Sci*, 2023. **24**(3).
317. Zhao, Q., et al., *A mitochondrial specific stress response in mammalian cells*. *EMBO J*, 2002. **21**(17): p. 4411-9.
318. Haynes, C.M., et al., *ClpP mediates activation of a mitochondrial unfolded protein response in C. elegans*. *Dev Cell*, 2007. **13**(4): p. 467-80.
319. Gispert, S., et al., *Loss of mitochondrial peptidase Clpp leads to infertility, hearing loss plus growth retardation via accumulation of CLPX, mtDNA and inflammatory factors*. *Hum Mol Genet*, 2013. **22**(24): p. 4871-87.
320. Quiros, P.M., et al., *Multi-omics analysis identifies ATF4 as a key regulator of the mitochondrial stress response in mammals*. *J Cell Biol*, 2017. **216**(7): p. 2027-2045.
321. Bhaskaran, S., et al., *Loss of mitochondrial protease ClpP protects mice from diet-induced obesity and insulin resistance*. *EMBO Rep*, 2018. **19**(3).
322. Maletzko, A., et al., *Increased presence of nuclear DNAJA3 and upregulation of cytosolic STAT1 and of nucleic acid sensors trigger innate immunity in the ClpP-null mouse*. *Neurogenetics*, 2021. **22**(4): p. 297-312.
323. Liu, Z., et al., *Aberrant mitochondrial aggregation of TDP-43 activated mitochondrial unfolded protein response and contributed to recovery of acetaminophen induced acute liver injury*. *Toxicol Res (Camb)*, 2024. **13**(1): p. tfae008.
324. Auburger, G., J. Key, and S. Gispert, *The Bacterial ClpXP-ClpB Family Is Enriched with RNA-Binding Protein Complexes*. *Cells*, 2022. **11**(15).

325. Key, J., S. Gispert, and G. Auburger, *Knockout Mouse Studies Show That Mitochondrial CLPP Peptidase and CLPX Unfoldase Act in Matrix Condensates near IMM, as Fast Stress Response in Protein Assemblies for Transcript Processing, Translation, and Heme Production*. *Genes (Basel)*, 2024. **15**(6).
326. Batsale, M., et al., *Biosynthesis and Functions of Very-Long-Chain Fatty Acids in the Responses of Plants to Abiotic and Biotic Stresses*. *Cells*, 2021. **10**(6).
327. Barnes-Velez, J.A., F.B. Aksoy Yasar, and J. Hu, *Myelin lipid metabolism and its role in myelination and myelin maintenance*. *Innovation (Camb)*, 2023. **4**(1): p. 100360.
328. Sen, N.E., S. Gispert, and G. Auburger, *PINK1 and Ataxin-2 as modifiers of growth*. *Oncotarget*, 2017. **8**(20): p. 32382-32383.
329. Liu, L., et al., *ER stress response during the differentiation of H9 cells induced by retinoic acid*. *Biochem Biophys Res Commun*, 2012. **417**(2): p. 738-43.
330. Li, J., et al., *Involvement of endoplasmic reticulum stress in all-trans-retinal-induced retinal pigment epithelium degeneration*. *Toxicol Sci*, 2015. **143**(1): p. 196-208.
331. Liu, Y., et al., *Retinoic acid receptor-related orphan receptor alpha stimulates adipose tissue inflammation by modulating endoplasmic reticulum stress*. *J Biol Chem*, 2017. **292**(34): p. 13959-13969.
332. Chen, Y., et al., *Pin1 Downregulation Is Involved in Excess Retinoic Acid-Induced Failure of Neural Tube Closure*. *Int J Mol Sci*, 2024. **25**(11).
333. Pu, J., et al., *All-trans retinoic acid protects piglets from TGEV-induced diarrhea and intestinal epithelial apoptosis by modulating redox status and endoplasmic reticulum stress pathways*. *J Anim Sci*, 2025.
334. Yoshida, H., et al., *XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor*. *Cell*, 2001. **107**(7): p. 881-91.
335. Ali, M.M., et al., *Structure of the Ire1 autophosphorylation complex and implications for the unfolded protein response*. *EMBO J*, 2011. **30**(5): p. 894-905.
336. Eletto, D., et al., *Protein disulfide isomerase A6 controls the decay of IRE1alpha signaling via disulfide-dependent association*. *Mol Cell*, 2014. **53**(4): p. 562-576.
337. Lindholm, D., H. Wootz, and L. Korhonen, *ER stress and neurodegenerative diseases*. *Cell Death Differ*, 2006. **13**(3): p. 385-92.
338. Yoshida, H., et al., *XBP1 is critical to protect cells from endoplasmic reticulum stress: evidence from Site-2 protease-deficient Chinese hamster ovary cells*. *Cell Struct Funct*, 2006. **31**(2): p. 117-25.
339. de Mena, L., J. Lopez-Scarim, and D.E. Rincon-Limas, *TDP-43 and ER Stress in Neurodegeneration: Friends or Foes?* *Front Mol Neurosci*, 2021. **14**: p. 772226.
340. Wang, C., et al., *Dual Role of Inositol-requiring Enzyme 1alpha-X-box Binding protein 1 Signaling in Neurodegenerative Diseases*. *Neuroscience*, 2022. **505**: p. 157-170.
341. Matus, S., et al., *XBP-1 deficiency in the nervous system reveals a homeostatic switch to activate autophagy*. *Autophagy*, 2009. **5**(8): p. 1226-8.
342. Wu, R., et al., *Involvement of the IRE1alpha-XBP1 pathway and XBP1s-dependent transcriptional reprogramming in metabolic diseases*. *DNA Cell Biol*, 2015. **34**(1): p. 6-18.
343. Luo, X., et al., *Spliced or Unspliced, That Is the Question: The Biological Roles of XBP1 Isoforms in Pathophysiology*. *Int J Mol Sci*, 2022. **23**(5).
344. Wen, X.Y., et al., *Identification of c-myc promoter-binding protein and X-box binding protein 1 as interleukin-6 target genes in human multiple myeloma cells*. *Int J Oncol*, 1999. **15**(1): p. 173-8.
345. Zhou, X., H. Jiang, and J. Hou, *Coordination of upregulated XBP-1 and downregulated c-myc during myeloma cell differentiation induced by 2-methoxyestradiol*. *Leuk Res*, 2007. **31**(9): p. 1259-65.
346. Xie, H., et al., *IRE1alpha RNase-dependent lipid homeostasis promotes survival in Myc-transformed cancers*. *J Clin Invest*, 2018. **128**(4): p. 1300-1316.
347. Zhao, N., et al., *Pharmacological targeting of MYC-regulated IRE1/XBP1 pathway suppresses MYC-driven breast cancer*. *J Clin Invest*, 2018. **128**(4): p. 1283-1299.
348. Dong, H., et al., *The IRE1 endoplasmic reticulum stress sensor activates natural killer cell immunity in part by regulating c-Myc*. *Nat Immunol*, 2019. **20**(7): p. 865-878.
349. Abdullah, A., et al., *IRE1alpha is critical for Kaempferol-induced neuroblastoma differentiation*. *FEBS J*, 2019. **286**(7): p. 1375-1392.

350. Wiedemeyer, R., et al., *Ataxin-2 promotes apoptosis of human neuroblastoma cells*. *Oncogene*, 2003. **22**(3): p. 401-11.
351. Junjappa, R.P., et al., *IRE1alpha Implications in Endoplasmic Reticulum Stress-Mediated Development and Pathogenesis of Autoimmune Diseases*. *Front Immunol*, 2018. **9**: p. 1289.
352. Halbach, M.V., et al., *Atxn2 Knockout and CAG42-Knock-in Cerebellum Shows Similarly Dysregulated Expression in Calcium Homeostasis Pathway*. *Cerebellum*, 2017. **16**(1): p. 68-81.
353. Tada, M., M. Nishizawa, and O. Onodera, *Roles of inositol 1,4,5-trisphosphate receptors in spinocerebellar ataxias*. *Neurochem Int*, 2016. **94**: p. 1-8.
354. He, L., et al., *ATM blocks tunicamycin-induced endoplasmic reticulum stress*. *FEBS Lett*, 2009. **583**(5): p. 903-8.
355. Hetz, C., et al., *XBP-1 deficiency in the nervous system protects against amyotrophic lateral sclerosis by increasing autophagy*. *Genes Dev*, 2009. **23**(19): p. 2294-306.
356. Reichlmeir, M., et al., *The ataxia-telangiectasia disease protein ATM controls vesicular protein secretion via CHGA and microtubule dynamics via CRMP5*. *Neurobiol Dis*, 2024. **203**: p. 106756.
357. Mishra, P., et al., *Gene editing improves endoplasmic reticulum-mitochondrial contacts and unfolded protein response in Friedreich's ataxia iPSC-derived neurons*. *Front Pharmacol*, 2024. **15**: p. 1323491.
358. Valenzuela, V., et al., *Artificial enforcement of the unfolded protein response reduces disease features in multiple preclinical models of ALS/FTD*. *Mol Ther*, 2025. **33**(3): p. 1226-1245.
359. Shen, D., et al., *Differential neuronal vulnerability to C9orf72 repeat expansion driven by Xbp1-induced endoplasmic reticulum-associated degradation*. *Cell Rep*, 2025. **44**(4): p. 115459.
360. Grima, N., et al., *Multi-region brain transcriptomic analysis of amyotrophic lateral sclerosis reveals widespread RNA alterations and substantial cerebellum involvement*. *Mol Neurodegener*, 2025. **20**(1): p. 40.
361. Li, Y., D. Liu, and S. Li, *IRE1/Xbp1 promotes the clearance of poly(GR) dipeptide repeats in amyotrophic lateral sclerosis*. *J Biol Chem*, 2025. **301**(11): p. 110764.
362. Matsui, T., *Transcriptional regulation of a Purkinje cell-specific gene through a functional interaction between ROR alpha and RAR*. *Genes Cells*, 1997. **2**(4): p. 263-72.
363. Jetten, A.M., S. Kurebayashi, and E. Ueda, *The ROR nuclear orphan receptor subfamily: critical regulators of multiple biological processes*. *Prog Nucleic Acid Res Mol Biol*, 2001. **69**: p. 205-47.
364. Agudo, M., et al., *A retinoic acid receptor beta agonist (CD2019) overcomes inhibition of axonal outgrowth via phosphoinositide 3-kinase signalling in the injured adult spinal cord*. *Neurobiol Dis*, 2010. **37**(1): p. 147-55.
365. Chen, C.T., et al., *Loss of RAR-related orphan receptor alpha (RORalpha) selectively lowers docosahexaenoic acid in developing cerebellum*. *Prostaglandins Leukot Essent Fatty Acids*, 2020. **152**: p. 102036.
366. Petkovich, M. and P. Chambon, *Retinoic acid receptors at 35 years*. *J Mol Endocrinol*, 2022. **69**(4): p. T13-T24.
367. Chauvet, C., et al., *The gene encoding human retinoic acid-receptor-related orphan receptor alpha is a target for hypoxia-inducible factor 1*. *Biochem J*, 2004. **384**(Pt 1): p. 79-85.
368. Boukhtouche, F., et al., *Human retinoic acid receptor-related orphan receptor alpha1 overexpression protects neurones against oxidative stress-induced apoptosis*. *J Neurochem*, 2006. **96**(6): p. 1778-89.
369. Serra, H.G., et al., *RORalpha-mediated Purkinje cell development determines disease severity in adult SCA1 mice*. *Cell*, 2006. **127**(4): p. 697-708.
370. Fujita, K., et al., *Developmental YAPdeltaC determines adult pathology in a model of spinocerebellar ataxia type 1*. *Nat Commun*, 2017. **8**(1): p. 1864.
371. Watanave, M., et al., *Pharmacological enhancement of retinoid-related orphan receptor alpha function mitigates spinocerebellar ataxia type 3 pathology*. *Neurobiol Dis*, 2019. **121**: p. 263-273.
372. Ajayi, A., et al., *Expanded ataxin-7 cause toxicity by inducing ROS production from NADPH oxidase complexes in a stable inducible Spinocerebellar ataxia type 7 (SCA7) model*. *BMC Neurosci*, 2012. **13**: p. 86.
373. Sanz, A.B., et al., *Cooperation between SAGA and SWI/SNF complexes is required for efficient transcriptional responses regulated by the yeast MAPK Slt2*. *Nucleic Acids Res*, 2016. **44**(15): p. 7159-72.
374. Napierala, J.S., et al., *Reverse Phase Protein Array Reveals Correlation of Retinoic Acid Metabolism With Cardiomyopathy in Friedreich's Ataxia*. *Mol Cell Proteomics*, 2021. **20**: p. 100094.
375. Fernandes, N.D., Y. Sun, and B.D. Price, *Activation of the kinase activity of ATM by retinoic acid is required for CREB-dependent differentiation of neuroblastoma cells*. *J Biol Chem*, 2007. **282**(22): p. 16577-84.

376. Jarvis, C.I., et al., *Age-related phenotypes in the staggerer mouse expand the RORalpha nuclear receptor's role beyond the cerebellum*. Mol Cell Endocrinol, 2002. **186**(1): p. 1-5.
377. Boukhtouche, F., J. Mariani, and A. Tedgui, *The "CholesteROR" protective pathway in the vascular system*. Arterioscler Thromb Vasc Biol, 2004. **24**(4): p. 637-43.
378. Guissart, C., et al., *Dual Molecular Effects of Dominant RORA Mutations Cause Two Variants of Syndromic Intellectual Disability with Either Autism or Cerebellar Ataxia*. Am J Hum Genet, 2018. **102**(5): p. 744-759.
379. Talarico, M., et al., *RORA-neurodevelopmental disorder: A unique triad of developmental disabilities, cerebellar anomalies, and myoclonic seizures*. Genet Med, 2025. **27**(4): p. 101347.
380. Kaehler, C., et al., *Ataxin-2-like is a regulator of stress granules and processing bodies*. PLoS One, 2012. **7**(11): p. e50134.
381. Kiehl, T.R., et al., *Generation and characterization of Sca2 (ataxin-2) knockout mice*. Biochem Biophys Res Commun, 2006. **339**(1): p. 17-24.
382. Pfeffer, M., et al., *Impact of Ataxin-2 knock out on circadian locomotor behavior and PER immunoreaction in the SCN of mice*. Chronobiol Int, 2017. **34**(1): p. 129-137.
383. Kiehl, T.R., H. Shibata, and S.M. Pulst, *The ortholog of human ataxin-2 is essential for early embryonic patterning in C. elegans*. J Mol Neurosci, 2000. **15**(3): p. 231-41.
384. Gadgil, A. and K.D. Raczynska, *U7 snRNA: A tool for gene therapy*. J Gene Med, 2021. **23**(4): p. e3321.
385. Liu, J.L. and J.G. Gall, *U bodies are cytoplasmic structures that contain uridine-rich small nuclear ribonucleoproteins and associate with P bodies*. Proc Natl Acad Sci U S A, 2007. **104**(28): p. 11655-9.
386. Jacobs, E.Y., et al., *Coiled bodies preferentially associate with U4, U11, and U12 small nuclear RNA genes in interphase HeLa cells but not with U6 and U7 genes*. Mol Biol Cell, 1999. **10**(5): p. 1653-63.
387. Gadgil, A., et al., *ALS-linked FUS mutants affect the localization of U7 snRNP and replication-dependent histone gene expression in human cells*. Sci Rep, 2021. **11**(1): p. 11868.
388. Courel, M., et al., *GC content shapes mRNA storage and decay in human cells*. Elife, 2019. **8**.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.