

Review

# Intrinsic Disorder in Proteins with Pathogenic Repeat Expansions

April L. Darling<sup>1,2,\*</sup>  and Vladimir N. Uversky<sup>1,3,\*</sup> 

<sup>1</sup> Department of Molecular Medicine, College of Medicine, Byrd Alzheimer's Institute, University of South Florida, Tampa, FL 33612, USA

<sup>2</sup> James A. Haley Veteran's Hospital, Tampa, FL 33612, USA

<sup>3</sup> Institute for Biological Instrumentation of the Russian Academy of Sciences, Pushchino, Moscow Region 142290, Russia

\* Correspondence: aldarlin@mail.usf.edu (A.L.D.); vuversky@health.usf.edu (V.N.U.); Tel.: +1-813-396-9249 (A.L.D.); +1-813-974-5816 (V.N.U.)

Received: 8 November 2017; Accepted: 21 November 2017; Published: 24 November 2017

**Abstract:** Intrinsically disordered proteins and proteins with intrinsically disordered regions have been shown to be highly prevalent in disease. Furthermore, disease-causing expansions of the regions containing tandem amino acid repeats often push repetitive proteins towards formation of irreversible aggregates. In fact, in disease-relevant proteins, the increased repeat length often positively correlates with the increased aggregation efficiency and the increased disease severity and penetrance, being negatively correlated with the age of disease onset. The major categories of repeat extensions involved in disease include poly-glutamine and poly-alanine homorepeats, which are often times located in the intrinsically disordered regions, as well as repeats in non-coding regions of genes typically encoding proteins with ordered structures. Repeats in such non-coding regions of genes can be expressed at the mRNA level. Although they can affect the expression levels of encoded proteins, they are not translated as parts of an affected protein and have no effect on its structure. However, in some cases, the repetitive mRNAs can be translated in a non-canonical manner, generating highly repetitive peptides of different length and amino acid composition. The repeat extension-caused aggregation of a repetitive protein may represent a pivotal step for its transformation into a proteotoxic entity that can lead to pathology. The goals of this article are to systematically analyze molecular mechanisms of the proteinopathies caused by the poly-glutamine and poly-alanine homorepeat expansion, as well as by the polypeptides generated as a result of the microsatellite expansions in non-coding gene regions and to examine the related proteins. We also present results of the analysis of the prevalence and functional roles of intrinsic disorder in proteins associated with pathological repeat expansions.

**Keywords:** protein repeat expansion; intrinsically disordered protein; intrinsically disordered protein region; homorepeats; protein aggregation; proteinopathies

## 1. Introduction

It is clear now that the protein universe includes not only globular, transmembrane, and fibrous proteins but also intrinsically disordered proteins (IDPs) and hybrid proteins with ordered domains and intrinsically disordered protein regions (IDPRs) [1,2]. In fact, there is a growing amount of evidence supporting the idea that many protein regions, and even entire proteins, lack stable tertiary and/or secondary structure in solution, and instead exist as dynamic conformational ensembles of interconverting structures [3–7]. Although these IDPs and IDPRs fail to form specific 3D structures, they are biologically active [2,8–13].

Being typically involved in regulation, signaling and control pathways, such as those involved in the cell cycle, IDPs/IDPRs are characterized by specific functionality [14,15] that complement the functional repertoire of ordered proteins, which have evolved mainly to carry out efficient catalytic and transport functions. Some illustrative biological activities of IDPs/IDPRs include regulation of cell division, transcription and translation, signal transduction, protein phosphorylation and other posttranslational modifications, storage of small molecules, chaperone action, and regulation of the self-assembly of large multi-protein complexes such as the ribosome [2,9,11,14–25]. As a matter of fact, the lack of rigid globular structure under physiological conditions provides IDPs/IDPRs with a remarkable set of unique functional advantages [26–31]. For example, large conformational plasticity allows IDPs/IDPRs to be promiscuous binders and interact efficiently with multiple different targets [2,12,32,33]. The functional importance of being disordered has been intensively analyzed and systemized in several dedicated reviews [2,10,18,24,34]. Many IDPs/IDPRs are known to undergo a disorder-to-order transition upon functioning [10,24,34,35]. When IDPRs bind to signaling partners, the free energy required to bring about the disorder to order transition takes away from the interfacial, contact free energy, with the net result that a highly specific interaction can be combined with a low net free energy of association [10,36]. High specificity coupled with low affinity seems to be a useful pair of properties for a signaling interaction so that the signaling interaction is reversible. In addition, a disordered protein can readily bind to multiple partners by changing shape to associate with different targets [32,33]. In addition to decoupled high binding specificity and low affinity, disorder has several clear advantages for functions in signaling, regulation, and control [2,8,10,11,13,17–20,25,37–39].

Many IDPs/IDPRs are involved in the pathogenesis of numerous human diseases (disorders), giving rise to the “disorder in disorders” or D<sup>2</sup> concept [40]. According to this concept, involvement of IDPs/IDPRs in the development of various diseases is defined by the unique structural and functional properties of these representatives of protein universe. Such diseases originate not only from the misfolding of causative IDPs/IDPRs but also from their misidentification, misregulation, and missignaling. Among human maladies associated with misbehavior of IDPs/IDPRs are various neurodegenerative diseases [41–43].

The absence of ordered structure in IDPs/IDPRs proteins has been associated with some specific features of their amino acid sequences, such as the presence of numerous uncompensated charged groups (often negative); i.e., a high net charge at neutral pH, which is a result of the extreme pI values in such proteins and a low content of hydrophobic amino acid residues [9]. In fact, IDPs/IDPRs were shown to be significantly depleted in order-promoting amino acid residues, such as Trp, Tyr, Phe, Ile, Leu, Val, Cys, and Asn, and enriched in disorder-promoting amino acids, such as Ala, Arg, Gly, Gln, Ser, Glu, Lys, and Pro [10,44–46]. Furthermore, IDPs/IDPRs are typically characterized by low sequence complexity and might contain numerous sequence repeats [37,47–49].

Amino acid tandem repeats are abundant in eukaryotes, where they are found within many physiologically important proteins playing important roles in protein functionality [50,51]. In general, the length of units in protein repeats can be different, with a special type being repeats with the repetitive unit equals to one amino acid residue (homorepeats). Repeats, especially those with poly-glutamine and poly-alanine sequences, are overrepresented in DNA-binding proteins and transcription factors [50], which, compared to the average protein, are engaged in more protein-protein interactions [50,52]. It is likely that proteins utilize homorepeats for non-specific protein-protein interactions or in modulation of specific interactions [53]. However, although homorepeats are engaged in important physiological interactions, they are prone for expansion via strand slippage of encoding DNA, which can lead to increased aggregation propensity of resulting proteins and subsequent disease. It was also pointed out that among various types of the protein repeats, the homorepeats have the highest tendency to aggregate [54].

Curiously, several human diseases are associated with pathological expansions of repeated sequences in specific proteins [55,56]. Most known homorepeat expansions that correlate with disease are found in exons, where translation occurs in a canonical fashion leading to long stretches of peptide homorepeats [55,57]. In most homorepeat disorders, length of the causing homorepeat expansion is correlated with disease severity and penetrance. Genetically, homorepeat disorders are characterized by the lack of conventional Mendelian transmission and exhibit a phenomenon called anticipation, where the following generation is likely to inherit a longer repeat than the previous one, and this results in increased disease severity with earlier onset [58–60]. This is because at the genetic level, the expanded repeats are characterized by meiotic or intergenerational instability and change in size when transmitted from parents to offspring [61]. On the other hand, mitotic instability or somatic mosaicism of such homorepeats defines their size variation within the tissues of an affected individual [61].

In rarer cases, the repeat expansions occur in non-coding regions that normally are not translated. However, some expansions have been shown to lead to unique peptide species through non-canonical mechanisms of translation [62]. In fact, these regions, like those found in exons, commonly undergo frame-shifts that lead to the production of several repeat peptide species [62]. This leads to the increased complication that the interaction between different translation products must be considered because biochemical interactions between different species could alter aggregation propensity, depending on the species, and could cause the peptides to undergo structural changes.

The other correlative feature of peptide repeat length is structure, with the longer repeats generally corresponding to increased disorder [63]. These long disordered stretches of repeat peptides have a propensity to aggregate, which contrasts with shorter peptides that remain diffuse throughout the cell. The appearance of large aggregates leads to a cascade of cellular events that causes toxicity and cell death. Cell death from aggregated repeat peptides is emerging as a common feature in many neurological and developmental diseases but the mechanism that causes it is not yet known.

Proteins affected by repeat expansion mutations are listed in Table 1. The goal of this work is to systematically examine proteins associated with proteinopathies caused by the poly-glutamine and poly-alanine homorepeat expansions, as well as by the polypeptides generated as a result of the microsatellite expansions in non-coding gene regions, with the major focus on the roles of intrinsic disorder in proteins associated with pathological repeat expansions. Here, we first discuss molecular mechanisms of various pathologies associated with repeat expansion and represent related proteins. Next, we represent the results of a systematic analysis of the intrinsic disorder status and the presence of disorder-based functional features (such as sites of posttranslational modifications and disorder-based protein binding sites, known as molecular recognition features, MoRFs) in proteins with pathological repeat expansions.

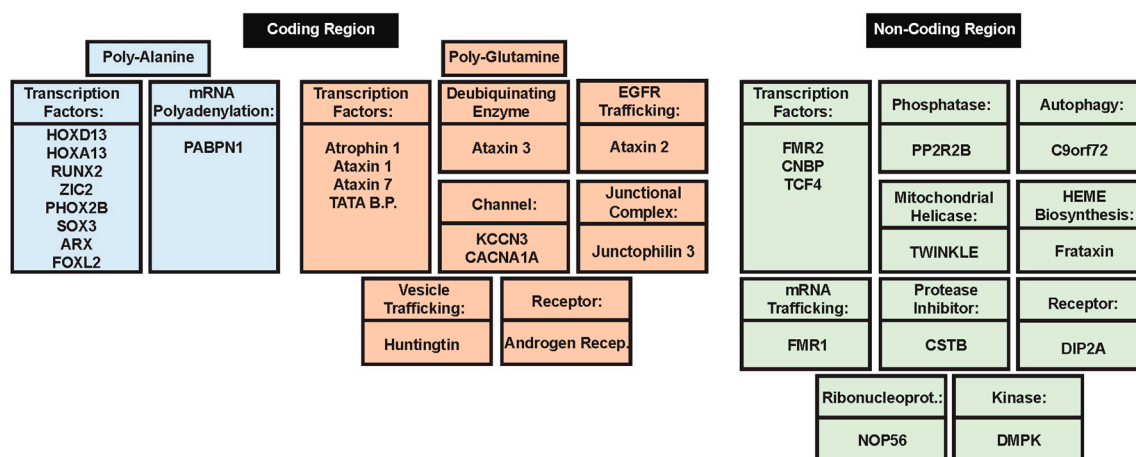
**Table 1.** Major characteristics of genes with pathological repeat expansions and proteins they encode.

	Repeat Location	Gene	Disease <sup>a</sup>	Repeat Sequence	WT Length	Pathogenic Length <sup>b</sup>	% Disorder <sup>c</sup>	References
Poly-alanine	Exon	<i>HOXD13</i>	SPD II	GCG	15	>21	33.82	[64–66]
	Exon	<i>HOXA13</i>	HFGS	GCG	12	>17	34.28	[67]
	Exon	<i>RUNX2</i>	CCD	GCG	17	>26	62.96	[65,68]
	Exon	<i>ZIC2</i>	HPE	GCG	9	–	54.89	[69,70]
	Exon	<i>PHOX2B</i>	CCHS	GCG	20	–	54.15	[71–73]
	Exon-X chrom.	<i>SOX3</i>	XLMR + GHD	GCG	15	>25	51.12	[74,75]
	Exon-X chrom.	<i>ARX</i>	XLMR	GCG	16	>17, >22	59.07	[76,77]
	Exon	<i>FOXL2</i>	BPEIS	GCG	14	>21, >24	47.34	[78,79]
	Exon	<i>PABPN1</i>	OPMD	GCG	10	>11, >16	59.80	[80,81]
Poly-glutamine	Exon	<i>KCNN3</i>	Schizo. <sup>d</sup>	CAG	–	–	37.50	[82]
	Exon	<i>JPH3</i>	HDL2	CAG/CTG	6 to 28	>41	51.07	[83–85]
	Exon	<i>HTT</i>	HD	CAG	6 to 35	>35	19.10	[86,87]
	Exon	<i>ATN1</i>	DRPLA	CAG	3 to 36	>48	86.05	[88,89]
	Exon	<i>AR</i>	SBMA	CAG	9 to 36	>37	54.13	[90]
	Exon	<i>ATXN1</i>	SCA1	CAG	6 to 39	>39	54.97	[91]
	Exon	<i>ATXN2</i>	SCA2	CAG	14 to 32	>33	79.13	[92,93]
	Exon	<i>ATXN3</i>	SCA3	CAG	12 to 40	>54	42.03	[82,94]
	Exon	<i>CACNA1A</i>	SCA6	CAG	4 to 18	>20	42.08	[95,96]
	Exon	<i>ATXN7</i>	SCA7	CAG	7 to 17	>33	71.30	[97,98]
	Exon	<i>TBP</i>	SCA17	CAG	25 to 42	>44	46.31	[99,100]
Non-coding	5' UTR	<i>PPP2R2B</i>	SCA12	CAG	7 to 32	>54	7.67	[101,102]
	5' UTR-X chrom.	<i>FMR1</i>	FXMR, FXTAS	CGG	6 to 55	>200, >55	38.29	[103]
	5' UTR	<i>DIP2B</i>	FRA12A MR	CGG	6 to 23	>200	19.73	[104]
	5' UTR-X chrom.	<i>FMR2</i>	FRA12A MR	GCC	–	>200	58.12	[105–107]
	5' UTR	<i>C9orf72</i>	C9ALS/FTD	GGGGCC	–	Unknown	2.70	[108,109]
	Intron	<i>FXN</i>	FRDA	GAA	7 to 22	>66	40.00	[110,111]
	Intron	<i>CNBP/zfn9</i>	DM2	CCTG	<27	>75	19.77	[112,113]
	Intron	<i>ATXN10</i>	SCA10	ATTCT	10 to 29	>279	5.26	[114,115]
	Intron	<i>NOP56</i>	SCA36	GGCCTG	3 to 8	>1500	26.26	[116]
	Intron	<i>TCF4</i>	FECD	CTG	–	>50	88.01	[117,118]
	3' UTR	<i>DMPK</i>	DM1	CTG	5 to 37	>50	14.63	[119,120]
	3' UTR	<i>ATXN8OS</i>	SCA8 <sup>e</sup>	CTG	6 to 37	>74	68.00 <sup>f</sup>	[121–123]
	Exon	<i>ATXN8</i>	SCA8 <sup>e</sup>	CAG	15 to 50	71 to 1300	100.00 <sup>f</sup>	[124]
	Promoter	<i>CSTB</i>	EPM1	CCCCGCCCCGCG	2 to 3	>14	40.82	[125,126]

<sup>a</sup> SPD II, synpolydactyly; HFGS, hand-foot genital syndrome; CCD, cleidocranial dysplasia; HPE, holoprosencephaly cephalic disorder; CCHS, congenital central hypoventilation syndrome; XLMR + GHD, X-linked mental retardation with isolated growth hormone deficiency; XLMR, ARX-nonsyndromic X-linked mental retardation; BPEIS, blepharophimosis, ptosis, and epicanthus inversus syndrome; OPMD, oculopharyngeal muscular dystrophy; Schizo., schizophrenia; HDL2, huntington's disease-like 2; HD, Huntington's disease; DRPLA, dentatorubral-pallidoluysian; SBMA, spinal and bulbar muscular atrophy; SCA, spinocerebellar ataxia; FXMR, fragile X mental retardation; FXTAS, fragile X-associated tremor/ataxia syndrome; FRA12A MR, fragile X mental retardation; FRA12A MR, fragile X mental retardation; ALS, amyotrophic lateral sclerosis; FTD, frontotemporal dementia; FRDA, Friedreich ataxia; DM, myotonic dystrophy; FECD, fuchs endothelial corneal dystrophy; EPM1, myoclonus epilepsy of Unverricht-Lundborg type, WT and pathogenic length refer to number of sequence repeats. <sup>b</sup> Pathogenic length indicates the threshold of the repeat length, above which the protein-carrier will cause development of pathology. <sup>c</sup> MobiDB-based predicted consensus disorder content is shown for query proteins (<http://mobid.bio.unipd.it/>) [127,128]. <sup>d</sup> Although CAG repeat tract length in *KCNN3* was correlated with schizophrenia, this is not a pathological repeat expansion and a cause of disease. <sup>e</sup> SCA8 is caused by the bidirectional transcription at the SCA8 locus containing *ATXN8OS* and *ATXN8* genes and therefore considered as the 'CTG\*CAG' repeat expansion disease, referring to the complementary base pairs of the *ATXN8OS* and *ATXN8* genes. <sup>f</sup> For *ATXN8OS* and ataxin-8 proteins, disorder content was calculated as an averaged value of the overall percent of residues predicted to be disordered by PONDR<sup>®</sup> VLXT, PONDR<sup>®</sup> VL3 and PONDR<sup>®</sup> VSL2.

## 2. Polyalanine Repeat Expansions

Polyalanine (poly-Ala) tract expansions are linked to 9 inherited human diseases, such as blepharophimosis-ptosis-epicanthus inversus syndrome (BPEIS), cleidocranial dysplasia (CCD), congenital central hypo-ventilation syndrome (CCHS); hand-foot-genital syndrome (HFGS), holoprosencephaly (HPE), oculopharyngeal muscular dystrophy (OPMD), synpolydactyly syndrome (SPD), X-linked mental retardation and abnormal genitalia (XLAG), and X-linked mental retardation and growth hormone deficit (XLMR + GHD) [129]. At the genetic level, the expansion of homopolymeric alanine is caused by the expansion of translated GCN trinucleotide repeats (where N refers to any of the four bases, among which GCG is significantly over-represented in the poly-Ala coding sequences and [130] in the disease-associated genes. It was emphasized that genes coding for poly-Ala stretches longer than four alanines are rather common in human genome that contains 494 such proteins possessing 604 poly-Ala domains [130]. Importantly, as shown in Figure 1, transcription factors are involved in 8 of the 9 poly-Ala expansion diseases and account for 36% of human proteins with poly-Ala tracts [131]. Poly-Ala repeats hang on the cryptic edge of toxicity depending on their length, which is directly correlated with their structure. Unlike other repeat expansions seen in disease, there is a low degree of polymorphism for poly-Ala tract repeats, most likely due to an altered mechanism of expansion as opposed to expansion of other repeats [76]. Poly-Ala tracts are thought to extend by unequal allelic homologous recombination during meiosis, while other disease relevant expansions extend due to DNA polymerase slippage during translation [70]. These observations, in addition to the fact that proteins containing poly-Ala tracts are highly conserved in mammals, makes it conceivable that diminutive extensions to alanine tracts are sufficient to cause cellular dysfunction with subsequent disease pathology [129].



**Figure 1.** Major known functions of proteins with pathogenic repeat expansions. Proteins with pathogenic expansions have varied functions depending on the type of expansion present. Poly-alanine (poly-Ala) expansions have the least variability in functions, with 8 of the 9 engaging in some sort of transcription regulation. Poly-glutamine (polyQ) expansions that cause pathology have more varied functions, but the majority participate in transcription regulation as well. Pathogenic repeats in non-coding regions occur in genes encoding proteins with the most varied functions. They include everything from catalytic proteins to receptors. Since the repeat extension occurs in the non-coding region of the gene, it is conceivable that there are not more synonymous functions among pathogenic repeat proteins in non-coding regions.

### 2.1. Molecular Mechanisms of Poly-Ala Expansion Diseases

Extension of poly-Ala tracts lead to structural changes in the repeat peptide that may indicate mechanisms of disease pathogenesis. Shorter length peptides are predominantly disordered with short  $\alpha$ -helical sections, which, upon expansion, become more prevalent [132]. Increased length



tends to push the peptides towards either the degradation or aggregation pathway depending on repeat number, with the longest tending towards aggregation due to length dependent formation of stable beta sheets that are resistant to degradation [68]. In fact, poly-Ala repeat peptides involved in disease have been shown to form large intracellular inclusions of the aggregated peptides. These large inclusions cause cellular dysfunction; however, this is not observed unless their length exceeds 19 repeats [65,76]. Because poly-Ala extensions cause misfolding and subsequent protein aggregation, they are now grouped into a growing class of maladies termed as protein misfolding diseases.

All poly-Ala tract expansions involved in disease function in the nucleus as transcription factors, with one exception. The exception is PABP-2, a polyadenylate-binding protein 2 (or polyadenylate-binding nuclear protein 1) responsible for controlling the length of polyadenylate tails after mRNA processing. Because of the different functions of PABP-2 in contrast to the other 8 genes involved in poly-Ala expansion disorders, the molecular mechanism of disease pathogenesis varies. Extension of the poly-Ala containing transcription factors leads to formation of dense aggregates that mis-localize to the cytoplasm. These aggregates can contain both the mutant extended peptide as well as the wild-type peptide, indicating that the mutant can sequester the normal functioning protein thus producing a dominant effect [77,133,134].

In contrast, expansion of the poly-Ala tract in PABP-2 does not cause protein mis-localization. Instead, PABP-2 remains in the nucleus, where it can perform its normal physiological function. However, only a very small extension to the poly-Ala tract causes the protein to aggregate and appear as large nuclear inclusions. It seems that the inclusions are resistant to degradation once they reach a certain size as evidence by their co-localization with ubiquitin and proteasomal subunits as well as the increase in beta sheet structures [77,133,134]. Although there is some evidence regarding the mechanism by which poly-Ala repeats leads to disease pathogenesis, it is not well understood.

## 2.2. Genes Associated with Poly-Ala Expansion Diseases

Poly-Ala tract expansions in transcription factors are commonly associated with birth defects in humans that cause malformations of the brain, digits, limbs, and heart [64]. For example, homeobox genes involved in the regulation of patterning during limb development can contribute to developmental defects when expansions of their poly-Ala tracts occur above identified thresholds [66]. One such gene is homeobox D13 (*HOXD13*). *HOXD13* protein contains a poly-Ala tract in the first exon, which upon expansion above an 8-alanine threshold, is associated with human synpolydactyly (SPD) [64–66]. SPD is an autosomal-dominant disease resulting in limb malformations, usually including digit duplication [64,66]. The mechanism by which the expansion contributes to malformation is unclear, but the localization of the protein changes from nuclear to cytoplasmic as a result of poly-Ala tract expansion, with the cytoplasmic protein appearing as amorphous aggregates [64–66].

Another HOX gene involved in developmental abnormalities due to a poly-Ala expansion is *HOXA13*. Mutations that cause an additional 6 alanines in the poly-Ala stretch of the corresponding protein have been linked to hand-foot-genital syndrome (HFGS) [67]. HFGS is another dominantly inherited developmental defect that causes malformations in the genitourinary tract and distal limbs [66]. The expansion causes a disruption in *HOXA13* protein-protein interactions, but it is not clear exactly how developmental malformations occur [67]. Similar to the manifestations of *HOXD13* poly-Ala expansion, *HOXA13* expansion also results in mislocalization of the protein to the cytoplasm as well as misfolding when expansion occurs [65].

Homeobox genes represent an additional set of genes that are developmentally important, specifically being highly involved in neural development. Therefore, mutations associated with homeobox genes can lead to devastating neurological defects. Two such genes, *ARX* and *SOX3*, are involved in development of the central nervous system and function as transcription factors containing several poly-Ala stretches [71,72]. Expansion of the poly-Ala stretch in *SOX3* protein is associated with X-linked mental retardation with growth hormone deficiency as well as X-linked

hypopituitarism [72,74]. The latter is due to two specific poly-Ala expansions, addition of 7 or 11 alanines, with the longer being associated with a more severe disease phenotype [74].

ARX is exceedingly dangerous due to the presence of multiple poly-Ala tracts, potentiating the risk of disease due to expansion. Because ARX has multiple stretches with expansion potential, it is associated with several developmental diseases of varying severity. The most common malady arising from the aberrant poly-Ala expansion is ARX-linked mental retardation, a non-syndromic form of X-linked mental retardation; however, syndromic X-linked mental retardation is also a common outcome [77,133,134]. Other disorders that have been linked to ARX poly-alanine tract expansions include hydrocephaly with abnormal genitalia, myoclonic epilepsy with spasticity and mental retardation, Partington syndrome, and X-linked infantile spasms [77,133,134].

Two additional transcription factors associated with developmental defects due to poly-Ala expansions include RUNX2 and ZIC2. RUNX2 is a bone-specific transcription factor with a poly-alanine tract [77,133,134]. When expansion occurs above a threshold of 10 repeats cleidocranial dysplasia, a bone developmental defect, occurs [68,131]. ZIC2 is a zinc finger transcription repressor that has been extensively linked to holoprosencephaly, a developmental defect where the midline of the brain does not properly form and therefore there is no separation of the brain into hemispheres [73,131]. Similar to other poly-Ala expansion disorders, expansion in these two transcription factors causes protein mislocalization, misfolding, and aggregation [68].

Development and regulation of the autonomic nervous system occur by various transcription factors in addition to homeobox genes. PHOX2B is an example that is involved in autonomic nervous system development [72]. PHOX2B contains a 20-residue C-terminal poly-alanine tract that, upon expansion, can result in congenital central hypoventilation syndrome (CCHS) [72,73]. CCHS is a disorder that effects breathing due to a disruption in autonomic nervous system regulation, which is especially problematic when sleeping [72]. The poly-Ala expansion in PHOX2B results in nuclear localization of the repeat peptide and defects in nuclear import [72]. Unlike in the previously mentioned poly-Ala expansion related disorders, expansion in PHOX2B causes nuclear localization of the protein as opposed to cytoplasmic, offering evidence that not all poly-Ala tract expansions in transcription factors result in similar cellular defects.

Transcription factors in the Forkhead family are known to play important roles in the maintenance of differentiated cells lines, embryogenesis, and tumorigenesis [132]. Therefore, they are commonly involved in human developmental diseases with most of them resulting in aberrant ocular manifestations [78]. One member of this family, FOXL2, is a transcription factor involved in both eye and ovary development [77]. It contains a poly-alanine stretch and upon expansion is correlated with blepharophimosis syndrome (BPES), a developmental disorder that causes defects in the ovary and eyelid [78,135]. The WT protein functions in the nucleus and appears diffuse throughout, but poly-Ala expansions cause FOXL2 to mislocalize to the cytoplasm in aggregated form and diminish its role as a transcription regulator [78].

The only gene associated with disease due to poly-Ala repeat expansions that is not a transcription factor is *PABPN1* encoding PABP-2 protein. This gene contains the shortest poly-alanine expansion associated with disease, with the addition of only 2 alanines sufficient to cause a disease phenotype [65]. Like in most other cases, the repeat length closely correlates with disease penetrance and severity, and extended stretches lead to muscle weakness with accompanying nuclear inclusions that are slow to evolve [77]. Expansions in the poly-alanine tract of PABP-2 have been linked to oculopharyngeal muscular dystrophy (OPMD), which is an adult-onset disorder marked by the presentation of progressive dysphagia, eyelid ptosis, and proximal limb weakness [81]. In addition, the skeletal muscle of those affected contains intranuclear filament inclusions that contain PABP-2 and biopsies show that inclusions are accompanied by myopathic and neurogenic changes [81]. OPMD is the only poly-Ala expansion disorder where onset occurs in adulthood and its presentation, pathology, and onset is more similar to polyglutamine expansion diseases than to other poly-Ala tract expansions [77].

PABP-2 demonstrates the ability of poly-Ala expansions to cause dysfunction even when they are not part of a transcription regulatory protein.

Therefore, while most proteins-carriers of pathogenic poly-Ala expansions behave similarly and engage in analogous functions, there are exceptions. Those exceptions demonstrate the ability for poly-Ala tract extensions to cause pathology in a multitude of ways, depending on the functionality and structure of the translated protein product. The very short extensions needed for poly-Ala tract expansion disease phenotypes to penetrate is directly correlative to structural changes in the resultant translational product. This highlights the fact that the structural transition of the protein containing the extended alanine repeat region may be the defining step that leads to cellular dysfunction and subsequent disease presentation.

### 3. Poly-Glutamine Repeat Expansions

Currently, there are at least twelve known hereditary diseases in which the expansion of a CAG repeat in the gene leads to neurodegeneration [136,137]. Table 1 shows that these poly-glutamine repeat diseases includes Huntington's disease, Huntington's disease-like 2, Kennedy disease (also known as spinal and bulbar muscular atrophy, SBMA), dentatorubral-pallidoluysian atrophy (DRPLA), spinocerebellar ataxia type 1 (SCA1), spinocerebellar ataxia type 2 (SCA2), SCA3 (also known as Machado-Joseph disease, MJD), SCA6, SCA7, SCA17, and schizophrenia. Note that although the CAG repeat tract length is somehow correlated with schizophrenia, this is not a pathological repeat expansion and a cause of disease. Similarly, expanded poly-glutamine (polyQ) tracts may occur in the case of the *JPH3* gene too, but the related HDL2 is not a typical polyQ disease. The majority of these diseases are accompanied by the progressive death of neurons, with insoluble, granular, and fibrous deposits being found in the cell nuclei of the affected neurons. The neurotoxicity in these diseases is due to the expansion of the (CAG)<sub>N</sub>-encoded polyQ repeat, which leads to the formation of amyloid fibrils and neuronal death. As a matter of fact, polyQ repeat expansions represent the most well studied group of trinucleotide repeat expansions involved in disease, with the discovery of the link between repeat expansion regions and disease being made when a polyQ expansion in the gene that encodes the androgen receptor was linked to SBMA [90,138]. The CAG trinucleotide repeat is highly unstable and therefore the repeat tract length has a high level of polymorphism across affected individuals as well as across different tissue types. Similar to poly-Ala expansions, most polyQ expansions occur in proteins that share common functions. Most disease-related proteins with polyQ expansions are involved in the regulation of neurogenesis or transcription in a DNA dependent manner [66]. In addition, most proteins with polyQ expansions are engaged in physiologically and functionally important promiscuous binding and interact with multiple partners [139]. PolyQ containing proteins have the potential to cause cellular dysfunction in a variety of ways. However, because of the multitude of functional interactions that most of such homorepeat-containing proteins participate in, their ability to aggregate in particular has several pathological implications.

#### 3.1. Molecular Mechanisms of PolyQ Repeat Expansion Diseases

PolyQ expansion diseases are considered protein misfolding diseases that arise by a toxic gain of function mechanism that is not well understood (see Figure 1). The categorization as protein misfolding disease comes from the fact that polyQ expansions are associated with highly stable  $\beta$ -rich amyloid-like protein inclusions [140]. Patients affected by the polyQ-expansion-related diseases present with polyQ-containing intracellular inclusions, which serve as a hallmark of this category of diseases [139]. These inclusions have been identified as both nuclear and cytoplasmic, and in addition to proteins with polyQ repeat expansions, contain ubiquitin, chaperone proteins, proteasome units, and various proteinaceous complexes with which the functional proteins are known to be associated [140,141]. Furthermore, these polyQ-expansion-containing proteins become resistant to degradation once they form these large inclusions.



Proteins with polyQ expanded repeats can cause pathology in several ways. First, expansion of the homorepeat region increases the probability that the polyQ containing protein will interact with itself, thereby forming pathological aggregates and deposits [142]. Additionally, when proteins with polyQ expansions aggregate they sequester other polyQ containing proteins, both of biological and pathological repeat lengths, rendering them unable to perform their biological function [143,144]. Aggregation can also cause the sequestration of non-repeat containing proteins such as molecular chaperones, which can be trapped by aggregates when unable to facilitate proper folding [145]. Since the homorepeat region is often involved in protein-protein interactions, expansion accompanied by an increased propensity to aggregate can alter the binding of biological partners and thus lead to pathology [144,145].

However, while it is simplistic to consider that these aggregates are pathological, there are a few perplexing examples of polyQ expansion diseases that cause neuronal toxicity in the absence of any visible intracellular inclusions [146]. In fact, some studies have shown that large amyloid-like inclusions of polyQ, instead of being cytotoxic, play a protective role in the cell by sequestering misfolded toxic proteins [147]. Therefore, these large aggregates may not be the toxic species, and instead the small soluble  $\beta$ -sheet rich oligomers may be the species responsible for pathology [146].

Mechanistically, the linkage of the CAG repeat expansions to cytotoxicity involves the tendency of longer polyQ sequences, regardless of protein context, to form insoluble aggregates [148–156]. Some biophysical properties of a series of simple polyglutamine peptides have been analyzed to gain information on potential mechanisms of cytotoxicity [154]. In this study, the close similarity of the far-UV CD spectra of polyQ peptides with repeat lengths of 5, 15, 28 and 44 residues to each other and to that of a polypeptide with a high degree of random coil structure suggested that the length-dependence of disease is not related to a conformational change in the monomeric states of proteins with expanded polyQ sequences [154]. However, spontaneous formation of amyloid-like fibrils was dramatically accelerated for polyQ peptides with repeat lengths exceeding 37 residues [154].

### 3.2. Genes Associated with PolyQ Expansion Diseases

The human genome contains many genes with CAG repeat stretches that are translated into polyQ repeats involved in neurogenesis, transcription factor regulation, and modulate the binding of transcription factor co-activators [157]. All polyQ containing proteins have not been linked to pathology, but several have successfully been identified in repeat expansion diseases. All polyQ expansion-related diseases are inherited in an autosomal dominant manner except for SBMA, which is X-linked. In addition, expansion of the polyQ tract does not lead to a disease phenotype unless a certain threshold repeat number is reached. Unlike in poly-Ala expansions, the pathogenic repeat length is significantly longer and has a stronger inverse relationship with disease severity, age of onset, and penetrance [61].

The most well studied polyQ expanded gene is HTT, which upon expansion of the homopeptide region, is responsible for the pathogenesis associated with Huntington's disease. Huntington's disease is a dominantly inherited motor neuron disease that generally affects middle aged adults but can also present as early-onset and in a juvenile-form if repeat lengths exceeds 70. The decline in motor functioning generally begins with chorea and progresses over an average period of 10–15 years [148]. Huntington's disease ultimately results in death, most commonly from bulbar dysfunction and its related complications [148].

The gene product from *HTT* is the Huntingtin protein, a highly interactive protein. Huntingtin contains several hydrophobic alpha-helices responsible for the mediation of several protein-protein interactions [86]. Using a yeast two-hybrid system, it was shown that Huntingtin directly interacts with 186 other proteins in its interaction network [87]. Because of the hydrophobic nature of many structural features in Huntingtin and its physiologically important promiscuous binding propensity, long extensions in its N-terminus tend to be poorly tolerated by the cell.

Huntington's Disease Like-2's (HDL2) only known genetic link comes from a CAG expansion in the *Junctophilin-3* gene, which is detected in 100% of cases [83,84]. The expansion is inherited in an autosomal dominant manner and can be traced back to Africa [158]. When the expansion exceeds the threshold of 50 repeats, it results in disease [159]. HDL2's clinical manifestations are very similar to those seen in Huntington's disease, but motor and cognitive symptoms are more variable between patients [158,159]. The disease generally presents itself with chorea, and its progression results in fatality after an average of 15–20 years, although this is inversely correlated with repeat length [84,87].

Junctophilin-3 is normally expressed in the brain and functions to enable the establishment of a junctional complex established between the cytoplasmic membrane and endoplasmic reticulum membrane [86]. The exact mechanism by which the repeat expansion leads to cell death is not known; however there is a decrease in the expression of Junctophilin-3 when the mutation occurs, suggesting that haploinsufficiency of the gene may be key to driving disease pathogenesis in HDL2. The expression is down-regulated when the expansion occurs due to the sequestration of the wild type version of the gene into aggregates of the mutant protein [83].

*Atrophin 1 (ATN1)* is a gene coding transcription repressor, that upon expansion of a CAG repeat, leads to dentatorubral and pallidoluysian atrophy (DRPLA). DRPLA is a neurodegenerative disease characterized by epilepsy, cerebral ataxia, dementia, chorea, and myoclonus [160]. Healthy individuals have a repeat length between 7–23 in ATN1, but when the expansion exceeds 48, disease ensues [160]. There is an inverse correlation between repeat size and age of onset and disease progression [160].

The gene that encodes the androgen receptor has been convincingly linked to spinal and bulbar muscular atrophy (SBMA). SBMA is a slowly progressing motorneuron disease possessing X-linked inheritance; therefore, only males are affected. Disease presentation includes muscle weakness and atrophy, gynecomastia, testicular atrophy, reduced fertility, and mild androgen insensitivity [90,161,162]. The X-linked inheritance of SBMA stems from the fact that disease is genetically linked to a CAG expansion in the *AR* gene (encoding the androgen receptor) located on the proximal arm of the X-chromosome. The polyQ expansion is located on the amino-terminus of the androgen receptor and upon elongation causes the translated protein to assume an altered structure going from an unfolded state to a stable beta sheet structure [90,161,162]. The change in the structure of the protein is believed to be in favor of the rate limiting structure of aggregation, a soluble oligomer capable of seeding additional aggregation reactions [162]. In fact, histological staining has revealed large insoluble fibers present in the nucleus of SBMA patients [163]. In the human androgen receptor, there are three polyglutamine repeats ranging in size from 5 to 22 residues, stretches of seven prolines and five alanines, and a polyglycine repeat of 24 residues. Polymorphisms of the largest polyglutamine and the polyglycine repeats of this protein were found in a number of human diseases, such as prostate cancer, benign prostatic hyperplasia, male infertility, and rheumatoid arthritis [164].

The seven remaining genes that undergo disease causing expansion of their polyQ regions are all involved in spinocerebral ataxia (SCA), but different genes lead to different forms of SCA. SCA is a dominantly inherited disorder with the primary feature being ataxia, which involves problems with balance, speech, and eye movements. There have been 40 characterized SCAs which differ in age of onset and disease presentation, and so far, 28 have been genetically linked [165].

SCA1, 2, 3, and 17 are all linked to expansions in the coding regions of related *ATXN* genes. SCA1, 3, and 17 are all associated with *ATXN* genes that specifically encode for a nuclear version of the ataxin protein, while SCA2 is associated with a cytoplasmic ataxin protein [140]. The cytoplasmic protein is referred to as ataxin 2 and causes a version of ataxia that resembles Parkinson's disease but has associated eye degeneration as well [93]. SCA1 is specifically caused by a CAG expansion in *ATXN1* [166]. It causes peripheral neuropathy and hypometric cascades [97,167]. CAG expansions in *ATXN3* lead to SCA3, which is characterized by peripheral neuropathy and ophthalmic change [97]. Expansions in the CAG repeat region in *ATXN7* cause SCA7. SCA7 is mainly a degenerative eye disorder characterized by retinal degeneration with associated visual loss [167]. All three genes involved in the above listed in the nuclear ataxin associated SCAs are involved in functions where

DNA binding is necessary, and upon expansion their localization changes from nuclear to cytoplasmic, meaning the protein is no longer able to perform its wild type functions.

Another nuclear protein involved in spinocerebellar ataxia that undergoes mis-localization to the cytoplasm upon expansion is the TATA-box-binding protein, TBP [97]. This protein is encoded by the *TBP* gene, in which expansion of either CAG or CAA repeat regions (both codes for the polyQ tracts in the corresponding protein) leads to SCA17 [141]. SCA17 is an ataxia with symptoms ranging from involuntary movements and dementia to psychosis [97]. It can present in children in their first two years and lead to developmental delays and an early death [163].

Not only are DNA-binding proteins involved in ataxia upon extension of CAG repeats, but certain phosphatases and channel proteins are also engaged. For example, CAG repeat expansions in the gene *CACNA1A* that encodes for the P/Q voltage-dependent calcium channel can cause SCA6 [168]. These expansions are generally very small and lead to a very slowly progressing disease that is presented as pure ataxia and occurs for the lifetime of the patient [157,168]. Similarly, CAG long repeat expansions in the *KCNN3* gene that affect the N-terminal region of the small conductance calcium-activated potassium channel KCNN3 (also known as hKCa3 or SK3) might be related to the pathology of schizophrenia and bipolar disorder [169–172]. Curiously, it was also reported that the polymorphism of schizophrenia symptom can be associated with both the CAG repeat numbers and the difference in allele sizes [82,170].

One of the general trends found in many polyQ extension-related diseases is a noticeable correlation between the number of CAG repeats and the probability of disease onset. For example, in the 3142-residue-long huntingtin, polyQ repeat encoded by the CAG repeat expansion of the exon 1 varies between 16 and 37 residues in healthy individuals, whereas patients with Huntington's disease have repeats of >38 glutamine residues [173]. Similarly, the age of onset and the severity of the progression of SCA1 are both directly linked to the length of the polyQ tract in ataxin-1 [174–176], with the length of the polyQ tract exceeding a threshold of 39–44 glutamine residues being associated with the formation of granular or fibrillar intranuclear aggregates of ataxin-1 and eventual cell death [177,178]. In SBMA, which is associated with the polyQ expansion of the androgen receptor, healthy individuals have a polyQ segment of 15 to 31 residues, whereas the SBMA afflicted individuals have 40–62 glutamine residues [179]. Finally, the age of onset of the DRPLA is inversely correlated with the length of the polyQ track repeat size in atrophin-1, which varies from 7–23 in normal individuals and is expanded to 49–75 in DRPLA patients [88].

PolyQ-related diseases have received the most attention among any repeat expansion disease due to a higher prevalence in the population. These diseases, while considered protein misfolding diseases, have many possible mechanisms that lead to cell pathology. However, there is a common theme of loss of function of the wild type protein in which the extension occurs. Because of the complex nature of the possibly toxicity mechanisms created by polyQ extensions, much more work needs to be done to understand their role in disease to develop effective therapies for those suffering with disease.

#### 4. Non-Coding Region Repeat Expansion

Microsatellites are tandem arrays of short (usually <10 bp) units commonly found in eukaryotic genomes [180]. Microsatellite expansions in non-coding regions of genes are the most peculiar type of genetic alteration, since despite the lack of start codons, in some cases, non-canonical translation still ensues generating polypeptides with highly repeated sequences. This aberrant translation complicates the cellular mechanisms of pathology by throwing further insults onto an already injured cell. Generally, these polypeptides are involved in a gain of function toxicity, and their translation is often correlated with promoter methylation of the gene they are located on, rendering it inactive. Therefore, expansions in the noncoding region of genes are especially dangerous and lead to several developmental and neurological diseases.

#### 4.1. Molecular Mechanisms of Diseases Associated with Non-Coding Region Repeat Expansions

The mechanisms of pathology for non-coding repeat expansions are not as straight forward as expansions in coding regions since non-canonical translational processes often occur, and there is still little understood about these processes themselves. In addition, not all non-coding expansions have identified repeat peptide products associated with DNA expansions, so while it is plausible that aberrant peptides may arise in all cases, this has not been confirmed. However, for a few expansions, such as those in the 5' UTR of C9orf72, there is confirmation of peptide translation due to expansions in non-coding regions [108,109,181]. These aberrant peptide products may contribute to cell death through a gain of function toxicity.

In addition, expansions in non-coding regions of DNA lead to the transcription of long mRNA transcripts known to form stable structures that are toxic to the cell. The mRNA itself can cause damage, but the main toxicity is due to its sequestration of RNA binding proteins [55]. Once the RNA transcript has successfully sequestered the RNA binding protein, it forms nuclear foci in cells that are visible using fluorescent histology. In addition, expansions in non-coding regions often lead to loss of expression of the translated protein, which causes a loss of function toxicity [55]. It is not clear whether one of these mechanisms is suffice to cause some expansion diseases or if multiple mechanisms ensue that add insult to injury and lead to cell death [55].

#### 4.2. Genes Associated with Non-Coding Region Repeat Expansions

Many of the genes involved in disease that undergo expansions in non-coding regions contain what are known as fragile sites. These fragile sites are specific loci on chromosomes which, following partial inhibition of DNA synthesis, during metaphase, appear as visible gaps [55]. In addition, the sites are associated with the activation of the DNA damage response at stalled replication forks and are considered to create a high level of genome instability [55]. Fragile sites can be classified into two main categories depending on the frequency with which they occur in the population. They can be classified as either common fragile sites or rare fragile sites. Rare fragile sites occur in less than 5% of the population and have increased incidents of breakage that are generally associated with expansion of nucleotide repeats [55]. Rare fragile sites are then further classified by the conditions by which they are induced when in cell culture, with folate-sensitive fragile sites representing the largest group. These folate-sensitive fragile sites have a high prevalence of being located on the X-chromosome and upon expansions of nucleotide repeats, are involved in several diseases.

Fragile X Mental Retardation (FXMR), the most prevalent form of mental retardation in males, has been genetically linked to a fragile site on the X-chromosome known as FRAXA [182]. FXMR causes intellectual disabilities stemming from defects in cognitive development and learning [183]. It is generally more severe in males due to the X-linked inheritance pattern and can cause the characteristic appearance of a long face and prominent forehead and ears in those severely affected [183]. Individuals affected with this disorder also tend to engage in behavioral abnormalities such as hyperactivity, especially at adolescent age, and commonly mimic symptoms that appear in autism [182].

The FRAXA site implicated in FXMR is located on *FMR1* [183–185]. *FMR1* encodes an RNA binding protein that has both a nuclear export and import signal, implicating that it may have some role in the nuclear transport of mRNA [184]. Upon an expansion of CGG in FRAXA, the promoter region of *FMR1* is methylated leading to the loss of gene expression. The loss of expression of *FMR1* is the driving factor for disease because the expansion in the absence of downregulation does not lead to a disease phenotype [186].

Although methylation of the promoter region in *FMR1* is required for pathology, the phenotype produced from the CGG expansion is dependent on length and comes in several forms. The first form is present in healthy individuals and contains 6–40 repeats, followed by the intermediate form which has 41–60 repeats, both of which do not lead to any disease phenotypes [182,183]. When the expansion is extended to 61–200 repeats, it is called a pre-mutation and is involved in less severe diseases than FXMR [183]. The permutation is involved in Fragile X-associated tremor/ataxia syndrome,

which is a late onset disease characterized by motor degeneration and FMR1-related primary ovarian insufficiency [183,187–189].

FRA12A is another rare folate-sensitive fragile site located on the *FMR2* gene of the X-chromosome and is also associated with mental retardation [106,107,110]. The *FMR2* transcript is expressed in placenta and adult brains and is found in high levels in the fetal brain [106]. The gene is translated into a 1311-amino acid protein that is nuclear localized and possess putative transcription transactivation potential [106]. Like in *FMR1* expansions, expansion in FRA12A in *FMR2* is not sufficient to causes disease and must also accompany methylation of the promoter region of the gene and down-regulation [186]. The degree of the amplification of the GCC region is also classified in *FMR2* as either normal, permutation, or full mutation with only the full mutation leading to disease [107]. The mental retardation associated with this expansion is similar to FXMR except it is generally much milder [106].

The last rare folate-sensitive fragile site that undergoes disease causing CGG nucleotide expansions and subsequent methylation is referred to as FRA12A [104]. It is located on the 5' UTR region of *DIP2b*, which encodes a protein involved in DNA methylation [104]. Upon expansion of CGG, methylation of the promoter region of *DIP2b* occurs leading to loss of expression that results in a disease phenotype [104]. The expansion induced silencing of *DIP2b* leads to mental retardation that is milder than that seen in FXMR.

The fragile nature of the expansion sites in non-coding regions is becoming more and more evident. Recently, it was found that another expansion involved in disease located at a locus on the *FXN* gene for frataxin displays chromosomal fragility [190]. Expansions of GAA occur in the first intron of *FXN* at an Alu repeat region and lead to a reduction of GAA in the transcribed product due to the impediment of elongation during transcription [110]. This phenomenon can be exacerbated upon an increase in the repeat tract length [110]. Long expansions in frataxin lead to the most common form of ataxia, Friedreich's Ataxia (FRDA). FRDA is an autosomal recessive disease characterized by degeneration in the central and peripheral nervous system as well as the heart [111]. It is one of the more severe forms of ataxia and generally reduces mobility and causes early death most commonly through cardiac complications [111].

One of the most recently discovered repeat expansions involved in disease is a G<sub>4</sub>C<sub>2</sub> hexanucleotide repeat expansion (HRE) found on the 5' untranslated region of *C9orf72* [108,109,181]. This expansion is the major genetic cause of Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Dementia (FTD). ALS is a motor neuron disease characterized by degeneration of upper and lower motor neurons that leads to paralysis and eventually death. FTD is a neurodegenerative disorder where neuron death causes atrophy of the frontal lobe in the brain and leads to a loss in executive functioning and changes in behavior and personality. Since the discovery of the genetic link, rapid progress has been made to identify the molecular mechanism involved in disease.

The *C9orf72* (C9) HRE leads to the partial loss in functioning of the C9 protein, a multifunctional homologue of DENN proteins [191]. DENN proteins function as guanine nucleotide exchange factors for small GTPases [192]. Due to the similarity in structure of DENN proteins to C9, it is predicted that C9 engages in similar functioning, specifically acting as a guanine nucleotide exchange factor for RAB [191,192]. C9 has also been found to play a functional role in endosomal trafficking and autophagy in neurons [191,192]. HRE expansions in C9 lead to decreased expression of the translational product which, in zebrafish, leads to age dependent motor deficits [193]. In addition, the expansions are transcribed into long stretches of mRNA which form nuclear foci in cells that cause toxicity through the sequestration of RNA binding proteins [57].

The last known mechanism by which C9 expansions may lead to pathology is through a gain of function toxicity mechanism. The gain of function comes from the non-canonical repeat associated non-ATG (RAN) translation products of the C9 expansion [194,195]. These products contain tandem peptide repeats and are termed dipeptide repeat (DPR) proteins. The C9 expansion results in six DPRs, one from each of the three reading frames of the sense mRNA (poly-GA/GP/GR) and one from



each of the antisense (poly-PR/PG/PA). Note that although C9 expansions are translated into the six reading frames and although all these reading frames are utilized in protein biosynthesis, only five different DPRs are synthesized. This is because at the protein level, it is impossible to discriminate poly-GP and poly-PG generated from the sense and antisense mRNAs. The sense and antisense peptides are sometimes translated in the same cell and have been shown to cause toxicity through various mechanisms such as blocking nuclear transport and impairing the assembly of membrane-less organelles [196–200]. These aberrant peptides are newly defined products of intronic repeat expansions and offer evidence into the possibility of the translation of non-coding DNA upon expansion of repeat regions. Therefore, further studies should be performed to try to identify similar peptide products in non-coding microsatellite expansions that are implicated in disease.

Like in C9, many of the disease-causing expansions in non-coding regions lead to neuro and muscular degenerations. For example, myotonic dystrophy (DM), the most common form of muscular dystrophy that occurs in adults, is linked to DNA repeat expansions [112]. The symptoms of disease include myotonia, muscular dystrophy, cataracts, diabetes, and cardiac conduction defects [201]. DM comes in several forms, and the forms DM1 and DM2 are linked to regions of repeat extensions. DM1 and DM2 are distinguishable by the fact that DM2 is generally a later onset disease and is not present from birth like DM1 [201]. DM1 is associated with CTG expansions on the 3' end of the gene DMPK, while DM2 is genetically linked to CCTG expansions on the first intron of the gene ZNF9 [112,201]. Both diseases are inherited in a dominant manner and present with multisystemic clinical features [201].

Other form of spinocerebellar ataxias (SCA) are associated with repeats in non-coding regions as opposed to the majority which are linked to polyQ expansions. These include SCA8, 10 and 36, which are dominantly inherited and characterized by seizures, cerebellar ataxia, and anticipation [115,116,202]. SCA36 can be further classified by distinct tongue atrophy and motor neuron degeneration [202], while SCA8 commonly has oculomotor incoordination as a main symptom [123]. SCA8 was the first form of SCA that was linked to a non-CAG repeat expansion [123]. It is genetically linked to a CTG expansion of the *ATXN8OS* gene [122,123] and a complementary CAG repeat expansion in the *ATXN8* gene [124]. In fact, SCA8 is caused by the bidirectional transcription at the SCA8 locus containing *ATXN8OS* (*ataxin 8 opposite strand*) and *ATXN8* genes and is therefore considered as the 'CTG\* CAG' repeat expansion disease, referring to the complementary expanded base pairs of the *ATXN8OS* (CTG) and *ATXN8* genes (CAG) [124]. Although ataxin-8 protein encoded by the *ATXN8* gene represents a polyglutamine protein, SCA8 is not considered as typical polyQ disease. SCA10 is linked to an ATTCT expansion in the ninth intron of ataxin 10 [115]. The expansion on SCA10 does not have to be continuous to cause disease. In fact, it was shown that several repeat interruptions of varying lengths and sequences can be present in an individual expressing the disease phenotype [115]. SCA36 is linked to a GGCCCTG hexanucleotide expansion on *nop56* [202]. The disease causing repeat expansion seen on *nop56* mRNA exceeds 1500 repeats and is longer than that seen in any other neuromuscular disorder associated with repeat expansions and shows the most dramatic shift from the wild type number of 3–8 repeats to the disease-causing expansion [202].

Progressive myoclonus epilepsy (EPM1) is a mitochondrial myopathy, meaning that the cell pathology is produced from the mitochondria and without sufficient energy production, high energy consuming tissue is compromised. EPM1 is inherited as an autosomal recessive disorder and is characterized by severe, stimulus-sensitive myoclonus and tonic-clonic seizures and myoclonus that is stimulus-sensitive [203]. The disease has been genetically linked to an extension of a repeat region in the promoter region of the cystatin B gene [125]. Cystatin B, without an extension mutation, generally only contains two copies of the dodecamer repeat CCCC GCCCGCG, but extension beyond 14 repeats leads to EPM1 [125]. However, unlike in most repeat expansion diseases, the length of extension is not correlated with age of onset or severity of disease most likely due to the fact that it occurs in the promoter region [204]. This also suggests that once the repeat is beyond a critical threshold, gene expression is reduced and disease ensues at the same rate regardless of the expansion size [125].

Another repeat expansion disease linked to the non-coding region is involved in ocular degeneration, specifically Fuchs' endothelial corneal dystrophy (FECD). FECD is an inherited degenerative disease that affects the corneal endothelium which helps to maintain corneal clarity [117]. It generally is an asymptomatic disease in the early stages but later stages present with corneal edema, associated eye pain, and vision loss [117]. FECD has been genetically linked to an intronic CTG expansion in a transcription factor, namely, TCF4 [117]. Cells containing the repeat expansion contain RNA foci and have reduced expression of TCF4, both of which may contribute to cellular pathogenesis.

CAG repeat expansions in the gene *PPP2R2B*, which encodes for protein phosphatase 2, has been genetically linked to spinocerebellar ataxia type 12 (SCA12) [205]. Although CAG repeat expansion mutations are located in exon 7 of *PPP2R2B*, there is no evidence that this CAG expansion results in polyQ production [206]. In fact, it was emphasized that the CAG expansion in *PPP2R2B* has a promoter function [206], and it was also mentioned that this expansion occurs in a 5'-untranslated region of the of *PPP2R2B* gene [102]. CAG repeats numbers 7–28 in normal individuals and 55–78 in SCA12 patients [206]. SCA12 is an ataxia that most closely resembles Parkinson's disease with symptoms such as loss of movement, tremors, and dementia [90] and which is relatively rare worldwide [207].

Repeat expansion mutations found in non-coding regions are the most variable due to the diversity in the functionality of genes that involved in disease. However, like in the other expansion mutation categories, non-coding expansions still share some common traits with a few exceptions. Most of the non-coding region repeats have an inverse correlation between repeat size and disease severity and penetrance, with the exception of the cystatin B extension. Also, like the polyQ and poly-Ala repeats, most associated diseases are neurodegenerative or neuromuscular in nature. The exception to this is the TCF4 extension, which is associated with opthalmic problems; however these are generally closely associated with neurodegenerative disease presentation.

## 5. Intrinsic Disorder in Proteins Associated with Pathological Repeat Expansions

Another important feature linking various diseases associated with the pathogenic repeat expansions is the presence of noticeable disorder in the corresponding carrier proteins, even before the introduction of repeat expansion mutations. This observation is illustrated by Figure 2 that represents the PONDR<sup>®</sup> VSL2 predictor, which is one of the more accurate stand-alone tools for evaluating intrinsic disorder status in a target protein, being statistically better for proteins containing both ordered and disordered regions [208,209]. Additional information on the intrinsic disorder status and on the presence of disorder-based functional features (such as sites of posttranslational modifications and disorder-based protein binding sites, known as molecular recognition features, MoRFs) for the majority of proteins considered in this review is presented in Supplementary Materials (see Figure S1) as outputs of the D<sup>2</sup>P<sup>2</sup> database (<http://d2p2.pro/>) [210] that provides disorder evaluations by several computational tools (such as IUPred [211], PONDR<sup>®</sup> VLXT [212], PrDOS [213], PONDR<sup>®</sup> VSL2B [208,209], PV2 [210], and ESpritz [214]). These D<sup>2</sup>P<sup>2</sup> outputs of multiple disorder predictors are complemented with some important disorder-related functional information (such as location of various curated PTMs and ANCHOR-predicted disorder-based protein-protein interaction sites [215,216], known as molecular recognition features, MoRFs, see [35,217–219]). Therefore, for each protein, D<sup>2</sup>P<sup>2</sup> represents location of IDPRs predicted by various computational tools (shown by differently colored bars). Next, positions of known and predicted functional domains are indicated. This is followed by a blue-green-white bar in the middle of the plot that shows the predicted disorder agreement between nine predictors, with blue and green parts corresponding to disordered regions by consensus. Yellow bars show locations of the predicted disorder-based binding sites (molecular recognition features, MoRFs), whereas differently colored circles at the bottom of the plot show location of various PTMs. Note that D<sup>2</sup>P<sup>2</sup> information is not available for KCNN3 and CACNA1A, both associated with the polyQ expansion diseases. We also utilized the outputs of the MobiDB database (<http://mobid.bio.unipd.it/>) [127,128], for further characterization of the disorder status of query proteins. This tool was used because MobiDB generates consensus disorder scores by aggregating

the output from ten predictors, such as two versions of IUPred [211], two versions of ESpritz [214], two versions of DisEMBL [220], JRONN [221], PONDR<sup>®</sup> VSL2B [208,209,222], and GlobPlot [223]. MobiDB also has manually curated annotations related to protein function and structure derived from UniProt [224] and DisProt [225], as well as from Pfam [226] and PDB [227]. Sections below provide a brief overview of the disorder status of all 33 proteins listed in Table 1.

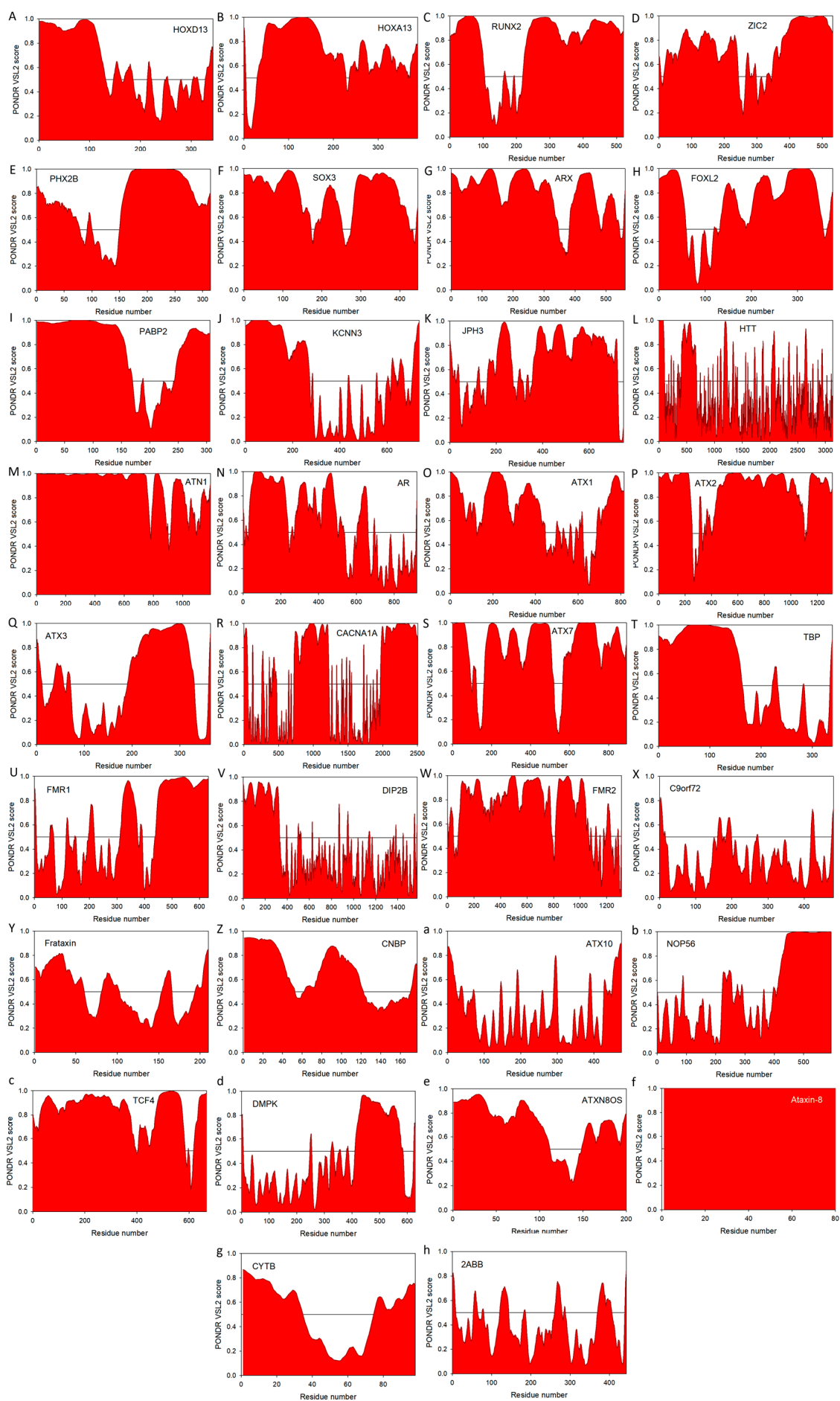
### 5.1. Disorder in Proteins Associated with Poly-Ala Expansion Diseases

Since the majority of proteins related to the poly-Ala expansion diseases are transcription factors, it is not surprising to find that they are expected to be highly disordered proteins. In fact, this is in agreement with the known notion that eukaryotic transcription factors and other proteins involved in the transcription regulation are, in general, highly disordered [228–232]. Furthermore, in addition to be associated with various poly-Ala tract extension-related diseases, many of these proteins are related to the pathogenesis of different cancers, clearly indicating important roles of these proteins in regulation of a multitude of diverse biological processes, which is another important functional characteristic of IDPs.

Figure 2A and Figure S1A show that homeobox protein HOXD13 (UniProt ID: P35453) is a highly disordered protein, with the MobiDB-based predicted consensus disorder content of 33.82%. This protein is associated not only with the poly-Ala tract expansion-based human synpolydactyly (SPD) [233], but deregulation of HOXD13 expression has been detected in breast cancer, melanoma, cervical cancer, astrocytomas, and, more recently, neoplastic tissue samples from 79 different tumor categories, being especially prominent in pancreatic cancer [234,235].

High disorder status of the homeobox protein HOXA13 (UniProt ID: P31271) is illustrated by Figure 2B and Figure S1B and is supported by MobiDB, which indicated that HOXA13 is another highly disordered protein with the consensus disorder content of 34.28%. Similar to HOXD13, HOXA13 might be related to the pathogenesis of both poly-Ala tract expansion-related hand-foot-genital syndrome (HFGS) [67,236] and some types of cancer, such as thyroid [237] and gastric cancers [238] characterized by the aberrant expression of HOXA13, both at gene and protein levels. Both HOXD13 and HOXA13 transcription factors were shown to form DNA-binding trimeric complexes with the TALE superclass proteins MEIS1A and MEIS1B [239]. Curiously, it was shown that multiple peptides derived from HOXD13 and HOXA13 can efficiently interact with MEIS proteins [239], suggesting the presence of complex HOXD13-MEIS and HOXA13-MEIS interfaces, which potentially originated as a result of folding-upon-binding reaction [33,240–242].

According to Figure 2C and Figure S1C, as well as based on the MobiDB analysis that showed the disordered residue content of 62.96%, it is clear that runt-related transcription factor 2, RUNX2 (UniProt ID: Q13950), is the most disordered protein with the pathogenic poly-Ala expansion. In addition to the poly-Ala tract (residues 73–89) this protein has a polyQ tract (residues 49–71) and a Pro/Ser/Thr-rich domain (residues 237–521). D<sup>2</sup>P<sup>2</sup> shows also that RUNX2 has a multitude of disorder-based binding sites (see Figure S1C), some of which coincide with the binding regions known to be involved in interaction with FOXO1, KAT6A, and KAT6B (residues 242–258, 336–439, and 374–468, respectively). Importantly, it was also shown that not only poly-Ala expansion, but also deletion within the poly-Ala tract (reducing its length from 17 to 11 alanines, the 11A allele) of RUNX2 might be pathogenic, being able to significantly enhance fracture risk in post-menopausal females in a site-selective manner related to intramembranous bone ossification [243].



**Figure 2.** Evaluation of intrinsic disorder propensities of 33 proteins associated with the proteins caused by the nucleotide expansions. Intrinsic disorder predisposition was evaluated by PONDR® VSL2 predictor, which is one of the more accurate stand-alone tools for prediction of the intrinsic disorder status of a target protein. This tool is known to be statistically better for proteins containing both ordered and disordered regions [208,209]. (A) Homeobox protein HOXD13 (UniProt ID: P35453); (B) Homeobox protein HOXA13 (UniProt ID: P31271); (C) runt-related transcription factor 2, RUNX2 (UniProt ID: Q13950); (D) Zinc finger protein ZIC2 (UniProt ID: O95409); (E) Paired mesoderm homeobox protein 2B (UniProt ID: Q99453); (F) Transcription factor SOX3 (UniProt ID: P41225); (G) Homeobox protein ARX (UniProt ID: Q96QS3); (H) Human FOXL2 (UniProt ID: P58012); (I) PABP2/PABPN1 (UniProt ID: Q86U42); (J) Small conductance calcium-activated potassium channel protein 3 (SK3, UniProt ID: Q9UGI6); (K) Human junctophilin-3 (JP-3, UniProt ID: Q8WXH2); (L) Human huntingtin (UniProt ID: P42858); (M) Atrophin-1 (UniProt ID: P54259); (N) Human androgen receptor (AR, UniProt ID: P10275); (O) Ataxin-1 (UniProt ID: P54253); (P) Ataxin-2 (UniProt ID: Q99700); (Q) Human ataxin-3 (UniProt ID: P54252); (R) Voltage-dependent P/Q-type calcium channel subunit  $\alpha$ 1A (CACNA1A, UniProt ID: O00555); (S) Ataxin-7 (UniProt ID: O15265); (T) TATA-box-binding protein (TBP, UniProt ID: P20226); (U) Synaptic functional regulator FMR1 (UniProt ID: Q06787); (V) Disco-interacting protein 2 homolog B (UniProt ID: Q9P265); (W) AF4/FMR2 family member 2 (UniProt ID: P51816); (X) C9orf72 (UniProt ID: Q96LT7); (Y) Frataxin (UniProt ID: Q16595); (Z) Cellular nucleic acid-binding protein (CNBP, UniProt ID: P62633); (a) Ataxin-10 (UniProt ID: Q9UBB4); (b) Nucleolar protein 56 (UniProt ID: O00567); (c) Transcription factor 4 (UniProt ID: P15884); (d) Myotonin-protein kinase (UniProt ID: Q09013); (e) Ataxin-8 (UniProt ID: Q156A1); (f) ATXN8OS protein (UniProt ID: P0DMR3); (g) Cystatin-B (UniProt ID: P04080); (h) Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit  $\beta$  isoform (PPP2R2B, UniProt ID: Q00005). In this analysis, scores above 0.5 correspond to intrinsic disorder.

Zinc finger protein ZIC2 (UniProt ID: O95409) has a MobiDB-defined disorder content of 54.89%. Figure 2D and Figure S1D illustrate that this protein has long IDPRs located in its N- and C-terminal tails. The protein is predicted to have a multitude of functional domains, possess several disorder-based binding regions, and have several PTM sites (see Figure S1D). Zic2 has multiple regions with compositional biases, such as a poly-Gly region (residues 490–508), two poly-His regions (residues 20–23 and 231–239), and four poly-Ala regions (residues 25–33, 89–97, 226–230, and 456–470). It has multiple zinc-finger domains (C2H2-type 1, 2, 3, 4, and 5, residues 256–291, 300–327, 332–357, 363–387, and 393–415, respectively) needed for transcription activation. Furthermore, a 100–255 region of ZIC2 known to be necessary for interaction with MDFIC and transcriptional activation or repression is predicted to have multiple disorder-based protein-protein interaction sites (see Figure S1D). Finally, similar to HOXD13 and HOXA13, deregulated expression of ZIC2 was shown to be associated with hepatocellular carcinoma, with this protein being required for the self-renewal maintenance of liver cancer stem cells [244].

Paired mesoderm homeobox protein 2B (PHOX2B homeodomain protein, UniProt ID: Q99453) is expected to have 54.14% disordered residues (as evaluated by the MobiDB-based consensus disorder content). In agreement with these MobiDB predictions, Figure 2E and Figure S1E show high levels of intrinsic disorder in N- and C-terminal regions of this protein. There are two poly-Ala tracts in human PHOX2B (residues 159–167 and 241–260) complemented by a poly-Gly region (residues 212–217). In addition to involvement in the poly-Ala expansion-related congenital central hypoventilation syndrome [245–247], mutations in PHOX2B are associated with neuroblastoma-2 [248,249]. Figure S1E shows that PHOX2B has several C-terminally-located disorder-based protein-protein interaction sites and also possess several PTM sites.

With its MobiDB consensus disorder score of 51.12%, transcription factor SOX3 (Sex-determining region Y-box3, UniProt ID: P41225) definitely belongs to the category of highly disordered proteins. This is further illustrated by Figure 2F and Figure S1F both showing high levels of intrinsic disorder almost evenly distributed throughout the entire protein sequence. Human SOX3 has a poly-Gly



and poly-Pro tracts (residues 129–133 and 290–294, respectively) and four poly-Ala regions (residues 234–248, 324–330, 340–347, and 353–364). The protein is predicted to have 9 MoRF regions and multiple phosphorylation sites (see Figure S1F). Besides being associated with X-linked mental retardation with growth hormone deficiency (via its poly-Ala tract expansion mutations) [250], as well as with X-linked hypopituitarism (via its over- and under-dosage) [251] and SOX3 copy number variation-related 46, XX sex reversal 3 (SRXX3) [252], SOX3 overexpression was shown to play a crucial role in tumor progression [253–257], placing this transcription factors into the oncogene category.

Homeobox protein ARX (Aristaless-related homeobox, UniProt ID: Q96QS3) is predicted by MobiDB consensus to have 59.07% disordered residues. It is not surprising since ARX has two long Ala-rich regions (residues 100–155 and 425–544), as well as a long Pro-rich region (residues 395–459) and a long Glu-rich region (residues 224–253). In fact, Figure 2G and Figure S1G indicate that intrinsic disorder is spread over the entire protein sequence and Figure S1G shows that this disorder is of functional importance, since ARX is predicted to have 8 MoRFs (two of very significant length, 44 and 157 residues), as well as several phosphorylation and methylation sites. Again, besides being related to the poly-Ala expansion-driven X-linked mental retardation [258], mutations ARX are related to agenesis of the corpus callosum in females and X-linked lissencephaly with abnormal genitalia in males [259], early infantile epileptic encephalopathy-1 [260,261], Partington syndrome [261], and X-linked lissencephaly-2 [259,262]. It was also pointed out that duplication mutation of ARX can cause benign bilateral cystic-like cavities in the cerebral and cerebellar hemispheres [263].

Figure 2H and Figure S1H show that human forkhead box protein L2 (FOXL2, UniProt ID: P58012) is predicted to have very high levels of intrinsic disorder, possessing the MobiDB consensus disorder score of 47.34%. As a matter of fact, it is unlikely that any significant part of this protein can spontaneously gain ordered structure. Instead, several regions of FOXL2 can fold at interactions with specific binding partners and this protein is shown to have multiple sites of phosphorylation, acetylation and SUMOylation (see Figure S1H). As the majority proteins discussed in this section, human FOXL2 contains a poly-Gly, a poly-Pro, and two poly-Ala stretches (residues 35–43, 284–292, 221–234, and 301–304, respectively). Besides multiple point mutations several different poly-Ala tract expansions associated with blepharophimosis, ptosis, and epicanthus inversus syndrome [264] several mutations in this proteins are directly linked to the premature ovarian failure 3 [265,266]. Furthermore, C134W mutation in FOXL2 was shown to be one of the causative agents of the adult granulosa cell tumor, one of the malignant ovarian sex cord-stromal tumors [267].

A last member of the diverse family of proteins with pathogenic poly-Ala expansions is polyadenylate-binding protein 2/polyadenine-binding protein nuclear-1 (PABP2/PABPN1, UniProt ID: Q86U42) that is predicted to be highly disordered by MobiDB (consensus disorder score of 59.80%), PONDR® VSL2 (Figure 2I) and D<sup>2</sup>p<sup>2</sup> (Figure S1I). In fact, all computational tools agree that the first 160 and the last 60 residues of PABP2/PABPN1 are expected to be disordered, whereas central region consisting of residues 170–250 should be ordered. In agreement with these predictions, X-ray structure was solved for the RNA binding domain also known as RNA recognition motif (RRM) of this protein spanning residues 167–254 (see, e.g., PDB ID: 3B4D; [268]). Figure S1I suggests that N- and C-terminally located intrinsic disorder is differently used in functionality of PABP2/PABPN1, with long N-terminal IDPR clearly serving protein-protein interaction roles (it has all 6 MoRFs found in this protein and possesses multiple phosphorylation and ubiquitination sites), whereas disordered C-tail mostly functions in PABP2/PABPN1 regulation, possessing a whole host of methylation sites. In agreement with this hypothesis, known protein interaction sites, such as regions needed for interaction with SKIP (residues 2–145), stimulation of poly(A) polymerase alpha (PAPOLA, residues 119–147) and coiled-coil-based interaction (residues 115–151), are all located within the N-terminal IDPR. Being the only non-transcription factor with the pathogenic poly-Ala tract, PABP2/PABPN1 does not form an exception from the multipathogenicity rule, being associated with oculopharyngeal muscular dystrophy via the expansion mutations in its poly-Ala tract [269] and also being involved in metastatic duodenal cancer [270] and non-small cell lung cancer [271].

## 5.2. Disorder in Proteins Associated with PolyQ Expansion Diseases

Surprisingly, despite being a multi-pass transmembrane protein (it has six transmembrane  $\alpha$ -helical regions, residues 293–313, 320–340, 371–391, 410–430, 459–479, and 528–548 and an intramembrane pore-forming region, residues 499–519), small conductance calcium-activated potassium channel protein 3 (SK3, UniProt ID: Q9UGI6) is predicted to have high levels of intrinsic disorder (it has a MobiDB consensus disorder score of 37.50%). Figure 2J shows that disorder is preferentially concentrated within the 270 N-terminal residues. Since SK3 is one of the two proteins considered in this article for which D<sup>2</sup>P<sup>2</sup> information is not available as of yet, we formed a relationship between the SK3 intrinsic disorder and function directly using the ANCHOR algorithm for prediction of the disorder-based protein-protein interaction sites [215,216]. This analysis revealed that human SK3 has six MoRFs, all located within the long disordered N-terminal tail (residues 1–29, 43–66, 86–149, 160–205, 210–228, and 235–241). The N-terminal region of SK3 has a long Q-rich region (residues 30–99) that includes two polyQ tracts (residues 30–41 and 67–85) and a Pro-rich region (residues 42–64). Furthermore, there are a polyQ and a poly-Ser regions in the C-terminal tail of the protein (residues 688–692 and 732–735, respectively). Besides being linked to schizophrenia via its polyQ tract expansion [272], SK3 is related to the pathogenesis of several types of cancer, being involved in regulation of human cancer cell migration and bone metastases [273–276].

According to the MobiDB analysis, human junctophilin-3 (JP-3, UniProt ID: Q8WXH2) has a consensus disorder score of 51.07%. This is in line with the outputs of PONDR<sup>®</sup> VSL2 (Figure 2K) and D<sup>2</sup>P<sup>2</sup> (Figure S1K), which clearly show very high levels of intrinsic disorder in this protein (especially in its C-terminal part). The N-tail of JP-3 contains a long Gly-rich region (residues 4–143), whereas an Ala-rich segment is located in the central part of this protein (residues 366–416). Furthermore, JP-3 contains a series of 8 MORN repeats (residues 15–37, 39–60, 61–82, 83–105, 107–129, 130–152, 288–310, and 311–333), which are membrane occupation and recognition nexus regions potentially involved in interaction with phospholipids and contributing to the binding of plasma membrane. The fact that there are 17 MoRFs and multiple phosphorylation sites in JP-3 (see Figure S1K) clearly shows that disorder is crucial for functionality of this protein. CAG/CTG expansion in the gene encoding junctophilin-3 is related to the Huntington's disease-like 2 pathology [158,277,278].

Figure 2L and Figure S1L show that human huntingtin (UniProt ID: P42858) is predicted to be moderately disordered. In fact, according to the MobiDB consensus analysis, the disorder content of this protein is 19.1%, with the majority of disordered regions being concentrated in its N-terminal region. Importantly, this N-terminal region with a high disorder content is a home for the polyQ expansion track. Furthermore, the N-terminus of huntingtin is known to be responsible for interaction with several nuclear proteins such as HYPA/FBP-11, which functions in pre-mRNA processing (spliceosome function) [279]; nuclear receptor co-repressor protein (NCoR) [280], which plays a role in the repression of gene activity; and p53 [281], a tumor suppressor involved in regulation of the cell cycle, and also contains multiple binding sites for other proteins with nuclear functions. The fact that huntingtin includes a PXDLS motif that serves as a binding site for the transcriptional corepressor C-terminal binding protein (CtBP) [282] suggests that this protein may also play a role in transcriptional repression. Huntingtin was shown to be a very promiscuous binder, being engaged in interaction with more than 200 proteins [283]. One of these huntingtin interactors, huntingtin yeast-two hybrid protein K (HYPK), was indeed identified as a typical IDP [283]. The major pathological involvement of huntingtin is its defining role in the Huntington's disease development, which is therefore considered as a single gene degenerative disorder [284].

The *DRPLA* gene encoding for atrophin-1 (UniProt ID: P54259) is widely expressed in brain and other tissues [89,160,285]. Although the predicted molecular mass of the atrophin-1 is 124 kDa, this protein migrates on SDS-PAGE at about 200 kDa [286], suggesting the high levels of intrinsic disorder. In agreement with these experimental data are the results of disorder prediction for this protein by multiple computational tools. For example, MobiDB suggests that this protein contains 86.05% disordered residues, whereas Figure 2M and Figure S1M also illustrate a very disordered

nature of human atrophin-1. Protein has both nuclear localization and export signals (residues 16–32 and 1033–1041, respectively) and multiple regions with composition biases: four poly-Pro regions (residues 302–305, 442–447, 509–512, and 709–712), three poly-Ser regions (376–382, 386–397, and 569–579), a Glu/Ser-rich region (residues 73–82), a poly-His region (residues 479–483), a polyQ tract (residues 484–502), an Ala/Arg-rich region (residues 807–820), and two Arg/Glu-rich regions with mixed charges (residues 821–832 and 930–939). Two regions of human atrophin-1 were established to play a role in interaction with BAIAP2 (residues 517–567) and FAT1 (residues 879–894) [287]. Figure S1M shows that almost the entire protein can be engaged in the disorder-based interactions with protein partners. Finally, atrophin-1 is heavily decorated with a multitude of different PTMs. Human atrophin-1 is related to dentatorubral-pallidoluysian atrophy via pathological expansion of its polyQ tract [288] and was also shown to be at the center of the protein-protein interaction network related to the serrated colorectal carcinoma [289].

Human androgen receptor (AR, UniProt ID: P10275) is a 919 residue-long protein that migrates in SDS-PAGE as a polypeptide with an apparent molecular weight of 110 kDa [290]. The protein can be separated on a modulating N-terminal domain (NTD) that includes functional AF1 transactivation domain (residues 142–485), a conserved centrally-located DNA-binding domain (DBD) consisting of two zinc-coordinated modules, and a C-terminally located ligand-binding domain (LBD) [290]. Human AR has a Gln-rich region (residues 58–120) that includes two polyQ tracts (residues 58–80 and 86–91), another polyQ stretch (residues 195–199), as well as a poly-Pro, a poly-Ala, and a poly-Gly region (residues 374–383, 398–404, and 451–473, respectively). All these regions are located within the NTD. According to MobiDB, AR has a consensus intrinsic disorder score of 54.13%. Intrinsic disorder is preferentially concentrated in the N-terminal half of this protein (see Figure 2N and Figure S1N). In agreement with these predictions, experimental analysis of a region of the androgen receptor N-terminal domain lacking the largest polyglutamine stretch, but containing the remaining repeats, showed that it lacked stable tertiary structure in aqueous solutions [164]. Detailed conformational studies using a combination of experimental and computational techniques revealed that the AF1 transactivation domain is in the molten globule-like conformation [291,292]. PolyQ tract expansion of AR is related to spinal and bulbar muscular atrophy X-linked 1 [293], whereas multiple mutations preferentially located within the C-terminal half of this protein are associated with several androgen insensitivity syndromes [294]. Furthermore, AR abnormalities are identified in benign prostatic hyperplasia and prostate cancer [294].

Ataxin-1 (UniProt ID: P54253) is a 815 residue-long chromatin-binding factor that repress Notch signaling [295], interacts with RNA via the 540–766 region [177] and is predicted by MobiDB to contain 54.97% disordered residues. Ataxin-1 has a self-association domain (residues 494–604) that partially overlaps with the AXH domain, a nuclear localization signal (residues 794–794) and a polyQ stretch (residues 197–225). The C-terminal region of this protein is known to interact with a ubiquitin-specific protease USP7 [296]. According to Figure 2O and Figure S1O, the N-terminal half of this protein is more disordered than its C-terminal half containing the AXH domain (residues 562–693) that is characterized by a significant sequence and structural similarity to the transcription factor HMG-box containing protein 1 (HBP1) [297]. The structure of this AXH domain is known (e.g., see PDB ID: 1OA8 [297]), which is the only structurally characterized part of ataxin-1. In fact, almost the entire 450-residue-long N-terminal domain is predicted to be mostly disordered, whereas in the C-terminal half, extensive intrinsic disorder is present in the C-tail region (residues 700–815). Figure S1O shows that both long N- and C-terminally located IDPRs contain multiple MoRFs and PTM sites. Expansion of a CAG trinucleotide repeat, which codes for glutamine in the ataxin-1, is the causing a factor of an autosomal dominant neurodegenerative disease, spinocerebellar ataxia type 1 (SCA1) [298]. Furthermore, this protein controls the epithelial-mesenchymal transition of cervical cancer cells [299].

A 1313 residue-long human ataxin-2 (UniProt ID: Q99700) is involved in EGF receptor trafficking [300]. It is predicted to have a MobiDB consensus disorder score of 79.13% and shows widespread disorder throughout the entire sequence that clearly has functional importance due to the

presence of multiple MoRFs and astonishing number of PTM sites (Figure 2P and Figure S1P). The amino acid sequence of human ataxin-2 has multiple regions with strong compositional biases, such as three Pro-rich regions (residues 47–158, 551–734, and 929–1085), a poly-Pro stretch (residues 55–64), a polyQ tract (residues 166–187) and a poly-Ser region (residues 213–223). Ataxin-2 is a highly basic protein except for one acidic region (residues 254–475) that contains two predicted globular domains, Lsm (Like Sm, amino acid 254–345) and LsmAD (Lsm-associated domain, amino acid 353–475) [301]. The LsmAD domain contains a clathrin-mediated *trans*-Golgi signal (YDS, amino acid 414–416) and an endoplasmic reticulum (ER) exit signal (ERD, amino acid 426–428). This domain is composed mainly of  $\alpha$ -helices according to the results from secondary structure prediction servers. The rest of ataxin-2 outside of the Lsm and LsmAD domains is only weakly conserved in eukaryotic ataxin-2 homologues and is predicted to be highly disordered [301]. Curiously, polyQ tract expansion in human ataxin-2 is associated with two neurodegenerative diseases, amyotrophic lateral sclerosis 13 associated with the intermediate expansions of CAG repeat (between 24 and 35 repeats) [302] and spinocerebellar ataxia 2 (SCA2) [92,93,303]. Furthermore, this protein may play a role in pathology of Parkinson's disease likely via perturbations of RNA-binding and poly(A) RNA-binding functions of several groups of proteins [304,305]. It is also related to primary open-angle glaucoma susceptibility [306], and its levels are reduced in neuroblastoma tumors with amplified MYCN [307].

Human ataxin-3 (UniProt ID: P54252) is a 364 residue-long deubiquitinating enzyme with a wide spectrum of functions related to maintenance of protein homeostasis, cytoskeleton regulation, degradation of misfolded chaperone substrates, myogenesis, and transcription [308–311]. Despite being an enzyme that possesses a catalytic Josephin domain (residues 1–180), ataxin-3 is predicted to have really high levels of intrinsic disorder. Figure 2Q and Figure S1Q show that disorder is mostly contained within the C-terminal half of protein, which is characterized by the MobiDB-based consensus disorder score of 42.03%. Analysis of human ataxin-3 by a multitude of biophysical and biochemical techniques supported results of these computational analyses and suggested that this protein is indeed composed of a structured N-terminal domain followed by a flexible tail [312]. The Josephin domain is highly conserved from nematodes to human and is also found in plants [313]. The intrinsically disordered C-tail is non-conserved contains long stretches of low complexity regions [313], including a polyQ tract (residues 292–305) that contains 12–40 glutamines in normal individuals and is expanded to 55–84 glutamines in the pathogenic form associated with spinocerebellar ataxia 3 (SCA3) [136,137]. Importantly, besides its involvement in SCA3, ataxin-3 has some other pathological functions, with decreased expression being correlated with the clinicopathologic features of gastric cancer [314].

Voltage-dependent P/Q-type calcium channel subunit  $\alpha$ 1A (CACNA1A, UniProt ID: O00555) is a 2505 residue-long polypeptide with a multitude of transmembrane regions and two long cytoplasmic domains (residues 715–1242 and 1808–2505), both predicted to be highly disordered (see Figure 2R). This observation is also in line with the high MobiDB consensus disorder score of this protein (42.08%). CACNA1A is a second protein for which no disorder-related information is provided by D<sup>2</sup>P<sup>2</sup>. However, similar to many other proteins considered in this article CACNA1A is characterized by the presence of several regions with compositional biases, such as poly-Gly (residues 13–18), two poly-Glu stretches (residues 727–732 and 1204–1207), poly-Arg, poly-His, and poly-Pro regions (residues 1002–1007, 2211–2220, and 2221–2224, respectively) in addition to the polyQ tract (residues 2314–2324). Involvement of CACNA1A in spinocerebellar ataxia type-6 (SCA6) is related to the trinucleotide CAG repeat expansion within its exon 47 [315] from the normal 4–16 to the 21–28 pathological SCA6-related repeats [316]. Pathological CACNA1A mutations were found to be associated with familial hemiplegic migraine type-1 [317–320], episodic ataxia type-2 [317], and early infantile epileptic encephalopathy type-42 [321,322]. CACNA1A is also associated with the exfoliation syndrome, which is the most common cause of open-angle glaucoma worldwide [323]. Furthermore, bioinformatics meta-analysis of public microarray datasets revealed that together with other members of the voltage-gated calcium channel family, CACNA1A can be implicated in the development and progression of diverse types of cancer and might undergo dramatic up-regulation in breast cancer [324].



Ataxin-7 is a 892 residue-long protein (UniProt ID: O15265) serving as a component of the STAGA transcription coactivator-HAT complex [167] that includes SPT3, TAF9, ADA, and GCN5 acetyltransferase [325]. Although human protein has a molecular mass of 95.4 kDa, at the SDS-PAGE it migrates at about 110 kDa [325], suggesting that ataxin-7 possesses significant amount of intrinsic disorder. In agreement with these observations and similar to other ataxins, this protein is predicted to be highly disordered, being characterized by the MobiDB consensus disorder score of 71.30% and being mostly disordered in the PONDR<sup>®</sup> VSL2 and D<sup>2</sup>P<sup>2</sup> plots (see Figure 2S and Figure S1S). In agreement with these high levels of predicted intrinsic disorder, the amino acid sequence of human ataxin-7 is heavily enriched in regions with compositional biases, such as two poly-Ala (residues 16–20 and 23–28), Gln-rich (residues 30–49), polyQ (residues 30–49), two Pro-rich (residues 40–65 and 402–486), two poly-Pro (residues 40–65 and 51–55), and two Ser-rich regions (residues 171–219 and 640–851) containing five poly-Ser tracts (residues 171–174, 213–219, 647–654, 717–730, and 840–845). Expansions of the polyQ tract from 4–35 to 36–306 repeats are associated with the spinocerebellar ataxia type-7 (SCA7) [326]. Furthermore, Lys264Arg mutation in ataxin-7 is among several common non-synonymous SNPs associated with breast cancer susceptibility [327], and a fusion between ataxin-7 and DNA repair protein Rad51C is expressed in colorectal tumors [328], whereas spleen-specific isoform of ataxin-7 was suggested to serve as a potential marker of the lymphoma-affected spleen [329].

The TATA-box-binding protein (TBP, UniProt ID: P20226) is a 339 amino acid-long general transcription factor engaged in the formation of the DNA-binding multiprotein factor TFIID related to the activation of eukaryotic genes transcribed by RNA polymerase II [330–334], as well as several other transcription factor complexes, such as a BRF2-containing transcription factor complex regulating the RNA polymerase III-mediated transcription [335] and the SL1/TIF-IB complex engaged in the assembly of the pre-initiation complex (PIC) during RNA polymerase I-dependent transcription [336]. Therefore, being required for transcriptional initiation by the three major RNA polymerases (RNAP I, II, and III) in eukaryotic nuclei, TBP is involved in the expression of most eukaryotic genes [337]. A part of the polyQ expansion-related pathology is abnormal interaction of TBP with the general transcription factor TFIIB and reduced DNA binding [338]. As many other transcription factors, TBP is predicted to be highly disordered (see Figure 2T and Figure S1T), possessing a wholly disordered N-tail (residues 1–160) that contains a polyQ tract (residues 55–95), and being characterized by the overall MobiDB consensus disorder score of 46.31%. This disorder distribution within the TBP sequence follows its domain organization, with the C-terminal domain that mediates virtually all of the transcriptionally relevant interactions of TBP being highly conserved among eukaryotes and [339], and with the N-terminal domain being evolutionarily divergent and showing sequence conservation only in vertebrates. In agreement with high levels of intrinsic disorder, human TBP, a protein with the calculated molecular mass of 37.7 kDa, was shown to possess an apparent molecular mass of ~49 kDa [340].

### 5.3. Disorder in Proteins Associated with Non-Coding Region Repeat Expansions

Since these proteins do not have pathological expansions in their coding regions, their intrinsic disorder status will be considered very briefly, with the exception being made for C9orf72, because of the known fact that although the GGGGCC hexanucleotide repeat expansion is located within the non-coding region of the C9ORF72 gene, it can be bi-directionally transcribed, and both sense and antisense repeat RNAs can be translated into the dipeptide repeat proteins (DPRs or C9RANT proteins) via the repeat-associated non-ATG (RAN) translation.

Synaptic functional regulator FMR1 (UniProt ID: Q06787) is a 632 residue-long polyribosome-associated RNA-binding protein regulating alternative mRNA splicing, mRNA stability, mRNA dendritic transport and postsynaptic local protein synthesis of a subset of mRNAs [341–345] among a myriad of other functions. FMR1 is predicted to be rather highly disordered, possessing MobiDB score of 38.92% and highly disordered C-terminal region that has multiple MoRFs and is heavily decorated



with different PTMs (see Figure 2U and Figure S1U). However, it does not possess regions with compositional bias.

Disco-interacting protein 2 homolog B (UniProt ID: Q9P265) is a moderately disordered 1576 residue-long protein involved in DNA methylation [346], with the MobiDB score of 19.73% and highly disordered N-terminal tail with several disorder-based protein binding regions and multiple sites of different PTMs (see Figure 2V and Figure S1V). This protein has a Ser-rich (residues 145–234), a poly-Ala (residues 1118–1121), and two poly-Val regions (residues 1503–1506 and 1534–1540).

AF4/FMR2 family member 2 (UniProt ID: P51816) is a 1311 residue-long RNA binding protein involved in alternative splicing regulation [347]. FMR2 is predicted by MobiDB to have 58.12% disordered residues occupying the N-terminal and central parts of this protein that have multiple disorder-based binding regions and several PTM sites (see Figure 2W and Figure S1W). There is no compositional bias in this protein.

Although the 481 residue-long protein C9orf72 (UniProt ID: Q96LT7) is the most ordered protein analyzed in this study (it has a MobiDB score of 2.70% and is expected to have short disordered tails, three to four short disordered loops, and a long disordered/flexible region located between the residues 130 and 210 as shown in Figure 2X and Figure S1X), it clearly deserves more attention in relation to the topic of this article. Besides the fact that the GGGGCC ( $G_4C_2$ ) hexanucleotide repeat expansions in the non-coding intronic region between the non-coding exons 1a and 1b of the *C9ORF72* gene represent the major cause of ALS and FTD [348,349] and that these expansions can vary from 10 to thousands of repeats [108,109,181,350,351], the expanded GGGGCC hexanucleotide repeats can be bi-directionally transcribed, and both sense and antisense repeat RNAs are engaged in the formation of RNA foci [352–355]. Furthermore, resulting hexanucleotide repeat RNA can be translated in a series of the dipeptide repeat proteins (DPRs or C9RANT proteins) by the RAN translation and these DPRs are commonly found as major constituents of proteinaceous inclusions throughout the CNS of the ALS and FTD patients [356]. This RAN translation of sense GGGGCC repeat RNAs generates three different proteins, poly(GA), poly(GR), and poly(GP) [194,195], and poly(PA), poly(PR), and poly(GP) proteins are synthesized as a result of translation of the antisense RNAs [353,355,357]. It is known that repeat-containing proteins are often intrinsically disordered, with the more perfect repeats being more disordered [47]. In agreement with these earlier observations, all DPRs synthesized as a result of the RAN translation of the sense and antisense GGGGCC hexanucleotide repeat RNAs, poly(GA), poly(GR), poly(GP), poly(PA), and poly(PR), were predicted to be highly disordered [358]. It was also emphasized that based on their positions within the CH-CDF phase space, DPRs can be classified either as native molten globules (Poly(GA) and Poly(PA)) or native coils or pre-molten globules (Poly(GP), Poly(GR), and Poly(PR)) [358].

Frataxin is a 210 residue-long iron-binding protein (UniProt ID: Q16595) that takes part in the heme biosynthesis [359] and biosynthesis and repair of iron-sulfur clusters [360] and acts as an iron chaperone [361]. Despite being involved in catalytic detoxification of redox-active iron [362], frataxin is predicted to have a MobiDB score of 40.00%, with the majority of disordered residues being concentrated within the N- and C-tails containing several MoRFs and PTMs (see Figure 2Y and Figure S1Y). There are no regions with compositional bias in this protein.

The cellular nucleic acid-binding protein is a 177 residue-long single-stranded DNA-binding protein (CNBP, UniProt ID: P62633) with moderate level of intrinsic disorder as evidenced by the MobiDB score of 19.77% and presence of several IDPRs enriched in MoRFs and PTM sites (see Figure 2Z and Figure S1Z). CNBP has an Arg/Gly-rich region (residues 22–42) seven zinc finger motifs of the CCHC type.

Ataxin-10 is a 475 residue-long protein (UniProt ID: Q9UBB4) needed for the survival of cerebellar neurons. With the MobiDB score of 5.26%, mostly short IDPRs (see Figure 2a and Figure S1a), and lack of compositional biases, this protein is the most ordered ataxin considered in this article.

Nucleolar protein 56 is a 594 residue-long protein (UniProt ID: O00567) serving as a core component of the box C/D small nucleolar ribonucleoprotein (snoRNP) particles [363] and related

to the early to middle stages of the biogenesis of 60S ribosomal subunit [364]. It is characterized by the moderate-to-high overall levels of intrinsic disorder, has a MobiDB score of 26.26%, and possesses a highly disordered C-terminal domain that contains multiple MoRFs, a multitude of diverse PTMs (see Figure 2b and Figure S1b), and a long Lys-rich stretch (residues 438–589).

Transcription factor 4 is a 667 residue-long transcription factor (UniProt ID: P15884) known for its binding to the immunoglobulin enhancer Mu-E5/KE5-motif and involvement in the initiation of neuronal differentiation. This is one of the most disordered proteins analyzed in this study, being characterized by a MobiDB score of 88.01%, containing long IDPRs and a whole host of different PTMs, using almost an entire sequence for the disorder-based protein-protein and protein-DNA interactions (see Figure 2c and Figure S1c), and containing a short poly-Ser region (residues 228–231).

Myotonin-protein kinase is a 629 residue-long non-receptor serine/threonine protein kinase (UniProt ID: Q09013) needed for the upkeep of skeletal muscle structure and function. It has a moderate intrinsic disorder content of 14.63%, with the majority of disordered residues, disorder-based binding regions, and PTM sites being present in the C-terminal part of this protein (see Figure 2d and Figure S1d). There are no compositional bias regions in this kinase.

Ataxin-8 and ATXN8OS are 80 and 200 residue-long proteins, respectively. Biological functions of the ataxin-8 (UniProt ID: Q156A1) and ATXN8OS proteins (UniProt ID: P0DMR3) are unknown. ATXN8OS is a putative protein with the disorder content of 68% (see Figure 2e). It is predicted to contain a couple of MoRFs located in the N-terminal region. Since ataxin-8 is simply a polyglutamine polypeptide (in fact, according to UniProt, it contains only 80 glutamine residues); it is not surprising that this protein is predicted to be completely disordered (see Figure 2f).

Cystatin-B is a short, 98 residue-long, protein (UniProt ID: P04080) that serves as an inhibitor of intracellular thiol proteinase. Cystatin-B is predicted by MobiDB to have 40.82% disordered residues and is expected to have flexible tails and a less flexible central region (see Figure 2g and Figure S1g).

Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B  $\beta$  isoform (PPP2R2B, UniProt ID: Q00005) is a 443 residue-long protein modulating substrate selectivity and catalytic activity of the PP2A phosphatase. It is characterized by a MobiDB score of 7.67% and has several short IDPRs, one short MoRF, and several PTM sites (see Figure 2h and Figure S1h). There are also seven WD repeats that are 40-residue-long conserved domains containing a centrally located Trp-Asp motif.

## 6. Summary

IDPs and IDPRs are characteristically comprised of low complexity domains, often containing repetitive amino acid sequences [47]. Expansion of these repetitive domains frequently results in pathology seen in several neurological and neuromuscular diseases. Many of the genes linked to repeat expansion diseases are translated into proteins predicted to contain a large percentage of disordered regions. Even expansions in regions that are not normally translated, such as those in C9orf72, result in aberrant dipeptide repeat products predicted to be disordered. The resultant disorder in these peptides often decreases in a length-dependent manner, causing many repetitive proteins to aggregate and sequester other proteins into the aggregated structures [365].

Trinucleotide repeats are the most common type seen in coding regions of genes, with polyQ being the most commonly occurring followed by polyAla [65]. Both pathological extensions result in diseases classified as protein misfolding disorders and share many pathological characteristics such as intracellular protein inclusions and the appearance of RNA foci. Since both Ala and Gln repeat expansions correlate with protein aggregation and formation of the  $\beta$ -structure-enriched amyloid-like fibrils [366,367], it is likely that the fibrillation process triggered by expanded repeats of these two completely different residues has some common molecular mechanisms. In agreement with this hypothesis, computational analysis of non-aggregated polyAla and polyQ peptides composed of 7, 10, 14 or 20 amino acids revealed that both types of these homo-oligopeptides are characterized by the presence of similar secondary structural elements (type I and type III  $\beta$ -turns, antiparallel  $\beta$ -strand,  $\alpha$ -helix, and  $3_{10}$ -helix) and that characteristic H-bonding patterns containing  $i-i+3$  and

$i-i + 4$  H-bonds are formed [368]. Furthermore, both polyAla and polyQ repeats were shown to form coiled-coil structures, the stability of which increased with the length of the expansion, and which were also able to form higher-order multimers and aggregates in vitro [369]. Also, it was pointed out that in the majority of expanded CAG and GCG repeat proteins, the polyQ or polyAla sequence is typically located within the protein, thereby possessing specific N- and C-terminal flanking sequences that may play a crucial role in regulation of the polyQ and polyAla aggregation propensity [367]. Finally, many of the proteins involved in both types of trinucleotide repeat expansions are associated with signaling, regulation, and RNA metabolism. Many of the diseases associated with both expansion types are developmental, neuromuscular, and neurodegenerative in nature [365].

The similarities shared by the two common trinucleotide repeats involved in disease are more abundant than the characteristics that set them apart. For example, trinucleotide repeats exceeding specific thresholds show replication-related instability that increases in a repeat length-dependent manner, with most instabilities causing repeat expansion [370]. Due to the effect of increased probability of repeat expansion on replication, and due to the fact that the increase in the rate of mutations that add additional codons and thereby increase the expansion length become more likely with each new generation, in successive generations, the age of the onset of trinucleotide expansion diseases typically becomes younger, whereas the severity of these maladies increases [370]. Importantly, one should keep in mind that expanded repeats can undergo further expansion-biased somatic instability, leading to further increases in the expansion length [371–375]. Furthermore, pathogenic processes associated with the expansions of both the polyQ and polyAla repeats are usually age-dependent (i.e., they do not happen before a particular age). This is likely due to the presence of the aforementioned somatic instability of the trinucleotide expansions [372] and because of other age-related processes, such as impairment of proteostasis [376–381].

However, not everything is similar for the repeat-containing proteins. In fact, one stark difference between polyQ and polyAla repeats is the threshold size of the repeat needed to cause disease. PolyQ repeats require much longer extensions for disease presentation, while polyAla repeats often can cause disease with negligible extensions. This size difference may have to do with the fact that extensions in polyAla repeats induce structural changes more readily than those in polyQ repeats because of the biophysical properties of the repeated amino acids in the peptide sequence [65].

Repeats in non-coding regions are more diverse than those seen in coding regions. The functions of the genes that are involved in disease vary widely as compared to those found in coding regions that are involved in disease. Often times in non-coding expansion diseases, there is a knock down in expression of the gene involved. However, in most cases this is not sufficient to cause disease and therefore there must be other mechanisms at play. For example, we now know that in the case of C9orf72, not only is the expression of the protein reduced, but aberrant mRNA and repeat peptide species are also produced [108,109,181].

The peptide species produced through non-canonical translation of the repeat region in C9orf72 are predicted to be disordered at short repeat lengths, but upon extension have been shown to form toxic stable aggregates in cells that cause the sequestration of proteins with low complexity domains [196]. Therefore, it is important to consider that these species may account for, at least in part, the pathology that is associated with the expansion of repeats in non-coding regions. In addition, it is important to note that the expansion in the 5' UTR of C9orf72 results in five different dipeptide repeat species. Three of the peptides are produced from the three available reading frames of the repeat region and their antisense partners make up the remaining three species present in disease. This could theoretically be the case for all expansions that occur in non-coding regions and should be considered for other similar cases. All of the repeat regions may also have the propensity to undergo translation into toxic repeat peptide species with unique biochemical properties. The predicted peptide species for each non-coding expansion linked to disease can be found in Table 2. Although, except for the DPRs generated as a result of RAN translation of the mRNA produced as a result of hexanucleotide expansion in the 5' UTR of the C9orf72 gene, the presence of such polypeptides was not demonstrated

as of yet, so one cannot exclude that at least some of these species are present in cells affected by the non-coding expansion mutations and can therefore contribute to the pathology of related diseases.

**Table 2.** Potential translation products of non-coding repeat expansions.

Gene	Repeat Sequence	Sense Translation	Antisense Translation
<i>FXS</i>	CGG-CGG-CGG-CGG	R-R-R-R	A-A-A-A
	GGC-GGC-GGC-GGC	G-G-G-G	P-P-P-P
	GCG-GCG-GCG-GCG	A-A-A-A	R-R-R-R
<i>DIP2B</i>	CGG-CGG-CGG-CGG	R-R-R-R	A-A-A-A
	GGC-GGC-GGC-GGC	G-G-G-G	P-P-P-P
	GCG-GCG-GCG-GCG	A-A-A-A	R-R-R-R
<i>FMR2</i>	GCC-GCC-GCC-GCC	A-A-A-A	R-R-R-R
	CCG-CCG-CCG-CCG	P-P-P-P	G-G-G-G
	CGC-CGC-CGC-CGC	R-R-R-R	A-A-A-A
<i>C9orf72</i>	GGG-GCC-GGG-GCC	G-A-G-A	P-R-P-R
	GGG-CCG-GGG-CCG	G-P-G-P	P-G-P-G
	GGC-CGG-GGC-CGG	G-R-G-R	P-A-P-A
<i>FXN</i>	GAA-GAA-GAA-GAA	E-E-E-E	L-L-L-L
	AAG-AAG-AAG-AAG	K-K-K-K	F-F-F-F
	AGA-AGA-AGA-AGA	R-R-R-R	S-S-S-S
<i>CNBP</i>	CCT-GCC-TGC-CTG	P-A-C-L	G-R-T-G-G
	CTG-CCT-GCC-TGC	L-P-A-C	G-G-R-T-G
	TGC-CTG-CCT-GCC	C-L-P-A	T-G-G-G-R
<i>ATXN10</i>	ATT-CTA-TTC-TAT-TCT	I-L-F-F-T	STOP-D-K-I-R
	TTC-TAT-TCT-ATT-CTA	F-F-S-I-L	K-I-R-STOP-D
	TCT-ATT-CTA-TTC-TAT	S-I-L-F-F	R-STOP-D-K-I
<i>TCF4</i>	CTG-CTG-CTG-CTG	L-L-L-L	D-D-D-D
	TGC-TGC-TGC-TGC	C-C-C-C	P-P-P-P
	GCT-GCT-GCT-GCT	A-A-A-A	R-R-R-R
<i>DMPK</i>	CTG-CTG-CTG-CTG	L-L-L-L	D-D-D-D
	TGC-TGC-TGC-TGC	C-C-C-C	P-P-P-P
	GCT-GCT-GCT-GCT	A-A-A-A	R-R-R-R
<i>JPH3</i>	CTG-CTG-CTG-CTG	L-L-L-L	D-D-D-D
	TGC-TGC-TGC-TGC	C-C-C-C	P-P-P-P
	GCT-GCT-GCT-GCT	A-A-A-A	R-R-R-R
<i>ATXN8</i>	CTG-CTG-CTG-CTG	L-L-L-L	D-D-D-D
	TGC-TGC-TGC-TGC	C-C-C-C	P-P-P-P
	GCT-GCT-GCT-GCT	A-A-A-A	R-R-R-R
<i>CSTB</i>	CCC-CGC-CCC-GCG	P-R-P-A	G-A-G-R
	CCC-GCC-CCG-CGC	P-A-P-R	G-R-G-A
	CCG-CCC-CGC-GCC	P-P-R-A	G-G-A-R

Curiously, it has been demonstrated that the repeat-associated non-ATG translation can occur not only for the mRNAs produced as a result of the non-coding region repeat expansions, but also for other RNAs containing expanded CAG and CTG trinucleotide repeats, resulting in expression of homopolymeric expansion proteins in all three reading frames [382]. Among characteristic examples of these RAN translation events are biosynthesis of homopolymeric polyglutamine, polyalanine, and polyserine proteins in the absence of an ATG codon. It was also pointed out that polyAla and polyserine proteins can contribute to the pathogenesis of some of the CAG expansion-associated polyQ diseases [382]. For example, in SCA8, SCA8<sub>GCA-Ala</sub> expansion protein was found in cerebellar Purkinje cells, whereas in DM1, the DM1<sub>CAG-Gln</sub> expansion protein was found in heart, myoblasts, and skeletal muscles [382].

Diseases found in microsatellite expansion regions are numerous and, in most cases, devastating. The broad commonalities between the functions of the genes involved, the structural change of the translated products as a result of the expansion, and the resulting disease phenotypes are overwhelmingly evident. This points to the question of whether there is a shared mechanism that occurs in most disease as a result of the expansions. Elucidation of the common mechanisms involved in pathology of each expansion disorder may shed light on this question and open new avenues for a broader therapeutic approach to treating repeat expansion disorders.

**Supplementary Materials:** Supplementary materials are available online. Figure S1: Intrinsic disorder propensity and some important disorder-related functional information generated for human proteins encoded by genes with nucleotide expansions by the D2P2 database (<http://d2p2.pro/>). Here, the outputs of several disorder predictors are shown by differently colored bars, whereas the blue-green-and-white bar in the middle of the plot shows the predicted disorder agreement between nine predictors, with blue and green parts corresponding to disordered regions by consensus. Yellow bar shows the location of the predicted disorder-based binding sites (molecular recognition features, MoRFs), whereas colored circles at the bottom of the plot show location of various PTMs.

**Acknowledgments:** This work was supported in part by a grant “Studies of structural and functional properties of metal binding and intrinsically disordered proteins” from the Program of the Russian Academy of Sciences “Molecular and Cellular Biology” (registration number 11411870001).

**Author Contributions:** V.N.U. conceived the idea; A.L.D. and V.N.U. collected and analyzed literature data; V.N.U. conducted bioinformatics analysis; A.L.D. and V.N.U. wrote the paper.

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

1. Turoverov, K.K.; Kuznetsova, I.M.; Uversky, V.N. The protein kingdom extended: Ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation. *Prog. Biophys. Mol. Biol.* **2010**, *102*, 73–84. [[CrossRef](#)] [[PubMed](#)]
2. Wright, P.E.; Dyson, H.J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999**, *293*, 321–331. [[CrossRef](#)] [[PubMed](#)]
3. Romero, P.; Obradovic, Z.; Kissinger, C.R.; Villafranca, J.E.; Garner, E.; Guillot, S.; Dunker, A.K. Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput.* **1998**, *3*, 437–448.
4. Dunker, A.K.; Obradovic, Z.; Romero, P.; Garner, E.C.; Brown, C.J. Intrinsic protein disorder in complete genomes. *Genome Inform. Ser. Workshop Genome Inform.* **2000**, *11*, 161–171. [[PubMed](#)]
5. Ward, J.J.; Sodhi, J.S.; McGuffin, L.J.; Buxton, B.F.; Jones, D.T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **2004**, *337*, 635–645. [[CrossRef](#)] [[PubMed](#)]
6. Xue, B.; Dunker, A.K.; Uversky, V.N. Orderly order in protein intrinsic disorder distribution: Disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.* **2012**, *30*, 137–149. [[CrossRef](#)] [[PubMed](#)]
7. Peng, Z.; Yan, J.; Fan, X.; Mizianty, M.J.; Xue, B.; Wang, K.; Hu, G.; Uversky, V.N.; Kurgan, L. Exceptionally abundant exceptions: Comprehensive characterization of intrinsic disorder in all domains of life. *Cell. Mol. Life Sci.* **2015**, *72*, 137–151. [[CrossRef](#)] [[PubMed](#)]
8. Dunker, A.K.; Garner, E.; Guillot, S.; Romero, P.; Albrecht, K.; Hart, J.; Obradovic, Z.; Kissinger, C.; Villafranca, J.E. Protein disorder and the evolution of molecular recognition: Theory, predictions and observations. *Pac. Symp. Biocomput.* **1998**, *3*, 473–484.
9. Uversky, V.N.; Gillespie, J.R.; Fink, A.L. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* **2000**, *41*, 415–427. [[CrossRef](#)]
10. Dunker, A.K.; Lawson, J.D.; Brown, C.J.; Williams, R.M.; Romero, P.; Oh, J.S.; Oldfield, C.J.; Campen, A.M.; Ratliff, C.M.; Hipps, K.W.; et al. Intrinsically disordered protein. *J. Mol. Graph. Model.* **2001**, *19*, 26–59. [[CrossRef](#)]
11. Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **2002**, *27*, 527–533. [[CrossRef](#)]
12. Uversky, V.N.; Dunker, A.K. Understanding protein non-folding. *Biochim. Biophys. Acta* **2010**, *1804*, 1231–1264. [[CrossRef](#)] [[PubMed](#)]



13. Uversky, V.N. Unusual biophysics of intrinsically disordered proteins. *Biochim. Biophys. Acta* **2013**, *1834*, 932–951. [[CrossRef](#)] [[PubMed](#)]
14. Iakoucheva, L.M.; Brown, C.J.; Lawson, J.D.; Obradovic, Z.; Dunker, A.K. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* **2002**, *323*, 573–584. [[CrossRef](#)]
15. Dunker, A.K.; Cortese, M.S.; Romero, P.; Iakoucheva, L.M.; Uversky, V.N. Flexible nets: The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **2005**, *272*, 5129–5148. [[CrossRef](#)] [[PubMed](#)]
16. Dunker, A.K.; Obradovic, Z. The protein trinity—Linking function and disorder. *Nat. Biotechnol.* **2001**, *19*, 805–806. [[CrossRef](#)] [[PubMed](#)]
17. Dunker, A.K.; Brown, C.J.; Obradovic, Z. Identification and functions of usefully disordered proteins. *Adv. Protein Chem.* **2002**, *62*, 25–49. [[PubMed](#)]
18. Dunker, A.K.; Brown, C.J.; Lawson, J.D.; Iakoucheva, L.M.; Obradovic, Z. Intrinsic disorder and protein function. *Biochemistry* **2002**, *41*, 6573–6582. [[CrossRef](#)] [[PubMed](#)]
19. Uversky, V.N. Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.* **2002**, *11*, 739–756. [[CrossRef](#)] [[PubMed](#)]
20. Uversky, V.N. What does it mean to be natively unfolded? *Eur. J. Biochem.* **2002**, *269*, 2–12. [[CrossRef](#)] [[PubMed](#)]
21. Uversky, V.N.; Oldfield, C.J.; Dunker, A.K. Showing your id: Intrinsic disorder as an id for recognition, regulation and cell signaling. *J. Mol. Recognit.* **2005**, *18*, 343–384. [[CrossRef](#)] [[PubMed](#)]
22. Dunker, A.K.; Silman, I.; Uversky, V.N.; Sussman, J.L. Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* **2008**, *18*, 756–764. [[CrossRef](#)] [[PubMed](#)]
23. Uversky, V.N. The mysterious unfoldome: Structureless, underappreciated, yet vital part of any given proteome. *J. Biomed. Biotechnol.* **2010**, *2010*, 568068. [[CrossRef](#)] [[PubMed](#)]
24. Dyson, H.J.; Wright, P.E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208. [[CrossRef](#)] [[PubMed](#)]
25. Tompa, P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* **2005**, *579*, 3346–3354. [[CrossRef](#)] [[PubMed](#)]
26. Dunker, A.K.; Bondos, S.E.; Huang, F.; Oldfield, C.J. Intrinsically disordered proteins and multicellular organisms. *Semin. Cell Dev. Biol.* **2015**, *37*, 44–55. [[CrossRef](#)] [[PubMed](#)]
27. Van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R.J.; Daughdrill, G.W.; Dunker, A.K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D.T.; et al. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **2014**, *114*, 6589–6631. [[CrossRef](#)] [[PubMed](#)]
28. Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V.N. Introducing protein intrinsic disorder. *Chem. Rev.* **2014**, *114*, 6561–6588. [[CrossRef](#)] [[PubMed](#)]
29. Fuxreiter, M.; Toth-Petroczy, A.; Kraut, D.A.; Matouschek, A.; Lim, R.Y.; Xue, B.; Kurgan, L.; Uversky, V.N. Disordered proteinaceous machines. *Chem. Rev.* **2014**, *114*, 6806–6843. [[CrossRef](#)] [[PubMed](#)]
30. Oldfield, C.J.; Dunker, A.K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **2014**, *83*, 553–584. [[CrossRef](#)] [[PubMed](#)]
31. Jakob, U.; Kriwacki, R.; Uversky, V.N. Conditionally and transiently disordered proteins: Awakening cryptic disorder to regulate protein function. *Chem. Rev.* **2014**, *114*, 6779–6805. [[CrossRef](#)] [[PubMed](#)]
32. Hsu, W.L.; Oldfield, C.; Meng, J.; Huang, F.; Xue, B.; Uversky, V.N.; Romero, P.; Dunker, A.K. Intrinsic protein disorder and protein-protein interactions. *Pac. Symp. Biocomput.* **2012**, 116–127. [[CrossRef](#)]
33. Oldfield, C.J.; Meng, J.; Yang, J.Y.; Yang, M.Q.; Uversky, V.N.; Dunker, A.K. Flexible nets: Disorder and induced fit in the associations of p53 and 14–3–3 with their partners. *BMC Genom.* **2008**, *9* (Suppl. 1), S1. [[CrossRef](#)] [[PubMed](#)]
34. Dyson, H.J.; Wright, P.E. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54–60. [[CrossRef](#)]
35. Oldfield, C.J.; Cheng, Y.; Cortese, M.S.; Romero, P.; Uversky, V.N.; Dunker, A.K. Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* **2005**, *44*, 12454–12470. [[CrossRef](#)] [[PubMed](#)]
36. Schulz, G.E. Nucleotide binding proteins. In *Molecular Mechanism of Biological Recognition*; Balaban, M., Ed.; Elsevier/North-Holland Biomedical Press: New York, NY, USA, 1979; pp. 79–94.

37. Ng, K.P.; Potikyan, G.; Savene, R.O.; Denny, C.T.; Uversky, V.N.; Lee, K.A. Multiple aromatic side chains within a disordered structure are critical for transcription and transforming activity of EWS family oncoproteins. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 479–484. [[CrossRef](#)] [[PubMed](#)]
38. Cortese, M.S.; Uversky, V.N.; Dunker, A.K. Intrinsic disorder in scaffold proteins: Getting more from less. *Prog. Biophys. Mol. Biol.* **2008**, *98*, 85–106. [[CrossRef](#)] [[PubMed](#)]
39. Uversky, V.N. Intrinsic disorder-based protein interactions and their modulators. *Curr. Pharm. Des.* **2013**, *19*, 4191–4213. [[CrossRef](#)] [[PubMed](#)]
40. Uversky, V.N.; Oldfield, C.J.; Dunker, A.K. Intrinsically disordered proteins in human diseases: Introducing the D2 concept. *Annu. Rev. Biophys.* **2008**, *37*, 215–246. [[CrossRef](#)] [[PubMed](#)]
41. Uversky, V.N. Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. *Front. Aging Neurosci.* **2015**, *7*, 18. [[CrossRef](#)] [[PubMed](#)]
42. Uversky, V.N. The triple power of D(3): Protein intrinsic disorder in degenerative diseases. *Front. Biosci. (Landmark Ed.)* **2014**, *19*, 181–258. [[CrossRef](#)] [[PubMed](#)]
43. Uversky, V.N. Intrinsic disorder in proteins associated with neurodegenerative diseases. *Front. Biosci. (Landmark Ed.)* **2009**, *14*, 5188–5238. [[CrossRef](#)] [[PubMed](#)]
44. Williams, R.M.; Obradovi, Z.; Mathura, V.; Braun, W.; Garner, E.C.; Young, J.; Takayama, S.; Brown, C.J.; Dunker, A.K. The protein non-folding problem: Amino acid determinants of intrinsic order and disorder. *Pac. Symp. Biocomput.* **2001**, 89–100. [[CrossRef](#)]
45. Radivojac, P.; Iakoucheva, L.M.; Oldfield, C.J.; Obradovic, Z.; Uversky, V.N.; Dunker, A.K. Intrinsic disorder and functional proteomics. *Biophys. J.* **2007**, *92*, 1439–1456. [[CrossRef](#)] [[PubMed](#)]
46. Vacic, V.; Uversky, V.N.; Dunker, A.K.; Lonardi, S. Composition profiler: A tool for discovery and visualization of amino acid composition differences. *BMC Bioinform.* **2007**, *8*, 211. [[CrossRef](#)] [[PubMed](#)]
47. Jorda, J.; Xue, B.; Uversky, V.N.; Kajava, A.V. Protein tandem repeats—The more perfect, the less structured. *FEBS J.* **2010**, *277*, 2673–2682. [[CrossRef](#)] [[PubMed](#)]
48. Simon, M.; Hancock, J.M. Tandem and cryptic amino acid repeats accumulate in disordered regions of proteins. *Genome Biol.* **2009**, *10*, R59. [[CrossRef](#)] [[PubMed](#)]
49. Tompa, P. Intrinsically unstructured proteins evolve by repeat expansion. *BioEssays* **2003**, *25*, 847–855. [[CrossRef](#)] [[PubMed](#)]
50. Alba, M.M.; Guigo, R. Comparative analysis of amino acid repeats in rodents and humans. *Genome Res.* **2004**, *14*, 549–554. [[CrossRef](#)] [[PubMed](#)]
51. Kalita, M.K.; Ramasamy, G.; Duraisamy, S.; Chauhan, V.S.; Gupta, D. Protorepeatsdb: A database of amino acid repeats in genomes. *BMC Bioinform.* **2006**, *7*, 336. [[CrossRef](#)] [[PubMed](#)]
52. Li, X.; Wang, W.; Wang, J.; Malovannaya, A.; Xi, Y.; Li, W.; Guerra, R.; Hawke, D.H.; Qin, J.; Chen, J. Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes. *Mol. Syst. Biol.* **2015**, *11*, 775. [[CrossRef](#)] [[PubMed](#)]
53. Mier, P.; Alanis-Lobato, G.; Andrade-Navarro, M.A. Context characterization of amino acid homorepeats using evolution, position, and order. *Proteins* **2017**, *85*, 709–719. [[CrossRef](#)] [[PubMed](#)]
54. Kajava, A.V. Tandem repeats in proteins: From sequence to structure. *J. Struct. Biol.* **2012**, *179*, 279–288. [[CrossRef](#)] [[PubMed](#)]
55. Gatchel, J.R.; Zoghbi, H.Y. Diseases of unstable repeat expansion: Mechanisms and common principles. *Nat. Rev. Genet.* **2005**, *6*, 743–755. [[CrossRef](#)] [[PubMed](#)]
56. La Spada, A.R.; Taylor, J.P. Repeat expansion disease: Progress and puzzles in disease pathogenesis. *Nat. Rev. Genet.* **2010**, *11*, 247–258. [[CrossRef](#)] [[PubMed](#)]
57. Usdin, K. The biological effects of simple tandem repeats: Lessons from the repeat expansion diseases. *Genome Res.* **2008**, *18*, 1011–1019. [[CrossRef](#)] [[PubMed](#)]
58. Koshy, B.T.; Zoghbi, H.Y. The CAG/polyglutamine tract diseases: Gene products and molecular pathogenesis. *Brain Pathol.* **1997**, *7*, 927–942. [[CrossRef](#)] [[PubMed](#)]
59. Ashley, C.T., Jr.; Warren, S.T. Trinucleotide repeat expansion and human disease. *Annu. Rev. Genet.* **1995**, *29*, 703–728. [[CrossRef](#)] [[PubMed](#)]
60. Carpenter, N.J. Genetic anticipation. Expanding tandem repeats. *Neurol. Clin.* **1994**, *12*, 683–697. [[PubMed](#)]

61. La Spada, A.R. Trinucleotide repeat instability: Genetic features and molecular mechanisms. *Brain Pathol.* **1997**, *7*, 943–963. [[CrossRef](#)] [[PubMed](#)]
62. Pearson, C.E. Repeat associated non-atg translation initiation: One DNA, two transcripts, seven reading frames, potentially nine toxic entities! *PLoS Genet.* **2011**, *7*, e1002018. [[CrossRef](#)] [[PubMed](#)]
63. Dosztanyi, Z.; Chen, J.; Dunker, A.K.; Simon, I.; Tompa, P. Disorder and sequence repeats in hub proteins and their implications for network evolution. *J. Proteome Res.* **2006**, *5*, 2985–2995. [[CrossRef](#)] [[PubMed](#)]
64. Xin, Q.; Li, L.; Li, J.; Qiu, R.; Guo, C.; Gong, Y.; Liu, Q. Eight-alanine duplication in homeobox d13 in a chinese family with synpolydactyly. *Gene* **2012**, *499*, 48–51. [[CrossRef](#)] [[PubMed](#)]
65. Albrecht, A.N.; Kornak, U.; Boddich, A.; Suring, K.; Robinson, P.N.; Stiege, A.C.; Lurz, R.; Stricker, S.; Wanker, E.E.; Mundlos, S. A molecular pathogenesis for transcription factor associated poly-alanine tract expansions. *Hum. Mol. Genet.* **2004**, *13*, 2351–2359. [[CrossRef](#)] [[PubMed](#)]
66. Muragaki, Y.; Mundlos, S.; Upton, J.; Olsen, B.R. Altered growth and branching patterns in synpolydactyly caused by mutations in hoxd13. *Science* **1996**, *272*, 548–551. [[CrossRef](#)] [[PubMed](#)]
67. Utsch, B.; Becker, K.; Brock, D.; Lentze, M.J.; Bidlingmaier, F.; Ludwig, M. A novel stable polyalanine [poly(A)] expansion in the HOXA13 gene associated with hand-foot-genital syndrome: Proper function of poly(A)-harbouring transcription factors depends on a critical repeat length? *Hum. Genet.* **2002**, *110*, 488–494. [[CrossRef](#)] [[PubMed](#)]
68. Shibata, A.; Machida, J.; Yamaguchi, S.; Kimura, M.; Tatematsu, T.; Miyachi, H.; Matsushita, M.; Kitoh, H.; Ishiguro, N.; Nakayama, A.; et al. Characterisation of novel runx2 mutation with alanine tract expansion from japanese cleidocranial dysplasia patient. *Mutagenesis* **2016**, *31*, 61–67. [[CrossRef](#)] [[PubMed](#)]
69. Paulussen, A.D.; Schrandt-Stumpel, C.T.; Tserpelis, D.C.; Spee, M.K.; Stegmann, A.P.; Mancini, G.M.; Brooks, A.S.; Collee, M.; Maat-Kievit, A.; Simon, M.E.; et al. The unfolding clinical spectrum of holoprosencephaly due to mutations in shh, zic2, six3 and tgif genes. *Eur. J. Hum. Genet.* **2010**, *18*, 999–1005. [[CrossRef](#)] [[PubMed](#)]
70. Cohen, M.M., Jr. Holoprosencephaly: Clinical, anatomic, and molecular dimensions. *Birth Defects Res. Part A Clin. Mol. Teratol.* **2006**, *76*, 658–673. [[CrossRef](#)] [[PubMed](#)]
71. Klaskova, E.; Drabek, J.; Hobzova, M.; Smolka, V.; Seda, M.; Hyjanek, J.; Slavkovsky, R.; Stranska, J.; Prochazka, M. Significant phenotype variability of congenital central hypoventilation syndrome in a family with polyalanine expansion mutation of the PHOX2B gene. *Biomed. Pap. Med. Fac. Univ. Palacky Olomouc Czech. Repub.* **2016**, *160*, 495–498. [[CrossRef](#)] [[PubMed](#)]
72. Di Lascio, S.; Belperio, D.; Benfante, R.; Fornasari, D. Alanine expansions associated with congenital central hypoventilation syndrome impair PHOX2B homeodomain-mediated dimerization and nuclear import. *J. Biol. Chem.* **2016**, *291*, 13375–13393. [[CrossRef](#)] [[PubMed](#)]
73. Bachetti, T.; Di Duca, M.; Della Monica, M.; Grappone, L.; Scarano, G.; Ceccherini, I. Recurrence of CCHS associated PHOX2B poly-alanine expansion mutation due to maternal mosaicism. *Pediatr. Pulmonol.* **2014**, *49*, E45–E47. [[CrossRef](#)] [[PubMed](#)]
74. Wong, J.; Farlie, P.; Holbert, S.; Lockhart, P.; Thomas, P.Q. Polyalanine expansion mutations in the X-linked hypopituitarism gene SOX3 result in aggresome formation and impaired transactivation. *Front. Biosci.* **2007**, *12*, 2085–2095. [[CrossRef](#)] [[PubMed](#)]
75. Albrecht, A.; Mundlos, S. The other trinucleotide repeat: Polyalanine expansion disorders. *Curr. Opin. Genet. Dev.* **2005**, *15*, 285–293. [[CrossRef](#)] [[PubMed](#)]
76. Cossee, M.; Faivre, L.; Philippe, C.; Hichri, H.; de Saint-Martin, A.; Laugel, V.; Bahi-Buisson, N.; Lemaitre, J.F.; Leheup, B.; Delobel, B.; et al. Arx polyalanine expansions are highly implicated in familial cases of mental retardation with infantile epilepsy and/or hand dystonia. *Am. J. Med. Genet. Part A* **2011**, *155*, 98–105. [[CrossRef](#)] [[PubMed](#)]
77. Lisik, M.; Sieron, A.L. [Arx—One gene—Many phenotypes]. *Neurologia i Neurochirurgia Polska* **2008**, *42*, 338–344. [[PubMed](#)]
78. Fan, J.; Zhou, Y.; Huang, X.; Zhang, L.; Yao, Y.; Song, X.; Chen, J.; Hu, J.; Ge, S.; Song, H.; et al. The combination of polyalanine expansion mutation and a novel missense substitution in transcription factor FOXL2 leads to different ovarian phenotypes in blepharophimosis-ptosis-epicanthus inversus syndrome (BPES) patients. *Hum. Reprod. (Oxf. Engl.)* **2012**, *27*, 3347–3357. [[CrossRef](#)] [[PubMed](#)]
79. Beysen, D.; De Paepe, A.; De Baere, E. Foxl2 mutations and genomic rearrangements in BPES. *Hum. Mutat.* **2009**, *30*, 158–169. [[CrossRef](#)] [[PubMed](#)]

80. Muller, T.; Schroder, R.; Zierz, S. Gcg repeats and phenotype in oculopharyngeal muscular dystrophy. *Muscle Nerve* **2001**, *24*, 120–122. [[CrossRef](#)]
81. Brais, B. Oculopharyngeal muscular dystrophy: A late-onset polyalanine disease. *Cytogenet. Genome Res.* **2003**, *100*, 252–260. [[CrossRef](#)] [[PubMed](#)]
82. Ivkovic, M.; Rankovic, V.; Tarasjev, A.; Orolicki, S.; Damjanovic, A.; Paunovic, V.R.; Romac, S. Schizophrenia and polymorphic CAG repeats array of calcium-activated potassium channel (KCNN3) gene in serbian population. *Int. J. Neurosci.* **2006**, *116*, 157–164. [[CrossRef](#)] [[PubMed](#)]
83. Holmes, S.E.; O'Hearn, E.; Rosenblatt, A.; Callahan, C.; Hwang, H.S.; Ingersoll-Ashworth, R.G.; Fleisher, A.; Stevanin, G.; Brice, A.; Potter, N.T.; et al. A repeat expansion in the gene encoding junctophilin-3 is associated with huntington disease-like 2. *Nat. Genet.* **2001**, *29*, 377–378. [[CrossRef](#)] [[PubMed](#)]
84. Margolis, R.L.; Holmes, S.E.; Rosenblatt, A.; Gourley, L.; O'Hearn, E.; Ross, C.A.; Seltzer, W.K.; Walker, R.H.; Ashizawa, T.; Rasmussen, A.; et al. Huntington's disease-like 2 (HDL2) in North America and Japan. *Ann. Neurol.* **2004**, *56*, 670–674. [[CrossRef](#)] [[PubMed](#)]
85. Todd, P.K.; Paulson, H.L. RNA-mediated neurodegeneration in repeat expansion disorders. *Ann. Neurol.* **2010**, *67*, 291–300. [[CrossRef](#)] [[PubMed](#)]
86. Andrade, M.A.; Bork, P. Heat repeats in the huntington's disease protein. *Nat. Genet.* **1995**, *11*, 115–116. [[CrossRef](#)] [[PubMed](#)]
87. Goehler, H.; Lalowski, M.; Stelzl, U.; Waelter, S.; Stroedicke, M.; Worm, U.; Droege, A.; Lindenberg, K.S.; Knoblich, M.; Haenig, C.; et al. A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to huntington's disease. *Mol. Cell* **2004**, *15*, 853–865. [[CrossRef](#)] [[PubMed](#)]
88. Nagafuchi, S.; Yanagisawa, H.; Ohsaki, E.; Shirayama, T.; Tadokoro, K.; Inoue, T.; Yamada, M. Structure and expression of the gene responsible for the triplet repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA). *Nat. Genet.* **1994**, *8*, 177–182. [[CrossRef](#)] [[PubMed](#)]
89. Koide, R.; Ikeuchi, T.; Onodera, O.; Tanaka, H.; Igarashi, S.; Endo, K.; Takahashi, H.; Kondo, R.; Ishikawa, A.; Hayashi, T.; et al. Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). *Nat. Genet.* **1994**, *6*, 9–13. [[CrossRef](#)] [[PubMed](#)]
90. La Spada, A.R.; Wilson, E.M.; Lubahn, D.B.; Harding, A.E.; Fischbeck, K.H. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **1991**, *352*, 77–79. [[CrossRef](#)] [[PubMed](#)]
91. Chong, S.S.; McCall, A.E.; Cota, J.; Subramony, S.H.; Orr, H.T.; Hughes, M.R.; Zoghbi, H.Y. Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nat. Genet.* **1995**, *10*, 344–350. [[CrossRef](#)] [[PubMed](#)]
92. Imbert, G.; Saudou, F.; Yvert, G.; Devys, D.; Trottier, Y.; Garnier, J.M.; Weber, C.; Mandel, J.L.; Cancel, G.; Abbas, N.; 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*, 285–291. [[CrossRef](#)] [[PubMed](#)]
93. Sanpei, K.; Takano, H.; Igarashi, S.; Sato, T.; Oyake, M.; Sasaki, H.; Wakisaka, A.; Tashiro, K.; Ishida, Y.; Ikeuchi, T.; 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*, 277–284. [[CrossRef](#)] [[PubMed](#)]
94. Durr, A.; Stevanin, G.; Cancel, G.; Duyckaerts, C.; Abbas, N.; Didierjean, O.; Chneiweiss, H.; Benomar, A.; Lyon-Caen, O.; Julien, J.; et al. Spinocerebellar ataxia 3 and machado-joseph disease: Clinical, molecular, and neuropathological features. *Ann. Neurol.* **1996**, *39*, 490–499. [[CrossRef](#)] [[PubMed](#)]
95. Riess, O.; Schols, L.; Bottger, H.; Nolte, D.; Vieira-Saecker, A.M.; Schimming, C.; Kreuz, F.; Macek, M., Jr.; Krebsova, A.; Macek, M.S.; et al. Sca6 is caused by moderate CAG expansion in the alpha1a-voltage-dependent calcium channel gene. *Hum. Mol. Genet.* **1997**, *6*, 1289–1293. [[CrossRef](#)] [[PubMed](#)]
96. Jodice, C.; Mantuano, E.; Veneziano, L.; Trettel, F.; Sabbadini, G.; Calandriello, L.; Francia, A.; Spadaro, M.; Pierelli, F.; Salvi, F.; et al. Episodic ataxia type 2 (EA2) and spinocerebellar ataxia type 6 (SCA6) due to CAG repeat expansion in the CACNA1A gene on chromosome 19p. *Hum. Mol. Genet.* **1997**, *6*, 1973–1978. [[CrossRef](#)] [[PubMed](#)]
97. David, G.; Abbas, N.; Stevanin, G.; Durr, A.; Yvert, G.; Cancel, G.; Weber, C.; Imbert, G.; Saudou, F.; Antoniou, E.; et al. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat. Genet.* **1997**, *17*, 65–70. [[CrossRef](#)] [[PubMed](#)]



98. David, G.; Durr, A.; Stevanin, G.; Cancel, G.; Abbas, N.; Benomar, A.; Belal, S.; Lebre, A.S.; Abada-Bendib, M.; Grid, D.; et al. Molecular and clinical correlations in autosomal dominant cerebellar ataxia with progressive macular dystrophy (SCA7). *Hum. Mol. Genet.* **1998**, *7*, 165–170. [[CrossRef](#)] [[PubMed](#)]
99. Nakamura, K.; Jeong, S.Y.; Uchiyama, T.; Anno, M.; Nagashima, K.; Nagashima, T.; Ikeda, S.; Tsuji, S.; Kanazawa, I. SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in tata-binding protein. *Hum. Mol. Genet.* **2001**, *10*, 1441–1448. [[CrossRef](#)] [[PubMed](#)]
100. Maltecca, F.; Filla, A.; Castaldo, I.; Coppola, G.; Fragassi, N.A.; Carella, M.; Bruni, A.; Cocozza, S.; Casari, G.; Servadio, A.; et al. Intergenerational instability and marked anticipation in SCA-17. *Neurology* **2003**, *61*, 1441–1443. [[CrossRef](#)] [[PubMed](#)]
101. Fujigasaki, H.; Verma, I.C.; Camuzat, A.; Margolis, R.L.; Zander, C.; Lebre, A.S.; Jamot, L.; Saxena, R.; Anand, I.; Holmes, S.E.; et al. SCA12 is a rare locus for autosomal dominant cerebellar ataxia: A study of an indian family. *Ann. Neurol.* **2001**, *49*, 117–121. [[CrossRef](#)]
102. Holmes, S.E.; O'Hearn, E.E.; McInnis, M.G.; Gorelick-Feldman, D.A.; Kleiderlein, J.J.; Callahan, C.; Kwak, N.G.; Ingersoll-Ashworth, R.G.; Sherr, M.; Sumner, A.J.; et al. Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. *Nat. Genet.* **1999**, *23*, 391–392. [[PubMed](#)]
103. Gray, S.J.; Gerhardt, J.; Doerfler, W.; Small, L.E.; Fanning, E. An origin of DNA replication in the promoter region of the human fragile X mental retardation (FMR1) gene. *Mol. Cell. Biol.* **2007**, *27*, 426–437. [[CrossRef](#)] [[PubMed](#)]
104. Winnepeninckx, B.; Debacker, K.; Ramsay, J.; Smeets, D.; Smits, A.; FitzPatrick, D.R.; Kooy, R.F. Cgg-repeat expansion in the DIP2B gene is associated with the fragile site FRA12A on chromosome 12Q13.1. *Am. J. Hum. Genet.* **2007**, *80*, 221–231. [[CrossRef](#)] [[PubMed](#)]
105. Gecz, J.; Bielby, S.; Sutherland, G.R.; Mulley, J.C. Gene structure and subcellular localization of FMR2, a member of a new family of putative transcription activators. *Genomics* **1997**, *44*, 201–213. [[CrossRef](#)] [[PubMed](#)]
106. Gecz, J.; Gedeon, A.K.; Sutherland, G.R.; Mulley, J.C. Identification of the gene FMR2, associated with fraxe mental retardation. *Nat. Genet.* **1996**, *13*, 105–108. [[CrossRef](#)] [[PubMed](#)]
107. Gu, Y.; Shen, Y.; Gibbs, R.A.; Nelson, D.L. Identification of FMR2, a novel gene associated with the fraxe CCG repeat and CPG island. *Nat. Genet.* **1996**, *13*, 109–113. [[CrossRef](#)] [[PubMed](#)]
108. DeJesus-Hernandez, M.; Mackenzie, I.R.; Boeve, B.F.; Boxer, A.L.; Baker, M.; Rutherford, N.J.; Nicholson, A.M.; Finch, N.A.; Flynn, H.; Adamson, J.; et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **2011**, *72*, 245–256. [[CrossRef](#)] [[PubMed](#)]
109. Renton, A.E.; Majounie, E.; Waite, A.; Simon-Sanchez, J.; Rollinson, S.; Gibbs, J.R.; Schymick, J.C.; Laaksovirta, H.; van Swieten, J.C.; Myllykangas, L.; et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **2011**, *72*, 257–268. [[CrossRef](#)] [[PubMed](#)]
110. Grabczyk, E.; Usdin, K. The GAA\*TTC triplet repeat expanded in friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. *Nucleic Acids Res.* **2000**, *28*, 2815–2822. [[CrossRef](#)] [[PubMed](#)]
111. Campuzano, V.; Montermini, L.; Molto, M.D.; Pianese, L.; Cossee, M.; Cavalcanti, F.; Monros, E.; Rodius, F.; Duclos, F.; Monticelli, A.; et al. Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **1996**, *271*, 1423–1427. [[CrossRef](#)] [[PubMed](#)]
112. Liquori, C.L.; Ricker, K.; Moseley, M.L.; Jacobsen, J.F.; Kress, W.; Naylor, S.L.; Day, J.W.; Ranum, L.P. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* **2001**, *293*, 864–867. [[CrossRef](#)] [[PubMed](#)]
113. Day, J.W.; Ricker, K.; Jacobsen, J.F.; Rasmussen, L.J.; Dick, K.A.; Kress, W.; Schneider, C.; Koch, M.C.; Beilman, G.J.; Harrison, A.R.; et al. Myotonic dystrophy type 2: Molecular, diagnostic and clinical spectrum. *Neurology* **2003**, *60*, 657–664. [[CrossRef](#)] [[PubMed](#)]
114. Matsuura, T.; Fang, P.; Pearson, C.E.; Jayakar, P.; Ashizawa, T.; Roa, B.B.; Nelson, D.L. Interruptions in the expanded ATTCT repeat of spinocerebellar ataxia type 10: Repeat purity as a disease modifier? *Am. J. Hum. Genet.* **2006**, *78*, 125–129. [[CrossRef](#)] [[PubMed](#)]



115. Matsuura, T.; Yamagata, T.; Burgess, D.L.; Rasmussen, A.; Grewal, R.P.; Watase, K.; Khajavi, M.; McCall, A.E.; Davis, C.F.; Zu, L.; et al. Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nat. Genet.* **2000**, *26*, 191–194. [[CrossRef](#)] [[PubMed](#)]
116. Obayashi, M.; Stevanin, G.; Synofzik, M.; Monin, M.L.; Duyckaerts, C.; Sato, N.; Streichenberger, N.; Vighetto, A.; Desestret, V.; Tesson, C.; et al. Spinocerebellar ataxia type 36 exists in diverse populations and can be caused by a short hexanucleotide GGCCTG repeat expansion. *J. Neurol. Neurosurg. Psychiatry* **2015**, *86*, 986–995. [[CrossRef](#)] [[PubMed](#)]
117. Mootha, V.V.; Hussain, I.; Cunnusamy, K.; Graham, E.; Gong, X.; Neelam, S.; Xing, C.; Kittler, R.; Petroll, W.M. TCF4 triplet repeat expansion and nuclear RNA foci in fuchs' endothelial corneal dystrophy. *Investig. Ophthalmol. Vis. Sci.* **2015**, *56*, 2003–2011. [[CrossRef](#)] [[PubMed](#)]
118. Nakano, M.; Okumura, N.; Nakagawa, H.; Koizumi, N.; Ikeda, Y.; Ueno, M.; Yoshii, K.; Adachi, H.; Aleff, R.A.; Butz, M.L.; et al. Trinucleotide repeat expansion in the tcf4 gene in fuchs' endothelial corneal dystrophy in japanese. *Investig. Ophthalmol. Vis. Sci.* **2015**, *56*, 4865–4869. [[CrossRef](#)] [[PubMed](#)]
119. Fu, Y.H.; Pizzuti, A.; Fenwick, R.G., Jr.; King, J.; Rajnarayan, S.; Dunne, P.W.; Dubel, J.; Nasser, G.A.; Ashizawa, T.; de Jong, P.; et al. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **1992**, *255*, 1256–1258. [[CrossRef](#)] [[PubMed](#)]
120. Mahadevan, M.; Tsilfidis, C.; Sabourin, L.; Shutler, G.; Amemiya, C.; Jansen, G.; Neville, C.; Narang, M.; Barcelo, J.; O'Hoy, K.; et al. Myotonic dystrophy mutation: An unstable CTG repeat in the 3' untranslated region of the gene. *Science* **1992**, *255*, 1253–1255. [[CrossRef](#)] [[PubMed](#)]
121. Tsilfidis, C.; MacKenzie, A.E.; Mettler, G.; Barcelo, J.; Korneluk, R.G. Correlation between CTG trinucleotide repeat length and frequency of severe congenital myotonic dystrophy. *Nat. Genet.* **1992**, *1*, 192–195. [[CrossRef](#)] [[PubMed](#)]
122. Worth, P.F.; Houlden, H.; Giunti, P.; Davis, M.B.; Wood, N.W. Large, expanded repeats in SCA8 are not confined to patients with cerebellar ataxia. *Nat. Genet.* **2000**, *24*, 214–215. [[CrossRef](#)] [[PubMed](#)]
123. Ikeda, Y.; Dalton, J.C.; Moseley, M.L.; Gardner, K.L.; Bird, T.D.; Ashizawa, T.; Seltzer, W.K.; Pandolfo, M.; Milunsky, A.; Potter, N.T.; et al. Spinocerebellar ataxia type 8: Molecular genetic comparisons and haplotype analysis of 37 families with ataxia. *Am. J. Hum. Genet.* **2004**, *75*, 3–16. [[CrossRef](#)] [[PubMed](#)]
124. Ikeda, Y.; Daughters, R.S.; Ranum, L.P. Bidirectional expression of the SCA8 expansion mutation: One mutation, two genes. *Cerebellum* **2008**, *7*, 150–158. [[CrossRef](#)] [[PubMed](#)]
125. Lalioti, M.D.; Mirotso, M.; Buresi, C.; Peitsch, M.C.; Rossier, C.; Ouazzani, R.; Baldy-Moulinier, M.; Bottani, A.; Malafosse, A.; Antonarakis, S.E. Identification of mutations in cystatin b, the gene responsible for the unverricht-lundborg type of progressive myoclonus epilepsy (epm1). *Am. J. Hum. Genet.* **1997**, *60*, 342–351. [[PubMed](#)]
126. Lalioti, M.D.; Scott, H.S.; Genton, P.; Grid, D.; Ouazzani, R.; M'Rabet, A.; Ibrahim, S.; Gouider, R.; Dravet, C.; Chkili, T.; et al. A pcr amplification method reveals instability of the dodecamer repeat in progressive myoclonus epilepsy (epm1) and no correlation between the size of the repeat and age at onset. *Am. J. Hum. Genet.* **1998**, *62*, 842–847. [[CrossRef](#)] [[PubMed](#)]
127. Di Domenico, T.; Walsh, I.; Martin, A.J.; Tosatto, S.C. Mobidb: A comprehensive database of intrinsic protein disorder annotations. *Bioinformatics* **2012**, *28*, 2080–2081. [[CrossRef](#)] [[PubMed](#)]
128. Potenza, E.; Domenico, T.D.; Walsh, I.; Tosatto, S.C. Mobidb 2.0: An improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Res.* **2014**, *43*, D315–D320. [[CrossRef](#)] [[PubMed](#)]
129. Amiel, J.; Trochet, D.; Clement-Ziza, M.; Munnich, A.; Lyonnet, S. Polyalanine expansions in human. *Hum. Mol. Genet.* **2004**, *13*, R235–R243. [[CrossRef](#)] [[PubMed](#)]
130. Lavoie, H.; Debeane, F.; Trinh, Q.D.; Turcotte, J.F.; Corbeil-Girard, L.P.; Dicaire, M.J.; Saint-Denis, A.; Page, M.; Rouleau, G.A.; Brais, B. Polymorphism, shared functions and convergent evolution of genes with sequences coding for polyalanine domains. *Hum. Mol. Genet.* **2003**, *12*, 2967–2979. [[CrossRef](#)] [[PubMed](#)]
131. Yamasaki, M.; Kanemura, Y. Molecular biology of pediatric hydrocephalus and hydrocephalus-related diseases. *Neurol. Med. Chir.* **2015**, *55*, 640–646. [[CrossRef](#)] [[PubMed](#)]
132. Bernacki, J.P.; Murphy, R.M. Length-dependent aggregation of uninterrupted polyalanine peptides. *Biochemistry* **2011**, *50*, 9200–9211. [[CrossRef](#)] [[PubMed](#)]
133. Ruggieri, M.; Pavone, P.; Scapagnini, G.; Romeo, L.; Lombardo, I.; Li Volti, G.; Corsello, G.; Pavone, L. The aristaless (arx) gene: One gene for many “interneuronopathies”. *Front. Biosci. (Elite Ed.)* **2010**, *2*, 701–710. [[CrossRef](#)] [[PubMed](#)]

134. Sherr, E.H. The arx story (epilepsy, mental retardation, autism, and cerebral malformations): One gene leads to many phenotypes. *Curr. Opin. Pediatr.* **2003**, *15*, 567–571. [[CrossRef](#)] [[PubMed](#)]
135. Beysen, D.; Moumne, L.; Veitia, R.; Peters, H.; Leroy, B.P.; De Paepe, A.; De Baere, E. Missense mutations in the forkhead domain of FOXL2 lead to subcellular mislocalization, protein aggregation and impaired transactivation. *Hum. Mol. Genet.* **2008**, *17*, 2030–2038. [[CrossRef](#)] [[PubMed](#)]
136. Cummings, C.J.; Zoghbi, H.Y. Trinucleotide repeats: Mechanisms and pathophysiology. *Annu. Rev. Genom. Hum. Genet.* **2000**, *1*, 281–328. [[CrossRef](#)] [[PubMed](#)]
137. Cummings, C.J.; Zoghbi, H.Y. Fourteen and counting: Unraveling trinucleotide repeat diseases. *Hum. Mol. Genet.* **2000**, *9*, 909–916. [[CrossRef](#)] [[PubMed](#)]
138. Fischbeck, K.H.; Souders, D.; La Spada, A. A candidate gene for X-linked spinal muscular atrophy. *Adv. Neurol.* **1991**, *56*, 209–213. [[PubMed](#)]
139. Ferrigno, P.; Silver, P.A. Polyglutamine expansions: Proteolysis, chaperones, and the dangers of promiscuity. *Neuron* **2000**, *26*, 9–12. [[CrossRef](#)]
140. La Spada, A.R.; Weydt, P.; Pineda, V.V. Frontiers in neuroscience huntington's disease pathogenesis: Mechanisms and pathways. In *Neurobiology of Huntington's Disease: Applications to Drug Discovery*; Lo, D.C., Hughes, R.E., Eds.; CRC Press/Taylor & Francis LLC.: Boca Raton, FL, USA, 2011.
141. Brouwer, J.R.; Willemsen, R.; Oostra, B.A. Microsatellite repeat instability and neurological disease. *BioEssays* **2009**, *31*, 71–83. [[CrossRef](#)] [[PubMed](#)]
142. Li, S.H.; Li, X.J. Aggregation of n-terminal huntingtin is dependent on the length of its glutamine repeats. *Hum. Mol. Genet.* **1998**, *7*, 777–782. [[CrossRef](#)] [[PubMed](#)]
143. Kazantsev, A.; Preisinger, E.; Dranovsky, A.; Goldgaber, D.; Housman, D. Insoluble detergent-resistant aggregates form between pathological and nonpathological lengths of polyglutamine in mammalian cells. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11404–11409. [[CrossRef](#)] [[PubMed](#)]
144. Perez, M.K.; Paulson, H.L.; Pendse, S.J.; Saionz, S.J.; Bonini, N.M.; Pittman, R.N. Recruitment and the role of nuclear localization in polyglutamine-mediated aggregation. *J. Cell Biol.* **1998**, *143*, 1457–1470. [[CrossRef](#)] [[PubMed](#)]
145. Yang, H.; Hu, H.Y. Sequestration of cellular interacting partners by protein aggregates: Implication in a loss-of-function pathology. *FEBS J.* **2016**, *283*, 3705–3717. [[CrossRef](#)] [[PubMed](#)]
146. Poirier, M.A.; Jiang, H.; Ross, C.A. A structure-based analysis of huntingtin mutant polyglutamine aggregation and toxicity: Evidence for a compact beta-sheet structure. *Hum. Mol. Genet.* **2005**, *14*, 765–774. [[CrossRef](#)] [[PubMed](#)]
147. Wolfe, K.J.; Cyr, D.M. Amyloid in neurodegenerative diseases: Friend or foe? *Semin. Cell Dev. Biol.* **2011**, *22*, 476–481. [[CrossRef](#)] [[PubMed](#)]
148. Zoghbi, H.Y.; Orr, H.T. Polyglutamine diseases: Protein cleavage and aggregation. *Curr. Opin. Neurobiol.* **1999**, *9*, 566–570. [[CrossRef](#)]
149. Ross, C.A.; Wood, J.D.; Schilling, G.; Peters, M.F.; Nucifora, F.C., Jr.; Cooper, J.K.; Sharp, A.H.; Margolis, R.L.; Borchelt, D.R. Polyglutamine pathogenesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **1999**, *354*, 1005–1011. [[CrossRef](#)] [[PubMed](#)]
150. Preisinger, E.; Jordan, B.M.; Kazantsev, A.; Housman, D. Evidence for a recruitment and sequestration mechanism in huntington's disease. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **1999**, *354*, 1029–1034. [[CrossRef](#)] [[PubMed](#)]
151. Wanker, E.E. Protein aggregation and pathogenesis of huntington's disease: Mechanisms and correlations. *Biol. Chem.* **2000**, *381*, 937–942. [[CrossRef](#)] [[PubMed](#)]
152. McCampbell, A.; Taylor, J.P.; Taye, A.A.; Robitschek, J.; Li, M.; Walcott, J.; Merry, D.; Chai, Y.; Paulson, H.; Sobue, G.; et al. CREB-binding protein sequestration by expanded polyglutamine. *Hum. Mol. Genet.* **2000**, *9*, 2197–2202. [[CrossRef](#)] [[PubMed](#)]
153. McCampbell, A.; Fischbeck, K.H. Polyglutamine and CBP: Fatal attraction? *Nat. Med.* **2001**, *7*, 528–530. [[CrossRef](#)] [[PubMed](#)]
154. Chen, S.; Berthelie, V.; Yang, W.; Wetzels, R. Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity. *J. Mol. Biol.* **2001**, *311*, 173–182. [[CrossRef](#)] [[PubMed](#)]

155. Chen, S.; Bertheliev, V.; Hamilton, J.B.; O’Nuallain, B.; Wetzel, R. Amyloid-like features of polyglutamine aggregates and their assembly kinetics. *Biochemistry* **2002**, *41*, 7391–7399. [[CrossRef](#)] [[PubMed](#)]
156. Perutz, M.F.; Pope, B.J.; Owen, D.; Wanker, E.E.; Scherzinger, E. Aggregation of proteins with expanded glutamine and alanine repeats of the glutamine-rich and asparagine-rich domains of sup35 and of the amyloid beta-peptide of amyloid plaques. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 5596–5600. [[CrossRef](#)] [[PubMed](#)]
157. Butland, S.L.; Devon, R.S.; Huang, Y.; Mead, C.L.; Meynert, A.M.; Neal, S.J.; Lee, S.S.; Wilkinson, A.; Yang, G.S.; Yuen, M.M.; et al. CAG-encoded polyglutamine length polymorphism in the human genome. *BMC Genom.* **2007**, *8*, 126. [[CrossRef](#)] [[PubMed](#)]
158. Krause, A.; Mitchell, C.; Essop, F.; Tager, S.; Temlett, J.; Stevanin, G.; Ross, C.; Rudnicki, D.; Margolis, R. Junctophilin 3 (JPH3) expansion mutations causing huntington disease like 2 (HDL2) are common in south african patients with african ancestry and a huntington disease phenotype. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **2015**, *168*, 573–585. [[CrossRef](#)] [[PubMed](#)]
159. Seixas, A.I.; Holmes, S.E.; Takeshima, H.; Pavlovich, A.; Sachs, N.; Pruitt, J.L.; Silveira, I.; Ross, C.A.; Margolis, R.L.; Rudnicki, D.D. Loss of junctophilin-3 contributes to huntington disease-like 2 pathogenesis. *Ann. Neurol.* **2012**, *71*, 245–257. [[CrossRef](#)] [[PubMed](#)]
160. Margolis, R.L.; Li, S.H.; Young, W.S.; Wagster, M.V.; Stine, O.C.; Kidwai, A.S.; Ashworth, R.G.; Ross, C.A. Drpla gene (atrophin-1) sequence and mRNA expression in human brain. *Brain Res. Mol. Brain Res.* **1996**, *36*, 219–226. [[CrossRef](#)]
161. Giorgetti, E.; Lieberman, A.P. Polyglutamine androgen receptor-mediated neuromuscular disease. *Cell. Mol. Life Sci.* **2016**, *73*, 3991–3999. [[CrossRef](#)] [[PubMed](#)]
162. La Spada, A. Spinal and bulbar muscular atrophy. In *Genereviews(r)*; Pagon, R.A., Adam, M.P., Ardinger, H.H., Wallace, S.E., Amemiya, A., Bean, L.J.H., Bird, T.D., Ledbetter, N., Mefford, H.C., Smith, R.J.H., et al. Eds.; Gene Reviews is a registered trademark of the University of Washington; University of Washington: Seattle, WA, USA, 1993.
163. Sun, Y.M.; Lu, C.; Wu, Z.Y. Spinocerebellar ataxia: Relationship between phenotype and genotype—A review. *Clin. Genet.* **2016**, *90*, 305–314. [[CrossRef](#)] [[PubMed](#)]
164. McEwan, I.J. Structural and functional alterations in the androgen receptor in spinal bulbar muscular atrophy. *Biochem. Soc. Trans.* **2001**, *29*, 222–227. [[CrossRef](#)] [[PubMed](#)]
165. Duenas, A.M.; Goold, R.; Giunti, P. Molecular pathogenesis of spinocerebellar ataxias. *Brain* **2006**, *129*, 1357–1370. [[CrossRef](#)] [[PubMed](#)]
166. Sobczak, K.; Krzyzosiak, W.J. Patterns of CAG repeat interruptions in SCA1 and SCA2 genes in relation to repeat instability. *Hum. Mutat.* **2004**, *24*, 236–247. [[CrossRef](#)] [[PubMed](#)]
167. Nakamura, Y.; Tagawa, K.; Oka, T.; Sasabe, T.; Ito, H.; Shiwaku, H.; La Spada, A.R.; Okazawa, H. Ataxin-7 associates with microtubules and stabilizes the cytoskeletal network. *Hum. Mol. Genet.* **2012**, *21*, 1099–1110. [[CrossRef](#)] [[PubMed](#)]
168. Seidel, K.; Brunt, E.R.; de Vos, R.A.; Dijk, F.; van der Want, H.J.; Rub, U.; den Dunnen, W.F. The p62 antibody reveals various cytoplasmic protein aggregates in spinocerebellar ataxia type 6. *Clin. Neuropathol.* **2009**, *28*, 344–349. [[CrossRef](#)] [[PubMed](#)]
169. Chandy, K.G.; Fantino, E.; Wittekindt, O.; Kalman, K.; Tong, L.L.; Ho, T.H.; Gutman, G.A.; Crocq, M.A.; Ganguli, R.; Nimgaonkar, V.; et al. Isolation of a novel potassium channel gene HSKCA3 containing a polymorphic CAG repeat: A candidate for schizophrenia and bipolar disorder? *Mol. Psychiatry* **1998**, *3*, 32–37. [[CrossRef](#)] [[PubMed](#)]
170. Ritsner, M.; Modai, I.; Ziv, H.; Amir, S.; Halperin, T.; Weizman, A.; Navon, R. An association of CAG repeats at the KCNN3 locus with symptom dimensions of schizophrenia. *Biol. Psychiatry* **2002**, *51*, 788–794. [[CrossRef](#)]
171. Dror, V.; Shamir, E.; Ghanshani, S.; Kimhi, R.; Swartz, M.; Barak, Y.; Weizman, R.; Avivi, L.; Litmanovitch, T.; Fantino, E.; et al. HKCA3/KCNN3 potassium channel gene: Association of longer CAG repeats with schizophrenia in israeli ashkenazi jews, expression in human tissues and localization to chromosome 1q21. *Mol. Psychiatry* **1999**, *4*, 254–260. [[CrossRef](#)] [[PubMed](#)]

172. Gargus, J.J.; Fantino, E.; Gutman, G.A. A piece in the puzzle: An ion channel candidate gene for schizophrenia. *Mol. Med. Today* **1998**, *4*, 518–524. [[CrossRef](#)]
173. Perutz, M.F. Glutamine repeats and inherited neurodegenerative diseases: Molecular aspects. *Curr. Opin. Struct. Biol.* **1996**, *6*, 848–858. [[CrossRef](#)]
174. Ross, C.A. Polyglutamine pathogenesis: Emergence of unifying mechanisms for huntington's disease and related disorders. *Neuron* **2002**, *35*, 819–822. [[CrossRef](#)]
175. Bates, G. Huntingtin aggregation and toxicity in huntington's disease. *Lancet* **2003**, *361*, 1642–1644. [[CrossRef](#)]
176. Soto, C. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat. Rev. Neurosci.* **2003**, *4*, 49–60. [[CrossRef](#)] [[PubMed](#)]
177. Yue, S.; Serra, H.G.; Zoghbi, H.Y.; Orr, H.T. The spinocerebellar ataxia type 1 protein, ataxin-1, has RNA-binding activity that is inversely affected by the length of its polyglutamine tract. *Hum. Mol. Genet.* **2001**, *10*, 25–30. [[CrossRef](#)] [[PubMed](#)]
178. Klement, I.A.; Skinner, P.J.; Kaytor, M.D.; Yi, H.; Hersch, S.M.; Clark, H.B.; Zoghbi, H.Y.; Orr, H.T. Ataxin-1 nuclear localization and aggregation: Role in polyglutamine-induced disease in SCA1 transgenic mice. *Cell* **1998**, *95*, 41–53. [[CrossRef](#)]
179. Gusella, J.; MacDonald, M. No post-genetics era in human disease research. *Nat. Rev. Genet.* **2002**, *3*, 72–79. [[CrossRef](#)] [[PubMed](#)]
180. Charlesworth, B.; Sniegowski, P.; Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **1994**, *371*, 215–220. [[CrossRef](#)] [[PubMed](#)]
181. Gitler, A.D.; Tsuiji, H. There has been an awakening: Emerging mechanisms of C9orf72 mutations in FTD/ALS. *Brain Res.* **2016**, *1647*, 19–29. [[CrossRef](#)] [[PubMed](#)]
182. Yu, S.; Pritchard, M.; Kremer, E.; Lynch, M.; Nancarrow, J.; Baker, E.; Holman, K.; Mulley, J.C.; Warren, S.T.; Schlessinger, D.; et al. Fragile x genotype characterized by an unstable region of DNA. *Science* **1991**, *252*, 1179–1181. [[CrossRef](#)] [[PubMed](#)]
183. Saul, R.A.; Tarleton, J.C. FMR1-related disorders. In *Genereviews(r)*; Pagon, R.A., Adam, M.P., Ardinger, H.H., Wallace, S.E., Amemiya, A., Bean, L.J.H., Bird, T.D., Ledbetter, N., Mefford, H.C., Smith, R.J.H., et al. Eds.; Gene Reviews is a registered trademark of the University of Washington; University of Washington: Seattle, WA, USA, 1993.
184. Eberhart, D.E.; Malter, H.E.; Feng, Y.; Warren, S.T. The fragile x mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear export signals. *Hum. Mol. Genet.* **1996**, *5*, 1083–1091. [[CrossRef](#)] [[PubMed](#)]
185. Verkerk, A.J.; Pieretti, M.; Sutcliffe, J.S.; Fu, Y.H.; Kuhl, D.P.; Pizzuti, A.; Reiner, O.; Richards, S.; Victoria, M.F.; Zhang, F.P.; et al. Identification of a gene (FMR-1) containing a cgg repeat coincident with a breakpoint cluster region exhibiting length variation in fragile x syndrome. *Cell* **1991**, *65*, 905–914. [[CrossRef](#)]
186. Knight, S.J.; Hirst, M.C.; Roche, A.; Christodoulou, Z.; Huson, S.M.; Winter, R.; Fitchett, M.; McKinley, M.J.; Lindenbaum, R.H.; Nakahori, Y.; et al. Molecular studies of the fragile x syndrome. *Am. J. Med. Genet.* **1992**, *43*, 217–223. [[CrossRef](#)] [[PubMed](#)]
187. Hagerman, P.J.; Hagerman, R.J. Fragile X-associated tremor/ataxia syndrome (FXTAS). *Ment. Retard. Dev. Disabil. Res. Rev.* **2004**, *10*, 25–30. [[CrossRef](#)] [[PubMed](#)]
188. Murray, A.; Webb, J.; Grimley, S.; Conway, G.; Jacobs, P. Studies of fraxa and fraxe in women with premature ovarian failure. *J. Med. Genet.* **1998**, *35*, 637–640. [[CrossRef](#)] [[PubMed](#)]
189. Sullivan, A.K.; Marcus, M.; Epstein, M.P.; Allen, E.G.; Anido, A.E.; Paquin, J.J.; Yadav-Shah, M.; Sherman, S.L. Association of FMR1 repeat size with ovarian dysfunction. *Hum. Reprod. (Oxf. Engl.)* **2005**, *20*, 402–412. [[CrossRef](#)] [[PubMed](#)]
190. Kumari, D.; Hayward, B.; Nakamura, A.J.; Bonner, W.M.; Usdin, K. Evidence for chromosome fragility at the frataxin locus in friedreich ataxia. *Mutat. Res.* **2015**, *781*, 14–21. [[CrossRef](#)] [[PubMed](#)]
191. Levine, T.P.; Daniels, R.D.; Gatta, A.T.; Wong, L.H.; Hayes, M.J. The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. *Bioinformatics* **2013**, *29*, 499–503. [[CrossRef](#)] [[PubMed](#)]
192. Farg, M.A.; Sundaramoorthy, V.; Sultana, J.M.; Yang, S.; Atkinson, R.A.; Levina, V.; Halloran, M.A.; Gleeson, P.A.; Blair, I.P.; Soo, K.Y.; et al. C9orf72, implicated in amyotrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. *Hum. Mol. Genet.* **2014**, *23*, 3579–3595. [[CrossRef](#)] [[PubMed](#)]



193. Ciura, S.; Lattante, S.; Le Ber, I.; Latouche, M.; Tostivint, H.; Brice, A.; Kabashi, E. Loss of function of C9orf72 causes motor deficits in a zebrafish model of amyotrophic lateral sclerosis. *Ann. Neurol.* **2013**, *74*, 180–187. [[CrossRef](#)] [[PubMed](#)]
194. Ash, P.E.; Bieniek, K.F.; Gendron, T.F.; Caulfield, T.; Lin, W.L.; DeJesus-Hernandez, M.; van Blitterswijk, M.M.; Jansen-West, K.; Paul, J.W., 3rd; Rademakers, R.; et al. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to C9FTD/ALS. *Neuron* **2013**, *77*, 639–646. [[CrossRef](#)] [[PubMed](#)]
195. Mori, K.; Weng, S.M.; Arzberger, T.; May, S.; Rentzsch, K.; Kremmer, E.; Schmid, B.; Kretzschmar, H.A.; Cruts, M.; Van Broeckhoven, C.; et al. The C9ORF72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTL/ALS. *Science* **2013**, *339*, 1335–1338. [[CrossRef](#)] [[PubMed](#)]
196. Lee, K.H.; Zhang, P.; Kim, H.J.; Mitrea, D.M.; Sarkar, M.; Freibaum, B.D.; Cika, J.; Coughlin, M.; Messing, J.; Molliex, A.; et al. C9ORF72 dipeptide repeats impair the assembly, dynamics, and function of membrane-less organelles. *Cell* **2016**, *167*, 774–788. [[CrossRef](#)] [[PubMed](#)]
197. Shi, K.Y.; Mori, E.; Nizami, Z.F.; Lin, Y.; Kato, M.; Xiang, S.; Wu, L.C.; Ding, M.; Yu, Y.; Gall, J.G.; et al. Toxic prn poly-dipeptides encoded by the C9orf72 repeat expansion block nuclear import and export. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E1111–e1117. [[CrossRef](#)] [[PubMed](#)]
198. Zhang, K.; Donnelly, C.J.; Haeusler, A.R.; Grima, J.C.; Machamer, J.B.; Steinwald, P.; Daley, E.L.; Miller, S.J.; Cunningham, K.M.; Vidensky, S.; et al. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **2015**, *525*, 56–61. [[CrossRef](#)] [[PubMed](#)]
199. Kwon, I.; Xiang, S.; Kato, M.; Wu, L.; Theodoropoulos, P.; Wang, T.; Kim, J.; Yun, J.; Xie, Y.; McKnight, S.L. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **2014**, *345*, 1139–1145. [[CrossRef](#)] [[PubMed](#)]
200. Lin, Y.; Mori, E.; Kato, M.; Xiang, S.; Wu, L.; Kwon, I.; McKnight, S.L. Toxic PR poly-dipeptides encoded by the C9orf72 repeat expansion target LC domain polymers. *Cell* **2016**, *167*, 789–802. [[CrossRef](#)] [[PubMed](#)]
201. Ranum, L.P.; Day, J.W. Myotonic dystrophy: Clinical and molecular parallels between myotonic dystrophy type 1 and type 2. *Curr. Neurol. Neurosci. Rep.* **2002**, *2*, 465–470. [[CrossRef](#)] [[PubMed](#)]
202. Kobayashi, H.; Abe, K.; Matsuura, T.; Ikeda, Y.; Hitomi, T.; Akechi, Y.; Habu, T.; Liu, W.; Okuda, H.; Koizumi, A. Expansion of intronic GGCTG hexanucleotide repeat in nop56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. *Am. J. Hum. Genet.* **2011**, *89*, 121–130. [[CrossRef](#)] [[PubMed](#)]
203. Lalioti, M.D.; Scott, H.S.; Antonarakis, S.E. Altered spacing of promoter elements due to the dodecamer repeat expansion contributes to reduced expression of the cystatin B gene in EPM1. *Hum. Mol. Genet.* **1999**, *8*, 1791–1798. [[CrossRef](#)] [[PubMed](#)]
204. Virtaneva, K.; D’Amato, E.; Miao, J.; Koskiniemi, M.; Norio, R.; Avanzini, G.; Franceschetti, S.; Michelucci, R.; Tassinari, C.A.; Omer, S.; et al. Unstable minisatellite expansion causing recessively inherited myoclonus epilepsy, EPM1. *Nat. Genet.* **1997**, *15*, 393–396. [[CrossRef](#)] [[PubMed](#)]
205. Williams, A.J.; Paulson, H.L. Polyglutamine neurodegeneration: Protein misfolding revisited. *Trends Neurosci.* **2008**, *31*, 521–528. [[CrossRef](#)] [[PubMed](#)]
206. O’Hearn, E.; Holmes, S.E.; Margolis, R.L. Spinocerebellar ataxia type 12. *Handb. Clin. Neurol.* **2012**, *103*, 535–547. [[PubMed](#)]
207. Dong, Y.; Wu, J.J.; Wu, Z.Y. Identification of 46 CAG repeats within PPP2R2B as probably the shortest pathogenic allele for SCA12. *Parkinsonism Relat. Disord.* **2015**, *21*, 398–401. [[CrossRef](#)] [[PubMed](#)]
208. Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Dunker, A.K. Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins* **2005**, *61* (Suppl. 7), 176–182. [[CrossRef](#)] [[PubMed](#)]
209. Peng, K.; Radivojac, P.; Vucetic, S.; Dunker, A.K.; Obradovic, Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinform.* **2006**, *7*, 208. [[CrossRef](#)] [[PubMed](#)]
210. Oates, M.E.; Romero, P.; Ishida, T.; Ghalwash, M.; Mizianty, M.J.; Xue, B.; Dosztanyi, Z.; Uversky, V.N.; Obradovic, Z.; Kurgan, L.; et al. D(2)p(2): Database of disordered protein predictions. *Nucleic Acids Res.* **2013**, *41*, D508–D516. [[CrossRef](#)] [[PubMed](#)]
211. Dosztanyi, Z.; Csizmok, V.; Tompa, P.; Simon, I. Iupred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **2005**, *21*, 3433–3434. [[CrossRef](#)] [[PubMed](#)]



212. Romero, P.; Obradovic, Z.; Li, X.; Garner, E.C.; Brown, C.J.; Dunker, A.K. Sequence complexity of disordered protein. *Proteins* **2001**, *42*, 38–48. [[CrossRef](#)]
213. Ishida, T.; Kinoshita, K. Prdos: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **2007**, *35*, W460–W464. [[CrossRef](#)] [[PubMed](#)]
214. Walsh, I.; Martin, A.J.; Di Domenico, T.; Tosatto, S.C. Espritz: Accurate and fast prediction of protein disorder. *Bioinformatics* **2012**, *28*, 503–509. [[CrossRef](#)] [[PubMed](#)]
215. Meszaros, B.; Simon, I.; Dosztanyi, Z. Prediction of protein binding regions in disordered proteins. *PLoS Comput. Biol.* **2009**, *5*, e1000376. [[CrossRef](#)] [[PubMed](#)]
216. Dosztanyi, Z.; Meszaros, B.; Simon, I. Anchor: Web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **2009**, *25*, 2745–2746. [[CrossRef](#)] [[PubMed](#)]
217. Mohan, A.; Oldfield, C.J.; Radivojac, P.; Vacic, V.; Cortese, M.S.; Dunker, A.K.; Uversky, V.N. Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.* **2006**, *362*, 1043–1059. [[CrossRef](#)] [[PubMed](#)]
218. Vacic, V.; Oldfield, C.J.; Mohan, A.; Radivojac, P.; Cortese, M.S.; Uversky, V.N.; Dunker, A.K. Characterization of molecular recognition features, MoRFs, and their binding partners. *J. Proteome Res.* **2007**, *6*, 2351–2366. [[CrossRef](#)] [[PubMed](#)]
219. Cheng, Y.; Oldfield, C.J.; Meng, J.; Romero, P.; Uversky, V.N.; Dunker, A.K. Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. *Biochemistry* **2007**, *46*, 13468–13477. [[CrossRef](#)] [[PubMed](#)]
220. Linding, R.; Jensen, L.J.; Diella, F.; Bork, P.; Gibson, T.J.; Russell, R.B. Protein disorder prediction: Implications for structural proteomics. *Structure* **2003**, *11*, 1453–1459. [[CrossRef](#)] [[PubMed](#)]
221. Yang, Z.R.; Thomson, R.; McNeil, P.; Esnouf, R.M. Ronn: The bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **2005**, *21*, 3369–3376. [[CrossRef](#)] [[PubMed](#)]
222. Peng, Z.L.; Kurgan, L. Comprehensive comparative assessment of in-silico predictors of disordered regions. *Curr. Protein Pept. Sci.* **2012**, *13*, 6–18. [[CrossRef](#)] [[PubMed](#)]
223. Linding, R.; Russell, R.B.; Neduva, V.; Gibson, T.J. Globplot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* **2003**, *31*, 3701–3708. [[CrossRef](#)] [[PubMed](#)]
224. Apweiler, R.; Bairoch, A.; Wu, C.H.; Barker, W.C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; et al. Uniprot: The universal protein knowledgebase. *Nucleic Acids Res.* **2004**, *32*, D115–D119. [[CrossRef](#)] [[PubMed](#)]
225. Sickmeier, M.; Hamilton, J.A.; LeGall, T.; Vacic, V.; Cortese, M.S.; Tantos, A.; Szabo, B.; Tompa, P.; Chen, J.; Uversky, V.N.; et al. Disprot: The database of disordered proteins. *Nucleic Acids Res.* **2007**, *35*, D786–D793. [[CrossRef](#)] [[PubMed](#)]
226. Finn, R.D.; Bateman, A.; Clements, J.; Coghill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2014**, *42*, D222–D230. [[CrossRef](#)] [[PubMed](#)]
227. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [[CrossRef](#)] [[PubMed](#)]
228. Liu, J.G.; Perumal, N.B.; Oldfield, C.J.; Su, E.W.; Uversky, V.N.; Dunker, A.K. Intrinsic disorder in transcription factors. *Biochemistry* **2006**, *45*, 6873–6888. [[CrossRef](#)] [[PubMed](#)]
229. Toth-Petroczy, A.; Oldfield, C.J.; Simon, I.; Takagi, Y.; Dunker, A.K.; Uversky, V.N.; Fuxreiter, M. Malleable machines in transcription regulation: The mediator complex. *PLoS Comput. Biol.* **2008**, *4*, e1000243. [[CrossRef](#)] [[PubMed](#)]
230. Dunker, A.K.; Uversky, V.N. Drugs for ‘protein clouds’: Targeting intrinsically disordered transcription factors. *Curr. Opin. Pharmacol.* **2010**, *10*, 782–788. [[CrossRef](#)] [[PubMed](#)]
231. Westerheide, S.D.; Raynes, R.; Powell, C.; Xue, B.; Uversky, V.N. HSF transcription factor family, heat shock response, and protein intrinsic disorder. *Curr. Protein Pept. Sci.* **2012**, *13*, 86–103. [[CrossRef](#)] [[PubMed](#)]
232. Minezaki, Y.; Homma, K.; Kinjo, A.R.; Nishikawa, K. Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J. Mol. Biol.* **2006**, *359*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
233. Malik, S.; Grzeschik, K.H. Synpolydactyly: Clinical and molecular advances. *Clin. Genet.* **2008**, *73*, 113–120. [[CrossRef](#)] [[PubMed](#)]

234. Cantile, M.; Franco, R.; Tschan, A.; Baumhoer, D.; Zlobec, I.; Schiavo, G.; Forte, I.; Bihl, M.; Liguori, G.; Botti, G.; et al. HOX D13 expression across 79 tumor tissue types. *Int. J. Cancer* **2009**, *125*, 1532–1541. [[CrossRef](#)] [[PubMed](#)]
235. Cantile, M.; Franco, R.; Forte, I.; Cerrone, M.; Anniciello, A.; Liguori, G.; Manna, A.; Corrado, A.; Aquino, G.; Terracciano, L.; et al. HOX D13 expression across 79 tumor tissue types and its prognostic role in pancreatic cancer. *Virchows Arch.* **2009**, *455*, 391.
236. Lappin, T.R.; Grier, D.G.; Thompson, A.; Halliday, H.L. HOX genes: Seductive science, mysterious mechanisms. *Ulster Med. J.* **2006**, *75*, 23–31. [[PubMed](#)]
237. Cantile, M.; Scognamiglio, G.; La Sala, L.; La Mantia, E.; Scaramuzza, V.; Valentino, E.; Tatangelo, F.; Losito, S.; Pezzullo, L.; Chiofalo, M.G.; et al. Aberrant expression of posterior HOX genes in well differentiated histotypes of thyroid cancers. *Int. J. Mol. Sci.* **2013**, *14*, 21727–21740. [[CrossRef](#)] [[PubMed](#)]
238. Chang, S.; Liu, J.S.; Guo, S.C.; He, S.C.; Qiu, G.L.; Lu, J.; Wang, J.; Fan, L.; Zhao, W.; Che, X.M. Hottip and HOXA13 are oncogenes associated with gastric cancer progression. *Oncol. Rep.* **2016**, *35*, 3577–3585. [[CrossRef](#)] [[PubMed](#)]
239. Williams, T.M.; Williams, M.E.; Innis, J.W. Range of HOX/TALE superclass associations and proteins domain requirements for HOXA13 : MEIS interaction. *Dev. Biol.* **2005**, *277*, 457–471. [[CrossRef](#)] [[PubMed](#)]
240. Gunasekaran, K.; Tsai, C.J.; Nussinov, R. Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers. *J. Mol. Biol.* **2004**, *341*, 1327–1341. [[CrossRef](#)] [[PubMed](#)]
241. Fong, J.H.; Panchenko, A.R. Intrinsic disorder and protein multibinding in domain, terminal, and linker regions. *Mol. Biosyst.* **2010**, *6*, 1821–1828. [[CrossRef](#)] [[PubMed](#)]
242. Wu, Z.H.; Hu, G.; Yang, J.Y.; Peng, Z.L.; Uversky, V.N.; Kurgan, L. In various protein complexes, disordered protomers have large per-residue surface areas and area of protein-, DNA- and RNA-binding interfaces. *FEBS Lett.* **2015**, *589*, 2561–2569. [[CrossRef](#)] [[PubMed](#)]
243. Morrison, N.A.; Stephens, A.S.; Osato, M.; Pasco, J.A.; Fozzard, N.; Stein, G.S.; Polly, P.; Griffiths, L.R.; Nicholson, G.C. Polyalanine repeat polymorphism in runx2 is associated with site-specific fracture in post-menopausal females. *PLoS ONE* **2013**, *8*, e72740. [[CrossRef](#)] [[PubMed](#)]
244. Zhu, P.P.; Wang, Y.Y.; He, L.; Huang, G.L.; Du, Y.; Zhang, G.; Yan, X.L.; Xia, P.Y.; Ye, B.Q.; Wang, S.; et al. ZIC2-dependent OCT4 activation drives self-renewal of human liver cancer stem cells. *J. Clin. Investig.* **2015**, *125*, 3795–3808. [[CrossRef](#)] [[PubMed](#)]
245. Sasaki, A.; Kanai, M.; Kijima, K.; Akaba, K.; Hashimoto, M.; Hasegawa, H.; Otaki, S.; Koizumi, T.; Kusuda, S.; Ogawa, Y.; et al. Molecular analysis of congenital central hypoventilation syndrome. *Hum. Genet.* **2003**, *114*, 22–26. [[CrossRef](#)] [[PubMed](#)]
246. Amiel, J.; Laudier, B.; Attie-Bitach, T.; Trang, H.; de Pontual, L.; Gener, B.; Trochet, D.; Etchevers, H.; Ray, P.; Simonneau, M.; et al. Polyalanine expansion and frameshift mutations of the paired-like homeobox gene PHOX2B in congenital central hypoventilation syndrome. *Nat. Genet.* **2003**, *33*, 459–461. [[CrossRef](#)] [[PubMed](#)]
247. Trochet, D.; O'Brien, L.M.; Gozal, D.; Trang, H.; Nordenskjold, A.; Laudier, B.; Svensson, P.J.; Uhrig, S.; Cole, T.; Munnich, A.; et al. PHOX2B genotype allows for prediction of tumor risk in congenital central hypoventilation syndrome. *Am. J. Hum. Genet.* **2005**, *76*, 421–426. [[CrossRef](#)] [[PubMed](#)]
248. Bourdeaut, F.; Trochet, D.; Janoueix-Lerosey, I.; Ribeiro, A.; Deville, A.; Coz, C.; Michiels, J.F.; Lyonnet, S.; Amiel, J.; Delattre, O. Germline mutations of the paired-like homeobox 2b (PHOX2B) gene in neuroblastoma. *Cancer Lett.* **2005**, *228*, 51–58. [[CrossRef](#)] [[PubMed](#)]
249. Trochet, D.; Bourdeaut, F.; Janoueix-Lerosey, I.; Deville, A.; de Pontual, L.; Schleiermacher, G.; Coze, C.; Philip, N.; Frebourg, T.; Munnich, A.; et al. Germline mutations of the paired-like homeobox 2b (PHOX2B) gene in neuroblastoma. *Am. J. Hum. Genet.* **2004**, *74*, 761–764. [[CrossRef](#)] [[PubMed](#)]
250. Laumonnier, F.; Ronce, N.; Hamel, B.C.; Thomas, P.; Lespinasse, J.; Raynaud, M.; Paringaux, C.; Van Bokhoven, H.; Kalscheuer, V.; Fryns, J.P.; et al. Transcription factor SOX3 is involved in x-linked mental retardation with growth hormone deficiency. *Am. J. Hum. Genet.* **2002**, *71*, 1450–1455. [[CrossRef](#)] [[PubMed](#)]
251. Woods, K.S.; Cundall, M.; Turton, J.; Rizotti, K.; Mehta, A.; Palmer, R.; Wong, J.; Chong, W.K.; Al-Zyoud, M.; El-Ali, M.; et al. Over- and underdosage of SOX3 is associated with infundibular hypoplasia and hypopituitarism. *Am. J. Hum. Genet.* **2005**, *76*, 833–849. [[CrossRef](#)] [[PubMed](#)]

252. Sutton, E.; Hughes, J.; White, S.; Sekido, R.; Tan, J.; Arboleda, V.; Rogers, N.; Knowler, K.; Rowley, L.; Eyre, H.; et al. Identification of SOX3 as an XX male sex reversal gene in mice and humans. *J. Clin. Investig.* **2011**, *121*, 328–341. [[CrossRef](#)] [[PubMed](#)]
253. Li, K.; Wang, R.W.; Jiang, Y.G.; Zou, Y.B.; Guo, W. Overexpression of SOX3 is associated with diminished prognosis in esophageal squamous cell carcinoma. *Ann. Surg. Oncol.* **2013**, *20* (Suppl. 3), S459–S466. [[CrossRef](#)] [[PubMed](#)]
254. Vural, B.; Chen, L.C.; Saip, P.; Chen, Y.T.; Ustuner, Z.; Gonen, M.; Simpson, A.J.; Old, L.J.; Ozbek, U.; Gure, A.O. Frequency of SOX group b (SOX1, 2, 3) and zic2 antibodies in turkish patients with small cell lung carcinoma and their correlation with clinical parameters. *Cancer* **2005**, *103*, 2575–2583. [[CrossRef](#)] [[PubMed](#)]
255. Kim, R.; Trubetskoy, A.; Suzuki, T.; Jenkins, N.A.; Copeland, N.G.; Lenz, J. Genome-based identification of cancer genes by proviral tagging in mouse retrovirus-induced T-cell lymphomas. *J. Virol.* **2003**, *77*, 2056–2062. [[CrossRef](#)] [[PubMed](#)]
256. Xia, Y.; Papalopulu, N.; Vogt, P.K.; Li, J. The oncogenic potential of the high mobility group box protein SOX3. *Cancer Res.* **2000**, *60*, 6303–6306. [[PubMed](#)]
257. Gure, A.O.; Stockert, E.; Scanlan, M.J.; Keresztes, R.S.; Jager, D.; Altorki, N.K.; Old, L.J.; Chen, Y.T. Serological identification of embryonic neural proteins as highly immunogenic tumor antigens in small cell lung cancer. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 4198–4203. [[CrossRef](#)] [[PubMed](#)]
258. Bienvenu, T.; Poirier, K.; Friocourt, G.; Bahi, N.; Beaumont, D.; Fauchereau, F.; Ben Jeema, L.; Zemni, R.; Vinet, M.C.; Francis, F.; et al. ARX, a novel Prd-class-homeobox gene highly expressed in the telencephalon, is mutated in x-linked mental retardation. *Hum. Mol. Genet.* **2002**, *11*, 981–991. [[CrossRef](#)] [[PubMed](#)]
259. Kato, M.; Das, S.; Petras, K.; Kitamura, K.; Morohashi, K.; Abuelo, D.N.; Barr, M.; Bonneau, D.; Brady, A.F.; Carpenter, N.J.; et al. Mutations of ARX are associated with striking pleiotropy and consistent genotype-phenotype correlation. *Hum. Mutat.* **2004**, *23*, 147–159. [[CrossRef](#)] [[PubMed](#)]
260. Turner, G.; Partington, M.; Kerr, B.; Mangelsdorf, M.; Gecz, J. Variable expression of mental retardation, autism, seizures, and dystonic hand movements in two families with an identical ARX gene mutation. *Am. J. Med. Genet.* **2002**, *112*, 405–411. [[CrossRef](#)] [[PubMed](#)]
261. Stromme, P.; Mangelsdorf, M.E.; Shaw, M.A.; Lower, K.M.; Lewis, S.M.; Bruyere, H.; Lutchera, V.; Gedeon, A.K.; Wallace, R.H.; Scheffer, I.E.; et al. Mutations in the human ortholog of aristaless cause x-linked mental retardation and epilepsy. *Nat. Genet.* **2002**, *30*, 441–445. [[CrossRef](#)] [[PubMed](#)]
262. Kitamura, K.; Yanazawa, M.; Sugiyama, N.; Miura, H.; Iizuka-Kogo, A.; Kusaka, M.; Omichi, K.; Suzuki, R.; Kato-Fukui, Y.; Kamiirisa, K.; et al. Mutation of ARX causes abnormal development of forebrain and testes in mice and x-linked lissencephaly with abnormal genitalia in humans. *Nat. Genet.* **2002**, *32*, 359–369. [[CrossRef](#)] [[PubMed](#)]
263. Stromme, P.; Bakke, S.J.; Dahl, A.; Gecz, J. Brain cysts associated with mutation in the aristaless related homeobox gene, ARX. *J. Neurol. Neurosurg. Psychiatry* **2003**, *74*, 536–538. [[CrossRef](#)] [[PubMed](#)]
264. Verdin, H.; De Baere, E. Blepharophimosis, ptosis, and epicanthus inversus. In *Genereviews(r)*; Pagon, R.A., Adam, M.P., Ardinger, H.H., Wallace, S.E., Amemiya, A., Bean, L.J.H., Bird, T.D., Ledbetter, N., Mefford, H.C., Smith, R.J.H., et al. Eds.; University of Washington: Seattle, WA, USA, 1993.
265. Harris, S.E.; Chand, A.L.; Winship, I.M.; Gersak, K.; Aittomaki, K.; Shelling, A.N. Identification of novel mutations in FOXL2 associated with premature ovarian failure. *Mol. Hum. Reprod.* **2002**, *8*, 729–733. [[CrossRef](#)] [[PubMed](#)]
266. Laissue, P.; Lakhal, B.; Benayoun, B.A.; Dipietromaria, A.; Braham, R.; Elghezal, H.; Philibert, P.; Saad, A.; Sultan, C.; Fellous, M.; et al. Functional evidence implicating FOXL2 in non-syndromic premature ovarian failure and in the regulation of the transcription factor osr2. *J. Med. Genet.* **2009**, *46*, 455–457. [[CrossRef](#)] [[PubMed](#)]
267. Fuller, P.J.; Leung, D.; Chu, S. Genetics and genomics of ovarian sex cord-stromal tumors. *Clin. Genet.* **2017**, *91*, 285–291. [[CrossRef](#)] [[PubMed](#)]
268. Ge, H.; Zhou, D.; Tong, S.; Gao, Y.; Teng, M.; Niu, L. Crystal structure and possible dimerization of the single RRM of human PABPN1. *Proteins* **2008**, *71*, 1539–1545. [[CrossRef](#)] [[PubMed](#)]
269. Van der Sluijs, B.M.; van Engelen, B.G.; Hoefsloot, L.H. Oculopharyngeal muscular dystrophy (OPMD) due to a small duplication in the PABPN1 gene. *Hum. Mutat.* **2003**, *21*, 553. [[CrossRef](#)] [[PubMed](#)]

270. Ohshima, K.; Kanto, K.; Hatakeyama, K.; Ide, T.; Wakabayashi-Nakao, K.; Watanabe, Y.; Sakura, N.; Terashima, M.; Yamaguchi, K.; Mochizuki, T. Exosome-mediated extracellular release of polyadenylate-binding protein 1 in human metastatic duodenal cancer cells. *Proteomics* **2014**, *14*, 2297–2306. [[CrossRef](#)] [[PubMed](#)]
271. Ichinose, J.; Watanabe, K.; Sano, A.; Nagase, T.; Nakajima, J.; Fukayama, M.; Yatomi, Y.; Ohishi, N.; Takai, D. Alternative polyadenylation is associated with lower expression of PABPN1 and poor prognosis in non-small cell lung cancer. *Cancer Sci.* **2014**, *105*, 1135–1141. [[CrossRef](#)] [[PubMed](#)]
272. Grube, S.; Gerchen, M.F.; Adamcio, B.; Pardo, L.A.; Martin, S.; Malzahn, D.; Papiol, S.; Begemann, M.; Ribbe, K.; Friedrichs, H.; et al. A CAG repeat polymorphism of KCNN3 predicts SK3 channel function and cognitive performance in schizophrenia. *EMBO Mol. Med.* **2011**, *3*, 309–319. [[CrossRef](#)] [[PubMed](#)]
273. Gueguinou, M.; Harnois, T.; Crottes, D.; Uguen, A.; Deliot, N.; Gambade, A.; Chantome, A.; Haelters, J.P.; Jaffres, P.A.; Jourdan, M.L.; et al. SK3/TRPC1/Orai1 complex regulates SOCE-dependent colon cancer cell migration: A novel opportunity to modulate anti-EGFR mab action by the alkyl-lipid ohmline. *Oncotarget* **2016**, *7*, 36168–36184. [[CrossRef](#)] [[PubMed](#)]
274. Steinestel, K.; Eder, S.; Ehinger, K.; Schneider, J.; Genze, F.; Winkler, E.; Wardelmann, E.; Schrader, A.J.; Steinestel, J. The small conductance calcium-activated potassium channel 3 (SK3) is a molecular target for Edelfosine to reduce the invasive potential of urothelial carcinoma cells. *Tumour Biol.* **2016**, *37*, 6275–6283. [[CrossRef](#)] [[PubMed](#)]
275. Clarysse, L.; Gueguinou, M.; Potier-Cartereau, M.; Vandecasteele, G.; Bougnoux, P.; Chevalier, S.; Chantome, A.; Vandier, C. cAMP-PKA inhibition of SK3 channel reduced both  $Ca^{2+}$  entry and cancer cell migration by regulation of SK3-Orai1 complex. *Pflugers Arch.* **2014**, *466*, 1921–1932. [[CrossRef](#)] [[PubMed](#)]
276. Chantome, A.; Potier-Cartereau, M.; Clarysse, L.; Fromont, G.; Marionneau-Lambot, S.; Gueguinou, M.; Pages, J.C.; Collin, C.; Oullier, T.; Girault, A.; et al. Pivotal role of the lipid raft SK3-Orai1 complex in human cancer cell migration and bone metastases. *Cancer Res.* **2013**, *73*, 4852–4861. [[CrossRef](#)] [[PubMed](#)]
277. Stevanin, G.; Camuzat, A.; Holmes, S.E.; Julien, C.; Sahloul, R.; Dode, C.; Hahn-Barma, V.; Ross, C.A.; Margolis, R.L.; Durr, A.; et al. CAG/CTG repeat expansions at the huntington's disease-like 2 locus are rare in huntington's disease patients. *Neurology* **2002**, *58*, 965–967. [[CrossRef](#)] [[PubMed](#)]
278. Walker, R.H.; Rasmussen, A.; Rudnicki, D.; Holmes, S.E.; Alonso, E.; Matsuura, T.; Ashizawa, T.; Davidoff-Feldman, B.; Margolis, R.L. Huntington's disease—Like 2 can present as chorea-acanthocytosis. *Neurology* **2003**, *61*, 1002–1004. [[CrossRef](#)] [[PubMed](#)]
279. Faber, P.W.; Barnes, G.T.; Srinidhi, J.; Chen, J.; Gusella, J.F.; MacDonald, M.E. Huntingtin interacts with a family of WW domain proteins. *Hum. Mol. Genet.* **1998**, *7*, 1463–1474. [[CrossRef](#)] [[PubMed](#)]
280. Boutell, J.M.; Thomas, P.; Neal, J.W.; Weston, V.J.; Duce, J.; Harper, P.S.; Jones, A.L. Aberrant interactions of transcriptional repressor proteins with the huntington's disease gene product, huntingtin. *Hum. Mol. Genet.* **1999**, *8*, 1647–1655. [[CrossRef](#)] [[PubMed](#)]
281. Steffan, J.S.; Kazantsev, A.; Spasic-Boskovic, O.; Greenwald, M.; Zhu, Y.Z.; Gohler, H.; Wanker, E.E.; Bates, G.P.; Housman, D.E.; Thompson, L.M. The huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6763–6768. [[CrossRef](#)] [[PubMed](#)]
282. Schaeper, U.; Boyd, J.M.; Verma, S.; Uhlmann, E.; Subramanian, T.; Chinnadurai, G. Molecular cloning and characterization of a cellular phosphoprotein that interacts with a conserved C-terminal domain of adenovirus E1A involved in negative modulation of oncogenic transformation. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10467–10471. [[CrossRef](#)] [[PubMed](#)]
283. Raychaudhuri, S.; Majumder, P.; Sarkar, S.; Giri, K.; Mukhopadhyay, D.; Bhattacharyya, N.P. Huntingtin interacting protein HYPK is intrinsically unstructured. *Proteins* **2008**, *71*, 1686–1698. [[CrossRef](#)] [[PubMed](#)]
284. Nopoulos, P.C. Huntington disease: A single-gene degenerative disorder of the striatum. *Dialogues Clin. Neurosci.* **2016**, *18*, 91–98. [[PubMed](#)]
285. Onodera, O.; Oyake, M.; Takano, H.; Ikeuchi, T.; Igarashi, S.; Tsuji, S. Molecular cloning of a full-length cDNA for dentatorubral-pallidoluysian atrophy and regional expressions of the expanded alleles in the CNS. *Am. J. Hum. Genet.* **1995**, *57*, 1050–1060. [[PubMed](#)]



286. Yazawa, I.; Nukina, N.; Hashida, H.; Goto, J.; Yamada, M.; Kanazawa, I. Abnormal gene product identified in hereditary dentatorubral-pallidoluysian atrophy (DRPLA) brain. *Nat. Genet.* **1995**, *10*, 99–103. [[CrossRef](#)] [[PubMed](#)]
287. Hou, R.; Sibinga, N.E. Atrophin proteins interact with the fat1 cadherin and regulate migration and orientation in vascular smooth muscle cells. *J. Biol. Chem.* **2009**, *284*, 6955–6965. [[CrossRef](#)] [[PubMed](#)]
288. Suzuki, Y.; Yazawa, I. Pathological accumulation of Atrophin-1 in dentatorubralpallidoluysian atrophy. *Int. J. Clin. Exp. Pathol.* **2011**, *4*, 378–384. [[PubMed](#)]
289. Chen, H.; Fang, Y.; Zhu, H.; Li, S.; Wang, T.; Gu, P.; Fang, X.; Wu, Y.; Liang, J.; Zeng, Y.; et al. Protein-protein interaction analysis of distinct molecular pathways in two subtypes of colorectal carcinoma. *Mol. Med. Rep.* **2014**, *10*, 2868–2874. [[CrossRef](#)] [[PubMed](#)]
290. Claessens, F.; Denayer, S.; Van Tilborgh, N.; Kerkhofs, S.; Helsen, C.; Haelens, A. Diverse roles of androgen receptor (AR) domains in AR-mediated signaling. *Nucl. Recept. Signal.* **2008**, *6*, e008. [[CrossRef](#)] [[PubMed](#)]
291. Lavery, D.N.; McEwan, I.J. Structural characterization of the native NH2-terminal transactivation domain of the human androgen receptor: A collapsed disordered conformation underlies structural plasticity and protein-induced folding. *Biochemistry* **2008**, *47*, 3360–3369. [[CrossRef](#)] [[PubMed](#)]
292. McEwan, I.J.; Lavery, D.; Fischer, K.; Watt, K. Natural disordered sequences in the amino terminal domain of nuclear receptors: Lessons from the androgen and glucocorticoid receptors. *Nucl. Recept. Signal.* **2007**, *5*, e001. [[CrossRef](#)] [[PubMed](#)]
293. Echaniz-Laguna, A.; Rouso, E.; Anheim, M.; Cossee, M.; Tranchant, C. A family with early-onset and rapidly progressive x-linked spinal and bulbar muscular atrophy. *Neurology* **2005**, *64*, 1458–1460. [[CrossRef](#)] [[PubMed](#)]
294. Shukla, G.C.; Plaga, A.R.; Shankar, E.; Gupta, S. Androgen receptor-related diseases: What do we know? *Andrology* **2016**, *4*, 366–381. [[CrossRef](#)] [[PubMed](#)]
295. Tong, X.; Gui, H.; Jin, F.; Heck, B.W.; Lin, P.; Ma, J.; Fondell, J.D.; Tsai, C.C. Ataxin-1 and brother of ataxin-1 are components of the notch signalling pathway. *EMBO Rep.* **2011**, *12*, 428–435. [[CrossRef](#)] [[PubMed](#)]
296. Hong, S.; Kim, S.J.; Ka, S.; Choi, I.; Kang, S. USP7, a ubiquitin-specific protease, interacts with ataxin-1, the SCA1 gene product. *Mol. Cell. Neurosci.* **2002**, *20*, 298–306. [[CrossRef](#)] [[PubMed](#)]
297. Chen, Y.W.; Allen, M.D.; Veprintsev, D.B.; Lowe, J.; Bycroft, M. The structure of the AXH domain of spinocerebellar ataxin-1. *J. Biol. Chem.* **2004**, *279*, 3758–3765. [[CrossRef](#)] [[PubMed](#)]
298. Servadio, A.; Koshy, B.; Armstrong, D.; Antalffy, B.; Orr, H.T.; Zoghbi, H.Y. Expression analysis of the ataxin-1 protein in tissues from normal and spinocerebellar ataxia type 1 individuals. *Nat. Genet.* **1995**, *10*, 94–98. [[CrossRef](#)] [[PubMed](#)]
299. Kang, A.R.; An, H.T.; Ko, J.; Kang, S. Ataxin-1 regulates epithelial-mesenchymal transition of cervical cancer cells. *Oncotarget* **2017**, *8*, 18248–18259. [[CrossRef](#)] [[PubMed](#)]
300. Nonis, D.; Schmidt, M.H.; van de Loo, S.; Eich, F.; Dikic, I.; Nowock, J.; Auburger, G. Ataxin-2 associates with the endocytosis complex and affects EGF receptor trafficking. *Cell. Signal.* **2008**, *20*, 1725–1739. [[CrossRef](#)] [[PubMed](#)]
301. Albrecht, M.; Golatta, M.; Wullner, U.; Lengauer, T. Structural and functional analysis of ataxin-2 and ataxin-3. *Eur. J. Biochem.* **2004**, *271*, 3155–3170. [[CrossRef](#)] [[PubMed](#)]
302. Elden, A.C.; Kim, H.J.; Hart, M.P.; Chen-Plotkin, A.S.; Johnson, B.S.; Fang, X.; Armakola, M.; Geser, F.; Greene, R.; Lu, M.M.; et al. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* **2010**, *466*, 1069–1075. [[CrossRef](#)] [[PubMed](#)]
303. Pulst, S.M.; Nechiporuk, A.; Nechiporuk, T.; Gispert, S.; Chen, X.N.; Lopes-Cendes, I.; Pearlman, S.; Starkman, S.; Orozco-Diaz, G.; Lunkes, A.; et al. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat. Genet.* **1996**, *14*, 269–276. [[CrossRef](#)] [[PubMed](#)]
304. Nkilaiza, A.; Chartier-Harlin, M.C. ATXN2 a culprit with multiple facets. *Oncotarget* **2017**, *8*, 34028. [[CrossRef](#)] [[PubMed](#)]
305. Nkilaiza, A.; Mutez, E.; Simonin, C.; Lepretre, F.; Duflo, A.; Figeac, M.; Villenet, C.; Semaille, P.; Comptaer, T.; Genet, A.; et al. RNA-binding disturbances as a continuum from spinocerebellar ataxia type 2 to parkinson disease. *Neurobiol. Dis.* **2016**, *96*, 312–322. [[CrossRef](#)] [[PubMed](#)]
306. Bailey, J.N.; Loomis, S.J.; Kang, J.H.; Allingham, R.R.; Gharahkhani, P.; Khor, C.C.; Burdon, K.P.; Aschard, H.; Chasman, D.I.; Igo, R.P., Jr.; et al. Genome-wide association analysis identifies TXNRD2, ATXN2 and FOXC1 as susceptibility loci for primary open-angle glaucoma. *Nat. Genet.* **2016**, *48*, 189–194. [[CrossRef](#)] [[PubMed](#)]



307. Wiedemeyer, R.; Westermann, F.; Wittke, I.; Nowock, J.; Schwab, M. Ataxin-2 promotes apoptosis of human neuroblastoma cells. *Oncogene* **2003**, *22*, 401–411. [[CrossRef](#)] [[PubMed](#)]
308. Mao, Y.; Senic-Matuglia, F.; Di Fiore, P.P.; Polo, S.; Hodsdon, M.E.; De Camilli, P. Deubiquitinating function of ataxin-3: Insights from the solution structure of the Josephin Domain. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12700–12705. [[CrossRef](#)] [[PubMed](#)]
309. Seki, T.; Gong, L.; Williams, A.J.; Sakai, N.; Todi, S.V.; Paulson, H.L. JOSD1, a membrane-targeted deubiquitinating enzyme, is activated by ubiquitination and regulates membrane dynamics, cell motility, and endocytosis. *J. Biol. Chem.* **2013**, *288*, 17145–17155. [[CrossRef](#)] [[PubMed](#)]
310. Tzvetkov, N.; Breuer, P. Josephin Domain-containing proteins from a variety of species are active de-ubiquitination enzymes. *Biol. Chem.* **2007**, *388*, 973–978. [[CrossRef](#)] [[PubMed](#)]
311. Li, F.; Macfarlan, T.; Pittman, R.N.; Chakravarti, D. Ataxin-3 is a histone-binding protein with two independent transcriptional corepressor activities. *J. Biol. Chem.* **2002**, *277*, 45004–45012. [[CrossRef](#)] [[PubMed](#)]
312. Masino, L.; Musi, V.; Menon, R.P.; Fusi, P.; Kelly, G.; Frenkiel, T.A.; Trottier, Y.; Pastore, A. Domain architecture of the polyglutamine protein ataxin-3: A globular domain followed by a flexible tail. *FEBS Lett.* **2003**, *549*, 21–25. [[CrossRef](#)]
313. Albrecht, M.; Hoffmann, D.; Evert, B.O.; Schmitt, I.; Wullner, U.; Lengauer, T. Structural modeling of ataxin-3 reveals distant homology to adaptins. *Proteins* **2003**, *50*, 355–370. [[CrossRef](#)] [[PubMed](#)]
314. Zeng, L.X.; Tang, Y.; Ma, Y. Ataxin-3 expression correlates with the clinicopathologic features of gastric cancer. *Int. J. Clin. Exp. Med.* **2014**, *7*, 973–981. [[PubMed](#)]
315. Zhuchenko, O.; Bailey, J.; Bonnen, P.; Ashizawa, T.; Stockton, D.W.; Amos, C.; Dobyns, W.B.; Subramony, S.H.; Zoghbi, H.Y.; Lee, C.C. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the  $\alpha$ 1A-voltage-dependent calcium channel. *Nat. Genet.* **1997**, *15*, 62–69. [[CrossRef](#)] [[PubMed](#)]
316. Terwindt, G.; Kors, E.; Haan, J.; Vermeulen, F.; Van den Maagdenberg, A.; Frants, R.; Ferrari, M. Mutation analysis of the CACNA1A calcium channel subunit gene in 27 patients with sporadic hemiplegic migraine. *Arch. Neurol.* **2002**, *59*, 1016–1018. [[CrossRef](#)] [[PubMed](#)]
317. Ophoff, R.A.; Terwindt, G.M.; Vergouwe, M.N.; van Eijk, R.; Oefner, P.J.; Hoffman, S.M.; Lamerdin, J.E.; Mohrenweiser, H.W.; Bulman, D.E.; Ferrari, M.; et al. Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the  $\text{Ca}^{2+}$  channel gene *cacn1a4*. *Cell* **1996**, *87*, 543–552. [[CrossRef](#)]
318. Carrera, P.; Piatti, M.; Stenirri, S.; Grimaldi, L.M.; Marchioni, E.; Curcio, M.; Righetti, P.G.; Ferrari, M.; Gelfi, C. Genetic heterogeneity in Italian families with familial hemiplegic migraine. *Neurology* **1999**, *53*, 26–33. [[CrossRef](#)] [[PubMed](#)]
319. Ducros, A.; Denier, C.; Joutel, A.; Cecillon, M.; Lescoat, C.; Vahedi, K.; Darcel, F.; Vicaut, E.; Bousser, M.G.; Tournier-Lasserre, E. The clinical spectrum of familial hemiplegic migraine associated with mutations in a neuronal calcium channel. *N. Engl. J. Med.* **2001**, *345*, 17–24. [[CrossRef](#)] [[PubMed](#)]
320. Stam, A.H.; Vanmolkot, K.R.; Kremer, H.P.; Gartner, J.; Brown, J.; Leshinsky-Silver, E.; Gilad, R.; Kors, E.E.; Frankhuizen, W.S.; Ginjaar, H.B.; et al. CACNA1A R1347Q: A frequent recurrent mutation in hemiplegic migraine. *Clin. Genet.* **2008**, *74*, 481–485. [[CrossRef](#)] [[PubMed](#)]
321. Epi4K Consortium De novo mutations in SLC1A2 and CACNA1A are important causes of epileptic encephalopathies. *Am. J. Hum. Genet.* **2016**, *99*, 287–298.
322. Reinson, K.; Oiglane-Shlik, E.; Talvik, I.; Vaher, U.; Ounapuu, A.; Ennok, M.; Teek, R.; Pajusalu, S.; Murumets, U.; Tomberg, T.; et al. Biallelic CACNA1A mutations cause early onset epileptic encephalopathy with progressive cerebral, cerebellar, and optic nerve atrophy. *Am. J. Med. Genet. A* **2016**, *170*, 2173–2176. [[CrossRef](#)] [[PubMed](#)]
323. Aung, T.; Ozaki, M.; Mizoguchi, T.; Allingham, R.R.; Li, Z.; Haripriya, A.; Nakano, S.; Uebe, S.; Harder, J.M.; Chan, A.S.; et al. A common variant mapping to CACNA1A is associated with susceptibility to exfoliation syndrome. *Nat. Genet.* **2015**, *47*, 387–392. [[CrossRef](#)] [[PubMed](#)]
324. Wang, C.Y.; Lai, M.D.; Phan, N.N.; Sun, Z.; Lin, Y.C. Meta-analysis of public microarray datasets reveals voltage-gated calcium gene signatures in clinical cancer patients. *PLoS ONE* **2015**, *10*, e0125766. [[CrossRef](#)] [[PubMed](#)]

325. Palhan, V.B.; Chen, S.; Peng, G.H.; Tjernberg, A.; Gamper, A.M.; Fan, Y.; Chait, B.T.; La Spada, A.R.; Roeder, R.G. Polyglutamine-expanded ataxin-7 inhibits staga histone acetyltransferase activity to produce retinal degeneration. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 8472–8477. [[CrossRef](#)] [[PubMed](#)]
326. Stevanin, G.; Durr, A.; Brice, A. Clinical and molecular advances in autosomal dominant cerebellar ataxias: From genotype to phenotype and physiopathology. *Eur. J. Hum. Genet.* **2000**, *8*, 4–18. [[CrossRef](#)] [[PubMed](#)]
327. Milne, R.L.; Burwinkel, B.; Michailidou, K.; Arias-Perez, J.I.; Zamora, M.P.; Menendez-Rodriguez, P.; Hardisson, D.; Mendiola, M.; Gonzalez-Neira, A.; Pita, G.; et al. Common non-synonymous SNPs associated with breast cancer susceptibility: Findings from the breast cancer association consortium. *Hum. Mol. Genet.* **2014**, *23*, 6096–6111. [[CrossRef](#)] [[PubMed](#)]
328. Kalvala, A.; Gao, L.; Aguila, B.; Dotts, K.; Rahman, M.; Nana-Sinkam, S.P.; Zhou, X.; Wang, Q.E.; Amann, J.; Otterson, G.A.; et al. Rad51C-ATXN7 fusion gene expression in colorectal tumors. *Mol. Cancer* **2016**, *15*, 47. [[CrossRef](#)] [[PubMed](#)]
329. Bharti, B.; Mishra, R. Spleen-specific isoforms of Pax5 and ataxin-7 as potential proteomic markers of lymphoma-affected spleen. *Mol. Cell. Biochem.* **2015**, *402*, 181–191. [[CrossRef](#)] [[PubMed](#)]
330. He, Y.; Yan, C.; Fang, J.; Inouye, C.; Tjian, R.; Ivanov, I.; Nogales, E. Near-atomic resolution visualization of human transcription promoter opening. *Nature* **2016**, *533*, 359–365. [[CrossRef](#)] [[PubMed](#)]
331. LeRoy, G.; Orphanides, G.; Lane, W.S.; Reinberg, D. Requirement of RSF and fact for transcription of chromatin templates in vitro. *Science* **1998**, *282*, 1900–1904. [[CrossRef](#)] [[PubMed](#)]
332. Hoffman, A.; Sinn, E.; Yamamoto, T.; Wang, J.; Roy, A.; Horikoshi, M.; Roeder, R.G. Highly conserved core domain and unique N terminus with presumptive regulatory motifs in a human tata factor (TFIID). *Nature* **1990**, *346*, 387–390. [[CrossRef](#)] [[PubMed](#)]
333. Peterson, M.G.; Tanese, N.; Pugh, B.F.; Tjian, R. Functional domains and upstream activation properties of cloned human TATA binding protein. *Science* **1990**, *248*, 1625–1630. [[CrossRef](#)] [[PubMed](#)]
334. Kao, C.C.; Lieberman, P.M.; Schmidt, M.C.; Zhou, Q.; Pei, R.; Berk, A.J. Cloning of a transcriptionally active human TATA binding factor. *Science* **1990**, *248*, 1646–1650. [[CrossRef](#)] [[PubMed](#)]
335. Gouge, J.; Satia, K.; Guthertz, N.; Widya, M.; Thompson, A.J.; Cousin, P.; Dergai, O.; Hernandez, N.; Vannini, A. Redox signaling by the RNA polymerase III TFIIB-related factor Brf2. *Cell* **2015**, *163*, 1375–1387. [[CrossRef](#)] [[PubMed](#)]
336. Friedrich, J.K.; Panov, K.I.; Cabart, P.; Russell, J.; Zomerdijs, J.C. TBP-TAF complex SL1 directs Rna polymerase I pre-initiation complex formation and stabilizes upstream binding factor at the rDNA promoter. *J. Biol. Chem.* **2005**, *280*, 29551–29558. [[CrossRef](#)] [[PubMed](#)]
337. Hochheimer, A.; Tjian, R. Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. *Genes Dev.* **2003**, *17*, 1309–1320. [[CrossRef](#)] [[PubMed](#)]
338. Friedman, M.J.; Shah, A.G.; Fang, Z.H.; Ward, E.G.; Warren, S.T.; Li, S.; Li, X.J. Polyglutamine domain modulates the TBP-TFIIB interaction: Implications for its normal function and neurodegeneration. *Nat. Neurosci.* **2007**, *10*, 1519–1528. [[CrossRef](#)] [[PubMed](#)]
339. Burley, S.K. The TATA box binding protein. *Curr. Opin. Struct. Biol.* **1996**, *6*, 69–75. [[CrossRef](#)]
340. Lescure, A.; Lutz, Y.; Eberhard, D.; Jacq, X.; Krol, A.; Grummt, I.; Davidson, I.; Chambon, P.; Tora, L. The N-terminal domain of the human TATA-binding protein plays a role in transcription from TATA-containing RNA polymerase II and III promoters. *EMBO J.* **1994**, *13*, 1166–1175. [[PubMed](#)]
341. Kenny, P.J.; Zhou, H.; Kim, M.; Skariah, G.; Khetani, R.S.; Drnevich, J.; Arcila, M.L.; Kosik, K.S.; Ceman, S. MOV10 and fmrp regulate AGO2 association with microRNA recognition elements. *Cell Rep.* **2014**, *9*, 1729–1741. [[CrossRef](#)] [[PubMed](#)]
342. Ascano, M., Jr.; Mukherjee, N.; Bandaru, P.; Miller, J.B.; Nusbaum, J.D.; Corcoran, D.L.; Langlois, C.; Munschauer, M.; Dewell, S.; Hafner, M.; et al. FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* **2012**, *492*, 382–386. [[CrossRef](#)] [[PubMed](#)]
343. Bechara, E.G.; Didiot, M.C.; Melko, M.; Davidovic, L.; Bensaid, M.; Martin, P.; Castets, M.; Pognonec, P.; Khandjian, E.W.; Moine, H.; et al. A novel function for fragile X mental retardation protein in translational activation. *PLoS Biol.* **2009**, *7*, e16. [[CrossRef](#)] [[PubMed](#)]
344. Didiot, M.C.; Tian, Z.; Schaeffer, C.; Subramanian, M.; Mandel, J.L.; Moine, H. The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Res.* **2008**, *36*, 4902–4912. [[CrossRef](#)] [[PubMed](#)]

345. Antar, L.N.; Li, C.; Zhang, H.; Carroll, R.C.; Bassell, G.J. Local functions for FMRP in axon growth cone motility and activity-dependent regulation of filopodia and spine synapses. *Mol. Cell. Neurosci.* **2006**, *32*, 37–48. [[CrossRef](#)] [[PubMed](#)]
346. Hayashi, T.; Lombaert, I.M.; Hauser, B.R.; Patel, V.N.; Hoffman, M.P. Exosomal microRNA transport from salivary mesenchyme regulates epithelial progenitor expansion during organogenesis. *Dev. Cell* **2017**, *40*, 95–103. [[CrossRef](#)] [[PubMed](#)]
347. Bensaid, M.; Melko, M.; Bechara, E.G.; Davidovic, L.; Berretta, A.; Catania, M.V.; Gecz, J.; Lalli, E.; Bardoni, B. FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure. *Nucleic Acids Res.* **2009**, *37*, 1269–1279. [[CrossRef](#)] [[PubMed](#)]
348. Majounie, E.; Renton, A.E.; Mok, K.; Doppert, E.G.; Waite, A.; Rollinson, S.; Chio, A.; Restagno, G.; Nicolaou, N.; Simon-Sanchez, J.; et al. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: A cross-sectional study. *Lancet Neurol.* **2012**, *11*, 323–330. [[CrossRef](#)]
349. Chio, A.; Borghero, G.; Restagno, G.; Mora, G.; Drepper, C.; Traynor, B.J.; Sendtner, M.; Brunetti, M.; Ossola, I.; Calvo, A.; et al. Clinical characteristics of patients with familial amyotrophic lateral sclerosis carrying the pathogenic GGGGCC hexanucleotide repeat expansion of C9orf72. *Brain* **2012**, *135*, 784–793. [[CrossRef](#)] [[PubMed](#)]
350. Daoud, H.; Suhail, H.; Sabbagh, M.; Belzil, V.; Szuto, A.; Dionne-Laporte, A.; Khoris, J.; Camu, W.; Salachas, F.; Meininger, V.; et al. C9orf72 hexanucleotide repeat expansions as the causative mutation for chromosome 9p21-linked amyotrophic lateral sclerosis and frontotemporal dementia. *Arch. Neurol.* **2012**, *69*, 1159–1163. [[CrossRef](#)] [[PubMed](#)]
351. Garcia-Redondo, A.; Dols-Icardo, O.; Rojas-Garcia, R.; Esteban-Perez, J.; Cordero-Vazquez, P.; Munoz-Blanco, J.L.; Catalina, I.; Gonzalez-Munoz, M.; Varona, L.; Sarasola, E.; et al. Analysis of the C9orf72 gene in patients with amyotrophic lateral sclerosis in Spain and different populations worldwide. *Hum. Mutat.* **2013**, *34*, 79–82. [[CrossRef](#)] [[PubMed](#)]
352. Cooper-Knock, J.; Bury, J.J.; Heath, P.R.; Wyles, M.; Higginbottom, A.; Gelsthorpe, C.; Highley, J.R.; Hautbergue, G.; Rattray, M.; Kirby, J.; et al. C9orf72 GGGGCC expanded repeats produce splicing dysregulation which correlates with disease severity in amyotrophic lateral sclerosis. *PLoS ONE* **2015**, *10*, e0127376. [[CrossRef](#)] [[PubMed](#)]
353. Zu, T.; Liu, Y.; Banez-Coronel, M.; Reid, T.; Pletnikova, O.; Lewis, J.; Miller, T.M.; Harms, M.B.; Falchook, A.E.; Subramony, S.H.; et al. RAN proteins and RNA FOCI from antisense transcripts in C9orf72 ALS and frontotemporal dementia. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E4968–E4977. [[CrossRef](#)] [[PubMed](#)]
354. Mizielinska, S.; Lashley, T.; Norona, F.E.; Clayton, E.L.; Ridler, C.E.; Fratta, P.; Isaacs, A.M. C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA FOCI. *Acta Neuropathol.* **2013**, *126*, 845–857. [[CrossRef](#)] [[PubMed](#)]
355. Gendron, T.F.; Bieniek, K.F.; Zhang, Y.J.; Jansen-West, K.; Ash, P.E.; Caulfield, T.; Daugherty, L.; Dunmore, J.H.; Castaneda-Casey, M.; Chew, J.; et al. Antisense transcripts of the expanded C9orf72 hexanucleotide repeat form nuclear RNA FOCI and undergo repeat-associated non-ATG translation in C9FTD/ALS. *Acta Neuropathol.* **2013**, *126*, 829–844. [[CrossRef](#)] [[PubMed](#)]
356. Al-Sarraj, S.; King, A.; Troakes, C.; Smith, B.; Maekawa, S.; Bodi, I.; Rogelj, B.; Al-Chalabi, A.; Hortobagyi, T.; Shaw, C.E. P62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked ftd and MND/ALS. *Acta Neuropathol.* **2011**, *122*, 691–702. [[CrossRef](#)] [[PubMed](#)]
357. Mori, K.; Arzberger, T.; Grasser, F.A.; Gijssels, I.; May, S.; Rentzsch, K.; Weng, S.M.; Schludi, M.H.; van der Zee, J.; Cruts, M.; et al. Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathol.* **2013**, *126*, 881–893. [[CrossRef](#)] [[PubMed](#)]
358. Santamaria, N.; Alhothali, M.; Alfonso, M.H.; Breydo, L.; Uversky, V.N. Intrinsic disorder in proteins involved in amyotrophic lateral sclerosis. *Cell. Mol. Life Sci.* **2017**, *74*, 1297–1318. [[CrossRef](#)] [[PubMed](#)]
359. Schoenfeld, R.A.; Napoli, E.; Wong, A.; Zhan, S.; Reutenauer, L.; Morin, D.; Buckpitt, A.R.; Taroni, F.; Lonnerdal, B.; Ristow, M.; et al. Frataxin deficiency alters heme pathway transcripts and decreases mitochondrial heme metabolites in mammalian cells. *Hum. Mol. Genet.* **2005**, *14*, 3787–3799. [[CrossRef](#)] [[PubMed](#)]

360. Yoon, T.; Cowan, J.A. Iron-sulfur cluster biosynthesis. Characterization of frataxin as an iron donor for assembly of [2Fe-2S] clusters in ISU-type proteins. *J. Am. Chem. Soc.* **2003**, *125*, 6078–6084. [[CrossRef](#)] [[PubMed](#)]
361. Bulteau, A.L.; O'Neill, H.A.; Kennedy, M.C.; Ikeda-Saito, M.; Isaya, G.; Szwed, L.I. Frataxin acts as an iron chaperone protein to modulate mitochondrial aconitase activity. *Science* **2004**, *305*, 242–245. [[CrossRef](#)] [[PubMed](#)]
362. O'Neill, H.A.; Gakh, O.; Park, S.; Cui, J.; Mooney, S.M.; Sampson, M.; Ferreira, G.C.; Isaya, G. Assembly of human frataxin is a mechanism for detoxifying redox-active iron. *Biochemistry* **2005**, *44*, 537–545. [[CrossRef](#)] [[PubMed](#)]
363. Watkins, N.J.; Lemm, I.; Ingelfinger, D.; Schneider, C.; Hossbach, M.; Urlaub, H.; Luhrmann, R. Assembly and maturation of the U3 snoRNP in the nucleoplasm in a large dynamic multiprotein complex. *Mol. Cell* **2004**, *16*, 789–798. [[CrossRef](#)] [[PubMed](#)]
364. Hayano, T.; Yanagida, M.; Yamauchi, Y.; Shinkawa, T.; Isobe, T.; Takahashi, N. Proteomic analysis of human Nop56p-associated pre-ribosomal ribonucleoprotein complexes. Possible link between Nop56p and the nucleolar protein treacle responsible for treacher collins syndrome. *J. Biol. Chem.* **2003**, *278*, 34309–34319. [[CrossRef](#)] [[PubMed](#)]
365. Orr, H.T.; Zoghbi, H.Y. Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* **2007**, *30*, 575–621. [[CrossRef](#)] [[PubMed](#)]
366. Scheuermann, T.; Schulz, B.; Blume, A.; Wahle, E.; Rudolph, R.; Schwarz, E. Trinucleotide expansions leading to an extended poly-L-alanine segment in the poly(A) binding protein PABPN1 cause fibril formation. *Protein Sci.* **2003**, *12*, 2685–2692. [[CrossRef](#)] [[PubMed](#)]
367. Wetzel, R. Physical chemistry of polyglutamine: Intriguing tales of a monotonous sequence. *J. Mol. Biol.* **2012**, *421*, 466–490. [[CrossRef](#)] [[PubMed](#)]
368. Leitgeb, B.; Kerenyi, A.; Bogar, F.; Paragi, G.; Penke, B.; Rakhely, G. Studying the structural properties of polyalanine and polyglutamine peptides. *J. Mol. Model.* **2007**, *13*, 1141–1150. [[CrossRef](#)] [[PubMed](#)]
369. Pelassa, I.; Cora, D.; Cesano, F.; Monje, F.J.; Montarolo, P.G.; Fiumara, F. Association of polyalanine and polyglutamine coiled coils mediates expansion disease-related protein aggregation and dysfunction. *Hum. Mol. Genet.* **2014**, *23*, 3402–3420. [[CrossRef](#)] [[PubMed](#)]
370. Walker, F.O. Huntington's disease. *Lancet* **2007**, *369*, 218–228. [[CrossRef](#)]
371. Lee, J.M.; Pinto, R.M.; Gillis, T.; St Claire, J.C.; Wheeler, V.C. Quantification of age-dependent somatic CAG repeat instability in HDH CAG knock-in mice reveals different expansion dynamics in striatum and liver. *PLoS ONE* **2011**, *6*, e23647. [[CrossRef](#)] [[PubMed](#)]
372. Swami, M.; Hendricks, A.E.; Gillis, T.; Massood, T.; Mysore, J.; Myers, R.H.; Wheeler, V.C. Somatic expansion of the huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum. Mol. Genet.* **2009**, *18*, 3039–3047. [[CrossRef](#)] [[PubMed](#)]
373. Dragileva, E.; Hendricks, A.; Teed, A.; Gillis, T.; Lopez, E.T.; Friedberg, E.C.; Kucherlapati, R.; Edelman, W.; Lunetta, K.L.; MacDonald, M.E.; et al. Intergenerational and striatal CAG repeat instability in huntington's disease knock-in mice involve different DNA repair genes. *Neurobiol. Dis.* **2009**, *33*, 37–47. [[CrossRef](#)] [[PubMed](#)]
374. Gonitell, R.; Moffitt, H.; Sathasivam, K.; Woodman, B.; Detloff, P.J.; Faull, R.L.; Bates, G.P. DNA instability in postmitotic neurons. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 3467–3472. [[CrossRef](#)] [[PubMed](#)]
375. Wheeler, V.C.; Lebel, L.A.; Vrbanc, V.; Teed, A.; Riele, H.; MacDonald, M.E. Mismatch repair gene MSH2 modifies the timing of early disease in HDH (Q111) striatum. *Hum. Mol. Genet.* **2003**, *12*, 273–281. [[CrossRef](#)] [[PubMed](#)]
376. Mizushima, N.; Levine, B.; Cuervo, A.M.; Klionsky, D.J. Autophagy fights disease through cellular self-digestion. *Nature* **2008**, *451*, 1069–1075. [[CrossRef](#)] [[PubMed](#)]
377. Powers, E.T.; Morimoto, R.I.; Dillin, A.; Kelly, J.W.; Balch, W.E. Biological and chemical approaches to diseases of proteostasis deficiency. *Annu. Rev. Biochem.* **2009**, *78*, 959–991. [[CrossRef](#)] [[PubMed](#)]
378. Hartl, F.U.; Bracher, A.; Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **2011**, *475*, 324–332. [[CrossRef](#)] [[PubMed](#)]
379. Koga, H.; Kaushik, S.; Cuervo, A.M. Protein homeostasis and aging: The importance of exquisite quality control. *Ageing Res. Rev.* **2011**, *10*, 205–215. [[CrossRef](#)] [[PubMed](#)]

380. Lopez-Otin, C.; Blasco, M.A.; Partridge, L.; Serrano, M.; Kroemer, G. The hallmarks of aging. *Cell* **2013**, *153*, 1194–1217. [[CrossRef](#)] [[PubMed](#)]
381. Tanaka, K.; Matsuda, N. Proteostasis and neurodegeneration: The roles of proteasomal degradation and autophagy. *Biochim. Biophys. Acta* **2014**, *1843*, 197–204. [[CrossRef](#)] [[PubMed](#)]
382. Zu, T.; Gibbens, B.; Doty, N.S.; Gomes-Pereira, M.; Huguet, A.; Stone, M.D.; Margolis, J.; Peterson, M.; Markowski, T.W.; Ingram, M.A.; et al. Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 260–265. [[CrossRef](#)] [[PubMed](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).