

Decoding the role of coiled-coil motifs in human prion-like proteins

Molood Behbahanipour, Javier García-Pardo , and Salvador Ventura *

Institut De Bioteconologia I De Biomedicina (Ibb) and Departament De Bioquímica I Biologia Molecular, Universitat Autònoma De Barcelona, Barcelona, Spain

ABSTRACT

Prions are self-propagating proteins that cause fatal neurodegenerative diseases in humans. However, increasing evidence suggests that eukaryotic cells exploit prion conformational conversion for functional purposes. A recent study delineated a group of twenty prion-like proteins in humans, characterized by the presence of low-complexity glutamine-rich sequences with overlapping coiled-coil (CCs) motifs. This is the case of Mediator complex subunit 15 (MED15), which is overexpressed in a wide range of human cancers. Biophysical studies demonstrated that the prion-like domain (PrLD) of MED15 forms homodimers in solution, sustained by CCs interactions. Furthermore, the same coiled-coil (CC) region plays a crucial role in the PrLD structural transition to a transmissible β -sheet amyloid state. In this review, we discuss the role of CCs motifs and their contribution to amyloid transitions in human prion-like domains (PrLDs), while providing a comprehensive overview of six predicted human prion-like proteins involved in transcription, gene expression, or DNA damage response and associated with human disease, whose PrLDs contain or overlap with CCs sequences. Finally, we try to rationalize how these molecular signatures might relate to both their function and involvement in disease.

ARTICLE HISTORY

Received 3 July 2021
Revised 22 July 2021
Accepted 25 July 2021

KEYWORDS

Prions; amyloids; prion-like domains; coiled-coils; protein aggregation




INTRODUCTION

Prusiner and co-workers coined the term prion in 1982. They reported the purification of infectious particles from scrapie-infected hamster brains for the first time and demonstrated that they consisted of a specific type of pathogenic element that causes fatal neurodegenerative diseases in mammals and humans [1,2]. It has been assumed that the toxicity of prions relies on their intrinsic capacity to lose their native conformation (either in ordered or disordered regions) and acquire a β sheet-rich secondary structure with significant aggregation and self-propagating capacities. However, almost forty years after this seminal discovery, increasing evidence indicates that prion-like conformational conversion is not always pathogenic. On the contrary, it can be exploited for functional roles [3]. The best-characterized examples of non-pathogenic prions are those identified in yeast and filamentous fungi. Both yeast and fungal prions are multi-domain proteins with regions enriched in Asn (N) and/or Gln (Q) residues [2,4]. Typically, these long Q/N-rich sequences map into unstructured regions of the protein termed prion domains (PrDs). Upon a conformational switch to a self-perpetuating conformation, the prionic traits can be inherited during cell division in a non-

Mendelian manner. This atypical mechanism of protein transmission provides increased tolerance to stress and facilitates adaptation to changing environments [5,6].

A less recognized feature of PrDs is the presence of protein sequence stretches with a propensity to form coiled-coils (CCs), often overlapping with the Q/N-rich or polyQ regions [3,7]. This super secondary structure has been associated with the formation of dimers, trimers, and higher-order oligomeric structures. Furthermore, it has been proposed that coiled-coil (CC) formation is an essential step preceding the physiological conformational switches to amyloid-like states in yeast prions [3,7].

Our group recently exploited different bioinformatics tools to systematically analyse the presence of prion-like domains (PrLDs) at the proteome level [2,3,8]. Using this approach, we identified a subset of human polypeptides that bear PrLDs with properties resembling those of yeast PrDs [3,8]. Subsequently, we explored whether these human domains contain regions with the propensity to fold into CCs using the COILS [9] and PARCOIL2 [10] algorithms. A total of 22 of these PrLDs were predicted to display sequences with a high propensity to fold into CCs, with 20 of them presenting polyQ tracts of different lengths.

*CONTACT Salvador Ventura  Salvador.Ventura@uab.es; Javier García-Pardo  javiergarciapardo@msn.com  Institut De Bioteconologia I De Biomedicina (Ibb) and Departament De Bioquímica I Biologia Molecular, Universitat Autònoma De Barcelona, 08193 Bellaterra, Barcelona, Spain

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Interestingly enough, we observed that the majority of the identified proteins, if not all, participate in transcriptional regulation. This is the case of the mediator complex subunit 15 (MED15). MED15 is part of the mediator multi-protein complex that regulates enhancer-driven gene transcription [3] and, under normal conditions, is located in the nucleus [3]. Thus, MED15 knockdown has been linked to reduced growth and decreased transcriptional activation, while MED15 is highly overexpressed in different human cancers, including head and neck squamous cell carcinomas, hepatocellular carcinoma, breast cancer, renal cell carcinoma, and testicular germ cell tumours. Indeed, patients with MED15 over-expression in tumour tissues exhibit bad prognosis, significantly shorter survival times, and more aggressive phenotypes [11]. MED15 contains a disordered N-terminal low-complexity region composed of discontinuous polyQ tracts and a strong propensity to form CCs that map to its PrLD.

Recent work has focused on elucidating the role of the CC motif for structural transitions of initially soluble α -helical regions to β -sheet amyloid states. This has been particularly challenging to study in the past in the context of complete PrLDs, and previous reports dealt with only relatively short peptides derived from yeast prions (*i.e.*, Ure2p PrD). Recently, Battle and co-workers have experimentally demonstrated, for the first time, that is, in fact, the CC motif of MED15 PrLD that mediates the transition towards a β -sheet amyloid state [3]. They proposed that this behaviour may prevent the establishment of relevant protein-protein interactions in pathological situations. Alternatively, MED15 self-association might result in phase separation and the consequent formation of transcriptional hubs. A third possibility is that the PrLD of MED15 would establish pathological interactions with other Q-rich proteins, leading to the formation of cytoplasmic insoluble aggregates. All in all, these recent results reinforce the hypothesis that aggregation of MED15 PrLD into amyloid fibrils involves a progressive conformational switch rather than a simple uncontrolled misfolding mechanism. It is feasible that the exact mechanism applies to the rest of the 19 identified polypeptides. Here we discuss this possibility and its potential implications for six of these predicted prion-like proteins associated with human disease.

LOW COMPLEXITY SEQUENCES IN PRION-LIKE PROTEINS

Prion-like proteins are rich in low complexity sequences. Recent work has shown that binding of

low complexity regions to their physiological partners is often accompanied by a local increase in the structuration of the binding region in a process known as ‘folding upon binding’. This binding and folding mechanism has been previously reported for different CCs, such as the GCN4 CC dimerization domain. This domain shows a two-state unfolding transition in which the binding appears coupled to folding [12]. Similarly, thermodynamic analyses with the CC oligomerization domain (SARAH) from serine/threonine mammalian sterile 20-like kinase (MST1) demonstrated that this domain is unstructured and folds upon binding to different partners [13,14]. However, other structural studies of the GCNA4 CC indicated that a pre-existing helical content promotes complex formation [14–16], which shows that segments of intrinsic helical propensity are important drivers of the interaction. As mentioned, a significant proportion of human prion-like proteins display sequences with a high propensity to fold into CCs that overlap with low complexity sequences. Thus, it is tempting to speculate that the presence of these preformed helical regions may speed up the binding to their partners. After this initial contact, further local folding may still take place within the complex.

COILED COILS MEDIATE FUNCTIONAL INTERACTIONS AND AMYLOID FORMATION

Sequences with Poly-Q repeat motifs are known to have a high propensity to form CCs. These sequences are frequent in yeast prions [4] and human proteins bearing PrLDs [3,7]. CC motifs were traditionally considered as molecular spacers between functional domains [17]. However, a growing body of evidence suggests that they frequently contain interaction regions and act as protein-protein interactors (PPI) and/or catalytic effectors [17]. CC forming proteins are widely distributed and have been shown to play a variety of biological roles [17,18], such as transcription regulation (*i.e.*, leucine zippers) [19]; modulation of chromatin [20]; chromosome dynamics [21]; kinetochore assembly [22]; cell cycle progression [23]; organization of the centrosome [24]; vesicle transport [25], organelles structuration and activation of cell-signalling cascades [26,27], among other functions. Examples of these functions are depicted in Figure 1. CC motifs are particularly abundant in RNA-binding proteins (RBPs) that are localized to liquid-liquid phase-separated (LLPS) neuronal granules [18]. This phase transition has been shown essential for appropriate RNA trafficking during local protein synthesis, but it is also associated with

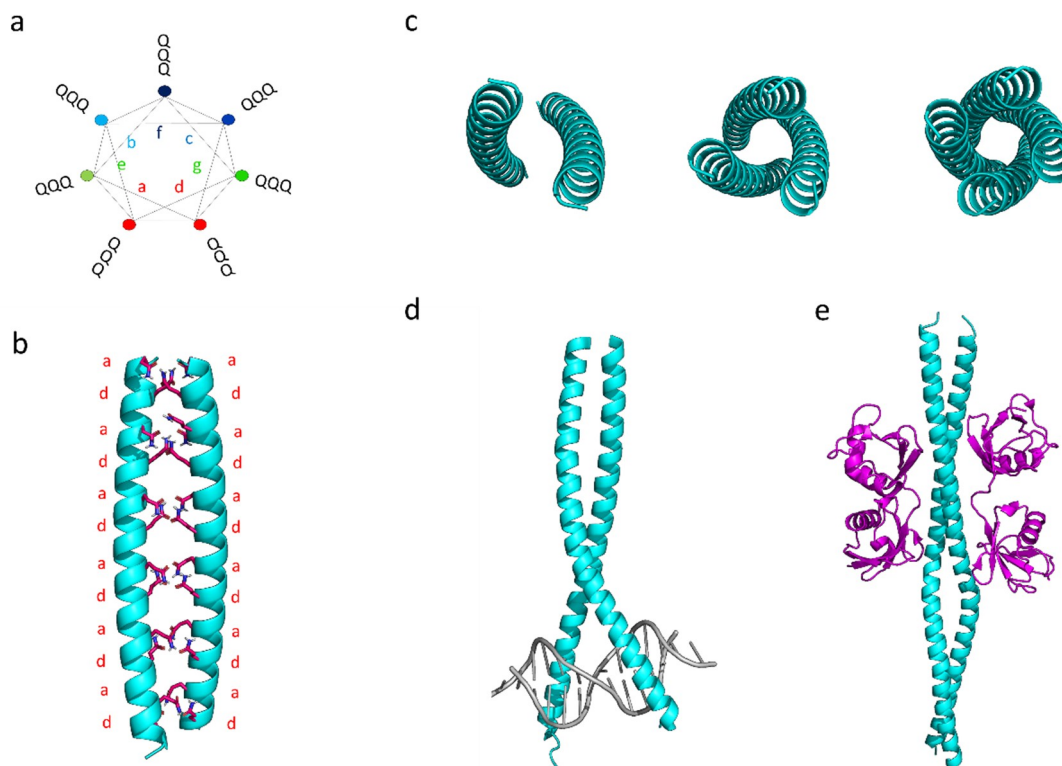


Figure 1. Structural features of coiled-coils built up by Poly-Q repeats. a) Helical wheel illustration for a regular heptad pattern composed built up by 21 Gln residues. The Gln residues are found in all seven (a-b-c-d-e-f-g) positions of the heptad repeat due to the particular chemical nature of this amino acid. b) Modelled structure of a canonical CC composed of 42 Gln residues. c) Structural representation of the oligomerization states driven by the supercoiling of the α -helices in a parallel orientation. d,e) Examples of functional protein-DNA and protein-protein interactions mediated by CCs. d) X-ray crystal structure of the leucine zipper Pap1 bZIP dimer bound to DNA. e) Crystal Structure of a diUb chain bound to the NEMO-UBAN domain. PDB codes are: 1GD2 and 2ZVN for d) and e), respectively.

misfolding and amyloid formation in various human neurological disorders [18].

The structure of CCs is now well understood, and this fold is ubiquitous in all domains of life (comprising up to 10% of all proteins in a given species) [25]. Naturally occurring CCs consist of at least two supercoiled α -helices. In their most prevalent form, they are built up from a heptad repeat *abcdefg*, in which the amino acids at positions *a* and *d* are mostly conserved hydrophobic residues (see Figure 1a). The resultant α -helices display a high amphipathic character, with both hydrophobic and polar faces. Yeast PrDs and human PrLDs possess sequence stretches in which Gln residues are abundant. Gln is both neutral and polar amino acid, with a long side chain that can be located on the surface of α -helices and at the theoretically hydrophobic a/d positions of the heptad repeats (see Figure 1a and Figure 1b). This dual behaviour allows for a variety of energetically favourable oligomerization states driven by the supercoiling of the α -helices (Figure 1c). Thus, it has been proposed that CC-mediated oligomerization of Q-rich proteins could drive to amyloid formation,

given appropriate stimulus or induced by changes in the environmental conditions. In support of this hypothesis, Hartmann and co-workers reported that insertions of two or six residues strain the supercoil and lead to the local formation of β -strands (also known as α - β coiled-coil) [28]. Additionally, experimental data indicate a possible direct structural conversion of CCs themselves to cross β -sheet amyloids. This structural shift was demonstrated in short peptides that form α -helical CCs when subjected to heating or pH changes [7], implying a structural shift from the intramolecular hydrogen-bonding network that maintains CCs to the typical intermolecular-bonding landscape present in the β -sheets amyloids [29].

As discussed above, a recent study by our group supports this mechanism being behind the formation of highly ordered amyloids by the PrLDs of MED15. Other human proteins that may respond to the exact mechanism were discovered in the same study but could not be discussed in detail. In the following sections, we describe the cases of TATA-box-Binding Protein (TBP), cAMP-response element-protein

plex. In this process, TBP binds to TATA-BOX sequences in two different steps. The first step requires the interaction of the inhibitory DNA-binding (IDB) surface from the C-terminal part of TBP to positively regulate the formation of the unstable DNA unbent complex. The second step involves a conformational switch that leads to forming a stable DNA bent complex [35]. The N-terminal region inhibits the formation of the bent complex and favours that of the unbent one, and conformational changes in this region are necessary to form the active assembly.

Recent cryo-EM structural studies have revealed many details concerning the early steps of TFIID assembly [36]. TBP interacts with up to 13 different TBP-associated factors (TAFs) to build up the TFIID complex, with six of them (TAF4, TAF5, TAF6, TAF9, TAF10, and TAF12) present in two copies. This large macromolecular complex of about 1.3 MDa is essential for recognizing core promoter sequences and recruiting the preinitiation complex (PIC) during basal transcription. Our previous analysis of TBP-associated factors (TAFs) revealed an overrepresentation of Q-rich sequences and overlapping CCs motifs [3]. This trend has been previously described by Fiumara and co-workers in yeast, who reported that many of the protein interactors of yeast functional prions contain CCs [7]. For instance, human TAF4 contains four glutamine-rich (Q-rich) domains involved in mediating interactions with transcriptional activators.

TBP has also been linked to the key pathological features of some neurodegenerative disorders, including neuronal intranuclear hyaline inclusion disease (NIH-ID) [37], spinocerebellar ataxia types 1, 2, 3 (SCA1, SCA2, SCA3) [38,39], dentatorubral-pallidoluysian atrophy (DRPLA) [38], Huntington's disease (HD) [40], and Alzheimer's disease (AD) [41]. In particular, expansion of the polyQ tract beyond 42 residues in TBP has been associated with the development of cerebellar ataxia (SCA17), an autosomal dominant and progressive neurodegenerative disease [42]. Pathological expansion in its polyQ domain has been proposed to induce neurodegeneration. This expansion caused TBP aggregation and precipitation within neurons [43,44], decreased its dimerization, altered its binding to other transcription factors. This behaviour is compatible with the presence of sequence stretches able to assemble into highly order amyloid-like structures. The emerging picture is that these so-called soft amyloid cores (SAC), are necessary for prion conversion in yeast as well as in human cells [3]. We identified a similar region in TBP that comprises residues 93-QQQQAVAAAQVQQSTSQQATQ-112 by using pWALTZ [45]. This sequence is located immediately

after the CC motif and connects with the adjacent doublet of TBP domains (Figure 2a). The presence of this SAC might facilitate amyloid formation by decreasing the energy barrier for the conformation transition, with the adjacent CC region facilitating the subsequent intermolecular contacts, as demonstrated for MED15 [3].

In summary, TBP plays an essential role in transcription initiation, and the presence of PrLD in this protein might be important for this function, through its binding to DNA or by its interaction and regulation of/by other transcriptional players. The questions are if, and how, mutations/expansions at its Q-rich region might promote amyloid conversion and which would be the role of CCs in this putative transition. Answering them, should help to understand the contribution of the PrLD identified in TBP to disease onset and progression.

CAMP-RESPONSE ELEMENT-BINDING PROTEIN (CBP)

cAMP-response element-protein-(CREB) binding protein (CBP) is a transcriptional coactivator localized in the cell nucleus. CBP has a lysine acetyltransferase domain (HAT in Figure 3a) located in the central region of the protein and catalyzes the acetylation of target proteins [46]. CBP interacts with a variety of cell signalling proteins, especially those with established pro-survival effects in neurons, such as the CRE-binding protein (CREB) [47] and plays a crucial role in critical biological processes, such as embryogenesis, development, differentiation, and apoptosis [48]. Thanks to its HAT motif, CBP is endowed with histone acetyltransferase activity, promoting the acetylation of histones, which influences chromatin condensation and is a key mechanism in regulating transcription [46]. On the other hand, the N- and C-terminal domains of CBP can act as transactivation domains, and the protein also contains three potential α -helical motifs containing the sequence LXXLL (amino acids 68–78, 355–365, and 2067–2077) called NR boxes, to mediate interactions with nuclear receptors. The C-terminal NR box lies within the C-terminal low complexity region of CBP and interacts with other coactivators, including activator for thyroid hormone and retinoid receptors (ACTR) and the steroid receptor coactivator 1 (SRC-1) [49–51].

Detailed sequence analysis of CBP showed that this protein contains up to six discontinuous segments with prion-like features (Figure 3a). In fact, the C-terminal PrLDs identified in this protein overlap with two predicted CC motifs and with a tract of 18 consecutive Gln residues near its carboxy-terminal tail (residues 2199–2216). This polyQ motif has been involved in the

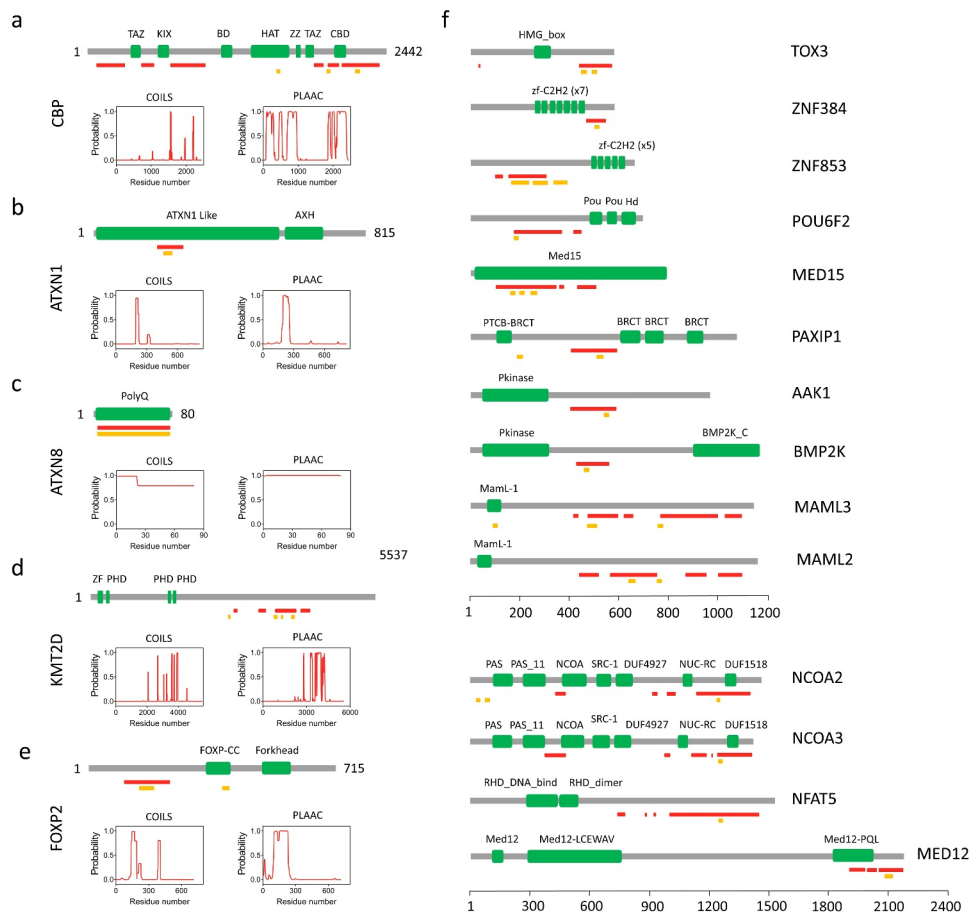


Figure 3. Human proteins containing PrLDs and overlapping CC regions. a-e) Linear representation of a) CBP, b) ATXN1, c) ATXN8, d) KMT2D, and e) FOXP2 showing the location of the Pfam domains (green) (3). The prion-like domains (red) based on PLAAC [34] predictions (0.8–1 score) and the CC regions (yellow) using COILS [9] predictor are indicated (0.8–1 score). In a-e), the COILS and PLAAC per-residue probability plots have been included. f) Linear representation of other human prion-like proteins with overlapping CCs. Note that the COILS score (> 0.3) has been considered for AAK1 and NFAT5.

sequestration of CBP to aggregates of several distinct polyQ proteins [52]. Indeed, CBP dysfunction has been identified in a number of neurological disorders, such as the Rubinstein–Taybi syndrome (RTS), AD, Amyotrophic Lateral Sclerosis (ALS) and polyQ-related diseases, in which CBP has been found in the inclusion bodies formed by polyglutamine-containing proteins in HD, DRPLA, SCA3, SCA7, and spinal and bulbar muscular atrophy (SBMA) [52–56]. CBP depletion from its normal location and sequestration into aggregates through interaction with expanded polyQ proteins has been suggested to be mediated by its C-terminal polyQ tract [52]. As a result, impaired CBP-mediated transcription and the lack of CBP-dependent acetylation are associated with increased cell death rates. It is tempting to propose that the C-terminal region of CBP and specifically its Q-rich, CC-containing, PrLD might be involved in a conformational switch, similar to the one described

for MED15, which might facilitate polyQ-mediated interactions and CBP aggregation.

ATAXIN-1 (ATXN1) AND ATAXIN-8 (ATXN8)

The spinocerebellar ataxias (SCAs) are a group of autosomal dominantly inherited progressive disorders associated with cerebellar degeneration and progressive ataxia [57]. Almost thirty different genetic causes of SCAs have been reported so far and are numbered chronologically in order of discovery [57]. Most of them have shown to progress with classic cerebellar signs [57]. However, many variants display disabling non-cerebellar features such as brainstem dysfunction, eye movement abnormalities, and visual loss [57].

Ataxin-1 (ATXN1) is an RNA binding protein with a role in transcriptional regulation. ATXN1 contains an RNA-binding motif (residues 541–767) comprising a globular AXH domain (residues 568–689) involved

in ATXN1 interactions. Predictions using PLAAC and COILS suggested that ATXN1 encodes prionic traits at the N-terminal region, with five CC heptad repeats overlapping with the detected PrLD (Figure 3b).

Mutations in the ataxin-1 gene have been associated with the development of spinocerebellar ataxia type 1 (SCA type 1). This protein is located in the nucleus and interacts with RBM17 (a splicing factor) and several transcription regulators, including SMRT, HDAC3 and Capicua [58–60]. Wild-type ATXN1 has a length of 816 residues. The presence of a trinucleotide repeat disorder caused by the expansion of the CAG repeat in the N-terminal region of ATXN1 leads to an expanded polyQ tract and disease development. This polyQ tract (residues 197–225) is located at the N-terminal region of the protein, mapping at the predicted PrLD and CC motifs. This expansion in ATXN1 favours its binding to the transcriptional repressor protein Capicua homolog via the RNA-binding protein and spliceosome component RBP7. This aberrant binding leads to a disruption in gene expression and splicing events in the affected neurons [57]. ATXN1 shuttles between the nucleus and cytoplasm. However, ATXN1 dynamics are altered by the Gln expansion and failure in nuclear export upon polyQ expansion dramatically reduces the retro-transport into the cytoplasm [61]. Regardless of the polyQ expansion, subtle increases in wild type ATXN1 levels could also lead to SCA1 hallmarks [62], whereas decreasing ATXN1 expression helps to reduce ATXN1's accumulation and mitigates cerebellar SCA1 pathogenesis [63–66]. Conversely, it also has been shown that loss of the protein increases the levels of BACE1 and A β pathology, indicating that ATXN1 levels is a potential factor for AD development [67]. Co-regulation by microRNAs (miRNAs) together with the RNA-binding protein PUMILIO1 (PUM1) are shown to fine-tune posttranscriptionally ATXN1 mRNA levels [62,63,68]. Posttranslational modifications such as phosphorylation, transglutamination, ubiquitination, and sumoylation are also required for the ATXN1 function [60]. Furthermore, the N-terminal α -helical CC domain in the PrLD of MED15 has been shown to efficiently promote spontaneous ATXN1 aggregation in vitro, implying that CC regions in modifiers are essential to promote aggregation and toxicity effects on polyQ diseases [69].

Much less is known about ATXN8. Expansions in this protein have been linked to spinocerebellar ataxia type 8 (SCA8) in humans. These polyQ expansions showed a reduced penetrance compared to the expansions observed for other SCA-associated disease genes [57]. A difference with ATXN1 is that the expansion on CAG repeats in ATXN8 leads to a nearly pure polyQ

protein. As shown in Figure 3c and according to PLAAC and COILS predictions, almost all the 80-residues of ATXN8 map to a PrLD with the potential to acquire a CC fold. It remains unclear how the structural features of proteins bearing long polyQ stretches are translated with toxic phenotypes. This might depend on the unique capacity of these proteins to target other cellular proteins through CC-to-CC interactions. The concept of CC-mediated aggregation also provides a plausible mechanism for polyQ-protein deposition in the brain.

LYSINE METHYLTRANSFERASE 2D (KMT2D)

The *KMT2D* gene (also called MLL4, ALR, or MLL2) is a member of the SET family of histone methyltransferase enzymes. *KMT2D* gene in human maps to 12q13.12, including over 19 kb pairs in length and contains up to 56 exons, encoding a large polypeptide with 5537 amino acids. The transcript of this gene has a molecular weight of 593 kDa and catalyzes the methylation (either mono-, di-, and trimethylation) on lysine 4 (K4) of the histone H3 protein (H3K4). This is an important histone modification that controls transcriptionally active promoters and enhancers [70], increasing the transcription of the gene packaged around the histones [71].

KMT2D contains a cluster of conserved domains at the C-terminal end, comprising a PHD (plant homeodomain)-zinc-finger like domain, two phenylalanine, and tyrosine (FY)-rich motifs (F/Y-rich N-terminus (FYRN) and F/Y-rich C-terminus (FYRC)) and a catalytic SET domain (SET: Su(var)3–9, Enhancer-of-zeste (E(z)), and Trithorax) that are needed for its enzymatic function [72]. *KMT2D* is a major regulator of cell-type-specific gene expression in cell differentiation during tissue development and embryogenesis [73,74]. In addition, *KMT2D* has also been related to tumour suppression and immune signalling [74–76].

In humans, *KMT2D* contains a high predicted disorder content of (55%), a common characteristic of nuclear proteins involved in transcription and chromatin organization, with a series of polyQ tracts at the C terminus end of the highly conserved central regions. These regions are part of a large low complexity segment, consisting of four discontinuous regions with predicted prion-like features (Figure 3d), with the largest prion-like region containing a set of three consecutive CC motifs that overlap with the polyQ repeats.

Most *KMT2D* mutations are assumed to cause truncated proteins that do not perform the function properly due to loss of the catalytic SET domain [72]. Mutated *KMT2D* has been linked to developmental

disorders including Kabuki syndrome [77], congenital heart disease [78], and multiple cancer types such as medulloblastoma, lymphoma, hepatocellular carcinoma, gastric cancer, breast cancer and prostate cancer [73,74]. This pleiotropic effect results from the fact that its pathogenic variants cause the interruption of histone methylation and abnormal enhancer regulation, leading to changes in transcription, thus affecting normal growth and development [74]. The exon 39 (3581–4510 aa) constitutes a mutational hotspot [79] and, interestingly enough, it contains the most extended prion-like domain, with missense disease-associated mutations clustering at the predicted CC and Q-rich motifs comprising residues 3897–3975 [70]. Whether such mutations predispose to disease by disrupting normal protein-protein interactions, mediated by KMT2D during chromatin remodelling, or promoting a conformational transition towards an aggregated state remains to be elucidated.

FORKHEAD BOX PROTEIN P2 (FOXP2)

FOXP2 is a multifunctional transcription factor expressed in various brain regions and peripheral organs during embryonic development and adulthood [80,81]. In humans, FOXP2 has 715 residues (80kDa) and is one of the most highly conserved proteins in vertebrate genomes [82]. It has been shown to impact gene regulation in multiple aspects of neuronal development [83–85]. Indeed, Forkhead-box protein P2 (FOXP2) was one of the first genes to be linked to human language disorder, characterized by a broader cognitive dysfunction and primary motor impairment. Thus, this protein is essential for the normal development of speech and language [86–88]. Human oncogenesis has also been linked to dysregulated FOXP2 function [89].

FOXP2 contains different sequence motifs, including a single C2H2 zinc finger (C2H2-ZF) motif and a leucine zipper (LZ) motif that mediates FOXP2 dimerization through CC formation. This domain has a critical role in DNA binding and promotes heterotypic and homotypic protein interactions with FOXP1/2/4 family members [90,91]. In addition, FOXP2 has a strongly conserved forkhead DNA-binding domain (FHD, 100-aa or ‘winged-helix’) at the C-terminus that can form a domain-swapped dimer, and two nuclear localization signals (NLS) [92,93]. As shown in [Figure 3e](#), a long and disordered predicted PrLD lies near the N-terminus in human FOXP2. It overlaps with a glutamine-rich region, that contains a long perfect polyQ repeat made of up to 40 CAG/CCG repeats (p.Q152-Q191) and a shorter imperfect repeat (p.Q200-

Q230) adjacent to it, with both Gln-rich regions displaying a high propensity to form CCs, according to COILS. Thus, FOXP2 shares all the sequential features shown to promote a CC to amyloid transition in the case of MED15 and it can be considered a potential functional human prion-like protein.

OTHER PREDICTED PRION-LIKE HUMAN PROTEINS WITH COILED-COILS

As previously mentioned, a total of 20 human prion-like domains (PrLDs) containing high CC propensity sequences that, in the majority of cases, overlapped with polyQ stretches were identified in the human proteome [3]. These predicted prion-like proteins work in the regulation of transcription, and they comprise key transcription coactivators, including MED12, MED15, MAML3 and MAML2, CBP, protein kinase and ATP binding activities like BMP2K, AAK1 or DNA-binding transcription factors such as TBP, NFAT5, NCOA3, NCOA2 and FOXP2 ([Figure 3f](#)). For most of them, it is unknown whether they can access a prionic state and, in this case, if α -helical CC within PrLDs would be self-sufficient mediators of functional prions aggregation or just intermediates or facilitators in the β -sheet formation process. In any case, we foresee an increasing interest in studying the verisimilitude of these transitions, provided the functional relevance of this prion-like proteins subset.

CONCLUSIONS

Recent observations suggest that the coincidence of low complexity regions and CC-forming sequences in the PrLDs of prion-like proteins is a frequent feature that might be critical for their function. In humans, these proteins are key functional mediators in transcription and are associated with divergent disorders such as neurodegenerative diseases or cancer. Although only a reduced set of functional prions have been deeply characterized, compared to the vast number of pathologic amyloid proteins, some specific traits can already be delineated. A feature common to almost all prion-like proteins is the spatial separation of the active globular and prionic domains. The prionic traits are concentrated in low complexity regions, which, as described here, might display a high propensity to form CCs and mediate protein-protein interactions, either directly or through a folding-upon-binding mechanism. Many human prion-like proteins interact with DNA and/or RNA, a property mediated by their globular domains, and that for some of them is associated with their liquid-liquid phase separation and the formation of membraneless compartments.

Although it is assumed that pathological conditions result from the inherent toxicity acquired by the aggregates of these proteins, our results with MED15 suggest an alternative explanation in which cytotoxicity is associated with the sequestration of the protein in the wrong compartment, its inactivation upon clustering with a consequent loss of crucial protein-protein interactions or conversely the establishment of undesired intermolecular interactions with dysregulate signalling pathways. The CCs would be important players in these processes, and indeed a significant number of pathological mutations map into these domains, either decreasing or increasing their helical propensity.

Despite progress in the last decades in the field of prions and amyloids, many questions remain to be answered. For example, we still need to address how human prion-like proteins participate in transcriptional mechanisms under specific environmental conditions and how dynamic are the stimuli that trigger their amyloid conversion *in vivo*. Another crucial question is how the conformational changes in PrLDs occur and how they lead to amyloid formation. The structural flexibility of low-complexity regions and the CC-mediated transition model provides a plausible mechanism to explain the functional to pathogenic conversions of at least a fraction of these polypeptides. Providing molecular evidence for such conformational transition in the proteins discussed in this review might uncover an additional layer of transcription regulation and help to understand why these proteins are indefectibly associated with human disease.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was funded by the Spanish Ministry of Science and Innovation (PID2019-105017RB-I00) to S.V., by ICREA, ICREA-Academia 2015 and 2020 to S.V. and by EU (PhasAge /H2020-WIDESPREAD-2020-5) to SV. J.G.-P. was supported by the Spanish Ministry of Science, Innovation and Universities with a postdoctoral grant Juan de la Cierva Incorporacion (IJC2019-041039-I).

AUTHOR CONTRIBUTIONS

M.B. contributed to the preparation and writing of the manuscript. J.G.P. and SV contributed to the conceptualization, writing, and revision of the manuscript.

ORCID

Javier García-Pardo  <http://orcid.org/0000-0001-9179-6371>
Salvador Ventura  <http://orcid.org/0000-0002-9652-6351>

References

- [1] Prusiner SB. Novel proteinaceous infectious particles cause scrapie. *Science*. 1982;216(4542):136–144.
- [2] Gil-Garcia M, Iglesias V, Pallares I, et al. Prion-like proteins: from computational approaches to proteome-wide analysis. *FEBS open bio*. 2021 May 31.1-12. doi: 10.1002/2211-5463.13213
- [3] Batlle C, Calvo I, Iglesias V, et al. MED15 prion-like domain forms a coiled-coil responsible for its amyloid conversion and propagation. *Commun Biol*. 2021;4(1):414.
- [4] Batlle C, Iglesias V, Navarro S, et al. Prion-like proteins and their computational identification in proteomes. *Expert Rev Proteomic*. 2017;14(4):335–350.
- [5] Wickner RB, Edsles HK, Shewmaker F, et al. Prions of fungi: inherited structures and biological roles. *Nature Rev Microbiol*. 2007;5(8):611–618.
- [6] Gemayel R, Chavali S, Pougach K, et al. Variable glutamine-rich repeats modulate transcription factor activity. *Mol Cell*. 2015;59(4):615–627.
- [7] Fiumara F, Fioriti L, Kandel ER, et al. Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell*. 2010;143(7):1121–1135.
- [8] Iglesias V, Paladin L, Juan-Blanco T, et al. In silico characterization of human prion-like proteins: beyond neurological diseases. *Front Physiol*. 2019;10:314.
- [9] Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. *Science*. 1991;252(5009):1162–1164.
- [10] McDonnell AV, Jiang T, Keating AE, et al. Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics*. 2006;22(3):356–358.
- [11] Wang KY, Duan CX, Zou XJ, et al. Increased mediator complex subunit 15 expression is associated with poor prognosis in hepatocellular carcinoma. *Oncol Lett*. 2018;15(4):4303–4313.
- [12] Dragan AI, Privalov PL. Unfolding of a leucine zipper is not a simple two-state transition. *J Mol Biol*. 2002;321(5):891–908.
- [13] Constantinescu Aruxandei D, Makbul C, Koturenkiene A, et al. Dimerization-induced folding of MST1 SARAH and the influence of the intrinsically unstructured inhibitory domain: low thermodynamic stability of monomer. *Biochemistry*. 2011;50(51):10990–11000.
- [14] Walavalkar NM, Gordon N, Williams DC Jr. Unique features of the anti-parallel, heterodimeric coiled-coil interaction between methyl-cytosine binding domain 2 (MBD2) homologues and GATA zinc finger domain containing 2A (GATAD2A/p66alpha). *J Biol Chem*. 2013;288(5):3419–3427.
- [15] Myers JK, Oas TG. Reinterpretation of GCN4-p1 folding kinetics: partial helix formation precedes dimerization in coiled coil folding. *J Mol Biol*. 1999;289(2):205–209.

- [16] Zitzewitz JA, Ibarra-Molero B, Fishel DR, et al. Prefolded secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system. *J Mol Biol.* 2000;296(4):1105–1116.
- [17] Truebestein L, Leonard TA. Coiled-coils: the long and short of it. *BioEssays.* 2016;38(9):903–916.
- [18] Ford LK, Fioriti L. Coiled-coil motifs of RNA-binding proteins: dynamicity in RNA regulation. *Front Cell Dev Biol.* 2020;8:1383.
- [19] Oshea EK, Rutkowski R, Kim PS. Evidence That the Leucine Zipper Is a Coiled Coil. *Science.* 1989;243(4890):538–542.
- [20] Miyaji-Yamaguchi M, Okuwaki M, Nagata K. Coiled-coil structure-mediated dimerization of template activating factor-I is critical for its chromatin remodeling activity. *J Mol Biol.* 1999;290(2):547–557.
- [21] Matityahu A, Onn I. A new twist in the coil: functions of the coiled-coil domain of structural maintenance of chromosome (SMC) proteins. *Curr Genet.* 2018;64(1):109–116.
- [22] Petrovic A, Keller J, Liu YH, et al. Structure of the MIS12 complex and molecular basis of its interaction with CENP-C at human kinetochores. *Cell.* 2016;167(4):1028–+.
- [23] Patzke S, Hauge H, Sioud M, et al. Identification of a novel centrosome/microtubule-associated coiled-coil protein involved in cell-cycle progression and spindle organization. *Oncogene.* 2005;24(7):1159–1173.
- [24] Salisbury JL. Centrosomes: coiled-coils organize the cell center. *Curr Biol.* 2003;13(3):R88–90.
- [25] Rose A, Schraegle SJ, Stahlberg EA, et al. Coiled-coil protein composition of 22 proteomes—differences and common themes in subcellular infrastructure and traffic control. *BMC Evol Biol.* 2005;5(1):66.
- [26] Maculins T, Garcia-Pardo J, Skenderovic A, et al. Discovery of protein-protein interaction inhibitors by integrating protein engineering and chemical screening platforms. *Cell Chem Biol.* 2020;27(11):1441–1451 e7.
- [27] Rahighi S, Ikeda F, Kawasaki M, et al. Specific recognition of linear ubiquitin chains by NEMO is important for NF-kappaB activation. *Cell.* 2009;136(6):1098–1109.
- [28] Hartmann MD, Mendler CT, Bassler J, et al. alpha/beta coiled coils. *eLife.* 2016;5:e11861.
- [29] Dong H, Hartgerink JD. Role of hydrophobic clusters in the stability of alpha-helical coiled coils and their conversion to amyloid-like beta-sheets. *Biomacromolecules.* 2007;8(2):617–623.
- [30] Huang S, Jj L, Yang S, et al. Neuronal expression of TATA box-binding protein containing expanded polyglutamine in knock-in mice reduces chaperone protein response by impairing the function of nuclear factor-Y transcription factor. *Brain.* 2011;134(7):1943–1958.
- [31] Purrello M, Di Pietro C, Mirabile E, et al. Physical mapping at 6q27 of the locus for the TATA box-binding protein, the DNA-binding subunit of TFIID and a component of SL1 and TFIIB, strongly suggests that it is single copy in the human genome. *Genomics.* 1994;22(1):94–100.
- [32] Lescure A, Lutz Y, Eberhard D, et al. 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(5):1166–1175.
- [33] Reid SJ, Rees MI, van Roon-Mom WM, et al. Molecular investigation of TBP allele length:: a SCA17 cellular model and population study. *Neurobiol Dis.* 2003;13(1):37–45.
- [34] Lancaster AK, Nutter-Upham A, Lindquist S, et al. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics.* 2014;30(17):2501–2502.
- [35] Zhao X, Herr W. A regulated two-step mechanism of TBP binding to DNA: a solvent-exposed surface of TBP inhibits TATA box recognition. *Cell.* 2002;108(5):615–627.
- [36] Patel AB, Louder RK, Greber BJ, et al. Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science.* 2018;362(6421):6421.
- [37] Takahashi J, Fukuda T, Tanaka J, et al. Neuronal intranuclear hyaline inclusion disease with polyglutamine-immunoreactive inclusions. *Acta Neuropathol.* 2000;99(5):589–594.
- [38] Uchihara T, Fujigasaki H, Koyano S, et al. Non-expanded polyglutamine proteins in intranuclear inclusions of hereditary ataxias—triple-labeling immunofluorescence study. *Acta Neuropathol.* 2001;102(2):149–152.
- [39] Perez MK, Paulson HL, Pendse SJ, et al. Recruitment and the role of nuclear localization in polyglutamine-mediated aggregation. *J Cell Biol.* 1998;143(6):1457–1470.
- [40] van Roon-Mom WM, Reid SJ, Jones AL, et al. Insoluble TATA-binding protein accumulation in Huntington's disease cortex. *Mol Brain Res.* 2002;109(1–2):1–10.
- [41] Reid SJ, van Roon-Mom WM, Wood PC, et al. TBP, a polyglutamine tract containing protein, accumulates in Alzheimer's disease. *Mol Brain Res.* 2004;125(1–2):120–128.
- [42] Nakamura K, Jeong S-Y, Uchihara T, et al. SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Hum Mol Genet.* 2001;10(14):1441–1448.
- [43] van Roon-Mom W, Reid S, Faull R, et al. TATA-binding protein in neurodegenerative disease. *Neuroscience.* 2005;133(4):863–872.
- [44] Nolte D, Sobanski E, Wissen A, et al. Spinocerebellar ataxia type 17 associated with an expansion of 42 glutamine residues in TATA-box binding protein gene. *J Neurol Neurosurg.* 2010;81(12):1396–1399.
- [45] Sabate R, Rousseau F, Schymkowitz J, et al. What makes a protein sequence a prion? *PLoS Comput Biol.* 2015;11(1):e1004013.
- [46] Chan HM, La Thangue NB. p300/CBP proteins: hATs for transcriptional bridges and scaffolds. *J Cell Sci.* 2001;114(13):2363–2373.
- [47] Mantamadiotis T, Lemberger T, Bleckmann SC, et al. Disruption of CREB function in brain leads to neurodegeneration. *Nat Genet.* 2002;31(1):47–54.
- [48] Giordano A, Avantaggiati ML. p300 and CBP: partners for life and death. *J Cell Physiol.* 1999;181(2):218–230.
- [49] Heery DM, Kalkhoven E, Hoare S, et al. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature.* 1997;387(6634):733–736.

- [50] Plevin MJ, Mills MM, Ikura M. The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *Trends Biochem Sci.* **2005**;30(2):66–69.
- [51] Sheppard HM, Harries JC, Hussain S, et al. Analysis of the steroid receptor coactivator 1 (SRC1)-CREB binding protein interaction interface and its importance for the function of SRC1. *Mol Cell Biol.* **2001**;21(1):39–50.
- [52] Nucifora FC, Sasaki M, Peters MF, et al. Interference by huntingtin and atrophin-1 with cbp-mediated transcription leading to cellular toxicity. *Science.* **2001**;291(5512):2423–2428.
- [53] McCampbell A, Taylor JP, Taye AA, et al. CREB-binding protein sequestration by expanded polyglutamine. *Hum Mol Genet.* **2000**;9(14):2197–2202.
- [54] Steffan JS, Bodai L, Pallos J, et al. Histone deacetylase inhibitors arrest polyglutamine-dependent neurodegeneration in *Drosophila*. *Nature.* **2001**;413(6857):739–743.
- [55] Rouaux C, J-p L, Boutillier A-L. Targeting CREB-binding protein (CBP) loss of function as a therapeutic strategy in neurological disorders. *Biochem Pharmacol.* **2004**;68(6):1157–1164.
- [56] McCampbell A, Taye AA, Whitty L, et al. Histone deacetylase inhibitors reduce polyglutamine toxicity. *Proc Natl Acad Sci USA* **2001**;98(26):15179–15184.
- [57] Klockgether T, Mariotti C, Paulson HL. Spinocerebellar ataxia. *Nat Rev Dis Primers.* **2019**;5(1):1–21.
- [58] Tsai -C-C, Kao H-Y, Mizutani A, et al. Ataxin 1, a SCA1 neurodegenerative disorder protein, is functionally linked to the silencing mediator of retinoid and thyroid hormone receptors. *Proc Natl Acad Sci USA.* **2004**;101(12):4047–4052.
- [59] Orr HT, Zoghbi HY. Trinucleotide repeat disorders. *Annu Rev Neurosci.* **2007**;30(1):575–621.
- [60] Rocha S, Vieira J, Vázquez N, et al. ATXN1 N-terminal region explains the binding differences of wild-type and expanded forms. *BMC Med Genomics.* **2019**;12(1):1–14.
- [61] Irwin S, Vandelft M, Pinchev D, et al. RNA association and nucleocytoplasmic shuttling by ataxin-1. *J Cell Sci.* **2005**;118(1):233–242.
- [62] Gennarino VA, Singh RK, White JJ, et al. Pumilio1 haploinsufficiency leads to SCA1-like neurodegeneration by increasing wild-type Ataxin1 levels. *Cell.* **2015**;160(6):1087–1098.
- [63] Nitschke L, Tewari A, Coffin SL, et al. miR760 regulates ATXN1 levels via interaction with its 5' untranslated region. *Genes Dev.* **2020**;34(17–18):1147–1160.
- [64] Park J, Al-Ramahi I, Tan Q, et al. RAS-MAPK-MSK1 pathway modulates ataxin 1 protein levels and toxicity in SCA1. *Nature.* **2013**;498(7454):325–331.
- [65] Friedrich J, Kordasiewicz HB, O'Callaghan B, et al. Antisense oligonucleotide-mediated ataxin-1 reduction prolongs survival in SCA1 mice and reveals disease-associated transcriptome profiles. *JCI Insight.* **2018**;3(21):21.
- [66] Keiser MS, Boudreau RL, Davidson BL. Broad therapeutic benefit after RNAi expression vector delivery to deep cerebellar nuclei: implications for spinocerebellar ataxia type 1 therapy. *Mol Ther.* **2014**;22(3):588–595.
- [67] Suh J, Romano DM, Nitschke L, et al. Loss of ataxin-1 potentiates Alzheimer's pathogenesis by elevating cerebral BACE1 transcription. *Cell.* **2019**;178(5):1159–1175. e17.
- [68] Lee Y, Samaco RC, Gatchel JR, et al. miR-19, miR-101 and miR-130 co-regulate ATXN1 levels to potentially modulate SCA1 pathogenesis. *Nat Neurosci.* **2008**;11(10):1137–1139.
- [69] Petrakis S, Raskó T, Russ J, et al. Identification of human proteins that modify misfolding and proteotoxicity of pathogenic ataxin-1. *PLoS Genet.* **2012**;8(8):e1002897.
- [70] Cuvertino S, Hartill V, Colyer A, et al. A restricted spectrum of missense KMT2D variants cause a multiple malformations disorder distinct from Kabuki syndrome. *Genet Med.* **2020**;22(5):867–877.
- [71] Prives C, Lowe SW. Mutant p53 and chromatin regulation. *Nature.* **2015**;525(7568):199–200.
- [72] Zhang J, Dominguez-Sola D, Hussein S, et al. Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. *Nat Med.* **2015**;21(10):1190.
- [73] Lee J-E, Wang C, Xu S, et al. H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *eLife.* **2013**;2:e01503.
- [74] Boniel S, Szymańska K, Śmigiel R, et al. Kabuki syndrome—clinical review with molecular aspects. *Genes (Basel).* **2021**;12(4):468.
- [75] Ortega-Molina A, Boss IW, Canela A, et al. The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nat Med.* **2015**;21(10):1199.
- [76] Cheon CK, Sohn YB, Ko JM, et al. Identification of KMT2D and KDM6A mutations by exome sequencing in Korean patients with Kabuki syndrome. *J Hum Genet.* **2014**;59(6):321–325.
- [77] Ng SB, Bigham AW, Buckingham KJ, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet.* **2010**;42(9):790–793.
- [78] Zaidi S, Choi M, Wakimoto H, et al. De novo mutations in histone-modifying genes in congenital heart disease. *Nature.* **2013**;498(7453):220–223.
- [79] Hannibal MC, Buckingham KJ, Ng SB, et al. Spectrum of MLL2 (ALR) mutations in 110 cases of Kabuki syndrome. *Am J Med Genet Part A.* **2011**;155(7):1511–1516.
- [80] Fisher SE. A molecular genetic perspective on speech and language. In G. Hickok & S. Small (Eds.), *Neurobiology of language*: Elsevier. Amsterdam. **2016**. p. 13–24.
- [81] Marcus GF, Fisher SE. FOXP2 in focus: what can genes tell us about speech and language? *Trends Cogn Sci.* **2003**;7(6):257–262.
- [82] Fisher SE, Scharff C. FOXP2 as a molecular window into speech and language. *Trends Genet.* **2009**;25(4):166–177.
- [83] Tsui D, Vessey JP, Tomita H, et al. FoxP2 regulates neurogenesis during embryonic cortical development. *J Neurosci.* **2013**;33(1):244–258.
- [84] Chiu YC, Li MY, Liu YH, et al. Foxp2 regulates neuronal differentiation and neuronal subtype specification. *Dev Neurobiol.* **2014**;74(7):723–738.
- [85] Clovis YM, Enard W, Marinaro F, et al. Convergent repression of Foxp2 3'UTR by miR-9 and miR-132 in embryonic mouse neocortex: implications for radial migration of neurons. *Development.* **2012**;139(18):3332–3342.
- [86] Mukamel Z, Konopka G, Wexler E, et al. Regulation of MET by FOXP2, genes implicated in higher cognitive

- dysfunction and autism risk. *J Neurosci.* **2011**;31(32):11437–11442.
- [87] Lai CS, Fisher SE, Hurst JA, et al. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature.* **2001**;413(6855):519–523.
- [88] Lieberman P. FOXP2 and human cognition. *Cell.* **2009**;137(5):800–802.
- [89] Herrero MJ, Gitton Y. The untold stories of the speech gene, the FOXP2 cancer gene. *Genes Cancer.* **2018**;9(1–2):11.
- [90] Li S, Weidenfeld J, Morrisey EE. Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. *Mol Cell Biol.* **2004**;24(2):809–822.
- [91] Häußermann K, Young G, Kukura P, et al. Dissecting FOXP2 oligomerization and DNA binding. *Angew Chem.* **2019**;131(23):7744–7749.
- [92] Stroud JC, Wu Y, Bates DL, et al. Structure of the forkhead domain of FOXP2 bound to DNA. *Structure.* **2006**;14(1):159–166.
- [93] Mizutani A, Matsuzaki A, Momoi MY, et al. Intracellular distribution of a speech/language disorder associated FOXP2 mutant. *Biochem Biophys Res Commun.* **2007**;353(4):869–874.